Weidong Huang · Leila Alem
Mark A. Livingston *Editors*

Human Factors in Augmented Reality Environments

Springer

Human Factors in Augmented Reality Environments

Weidong Huang • Leila Alem • Mark A. Livingston
Editors

Human Factors in Augmented Reality Environments

Springer

Editors
Weidong Huang
CSIRO ICT Centre
Marsfield, NSW, Australia

Leila Alem
CSIRO ICT Centre
Marsfield, NSW, Australia

Mark A. Livingston
U.S. Naval Research Laboratory
Washington, DC, USA

ISBN 978-1-4614-4204-2 ISBN 978-1-4614-4205-9 (eBook)
DOI 10.1007/978-1-4614-4205-9
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012946725

© Springer Science+Business Media New York 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Advances in hardware and networking have made possible a wide use of augmented reality (AR) technologies. Many innovative technologies have been invented to support AR, and many AR systems have been built to facilitate engaging user experience. Despite the attempts being made by researchers and engineers, the truth is that most of those technologies and systems remain in laboratory settings. Simply putting those hardware and technologies together does not make a "good" system for end users to use. New design principles and evaluation methods specific to this emerging area are urgently needed to keep up with the advance in technologies.

Human Factors in Augmented Reality Environments is the first book on human factors in AR, whose contributors include well-established researchers worldwide from diverse disciplines. This book is designed to systematically address the issues related to design, development, evaluation and application of AR systems. Topics include surveys, case studies, evaluation methods and metrics, HCI theories and design principles, human factors and lessons learned and experience obtained from developing, deploying or evaluating AR systems.

More specifically, this book includes 11 chapters which are broadly categorized into the following four parts:

1. Overview
2. Perception and cognition
3. Design principles and recommendations
4. User experience

Readers are encouraged to read individual chapters in each part to retrieve insights into the current state-of-the-art research and to explore further research questions in this area. It is our hope that while AR technology is rapidly progressing, equally enthusiastic efforts can be devoted to research on how the technology should be adapted for human everyday use and how our understanding of human factors should be applied in AR.

We wish to thank all the authors who submitted their valuable work for consideration. Without their effort and contribution, this book would not have come to fruition. Our sincere thanks and appreciation are also extended to Springer editors Susan Lagerstrom-Fife, Courtney Clark and Jennifer Maurer and staff members of the production team for their support throughout this project.

Marsfield, NSW, Australia	Weidong Huang
Marsfield, NSW, Australia	Leila Alem
Washington, DC, USA	Mark A. Livingston

Contents

Part I Overview

Part II Perception and Cognition

Part III Design Principles and Recommendations

Part IV User Experience

Contributors

Jane H. Barrow George Mason University, Fairfax, VA, USA

Christophe Bortolaso University of Toulouse, IRIT – Elipse, Toulouse Cedex 9, France

Syrine Charfi Ecole Nationale d'Ingénieurs de Tunis, Tunis, Tunisia

Elizabeth F. Churchill Yahoo! Research, Santa Clara, CA, USA

Arindam Dey University of South Australia, Adelaide, SA, Australia

Emmanuel Dubois University of Toulouse IRIT – Elipse, Toulouse Cedex 9, France

Henry Been-Lirn Duh Department of Electrical and Computer Engineering, National University of Singapore, Singapore

Joseph L. Gabbard Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Subhashini Ganapathy Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, Dayton, OH, USA

Yolande Kolstee AR Lab, Royal Academy of Art, The Hague, The Netherlands

Nai Li Department of Communications and New Media, National University of Singapore, Singapore

Mark A. Livingston U.S. Naval Research Laboratory, Washington, DC, USA

Nicholas Mariette University of New South Wales, Sydney, NSW, Australia

Thomas Olsson Unit of Human-Centered Technology, Tampere University of Technology, Finland

Marco de Sá Yahoo! Research, Santa Clara, CA, USA

Christian Sandor University of South Australia, Adelaide, SA, Australia

Dominique L. Scapin National Research Institute in Computer Science and Control (INRIA), Domaine de Voluceau – Rocquencourt, Le Chesnay Cedex, France

Ciara M. Sibley U.S. Naval Research Laboratory, Washington, DC, USA

J. Edward Swan II Mississippi State University, Starkville, MS, USA

Bruce H. Thomas University of South Australia, Adelaide, SA, Australia

Pier Paolo Valentini Department of Industrial Engineering, University of Rome “Tor Vergata”, Rome, Italy

Part I
Overview

Chapter 1
Issues in Human Factors Evaluations of Augmented Reality Systems

Mark A. Livingston

1 Importance of Human Factors to Augmented Reality

In his widely-accepted definition of augmented reality (AR), Azuma [1] cited three components of AR applications. First is the combination of real and virtual imagery. Second is the *registration* (alignment) of computer graphics with objects or locations in the real 3D environment. Third, and perhaps of primary concern in this volume, is that AR systems must be interactive in real time. As one focuses on this third component, the importance of human factors for AR systems may be considered an obvious and critical component of research in AR. One might think that the human factors of AR systems would thus be heavily studied; however, as this volume shows, there is a paucity of investigations regarding human factors in AR systems compared to the enabling technologies.

There may be good reasons for this. The technology, notably in tracking of the user's viewpoint (often the primary contributor to the success or failure of registration between real and virtual objects), has yet to meet the minimum requirements for success in many applications. This is likely due to one or both of two key factors. First, the tracking problem, described extensively in the literature [18], has no simple solution that can solve all the varied tracking instances. However, there are many good technologies (which are still being improved). This helps give rise to the second key factor in meeting the minimum tracking requirements for application success: a relative lack of exploration of what those minimum requirements are. While there are some examples of such human factors evaluations [9, 12, 13], each application and perhaps each tracking technology, with varied performance capabilities in diverse measures such as (static) accuracy, noise, and—most critically [7]—latency

---

M.A. Livingston (✉)
U.S. Naval Research Laboratory, Washington, DC 20375, USA
e-mail: mark.livingston@nrl.navy.mil

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_1,

or dynamic accuracy, engenders a further need for investigation under new experimental conditions.

On top of that complexity, the AR researcher may also pause to consider that tracking is only one critical component of an AR system. There are numerous choices for display technologies, ranging from head-worn to hand-held to projection systems and encompassing optical and video combiners for real and synthetic imagery. As will be seen in this volume, such displays present their own human factors challenges, many of which are only beginning to be understood. Another point emphasized in this volume is that the visual sense is not the only way in which merging of real and virtual may occur in an AR system; the definition given above purposefully avoided excluding audio, haptic, olfactory, and taste displays. Although the last two of these are rarely employed in either virtual environments or AR, it should be obvious that if such displays are to be usable, then their own human factors must be understood. This will enable solutions for these issues that meet users' requirements for the desired applications. Similarly, user interfaces, information presentation methods, and calibration techniques are all candidates for human factors evaluations.

2 Applications and Tasks for Evaluation

Sharp readers hopefully noted that the above discussion ended twice with the note that the application itself was a factor in the experimental conditions of human factors evaluations. This raises another point in the understanding of why relatively few human factors evaluations have been conducted for AR, and even offers a reason why many of the evaluations that have been conducted examined perceptual factors. Despite multiple cycles of media and industry attention, there has never really been a widely-accepted "killer application" for AR. Perhaps the emergence of mobile phones equipped with not only auto-stereo displays but also sufficient computational power to perform visual tracking will lead to location-based information and navigation being delivered in an AR application. One could argue that without a demonstration of success in an evaluation of an AR application, that claims of AR being the right technology to apply to a particular task are unsubstantiated.

But there have been technical successes for AR applications and recognition of the need for human factors to be incorporated into the development cycle. Literature on demonstrated applications gives several case studies that showed the potential for AR to benefit users in a diverse set of applications. The first AR application developed after the introduction of the technology in computer science circles [15, 16] was for fighter pilots. The "Super Cockpit" [4, 5] foreshadowed modern head-up displays still used in fighter jets and available in some passenger cars. As the key feature of the application was providing spatial awareness for the pilot, human factors were recognized as a critical element of the system development in the early work. The applications that helped revive the academic interest in AR also found human factors work critical to the success or lack thereof. The well-known Boeing project to use AR to guide manufacturing employees through the assembly of a wiring harness was a technical improvement, but low social acceptance doomed the project's overall success [2]. Medical AR applications are perhaps among the most ambitious projects in terms of the registration requirements. The Ultrasound AR project at the University of North Carolina at Chapel Hill [14] relied heavily on iterative evaluation by an expert (physician) who provided feedback on how the perception of the ultrasound data matched her expert knowledge of anatomical structure generally and her patients' pre-operative scans specifically. The neurosurgery AR project GE and Brigham and Women's Hospital [11] also heavily involved physicians in the planning and evaluation of the system. The Touring Machine [3] squarely addressed the human factors associated with the interface to information that could be made available at any location, inspiring the applications being developed today for mobile phones in urban environments. Users of these initial applications can now receive navigation information and could soon be able to view signposts placed by friends indicating where to meet or recommendations for what to order at a restaurant.

3 Measuring Perceptual and Cognitive Performance

Implicit in the foregoing discussion are two types of measurements. Low-level perceptual measures are perhaps most useful for evaluating individual components; they tend to indicate the quality of these hardware components. Perceptual measures perhaps make it easier to isolate individual system components; this may reflect overconfidence of our knowledge of the processing in the mind, but it seems to have served the field well thus far. Thus, it is easier to attribute poor performance on perceptual measures to unsuitability of the hardware or software providing perceptual cues rather than a misunderstanding or inappropriate response from the user. As will be seen in later chapters, different presentation techniques can vary widely in perceptual quality. Reasons for such differences ought to be explored to determine perceptual benefits of the various techniques.

Perceptual measures do have some advantages for acquiring data. They generally require tasks that are much shorter, since they rely more on basic reactions than thought (i.e. cognition). Also, they generally don't require specialized knowledge of the field of expertise in which the final application will occur. Thus the subject pool can be open to anyone, with restriction only on those whose perceptions are outside the bounds of "normal" human performance. (For example, color-blind subjects may not be eligible for studies of color's utility in visual presentation, and stereo-blind users may not be eligible for studies of many head-worn AR displays.) Short tasks lend themselves to numerous repetitions in a statistical evaluation, assisting the experimenter reach statistical validity. Since these tasks can be disconnected from specific applications, they are often applicable to multiple scenarios. One can argue that these tests are therefore much more useful for evaluating hardware suitable for inclusion in an application [10]. There will generally be clear or obvious influences on perceptions from background conditions. Background lighting can heavily influence visual perception, and analogous background noise can affect hearing, touch, and the other senses. Workload and attention can also affect perceptions; such confounding factors can be hard to identify and isolate in experimental design for both perceptual and cognitive tasks.

Measures of performance on cognitive tasks are generally more indicative of how users fare with the system. Ideally, careful and detailed study of perceptual factors first lead to optimal performance or configuration of component technologies, and then more cognitively-demanding tasks are used for evaluation of (complex pieces of) the system. Since it is all but impossible to separate cognitively-demanding tasks from their perceptual underpinnings, results from cognitive measures can sometimes be harder to interpret with regard to the detailed parameters of the system components; they must be evaluated in concert with perceptual factors to determine what changes would be most effective in improving users' performance with the overall system. The greatest advantage of cognitive tasks is that they are much more likely to be similar to the task users will perform in the final application; therefore, the information gained in a cognitive evaluation is likely to be much more relevant and precise about the utility of the application or its features. Of course, this will likely mean that the results are less broadly applicable to other cognitive tasks.

Designing a cognitive task for evaluation should come naturally from the application, whereas designing general low-level perceptual tasks that apply to a large class of AR applications might not be so easy [8]. Getting useful results with cognitive evaluations usually requires having subjects who are reasonably precise representatives of the intended end-user population. So a medical AR system must be evaluated by physicians (or physicians-in-training). Of course, medical applications also must be evaluated safely, which usually means medical phantoms rather than live patients in initial evaluations. Even consumer applications (e.g. navigation assistance from mobile phones) should be evaluated by people who are likely users (e.g. people who rely heavily on mobile phone applications). One could also do an evaluation with subjects who do not use mobile phone applications, but the results may indicate less about the details of the application than the willingness of the participants to become mobile phone users for the proposed application. Of course, this could also be valuable knowledge for a market survey, so the lesson is not to avoid such evaluations, but simply to be aware of what you are really evaluating and make sure that the data justify the results you claim.

From this dichotomy, one can begin to understand that there are many different types of evaluations, each of which can have different value and products that may be applied at various times in the development cycle; these can be applied to AR systems development [6]. *Expert evaluation* is a heuristic evaluation performed by experts in usability; often the user interface developers can perform this evaluation for themselves (after some practice being objective). One may also consult design guidelines that apply to the AR system being developed. While the cautions above regarding the applicability of data gathered on different tasks should be particularly noted when applying general design guidelines to systems with the novelty AR interfaces have (compared to most human–computer interfaces), this can be a valuable source of expert advice. Perceptual evaluations most naturally fit into the category of *user-based statistical evaluation*. The tasks in this type of evaluation are typically perceptual and/or low-level cognitive tasks, and the study usually focuses on a specific aspect of the application. In this category, the design of the task determines how generalizable the results are beyond the experimental scenario. As noted above, general perceptual tasks that test the usability of hardware may be more applicable. But again a cautionary tone is in order: if the effect of a feature rises to a cognitive level (e.g. the effect of latency in a tracking device), the dependence on the task will rise as well. *Formative evaluation* requires representative users to perform tasks that end-users will do with the application; this language virtually repeats the description of cognitive evaluations above, and task error rates and performance time are common measures. One may also evaluate how the AR application fits in the wider context of a job, affecting user fatigue, workflow outside of the AR application, or safety procedures. Thus both qualitative and quantitative measures are useful in this class of evaluations. Finally, when one is ready to test an AR application against a traditional method of performing a task, then one conducts a *summative evaluation*. This is the most costly type of evaluation, since the number of users (and their expertise) must be higher.

Research Hint #1:

Advanced readers should consider these evaluation types when reading this volume and consider what types of evaluations have been done and could be done in each area.

4 Interpreting the Results

The above discussion raised a few issues regarding the interpretation of the results of the various forms of human factors evaluations. One additional issue that is often lacking but has been offered is the relating of results to perceptual and cognitive theories. Work in depth perception with head-worn AR displays has been conducted this way for several years [17], and the remaining work in human factors of AR would be well-served by this approach to interpretation of results. One could easily see extending this to using perceptual and cognitive theories for designing experimental tasks and assisting with interface design. This is done implicitly far too often; it would enhance the understanding and theoretical grounding of the work if bases for experiments and hypotheses and the resulting insights were stated explicitly.

Research Hint #2:

Advanced readers may wish to consider the theoretical grounding and interpretation when reading this volume. Consideration of perceptual and cognitive theories that relate to the work described should help push the work to new insights and advances.

5 Summary

Thus we begin our discussion of human factors in AR. We have argued in this introduction that the perceptual and cognitive theories, tasks, and performance measures that have been applied or could be applied to AR should be fundamental to research in AR, but noted that this is not an argument that has permeated the AR community. This may be evidenced in the relative rarity of papers in premiere conferences and journals that conduct formal or even informal studies on the human factors associated with the AR system or its components. Perhaps the chapters in this volume will spark more efforts in this critical research.

Acknowledgements The author would like to thank LT Gregory O. Gibson for assistance in the preparation of this chapter.

References

1. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators and Virtual Environments **6**(4), 355–385 (1997)
2. Curtis, D., Mizell, D., Gruenbaum, P., Janin, A.: Several devils in the details: Making an AR application work in the airplane factory. In: Augmented Reality: Placing Artificial Object in Real Scenes (Proceedings of International Workshop on Augmented Reality), pp. 47–60 (1998)
3. Feiner, S., MacIntyre, B., Höllerer, T., Webster, A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In: International Symposium on Wearable Computing (ISWC), pp. 74–81 (1997)
4. Furness, LT T.A.: The application of head-mounted displays to airborne reconnaissance and weapon delivery. Tech. Rep. TR-69-241, U.S. Air Force Avionics Laboratory, Wright-Patterson AFB (1969)
5. Furness, T.A.: The super cockpit and its human factors challenges. In: Proceedings of the Human Factors Society 30th Annual Meeting, pp. 48–52. Human Factors and Ergonomics Society (1986)
6. Hix, D., Gabbard, J.L., Swan II, J.E., Livingston, M.A., Hollerer, T.H., Julier, S.J., Baillot, Y., Brown, D.: A cost-effective usability evaluation progression for novel interactive systems. In: Hawaii International Conference on System Sciences (HICSS-37) (2004)
7. Holloway, R.L.: Registration error analysis for augmented reality. Presence: Teleoperators and Virtual Environments **6**(4) (1997)
8. Livingston, M.A.: Evaluating human factors in augmented reality systems. IEEE Computer Graphics and Applications **25**(6), 12–15 (2005)
9. Livingston, M.A., Ai, Z.: The effect of registration error on tracking distant augmented objects. In: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008), pp. 77–86 (2008)
10. Livingston, M.A., Swan II, J.E., Julier, S.J., Baillot, Y., Brown, D., Rosenblum, L.J., Gabbard, J.L., Hollerer, T.H., Hix, D.: Evaluating system capabilities and user performance in the battlefield augmented reality system. In: Performance Metrics for Intelligent Systems Workshop (PerMIS'04) (2004)
11. Lorensen, W., Cline, H., Nafis, C., Kikinis, R., Altobelli, D., Gleason, L.: Enhancing reality in the operating room. In: Proceedings of IEEE Visualization, pp. 410–415, CP46 (1993)

12. Robertson, C.M., MacIntyre, B.: An evaluation of graphical context as a means for ameliorating the effects of registration error. In: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007), pp. 99–108 (2007)
13. Robertson, C.M., MacIntyre, B., Walker, B.N.: An evaluation of graphical context when the graphics are outside of the task area. In: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008), pp. 73–76 (2008)
14. State, A., Livingston, M.A., Hirota, G., Garrett, W.F., Whitton, M.C., MD, E.D.P., Fuchs, H.: Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In: SIGGRAPH 96 Conference Proceedings, Annual Conference Series, pp. 439–446. ACM SIGGRAPH, Addison Wesley (1996)
15. Sutherland, I.E.: The ultimate display. In: Proceedings of the IFIP Conference, pp. 506–508 (1965)
16. Sutherland, I.E.: A head-mounted three-dimensional display. In: AFIPS'68: Fall Joint Computer Conference, pp. 757–764 (1968)
17. Swan II, J.E., Livingston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., Hix, D.: A perceptual matching technique for depth judgments in optical, see-through augmented reality. In: Proceedings of IEEE Virtual Reality, pp. 19–26,317 (2006)
18. Welch, G., Foxlin, E.: Motion tracking: No silver bullet, but a respectable arsenal. IEEE Computer Graphics and Applications **22**(6), 24–38 (2002)

Chapter 2
Human Factors Research in Audio Augmented Reality

Nicholas Mariette

1 Introduction

To discuss research into human factors in audio augmented reality, it is helpful to specify the context and motivation for such investigations. Elucidating this scope requires some discretionary definitions and limitations due to the broad and finely-variable nature of the medium. We will start with the fundamentals.

AR itself is a medium for which it can be problematic to make a comprehensive definition. One widely-cited publication by Azuma [1] defines generic AR as having the following three characteristics:

1. Combines real and virtual.
2. Interactive in real time.
3. Registered in 3D.

A looser definition by Mallem [33] defines AR as enabling "spatial and temporal virtual and real worlds [to] co-exist, which aims to enhance user perception in his real environment." While the former definition might be considered the canonical one, the latter might better describe many recent instances of technology that have come to be known as AR.

A significant current trend in technology and for AR applications is the rapid development and adoption of mobile computing devices such as current generation smart phones that are capable of bringing AR "apps" to mass-markets. Aspects of mobile AR including design, cognition and user experience are discussed further in this book in Chaps. 5–7 and 9. It is these kinds of accessible AR instances, such as the common format of visual augmentations to live video from smart-phone cameras that are simultaneously bringing AR to larger audiences, as well as stretching the bounds of Azuma's three characteristics of AR. In particular, the "real" is often substantially

---

N. Mariette (✉)
University of New South Wales, Sydney, NSW, Australia
e-mail: n.mariette@unswalumni.com

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_2,

mediated (e.g. by the use of live video), while 3D registration might also be of low quality, or somewhat closer to 2D registration. As such, these new applications often conform better to Mallem's broader definition of AR. The same trend is occurring to the characteristics of *audio AR* technology, with many interesting and exciting new developments happening in the smart-phone application market.

As a curious aside, the broadest definition of AR is manifest in the literal meaning of the words, augmented reality. In other words, AR is any augmentation or enhancement to plain old, unadorned reality. Interestingly, a broader definition such as Mallem's that satisfactorily encompasses the present evolution of AR also embraces *older* technology not previously identified as AR, yet comfortably described by the literal phrase. For instance, the visual use of static or moving images projected onto unconventional surfaces, or the use of mobile transistor radios or early portable tape players with headphones could both be understood as presenting an augmented reality. Equally, these examples of realities augmented by technologically-delivered media both in some way enhance the perception of the person apprehending them.

With such broad possible meanings of the term AR, it is clearly difficult to commence a focussed discussion of human factors of these media. Nevertheless, it is clear that human factors are important to discuss. If any given augmentation of reality is to be understood as enhancing a user's perception within their real environment, this logically invites questions as to *how well* their perception is enhanced, in *what manner* it occurs, and *why* and *how* is it useful.

This chapter explores the study of human factors in audio augmented reality by first defining the media that exist under this name. This scope is limited somewhat to those kinds of audio AR that have significant human factors and consequential performance effects. We briefly look at human factors research in AR in general, then we survey how human factors have been investigated in research on audio AR, and what these studies might be able to contribute to the further development of the medium.

2 What Defines *Audio* Augmented Reality?

As is the case for general AR, a definition of the medium of *audio* augmented reality is both problematic, and necessary for a survey of human factors research.

2.1 A Broad Definition

In the most general sense, audio AR is simply the introduction of artificial sound material into the real world. In discussing telepresence applications, Cohen et al. [12] noted:

> One common example of augmented audio reality is sound reinforcement, as in a public address system.

Novo [43, p. 291] made a similar observation:

> The idea of mixing real and virtual environments, “joint reality” or “augmented reality,” was probably first realized with assisted reverberation systems.

This observation was also made by Härmä [16, pp. 795–823], who considered voice-only teleconference systems dating back to the 1950s as AR systems.

Note that this general concept, nevertheless has some specificity to it, given that the application of assisted reverberation (and others) considers the augmentation of the *aural* reality (the real sound sources and room acoustics). This suggests that augmented reality in an aural context means augmentation of the *aural* reality—i.e. that the sensory mode of the augmentation matches that of the augmented, however this may not always be the case. Next, we consider the range of characteristics that define the different types of audio AR, of which the above is only one example.

2.1.1 Predominant Sensory Modality

The modality of the reality that is being augmented with virtual audio—aural, visual, haptic, olfactory, gustatory, or any combination of these. This accounts for the possibility of cross-modal augmentation, the most common of which would be virtual audio augmenting the visual reality. Audio augmentation of the haptic reality has also been investigated. In cross-modal situations it is necessary for the link between stimuli to be clearly referenced to the same object.

2.1.2 Spatial Characteristic

Virtual audio may be processed as a mono, stereo, two-dimensional (2D) or three-dimensional (3D) audio signal. In fact, this is a somewhat artificial distinction, given that there is a complex connection between the creation of a signal—a *sound event*—and a listener’s auditory perception of the signal—the resulting *auditory event* [7, p. 2]. A sound event with one spatial characteristic might be perceived as an auditory event with a different spatial characteristic, depending on the circumstances. For instance, a mono signal presented to only one ear could be perceived as a near-field, 3D positional sound source.

For the purposes of this discussion, we shall consider only the situation where the spatial characteristic of the perceived auditory event matches the intention of the processing that generated the sound event. The nature of the auditory event is an important human factor in spatial audio reproduction. In general, mono and stereo virtual audio are perceived on the left–right axis through the centre of the head, while 2D and 3D virtual audio are perceived to be in a given direction and distance no closer than the perimeter of the head.

2.1.3 Presentation Means

This describes whether the virtual audio is rendered specifically for each individual user, or once for a group audience. Audio AR has been conceived both for individual users (generally presented over headphones), or for a collective group audience (generally presented via a loudspeaker array e.g. [57, 58]). This category of classification partly aligns with the classification by Lindeman and Noma [26] based on the "location of mixing," which can be in the environment (via speakers, or using passive, acoustic mixing through headphones), in the sensory subsystem (e.g. using bonephones or implant technology), or in a computer system (e.g. microphone feedthrough).

2.1.4 User Tracking Extent or Interactivity Mode

This describes whether or not the user's head-orientation is tracked and used in addition to position, to control rendering. It relates to Azuma's AR characteristic of real-time interaction and 3D registration of the stimuli. Many audio AR applications on smart phones only provide virtual audio in relation to position. This does not preclude 3D audio from being presented—e.g. a spatial sound-field at a given location, without a fixed orientation. However it does preclude *positional* 3D audio—which requires the 3D audio scene orientation to remain fixed relative to the physical world, even while the user turns their head. The availability of the user's head-orientation thus has a great effect on the affordances that may be realised by the medium. This in turn has a great effect on the tasks, usability, and higher-level applications that are possible. For this reason, the interactivity mode is important in the discussion of human factors research in audio AR media.

2.1.5 User Mobility

This describes the amount of freedom of the user to move around. It depends on whether the tracking system requires built infrastructure within a local area, or whether it enables free-roaming, as for global positioning system (GPS) receivers. This is an important, yet basic factor, however we have not used this to define the types of audio AR for this discussion. Some literature specifies whether or not a system is mobile, because much research to date has used tethered or limited-area tracking systems (e.g. [59]).

2.2 Taxonomy

Now we use the defining characteristics described above to develop a taxonomy of audio AR media terms from the literature, and to aid the discussion of human factors research. In practice, particular combinations of characteristics are more common than others. This chapter will focus on the most common audio AR media.

2.2.1 Augmented Reality Audio

Augmented Reality Audio (ARA) specifies augmentation of the auditory sensory modality, generally with virtual audio that has 2D or 3D spatial characteristics, individual presentation, and full position and orientation tracking extent. For instance, the medium known as Augmented Reality Audio (ARA; [17]) is distinctly focused on audio augmentations to the *audible* reality. The augmented sensory modality is important because new human factors arise during multi-modal or cross-modal perception.

2.2.2 Audio AR

This term has been used without precise identification of the augmented sensory modality, although augmentations of visual or auditory modes are the most common. Tools going by the term Audio AR have used all spatial characteristics from mono to full 3D, both individual or group presentation, and both extents of user tracking. Media known as Audio AR are usually more general, and often augment the *visible* reality, or have a less critical focus on audio-feedthrough [32, 42].

2.2.3 Augmented Audio Reality

This less common term was first used to describe the augmentation of *audible* reality [11, 12], a meaning that is supported by its literal interpretation, however it has also been used interchangeably with audio AR [24]. For this discussion, we will not include the term Augmented Audio Reality, which has some overlap with the other terms, yet is probably best aligned with the term ARA.

2.2.4 Spatial Audio AR

An important sub-group of Audio AR only uses 2D or 3D spatial characteristics, with full position and orientation tracking, denoted herein as Spatial Audio AR (SAAR).

2.2.5 Personal, Location-Aware Spatial Audio

A further sub-group of SAAR is the medium presented only to individual users, which I have previously termed Personal Location-Aware Spatial Audio (PLASA; [37]). PLASA by definition requires 2D or 3D spatial audio, with position and orientation tracking, and rendering for individual audition to headphones, or other personal transducers such as bonephones [51].

2.2.6 Locative Audio

This term groups a broader category of media that use only positional user tracking, thereby not facilitating 2D or 3D audio with absolute directionality in the world reference frame. This medium does not specify *how* the sound relates to location or how it is processed. As noted earlier, this does not preclude the use of spatial audio, however it will remain *oriented* within the *user's* reference frame, while being *positioned* relative to the absolute, *world* reference frame.

The term derives from "locative media" [14], which refers to electronic media that relate to the user's locational context. Locative media are closely related to the media of pervasive/ubiquitous computing, whereby services are made constantly available, and sensitive to user *contexts* (locational, social, technological and others). Locative audio can exist using relatively unsophisticated technology, since by definition it only requires mono audio and rough user position information.

During the last decade, many projects with artistic or utilitarian applications could be classed as locative audio, or non-spatial audio AR. Fewer projects could be classified as PLASA or SAAR.

3 Early Audio AR Media

This section presents a brief review of applications and evaluations of mobile audio AR.

Audio augmented reality was first proposed no later than 1993 by Cohen et al., who demonstrated a system for a stationary user that employed binaural spatial audio to augment the sight of a real telephone, with the virtual auditory image of a ringing telephone presented over headphones. The system used head-tracking, so the virtual sound was registered to the world reference frame.

This static AR concept has since been extended as mobile AR, by introducing indoors or outdoors body-motion-tracking. Also, even before the completion in early 1994 of the 24 satellite constellation for the Global Positioning System (GPS), Loomis et al. [29] proposed using GPS position tracking in their personal guidance system (PGS) for the visually impaired.

The PGS concept was to present a virtual acoustic display of externalised sound beacons within the traveler's auditory space, with an aim to "allow blind users to travel without assistance over unfamiliar territory." They also hoped to "permit the user to develop better cognitive representations of the environments" [29]. Although the PGS was not conceived as an audio AR system, it is a good example both of PLASA and SAAR.

Cohen [11] also identified GPS as an appropriate positioning technology for audio AR applications, with a design for a GPS Personal Guidance System. He made the visionary statement that "further synergetic effects are obtained by eventual leverage off emerging ubiquity of telecommunication networks (more bandwidth [like high-fidelity audio]) and GIS (geographic information system) databases, accessing terrain or street data." Even present day audio AR systems are yet to fully realise the potential of these synergies.

Over the following decade, several outdoor, GPS-based audio AR projects were implemented as fairly bulky packages, for example, backpack-based systems [18, 19], or roll-around cases [46]. In 2001–2004, the indoors LISTEN project [53] used high resolution, sub-decimetre tracking, and further reduced the worn system to passive tracking signal emitters and headphones, with remote tracking and spatial audio rendering. Mariette [37] reviews several other audio AR systems with sound art and pure research applications. The form factor of a system often has important repercussions on human factors and usage simply due to the affordances associated with a device's size and weight.

With cheap digital compasses, powerful portable computers, lightweight consumer GPS receivers (and soon, Galileo receivers,) affordable, portable outdoors audio AR systems may now be implemented. However, despite this great potential, there has been relatively little investigation of the usability and perceptual performance of many of these systems. Evaluation of audio AR systems has often been limited to basic verification of functionality.

4 Introduction to Human Factors in Audio AR

This section discusses the relationship between affordances of audio AR media, human characteristics and perceptual abilities, technical characteristics of devices and their combined meaning in regards to the human factors of the resulting media ecosystems. This provides a conceptual introduction to the specific survey of research that follows.

4.1 Affordances and Human Factors

An affordance of an object, device or system is something that it is possible to do with that thing. Gibson [15] defined an affordance as "a specific combination of the properties of [an object's] substance and its surfaces taken with reference to an animal." According to Jones [22], Gibson further suggested that an affordance concerns personal and environmental properties taken in reference to each other. Clearly, affordances involve inherent human capability and inherent physical possibilities of an object and its environment.

For example, a *sounding object* (i.e. an object that is radiating acoustic energy), along with the human ability for spatial hearing and the physical acoustics of air and the surrounding environment affords the act of localising the sound and navigating to its origin if so desired.

A human factor is a characteristic of one person or many people in relation to their use of an object/device or system, such as audio AR. Typically, a human factor describes any effect on task performance that is unique to the *interaction* between a human and a given implementation of a device that affords the task in question. Components of the task may then be broken down into fundamental tasks that are afforded by natural human physiology and physics, in a situation *without* a technical device or intervention.

4.2 Background: Human Factors in Visual AR

In a study of human factors in visual AR, Livingston [27] performed a *domain analysis* (from software engineering) to identify tasks performed using an AR system at various levels of human functionality such as perceptual, cognitive and higher level tasks. Chapter 4 in this book provides an example of this approach towards the pursuit of "X-ray vision" using visual AR. Chapter 5 also discusses the human factors of mobile AR that are related to embodied cognition, such as physical interaction.

Livingston claims that the two most important means of studying AR applications are: to compare them with traditional methods for achieving a given cognitive task (such as navigation by way-finding); and to understand the perceptual bases of the AR interface in order to improve its suitability for given tasks from a more fundamental level. This approach led to a strategy of testing the perception and task-based performance using only the well-designed features of an AR interface. By identifying "canonical" perceptual and cognitive sub-tasks of common higher-level cognitive tasks, techniques can be transferred "between applications and across perceptual mechanisms."

Given examples of canonical tasks are recognition, resolution or description of an object by itself or amongst others, prediction of future positions of an object, navigation to or in-relation-to an object, distance and direction identification, identification of alert signals, and manipulation of the position or orientation of objects. Most or all of these tasks can be performed to some degree via different senses or perceptual mechanisms such as visual, auditory or tactile modes.

After identifying the canonical perceptual tasks, a set of AR system and user performance requirements can be determined for those tasks, and thus the successful use of the system for higher-level tasks. In order to understand performance of perceptual tasks in AR, it is often necessary to compare with performance of an equivalent real-world task. For instance, resolution and identification of objects can be tested in using visual acuity tests such as the standard Snellen eye chart. In audition, the localization test is an analogous example.

If a canonical task is isolated from extraneous factors so as to function as a control condition in an experiment, then it can be analysed in terms of the user and system parameters that effect performance. While the performance of an isolated perceptual task using real stimuli relies only on an individual's ability, in AR, the equivalent perceptual task is mediated by the technology that implements the AR system. Thus, task performance in AR involves both the individual's inherent ability, and some effect of the technological and system parameters of the specific AR implementation.

In order to discuss the research of human factors of audio AR, we need knowledge of the human characteristics that are involved in interactions with an audio AR system. We also need to understand the technical characteristics of a given audio AR system implementation. Human factors then arise during interaction of natural human ability and system characteristics, for all tasks afforded by the human-device-environment "ecosystem."

4.3 Characteristics of Natural Spatial Hearing

Amongst audio AR media, the most sophisticated are those that use 2D or 3D spatial audio for the augmentation of reality in any sensory modality. For sonic stimuli that contain no spatial information (mono sound sources), humans have the fundamental ability to detect individual sound sources within the sonic gestalt. Higher-level abilities then include the ability to comprehend dynamic pitch, timbre, rhythm, loudness and higher-level semantic information from the detected sound sources.

When some spatial information is available in the sonic stimuli (as for stereo, 2D and 3D stimuli), in addition to source detection, humans have the ability to *localize* (determine the direction and distance of) the sound source in space to a varying degree of accuracy, depending on the particular circumstances. The fundamentals of human spatial hearing have been examined in detail in the research literature, and thoroughly reviewed by Blauert [7].

The innate human ability to hear sounds in space has been characterized beginning with the fundamental situation of a single, stationary sound source in *free-field, anechoic* conditions (an unenclosed, un-obstructed environment). Further studies have examined more complex situations involving multiple sound sources, source and/or listener motion, and reverberant environments. This has established the human ability and limitations involved in achieving particular tasks in relation to *real* sound sources, in ideal, and certain non-ideal situations.

Perception of dynamic, real sources or motion-interactive, virtual sources is affected by a further range of factors that require new ways of measuring perceptual performance in comparison to static spatial audio.

4.4 Human Factors of Synthetic Spatial Audio Presentation

The characterization of human spatial hearing abilities becomes more complicated when they are mediated by technology. Perceptual experiments occur in laboratories with seated, stationary listeners, and spatial stimuli that are either static, or interactive with head-turns only.

Free-field stimuli have been simulated on headphones using detailed measurements of the acoustic response between sound sources and listeners' ear canals. Such filters are known as *head-related transfer functions* (HRTFs) when represented in the frequency domain. These techniques enabled subsequent researchers to implement systems that produced real-time, head-turn-interactive synthetic binaural audio [3, p. 57] that is used in PLASA systems.

The range of factors is wide indeed, encompassing any environmental variable that might alter spatial hearing perception for real or synthetic situations. For example, recently studied factors include: multiple simultaneous source masking [5], stimulus temporal features [8] or spectral features [21], source motion [10], the reverberant acoustic environment [4], multi-modal stimuli (usually visual and aural; [60]), and the numerous limitations of spatial sound synthesis techniques [25, 39].

These factors become highly influential in the human factors of audio AR. New factors often require a new performance measure to quantify the perceptible effects of the stimulus factor in question. The range of performance measures could be organised along an axis from fundamental (such as source azimuth, elevation or distance) to high-level (such as the "presence," "realism" or "stability" of a stimulus). Some measures of performance are described below.

Lateralization describes the kind of localization that occurs when auditory events are heard within the head—e.g. for headphone-delivered synthetic sources generated using only interaural time and intensity differences [20].

Absolute localization relates points in the sound source space to those in the auditory space—i.e. from the physical to the perceptual [6, p. 38]. This ability can be further characterised by localization and localization blur in the horizontal plane, median plane and in distance. Factors noted to influence localization include the source position itself, signal type, familiarity, spectral content, amplitude envelope and duration, previous sonic events, listener head movements, critical listening experience and individual differences. Even the response method used to measure localization is itself a potentially influential factor [44].

Ambiguous localization or the "cone of confusion" is a phenomenon in human spatial hearing whereby the left–right symmetry of the human head and resulting ambiguity in the fundamental cues to localization means that errors can occur as *front–back* or *up–downconfusion*. The main cue to resolve these confusions is the variation of the binaural signals with head-movements. Wightman and Kistler [55] also showed that front–back disambiguation was achievable if intended sound source movements are known by the listener—for instance if they control the source position. This shows that front–back resolution is to some degree a cognitive process.

Externalization is the perceived quality of an auditory event by which the associated sound source seems located at some distance away from the listener's head. The converse perceptual quality is called "inside-the-head locatedness" (IHL; [6], 116–160). Blauert notes that while this occurs often for headphone-presented signals, it also occurs for particular natural situations, such as the experience of humming with one's mouth closed and ears blocked. Factors noted to affect IHL and distance perception include signal type, familiarity, prior sounds, and listener individual differences, experience and expectations. A classic example of the influence of signal type, familiarity and experience is that whispered speech is systematically perceived to be at closer distances than shouted speech, regardless of other distance cues, or the amount of experience with the system [45].

5 Human Factors of Motion-Interactive Synthetic Spatial Audio

This category of spatial audio synthesis includes mobile sources with or without head-tracking, or static sources using head-tracking. The human factors of head-tracked spatial audio all apply to SAAR and PLASA as well.

Minimum audible movement angle (MAMA) measures the smallest detectable source motion below which a source appears to be stationary. MAMA depends on source velocity, source azimuth and duration, with typical values between 5 and 30°. It also has implications for the minimum detectable latency of spatial audio in response to head-turns [54].

Localization response time (RT) is the time to localize a sound source. This depends on source azimuth, the listener's head's rotational velocity, stimulus duration, the definition of *when* localization has been achieved, and the total system latency to head-turns. RTs can also be used to assess performance of tasks other than localization, such as various pointing tasks. Fitts' Law [13] predicts that RTs for general pointing tasks increase with decreasing target size and increasing target distance. The pointing task is an analogous task to that of using a PLASA system to navigate to the location of a virtual sound source by walking, for which mean velocity (related to RT) has been shown to depend on the "capture radius" of the sound source [38].

Perceptible effects of total system latency to head-turns (TSL) are an important factor in head-tracked binaural rendering systems. TSL to head-turns is the time between when the user turns their head and the corresponding change of spatial audio rendering. This interactive rendering is relative to the world reference frame, rather than the listener's reference frame. The TSL is a technical system parameter that translates into a human factor in terms of its *detection threshold* (DT, or just-noticeable effect) and *difference limen* (DL, or just-noticeable difference). Various studies have found measurable perceptual effects of TSL to head-turns between 32 and 250.4 ms, for different tasks, covering DTs and DLs, subjective latency ratings, localization errors, front–back error rates and RTs. For instance, Brungart et al. [9] found that latencies up to 243 ms for continuous stimuli barely affected localization error, while RT increased significantly for latencies of 73 ms or greater, with stimulus durations over 500 ms. For continuous stimuli, listeners can achieve accurate localization by adapting their behaviour and reducing head-turn speed, which increases RTs, while for stimuli briefer than the TSL, latency has little effect on localization RTs, but increases localization error.

5.1 Perception in Applied Contexts

When spatial audio is used as an interface or display of information, as in SAAR, task-based performance becomes an important measure of quality. If task performance were not to be affected by variation of a fundamental performance measure (e.g. localization blur) up to some threshold value, there might be no reason to ensure system performance to that level. Functional assessment of this kind can vary substantially depending on the user tasks or perceptual aspects being studied, yet these evaluations are often revealing and valuable.

5.1.1 Research by Loomis et al.

Since 1985 [30], Loomis and colleagues have evaluated auditory localization performance afforded by the Personal Guidance System (introduced in Sect. 2.3), which was designed for outdoors, audio-based navigation assistance for visually impaired people. The use of “virtual acoustic display” is considered a potentially powerful interface thanks to direct spatial mapping of positional information, and research on spatial hearing abilities in specific mobile task contexts is directed towards achieving this potential.

Task-based performance investigation has included comparisons between spatial audio presentation and other modes such as speech-only interfaces. Further research sought to evaluate the human ability to use spatial audio navigation cues during motion. An active localization experiment [31] compared navigation to real and virtual sound sources and validated the concept of navigation using simple binaural spatial audio synthesis. It was found that azimuth bearing of synthetic spatial sounds is perceived relatively accurately, especially when utilising head-turn interaction. However, distance is often poorly estimated [49], therefore many later experiments examined distance perception of real and virtual sources.

Work by Loomis and colleagues is important to the research of mobile audio AR because it provides patterns for the design of novel experiments on motion-interactive, multimodal spatial audio perception.

6 Human Factors in Audio AR

Many projects have been developed and written up in audio AR since the early days of Cohen et al. [12] and Bederson [2]. Still there are relatively few that have strongly evaluated the relationship between various system design choices and the resulting user experience (which is discussed in Chap. 9). Most projects have been content with simple validation of the functionality of the system, discussion of potential usage, or a qualitative evaluation that doesn’t link system parameters to performance outcomes. This section presents a broad, selective survey of works that have provided more substantial investigation into human factors of audio AR media.

6.1 Locative Audio

Of the audio AR media, locative audio is probably the most prevalent, as well as presenting the most potential for popularization, given that without the requirement for head-tracking, it is the easiest to implement on mass-market devices such as smart-phones. At the same time, this medium is represented by relatively little literature containing any substantial evaluation, possibly due to its relative accessibility, ease of implementation and perhaps an impression of self-evidential functionality. Many cite the basic human factor that audio interfaces (in comparison to visual interfaces) provide a sensory information channel that may be received in a passive, background manner, as described by Sawhney and Schmandt [48] for instance.

Mantell et al. [34] introduced an interesting application for smart-phones called Navinko, designed as "a social network and navigation system enabled with audio augmented reality for cyclists in Tokyo." In this unique application, it was necessary to create an interface that was usable while riding a bicycle, and could inform users about nearby landmarks (Points Of Interest) and other users in terms of their momentary relative distance, direction and speed. They found that prior art systems were unable to provide comprehensible information about all these characteristics of multiple POIs simultaneously. One solution they found was to use a simulated Doppler effect on the sounds assigned to POIs, employing a familiar physical sonic effect as a metaphor. This provided a way of mapping extra physical interaction parameters (relative speed and distance) to sounds without adding extraneous simultaneous sounds. However, this innovation had the disadvantage of requiring continuous sounds, which prohibits some sounds that might be more semantically appropriate for some objects.

Williams et al. [56] evaluated a system that enabled school children to design and create "soundscapes" in the outdoor environment. The potential future impact of the technology on children's spatial practice is discussed and the concept of children "tagging" environmental hazards is raised.

This study provides a great deal of qualitative feedback on human factors of the individual and social user experience. For example, with a visual, screen-based interface on the user device, as well as the auditory interface, the visual feedback became a significant part of the experience:

> The children watched the screen often as they moved around, and enjoyed trying to steer the little dot that represented their position into the circles that represented sounds, rather than just using the headphones and coming across sounds.

Whilst the use of the screen as an extra user interface element in audio AR detracts from the purity of the medium, it is often desired for convenience or other reasons. When including this element, it is important to realise that it raises human factors of its own in terms of usability, attention and interaction quality.

6.2 ARA

As discussed previously in Sect. 2.2.2, an important question is which sensory modality of real-world stimuli is augmented. Different problems exist for augmentation of visible or audible objects. Both situations require perceptually accurate synthetic spatial audio, with precise and timely registration to real-world objects. However, augmentation of *acoustic* real objects requires perceptually matching the surrounding acoustical environment as well.

An example is the Wearable Augmented Reality Audio (WARA) research by Härmä et al. [17], in which synthetic audio augments the real acoustic environment around the listener. To achieve this *augmented audio reality*, Härmä et al. use headphones with embedded microphones facing outwards for active, real-time transmission of the sonic environment, mixed with the virtual audio. Perceptual evaluation in Härmä et al. [17] was limited but showed that in some cases, listeners found it very difficult to determine which sound sources were real or virtual.

Tikander [50] presented a study of the usability of an ARA headset, for long periods of use (20–40 h) in everyday life conditions, focused on issues with the active, equalised "hear-through" functionality of the device. While this device can be used to present virtual audio augmentations to the natural acoustic world, in this study no synthetic audio was added—this was suggested for future work. At the time, only the naturalness of the feed-through "pseudoacoustics" was investigated. Nevertheless, this is a valuable study in the human factors of a vital component of a complete ARA presentation system.

Many details of the usability of the pseudoacoustics were gathered. The most common annoyances related to sounds of the user being boosted due to an "occlusion effect" of internal sounds resonating in the ear canal—for example, while eating crunchy food. There were also handling inconveniences, social issues (due to the appearance of hearing being blocked by headphones), some discomfort with the in-ear headphones over time, and amplification of mechanical noises of the device, such as movement of the headset wires. Some loud sounds also caused undesirable distortion in the hear-through pseudoacoustic sound. Tests showed that mean opinion scores for annoyance of the experience improved after the field-trial period, in comparison to initial impressions. Acoustically, the headset was almost neutral and worked well for most occasions. Some difference to natural, open-ear listening was detectable in various circumstances, but was not clearly definable.

Martin et al. [40] validated two ARA headset systems with different earpieces, using localization errors, in comparison with standard virtual auditory space (VAS) technique and natural spatial hearing of real, free-field sound sources. They found that both headsets afforded the participants ability to localize sound sources in VAS as well as normal headsets. In regards to free-field localization, one earpiece afforded performance similar to normal hearing, although with poor low-frequency response, while the other earpiece interfered with normal localization, but provided a good low-frequency response.

Moustakas et al. [41] introduced an audio-only, two-player (or multiplayer) game prototype that was experienced as an ARA environment. The game consisted of hunting creatures that could only be identified by their sound, with further ambient sound objects in the real and virtual spaces, as well as the sonic presence of the other player(s).

They used questionnaires to assess the human factors of perceived user immersion; the perceptual significance of the real and virtual audio in terms of "game-play realization"; the effectiveness of interaction between players; and the overall expected and achieved user experiences.

The study showed that the game design achieved high levels of perceived user immersion, and great novelty to the players. Source localization ability was widely varied, and sound source identification was rated lower than other user experience factors that were examined. Self-reported (questionnaire results) and objective performance levels (ability to win games) were generally better for players with a background studying sound or music. This shows that listening skills may be important for faster adaptation and/or greater ability to use ARA systems.

In terms of presence, 80 % of players considered the presence of the remote player was significant within their immediate augmented audio environment. A slighter majority of 65 % felt that the real and virtual audio components equally supported the game scenario, while the rest felt the virtual component was dominant. This is evidence of the potential for high quality, immersive game-play built on ARA interaction environments.

Overall, the results of this evaluation were considered positive and encouraged the researchers to plan the development of more sophisticated game scenarios for ARA.

6.3 Spatial Audio AR and PLASA

In comparison with other forms of audio AR, relatively few SAAR or PLASA systems have been reported, particularly those with unrestricted mobility and outdoor functionality. There have been fewer published reports of human factors evaluation, particularly of a quantitative nature. However, several technical and perceptual factors are unique to this medium, so new experiments are required to understand the human factors of the new kinds of interactions that it affords.

6.3.1 Disambiguation of Front–Back Confusions by Body Motion inSAAR

In spatial hearing, front–back confusions are a well-known phenomenon whereby the listener incorrectly localizes a sound source to its mirrored position through the frontal plane (the vertical plane on the axis through the ears). This type of localization error can occur for real and synthetically spatialised sound sources. Experiments have shown that the listener can resolve front–back ambiguities by rotating their head; also that sound source movement can be used to resolve confusions if the listener is aware of the intended direction of source movement.

Mariette [36, 38] presented an experiment that shows mitigation of front–back confusions for synthetic binaural spatial audio interactive with body movement but not head-turns—i.e. SAAR without head-orientation tracking. This partly disabled mobile augmented reality system renders sound source *positions* relative to the *world* reference frame, (so the listener may walk past a stationary spatialised sound), but it renders instantaneous source *bearing* relative to the *listener's* reference frame.

As expected, front–back localization improves after the listener interacts with the spatialised sound by walking forward on a straight line past the source, so that it either looms towards, passes or recedes away from the listener. Dynamic localization cues of increasing source azimuth and changing source range enable the listener to constantly revise their judgment of a sound's location in front or behind them. An angular change in the sound source bearing while walking of between 12 and 16 □ give a significant improvement rate. This variation was observed to be similar to the expected static localization error rate for the rendering method that was used.

This suggests that higher resolution spatial sound synthesis will allow listeners to use smaller source azimuth and range changes to disambiguate front–back confusions.

Since the geometry of dynamic localization cues scales linearly with distance, the minimum source range for an acceptable localization correctness rate will then depend on the position tracking resolution. Expressed in another way, interactions exist between the audio AR system performance limitations of position tracking accuracy and computation power available for more accurate rendering. The weaker of these two specifications will determine the minimum source distance that allows acceptable front/back localization performance.

6.3.2 Validation of Navigation Using Motion-Interactive Spatial Audio

The first relevant study of navigation affordance was before the earliest use of the term audio AR, and involved motion-interactive, basic spatial audio used to represent world-stationary sound sources. Loomis et al. [31] created a "simple virtual sound display" with analogue hardware controlled by a 12MHz 80286 computer, video position tracking using a light source on the user's head, and head-orientation tracking using a fluxgate compass. Basic binaural spatial audio was rendered by synthesis of interaural time and intensity differences, with distance simulated using a filter for atmospheric attenuation and a variable ratio of direct to reverberated audio.

This experiment investigated human factors with regards to a navigation task, where participants had to "home-into" real or virtual sound sources around them. The real/virtual sound source condition had no significant effect on any of the performance measures: time to localise; distance error at perceived location; and absolute angular change during navigation. The path angular change was much larger for virtual stimuli (33°) than real stimuli (14 °), but this difference was not found to be statistically significant. Loomis et al. concluded that a simple virtual display could be "effective in creating the impression of external sounds to which subjects can readily locomote," but that more sophisticated displays might improve space perception and navigation performance.

That, and much other research has focused on basic verification of navigation ability. As Loomis et al. stated: "homing can be accomplished merely by keeping the sound in the median plane until maximum sound intensity is achieved."

A study by Jones et al. [23] investigated outdoors navigation of three routes on a university campus, using directional guidance provided by applying stereo-panning to music tracks according to the user's position, which was tracked using a GPS receiver. This study measured completion rate and mean time to complete each route, as well as ratings for the NASA *task load index* questionnaire. They found that simple panning was adequate for navigation of complicated routes, and users had a positive experience on several scales. They also found that the navigation completion rate suffered for routes on an open field in comparison to routes limited by pathways and buildings.

Some other studies such as those by Lokki et al. [28], Rutherford and Withington [47] and Walker and Lindsay [52] have used virtual auditory environments (VAEs) to evaluate navigation tasks using spatialised auditory beacons. However, these studies occurred indoors, with seated participants, so while they use related technology, they are not able to inform on human factors of audio AR.

6.3.3 Human Factors with Respect to Rendering Method and Head-TurnLatency in Navigation with SAAR

More recent studies premise that navigation *is* afforded by SAAR systems, then consider the human factors relating to technical specifications of the system, such as latency and rendering technique.

Mariette [35, 36] assessed human factors of an outdoor navigation task using a mobile SAAR system. One pilot study and a main experiment attempted to measure variations in the navigation performance afforded by the system. Two common technological limitations of such systems are the binaural rendering resolution, and latency between head-turns and corresponding rendered audio changes. The experiments investigated how several quantitative navigation performance measures and one subjective measure might be affected by the binaural rendering technique and head-turn latency. These are human factors unique to the SAAR environment, and technical parameters that are often limited by other aspects of the system design, such as available computing power, battery life, hardware financial budget, and the mobile device operating system.

In order to study the human factors of this situation, the task was to navigate from a central base position to the location of multiple virtual sound sources that were stationary in the world reference frame. This task was designed as a generalization of any navigation task using spatial audio beacons, since any continuous series of way-points (A–B–C–D–…) can be generalised as a series of point-to-point navigations (A–B, A–C, …). Thus the same experiment protocol could be used in the future to examine the effect of many other technical factors on human performance, and the supported affordance for any navigation activity.

For each stimulus, system parameters were varied, providing the technical factors under examination. Simultaneously, body position/orientation and head-orientation sensor data were recorded for later analysis. Novel objective performance measures were then used to assess the degradation of participants' navigation performance from the ideal, due to tested system parameter values.

The pilot experiment used only one participant, and found that a source *capture radius* of two meters significantly affected the user's navigation distance efficiency compared to other radii. The capture radius is specified as the distance from the exact source location within which the source is considered to be localised correctly. Decreasing capture radius significantly reduced distance efficiency ($p < 0.05$), showing that straighter navigation paths were supported by larger capture circles. Therefore, for better navigation performance in SAAR applications using similar technical specifications, source capture radius should be three meters or more.

The main experiment, using eight participants, found that render method significantly affected all performance measures except subjective stability rating, while head-turn latency only affected mean track curvature and subjective stability.

Overall results showed that regardless of sometimes severe system performance degradation, all eight participants successfully navigated to most source positions within a reasonable time limit of 60 s.

This experiment shows the unique human factors of mobile, SAAR in comparison to static spatial audio displays. When navigation performance is compared to static sound source localization, it is clear that improved localization performance is enabled by the perceptual feedback provided by system interactivity to user position movements and head-turns. While front–back confusions are common and azimuth errors are large in static experiments, in the SAAR, participants usually began their navigation in approximately the correct direction, and were almost always successful in walking to the source location.

Higher-precision binaural rendering significantly improved most navigation performance measures, while excessive head-turn latency only showed significant effects on stability rating and mean track curvature, and only then for the greatest latency level of 976 ms. No significant performance effects were found for latencies of 376 ms or worse.

The most interesting result was a statistically significant interaction effect between rendering method and latency, in which subjective stability degradation occurred for increased head-turn latency only when the higher-precision rendering was used. Apparently the lower resolution rendering mitigated the detrimental effect of high head-turn latency on perceived stability.

The primary conclusion for human factors of navigation using SAAR was that even systems with high head-turn latency or relatively low resolution rendering afforded successful navigation to positional sound sources. However, degradation of both specifications does have a significant detrimental effect on objective and subjective performance. Improved navigation performance is best supported by improving both system specifications at the same time, but within the parameters of this study, greater performance benefits were achievable by increasing rendering resolution than by reducing system latency. Mid-range latencies up to 376 ms TSL can be tolerated for any of the rendering methods used in the study.

6.4 Audio AR Implications for Human Factors of General Sound inSpace

Finally, in an almost circular extension of human factors research, audio AR has been used to understand the more general experience of sound in space. Kinayoglu [24] used audio AR to investigate how sound influences sense of place in terms of emotive, synaesthetic effects; attention and gaze behaviour; spatial orientation and sense of scale; audio–visual congruence in sense of place; and perception of personal and social space.

The experimental protocol compared the experience of space during free exploration of several zones within a university campus, separately using natural hearing and mobile, head-tracked SAAR. Participants rated their acoustic and multi-modal experience of each zone in terms of pleasantness, vibrancy/impressiveness, noisiness, relaxation, orientation, intimacy and familiarity of the environment. In some places, the sound design deliberately played with congruence with the surrounding visible environment. Results showed that soundscapes influence the experience of place, with effects on climatic, emotive, attentive, social and behavioural impressions of the spaces. In particular, sonic aesthetic qualities most strongly influenced emotional response to the environment, and semantic compatibility most strongly influenced the audio–visual congruence ratings, with incongruence resulting in anxiety and disconnectedness, even when using aesthetically pleasing sounds. This research shows that the semantic and aesthetic qualities of virtual audio content contribute to important human factors in the AR experience.

7 Conclusions

Human factors of audio AR media is a vital front of investigation in the field. Many experimental audio AR systems of various classifications have been presented to date, yet greater availability of consumer and commercial applications has only recently begun. At this stage in the development of the medium, it is important to understand how the component technologies, and variations of system design and specifications affect the user experience. This chapter has provided an introduction through classification of the media involved, a review of selected prior art, and a survey of human factors investigations. With continuing development along these lines, the utility and effectiveness of audio AR will further improve, and the field is likely to produce some thrilling future applications.

References

1. Azuma, R.: A survey of augmented reality. Presence: Teleoperators and Virtual Environment 6(4), 355–385 (1997)
2. Bederson, B.B.: Audio augmented reality: A prototype automated tour guide. In: CHI. pp. 210–211 (May 7–11 1995)
3. Begault, D.R.: 3-D Sound for Virtual Reality and Multimedia. Academic Press Professional, Cambridge, MA (1994)
4. Begault, D.R., Wenzel, E.M., Anderson, M.R.: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J of the Audio Eng Soc 49(10), 904–916 (2001)
5. Best, V.: Spatial Hearing with Simultaneous Sound Sources: A Psychophysical Investigation. Ph.D. thesis, Faculty of Medicine, University of Sydney (April 2004)
6. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA (1983)
7. Blauert, J.: Spatial hearing: The psychophysics of human sound localization (rev. ed.). MIT Press, Cambridge, MA (1997)
8. Brungart, D.S., Simpson, B.D.: Effects of temporal fine structure on the localization of broadband sounds: potential implications for the design of spatial audio displays. In: Int Conf on Aud Disp. Paris, France (June 24–27 2008)
9. Brungart, D.S., Simpson, B.D., McKinley, R.L., Kordik, A.J., Dallman, R.C., Ovenshire, D.A.: The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. In: Int Conf on Aud Disp. Sydney, Australia (July 6–9 2004)
10. Carlile, S., Best, V.: Discrimination of sound source velocity in human listeners. J of the Acoustical Soc of America 111(2), 1026 (February 2002)
11. Cohen, M.: Augmented Audio Reality: Design for a Spatial Sound GPS PGS. In: Center on Disabilities Virtual Reality Conf. California State University, Northridge (1994)
12. Cohen, M., Aoki, S., Koizumi, N.: Augmented audio reality: Telepresence/ar hybrid acoustic environments. In: IEEE Int Workshop on Robot and Human Communication. pp. 361–364 (1993)
13. Fitts, P.: The information capacity of the human motor system in controlling the amplitude of movement. J of Experimental Psychology 47(6), 381–91 (1954)
14. Galloway, A., Ward, M.: Locative Media as Socialising and Spatialising Practice: Learning from Archaeology. Leonardo Electronic Almanac 14(3) (2006)
15. Gibson, J.J.: In Perceiving, Acting, and Knowing: Toward an Ecological Psychology, chap. The theory of affordances. Lawrence Erlbaum Associates, Hillsdale, NJ (1977)
16. Härmä, A.: Ambient human-to-human communication. Handbook of Ambient Intelligence and Smart Environments pp. 795–823 (2010)
17. Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Nironen, H., Vesa, S.: Techniques and applications of wearable augmented reality audio. In: 114th Conv Audio Eng Soc. Amsterdam, The Netherlands (March 22–25 2003)
18. Helyer, N.: Sonic landscapes (22/8/2006 1999–2001), http://www.sonicobjects.com/index.php/sonicobjects/more/sonic_landscapes/
19. Holland, S., Morse, D.R., Gedenryd, H.: AudioGPS: spatial audio in a minimal attention interface. Personal and Ubiquitous Computing 6(4), 253–259 (2002)
20. Jeffress, L.A., Taylor, R.W.: Lateralization vs Localization. J of the Acoustical Soc of America 32, 936 (1960)
21. Jin, C., Corderoy, A., Carlile, S., van Schaik, A.: Contrasting monaural and interaural spectral cues for human sound localization. J of the Acoustical Soc of America 115, 3124 (2004)
22. Jones, K.S.: What is an affordance? Ecological Psychology 15(2), 107–114 (2003)
23. Jones, M., Bradley, G., Jones, S., Holmes, G.: Navigation-by-Music for Pedestrians: an Initial Prototype and Evaluation. Proc of the Int Symp on Intelligent Environments: Improving the quality of life in a changing world (April 5–7 2006)

24. Kinayoglu, G.: Using audio-augmented reality to assess the role of soundscape in environmental perception. In: eCAADe27. Istanbul (2009)
25. Kyriakakis, C.: Fundamental and technological limitations of immersive audio systems. Proc of the IEEE 86(5), 941–951 (1998)
26. Lindeman, R.W., Noma, H.: A classification scheme for multi-sensory augmented reality. In: Proc 2007 ACM symposium on Virtual reality software and technology. pp. 175–178. ACM (2007)
27. Livingston, M.A.: Evaluating human factors in augmented reality systems. Computer Graphics and Applications, IEEE 25(6), 6–9 (November/December 2005)
28. Lokki, T., Grohn, M., Savioja, L., Takala, T.: A case study of auditory navigation in virtual acoustic environments. In: Int Conf on Aud Disp (2000)
29. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Personal Guidance System for the Visually Impaired using GPS, GIS, and VR Technologies. In: VR Conf. California State University, Northridge (1993)
30. Loomis, J.: Digital map and navigation system for the visually impaired (July 25 1985), http://www.geog.ucsb.edu/pgs/papers/loomis_1985.pdf
31. Loomis, J., Hebert, C., Cicinelli, J.: Active localization of virtual sounds. J of the Acoustical Soc of America 88(4), 1757–1764 (October 1990)
32. Lyons, K., Gandy, M., Starner, T.: Guided by voices: An audio augmented reality system (2000)
33. Mallem, M.: Augmented reality: Issues, trends and challenges. In: 2nd Int Conf on Image Processing Theory Tools and Applications (IPTA). p. 8 (July 2010)
34. Mantell, J., Rod, J., Kage, Y., Delmotte, F., Leu, J.: Navinko: Audio augmented reality-enabled social navigation for city cyclists. In: Pervasive (2010)
35. Mariette, N.: CMMR/ICAD 2009, chap. Navigation Performance Effects of Render Method and Head-Turn Latency in Mobile Audio Augmented Reality, pp. 239–265. Lecture Notes in Computer Science 5954, Springer, Berlin/Heidelberg (2010)
36. Mariette, N.: Perceptual Evaluation of Personal, Location-Aware Spatial Audio. Ph.D. thesis, University of New South Wales, Sydney, Australia (March 2011)
37. Mariette, N.: From backpack to handheld: The recent trajectory of personal location aware spatial audio. In: 7th Int Digital Arts and Culture Conf. Perth, Australia (15th–18th September 2007)
38. Mariette, N.: Mitigation of binaural front–back confusions by body motion in audio augmented reality. In: 13th Int Conf on Aud Disp. Montréal, Canada (June 26–29 2007)
39. Martens, W.L.: Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. Acoustical Science and Technology 24(5), 220–232 (2003)
40. Martin, A., Jin, C., Schaik, A.: Psychoacoustic evaluation of systems for delivering spatialized augmented-reality audio. J of the Audio Eng Soc 57(12), 1016–1027 (December 2009)
41. Moustakas, N., Floros, A., Grigoriou, N.: Interactive audio realities: An augmented / mixed reality audio game prototype. In: 130th Conv Audio Eng Soc (May 13–16 2011)
42. Mynatt, E.D., Back, M., Want, R., Baer, M., Ellis, J.B.: Designing audio aura. In: Proc SIGCHI Conf on Human factors in Comp Sys. pp. 566–573. ACM Press/Addison-Wesley Publishing Co. (1998)
43. Novo, P.: Communication Acoustics, chap. Auditory Virtual Environments, pp. 277–297. Springer, Berlin–Heidelberg–New York (2005)
44. Pernaux, J.M., Emerit, M., Nicol, R.: Perceptual evaluation of binaural sound synthesis: the problem of reporting localization judgments. In: 114th Conv Audio Eng Soc (March 22–25 2003)
45. Philbeck, J.W., Mershon, D.H.: Knowledge about typical source output influences perceived auditory distance. J of the Acoustical Soc of America 111, 1980–1983 (2002)
46. Rozier, J., Karahalios, K., Donath, J.: Hear&There: An Augmented Reality System of Linked Audio. In: Int Conf on Aud Disp. Atlanta, Georgia (April 2000)

47. Rutherford, P., Withington, D.: The application of virtual acoustic techniques for the development of an auditory navigation beacon used in building emergency egress. In: Proc 2001 Int Conf on Aud Disp. Espoo, Finland (July 29–August 1 2001)
48. Sawhney, N., Schmandt, C.: Design of Spatialized Audio in Nomadic Environments. In: Proc Int Conf on Aud Disp (1997)
49. Speigle, J.M., Loomis, J.M.: Auditory distance perception by translating observers. In: Proc, IEEE Symp on Research Frontiers in Virtual Reality. pp. 92–99 (1993)
50. Tikander, M.: Usability issues in listening to natural sounds with an augmented reality audio headset. J Audio Eng Soc 57(6), 430–441 (2009)
51. Walker, B.N., Lindsay, J.: Navigation performance in a virtual environment with bonephones. In: Proc 11th Int Conf on Aud Disp. pp. 260–263. Limerick, Ireland (2005)
52. Walker, B., Lindsay, J.: Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Human Factors 48(2), 265 (2006)
53. Warusfel, O., Eckel, G.: LISTEN – augmenting everyday environments through interactive soundscapes. In: IEEE Virtual Reality (2004)
54. Wenzel, E.M.: Analysis of the role of update rate and system latency in interactive virtual acoustic environments. In: 103rd Conv Audio Eng Soc. New York, NY, USA (September 26–29 1997)
55. Wightman, F.L., Kistler, D.J.: Resolution of front-back ambiguity in spatial hearing by listener and source movement. J of the Acoustical Soc of America 105(5), 13 (1999)
56. Williams, M., Jones, O., Fleuriot, C., Wood, L.: Children and emerging wireless technologies: investigating the potential for spatial practice. In: Proc SIGCHI Conf on Human factors in Comp Sys, April. pp. 02–07 (2005)
57. Woo, D., Mariette, N., Helyer, N., Rizos, C.: Syren – a ship based location-aware audio experience. J of global positioning systems, Int Assoc of Chinese Professionals in GPS 4(1–2), 41–47 (2005), alberta, Canada
58. Woo, D., Mariette, N., Salter, J., Rizos, C., Helyer, N.: Audio Nomad. In: ION GNSS. Fort Worth, Texas (September 26–29 2006)
59. Zhou, Z., Cheok, A.D., Qiu, Y., Yang, X.: The role of 3-D sound in human reaction and performance in augmented reality environments. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 37(2), 262–272 (March 2007)
60. Zwiers, M., Van Opstal, A., Paige, G.: Plasticity in human sound localization induced by compressed spatial vision. Nature Neuroscience 6(2), 175–181 (2003)

Part II
Perception and Cognition

Chapter 3
Basic Perception in Head-Worn Augmented Reality Displays

Mark A. Livingston, Joseph L. Gabbard, J. Edward Swan II, Ciara M. Sibley, and Jane H. Barrow

1 Introduction

For many first-time users of augmented reality (AR) displays, the experience suffers compared to their expectations. While several human factors issues are responsible for this disconnect, abundant anecdotal evidence and numerous controlled laboratory studies have shown that part of the performance gap is in the low perceptual quality of the graphical presentation. Despite extensive research in producing photorealistic graphics, little work in AR has been demonstrated to have that level of visual realism. Reviewing the literature and our own experiences and research, we identified four fundamental areas in which basic perception of the virtual and real elements in the merged world may be lacking. Visual acuity captures issues of geometric resolution, limited contrast and distorted perception of colors reveal issues of color resolution and presentation. These challenges lead naturally to issues of text legibility. In many applications, depth segmentation raises issues regarding the quality of stereo imagery. In this chapter, we examine these four issues as they apply to head-worn AR displays.

M.A. Livingston (✉) • C.M. Sibley
U.S. Naval Research Laboratory, Washington, DC 20375, USA
e-mail: mark.livingston@nrl.navy.mil; ciara.sibley@nrl.navy.mil

J.L. Gabbard
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
e-mail: jgabbard@vt.edu

J.E. Swan II
Mississippi State University, Starkville, MS 39762, USA
e-mail: swan@acm.org

J.H. Barrow
George Mason University, Fairfax, VA 22030, USA
e-mail: jbarrow1@gmu.edu

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_3,

Despite numerous display options for augmented reality (AR), head-worn displays are still the popular choice for industrial, medical, and military applications. Among the most important advantages they offer is hands-free viewing of the environment and adding information into the environment from the user's point of view. Such head-worn displays come closest to instantiating Sutherland's original vision [35]. However, we have documented the reduction of fundamental capabilities of the human visual system when viewing the environment through head-worn AR displays. In this chapter, we will discuss the theoretical and ecological background of these issues, experimental designs to measure the practical implications for AR users, and report published and unpublished results from experiments conducted in our respective laboratories. Further, we argue for the importance of including this type of evaluation of AR systems both early and at regular intervals during the design and evolution of an AR system.

For many years, the Naval Research Laboratory and the Office of Naval Research sponsored research on various aspects of mobile AR systems. The central application in these programs was the Battlefield Augmented Reality System (BARS™) [23, 25]. Development of BARS was conducted through a usability engineering paradigm [9, 12], in which technical capabilities of the system and user performance on representative tasks with that system [26] are evaluated iteratively. (See Chaps. 8 and 9 for other perspectives on usability engineering and user experience applied to mobile AR.) BARS was envisioned to provide situation awareness (SA) [5] for a dismounted warfighter during military operations in urban terrain. Among the many usability engineering findings for BARS was the fundamental need to be able to read text labels presented through AR. Labels could identify streets or landmarks in the environment or be part of military standard icons giving information designed to provide SA. This was one impetus for the work on the legibility of text in head-worn AR displays discussed in this chapter.

One goal with the BARS hardware was to compare whether the AR condition was as natural as a real analog of the task was. This gave an objective basis for evaluating the quality of AR systems. We set up a series of comparisons, the simplest of which was inspired by early efforts to test the effective visual acuity of immersive head-worn displays [16]. This led us to create an AR Snellen eye chart [27]. We found that users (all with at least normal or corrected-to-normal vision) had their visual acuity decreased by wearing the AR display and looking at a real eye chart, and similarly decreased when looking at the graphical eye chart.

An extension to the BARS application was training in basic combat skills that might be required in an urban context [23]. In the course of evaluating this application, several observations were made about the quality of the head-worn video-overlay display [1]. In particular, subjects noted difficulty seeing the real walls in the environment, perceiving depth, and that the (real) targets were too small to see. This further motivated us to examine the perceptual capabilities users were able to achieve with head-worn AR displays and is another example of the value of the iterative approach to evaluation the performance of complex AR systems.

2 Measures of Visual Capabilities

In this section, we discuss measures of visual capabilities relevant to head-worn AR displays. The goal of this review is to provide a brief introduction to concepts from perception for AR researchers and users. We will conclude this section with a discussion of how AR affects these basic perception factors.

2.1 Visual Acuity and Contrast Sensitivity

The quantity that comes to mind for most people when asked about visual capabilities is *visual acuity*, the ability of the observer to resolve fine details in the visual field. This is quantified by determining the smallest stimulus (measured in angular size) that the observer can identify at a rate better than chance. This quantity might lead to an impression that our ability to recognize objects is a function only of the size; in truth, recognition is a function of both size and contrast. *Contrast sensitivity* describes the observer's ability to discern differences in the luminance (or color) values across the visual field; it measures the threshold of contrast required to accurately perceive the target. Since contrast is a ratio of foreground to total luminance, its value is in [0,1]. Sensitivity is the reciprocal of this threshold. For each spatial frequency, contrast sensitivity is measured; these are then connected into a curve which separates perceptible stimuli from imperceptible ones. Both axes are typically drawn with logarithmic scaling. The canonical shape of such a *contrast sensitivity function* (CSF) is shown in Fig. 3.1. Sensitivity forms a concave-down parabola; objects whose size and contrast fit under the curve are perceptible.

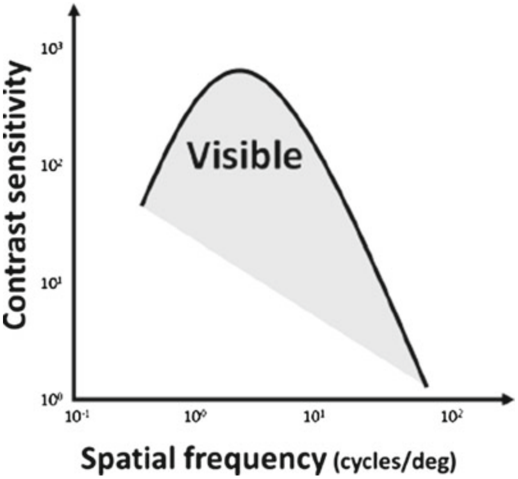


Fig. 3.1 The canonical shape of the contrast sensitivity function, graphed as a function of spatial frequency for size and contrast threshold for sensitivity

Normal visual acuity is considered to be the capability of resolving 1 min of visual arc [33]; this is most often measured (e.g. by an optometrist) by reading a Snellen chart at a standard distance of 20 ft (6 m), with the letters scaled appropriately so that the various arms, bars, and counters that differentiate letters occupy the requisite number of minutes of visual arc for 20/20 vision (1 min), 20/30 vision (1.5 min), etc. One can immediately see the challenge of the Snellen chart: not all letters are "equally spaced" in this design space. For example, the bar (middle, horizontal stroke) in a lower case "e" separates it from the lowercase "c," which is in turn separated by the counter (opening) from the lowercase "o." An "equally spaced" Snellen chart must ensure that the bar of the "e" and the counter of the "c" are equal in width. Another common example is that the length of the lower arm of a capital "E" must be exactly 1 min of arc to separate it from a capital "F," but this implies that the separation from a capital "L" will be different. A similar design issue is that sans-serif fonts should be used.

In response to this type of difficulty and to standardize across countries with different alphabets, one may instead use a rolling "E" chart. This chart shows a capital letter "E" in one of four orientations—i.e. with the arms pointing up, right, down, or left. But this again is not quite equal in all directions, unless the letter is perfectly square. The Landolt "C" chart allows a slight improvement on this; it orients a capital "C" with the opening on the top, right, bottom, or left of the figure. Again, a chart designer should take care to use a rotationally symmetric figure, even if the letter is not normally printed in such a fashion. This will prevent the observer from gaining a clue based on the aspect ratio of the figure. Another option for similar charts is to use sine-wave gratings. A small 2D image with a wave in one dimension and a constant value in the other dimension creates a figure of which one may query the observer for its orientation. The orientation of the waves can span up to (but not equal to) 180 °. A chart designer would likely select four cardinal orientations, such as waves that are horizontal, vertical, or along the two 45 ° diagonals.

When transferring these figures to a digital imaging device, another issue arises for the chart designer. The discrete nature of the display interferes with the production of the desired figures for recognition. For example, a diagonally-oriented sine wave is guaranteed to create aliasing on a digital display. While anti-aliasing techniques can mitigate the problem, a better strategy would avoid such a problem altogether. A similar difficulty exists for any letter or derived figure that uses curved strokes. Thus the Landolt "C" chart suffers from this difficulty as well. However, this can be mitigated by relaxing the requirement of using a known figure (such as a letter) in favor of simply a recognizable feature of an abstract figure. Thus a "squared" version of a "C" that uses only horizontal and vertical strokes but leaves a small opening is a reasonable design choice [7]. A rolling "E" could also be designed to fit these requirements, since the letter "E" requires only vertical and horizontal strokes in any of the four cardinal orientations.

However, the design challenges do not end with the shape of the figure. The relative brightness between the foreground and the background interacts with the size of the features. So, as noted above, recognition is a function of both size and contrast. Contrast sensitivity has been accepted as part of a comprehensive approach to describing visual capabilities [14] and can be crucial in clinical evaluations for cataracts and diabetic retinopathy. The standard minimum for contrast in optometric examinations is 0.8, although it is unclear how rigidly this standard is followed in clinical practice or by what contrast definition it is to be measured. Contrast is frequently expressed by the Michelson definition:

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}},$$

where $L_{\max}$ and $L_{\min}$ are, respectively, the maximum and minimum luminances in the image. According to some, the visual system measures these extreme values from a local region around the foreground features [29]. The contrast in the image influences the observer's visual acuity score; at higher levels of contrast, the human eye is capable of detecting smaller details. Snellen charts, sine-wave gratings, rolling "E" charts, and Landolt "C" charts all may be adapted to provide a convenient way to measure an observer's CSF.

2.2 Color Perception

The retinal responses to various wavelengths of light result in the spectrum of hues available to human *color perception*. The three types of cones (long, medium, and short wavelength—typically referred to as red, green, and blue, respectively) in the retina respond to different but overlapping ranges of wavelength of light, creating the effect that the visual system processes and interprets as color. The Commission Internationale de l'Éclairage (CIE) defined three standard primaries to describe color in 1931, leading to the CIE chromaticity diagram (Fig. 3.2). One problem with this diagram was that distance in the space did not have a uniform meaning to the perception of color. That is, a single distance could be a perceptually small difference in one region of the space while at the same time being a perceptually large difference in another region of the space. This led eventually to the CIE 1976 ($L^*a^*b^*$) color space (CIELAB). L denotes a luminance channel; a and b are chrominance channels (Fig. 3.3). The a axis moves from green to red; the b axis moves from blue to yellow. This description of color closely matches the opponent process theory [15] of how the human visual system process wavelengths of light into color. The model says that three relative measurements are acquired: one differentiating black and white (the luminance channel), one differentiating between red and green, and one differentiating between blue and yellow (the latter of which is itself derived from red and green). This space is (nearly) perceptually uniform, with distortions of perceptual difference thought to be approximately 1.6:1—i.e. about a 60 % change in distance might be interpreted as perceptually identical. While far from perfect, it represents a significant improvement over the estimated 1,000:1 ratio of perceptually similar distances that are present in the 1931 CIE chromaticity diagram. One curious observation about CIELAB is that the colors that may be specified well exceed normal human visual

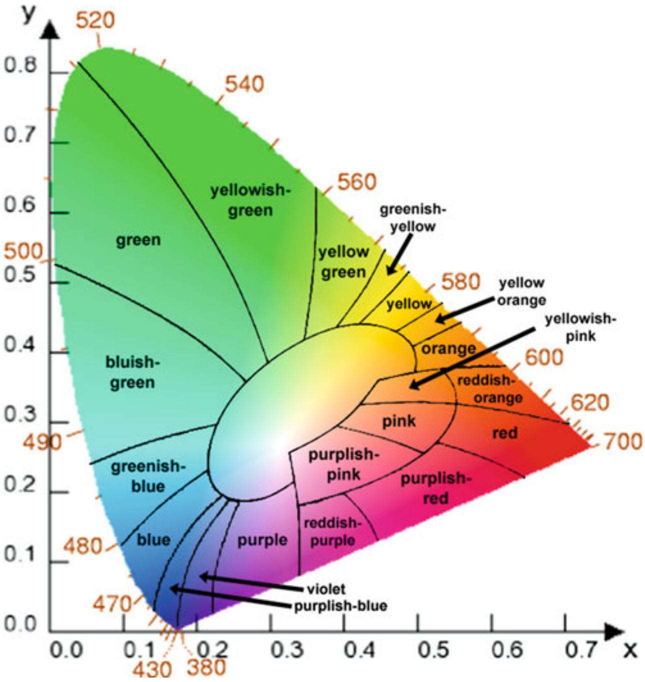


Fig. 3.2 Names may be superimposed on the regions of a color space, such as the 1931 CIE chromaticity diagram seen here. However, some of these names are far from common usage, and individuals will exhibit differences in the names

capacity for differentiating colors, which in turn typically surpasses the monitor gamut for a display device (or printer). We note in passing other color specifications, such as CIE Luv (a similar model to CIELAB, where L is lightness and u and v are chromaticity coordinates; hue-saturation-value (HSV) is another standard color space, built on a single chroma coordinate (hue), a saturation value that gives distance from achromatic value, and a brightness coordinate.

Color names may be ascribed to the regions of a color space; however, here again, problems in the 1931 CIE chromaticity diagram become obvious. The size of the regions is far from uniform. Individuals will exhibit significant differences in where the boundaries should be drawn between subjects and perhaps even within subjects, depending on any number of factors, including lighting and other physical issues, and even mood or other subjective issues. In addition, many of the "standard" names seen in Fig. 3.2 are far from being commonly-used terms to describe color. There are, however, color names that are consistently used. Of these, eight are chromatic and have one-word English names: red, green, blue, yellow, purple, orange, brown, and pink. (Achromatic terms black, gray, and white complete the list of basic color terms.) These colors were found to be maximally discriminable and unambiguously named, even across cultures [34]. Thus color naming can be a valid task that indicates color perception.

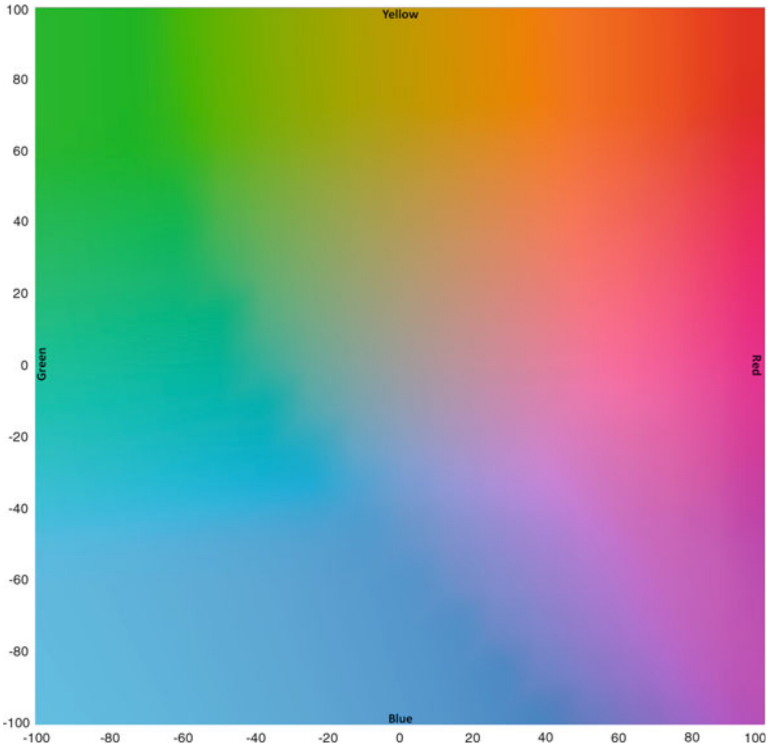


Fig. 3.3 The CIELAB space aimed to both mimic the physiological measurements of the eye and form a perceptually uniform color space. This slice at $L = 65$ shows the *a* axis running from *green* (*left*) to *red* (*right*) and the *b* axis running from *blue* (bottom) to *yellow* (*top*). The origin of the *ab* plane is the *gray* pointin the center of this plot. The portion of the space shown here ranges from – 100 to 100 in both *a* and *b* and corresponds to the portion drawn in later graphs

Color vision testing can be done in several ways. The most common method is the Ishihara pseudoisochromatic test [17], which uses a series of color plates. These 38 plates show mosaic images; the tiles of the mosaic have irregular but smooth shapes and appear in one of two colors, with varying size, brightness, and saturation to prevent determination of the figure without color discrimination. One color serves as the background, while the other color tiles draw a numeral to be recognized or a curved lines to be traced. These two hues are chosen such that people with normal color vision will differentiate them; however, people with color vision abnormalities will be unable to differentiate them. With a series of plates that test color combinations known to cause confusion for the various forms of color vision deficiencies, one can measure color vision capabilities. Versions of the test that use 14 or 24 plates are also common. The similar Dvorine pseudoisochromatic color test is based on the same conceptual design. Most figures consisted of a numeral or numerals, with three sizes of dots that did not vary in intensity. Other figures again used the path tracing task of Ishihara. Both sets of color plates (and other, similar sets) provide a test that is easy to administer and easy to take. However, the light source ought to properly match the intended source, and the colors must be carefully selected and may need to be varied for each individual. Finally, no accurate scoring criteria for classifying the type or extent of a defect are available.

Another strategy for color vision testing is to use a color arrangement test, of which the Farnsworth–Munsell test [6] is most relevant to our discussion below. A set of 100 colored tiles are placed before the subject, with an additional tile serving as a reference tile. (Later variations cut the number to 15 or other values.) The idea of the color arrangement test is that the colors can be naturally ordered by someone with normal color vision. In the typical procedure, a blue shade is identified as a reference color, and the subject is instructed to place next to it the most similar of the remaining tiles; this step is iteratively repeated with the most recently placed tile becoming the reference tile. A subject with normal color vision will trace out a smooth, approximately circular path in either the CIE 1931 chromaticity diagram or in CIELAB. (A large difference in hue between the reference tile and the intended last tile ensures that subjects with normal color vision will correctly identify the first tile to place next to the reference.) A subject with normal color vision will move around the circle, whereas a user with deficient color vision will make some characteristic steps across the circle rather than follow the circumference. The advantage of this method is that any color deficiency can be detected, rather than the finite preplanned anomalies detectable by color plates. This test can also be evaluated numerically by using the angles of the segments between the ordered tiles [37]. However, the light source must still match the intended source, and the abstract ordering task requires more mature thinking, as well as manual dexterity. Thus this test in not appropriate for very young subjects.

2.3 *Stereoacuity*

Humans (and other primates) have overlap between the visual fields of their two eyes, creating the ability to interpret 3D geometry from the 2D retinal measurements and the offset between the eyes. This yields two distinct depth cues, angle of convergence between the eyes to fixate on an object, and binocular disparity of the object in the two retinal images. Convergence angle has rather limited value, helping primarily at near field[1] distances [3]. However, binocular disparity is a much stronger cue at near field distances (enabling one or two orders of magnitude greater depth precision) and extends well beyond the near field, perhaps to several hundreds of meters from the observer. Despite this capability, the utility of stereo mechanisms in head-worn displays has not been studied extensively. This is in part due

---

[1] The near field is most often defined as either that which is within arm's length or the slightly more generous but standardized length of 2 m. Other definitions may extend the range.

to a lack of hardware features; until recently, displays that permitted adjustment of inter-pupillary distance (IPD) were limited to a few research systems and the occasional commercial offering. By comparison, it has long been trivial in rendering software to create arbitrarily-separated images for left and right eyes in binocular displays. Incorrect IPD distorts distance perception; if the IPD is set at the population mean of 65 mm, then an observer with a larger IPD would think the object were farther away than it really is; an observer with smaller IPD would think it were closer. Similarly, errors in the judged depth of nearby virtual objects have been measured as a function of changes in binocular vergence.

The ability to discern a difference in depth from stereo vision is known as *stereoacuity*. Many aspects of stereo perception are thought to exhibit individual differences. The range of normal human IPD is wide relative to the absolute size; the smallest "normal" value is approximately 53 mm, while the largest normal value is approximately 73 mm. Additionally, the distance in front of a user's eyesat which a head-worn display may rest varies (much as different people wear eyeglasses at different points on the bridge of the nose); this distance may affect how the IPD should be set. Finally, some people (estimated as high as 20 % [38] and as low as 2–5 % [2]) are *stereo-blind*, and thus receive no depth information from binocular disparity (though they often compensate well enough to obscure this). So while it is a measurable quantity, a test to evaluate the effect of head-worn AR displays should account for these individual differences or screen out stereo-blind subjects. Stereoacuity has been measured for stereoscopic depth at as little as 3 arcsec. This would at first seem to be in conflict with the 1 arcmin for normal visual acuity; it is clear that the human visual system has an extraordinary ability to accurately reconstruct the world through stereo vision.

2.4 *Effects of AR Displays*

AR displays alter these fundamental measures of perception in meaningful ways. The user's capabilities interact with the hardware in ways that are not necessarily intuitive or obvious to AR system designers. We revisit these basic capabilities, discussing how AR displays interact with each. We will use the terms *optical see-through* and *video overlay* to describe the two major variants of head-worn AR displays. Other displays, such as head-mounted projective displays (HMPDs) and hand-held displays are mentioned only briefly. For each metric, it is important to understand how the AR user's perception of both the surrounding real environment and the virtual entities may be affected by the display. This can be very different for optical see-through displays versus video overlay displays. We also introduce our terminology of a display *device* implying the entire system that mediates and/or relays images to the observer's eyes. Devices incorporate display *elements*, usually liquid crystal displays (LCDs) or—in more recent display devices—organic light emitting diodes (OLEDs). Display *optics* include various lenses and mirrors (which may be partially-silvered and thus translucent). All these components and their arrangement can affect the fundamental measures of perception.

2.4.1 Visual Acuity

It is a relatively straightforward calculation to convert the resolution in pixels and field of view (FOV) of a head-worn display into an estimate of visual acuity. But it is important to note that simply measuring the angular resolution of a head-worn AR display is not sufficient to characterize the visual acuity or the broader experience a user will have when wearing the display. While it would theoretically place a limit on the performance a user could achieve, the human visual system is quite exceptional at filling in and interpolating information. Thus if one uses a Snellen eye chart, the visual system may be able to infer the identity of recognizable shapes (i.e. letters) without having complete information. Abstract figures and shapes, or rotationally symmetric figures and shapes, may overcome this confounding factor in acquiring measurements.

There is a fundamental difference between perceiving virtual objects and perceiving the real environment as it relates to visual acuity. Optical see-through displays theoretically should not affect the visual acuity with regard to the real environment. If acuity is measured as a purely geometric assessment, this is the case. However, the measurement using a Snellen chart is of legibility, which is a recognition task. If the chart also tests contrast, then the optics play a critical role. Given the reality of some optical see-through displays, this is an inherent characteristic of the test, as described below. The visual acuity of the user in perceiving virtual objects will be affected by the angular resolution of the display elements, even in an optical see-through display.

Video overlay AR displays (whether head-worn or hand-held) present both the real and virtual portions of the scene on finite-resolution display elements, which means that both are subject to reduced visual acuity. Further, the limit of the visual acuity may be lowered further by the camera that acquires the real environment. Some commercial video overlay AR displays have incorporated cameras that had lower resolution than the display elements, so this is a practical consideration. We note that HMPDs offer graphics at a resolution determined by the projector's angular resolution. Legibility may be further diminished by the shape of the retro-reflective surface; any deviation from flat will likely distort the shapes and inhibit recognizing a letter or understanding an abstract shape's orientation. Similarly, spatial AR systems offer a visual acuity that is a function of not only the display resolution, but also the distance that a user stands from the display surface.

2.4.2 Contrast Sensitivity

The same cases apply to contrast: real environment versus virtual objects, optical see-through versus video overlay. Contrast can be difficult to measure for optical see-through displays, given the uncontrolled nature of the real environment. One may wish to consider the contrast ratio of two objects in the real environment as seen through the optical elements, the contrast within the graphics as relayed by those optics from the display devices, or the contrast between a virtual object and an object in the uncontrolled real environment. This last measurement can be quite challenging to acquire. The contrast a user perceives through an AR display depends on the display device and the optical elements that present the image to the user. Some displays significantly diminished the brightness of the real world so that the graphics did not need to be as bright. The Sony Glasstron used an electronic, global mask behind the semi-transparent mirror that reflected the graphical display into the eyes and through which light from the real environment passed. Thus the mask reduced brightness and contrast in the real world, in the hope of increasing the contrast between the real world and virtual objects. Mobile AR applications have an especially challenging task; sunlight is far brighter than any AR display (and than the user would want the AR display to be), so the mask or a similar filter is a critical element of a successful optical see-through display for outdoor use.

Video overlay AR displays have the same cases of real and virtual imagery to consider, but since all are presented on the same display elements and relayed through the same optics, the optical paths are not quite as different as in optical see-through. Contrast in the real environment will be limited by the dynamic range of the camera before the image is sent to the display elements, much in the way that the Glasstron mask or optical filters were meant to do for optical see-through displays. Although video overlay offers the possibility of matching the brightness of the real and virtual objects, this can be quite challenging to measure the real environment in real-time and reproduce the appearance. HMPDs and spatial AR often (but not always) require dark rooms, reducing the contrast in the real environment.

2.4.3 Color Perception

Just as a computer monitor and a printer have a gamut of colors that they can produce, so too do the display elements in AR displays (both optical see-through and video overlay) have a gamut of colors. Optical AR displays suffer from the partial transparency of graphics over an unknown real-world background; the combination can significantly change the color perceived by the user. A methodology to acquire measurements of this hue shift and measurements for an optical AR display are presented in Sect. 3.3.3. The perception of the color of real objects can be distorted by the electronic mask described above; in addition to reducing the apparent contrast, it may introduce a consistent tint to the colors and the context in which they are seen. By virtue of dimming the entire world, colors will appear quite different. Because of the contextual nature of color, the dimming and any hue shift can affect the virtual objects on the display surface as well as the real objects behind the display surface. Clear optics in theory can avoid these problems.

Color is highly contextual, and thus knowledge of the background and surrounding (2D) visual field, as is available in video overlay AR displays, can be extremely helpful in selecting colors for discriminability on the display. Video overlay AR displays are heavily dependent on the camera and display element for the richness of the color presented to the user. Between the cameras' limited range and the display elements' gamut, there can be a complex chain of modulation between a real object and the eye of the user. The virtual objects would have a shorter, simpler path to the eye, but they will still be affected by the context. Of course, if the video image passes through the graphics memory, then it can be inspected for values and the effect of the context can be mitigated by "pre-distorting" the colors of virtual objects according to an inverse function of the distortion. Not all commercial video overlay AR displays pass the camera image through the graphics processor, however.

2.4.4 Stereoacuity

Numerous challenges for proper perception of stereo imagery have been noted for head-worn AR displays, of both optical see-through and video overlay varieties. Few displays offer even some of the adjustments that are optimal for comfortable perception of stereo imagery: adjustment of the IPD, adapting the vergence angle, and alignment of the left-eye and right-eye displays. The rendering software should also adjust the IPD to match the display (which preferably will match the user). It can correct the alignment of the displays for the two eyes (as discussed in the next section). Software correction of the vergence angle is possible, although it potentially introduces perspective distortion into the imagery if the hardware does not match the angle. This latter capability is rare in head-worn display hardware.

Of course, hand-held displays generally ignore these issues, but binocular disparity is a powerful cue for depth at large distances (perhaps to several hundred meters), so the 2011 emergence of auto-stereo displays for mobile phones should encourage developers to consider these effects for hand-held displays as well.

3 Measurements of Visual Capabilities in AR

There have been a modest number of experiments to measure visual capabilities with head-worn AR displays. The general observations made at the end of Sect. 3.1 motivated or were discovered by the studies described in this section. We summarize these experiments and software (in approximate chronological order within the four experimental goals) and discuss some practical difficulties in collecting these measurements with AR head-worn displays.

3.1 Visual Acuity

A test of four optical see-through AR displays [40] investigated the smallest real targets visible from one meter with the display off and with the display showing a blank screen. The latter condition implied that the display emitted some light and, in the case the Sony Glasstron PLM-50, enabled the mask that reduced transmittance of the light entering from the environment. Two binocular displays showed differences in these two conditions. The Glasstron (33 ° measured horizontal FOV, NTSC resolution) allowed 1 mm targets (3.4 arcmin, or 20/69 Snellen score) with no power (mask off) but only 6 mm targets (20.6 arcmin, 20/412 Snellen) with power (and mask) on. Virtual I-O I-glasses (25 °, NTSC) enabled users to see 0.5 mm targets (1.7 arcmin, 20/34 Snellen) without power and 3 mm targets (10.3 arcmin, 20/206 Snellen) with power. A MicroOptical Corp. Clip-On CO-1 (10 °, QVGA) and MicroOptical Integrated EyeGlass (17 °, VGA) both allowed users to see 0.5 mm targets (1.7 arcmin, 20/34 Snellen), although a poorly positioned CO-1 was found to limit users to 12 mm (41.25 arcmin, 20/825 Snellen) targets.

The Augmented Reality Performance Assessment Battery (ARPAB) [18] included visual acuity tests for AR displays. It was used as a pre- and post-test of a distance estimation task in optical see-through AR with 20 subjects. With natural or corrected vision, subjects were found to range from 20/13 to 20/30 in visual acuity. The AR pre-test with a Sony Glasstron[2] (SVGA, 27 ° horizontal FOV) yielded 20/40 Snellen scores for 18 subjects; one subject scored 20/30, and one scored 20/50. After the distance estimation task, all subjects scored 20/40 on visual acuity with the Glasstron. The pre-test with a Microvision Nomad[3] (SVGA, ≈ 21 ° horizontal FOV) yielded mostly 20/30 and 20/40 scores (precise distribution not given), with one subject scoring 20/50. Only three participants scored differently on the post-test: one from 20/50 to 20/40, one from 20/40 to 20/30, and one from 20/30 to 20/50.

A Sony Glasstron LDI-D100B caused eight users with normal or corrected-to-normal vision (i.e. 20/20 or better) to drop at least one step measured with a Snellen chart (≈ 20∕30) looking through the optics of the display at the same real-world optometric chart [28]. All users scored 20/30 looking at a graphical chart. This test was extended to a 2D contrast sensitivity measure using a sine-wave chart [20]. The Glasstron notably reduced the contrast sensitivity of the user compared to his or her normal vision, though it should be noted that the contrast levels in this experiment were well below the standard for optometric exams. Notably, looking through the Glasstron at the real target was significantly worse than looking at graphical targets in the Glasstron. This Glasstron model used a similar LCD mask as described for the PLM-50. The maximum transparency was specified as 80 %, and the loss of brightness of the real world was seen in the increased contrast needed to see real targets. The Nomad 1000 also reduced contrast sensitivity both looking through the display at real targets and looking at graphical targets, but by a far smaller amount.

Video overlay AR systems mediate the real world through a camera, which limits the user to its spatial (and color) resolution. In testing a training application [4], subjects using video overlay with a camera[4] mounted to a Virtual Research V8 head-worn display (48 ° horizontal by 36 ° vertical, VGA) were found to have degraded visual acuity, but no quantitative data were reported.

---

[2] Model not reported, but given the year and reported specifications, most likely an LDI-D100B or similar.

[3] Model not reported, but given the year and reported resolution, most likely a Nomad 1000.

[4] The authors report using "an Auto Gain Control (AGC) and Electronic Light Control (ELC) Panasonic camera," with an FOV "compatible with the field-of-view of the HMD," but do not give precise specifications or models.

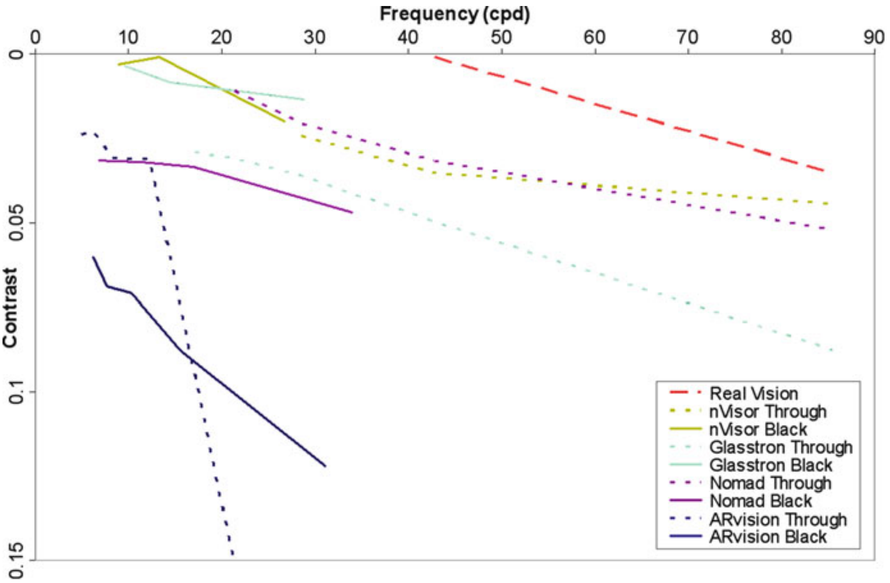


Fig. 3.4 The sampled CSF for four head-worn AR displays. The size of the target decreases as frequency (measured by cycles per degree) increases to the right. The contrast is typically graphed with lower contrast at the top of the vertical axis. In this test, the nVisorST performed well in both the see-through and graphical cases, the Glasstron provided sharp graphics, the Nomad provided clear see-through, while the ARvision struggled in both cases

Evaluation of a custom-built HMPD (52°, VGA) was done with a modified Landolt-C acuity test [7]. Users identified the location (up, down, left, or right) of an opening in a square (as described above) under three levels of light. The study found that the resolution of the display limited subjects to a visual acuity of 4.1 arcmin, or a Snellen score of 20/82 for all lighting levels. The type of retro-reflective material (needed for the HMPD) affected performance with low contrast targets.

Bringing together the design for measuring 2D contrast sensitivity and the modified Landolt-C task provided the ability to measure a portion of the contrast sensitivity function for head-worn displays. Figure 3.4 shows the sampled CSF for four head-worn displays. One observation from the graph is that the varying resolutions and measured contrasts of the displays make direct comparisons somewhat difficult. The line for each condition extends from the size of one pixel at the left; this is the smallest target for which testing was possible. The right extent denotes the size of a five-pixel target. The contrasts tested were all well below the standard for an optometric exam, and the heights of the various lines—which denote the contrast necessary to see a given size of target—are best judged relative to each other. The exception to this, however, is the natural vision condition; the monitor on which this test was performed could not exceed the capacity of the human visual system.

The optical see-through nVis nVisorST performed the best overall, with both the graphics and see-through cases yielding the closest performance to users' natural vision. The Glasstron LDI-D100B enabled high performance with the graphics. Again, the difficulty seeing the real world was due to the mask, so that greater contrast was needed to see real targets. The Nomad 1000 saw the reverse conditions; its clear optics enabled users to see roughly as small a target as any head-worn display condition. However, the monochrome red graphics did not enable the users to see virtual targets as well as the real world. Finally, the video overlay Trivisio ARvision display struggled in both the graphics condition, where the display quality is the limiting factor, and in the see-through condition, where the camera's resolution and contrast also influence the results.

3.2 Text Legibility

Leykin and Tuceryan [19] trained automatic classifiers to assess the readability of text labels over textured backgrounds. They collected training data from six human subjects, who rated the readability of 100 text-over-texture images on an eight-point Likert scale. They then tested several standard machine learning algorithms, finding that a support vector machine was the best, but that overall performance was limited by the large number of outliers (20 %) in the training data. Because the system analyzed the texture images and the features of the text itself, this method was considered appropriate only for AR systems that have video feedback (which usually is only the case for video overlay AR systems, but may be implemented more widely, as evidenced by the next experiment). Also, they limited their training and test data to grayscale presentation of text and texture.

Gabbard et al. [10] conducted a controlled study with an outdoor AR system. They studied six background textures commonly found in outdoor environments and six drawing styles, measuring response time and error for the task of identifying and reading a single numeral presented in a text sequence consisting mostly of (English uppercase) letters. Eighteen users wore a Glasstron PLM A55 bi-ocular optical see-through display (NTSC) to read text at three distances: beyond, on, and in front of a real background object (which displayed the texture). The display's focal distance was fixed (by the hardware) at 2 m; the backgrounds were (respectively) at 1, 2, and 4 m for the three distances. They found that subjects were fastest with a red brick background, perhaps because the text strings could be "framed" within a single brick. A "billboard" drawing style using blue text with a white background in the graphics rendering (which is the most opaque that the optical see-through display can be) yielded the least error among drawing styles. It was followed by a static green text color with no background in the graphics rendering. Dynamic drawing styles did not fare well, and distance of the text relative to the background real object did not show main effects.

In a follow-up study [11] using a Glasstron LDI-100B, the independent variables for the drawing of text were divided into text color (white, red, green, and cyan), drawing style (none, billboard, drop shadow, and outline), and active drawing style algorithm (maximum HSV complement, maximum brightness contrast). On each trial, users identified a letter that was repeated consecutively in one text block, and then counted occurrences of that letter in a second text block. Participants were most accurate with the building texture as the background, and least accurate using the red brick texture. A similar trend was found for response time. Text color showed no main effect on either error or response time, which was attributed to limited luminance capabilities of the Glasstron and the positive effects of the active text drawing styles. The billboard drawing style yielded lower accuracy and slower responses. Among the active styles they found a small but significant main effect; participants were most accurate when reading text drawn with maximum brightness contrast.

Renkewitz et al. [32] devised an experiment to find the optimal font size for displaying text in a head-worn display for outdoor AR applications. They used a video overlay AR system presented on an unknown HMD with SVGA resolution. They found that for a concrete wall texture, blue fonts needed a size of 8.3 points and red fonts needed a size of 8.1 points. For a bush texture, blue fonts needed 8.6 points and red fonts 9.0 points. Response times rapidly decreased until the font size was increased to 15 points, with slight gains at greater sizes. Thus they concluded that while 9 points was readable, it was not sufficient for fast reading. Selecting 2 s as the maximum acceptable reading time, they recommended fonts sizes of (concrete, blue) 23 points, (concrete, red) 42 points, (bush, blue) 34 points, and (bush, red) 29 points, equivalent to 0.92 °, 1.68 °, 1.36 °, and 1.18 ° (respectively).

Labels that overlap in an AR view become difficult to read; however, stereoscopic segmentation can perceptually separate multiple, overlapping labels [31]. Seventeen subjects read an airplane call sign in AR and, on a subsequent AR display, selected the object with that label. As the amount of overlap increased, response times increased; similarly, as the amount of overlap decreased, the error rate decreased. The viewing conditions varied disparity: ordered with depth, random with respect to depth, and constant with respect to depth. There was a significant interaction between the viewing condition and the amount of overlap. The ordered condition led to faster responses when the amount of overlap was high, implying that the targets in close proximity to other objects benefited from the ordering of disparity with depth. Dynamic scenes (including moving labels) yielded lower error rates than static scenes. An earlier experiment [30] found that correct vertical separation based on depth yielded lower error than no vertical separation and inverted vertical separation. Correct vertical separation meant that the closest target, which was lowest in the visual field, had the lowest label in the visual field. Inverted separation meant that the closest object was lowest in the visual field, but its label was highest among the labels. These two conditions did not show a significant difference in response time, and both were faster than no vertical separation.

3.3 Color Perception

In designing ARQuake [36], color selection was recognized as an important consideration. The dark colors of the original game were not conducive to display on a see-through AR display. Therefore, nine colors were tested (red, green, blue, cyan, magenta, yellow, purple, pink, and orange), each at four intensities and in four lighting conditions (standing in shade or sunlight, crossed with looking into shade or sunlight). Users assigned a Likert rating (1–10) for visibility and opaqueness. Nine color/intensity combinations were found to have a minimum score of six and mean score of at least seven in each lighting condition: three intensities of purple, two of blue, two of yellow, and two of green.

As noted above, video AR systems limit the user to the color resolution of the camera, modulated by the display's color gamut. A training system was used for testing color perception [4]. Success rate on a Dvorine pseudo-isochromatic color test for color blindness dropped from 97.3 % to 91.3 %, remained at that level during testing, and rose to 96.7 % in a post-test. Color identification dropped from 98.9 % accuracy to 62.2 % accuracy. Some adaptation occurred; after completion of the experimental task, color identification rose to 70.0 % accuracy while still wearing the AR display. Accurate (100.0 %) color perception returned after removing the display. No details were given on what constituted accuracy in color perception.

One can also quantify the perception of color through head-worn AR displays using a color naming task. The reduction of contrast in the Glasstron noted for the visual acuity and contrast sensitivity test appeared to also cause some color confusion near the white point of the CIE 1931 color space, especially when looking at real-world objects through the see-through optics [20]. Color samples near the boundaries of named regions were inconsistently labeled, with lighter colors progressively less consistent in their names. Darker colors were less salient in the graphics with a white real-world background.

Switching to a color matching task [22] gave objectivity and much greater precision in the data about color perception. Users were asked to control the chroma values (a and b of CIELAB) with a two-dimensional trackball. An initially gray target patch would change hue, and users could move through the color space in order to match the color of a reference patch. Colored bars around the target helped remind users which way to move through the space. Another helpful change was conceiving of the task in a perceptually uniform color space, such as CIELAB. With this experimental design, color error could be expressed as ΔE, a distance in color space which has been well-studied in perceptual literature for perceptually uniform color spaces such as CIELAB and CIE Luv. Setting up the matching task so that users could not receive unintended assistance by looking around the display required some careful arrangement of physical barriers. However, the result was a rich set of data.

There are two sets of graphs; the first (Fig. 3.5) shows the objective color distortion measured by a StellarNet EPP2000CXR spectrophotometer with CR2 cosine receptor. This data maps the color gamut of the display device under conditions of both see-through and graphics conditions. The graphics condition was further tested with cases of a black background and a white background; with the optical see-through, the background can heavily influence the perceived color of light that enters into the user's eye through each pixel of the graphical display. In this data, the nVisorST was shown to have only modest distortion away from the yellow-green and cyan corners of CIELAB space in the see-through condition; this was credited

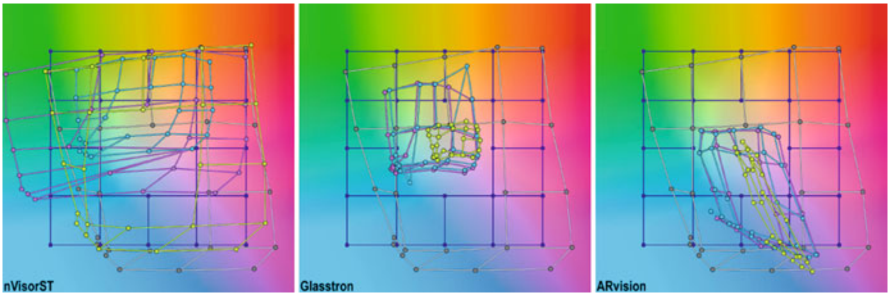


Fig. 3.5 Objective measurements of color distortion in the nVisorST (*left*), Glasstron (*center*), and ARvision (*right*) as determined by spectrophotometer measurements. Three visual conditions were of interest: see-through (*yellow graphs*)—a misnomer for the video overlay ARvision, but still a convenient label, graphics-on-black background (*magenta graphs*), and graphics-on-white background (*cyan graphs*). The see-through data should be compared to the measured color gamut of the reference monitor that was the real-world background; this data is mapped by the *gray graph*. The *blue reference grid* gives the definitions in CIELAB space of the colors sent to the displays

to the clear optics of the display, which caused little distortion of color. In the case of graphics-on-black background, the ambient light in the room pushed the objectively-measured color out of the blue half of the color space (perhaps not surprising, since blue accounts for little perceived intensity) and pulled it towards green (the strongest component of what we perceive as intensity). The white background increased the push away from blue region of CIELAB and pushed away from the red and green slightly as well. The Glasstron, again owing in part to the reduction of contrast, pulled all colors towards the center (gray) point (reducing apparent saturation of color), but also slightly towards the yellow half of the color space. The amount of distortion was approximately the same for the graphics-on-white and graphics-on-black background conditions; the pull towards the center intensified significantly in the see-through condition. Finally, the ARvision pulled the entire graph towards the blue half of the space and also reduced the saturation of colors, although less in the magenta corner of CIELAB than regions. Again, the distortion of color was similar in the case of the graphics-on-white background condition as in the graphics-on-black background condition (subject only to the display's capabilities). Analogous to the Glasstron, the video overlay (subject to the camera and display capabilities) intensified the distortion, but this in the red–green dimension of CIELAB; it slightly increased the shift towards blue but not the distortion in the blue–yellow dimension.

The second set of graphs (Fig. 3.6) shows the perceived colors. These are only compared to the monitor gamut of the reference display, the gray grid also shown in Fig. 3.5. Before examining the data, we note that user responses spanned nearly the entire portion of CIELAB graphed (a,b $\in [-100,100]$). Thus the relative patterns of distortion would seem not to be an effect of attempting to match CIELAB color specifications that are outside of normal human color perception. With that observation, we can turn our attention to the data. The nVisorST caused users to perceive

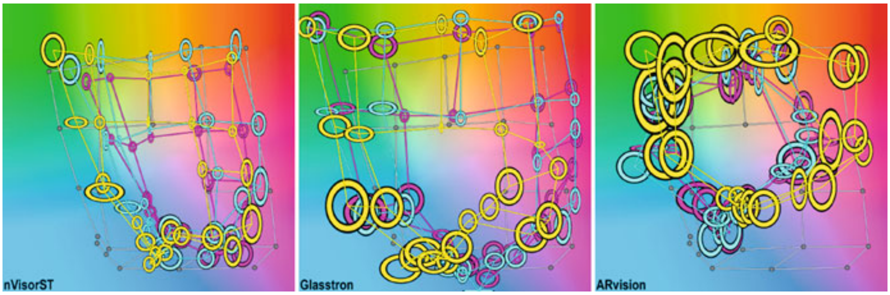


Fig. 3.6 The perceptual matching showed the color distortion in the nVisorST (*left*), Glasstron (*center*), and ARvision (*right*). The same three visual conditions are graphed with the same colors as in Fig. 3.5. All data should be compared to the measured color gamut of the reference monitor, again mapped by the gray graph. In this figure and Fig. 3.5, the CIELAB domain is represented by the L = 65 slice with a and b in [−100,100]. This portion of the domain is also depicted in Figs. 3.7 and 3.8 and thus may be used to help compare the data in all four graphs

less strength in the cyan corner in the see-through condition. There was notable variance in the answers in this condition as well, denoted by the scale of the circular data indicators. The graphics-on-white background condition showed a similar distortion away from the cyan region of CIELAB and a similar variance. The graphics-on-black condition caused further distortion away from cyan, but a little less variation in user responses. For the Glasstron, users appeared to generally overcompensate for the reduction in saturation, perceiving colors that were much further from the center of the color space than they should have. This was accompanied by an increase in the variation between users, especially as the perceived colors moved toward the outside of CIELAB. The patterns of distortion and variation were similar in the three visual conditions of see-through, graphics-on-black-background, and graphics-on-white background. Finally, for the ARvision, the stark distortion is away from the magenta corner of CIELAB; further, colors near the center of the space appeared to be perceived as more saturated than they were. The video overlay appeared to suffer from this effect more than the other two visual conditions. There was a notable increase in individual variation for nearly all colors in all three display conditions compared to the other two head-worn displays.

Gabbard et al. [13] applied the textured background from the text legibility experiments (described above) to create a testbed for measuring the effect of blending natural light reflected off real-world backgrounds with virtual light produced by an optical see-through display. They found large perceptual shifts (Fig. 3.7) between a “no-background” condition and brick, foliage, pavement, sidewalk, and white background conditions. The white background versus the remaining textured conditions showed the next largest set of changes. In terms of direction, the no-background condition pulled the colors towards the white point of the color space compared to the white background, which allowed the perceived colors to be distributed more over the color space. The foliage texture background pushed the colors away from the white point compared to the white background. The brick texture background

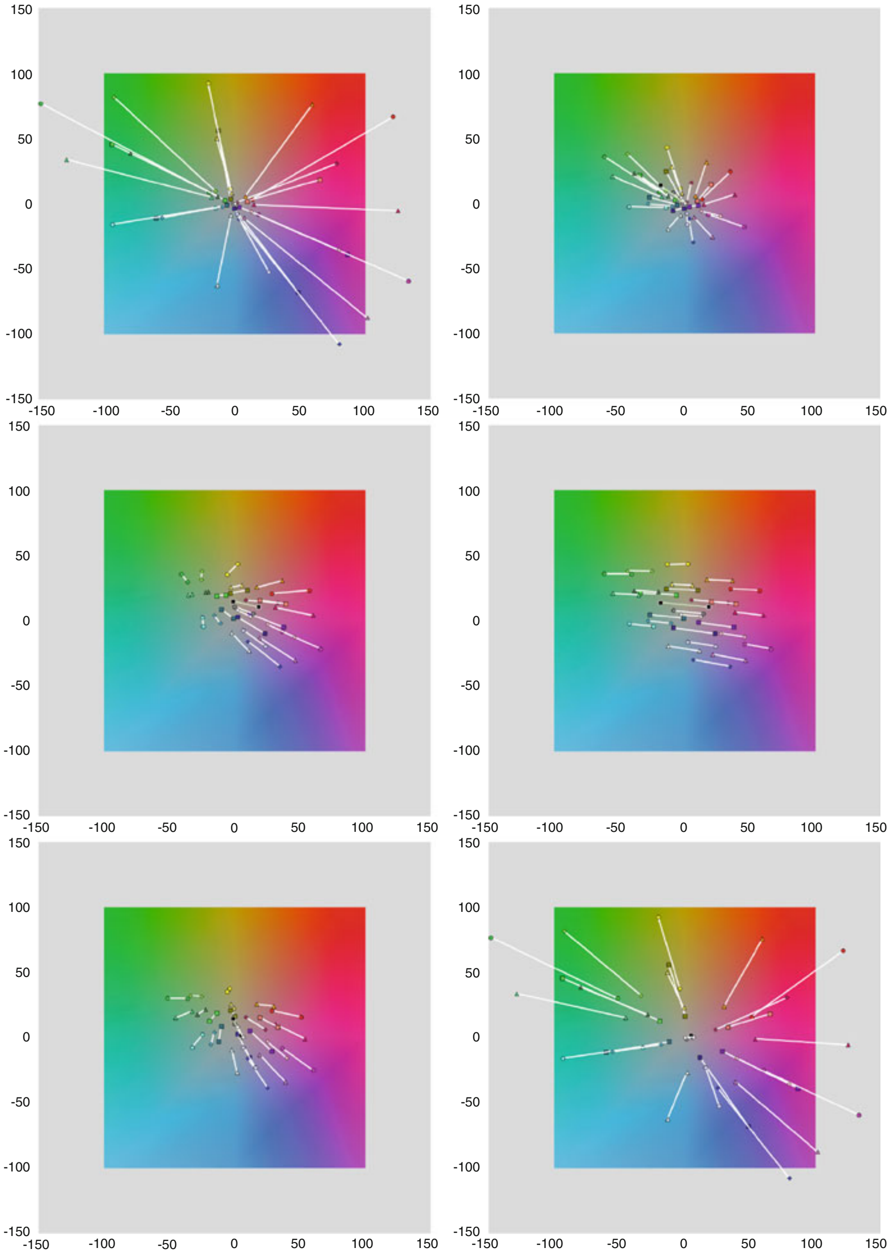


Fig. 3.7 Objective chromatic changes between textured backgrounds measured in [13], converted to CIELAB; colored portion of background corresponds to the background of Figs. 3.5, 3.6, 3.8, and 3.9. *Top left*: no-background (*outer*) versus white background (*inner*). *Top right*: foliage background (*outer*) versus white background (*inner*). *Center left*: brick background (*outer*) versus sidewalk background (*inner*). *Center right*: brick background (*right*) versus foliage background (*left*). *Bottom left*: pavement background (*outer*) versus sidewalk background (*inner*). *Bottom right*: pavement background (*inner*) versus no background (*outer*)

pushed the colors away from the green region of the color space compared to the foliage texture background, and the sidewalk texture background pulled the colors toward to white point compared to the brick texture background.

3.4 New Results on Color Perception

We present two previously unpublished studies of color perception in AR displays.

3.4.1 Farnsworth Color Arrangement

One simple strategy to test for distortion in color perception with AR displays is to conduct a standard color vision test with the AR display. This was done for a Sony Glasstron LDI-D100B, a Trivisio ARvision, and an nVis nVisorST. Before giving the results, it will help to define two types of color vision deficiency. The most common form of color blindness is poor red–green discrimination. The most common type of this deficiency is caused by a shift in the medium wavelength (colloquially, green) retinal receptors towards the long (red) wavelengths; subjects with this condition are called *deuteranomal*. It affects perhaps 5 % of the male population and is hereditary. A more rare form (0.01 % of the population) is caused by a defect in short wavelength (blue) receptors and affects the ability to differentiate blue and yellow hues; subjects with this condition are called *tritanomal*. This is also a hereditary deficiency, but shows no difference in gender. There are color vision deficiencies marked by the absence of one of the three types of cone receptors and deficiencies marked by a defect in one of the three types, as well as combinations of these, so numerous other types of color vision deficiencies exist, but these two will be sufficient to describe the results of this study.

Twenty-four subjects (eight for each display) completed the Farnsworth D-15 color arrangement test in four modalities. As a baseline case, each subject completed the test using a standard computer monitor. The subject completed a “see-through” version of the task on the same monitor, but with his or her vision mediated by the AR display. Note that this has very different meanings for the optical see-through Glasstron and nVisorST than it does for the video overlay ARvision. Two graphical conditions rounded out the set: seeing the Farnsworth test on the AR display with a white background, and seeing it with a black background. The order of these four display conditions was counterbalanced by a Latin square (repeated twice for each display’s eight subjects).

All 24 users passed the test in the baseline case, so we believe that all deviations from normal with the displays were due to the displays and their properties. With the nVisorST, all eight users tested as having normal color vision both looking through the display at the monitor (the see-through condition) and with the virtual graphics displayed over both black and white backgrounds. In all, there was no significant distortion of color perception with the nVisorST. Similarly, almost all users passed the test in all conditions with the Glasstron. One user made a minor error in the see-through condition, which for that user was the last condition mandated by the Latin square. This error resulted in a Confusion Index (C) of 1.39 and a Scatter Index(S) of 1.10 using the scoring method of Vingrys and King-Smith [37]. A value of 1.00 (in both indices) is considered normal, but it would require values near or above 2.00 in both to be considered to have a color vision deficiency. So although the user would still have been judged to pass the test, there is a hint of difficulty in the Glasstron see-through condition, although given that this was the final condition tested, fatigue may also have been a factor.

Turning to the results with the video overlay ARvision, we recall that the colors perceived are a function of the camera parameters *and* the display element's color gamut. The combined effect yielded a wide array of color vision results. We emphasize that all users tested as having normal color vision using a standard computer monitor, although minor errors occurred, as noted below. In the condition equivalent to see-through for the video overlay display—i.e. where the users saw the computer monitor through the camera and head-worn display—five users tested as normal but with Confusion Indices and Scatter Indices ranging from 1.10 to 1.69, the latter of which approaches the Confusion Index for tritanomal subjects. Another subject was scored as $C = 2.24$ and $S = 2.16$, which is consistent with being tritanomal, although the angle of crossing the circle was consistent with normal color vision. One subject did match the tritanomal profile in all three measures, and one matched the deuteranomal profile in all three scores. The results for the ARvision with the graphical test seen over white and black backgrounds were more encouraging. Five users tested as normal with the white background, while two tested as normal, but with C and S in the range of [1.11,1.65]. One user matched the tritanomal profile. With the black background, again five users tested as having normal color vision, two users tested as normal but with C and S in [1.10,1.39], and one user matched the tritanomal profile (which was the same user as for the white background).

Thus we can see that some users were transformed by some AR display conditions into partially color-blind subjects; given that these deficiencies are known to be hereditary and rare, this speaks to the limitations of the AR display (including display element for all displays, as well as the optics of the see-through displays and the cameras of the video overlay display).

3.4.2 CIELAB Measurements with Chromatic Backgrounds

We extended the experiment described above using the CIELAB space to include conditions with chromatic backgrounds; data with these backgrounds was acquired only for the nVisorST display. We extended our experimental design to four new display conditions: solid color backgrounds of (respectively) red, green, blue, and yellow. For comparison purposes, subjects again completed a natural vision condition in which there was no AR display used; they merely matched what they saw on one side of the barrier to the other side of the barrier. This baseline allows us to quantify differences and natural variation in performance of the perceptual

Table 3.1 The main effect of display condition on the distance (*ΔE* metric in CIELAB) and angle of the error (in radians)

	ΔE		Angle		Background
Display condition	Mean	SD	Mean	SD	Intensity
Natural vision	27.8397	21.8763	0.1855	1.7092	20.5
Red background	41.2221	27.9827	−0.2676	1.6762	22.9
Green background	53.7014	28.6392	−0.1540	1.0546	53.7
Blue background	44.1641	24.9506	0.8739	1.1027	13.4
Yellow background	56.6131	29.6001	−0.2579	1.4069	78.9

The absolute values for *ΔE* are not as meaningful as the relative size of the chromatic background conditions to the natural vision condition. Regarding the angular errors, it is interesting to note that only the green, blue, and yellow backgrounds caused significant differences from the natural vision condition. The intensity of the background may be a cause for the effects observed

matching task. Nineteen subjects completed the experiment with 22 color samples in each of the five display conditions.

We found main effects on both dependent measures, distance (*ΔE*) in color space and the direction of this error within the *ab* plane of CIELAB. Looking at the distance measure, we found a main effect of the display condition—F(4,72) = 80.444, p = 0.000. Not surprisingly, subjects were most accurate to the input colors in the natural vision condition. Since some of this error is inevitably due to an imperfect monitor gamut for the reference monitor, the more interesting finding is not the absolute error in this baseline condition, but that the four chromatic backgrounds all exhibited significantly higher error and higher variance in the responses (Table 3.1). Furthermore, the red and blue backgrounds exhibited significantly lower error than green and yellow. One could easily hypothesize that this may be an effect of the background intensity, although this would seem to be an incomplete explanation for the difference between the natural vision condition and the four chromatic backgrounds. There was also a significant main effect on the angle of this error—F(4,72) = 46.127, p = 0.000. The natural vision condition had a low mean and a high variance, which means the distribution of the direction of error was likely more similar to noise. The red, green, and yellow had a similar mean direction, although the green was more consistent (lower variance) than the other two, and the blue background induced an entirely different direction and was also somewhat more consistent about this effect than the red or yellow backgrounds. The variances are so high that it is unclear whether the mean angles for the various conditions are meaningful measures.

Figure 3.8 shows a plot of the mean and standard deviation for each color sample of this data in the same spirit as Fig. 3.6. The center of the circle is at the mean location for each color sample's response, and the scale of the circle in the *a* and *b* dimensions is a linear function of the standard deviation of the subject responses. While these graphs showed that we tried to sample colors away from the achromatic central region, there is often distortion further away from the center than the natural vision condition. This illustrates both the increased *ΔE* and the patterns of angular error. In this graph, we can see that the mean angle for the blue background diverges well away from the other conditions towards the yellow region. One might expect to

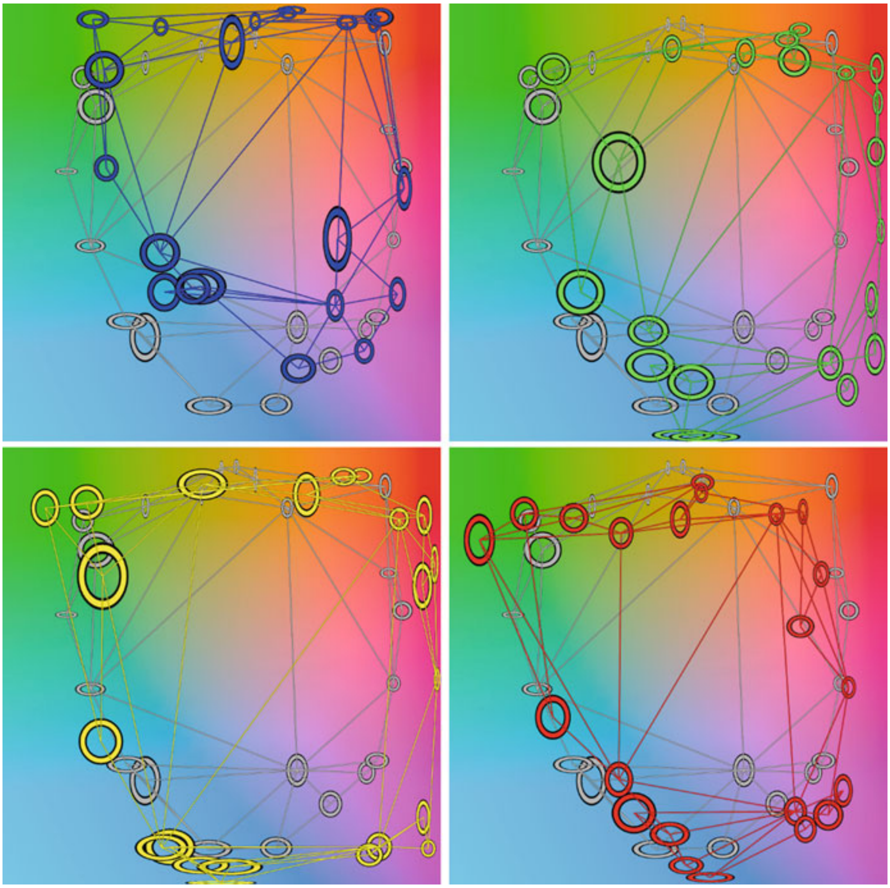


Fig. 3.8 The perceptual matching showed the color distortion in the nVisorST for four canonical chromatic backgrounds of solid color: *blue* (*upper left*), *green* (*upper right*), *yellow* (*lower left*), and *red* (*lower right*). Each graph is shown on CIELAB space and includes a gray reference grid of the matching as completed with no mediation by any AR display

see such an effect, pushing users towards the opposing color in human perception. This effect is not seen consistently with the other chromatic backgrounds, although it does apparently appear for colors near the opposing color with the green background (right side of the plot) and yellow background (bottom side of the plot). In order to produce hues that would be recognized by subjects as having "typical" appearance for the nominal background colors, we did not equalize the intensity of the background colors. Also, we note that the intensity of black (listed as the background for the natural vision condition in Table 3.1) was greater than the intensity of the blue background seen through the nVisorST display optics. While it is possible the optics did cause this difference as measured by the color meter, it is curious to observe such numbers. Thus we could hypothesize that the low intensity of the blue background is the cause of the unique direction for the mean error in this condition. One could also formulate theories about the saturation of the background as the cause of any of the differences observed.

We noted a significant main effect of sample color on ΔE—F(21,378) = 23.879, p = 0.000—and on angle F(21,378) = 14.362, p = 0.000. While the graphs indicate that various samples were moved by different distances and different directions, it is hard to make general statements about the patterns beyond those made above. Further, there were significant interactions between the display condition and the color samples for both ΔE—F(84,1512) = 4.301, p = 0.000—and on angle F(84,1512) = 3.492, p=0.000. These serve to solidify our assertion that the pattern of errors is quite complex and deserves further data collection and analysis. We also measured response time; we found a significant main effect of color sample—F(21,378) = 4.276, p = 0.000—but not of the display condition—F(4,72)=0.386, p=0.818. While we may have expected such an effect based purely on the starting condition for each trial of a gray field to match to a chromatic sample, there was no apparent pattern of the response time as a function of the distance of the color sample from the origin of the *ab* plane. We saw a significant interaction between display condition and color sample—F(84,1512) = 1.556, p = 0.001—but defer interpretation until such time as the main effect can be explained.

This type of data may be used to adapt the displayed colors so that they will be perceived as intended and appear as if matched to the lighting of the real environment [39]. Two important considerations are adaptation of the user over time and the system latency that may be introduced by pre-distorting the colors. If the latter must be done on a per-pixel basis to account for the background of each pixel, then the compensation algorithm may be an expensive rendering process. An implementation that takes full advantage of the programmable nature of the modern graphics processing unit (GPU) may alleviate this.

It is natural to compare the various backgrounds (black, white, see-through, red, green, blue, and yellow) for patterns to see what appears to behave similarly and what appears to behave differently. Figure 3.9 compares the data from all background conditions, displayed over CIELAB color space (Fig. 3.3). Recall that green shades are on the left, while red shades are on the right; blue shades are at the bottom, while yellow shades are at the top. We see that the green and blue backgrounds generally caused users to shift away from that color in the matching task (though exceptions exist). The yellow and red backgrounds seemed to cause users to shift away from the achromatic center of the slice of the color space. The colored backgrounds generally caused larger errors than the achromatic and see-through conditions, which saw little consistent pattern of errors. One may perhaps see in the bottom (blue) half of the reference sample space that the shift was toward the yellow half, and vice-versa. However, this data is rather sparse, and strong, consistent patterns would appear to require more data.

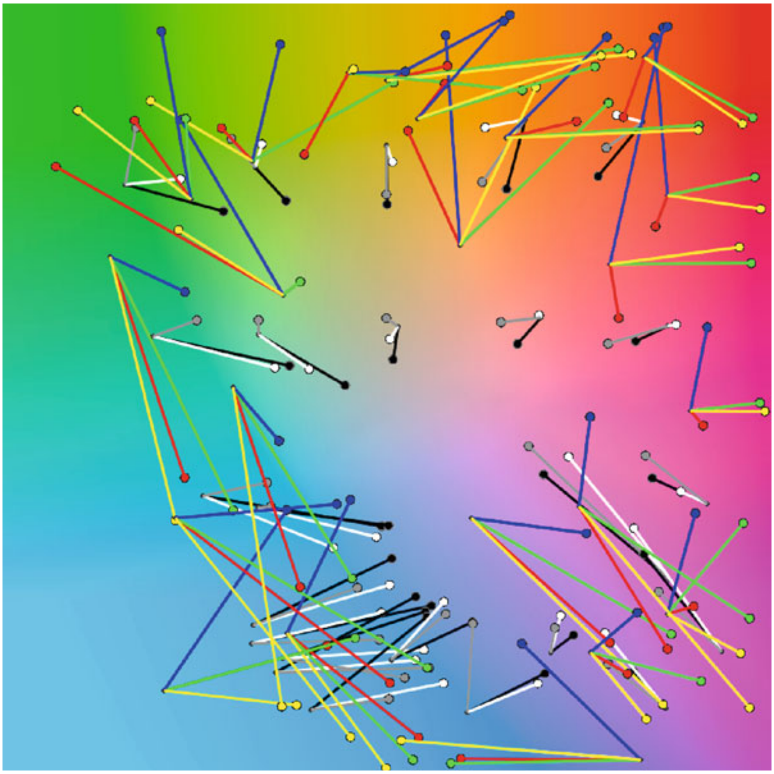


Fig. 3.9 In comparing the direction and magnitude of the color matching error under the various background conditions for the two color matching experiments, we see a few patterns emerge in CIELAB color space. The reference color is denoted by *small, dark gray circles with no outlines*, whereas the *circles with outlines* represent the mean color error (*ΔE*, represented by distance and angle in color space) in user matching. The *achromatic circles with gray or black outlines* indicate the *black* or *white* backgrounds according to their inner color, with the *gray inner circles with black outlines* indicating the see-through condition. In the first test, we used a set of 24 reference samples. The *colored circles with black outlines* indicate the colored backgrounds (*red*, *green*, *blue*, and *yellow*) from the second experiment, which used a new set of 22 reference samples

3.5 Stereoacuity

In order to perceive stereo images, the human visual system must fuse the input from the left and right eyes into a coherent picture of the three-dimensional world. Among the many issues created by stereo optical see-through displays is vertical disparity in the graphics relative to the real world. Such disparity will cause *diplopia* (double vision) for the user of a head-worn AR display, and even if the visual system manages to compensate for misalignment, eye strain and headaches are likely from extended use of a device with this disparity. This disparity was measured using nonius lines for a set of Glasstron displays [24] and corrected with three algorithms.

Two modified the six degree-of-freedom offset between the eyes to include (1) a vertical component to the IPD translation or (2) a pitch rotation; (3) correction by image shift was also studied. Notable variability between different devices of the same manufacturer and model were noted, and the correction methods did not yield equivalent visual angles, indicating the tolerance of the human visual system to adjust to errors (despite the potential for fatigue effects).

While correction of such vertical disparity is a necessary condition for proper perception of stereo, it is not sufficient for understanding the stereo capability provided by a head-worn display. By testing the depth of a virtual target to the depth of a real reference object, one can measure the stereoacuity users experience with an AR display. For two custom-built HMPDs, stereoacuity measurements were recorded in pilot tests [8]. With a 52 ° diagonal FOV and 640 ×480 graphical resolution, five subjects were determined to have stereoacuities between 1.1 and 6.2 arcmin with the real reference at 80 cm, between 1.4 and 3.0 arcmin with the reference at 150 cm, and between 0.6 and 0.8 arcmin with the reference at 300 cm. With a 42 ° diagonal FOV and 800 ×600 graphical resolution, five subjects were determined to have stereoacuities between 1.1 and 1.7 arcmin with the real reference at 80 cm, between 1.2 and 2.6 arcmin with the reference at 150 cm, and between 0.4 and 1.7 arcmin with the reference at 300 cm.

Stereoacuity also may be measured through application of a depth-matching task with horizontal disparity as the independent variable. This disparity normally gives the impression of depth to the human visual system. A test task of matching the apparent depth of a virtual target to a real reference object (provided on a monitor visible through the HMD) was applied to the nVisorST [21]. The results showed that subjects generally achieved a stereoacuity between 1.9 and 3.9 arcmin; this may be compared to the typical stereoacuity of 1 arcmin, although hyperacuity for stereoscopic depth can reach 3 arcsec. As may be inferred from the discussion above, lower contrast and smaller size of the object decreased the stereoacuity (i.e. raised the detection threshold measured in arcminutes). Regression of the mean disparity for each subject versus the subject's IPD showed an excellent linear fit, indicating that users were able to convincingly verge the real and virtual objects.

4 Discussion

Summarizing a diverse array of experiments such as those described above is destined to be a difficult task. But from each set of experiments, we learned important lessons about how head-worn AR displays affect basic perceptual capabilities of the human visual system. It is also important to note differences in the methods use to collect data.

First, we note the importance of evaluating the quality of seeing both the real environment and the virtual objects that the AR application places within it. This merging of real and virtual worlds is the fundamental characteristic of AR, and as such should be taken into account in any evaluation of the perceptual quality of an

AR technology or application. Thus the see-through condition for optical see-through displays and the background video in video overlay displays are critical conditions to be evaluated, just as the display elements and optics that (respectively) generate and mediate the view of the virtual objects are obvious targets of evaluations. It immediately follows that identification of the limiting factor in an optical or video AR display (display elements, optics, masks, and cameras, as applicable) is of great importance to the AR system designer. It is also worth noting that we reviewed primarily work on head-worn displays of optical see-through and video overlay types, adding comments about alternative displays in the few locations where data exists. As hand-held displays become more popular, there will be an increasing need for these types of evaluations to be conducted with hand-held displays.

The second defining feature of AR is the *registration*, or alignment, of the virtual objects to the real environment. While the geometric measurements in modeling and tracking the objects and the user are central to registration, being able to accurately discern the virtual and real objects is also a prerequisite. If a user is to be able to understand the relationship between real and virtual with a reasonable degree of precision, then it stands to reason that basic measures such as contrast sensitivity (incorporating both size and difference in brightness or color) contribute to the understanding of whether users will perceive objects to exist in the merged AR world.

A chief application of AR is to convey information about the surrounding world; as such, many useful AR applications overlay text on the real environment. If this text is to be useful, then it must be legible when presented in an AR display. While the contrast sensitivity measure will indicate this in an abstract context, reading language operates at a level above raw distinction of which (tiny) regions of an image (whether real or virtual) compose a letter; familiarity with a language breeds an ability to understand words without seeing all the letters. Thus the raw resolution required may be too strict a requirement; in the world of small mobile devices, this may be exploited to the benefit of the human visual system.

One underlying theme we detect in the conduct of the experiments described here is that there are many experimental tasks that AR can copy from the perception literature and everyday contact with specialists in aspects of perception (e.g. optometrists). Virtual eye charts can be traced to the early days of virtual environment research [16]. But the perception research literature may have superior recommendations over the clinical practices we experience in our personal lives. The measurement of contrast sensitivity versus visual acuity is one example of this. Measuring the precise distortion of color is another; this type of data is far superior in value to the results of standard color vision testing for the display manufacturer who works to improve the color quality of an AR display. Careful consideration of the task design may become more critical in shifting the emphasis of work to mobile displays; the style of use may not be as conducive to adapting optometric examinations as the style of use of head-worn displays.

Our emphasis on an iterative evaluation process is justified by the results of most of the basic perception tests. The experiments studied showed how AR displays frequently limit the basic perceptual capabilities of the users. But users may not always be able to identify the precise nature of such problems. In the BARS application mentioned above, we could identify numerous situations in which the display of text would help the user. When we built prototype applications and showed them to users, many of the negative comments indicated the difficulty subjects had seeing in displays, but how to address low visibility of text that is overlaid optically on top of a bright real environment with a combination of size and contrast is not likely to be prescribed by end users. At the same time, even something as simple as the time required to read text was shown to vary with the color contrast and size, which clearly has implications for the application.

Color presents an array of issues for AR displays. We reviewed data showing the color distortion that occurs with both optical and video displays. The former must compete with the uncontrolled background, and have their color gamut severely altered by the optical conditions of the display and the background. In this area, far more data is likely to be needed before detailed correction functions can be derived for the diverse set of circumstances proposed for AR applications, especially mobile applications. Even with regard to video overlay displays (which, in the form of mobile phones, are increasingly the choice for mobile applications), the issues of color contrast are not solved, although the greater control of the final image that reaches the eye has enabled greater progress. Here, both objective measurements taken with a spectrophotometer (also known as a color meter) and subjective measurements collected from human subjects contribute to our understanding and towards a solution. The contextual nature of color—in which adjacent colors and intensities affect the perception of the neighboring colors and intensities—implies the need to acquire both objective and subjective measurements, as well as an understanding of the application context.

With regard to stereoacuity, we saw two critical issues. First was the potential for improper stereo to lead to eye fatigue and headaches. While studies have also demonstrated the tolerance of the human visual system to errors in stereo displays, these factors clearly limit the use of improperly calibrated stereo displays. As AR is still considered a strong candidate for medical and manufacturing applications in which work is to be done at close distances, stereo would seem to be an important feature to correct in AR displays. The second critical issue is that the displays still limit the human visual system from applying the binocular cues for scene understanding, as evidenced by the studies conducted.

Perhaps the most important lesson to be learned from this review is the sparse amount of data that has been collected on these fundamental questions for AR display and thus AR systems. Replication and extension to new devices, filling in the gaps in data collection, and designing compensation or correction algorithms would all benefit the field. We also encourage AR researchers who conduct evaluations of AR applications to learn from the lessons taught to us by our users: AR applications may fail to meet expectations (of users and/or designers) for reasons that range from the "high-level" application soundness down to "low-level" issues of basic perception. Evaluators would be wise to conduct studies of basic perception when looking for the reasons an application fell short. Improved hardware will surely improve the results of the studies discussed here. But, as several studies showed, clever use of the limited resources can overcome the perceptual challenges and lead to greater application success.

References

1. Brown, D.G., Stripling, R., Coyne, J.T.: Augmented reality for urban skills training. In: IEEE Virtual Reality, pp. 249–252 (2006)
2. Bruce, V., Green, P.R., Georgeson, M.A.: Visual Perception: Physiology, Psychology, and Ecology, 3rd edn. Psychology Press (1996)
3. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: W. Epstein, S. Rogers (eds.) Handbook of Perception and Cognition, Vol. 5: Perception of Space and Motion, 2nd edn., pp. 69–117. Academic Press (1995)
4. Darken, R.P., Sullivan, J.A., Lennerton, M.: A chromakey augmented virtual environment for deployable training. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (2003)
5. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Human Factors **37**(1), 65–84 (1995)
6. Farnsworth, D.: The Farnsworth-Munsell 100-hue and dichotomous tests for color vision. Journal of the Optical Society of America **33**(10), 568–578 (1943)
7. Fidopiastis, C., Fuhrman, C., Meyer, C., Rolland, J.: Methodology for the iterative evaluation of prototype head-mounted displays in virtual environments: Visual acuity metrics. Presence: Teleoperators and Virtual Environments **14**(5), 550–562 (2005)
8. Fidopiastis, C.M.: User-centered virtual environment assessment and design for cognitive rehabilitation applications. Ph.D. thesis, Dept. of Modeling and Simulation, Univ. of Central Florida (2006)
9. Gabbard, J.L., Hix, D., Swan II, J.E., Livingston, M.A., Höllerer, T.H., Julier, S.J., Brown, D., Baillot, Y.: Usability engineering for complex interactive systems development. In: Human-Systems Integration Symposium (2003)
10. Gabbard, J.L., Swan II, J.E., Hix, D.: The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. Presence: Teleoperators and Virtual Environments **15**(1), 16–32 (2006)
11. Gabbard, J.L., Swan II, J.E., Hix, D., Jung Kim, S., Fitch, G.: Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In: IEEE Virtual Reality, pp. 35–42 (2007)
12. Gabbard, J.L., Swan II, J.E., Hix, D., Lanzagorta, M., Livingston, M.A., Brown, D., Julier, S.: Usability engineering: Domain analysis activities for augmented reality systems. In: The Engineering Reality of Virtual Reality (SPIE Volume 4660), pp. 445–457 (2002)
13. Gabbard, J.L., Zedlitz, J., Swan II, J.E., Winchester III, W.W.: More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In: IEEE Virtual Reality, pp. 79–86 (2010)
14. Ginsburg, A.P., Hendee, W.R.: Quantification of Visual Capability, pp. 52–71. Springer-Verlag (1992)
15. Hering, E.: Outlines of a Theory of the Light Sense. Harvard University Press, Cambridge, MA (1964)
16. Holloway, R., Fuchs, H., Robinett, W.: Virtual worlds research at the University of North Carolina at Chapel Hill as of February 1992. In: Visual Computing: Integrating Computer Graphics with Computer Vision, pp. 109–128. Springer-Verlag (1992)
17. Ishihara, S.: Tests for Colour-blindness. Hongo Harukicho, Handaya, Tokyo (1917)
18. Kirkley, Jr., S.E.H.: Augmented reality performance assessment battery (ARPAB): Object recognition, distance estimation and size estimation using optical see-through head-worn displays. Ph.D. thesis, Instructional Systems Technology, Indiana University (2003)
19. Leykin, A., Tuceryan, M.: Automatic determination of text readability over textured backgrounds for augmented reality systems. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 224–230 (2004)

20. Livingston, M.A.: Quantification of visual capabilities using augmented reality displays. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 3–12 (2006)
21. Livingston, M.A., Ai, Z., Decker, J.: A user study towards understanding stereo perception in head-worn augmented reality displays. In: IEEE International Symposium on Mixed and Augmented Reality (2009)
22. Livingston, M.A., Barrow, J.H., Sibley, C.M.: Quantification of contrast sensitivity and color perception using head-worn augmented reality displays. In: IEEE Virtual Reality, pp. 115–122 (2009)
23. Livingston, M.A., Brown, D., Julier, S.J., Schmidt, G.S.: Mobile augmented reality: Applications and human factors evaluations. In: NATO Human Factors and Medicine Panel Workshop on Virtual Media for Military Applications (2006)
24. Livingston, M.A., Lederer, A., Ellis, S.R., White, S.M., Feiner, S.K.: Vertical vergence calibration for augmented reality displays. In: IEEE Virtual Reality (Poster Session) (2006)
25. Livingston, M.A., Rosenblum, L.J., Julier, S.J., Brown, D., Baillot, Y., Swan II, J.E., Gabbard, J.L., Hix, D.: An augmented reality system for military operations in urban terrain. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (2002)
26. Livingston, M.A., Swan II, J.E., Julier, S.J., Baillot, Y., Brown, D., Rosenblum, L.J., Gabbard, J.L., Höllerer, T.H., Hix, D.: Evaluating system capabilities and user performance in the battlefield augmented reality system. In: Performance Metrics for Intelligent Systems Workshop (PerMIS'04) (2004)
27. Livingston, M.A., Zanbaka, C., Swan II, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: IEEE Virtual Reality (Poster Session), pp. 287–288 (2005)
28. Livingston, M.A., Zanbaka, C.A., Swan II, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: IEEE Virtual Reality 2005 (Poster Session), pp. 287–288 (2005)
29. Peli, E.: Contrast in complex images. Journal of the Optical Society of America A **7**(10), 2032–2040 (1990)
30. Peterson, S.D., Axholt, M., Ellis, S.R.: Label segregation by remapping stereoscopic depth in far-field augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 143–152 (2008)
31. Peterson, S.D., Axholt, M., Ellis, S.R.: Objective and subjective assessment of stereoscopically separated labels in augmented reality. Computers & Graphics **33**(1), 23–33 (2009)
32. Renkewitz, H., Kinder, V., Brandt, M., Alexander, T.: Optimal font size for head-mounted-displays in outdoor applications. In: International Conference on Information Visualisation, pp. 503–508 (2008)
33. Riggs, L.A.: Visual acuity. In: Vision and Visual Perception, pp. 321–349. John Wiley and Sons (1965)
34. Smallman, H.S., Boynton, R.M.: On the usefulness of basic colour coding in an information display. Displays **14**(3), 158–165 (1993)
35. Sutherland, I.E.: A head-mounted three-dimensional display. In: 1968 Fall Joint Computer Conference, vol. 33, pp. 757–764. Thompson Book Co. (1968)
36. Thomas, B., Close, B., Donoghue, J., Squires, J., Bondi, P.D., Piekarski, W.: First person indoor/outdoor augmented reality application: ARQuake. Personal and Ubiquitous Computing **6**(1), 75–86 (2002)
37. Vingrys, A.J., King-Smith, P.E.: A quantitative scoring technique for panel tests of color vision. Investigative Ophthalmology and Visual Science **29**(1), 50–63 (1988)
38. Ware, C.: Information Visualization: Perception for Design, 2nd edn. Morgan Kaufmann (2004)
39. Weiland, C., Braun, A.K., Heiden, W.: Colorimetric and photometric compensation for optical see-through displays. In: Intelligent and Ubiquitous Interaction Environments, Universal Access in Human-Computer Interaction, part of HCI International, LNCS Volume 5615, pp. 603–612 (2009)
40. Woods, R.L., Fetchenheuer, I., Vargas-Martín, F., Peli, E.: The impact of non-immersive head-mounted displays (HMDs) on the visual field. Journal of the Society for Information Display **11**(1), 191–198 (2003)

Chapter 4
Pursuit of “X-Ray Vision” for Augmented Reality

Mark A. Livingston, Arindam Dey, Christian Sandor, and Bruce H. Thomas

1 Introduction

One of the most intriguing capabilities envisioned for augmented reality (AR) systems is the notion of “X-ray vision,” or the ability to virtually “see through” one surface to what in reality is hidden from view. Many AR systems have been premised on this feature for the primary value provided by the application. Furness’ pioneering work [15, 16] on head-up displays for pilots was motivated in large part by the ability to virtually see through the solid cockpit walls and floor to a virtual copy of the real world hidden by the aircraft infrastructure. The ultrasound AR system [2] provided the ability to understand the data acquired by the ultrasound probe from inside the body; to do this requires a metaphor that ensures the user will understand the data to reside behind the visible skin surface, and in fact this was a problem noted early in the development of the system. We believe that the metaphor “X-ray vision” was first applied to this work [45].

There are two sides to this unique perceptual capability. There is the issue of the absolute distance of graphical entities within the coordinate system of the real environment. There is also the relative order of real and virtual surfaces within the merged environment. Neither of these perceptual capabilities seem to come naturally for AR users, and thus numerous visual metaphors have been conceived in order to give the impression of relative and absolute depth. Another issue that challenges the designer of X-ray visualization metaphors is the potential to overload the user with information [20, 28]. If the entire database of known objects is shown

---

M.A. Livingston (✉)
U.S. Naval Research Laboratory, Washington, DC 20375, USA
e-mail: mark.livingston@nrl.navy.mil

A. Dey • C. Sandor • B.H. Thomas
University of South Australia, Adelaide, SA, 5001, Australia
e-mail: Arindam.Dey@unisa.edu.au; chris.sandor@gmail.com; Bruce.Thomas@unisa.edu.au

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_4,

(generously assumed to be accurately registered to the real world), then the user could easily be overwhelmed with information and be unable to understand the depths of any object.

Developing X-ray vision AR systems is a difficult problem from a number of perspectives. First, X-ray vision is truly an unnatural act. This is not an ability people perform without the aid of perceptual and cognitive mechanisms, and as such there are few *metaphors* to which people can relate. Therefore, careful consideration of depth cues and their usability, as well as the introduction of new cues, are required. Second, the presentation of X-ray vision information to the user is more difficult than presenting depth in traditional three-dimensional (3D) graphics or even virtual environments (VE). In the latter, there is a rich history of experiments showing that depth is underestimated. This is due to the fact that the system is not in complete control of the visual information received by the user. This chapter will explore the perceptual background, visual metaphors, and empirical investigations into overcoming this challenging and important topic in AR research. We conclude with an analysis of the research and suggestions for further research.

2 Perceptual Background

Our perception of depth draws on a number of cues. These cues interact with the distance from us; they also interact with the properties of AR displays in unique ways. Cutting divided the environment into three regions: personal space (generally, within arm's length), action space (a distance at which one can reliably and accurately interact with other entities), and vista space (anything beyond action space).[1] Figure 4.1 shows his ordering of the cues within each of these regions. This section consists of a brief review of each depth cue and how they can be affected by AR displays. For more complete reviews of depth perception, we refer to reader to textbooks on perception [5, 42] or chapters [6, 7, 34] which detail depth cues, their relative strengths, and variations noted from the general patterns depicted in Fig. 4.1. Readers will also note that some authors delineate and categorize cues differently; we use the categories of Cutting [6] to match Fig. 4.1. We also note that depth cues do not, in general, function in isolation; *cue combination* leads the human visual system to understanding of depth. Our analysis here is limited to individual cues.

A few terms will be helpful in the following discussion. The *human visual system* should be understood as consisting of the complete pathway that begins with the retina, travels up the optic nerve, passes through the lateral geniculate nucleus, and ends in the visual cortex. Interested readers may consult a recommended textbook for details, but the important concept is that all of these components play a role in perception, and in depth perception in particular. Measurements taken in the retina

---

[1] Personal space, action space, and vista space are commonly termed as near-field, medium-field, and far-field distances, respectively. In this chapter, we will use the latter set of terms.

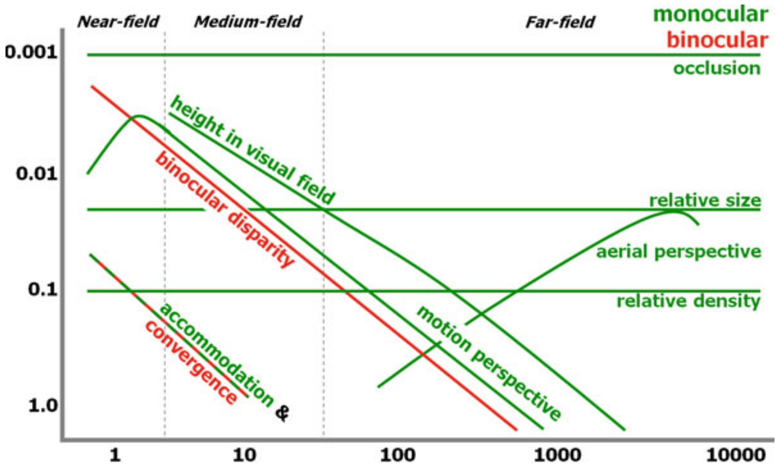


Fig. 4.1 Depth cues and the depth discrimination threshold (where smaller numbers indicate more potent cues) in the three fields of depth. In Heiko Hecht, Robert Schwartz, and Margaret Atherton, eds., Looking into Pictures: An Interdisciplinary Approach to Pictorial Space, top half of Fig. 11.2, page 223, ©2003 Massachusetts Institute of Technology, by permission of The MIT Press. This chart is modeled on one drawn by Cutting [6] using data from Nagata [34] and Cutting

will be compared along the pathway and processed in the cortex to interpret depth. *Ordinal depth* refers to identification of the relative depth of objects: which is closest, which is next after that, etc. *Metric depth* indicates that a measurement (e.g. in meters) can be ascertained by the observer, at whatever level of accuracy the observer is capable. Most of the cues are *monocular*; they apply to a single eye. Two cues are *binocular*; they require two eyes to be applied to a scene or a visual system. Figure 4.1 shows monocular cues in green and binocular cues in red. When considering distances in the environment, we can classify them as *egocentric* distances, indicating a distance from the observer's viewpoint, or *exocentric*, indicating a distance between two objects within the field of regard. All of the work in AR has studied egocentric distance for the AR user.

2.1 Occlusion

When a solid object prevents light rays from another object from reaching the observer directly, we define this as a case of *occlusion*. If the closer object, known as the *occluder*, blocks only some of the rays from the more distant, or *occluded*, object, we call this *partial occlusion*. This cue provides information about ordinal depth amongst the objects, but not about the metric depth. As depicted in Fig. 4.1, occlusion is generally the most powerful depth cue in any region of space. That this cue provides such powerful depth information derives from the fact that we are accustomed to a world populated by solid objects that are opaque. Of course, we do encounter transparent or translucent objects regularly, and situations such as windows have become sufficiently common that we incorporate them smoothly into our understanding. Other translucent media will be discussed under the aerial perspective cue.

Using occlusion as a depth cue also relies on an assumption often known as Helmholtz's rule: the contour of the occluder does not change direction where it intersects the contour of the occluded object. Such a coincidence would likely cause a misinterpretation of the depth order and/or the shapes of the respective objects. (The phenomenon of false contours gives another example of a possible misinterpretation.) One can imagine the difficulty of interpreting shape in a world made up of objects that are composed of wireframe outlines, with no solid surfaces. In such a world, the ambiguity of occlusion ordering would be high. Another assumption should be obvious but is worth noting for the discussion below. One must also be able to differentiate the two objects—i.e. there must be a difference in the brightness or color of the occluder and occluded object so that the contour is detected by the visual system.

Optical see-through (OST) AR displays often prevent the use of the occlusion cue by the human visual system. The optical combination of real and virtual is generally achieved through a partially-silvered mirror. Thus for every pixel in the virtual image, there is a real solid angle behind it that will show through the combiner. Wherever graphical objects are visible, the graphics do reduce the salience of the real world. Thus the graphics cannot be ordered in depth relative to the real objects merely with the occlusion cue. As noted above, for certain translucent objects we experience everyday, this situation may become sufficiently familiar to understand the geometric relationships. However, for most AR users of optical displays, the situation has yet to become familiar. A few research prototypes of optical displays have been demonstrated [23] with the capability (usually via a "masking" display in the optical path to the real world) to occlude the real world at pixels (or blocks of pixels) that are part of virtual objects. At this time, we are unaware of commercially-available optical displays with this feature. On the other hand, video-overlay AR displays can overwrite the video pixels completely and thus occlude the real world with the virtual objects. This is true both for those that use a form of chromakey replacement and those that use a graphical frame buffer and render directly onto the image. The occlusion cue is thus available in AR systems that use video displays. Also, since this cue is completely reliant on the 2D projection of the world onto the retina, there is no difference between head-worn and hand-held displays for this cue. All AR applications will suffer some loss of fidelity of this cue if they are unable to accurately align the graphics to the real environment, an error in *registration*.

2.2 Binocular Disparity

Because of the displacement between the two eyes, each eye sees a slightly different image of an object; in particular, there is a shift in the position on each retina. This translation is known as *binocular disparity*, and from it the visual system can compute the distance to the object. In terms borrowed from AR (or more precisely, from computer vision), this cue relies on a calibration of the distance between the eyes. Our visual system learns this calibration through everyday experience of the world. As also known in computer vision, the visual system must identify corresponding points on the object in order to compute the disparity. This cue interacts with convergence, discussed below, since that also can affect the apparent translation on the retina. As shown in Fig. 4.1, binocular disparity is a very strong cue at close distances; however, because the amount of translation decays with increasing distance (due to perspective foreshortening), its strength diminishes as an object moves farther away.

Providing correct binocular disparity provides a challenge for head-worn AR displays, but recent commercial offerings are able to meet this requirement. Many older systems did not permit the adjustment of the inter-pupillary distance (IPD) of the display hardware. Thus, the AR user would experience an incorrect binocular disparity without careful calibration of the device's fit for each user. For optical displays, measurement of the IPD and setting the hardware and software rendering to match provides correct binocular disparity. For video displays, the camera mountings on the display must be included on the display hardware mounting and moved in concert. For hand-held displays, there is little choice but to experience the disparity that is dependent on the distance to the display surface, which is generally held within the near-field (especially if one accepts the limitation to arm's length). Some mobile phones are now using autostereoscopic displays, which is an example of the possibility for a hand-held display to enable the user to experience binocular disparity that is dependent on the virtual distance to a virtual object rather than the real distance to the display surface. In this case, a calibration of the IPD and the distance to the display will permit proper binocular disparity for virtual objects. Another fundamental issue for depicting binocular disparity is the limited resolution of the display elements used for the virtual imagery in optical displays and both real and virtual imagery in video displays. Given that the visual angle occupied by each pixel barely supports normal human vision (cf. Chap. 3), it stands to reason that the retinal measurements of disparity suffer from this limited resolution as a consequence of the display hardware limitation.

2.3 Motion Parallax

The relative movement on the retina of stationary objects caused by egomotion of the observer is known as *motion parallax*. The similar term *motion perspective* is sometimes used as a synonym and sometimes applied to depth perceived for objects moving near or around an observer (whether parallel or perpendicular to the central view ray, or at any angle). In general, one may infer the relative motion and depths of objects moving relative to the observer. This cue is quite powerful in the near-field, and—in an analogous fashion to the decay in strength of binocular disparity—decays with increasing distance to the objects due to the relatively small motion on the retina that a distant object's relative motion will cause. Figure 4.1 shows another curious effect for motion parallax (and motion perspective, if considered separately), but one which has a simple explanation. The cue decreases in effectiveness at extremely close distances, because relatively large translations on the retina are difficult to track. Additionally, rapid motions on the retina do not lend themselves detection of differences, so interpreting relative depth of two nearby moving objects through motion perspective is challenging. Thus the graph of the sensitivity for this cue (Fig. 4.1) is not monotonic.

Since motion parallax is a purely geometric cue, both optical and video AR displays are capable of depicting the necessary information to make use of this cue. Of course, this statement assumes that the problems described under binocular disparity do not prevent understanding of the relative motion on the display surface. The limited resolution of the display hardware—and the camera of video displays—can present a significant problem for both nearby and distant objects, but most AR applications avoid displaying extremely close objects. Given the observation of the loss of fidelity at extremely close distances, this avoidance is probably advisable for proper depth perception. The loss of resolution for distant objects makes this cue ineffective for some of the studies described below. Hand-held displays convey this cue in theory, but the limited angular field of view (FOV) may limit the ability to apply this cue. Noise in the tracking that creates dynamic registration error will interfere with interpretation of smooth motion and thus with this cue.

2.4 Height in Visual Field

We are generally accustomed to seeing object sit on the ground. Imagine a desktop with objects sitting on the planar surface and the observer's eye above the plane. Then the *height in the visual field* would be a simple function of the egocentric distance from the observer to the objects and the perspective projection parameters of the human visual system. Just as with the innate calibration of the IPD, humans have an understanding of their height from the ground plane. However, this knowledge can be clouded by the situation (e.g. uneven ground). In addition, visual acuity decreases the precision of the measurements with increasing distance needed to turn height in the visual field into metric depth. Thus while ordinal depth may be computed via height in the visual field, metric depth presents more of a challenge, and the utility of this cue decreases with increasing distance to the object. Further, the sharp observer will note that Fig. 4.1 omits this cue from the near-field. This is based on the assumption that an observer's central view ray is parallel to the ground. Given the vertical limits of the human FOV (in particular, approximately 45° below horizontal), this means that an object must have a minimum distance equal to the height of the observer's eye in order to be in the FOV. Clearly, the observer can simply look down to remedy this situation for a nearby object (at the obvious cost of limiting the distance objects can be). In unusual situations such as looking down from a balcony, this cue has analogies (e.g. "height"—distance—away from the side of the building) that are likely of little practical value.

Height in the visual field is also a purely geometric cue, so AR displays—both hand-held and head-worn—generally handle this accurately. Again, limited display resolution will limit the utility of this cue in providing metric depth information for graphics in all AR displays and for real objects in video AR displays. Also, the limited FOV may move the effective region of this cue even farther from the user than the beginning of action space. Many head-worn displays do not offer a large vertical FOV. Hand-held displays, inasmuch as they are held farther from the eye than head-worn displays rest, suffer even greater loss of the effective distance of this cue. Registration error, especially vertical misalignment to the correct real-world location, will further reduce the usefulness of this as a cue in AR.

2.5 Relative Size

Artists have long understood the basic relationships of linear perspective: distant objects occupy a smaller area on the retina than nearby objects of the same size. Thus the *relative size* of the nearby object on the retina is larger. If the objects are recognized (e.g. another person, or a standard door frame) or their absolute size is otherwise known, then ordinal and metric depth may be ascertained from this cue. The key concept is that this is a *relative* cue; there must be a basis for comparison, either within the scene or from experience. Thus a tree, which may be quite recognizable but have an uncertainty associated with its height or with the breadth of its branches, provides a less accurate (but still useful) cue through relative size. The threshold to see a difference from relative size is higher than to discern occlusion, so while relative size is effective throughout the visual fields, it is less effective than occlusion, and in the near-field and medium-field, less effective than the other cues discussed above (Fig. 4.1).

Relative size is also a purely geometric cue, and as such is in theory achievable by both hand-held and head-worn AR displays. But like the other geometric cues, display and camera resolution as well as registration error can reduce its utility.

2.6 Relative Density

We turn our attention to a monocular geometric cue that is quite similar to relative size. A cluster of objects or features in a texture will have a characteristic spacing on the retina which is called *relative density*. This is perhaps most easily conceived as analogous to a checkerboard pattern used in computer vision calibration techniques. If such an object were of known size and placed in an AR environment, then the observer would be able to infer its distance by the perspective foreshortening effect on the texture. Generalize this to any pattern of objects for which a comparison can be made—again, either to another object (or group of objects) in the scene or to something from experience—and one has successfully constructed this cue. Relative density can also yield both ordinal and metric depth perception. It is less effective than relative size, perhaps by almost an order of magnitude (Fig. 4.1).

As relative density is a purely geometric cue, the only observation necessary for AR displays is that, owing to its lower effectiveness, it would seem likely to be even more sensitive to display (and camera) resolution and, potentially, dynamic registration error than relative size.

2.7 Convergence

The other binocular cue is the angle between the central view ray of the two eyes, a measure of *convergence*. By convention, large angles are those needed to fix nearby objects in both retinal images, whereas a focus on the horizon reduces the angle to 0° (parallel view rays for the two eyes). This cue interacts with binocular disparity, since both change the retinal position of objects. The human visual system is calibrated by experience to understand the difference, since convergence is an *oculomotor* cue—i.e. dependent on the muscles of the eye—whereas binocular disparity is not. As Fig. 4.1 indicates, this cue is effective only in the near-field and slightly into the medium-field. An exercise in trigonometry should convince the reader why this cue's effectiveness drops so dramatically with distance, but can be used to provide metric depth for nearby objects. The angular resolution of the oculomotor system would need to be very high.

This cue is typically completely lost with AR displays. Hand-held displays force the user to converge on a single display surface; thus any information about distance is to the display, not to the virtual objects depicted. Head-worn displays that provide separate display elements for each eye could in theory automatically adjust this angle or change the imagery in order to compensate appropriately for the geometric adjustment, but in practice, no displays do this. Because it is an oculomotor cue, it is nearly impossible for an AR display to correctly stimulate the visual system with this cue. For a video AR display, the cameras would have to move in concert with the oculomotor changes of the eyes. McCandless and Ellis [32] presented a quantitative analysis of eye position during changes in convergence, concluding that accurate depiction of a fixed stimulus distance in a binocular display requires real-time compensation for the observer's eye movements.

2.8 Accommodation

The eye naturally focuses on an object of interest; the change in the depth of the focus distance is called *accommodation*. A static focus distance may not provide much information, but changes do, albeit only in the near-field and some of the medium-field (Fig. 4.1), exactly equal to the effectiveness of convergence (hence the graph of their effectiveness looks to be dashed). This cue's effectiveness also decreases with increasing age. The information provides metric depth in the near-field.

This cue is also typically completely lost with AR displays. Most graphics (AR and other) are rendered with a pinhole camera model, thus creating infinite depth of field; thus, all objects in the FOV are in focus. While rendering methods exist for out-of-focus rendering, they can be computationally expensive and are thus rarely used in interactive systems. Post-processing methods can compute and add the proper amount of blur [24], but these are also rarely used. While recent research has demonstrated an optical display with multiple focal planes [25], this seems to be a long time from entering commercial markets. Video AR displays rely on the camera to provide the accommodation cue (i.e. focus) for the real environment, and generally cannot change this setting in real time. In theory, a real-time focus mechanism and an eye tracker could be coupled together via software control, but we cannot cite an example of this being implemented.

2.9 Aerial Perspective

We began our discussion with the strongest depth cue, the occlusion of distance objects by intervening objects and noted that translucent objects produced variation in the interpretations of depth. However, if the human visual system can attribute the appearance of an object to intervening translucent media, such as atmospheric haze, fog, or even “clear” air, then it invokes the cue of *aerial perspective*. The uniqueness of this cue is seen in two ways in Fig. 4.1. First, it is generally ineffective in the near-field and medium-field; second, the dominant trend is for its effectiveness *increase* with distance, until at great distances near the limits of the human visual system, it suffers a slight reversal of this trend. Both of these features may be explained by the amount of intervening media. Until there is a sufficient amount, the cue is ineffective, and as more intervening media accumulates, the effectiveness increases until the object is no longer discerned and the cue can no longer be applied. It is exceptionally difficult to characterize all instances of this effect; for example, an extremely thick fog may provide significant information in the near-field and block information beyond a few meters into the medium-field. The more typical case is that which is graphed in Fig. 4.1.

Graphics hardware has offered simulations of this effect for many years, although its use in AR is undocumented. Moreover, for this cue to be effective, it should match the real environment, which would require measures of the atmospheric media at and up to the application-critical distances from the user. While in theory such measurements are obtainable, in practice they would seem unworthy of the attention given that the maximum strength of this cue is less than relative size, and only approaches that level at great distances, beyond where most AR applications have required (or demonstrated) capability of accurately registering graphics. The nature of this effect would also seem to work against optical AR; this cue, when accurately depicted, will make objects’ apparent colors mix with the intervening media. This may be difficult to differentiate from the optical mixing of colors of virtual objects with the background colors. Video AR can overcome this mixing, but still may need to account for the color context to properly convey the subtle color shift of virtual objects due to this cue. Hand-held and head-worn display would appear to have little difference, although if the application wants to depict a thick fog in the near-field, the distance of the display from the user's eyes may also be critical to proper depiction of this cue.

3 Depth Estimation Protocols

In this section, we briefly review the tasks that have been used to study the perception of egocentric distance. A more complete review may be found in the perception literature [31], which will include protocols for measurement of exocentric distances. We note the applicability to AR for each measurement protocol.

3.1 Verbal Estimation

Perhaps the most straightforward way to measure an observer's perception of distance is to have the observer report the distance in some standard measurement metric, most commonly feet or meters. This protocol should be understood to include non-verbal methods of entering the data, such as typing a number into a keyboard. The key aspect is that the stimulus is constant, and the observer's response is (in theory) a continuous value (although the observer will discretize the response). The advantage of this protocol is that it can be applied in nearly any situation; it was used in several of the experiments discussed in Sect. 4.5. An inverse method of this protocol could be implemented by instructing an observer to place an object at a given egocentric distance. These methods are *open-loop*, since there is no feedback available to the user to determine whether an answer is accurate while the judgment is being made. This method requires that the user have a strong sense of the distance metric, which is a skill that an observer may not have acquired or practiced.

3.2 Forced-Choice

Another simple and widely-applicable method to measure perception is to ask the observer to choose between a small number of alternatives, such as which of two (or some small number) of objects is closest. This is, of course, a widely-used technique in user studies in general. Variations are easily-imagined and reasonable for evaluation tasks; for example, an observer may be asked whether a target is in front, between, or behind a set of reference objects distributed in depth. This protocol again offers the advantage of having few encumbrances in AR systems. However, it implies that ordinal depth measures are being collected, which may or may not be the desired protocol. This again is an open-loop task.

3.3 Perceptual Matching

A common method of estimating the metric distance of an object is to have the observer place it at the distance requested. The observer is given some method of controlling the distance to the object, and adjusts the location until it matches that of the reference object. This is a *closed-loop* task, because the user can see the changes being made to the target as adjustments are applied and thus receives feedback about the changing relative size. Since such motions are in themselves a depth cue, this visual feedback clearly informs the user as to the quality of the match. When implemented in an AR system, this implies the need for some interaction with the system. This can be as simple as manipulation with a mouse, or could even task the form of verbal commands such as “closer,” “farther,” and “stop.” This method is quite popular among the AR studies discussed in Sect. 4.5.

3.4 Reaching

For distance judgments to objects in the near-field (which is often defined as that the observer can reach), the user may simply be asked to reach out and touch the object. This can be done accurately even when the user is blind-folded. Thus a *blind reaching* protocol may be employed to determine the egocentric distance. This is in a sense a form of perceptual matching if performed while the observer can still see the object whose distance is being perceived, so a blind-fold and object removal are often employed in a reaching task. There remains the possibility of tactile feedback if the user is able to touch a table or other objects continuously or at intervals along the path to the object’s location. The user should be informed if the object has been removed while the blind-fold is in place, so that the lack of such feedback is not a surprise and does not induce an increase in the estimated distance. This technique is well within the capabilities of most AR systems, although the physical encumbrances such as various wires and hardware components of many AR systems must be carefully placed in order to avoid unintentional tactile cues. We are aware of only one study that has implemented this in AR up to the current time, however. One should also be very careful with the implementation of video AR systems; if they displace the camera viewpoint outside the eyes, the resulting distance perception will be heavily affected by the offset between the cameras’ viewpoints and the eyes’ viewpoints [4], assuming a stereo system. A monocular camera system used for biocular (same image to both eyes) display is sure to cause additional problems for the binocular depth cues. This offset is especially important in the near-field, where this estimation protocol can be applied.

3.5 Walking

People are quite good at walking to a location of a previously-seen target. The user can be shown a target and (similarly to reaching), after removal of the object and putting on a blind-fold, walk to the target's location. This task is often used in the perception literature and has been applied to AR. The physical encumbrances of AR systems can come into play, so care must be taken that the user will not trip over wires or other equipment. Further, it has been shown that the perceived distance can depend on the effort to reach the target [50], so if a heavy head-worn AR display or a backpack loaded with devices to support the AR system is being used, this may result in distorted distance perception. Even a hand-held device, if it tires the user's arms, may induce fatigue and thus distort this measure of perceived distance. The caution regarding effort does apply to any of the distance measures discussed here, but seems most applicable to systems where the user knows that walking to the target is imminent. Clearly, this measurement protocol is most applicable to "reasonable" distances, so the medium-field seems to benefit the most from such a measure. However, care must be taken (using AR or not) to make sure that—if the subject returns to the same starting position for each trial—the return path does not give subjects unintended feedback about their accuracy. This is especially true if subjects remove the blindfold between trials.

A variation on this concept is for the subject to conduct *imagined walking* to the location of the target object. This has also been used with success in the perception literature [35]. The subject is asked to start a timer (e.g. a stopwatch or a computer-implemented timer) as he imagines beginning to walk to the target. Then the subject stops the timer after the time has elapsed that he believes he would need to reach the target. This yields (perhaps surprisingly) an accurate estimate of the distance; the only need is to "calibrate" the timing data by having the subject (actually) walk at his or her normal pace for a known distance before the trials begin. This time can then be used to convert the time for each trial into a distance estimate. The factors of fatigue and effort may play a role in this estimated time.

3.6 Throwing

Another visually-directed action that has been successfully applied in the perception literature is throwing a small object that won't bounce (e.g. bean bag) to the target's location. While this clearly requires a minimum level of motor skill on the part of the observer, it has been shown to be as accurate as walking to a target's location. With regard to AR, this again requires a low level of encumbrance from the AR system in order to engage in the action of throwing. Assuming the object to be thrown is light and easily held in the hand, this would appear to be an easy requirement to meet. Mobile AR systems, however, may find some difficulty in applying this, as will those AR systems (mobile or otherwise) that employ hand-held display devices. The combination of walking and throwing has been used with success in the perception literature.

3.7 Triangulation

As a substitute for walking great distances, a user can walk a short distance at an oblique angle to the object (typically at an angle between the line of sight to the object and a direction perpendicular to the line of sight). The subject can be asked to point continuously at the object while walking on this path, known as *triangulation by pointing*. Alternatively, the subject may be asked to walk (without pointing) until told to stop; upon stopping, the subject is asked to turn and face the target or turn and begin walking toward the target. The direction indicated by the pointing, facing, or walking is then used to triangulate the measurement of perceived distance. This can reduce the effort and time associated with a walking protocol. These protocols all work well for AR systems, albeit with the caveats expressed above regarding physical encumbrances. The weight that the subject is carrying would seem to be less of a concern, since the walking distance is presumably not going to be very far; however, fatigue can still be a factor. One difficulty with this protocol in general is the low precision that results for larger distances. If there is a large range of distances being studied, the subjects may have greater capability for precision with the nearby distances, an argument that can be verified with simple trigonometry.

3.8 Size and Motion

The final methods of measuring perceived distance come directly from the application of depth cues and are thus thought to be less susceptible to interference from the cognitive level processing in the human visual system. The relationship between the size of an object and its distance is simple to express with trigonometric functions. This implies that the perceived size of an object implies an assessment of the perceived distance to that object. Thus asking a subject to assess the size of a distant object provides an indirect measure of the subject's perception of the distance to it. Of the above protocols, verbal estimation, forced-choice, and perceptual matching variations are easy to construct for this measurement.

In a variation of the triangulation protocol, a subject can be asked to view an object while moving. The object of known size serves as a reference for judging the distance that the observer moved; another simple trigonometric relationship converts the distance moved to a perceived distance to the object. The distance judgments through this protocol tend to be lower than through perceiving size.

Both of these protocols are amenable for AR systems. The resolution of the graphics may affect the perception of size and of displacement due to motion, as discussed above for the cues of relative size and motion parallax. For the motion estimate, the caveats regarding moving (expressed above for other protocols in which the observer moves) apply.

3.9 Summary of Protocols

As one can see, there are a number of protocols that have been used in the perception literature to judge the egocentric distance to objects. These protocols generally may be applied to AR systems; the primary limiting factors are the limitations for the depth cues themselves, noted in Sect. 4.2, and potential limitations on ease of movement for protocols that require the observer to move. Reasonable (if not easy) accommodations have been demonstrated in AR depth perception experiments summarized in Sect. 4.5. Interested readers should refer to the original papers for details. Loomis and Knapp [31] summarize the application and relative performance of these various measurement protocols in the perception literature. Some of the evaluations in AR described below studied the accuracy of estimates made with multiple protocols.

4 Visualization Metaphors

In response to the perceptual challenges, AR system designers have devised several visualization metaphors to convey ordinal and metric depth information. Some metaphors attempt to emulate the appearance of the real world within the virtual environment in order to allow the user to take advantage of the real-world cues to understand the depth relationships of virtual objects within the real world. Other methods introduce synthetic cues to convey depth relationships with graphical parameters instead of relying on analogies to the real world. Still other applications have simply presented the graphics in superposition and counted on other depth cues to overcome the conflict in occlusion. Some methods were developed in the context of mobile AR systems, which presents some unique design challenges (cf. Chap. 5), and we discuss applicability of the metaphors for a variety of displays and application scenarios. We review these methods in approximate order of their appearance in the literature; however, we allow that some techniques were not published upon first usage. In addition, it should be recognized that graphics in *superposition* over real surfaces have long represented occluded surfaces.

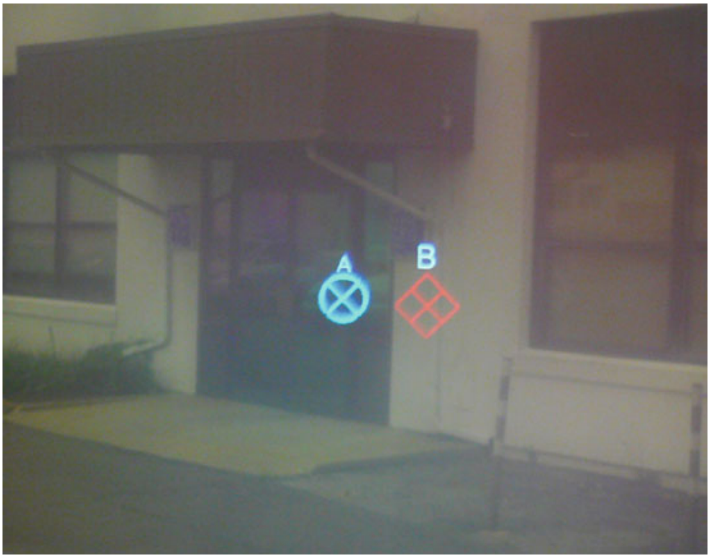


Fig. 4.2 Setting transparency of a virtual surface as a function of distance simulates the atmospheric perspective cue for the human visual system

4.1 Opacity

Perhaps the most natural metaphor for X-ray vision is to depict a surface that occludes others as being partially transparent [12]. If the observer can accept the (physically unrealistic) premise that the real surface has turned translucent, then the presence of the graphical objects can convey the correct depth ordering between the real and virtual surfaces. This becomes as direct an implementation of “X-ray vision” as the technology permits, although there is a significant qualitative difference between the appearance of the (real) occluder and (virtual) occluded surface. The technique may be extended to any number of hidden surfaces, and a filtering aspect can be added by enabling a user or an automated control mechanism for various hidden surfaces. The technique may also be inverted by setting transparency of a virtual surface as a function of distance from the observer [29]. This variation does not require modification or explicit representation of the real surface which hides the virtual object’s location, although it does benefit from it. Current 3D graphics hardware—specifically four-byte color and z-buffer depth ordering—simplifies the implementation of this technique considerably. This variation of the technique is illustrated in Fig. 4.2.

The challenge of this metaphor for human perception is that it can explicitly break the occlusion cue. Both the graphics corresponding to a hidden object and the real surface must to a certain degree be visible in the correct direction. Another natural interpretation is that the graphics, being visible while in competition with a real surface, must therefore be in front of the real surface, which is exactly the opposite of the geometric arrangement that X-ray vision condition attempts to convey. This technique is affected by the display capabilities as well. Most OST AR displays are not capable of occluding the real world with graphics on the display. Thus, all graphics are translucent and thus interact with the metaphor on which this technique tries to build. Video overlay displays can completely overwrite the video pixels with graphics, however. The cost of this capability is that the depicted occlusion relationship is thus completely reversed from the desired understanding. This gave rise to the next metaphor.

4.2 "Cutaway" or "Virtual Hole"

As noted above, despite the dimness of transparent graphics representing occluded objects, they often appeared to be in front of the real surfaces that were logically intended to hide them. This is especially true when the virtual surfaces are represented with wireframe outlines or small filled surfaces, whereas the occluder is a larger, smooth surface. Because the human visual system has a preference for continuous surfaces, the real surface's continuity "pushes" the graphics forward in front of the real surface [11], creating precisely the opposite of the intended perception.

This conflict can be resolved by creating more graphical representations to make it clear that the first set of objects in fact lies behind the real surface. The metaphor of a "cutaway" [12] or a "virtual hole" [2] can be used to create a context for the graphics. The hole breaks the real surface in a way that permits the virtual graphics to be perceived at the intended depth. A cutaway can take the form of the X-ray vision metaphor in the Superman comics which gave rise to the name for the capability, simply making it clear that the occluder is interrupted. (Figure 4.3 shows an implementation from the AR system for laparoscopic visualization [14].) Or such a virtual hole may have a full 3D structure to it, with sides and a virtual bottom that is clearly behind the virtual graphics that fit inside the hole [2, 41]. Either of these metaphors overcomes the perceptual preference for continuity of the surfaces and conveys the ordinal depth relationships, a fact verified by an experiment employing a physical hole [11].

4.3 Stipple

One classic illustration technique to represent hidden surfaces is to depict visible surfaces with solid lines, represent a set of hidden surfaces with dashed lines, and perhaps

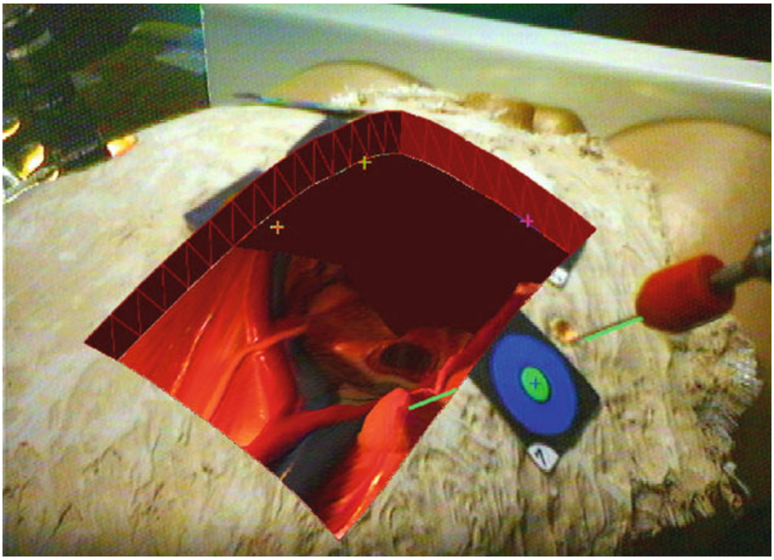


Fig. 4.3 The virtual hole metaphor in the AR system for laparoscopic surgery visualization [14]; image courtesy Department of Computer Science, University of North Carolina at Chapel Hill

extending to represent a more distant set of surfaces with dotted lines. This builds upon the depth cue of relative density of texture in the scene to create the impression of depth. Such stipple effects have been a tool for technical illustrators for many decades. They were among the earliest techniques introduced in AR for representing hidden surfaces as well [12]. Because simple stipple patterns have been a technique in hand-drawn illustration for so long, they are among the techniques implemented for drawing lines in computer graphics hardware as well. Extensions of this technique to filled polygonal representations of surfaces are also featured in graphics hardware. Figure 4.4 shows an example of this technique using filled polygons.

One difficulty with such techniques, however, is that the hardware implementation of the patterns typically uses screen-space or window-space addressing of the coordinates for the origin of the stipple pattern. This means that if the object moves or the view position or orientation changes even slightly, the stipple pattern will move with respect to the object. Thus the stipple pattern will appear to “crawl” along the lines or surfaces. While this does not prevent the user from understanding the depth relationships, it is likely to be distracting. Programmable graphics hardware allows for the implementation of an object-space stippling pattern [28].

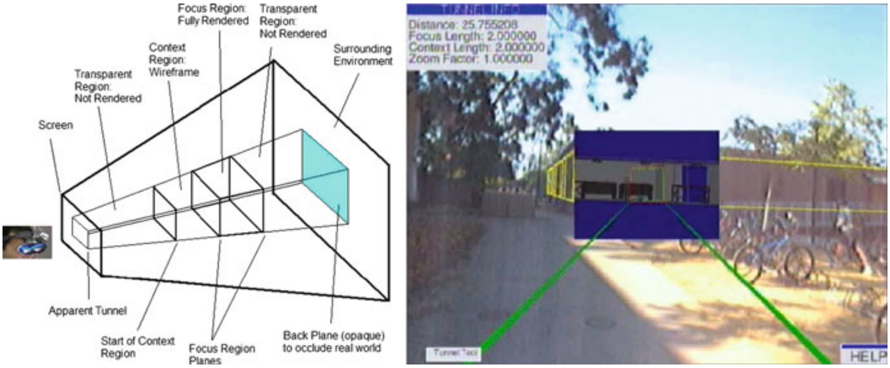
Stippling patterns would appear to suffer from the converse problem as the opacity-based representation: the stipple is by definition a non-continuous representation of a continuous surface. This visual cue can therefore be reconstructed into something continuous (which is the intention), but then also brought forward in depth (counter to the intention). The familiarity of many people with assembly instructions and other forms of technical illustration mitigates this conflict

Fig. 4.4 Using stipple effects available in graphics hardware can provide a depth cue. The decreasing density of the stipple pattern can indicate increasing distance from the viewpoint. This can be reminiscent of the relative density depth cue. One can see that overlapping silhouettes cause interference in the density of stipple patterns, limiting the effectiveness of the cue

somewhat; it is simply a convention that dashed lines are behind solid lines, and (for most people) that dotted lines are behind dashed lines. So the technique succeeds in spite of some perceptual ambiguity.

4.4 Shadow Projection

Height in the visual field—i.e., distance from the horizon—is a direct cue from the virtual object about its depth within the environment, based on perspective properties of the rendered display. An indirect form of this cue is the position of a shadow cast on the environment [46]. Understanding the shadow may also require understanding the size and shape of the virtual object, as these will also affect the form of the shadow. Once the shadow is understood to be caused by the object, it gives the observer some information about the position of the object above the ground (on the line from the light to the shadow) and thus the distance from the user; this is of course assisted by the existence of shadows cast by real objects, so that the location of the light source(s) may be reliably inferred. Accurately casting a shadow from a virtual object onto the real world requires a precise model of the real surface onto which the shadow must be cast, an accurate characterization of the light source(s), and the computational resources to compute the geometric interaction of the light(s) with the object and the surface. While these have been within reach of graphics hardware for a number of years [44], relatively few systems implement such cues.



Distance: 25.755208
Focus Length: 2.000000
Context Length: 2.000000
Zoom Factor: 1.000000

Fig. 4.5 The virtual tunnel extends the metaphor of the virtual hole to multiple surfaces, creating the effect of a tunnel. *Left*: Diagram showing the four regions of space used in rendering the tunnel. *Right*: An example of the tunnel from the user’s point of view. Images courtesy of Tobias Höllerer and the University of California at Santa Barbara [3]

4.5 *Virtual Tunnel*

Interactive X-ray vision can be achieved through a virtual tunnel [3]. In this tool, a frustum was formed with the user’s position at the apex and rendered rectangles parallel to the view plane. Occluded objects are rendered inside that frustum (Fig. 4.5). This type of rendering provides a feeling of looking down a tunnel, hence it is called a tunnel tool. Inside the tunnel there are four regions created by three planes. The first region, starting from the user’s position to the occluder, remains transparent. The second region is called the Context region, where the occluder is rendered with wireframe outlines, to provide the user a context of what she is looking through without providing too much information. Then the third region, called the Focus region, renders the occluded object with filled shapes. This is the region in which the user is interested. The fourth region, ranging from the end of Focus region to the opaque back plane, is again transparent. The back plane is used to occlude the real world. The user can slide the whole set of planes (regions) to get an interactive X-ray vision tool.

4.6 *Ground Grid*

Horizontal relationships, i.e. the context, between an occluder and an occluded object can be clarified by a ground grid [49]. Here, a virtual grid on the ground is presented in the occluded region, and the user can determine the size and location of the area hidden by an occluder by interpreting the grid lines. Grid lines may consist of concentric circles (Fig. 4.6) or a rectilinear grid corresponding to some

Fig. 4.6 A ground grid consisting of concentric circles can assist the user with resolving the depths of virtual and (if registration is accurate) real objects in the scene. This image is a concept sketch from 2002 and incorporates a line-based stipple technique

external coordinate system. While this metaphor can resolve the relationship, it may increase visual clutter and, to some extent, decrease the visibility of the occluded location.

4.7 Image-Based Techniques

Most of the previous X-ray vision metaphors presented occluded objects as floating on top of the occluder due to the naïve overlay of synthetic data on top of the real world imagery. Inspired by Focus and Context visualization techniques, Kalkofen et al. [21] introduced the notion of context-preserving X-Ray vision. This technique controls the removal of real-world information based on edge maps of the image of the occluder. Avery et al. [1] extended the idea to mobile AR. They detected and overlaid edges of the occluder on the occluded virtual objects to increase the perceptual reference between occluded objects and occluders (Fig. 4.7, left). The idea of edge-map based X-ray vision was extended and a closer understanding of human perception was employed by Sandor et al. [39] to design an X-ray vision based on multiple *saliency-maps* (Fig. 4.7, right). Along with the edges of the occluder, this X-ray metaphor additionally preserved hue, luminosity, and motion as salient features. Image analysis may reveal other salient features that can be used to communicate that a

Fig. 4.7 Two types of image-based techniques: (*left*) only Edge-map based and (*right*) three additional saliency-map based

single layer of real surface exists in front of the virtual objects superimposed on them [51]. Textures and highly saturated colors may produce visual saliency in the way that edges do; this can be exploited to determine what pixels should have an extra overlay on them to convey the proper depth ordering.

4.8 Tram Lines

Tram lines emerge from a similar concept as the ground grid; they give a direct cue of the distance along the tram lines and/or relative size of the area between the lines. These distance cues and relative size then give cues to the absolute or relative size of virtual objects near the lines. The tram lines were originally designed [26] to match the linear perspective cue provided by an indoor hallway (Fig. 4.8). The goal in the work was to assist with depth perception, not necessarily not X-ray vision perception. However, as noted above, one belief is that improvements in depth perception of objects—independent of each other—will in turn cause correct perception of relative depth judgments of visible and occluded surfaces, in terms of both ordinal depth and metric depth. Since relative size is considered to be a powerful cue at any distance, this would appear to be a promising method for expressing the distance of virtual objects. However, limited resolution of AR displays may limit the effectiveness of this cue.

4.9 Melting

In the same inspiration of X-ray vision as the image-based techniques, Sandor et al. [40] employed a space-distorting visualization that virtually melts the occluder to show the occluded objects (Fig. 4.9). While Melting provides a clearer view of the

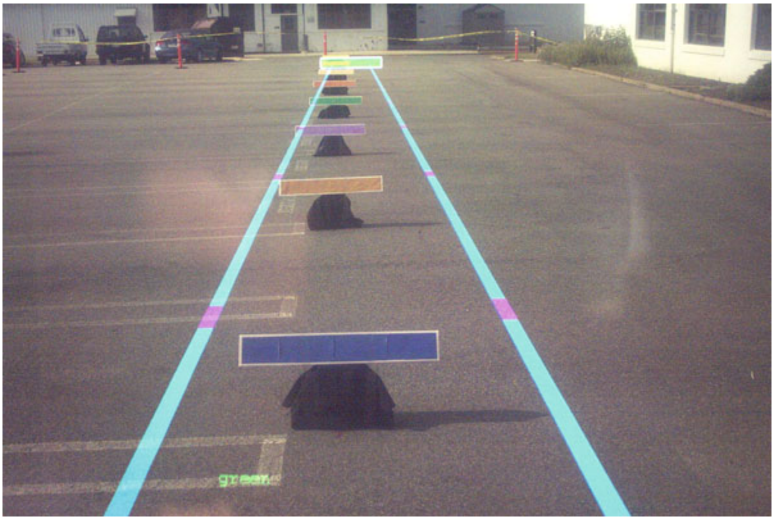


Fig. 4.8 Virtual tram lines aimed to re-create the linear perspective cue from an indoor hallway in the outdoor AR application context

occluded location, it does not preserve enough information of the occluder. Hence, the context of occlusion suffers.

4.10 Virtual Wall

Despite the success of the edge map technique for expressing the relative order of real and virtual surfaces, it can be of limited applicability due to the expense of computing the edge map or the need for special-purpose hardware. Thus a simpler set of "standard" edges could be applied based on the depth of an object [28]. Edges may be added (or subtracted) with each successive layer of depth behind another real surface. This cue does not build directly on any perceptual cues; it emerges from the occlusion cue, attempting to express a physical surface or set of physical surfaces that intervene between the observer and the virtual object of interest. In the initial implementation, additional "edges" in the virtual wall represent another layer of "occlusion" of the virtual object (Fig. 4.10).

5 Evaluations

Because the perception of depth occurs solely within the human visual system, it becomes critical to evaluate the perception of depth experienced by users of a particular application. Several studies have been conducted, using both ordinal and

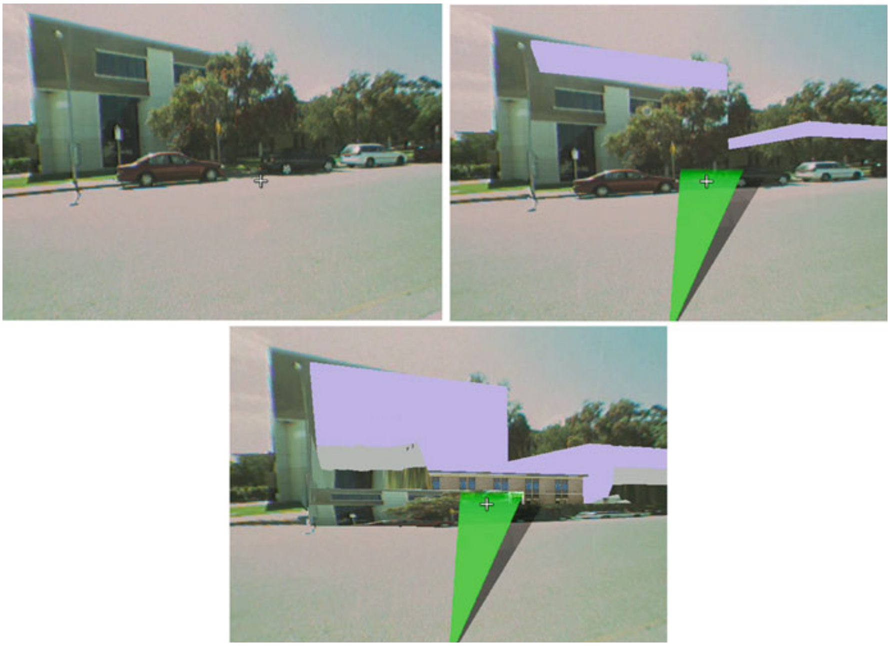


Fig. 4.9 The Melting metaphor completely removed the occluder to show occluded objects

Fig. 4.10 The Virtual wall aimed to provide the cues of the Edge map, but with a less-cluttered and computationally less-demanding representation

Table 4.1 Summary of evaluations (ordered chronologically)

Publication	Display used	Task site	Field of distance	Depth measure	X-Ray metaphors	Estimation protocols
Rolland et al. [36]	OST	In	N	Ord		C
Ellis et al. [10]	OST	In	N	Ord,Met		C M
Ellis and Menges [11]	OST	In	N	Met	…	M
McCandless et al. [33]	OST	In	N	Met		M
Rolland et al. [38]	OST	In	N	Ord,Met		C M
Livingston et al. [29]	OST	Out	F	Ord	1, 3, …	C
Kirkley [22]	OST	In	M	Met	+	V
Livingston et al.[30]	OST	In	M-F	Met		M
Jerome and Witmer [17]	OST	In	M	Met		M V
Swan et al. [48]	OST	In	M-F	Met	1	M
Fidopiastis [13]	HMPD	In	N-M	Met		M
Swan et al. [47]	OST	In	M-F	Met		M V W
Jones et al. [19]	OST	In	M	Met		W
Livingston et al. [26]	OST	In+Out	M-F	Met		M
Dey et al. [8]	HH-Vid	Out	F	Met	7,9	V
Singh et al. [43]	OST	In	N	Met	+	M R
Sandor et al. [39]	HH-Vid	Out	F	□	7	
Livingston et al. [28]	OST	Out	M-F	Ord	1,3,5,6,7,10	C
Jones et al. [18]	OST	In	M	Met		W

Displays used were either optical see-through (OST) of various brands, a head-mounted projective display (HMPD), or hand-held video (HH-Vid) AR displays. The site in which the experimental task implies in what type of environment the target was embedded; we note which experiments were indoor (In), outdoor (Out), or both. The field of distance was within the near-field (N), medium-field (M), or far-field (F), or may have crossed the boundary between two of these fields. Depth measures used could be either metric (Met) or ordinal (Ord), as defined in Sect. 4.2; some papers used experiments of both types, and one (∗) used subjective evaluation. We note whether X-ray vision was a condition in the experiment; we use the subsection numbers from Sect. 4.4 to indicate which X-ray vision metaphors were employed. In addition, a + indicates simple superposition was used, an ellipsis means that some number of depth cues were systematically tested or adapted to the AR system, and a blank indicates that X-ray vision was not a condition. In the case of both X-ray vision and depth cues, "use" may not imply controlled testing; see the text and the original papers for details. Finally, one or more estimation protocols were used to measure subjects' perceptions: forced-choice (C), perceptual matching (M), verbal report (V), direct walking (W), and/or blind reaching (R)

metric depth properties. There are several protocols for measuring depth perception; most of these can be adapted to AR scenarios.

We next review results of various studies. Table 4.1 shows that X-ray vision and metric depth perception have been investigated for a long time; however, most of the experiments were conducted during the last decade using an OST HMD and, until recently, mostly in an indoor location. There are a few key concepts to consider; we classify studies according to these principles in Table 4.1. Some studies have explicitly studied the metric depth perception of graphical objects that are at locations in the real world that are not occluded from view. Other studies have focused on the case of “X-ray vision,” or seeing a graphical object that is represented at a location that is occluded from view. Many of these studies use a baseline case of depth perception in the non-occluded case, which helps establish the relationship between the two cases. Also, the studies have varied in the field of depth they studied; early work focused on the near-field, but work progressed to the medium-field and the far-field in the last decade.

5.1 Work in the Near-Field

Rolland et al. [36] conducted a pilot study at 0.8 and 1.2 m, which asked users to identify whether a 40 mm cube or a 13 mm (diameter) cylinder was closer; the objects were both virtual, both real, or one of each. They found large variability in the perception of depth of virtual objects presented in a stereo optical AR display, attributing this to accommodation and convergence errors (since the display used collimated views for the two eyes) as well as differences in the size of shape of the objects. Subjects overestimated at the tested distances.

Ellis et al. [10] showed that perceived depth of near-field (1.08 m) virtual objects was linked to changes in binocular convergence. Subjects’ perception of the depth of the virtual object was correlated with the change in convergence induced by the presence of a real object either at or closer than the virtual object’s distance. The shortening of the estimated distance was less in the case of the real object being closer than the virtual object. One potential cause for this foreshortening of distances was the mismatch between accommodation and convergence. In this study, X-ray vision was expressed by simple superposition of the graphical objects. Ellis and Menges [11] summarized a series of AR depth judgment experiments, which used a perceptual matching task to examine near-field distances of 0.4–1.0 m and studied an occluder (the X-ray vision condition), convergence, accommodation, observer age, and monocular, bi-ocular, and stereo AR displays. They found that monocular viewing degraded the depth judgment, and that the X-ray vision condition caused a change in convergence angle which resulted in depth judgments being biased towards the observer. They also found that cutting a virtual hole in the occluder, which made the depth of the virtual object physically plausible, reduced the depth judgment bias compared to superposition. McCandless et al. [33] used the same experimental setup and task to additionally study motion parallax and AR system latency in monocular viewing conditions; they found that depth judgment errors increased systematically with increasing distance and latency. They constructed a model of the error showing that lateral projected position of the virtual object and depth judgment error were each linearly related to the latency.

Rolland et al. [38] compared forced-choice and perceptual matching tasks with a prototype stereo OST display; the display alleviated some conflict between accommodation and convergence noticed in earlier experiments. Four shapes (cube, cylinder, faceted cylinder, and octahedron) and three sizes of objects were used as stimuli; all were at a distance of 0.8 m. They found a discrimination threshold of 7 mm using constant stimuli and precision of rendered depth of 8–12 mm using adjustments, depth accuracy, and no consistent depth judgment bias.

Fidopiastis [13] developed protocols to test perceptual effects of head-mounted projective displays (HMPDs), including depth perception. Pilot testing at near-field distances (and edging into the medium-field) were conducted with two prototype displays. With a 52° diagonal FOV and 640 ×480 graphical resolution, five subjects were found to have mean signed errors between −7.7 and 9.9 mm at a distance of 800 mm, −5.5 to 12.5 mm at 1,500 mm, and −11.8 to 19.8 mm at 3,000 mm. With a 42° diagonal FOV and 800 ×600 graphical resolution, five subjects had mean signed errors between −3.0 and 7.7 mm at a distance of 800 mm, between −8.9 and 15.6 mm at 1,500 mm, and −34.7 mm and 32.2 mm at 3,000 mm. The variances in these errors (within subject) grew from a few millimeters at 800 mm to 10–30 mm at 1,500 mm to 20–70 mm at 3,000 mm. The research indicated that individual differences in the distance at which subjects would focus in dark environments may help predict some differences in depth perception accuracy.

Singh et al. [43] evaluated the effect of closed-loop (perceptual matching) and open-loop (blind reaching) depth perception tasks for near-field AR (34–50 cm) in the presence and absence of a salient occluder. They reported that a perceptual matching protocol (closed-loop) was significantly better in depth judgment than the blind reaching (open-loop) protocol. However, using both of the protocols, depth was mostly underestimated. The effect of the occluder was complex and interacted with the protocol and the distance. Without an occluder present, error tended to increase with distance. With the occluder present, an interaction with the convergence cue could produce or not produce errors depending on the distance.

5.2 Extension to Medium-Field and Far-Field

Livingston et al. [29] studied varied representations of occluded buildings using drawing styles of wireframe, filled, and filled-with-wireframe outlines, varied opacity (constant or based on distance), and varied intensity (constant or based on distance). As discussed in Sect. 4.4, the last two approximate the aerial perspective cue. Other variables included the use of a consistent ground plane (on and off) and presenting images with or without binocular disparity (in software, with fixed hardware IPD). Subjects were more accurate in identifying ordinal depth among buildings at 60–500 m with the ground plane consistent, but were found to have no significant difference in performance under conditions of filled-with-wireframe-outline drawing, opacity decreasing with object distance, and intensity decreasing with object distance. Subjects were found to get faster with practice, but not significantly more accurate. This could indicate that the cues were intuitive for users. A follow-up study [30] used a perceptual matching technique and found that similar errors were made when matching the depth of real objects and unoccluded virtual objects against real reference objects.

Kirkley [22] studied occluders (via superposition), the ground plane, and object type (real, realistic virtual, and abstract virtual) in both monocular (Microvision Nomad) and bi-ocular (Sony Glasstron) optical AR viewing at medium-field distances (3.0–33.5 m). He found that occluders increased error, placing objects on the ground plane decreased error, and judging the depth of real objects was most accurate. In most cases, users underestimated the distance to virtual objects seen in the head-worn AR displays. Jerome and Witmer [17] noted issues of registration, monocular viewing, and wide separation of repeated distances (i.e. becoming a memory task rather than a perceptual task) with Kirkley's work. They used an OST display, object distances of 1.5–25 m with eight objects (four small and four large, four abstract shapes and four familiar objects), and two protocols (perceptual matching and verbal report). They found that female subjects had significantly more error with virtual objects than with real objects, while male subjects showed no significant difference. They also found that error in judging distance to virtual objects was significantly reduced when preceded by a distance judgment of a real object. Finally, their subjects were more accurate with perceptual matching (of a physical robot) to a virtual distance than with verbal reporting of virtual distance. Both of these tests were done with hardware copies of the Battlefield Augmented Reality System [27], used in the two studies mentioned in the preceding paragraph and in the first study discussed in the next paragraph.

Swan et al. [48] explored medium- and far-field distances (5–45 m) in optical AR with variables such as position in the visual field, occlusion of the site of the virtual object, and practice on the task using a perceptual matching technique. Subjects underestimated inside of 23 m but overestimated beyond that distance. However, the subjects may have been using a combination of cues to arrive at estimates of depth of the virtual object with respect to the appropriate real object. These included relative size, disparity, brightness, aerial perspective approximations, or the perhaps-questionable convergence presented in the display (which lacked an IPD adjustment). Thus this apparent "cross-over" point from underestimation to overestimation may have derived from the environment or the equipment used. This cautionary argument is one motivating factor to compare the work against similar experiments. Data in [48] supported an estimate of an 8 % increase in the rate at which error increased with distance from the unoccluded to the occluded condition. While this pattern is not surprising, the precise amount again requires verification and qualification as to the conditions under which it occurs in general.

Swan et al. [47] examined the effect of the experimental protocol and object type (real objects with natural vision, real objects seen through optical AR, virtual objects, and combined real and virtual objects) at medium-field distances (3–7 m) with optical AR displays. The results indicated underestimation and implicated the restricted FOV and the inability for observers to scan the ground plane as explanations for the bias. Jones et al. [19] compared optical AR depth perception against VE depth perception and suggested that the virtual background contributes to the underestimation of depth in immersive VE. They found less underestimation in AR than in an analogous VE and no effect of using motion parallax in AR (versus standing still). Another study of metric depth with indoor AR using an OST display was reported by Jones et al. [18]. A blind walking protocol was used to measure depth perception in medium-field distances (3–7 m) and, similar to previous studies, found a consistent underestimation of distance.

5.3 Experiments Focused on Outdoor and Mobile AR

Livingston et al. [26] studied the tram lines they introduced and compared performance indoors (with strong linear perspective cues) and outdoors (without strong cues) using optical AR and perceptual matching at medium-field distances (4.8–38.6 m). They found a consistent underestimation of depth indoors (in contrast to earlier experiments in the same environment [48]) but overestimation outdoors. The presence of the tram lines decreased the estimated distance both indoors and outdoors, reducing error in the latter but increasing error in the former.

Melting [40] was compared with an edge map in an egocentric depth perception study of X-ray vision by Dey et al. [8]. The authors employed a verbal reporting protocol in their outdoor study, where participants had to guess the distance of an occluded object placed at far-field distances (69.7–117.0 m). Contradicting previous findings by Livingston et al. [26], they found that, like indoor AR environments, depth is underestimated in outdoor AR environments as well. Authors reported that the Melt visualization performed more accurately and faster than X-ray vision. However, Melt vision removes the occluder completely and eventually loses the important features of the occluder. Unlike other experiments of X-ray vision, this experiment was performed using a hand-held device.

In another evaluation using a hand-held display at an outdoor location, Sandor et al. [39] evaluated their *saliency-based* X-ray vision with the previously-presented edge-map metaphor. They found similar results while selecting a target object in the occluded location, though saliency-based X-ray preserved more information of the occluder. Another on-line survey was conducted, where both of these X-ray vision metaphors were applied to three different levels of brightness and edges of the occluder. They found that with higher brightness, edge-overlaid X-ray vision provided more information about the occluded region, whereas under mid-range or lower brightness, saliency-based X-ray vision provided more information about the occluded region. Dey et al. [9] found that subjects, while navigating in an urban environment, spent more time looking at their hand-held AR display with X-ray vision capabilities than subjects spent looking at traditional maps viewed on their hand-held display (whether oriented with North in the vertical direction or the user's current view direction in the vertical direction on the screen). The AR with X-ray vision condition also had the fewest context switches between the hand-held display and the environment.

Broad-based comparison of X-ray vision metaphors is rare in the literature. Livingston et al. [28] compared opacity, stipple, ground grid, edge map, virtual wall, and (a variant of) the virtual tunnel techniques for ordinal depth judgments of

Fig. 4.11 This customized and simplified version of the virtual tunnel concentrated on the number of planes through which the tunnel was being extended to get to the target object. It was the best method in a broad-based comparison of X-ray vision metaphors [28]

virtual icons amongst a set of real buildings given on a map (of which the nearest wall was visible). A customized virtual tunnel (Fig. 4.11) was found to yield the lowest error, followed by the virtual wall and the ground grid. Additionally, the virtual tunnel, virtual wall, and edge map techniques were found to bias subjects to underestimation, while the opacity technique led to overestimation. Users were fastest with the baseline case of no X-ray vision metaphor and the virtual tunnel; they were slowest with the edge map. It was noted that the edge map technique had the potential to be very slow or require dedicated hardware to maintain in a dynamic environment; this study was conducted in a static environment.

6 Discussion

We conclude our review with discussion of the trends evidenced by the visualization metaphors and studies of them. We present four graphs that summarize some of the experiments described above in the medium-field and far-field. While the medical applications demonstrate the need for X-ray vision to work for applications that are contained within the near field, the trend towards mobile applications of AR as the potential consumer application of the technology pushes the more distant fields to great importance.

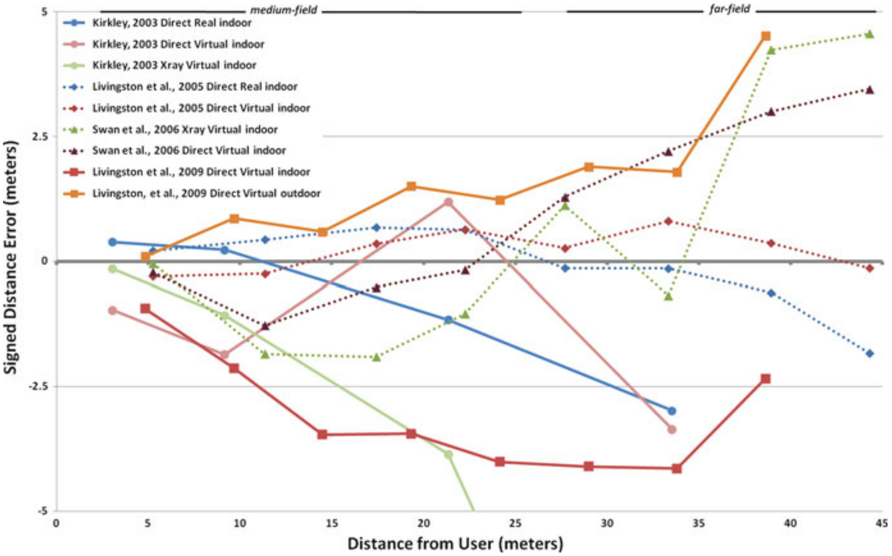


Fig. 4.12 A graph of data from experiments [22, 26, 30, 48] that were conducted in the medium-field and the neighboring part of the far-field. Each experiment compared varying conditions, but only two directly studied the X-ray condition. The lines on this graph are coded by symbols to differentiate the experiment and by color to identify equivalent conditions. Thus the two *green lines* for the Xray conditions should be similar; instead, they diverge, with the fourth point of [Kirkley, 2003 Xray Virtual indoor] at $(33.5, -13.7)$, well off the bottom of this graph. The *red lines* are all measures of depth estimation with virtual objects in spaces that are directly viewed by the subject; these lines also show quite a bit of divergence from each other. The two *blue lines* from estimation of distance to real objects are somewhat similar, but still do not quite agree on the nature of the errors. In this and the next four graphs, the *thick gray line* at zero indicates veridical perception

6.1 Metric Depth in Medium-Field and Far-Field

The first graph (Fig. 4.12) shows data from four experiments that were conducted in the medium-field and the adjacent portion of the far-field using head-worn OST displays. Each experiment compared different conditions; the point marker indicates which experiment is graphed with each line. Most of these conditions were visualizations of virtual objects in locations that were directly visible to the user; an X-ray vision metaphor was applied in only two conditions of two (separate) experiments. The data from Kirkley [22] is notable in that it diverges so far from the remainder of the data; the final data point $(33.5, -13.9)$ is cut off from the graph (which was zoomed to the majority of the data to enable legibility of the bulk of the data). The X-ray condition in Swan et al. [48] agrees with Kirkley's data for some medium-field distances. These tests were designed to understand the metric perception of distance to virtual objects. Of the four data sets for direct virtual object viewing indoors (red hues), one stands out as an outlier: the indoor data from Livingston et al. [26] near the bottom of the visible portion of the graph (although the final point from Kirkley’s equivalent condition appears to show the same behavior). That the same physical environment could give rise to such disparate data as the indoor data sets from Livingston et al. [30] and Swan et al. [48] (which were acquired in the same hallway) is curious or even suspicious. It underscores the potential difficulty in acquiring good data in these experiments, including the need to control potential confounding factors and to be concerned with the general ability of subjects with depth perception. One potential confound was the larger FOV of the OST display used in these experiments (a Sony Glasstron, Microvision Nomad 1000, and nVis nVisorST were used in various experiments).

The outdoor data from Livingston et al. [26] is also somewhat different, but the environment offers an obvious hypothesis as to why this may have occurred. The difference in the data may be an effect produced by the loss of the powerful perspective cue from the indoor to the outdoor environment, or it could (again) be an artifact of the different head-worn display. Due to the disruptive nature of collecting indoor data in an office environment, the environment variable was not counterbalanced, but merely separated by 7–14 days. With a perceptual task, one would assume that such a design would not produce order effects, but this graph does cause some concern.

Two data sets in this graph show a version of the task with a real target [22, 30]. The data with the real target appears to be quite reasonable; this would appear to validate that the users were in fact making use of the depth cues in the merged environment when the task was moved into AR. One could argue that the perceptual matching was not a depth-based task, but merely a size-matching task, however. Again, this argument underscores the importance of validating the task and experimental design. Taken together, the data in this graph tend to drift from below the veridical line (signed error equal to zero) in the medium-field to above the veridical line in the far-field. This actually contradicts the long-standing observation in virtual environments that users tend to underestimate depth. The majority of the data in this graph in the medium-field indicates underestimation, but it would appear that there is a slight tendency to overestimate depth to virtual objects as they move into the far-field. However, the data are far from in agreement even on this fundamental judgment. This is clearly an area that deserves more detailed exploration.

6.2 *Metric Depth Well into the Far-Field*

The second graph (Fig. 4.13) shows a metric depth study conducted well into the far-field [8]; this experiment used a hand-held display. As noted above, this graph shows a general trend of underestimation under these experimental conditions. The reasons for the difference between this graph and the previous graph are unexplored, but the variety of conditions could give rise to several hypotheses that could be tested. Clearly, the virtual cue akin to the grid lines and the ground grid was extremely helpful in helping users improve their accuracy with metric depth

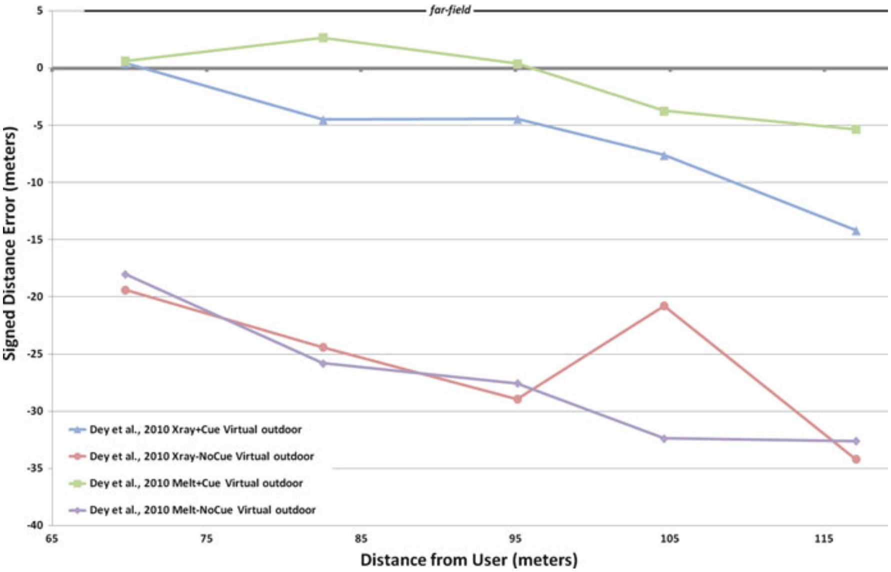


Fig. 4.13 A graph of data from an experiment [8] that was conducted in the far-field compared two visual metaphors for X-ray vision and a cue akin to a tram line. This cue was extremely helpful. All of these conditions were conducted with hand-held displays

estimation. Given the difficulty of this task in general and the accuracy achieved, it is likely that users simply relied on the cue and estimated distance by counting the units (tens of meters) from their location to the virtual object. If this is an acceptable mode of operation in the target application, then it becomes perhaps the most obviously beneficial technique of all those discussed here.

6.3 Ordinal Depth in the Medium-Field and Far-Field

The third graph (Fig. 4.14) returns to the boundary region between the medium-field and far-field; this data comes from a single experiment conducted on ordinal depth perception using a head-worn display [28]. In contrast to the original paper, this graph separates by X-ray vision metaphor and by the "zone" in which an object was found. A "zone" indicated the number of intervening real surfaces in the known environment that were between the user and the location of the virtual object; zone 1 had no intervening surfaces. The analysis of this graph should begin with the line representing no X-ray vision metaphor being applied, which exhibited the most error in the first three zones, found to be overestimation of the ordinal distance, then crossed over to a balance between overestimation and underestimation, and finally settled into the second-greatest underestimation in the farthest zone. Clearly, we

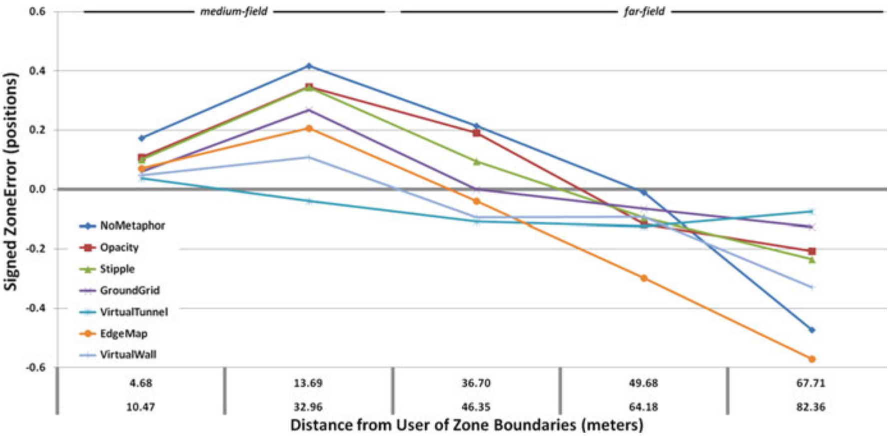


Fig. 4.14 A graph of data from an experiment [28] that was conducted in the medium-field and far-field compared several metaphors for X-ray vision. This experiment used a head-worn OST display. The occlusion “zones” were the distances (shown on the horizontal axis) for which the number of real surfaces intervening between the user and the virtual object location was constant. Since the users knew the environment, it is safe to assume that they were aware of these surfaces

would expect any X-ray vision metaphor to improve upon this performance. One should be careful regarding the first and last zones, since only one direction of error is possible in each, but relative performance differences are meaningful. The Virtual Tunnel stays the closest to correct ordering, with the Virtual Wall coming close to or exceeding its performance until the most distant zone. The Edge Map, by contrast, appeared to suffer from visual clutter. It is also interesting to note the similarity of the graphs for the Opacity and Stipple metaphors. The former simulates the natural depth cue of aerial perspective quite well, whereas the latter is a completely synthetic cue commonly used in technical illustration. This demonstrates that users can condition themselves to think in terms of metaphors quite effectively, and perhaps as well as we intuit natural cues for the human visual system.

6.4 Ordinal Depth Well into the Far-Field

The fourth graph (Fig. 4.15) shows the data from an early comparison of various X-ray vision metaphors [29] using a head-worn OST display with far-field distances. The two metaphors that emulated the aerial perspective cue (Opacity and Intensity fading with distance) both improved user performance, and their combination improved it yet further. Users were also aided by the use of a filled shape with a wireframe outline, rather than either representation alone. It is interesting to compare how the users fared with the cue of height in the visual field (represented by the

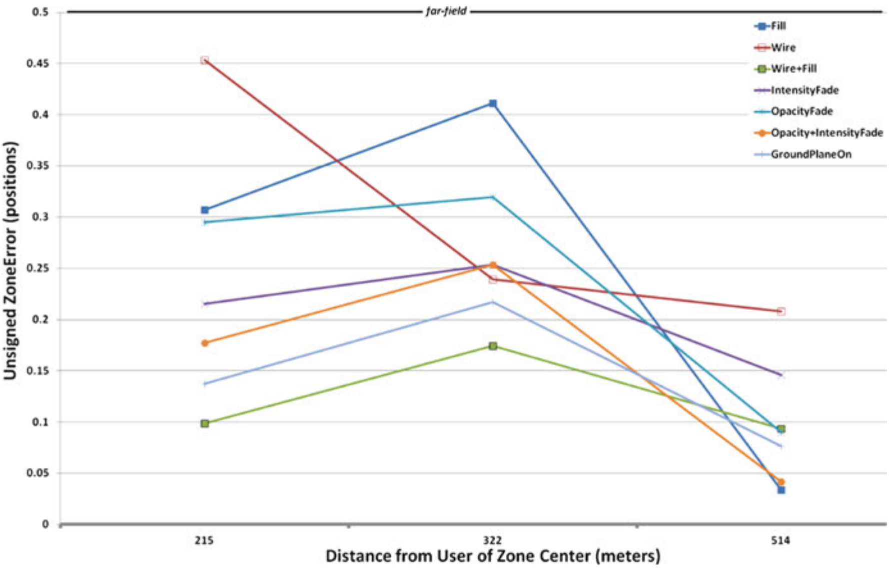


Fig. 4.15 A graph of data from an experiment [29] that was conducted in the far-field compared three metaphors for X-ray vision. This experiment also used a head-worn OST display. The distances (shown on the horizontal axis) were to the centroid of a building that was the target at the given distance. The targets and the distracting objects were mutually disjoint

use of a consistent ground plane, denoted “GroundPlaneOn” in the graph legend) in comparison to the X-ray vision metaphors. This early work showed the potential of AR to provide usable X-ray vision and helped to spark renewed interest in this area, including several of the works discussed in this chapter. This data is in unsigned error, so the graph is not as informative as the others.

6.5 *Near-Field Depth*

The final graph summarizes experiments conducted in the near-field (Fig. 4.16). While the experimental conditions differ so greatly that any general conclusions are difficult to draw, we can see that much of the data indicates underestimation of distance. The graph depicts data points in X-ray vision conditions with circle glyphs (and connecting lines for the Singh et al. [43] data). All of these points indicate underestimation, except for the Ellis et al. data point for monocular viewing by subjects with advanced age (greatest overestimation of the Ellis et al. data) [11]. This and the other monocular point (most underestimated of the Ellis et al. data) presumably suffered from the poor disparity and convergence cues in the early optical AR displays as well as the decreasing ability of humans to use accommodative cues with increasing age. All three of these cues are quite sensitive in the near-field.

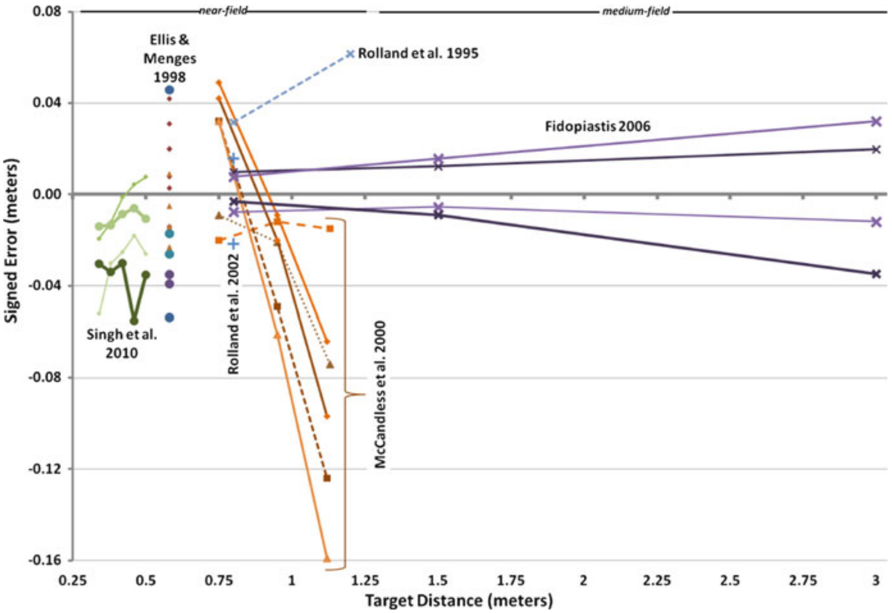


Fig. 4.16 A graph of data from six papers studying depth perception and/or X-ray vision in the near-field. Labels within the graph indicate the publications. The X-ray vision conditions (two line graphs in Singh et al. and six points in Ellis et al.) are indicated by *large circles* at the data points. This data largely indicates underestimation, a point discussed in the text

This may be similar to the explanation of the overestimation that Rolland et al. [36] found. McCandless et al. [33] also noted that the accommodative cue was slightly too close for the virtual objects even at the near distance and cited this as a potential source of bias in depth judgments. They also posited that the close proximity of the wall behind the experimental scene may have biased the depth judgments. These two factors may have caused the overestimation of distance for the nearest targets and perhaps some of the underestimation at the farther distances. This gives another specific example of how uncontrolled factors may affect experimental results.

6.6 *Summary Observations*

The first observation beyond the graphs is how little overlap there is between them. It would have been nice to show all this work on a single graph, but the distances would have been starkly disjoint sets, not to mention the near impossibility of equating the metric and ordinal measurements. A different graph would be needed to focus in on that data rather than lose it in the scale of even (Fig. 4.12). A few issues deserve special attention, given the summary of experiments above.

6.6.1 Underestimation vs. Overestimation

As noted above, there is a well-observed (though not yet well-understood) trend towards underestimation of distances in VEs. One can easily put forth theories that AR need not behave the same, primarily on the basis of the grounding in the real environment and natural depth cues that users receive from it. Still, we have noted that some experiments with (metric or ordinal) depth perception found underestimation in one or more conditions, a few found overestimation in one or more conditions, and many found a mixture of underestimation and overestimation. Clearly, this is an area that is ripe for further study to understand the phenomenon in AR.

6.6.2 Optical vs. Video

We have attempted to note precisely what type of AR display was used in the various experiments above, and the interaction of the two fundamental technologies of OST and video overlay with the natural depth cues used by the human visual system. Despite the relative ease with which we can reason about the expected effects, there is relatively little direct comparison of the benefits of the two types of displays for the perception of depth. (Some comparison for applications has been done [37].) This is another area that would benefit the field to have studied further.

6.6.3 Head-Worn vs. Hand-Held

An issue that may interact with the above question is the type of display used. Head-worn displays were classically the presumed form for AR, but mobile phones appear to have displaced them, as an example of hand-held displays. A priori, there is no reason to assume that the perceptual qualities of the two form factors are identical. This may account, for example, for the difference between Fig. 4.13 and the apparent differences between certain portions of data in Figs. 4.12 and 4.14. (We make this statement despite our acknowledgment of the difficulty of comparing the graphs.) Since each type of display has a large potential user base, it would again seem to be a rich area for research. One twist, however, is that when studying the effects of these types of displays, it may be more important than for the other empirical questions to embed the task within an expected or prototypical application context. For example, if mobile phones are to be used as AR navigation aids, then a consumer-level registration accuracy and task should be the basis for an evaluation. If a military application plans to integrate a head-worn AR capability into (for example) a pilot's helmet, then the depth perception or X-ray vision task should come in the context of his role in the battlespace.

6.6.4 Binocular Vision

One underexplored area in recent research is near-field localization; while this was heavily featured in early work, advances in displays—notably adjustments in binocular disparity and focal distance, have the potential to significantly improve the depth cues that are critical in the near field. This distance is important for a number of proposed AR applications, including medicine and manufacturing. Studies that determine the utility of adjustments in convergence, binocular disparity, and accommodation in display devices might provide some guidance for hardware designers.

6.6.5 Outdoor Environments

While a few recent experiments [8, 26, 28] studied outdoor X-ray vision, this is still an under-explored direction. The use of X-ray vision is not restricted to indoor environments. Recent advances in mobile phones and portable devices have led to their rise as AR platforms and opened new possibilities for outdoor mobile AR applications, such as navigation tools. It is required to conduct experiments on AR X-ray vision in outdoor locations. Unlike indoor environments, outdoor environments are practically impossible to fully control. The random variation of the environment is expected to have different effects on human perception, which are necessary to understand for the extension and improvement of X-ray vision. The contradictory results need to be clarified and quantified.

6.6.6 Comparison of Visual Metaphors

We note that a few tests directly compared two or more metaphors; however, the set of all possible combinations of comparison is far from complete. In part, this is due to the recent introduction of methods, but it also stems from the volume of data needed to compare more than a small number of X-ray vision metaphors. An ambitious study could fill this gap in the knowledge of the field, although an experimenter should clearly be prepared to exhibit patience in such an endeavor. Also, there are parameters associated with various techniques; ideally, each technique to be compared against another technique should be optimized for the task in order to achieve a fair comparison.

6.6.7 Intuitiveness and Interaction

Sutherland's original vision for the virtual world—which was in fact an augmented reality, not the immersive virtual worlds we conceive today with the term—was that the virtual objects would be indistinguishable from reality. We argue that to fulfill this vision, we must achieve visualization metaphors in AR that are as intuitive for the human visual system as natural depth cues. This may seem impossible, but we noted above that stipple effects appear to have achieved this status. How to test the intuition of the responses is not an easy question, but one obvious first choice is the reaction time with which a user can (accurately) determine the (metric or ordinal) depth. Many of the experiments described herein measured reaction time, but rarely with anything close to the speed of human vision in assessing depth. Many experiments used only a few repetitions of the task, and even fewer reported results with respect to repetition. This could give some insight into how easily and how well users can learn to accept the X-ray visualization metaphors. Similarly, only one tool [3] used interaction as a means to understand depth; while perhaps not appropriate for all applications, this is another area that is under-explored in the field.

7 Conclusion

X-ray vision is among the most flashy of the capabilities offered by AR technologies. As mobile applications begin to enter the consumer market and industrial and military applications continue to mature and find new uses, making AR systems work at the perceptual level embodied in the X-ray vision perception is critical. Even for (conceptually) simple applications such as navigation, which can be implemented without X-ray vision, accurate distance perception of graphical cues can be an obvious benefit to the user. It should be clear from the significant research efforts described here that the best methods are far from clear for even narrow applications, let alone across the field of AR. The analysis across experiments identified several directions for future research that are likely to bring great benefit to the field and scientific grounding to the techniques. Furthering the latter, we connected the techniques through the metaphors to the perceptual cues applied by the human visual system to the understanding of distance; it is important to understand the fundamental relationship between the AR cues and the visual system in order to determine the reasons for the success or failure of a technique and to suggest improvements in the case of the latter. It will also be increasingly important to push the tests of distance perception and X-ray vision closer to the applications envisioned, so that the metaphor of X-ray vision can truly be simulated in augmented reality.

Acknowledgements The authors would like to thank Steve Ellis, Cali Fidopiastis, Henry Fuchs, LT Gregory O. Gibson, Tobias Höllerer, Christian Jerome, Sonny Kirkley, Jannick Rolland, Andrei State, and Ed Swan for their assistance in the preparation of this chapter.

References

1. Avery, B., Sandor, C., Thomas, B.H.: Improving spatial perception for augmented reality X-ray vision. In: IEEE Virtual Reality, pp. 79–82 (2009)
2. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. Computer Graphics (Proceedings of SIGGRAPH'92) **26**(2), 203–210 (1992)

3. Bane, R., Höllerer, T.: Interactive tools for virtual X-ray vision in mobile augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 231–239 (2004)
4. Biocca, F.A., Rolland, J.P.: Virtual eyes can rearrange your body: Adaptation to visual displacement in see-through, head-mounted displays. Presence: Teleoperators and Virtual Environments **7**(3), 262–277 (1998)
5. Bruce, V., Green, P.R., Georgeson, M.A.: Visual Perception: Physiology, Psychology, and Ecology, 3rd edn. Psychology Press (1996)
6. Cutting, J.E.: Reconceiving perceptual space. In: H. Hecht, R. Schwartz, M. Atherton (eds.) Looking into Pictures: An Interdisciplinary Approach to Pictorial Space, chap. 11, pp. 215–238. MIT Press (2003)
7. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: Handbook of perception and cognition, Vol. 5: Perception of space of motion, pp. 69–117. Academic Press (1995)
8. Dey, A., Cunningham, A., Sandor, C.: Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In: 17th ACM Symposium on Virtual Reality Software and Technology, pp. 211–218 (2010)
9. Dey, A., Jarvis, G., Sandor, C., Wibowo, A.K., Mattila, V.V.: An evaluation of augmented reality X-ray vision for outdoor navigation. In: The 21st International Conference on Artificial Reality and Telexistence, pp. 28–32 (2011)
10. Ellis, S.R., Bucher, U.J., Menges, B.M.: The relationship of binocular convergence and errors in judged distance to virtual objects. In: Proceedings of the International Federaion of Automatic Control, pp. 297–301 (1995)
11. Ellis, S.R., Menges, B.M.: Localization of virtual objects in the near visual field. Human Factors **40**(3), 415–431 (1998)
12. Feiner, S.K., Seligmann, D.D.: Cutaways and ghosting: Satisfying visibility constraints in dynamic 3D illustrations. The Visual Computer **8**(5–6), 292–302 (1993)
13. Fidopiastis, C.M.: User-centered virtual environment assessment and design for cognitive rehabilitation applications. Ph.D. thesis, Department of Modeling and Simulation, University of Central Florida (2006)
14. Fuchs, H., Livingston, M.A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J.R., Rademacher, P., Drake, S.H., Meyer, A.A.: Augmented reality visualization for laparoscopic surgery. In: Proceedings of First International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 934–943 (1998)
15. Furness, L.T.A.: The application of head-mounted displays to airborne reconnaissance and weapon delivery. Tech. Rep. TR-69-241, U.S. Air Force Avionics Laboratory, Wright-Patterson AFB (1969)
16. Furness, T.A.: The super cockpit and its human factors challenges. In: Proceedings of the Human Factors Society 30th Annual Meeting, pp. 48–52. Human Factors and Ergonomics Society (1986)
17. Jerome, C.J., Witmer, B.G.: The perception and estimation of egocentric distance in real and augmented reality environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 49, pp. 2249–2252 (2005)
18. Jones, J.A., Swan II, J.E., Singh, G., Ellis, S.R.: Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 29–35 (2011)
19. Jones, J.A., Swan II, J.E., Singh, G., Kolstad, E., Ellis, S.R.: The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 9–14 (2008)
20. Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Hollerer, T., Sestito, S.: Information filtering for mobile augmented reality. In: Proceedings of IEEE and ACM International Symposium on Augmented Reality, pp. 3–11 (2000)
21. Kalkofen, D., Mendez, E., Schmalstieg, D.: Interactive focus and context visualization for augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 191–201 (2007)

22. Kirkley, Jr., S.E.H.: Augmented reality performance assessment battery (ARPAB: Object recognition, distance estimation, and size estimation using optical see-through head-worn displays. Ph.D. thesis, Instructional Systems Technology, Indiana University (2003)
23. Kiyokawa, K., Kurata, Y., Ohno, H.: An optical see-through display for mutual occlusion of real and virtual environments. Computers and Graphics **25**(5), 765–779 (2001)
24. Klein, G., Murray, D.: Simulating low-cost cameras for augmented reality compositing. IEEE Transactions on Visualization and Computer Graphics **16**(3), 369–380 (2010)
25. Liu, S., Hua, H., Cheng, D.: A novel prototype for an optical see-through head-mounted display with addressable focus cues. IEEE Transactions on Visualization and Computer Graphics **16**(3), 381–393 (2010)
26. Livingston, M.A., Ai, Z., Swan II, J.E., Smallman, H.S.: Indoor vs. outdoor depth perception for mobile augmented reality. In: IEEE Virtual Reality, pp. 55–62 (2009)
27. Livingston, M.A., Brown, D., Julier, S.J., Schmidt, G.S.: Military applications of augmented reality. In: NATO Human Factors and Medicine Panel Workshop on Virtual Media for Military Applications (2006)
28. Livingston, M.A., Karsch, K., Ai, Z., Gibson, LT G.O.: User interface design for military AR applications. Virtual Reality **15**(2–3), 175–184 (2011)
29. Livingston, M.A., Swan II, J.E., Gabbard, J.L., Höllerer, T.H., Hix, D., Julier, S.J., Baillot, Y., Brown, D.: Resolving multiple occluded layers in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 56–65 (2003)
30. Livingston, M.A., Zanbaka, C., Swan II, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: IEEE Virtual Reality, pp. 287–288 (2005)
31. Loomis, J.M., Knapp, J.M.: Visual perception of egocentric distance in real and virtual environments. In: L.J. Hettinger, M.W. Haas (eds.) Virtual and Adaptive Environments, pp. 21–46. Lawrence Erlbaum Associates (2003)
32. McCandless, J.W., Ellis, S.R.: The effect of eye position on the projected stimulus distance in a binocular head-mounted display. In: Stereoscopic Displays and Virtual Reality Systems VII, Proceedings of SPIE Vol. 3957, pp. 41–48 (2000)
33. McCandless, J.W., Ellis, S.R., Adelstein, B.D.: Localization of a time-delayed, monocular virtual object superimposed on a real environment. Presence: Teleoperators and Virtual Environments **9**(1), 15–24 (2000)
34. Nagata, S.: How to reinforce perception of depth in single two-dimensional pictures. In: Pictorial Communication in Virtual and Real Environments, chap. 35, pp. 527–545. Taylor & Francis, Inc. (1991)
35. Plumert, J.M., Kearney, J.K., Cremer, J.F., Recker, K.: Distance perception in real and virtual environments. ACM Transactions on Applied Perception **2**(3), 216–233 (2005)
36. Rolland, J.P., Ariely, D., Gibson, W.: Towards quantifying depth and size perception in virtual environments. Presence: Teleoperators and Virtual Environments **4**(1), 24–49 (1995)
37. Rolland, J.P., Fuchs, H.: Optical versus video see-through head-mounted displays. Presence: Teleoperators and Virtual Environments **9**(3), 287–309 (2000)
38. Rolland, J.P., Meyer, C., Arthur, K., Rinalducci, E.J.: Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence: Teleoperators and Virtual Environments **11**(6), 610–625 (2002)
39. Sandor, C., Cunningham, A., Dey, A., Mattila, V.V.: An augmented reality X-ray system based on visual saliency. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 27–36 (2010)
40. Sandor, C., Cunningham, A., Eck, U., Urquhart, D., Jarvis, G., Dey, A., Barbier, S., Marner, M., Rhee, S.: Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In: IEEE Virtual Reality (2010)
41. Schall, G., Mendez, E., Kruijff, E., Veas, E., Junghanns, S., Reitinger, B., Schmalstieg, D.: Handheld augmented reality for underground infrastructure visualization. Personal and Ubiquitous Computing **13**(4), 281–291 (2009)
42. Schiffman, H.R.: Sensation and Perception, 5th edn. John Wiley & Sons (2001)

43. Singh, G., Swan II, J.E., Jones, J.A., Ellis, S.R.: Depth judgment measures and occluding surfaces in near-field augmented reality. In: ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, pp. 149–156 (2010)
44. State, A., Hirota, G., Garrett, W.F., Chen, D.T., Livingston, M.A.: Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In: Computer Graphics Proceedings, Annual Conference Series (Proceedings of SIGGRAPH’96), pp. 429–438 (1996)
45. Stix, G.: See-through view: Virtual reality may guide physician’s hands. Scientific American **267**(3), 166 (1992)
46. Sugano, N., Kato, H., Tachibana, K.: The effects of shadow representation of virtual objects in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 76–83 (2003)
47. Swan II, J.E., Jones, A., Kolstad, E., Livingston, M.A., Smallman, H.S.: Egocentric depth judgments in optical, see-through augmented reality. IEEE Transactions on Visualization and Computer Graphics **13**(3), 429–442 (2007)
48. Swan II, J.E., Livingston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., Hix, D.: A perceptual matching technique for depth judgments in optical, see-through augmented reality. In: IEEE Virtual Reality, pp. 19–26 (2006)
49. Tsuda, T., Yamamoto, H., Kameda, Y., Ohta, Y.: Visualization methods for outdoor see-through vision. IEICE Transactions on Information and Systems **E89–D**(6), 1781–1789 (2006)
50. Witt, J.K., Proffitt, D.R., Epstein, W.: Perceiving distance: A role of effort and intent. Perception **33**(5), 577–590 (2004)
51. Zollmann, S., Kalkofen, D., Mendez, E., Reitmayr, G.: Image-based ghostings for single layer occlusions in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 19–26 (2010)

Chapter 5
Cognitive Issues in Mobile Augmented Reality: An Embodied Perspective

Nai Li and Henry Been-Lirn Duh

1 Introduction

Augmented reality (AR), which is characterized by overlaying virtual imagery onto a physical world scene, enables users to interact with virtual information situated in real space and real time. Driven by the advances of hardware and software, mobile AR emerges as a promising interface for supporting AR interaction. Wearable devices such as head-mounted displays (HMDs) and wrist-worn displays are important channels for people to get immediate access to AR contents while roaming the surrounding environment. Recent developments in mobile computing technologies also enable more handheld devices like mobile phones, PDAs and tablet PCs to be the platforms to implement AR. The convergence of AR and mobile devices delivers an innovative experience for users to explore the physical world.

Mobile AR reveals potential in a diversity of domains, including manufacturing, tourism, education, entertainment, and urban modeling [79, 114, 115]. Despite the discussion on technical issues of AR, the social influence of AR technology has received considerable attention in recent years [43, 63]. In order to develop AR applications with high effectiveness, the user-centred perspective should be incorporated, and more understanding on the impact of characteristics of AR systems on human activities is needed [126]. Human factors play a vital role in designing effective computing systems. To date, some researchers have explored human factors in

---

N. Li (✉)
Department of Communications and New Media, National University of Singapore,
Blk AS6, #03-41, 11 Computing Drive, Singapore
e-mail: g0900788@nus.edu.sg

H.B.-L. Duh
Department of Electrical and Computer Engineering, National University of Singapore,
21 Heng Mui Keng Terrace, #02-02-09, Singapore
e-mail: eledbl@nus.edu.sg

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_5,

AR from different perspectives with attempts to propose design guidelines [24, 31, 67]. Cognitive issues, which relate to users' cognitive process for understanding an AR environment when interacting with the system, are identified as an important category of human factors in AR [24]. Duh, Ma, and Billinghurst [24] suggested that gaining insights into cognitive issues underlying the effectiveness of existing AR systems is of significance for guiding and improving future design.

Mobile AR introduces more possibilities to AR interaction than stationary AR interfaces fixed to certain locations, but also presents new issues for developing effective AR systems. At present, our knowledge about human factors in mobile AR is very limited, and therefore it is crucial to comprehensively identify the opportunities and challenges posed by mobile AR interaction on the cognitive process. In this chapter, we provide an overview of cognitive issues involved in mobile AR systems through the lens of mobile AR interaction, in order to both clarify how mobile AR interaction concerns human cognition and offer guidance for mobile AR design in the future.

In the following sections, we will introduce the embodied perspective and explain our rationale for selecting it as the theoretical approach to organize cognitive issues in mobile AR interaction. We will further discuss the cognitive issues affecting mobile AR in detail, based on the findings of existing literature.

2 Embodied Cognition in Mobile AR Interaction

The development of computing technologies has revolutionized human–computer interaction (HCI) by driving people to actively and naturally interact with technologies within a broad range of contexts. The interaction penetrates to activities in people's daily life beyond working tasks, and ubiquitous and mobile computing makes it possible to flexibly manipulate technologies without geographical constraints. Instead of viewing the human mind as solely an information processor and cognition being isolated with action, several new theoretical approaches can be applied to the study of cognitive functioning in the field of HCI.

Embodiment, referring to "the property of our engagement with the world that allows us to make it meaningful," has been extended to the HCI domain in more recent years [23]. The embodied perspective of cognition suggests that cognitive processes are grounded in the bodily interaction in real space and real time [119]. The bodily engagement situated in specific physical and social environments can shape human cognitive process. As people's physical interaction style with the digital world becomes increasingly direct and inseparable from physical and social contexts, the embodied perspective serves as a suitable approach for analyzing HCI and generating implications for developing computing devices [51]. According to the embodied perspective, HCI involves users' constructing meaning and understanding by using technologies in physical and social environments rather than using technologies simply to implement tasks and process information.

Mobile AR demonstrates great potential for incorporating embodiment into users' interaction with computing technologies. It shifts the presentation of virtual information from onscreen displays to directly overlaying information onto the physical world. In particular, it entails the characteristic of context sensitivity for building meaningful relations between virtual information and the physical environment, which leads to a seamless merge of virtual and physical worlds [121]. Indeed, the role of physical environments as an integral part of activities is perceived as an important aspect of embodied cognition [119]. Mobile AR also offers a growing number of possibilities for users to physically interact with AR contents. By taking advantage of the features of different mobile interfaces, manipulation of AR contents has become a vital part of mobile AR interaction. There is a trend toward developing natural interaction techniques based on human manipulation skills and perceptions [25, 110]. The enhanced bodily engagement with virtual information expands users' capability of directly interacting with computing technologies to construct understanding of the setting. Apart from supporting individual activities, mobile AR also facilitates the establishment of spaces for shared experiences in indoor and outdoor environments. It enables social dynamics of meaningful construction in real world activities among multiple users whilst adding unique information produced by computing technologies. The interdependent roles of multiple users such as collaborators or competitors supported by mobile AR are also important in the effectiveness of shared activities [63, 124]. Specifically, mobile AR can augment the physical world and integrate physical resources with the mechanics of joint activities [80]. The movement and location of users can trigger certain events in mobile AR-supported activities. Therefore, users' engagement is important to mobile AR interaction, and the interaction is bound to both physical and social surroundings. Mobile AR systems should seek to strengthen users' construction of meaning and understanding in AR environments. These features of mobile AR provide evidence for using the embodied perspective to better understand mobile AR interaction and explain how the interaction supports cognitive functioning in the process of building understanding of mobile AR contexts.

On the basis of the characteristics of mobile AR interaction, we identify three primary categories of cognitive issues in mobile AR interaction: information presentation, physical interaction, and shared experience. The cognitive aspects related to each issue will be examined in the following sections.

3 Design for Mobile AR Interaction

In this section, we outline three categories of cognitive issues in mobile AR interaction including information presentation, physical interaction and shared experience. For each category, relevant aspects affecting users' cognitive functioning are discussed, and examples of mobile AR applications are presented in order to exemplify the role of cognitive issues involved in the use of systems.

3.1 Information Presentation in Mobile AR

The display of virtual information upon the physical world is a fundamental property of AR interfaces. Compared to virtual reality (VR), AR creates opportunities for enhancing real world objects and environments instead of replacing them. The realization of worth of mobile AR largely depends on maximizing the relevance of virtual overlays with regard to the physical world [48]. Annotations, which provide information relevant to one's physical surroundings, have become a mainstream component in current mobile AR systems in order to facilitate the understanding of the real world [121]. The meaningful integration of physical environments with mobile AR-supported activities can shape cognitive process in an embodied way [119]. To highlight the significance of arranging AR contents in a well-structured way, view management refers to "decisions that determine the spatial layout of the projections of objects on the view plane" [8]. Virtual information display is an integral part of the view supported by AR. In this section, amount, representation, placement and views combination of information are taken into account to clarify the impact of information presentation on cognitive functioning in mobile AR.

3.1.1 Amount

The co-existence of informative virtual elements and real world scenes in mobile AR interfaces delivers a unique opportunity for users to construct meaning in the physical world. Planning the amount of information available to users becomes a critical issue in mobile AR in order to support the cognitive process of making sense of information [8]. This concern has become salient as diverse platforms characterized with different properties like display spaces and field of view are gradually applied to mobile AR.

The synthetic scene may hinder cognitive functioning if visual complexity of real world background and virtual information is not well balanced in AR [102]. Presenting a large volume of information simultaneously can result in clutter display, which is characterized by information interference in a single display. Users may feel overwhelmed by the information and have difficulty focusing, thereby increasing their mental load when using mobile AR. By investigating the effect of cumulative cluster on human cognitive performances in AR, Stedmon et al. [102] contended that clutter display sets barriers for searching targets and the visual confusion has negative effects on users' understanding of information delivered by AR. The issue of information density displayed on handheld devices in navigation has also been investigated by Ganapathy et al. [33]. By showing labels of surrounding spots such as hotels, parks, and bridges, the scenario facilitated users' identification of points of interest around them. The authors found that users have certain preferences toward the amount of items presented on a single screen. Too many or too few items could increase the effort required to search and comprehend information. For military AR applications, the focused display was stressed in order to present the most relevant information to users during military operations; insufficient information negatively affected the maintenance of situation awareness, while too much information generated cognitive overload [73].

The cognitive effects of the amount of information also need to be addressed when applying mobile AR to realize X-ray visualization [92]. AR can not only display additional instructions in the real world, but also generate X-ray vision by visualizing the structure of invisible objects. The presentation of meaningful hidden structures requires delivering an appropriate amount of depth cues to aid the understanding of spatial location of occluded information with respect to the occluding scene, and preserving important information within occluding structures [55]. The amount of depth cues conveyed by mobile AR is vital for users to recognize the spatial relationship between occluded and occluding objects and thereby capture an accurate model of the environment [4]. Also, the display of virtual information onto the physical background can bring ambiguities to understand the real-world information [55]. Additional virtual information might occlude important information in the real-world scene, which leads to increased mental loads to make sense of the environment.

3.1.2 Representation

The representation of virtual information in mobile AR needs to be considered in order to increase users' efficiency of recognizing and comprehending information. The context-dependent characteristic of information representation has been acknowledged by a body of research [29, 121].

AR serves as a new form of representation to attach location-based information to the physical world scene [120]. Conventionally, the separation between information and physical spaces produces "cognitive distance" for users since they have to switch across spaces to extract targeting spots from information displays and then apply the information to real-world situations [60]. The transition of attention in these processes increases users' cognitive loads. For example, when offering navigation guidance for driving, "divided attention" caused by moving attention between information display and the road view could affect drivers' information processing and driving performance [60]. By overlaying virtual information about road conditions onto the windshield, mobile AR had stronger capabilities for narrowing the gap between geo-referenced information display and the physical space than a 2D bird's eye view map display; also, with AR people could more easily concentrate on the view in front of the car while gaining information regarding current and upcoming road conditions [60].

Although AR shows potential for promoting cognitive process in displaying geospatial data, the degree of enhancement varies depending on the methods used. The forms of virtual imagery in mobile AR can vary, including points, textual annotations, 2D graphics and 3D graphics. The representation of information impacts users' efforts to link virtual data with real locations in order to understand the physical world [108]. For example, 3D representations were found to be more capable of displaying spatial information than 2D representations; the high level of realism facilitated by 3D representations made the virtual scene more understandable [19]. For nearby buildings, plain and un-textured models were adequate to convey geospatial information [29]. Different AR presentation schemes of arrows for navigation guidance impacted users' interpretation of the distance [108]. Directly viewing spots in occluded areas in real time can allow people to identify their surroundings more easily compared to using a map with symbolic representations [111]. Investigations have also been carried out in order to obtain insights into how representations of information affect user experiences in urban planning activities. For instance, different representations including spheres, cylinders, and smoke were used to visualize the level of CO in outdoor environments with the support of a mobile AR system designed by White and Feiner [115]. They found that the type of representation affected users' cognitive and emotional reactions to the data. These users' reactions suggested that diverse types of representations should be designed to adapt to different contexts.

Additionally, the reference frame of information is critical to one's cognitive functioning in order to interpret surrounding environments supported with mobile AR. Egocentric and exocentric viewpoints serve as two primary reference frames in information presentation [76]. While the exocentric viewpoint provides a better overview of the surrounding context, the egocentric viewpoint presents information for local guidance based on the first-person perspective [2]. It is suggested that the egocentric viewpoint is more useful for inferring the spatial relationship between the user's current location and the plot of interest compared with the exocentric viewpoint [93]. To take advantage of both reference frames, Langlotz et al. [69] developed a mobile AR system that enabled users to first have a global view of a exocentric 2D map with nearby annotations, then when users got close to the annotated spot, they could switch the view to a first-person perspective in order to find a way to reach the actual spot.

3.1.3 Placement

Optimizing the placement of virtual information is an indispensible aspect to consider when managing the view in mobile AR. It is suggested the information layout can impact the understandability of information; appropriate arrangement of information helps users to connect the meaning of virtual information with the real-world view [48].

As an intuitive way of annotating physical objects, labeling plays a vital role in providing more information to support exploration in AR settings [125]. Label overlap and object occlusion are two crucial problems associated with placing annotations in mobile AR [125]. Visual clutter can cause meaning ambiguity in AR contexts and exert a negative impact on users' understanding of target objects [33]. In addition, the relative distance between the label and the target object could affect users' eye movement when reading information. In one study by Azuma and Furmanski [3], as the relative label distance increased, users needed to take longer time to read the label. Ganapathy et al. [33] provided empirical evidence that there is an acceptable maximum distance between the annotation and the target object to ensure the readability of information. In addition to the placement issue in each frame, the information layout in frame-to-frame transition is also a factor that relates to cognitive functioning. Unnatural changes of the positions of virtual information between two frames might lead to visual discontinuities; minimizing the visual discontinuity serves as a goal when managing the view in AR [8].

The dynamic feature of mobile AR should also be taken into account when planning the placement of virtual information. Mobile AR presents greater challenges for managing labeling placements than regular static backgrounds. Compared with the stationary setting, it is more difficult for wearable devices to locate spatial features of target objects such as position and orientation. This in turn affects the system's ability to identify the visible part of the object within a users' visual field and may lead to inappropriate label placements [75]. Also, the location of virtual information is crucial to users' visual attention. For mobile wearable AR, there are three main types of coordinate systems including *head-stabilized* where "information is fixed to the user's viewpoint," *body-stabilized* where "information is fixed to the user's body position," and *world-stabilized* where "information is fixed to real world locations" [10]. World-stabilized views enabled users to access to context-dependent information registered in the real world of the far field. Head-stabilized and body-stabilized views are commonly adopted in augmentations in the near field. In comparison with head-stabilized display, body-stabilized display has shown advantages in helping to understand the location of information and thereby enhancing the speed of searching information [10]. The influence of different spatial layouts of the virtual menu on the efficiency of selecting 3D objects also suggests that the placement issue should be considered when presenting information in mobile AR [117]. Rather than constantly concentrating on the information display, users of mobile AR supported by handheld devices usually switch their attention back and forth between the handheld display and the surrounding setting. The layout of information could impact the effectiveness of understanding the information at a glance [9].

3.1.4 View Combination

A set of visualization techniques such as *zooming and panning*, *overview and detail*, and *focus and context* are employed to combine multiple views on a single display [1, 2]. These techniques attempt to make full use of the display space and increase the efficiency of searching information [1].

Zooming and panning separates focused and contextual information temporally, which allow users to continuously zoom and pan the view in order to target desired information. The integration of zooming and panning with mobile AR assists users in getting detailed information in a consistent way [2, 14]. However, if people need to associate local details with the surrounding context to understand the meaning of information, zooming in to obtain details could result in the loss of global context, which could then increase users' required cognitive efforts [17].

Overview and detail separates focused and contextual information spatially by using two linked windows in a single view. The separation of two spaces requires additional visual navigation which may introduce new cognitive issues. Focus and context, such as fisheye view, seamlessly keeps focused information within the global context in one view via image distortion. It seeks to make the focus salient in the context and simultaneously enable users to grasp the spatial relation between focus and context. This type of visualization is useful when users have to frequently switch between focused and contextual information [7], but recognition and interpretation of distorted views are two primary issues affecting cognitive functioning [17]. The trade-off between the amount of distortion and mental load should not be ignored when applying distortion-oriented views to present concurrent information in mobile AR interfaces.

Rather than simply deliver information of immediate surroundings, a body of systems have sought to visualize information regarding off-screen and occluded objects by taking advantage of these techniques of view combination [2, 50, 93]. In response to the information presentation limitations of handheld devices, such as small screen size and narrow camera field of view, expanded views have been explored in mobile AR with attempts to navigate physical environments [50]. For instance, Sandor et al. [93] provided evidence that utilizing egocentric space-distorting visualizations in mobile AR is a strategy for displaying off-screen and occluded points of interest. The provision of focused and contextual information in a single view makes users more easily able to predict the spatial relation between the current environment and the point of interest without frequently switching attention across separated views. However, the distorted space could introduce new cognitive burdens for users due to increased effort required to interpret the distortion. Cognitive dissonance may arise from reading the reconstructed model of the real world [50].

3.2 Physical Interaction in Mobile AR

In more recent years, the interactive aspect of visualization has received increasing attention, and user-centred visualization has been specifically emphasized [1, 28]. According to the embodied perspective, the cognition is grounded in bodily engagement with technologies [119]. As more interactive techniques are available for mobile AR, considerations should be given to cognitive issues of physical interaction in mobile AR in order to enhance the effectiveness of systems. Navigation, direct manipulation, and content creation represent three typical physical actions engaged by users in mobile AR, and are described in this section.

3.2.1 Navigation

The advancement of mobile AR brings about innovative experiences to navigational activities by narrowing the gap between physical surroundings and abstract virtual representations. Spatial awareness, including “a person’s knowledge of self-location within the environment, of surrounding objects, of spatial relationships among objects and between objects and self, as well as the anticipation of the future spatial status of the environment,” is an important factor for evaluating the success of navigation [112]. Beyond just delivering straightforward information to users, the research on mobile AR has focused more attention on interaction dimensions in navigation in order to motivate users to do self-exploratory activities and enhance their spatial awareness of physical spaces [89].

Users’ ability to change viewpoints to browse information is crucial in order to explore the environment in navigation. Some mobile AR applications make it possible to actively transit across different viewpoints in real time to get desired information [2, 69]. By taking advantage of handheld devices, users can easily zoom in and out to search plots of interest, which overcomes the limited size of handheld device and enables users to gain desired information efficiently. Since the states before and after zooming can be varied in the view perspective or the field of view, the smoothness of transition is important to avoid introducing extra cognitive loads [2]. Photo-based AR interaction was designed to capture different viewpoints by taking snapshots in navigation, which allowed users to review previous viewpoints without physically revisiting those locations [103]. This method is identified as an effective way to reduce effort and time when investigating the environment.

The modality of information display also impacts users’ interaction in navigation. Users move their attention between information display and physical surroundings to cognitively map their present location and spatial relationships with targeting plots [112]. Mobile phones equipped with built-in projectors are applied to navigation in order to support information displayed on a mobile phone screen and projector [40]. A projector characterized by a big screen size and high resolution helps users effectively look for information. The combination of mobile phone screen size and projector display resolution reveals potential for increasing efficiency and comfortableness of navigation. Schöning, Rohs, Kratz, Löchtefeld and Krüger [97] examined the use of projector phones in augmenting information on a classical paper map. The projection display could overlay additional customized geographical information about plots of interest on the map, avoiding the split of attention between the screen display and the physical map in navigation. The user study indicated that the projection display was capable of enhancing the efficiency of navigation by reducing completion time and error rate [97].

Instead of simply aiding users in finding their destinations, mobile AR can convey rich location-based background information based on users’ interests during the navigation. Presenting information at two stages lets users browse general information at first; then after selecting certain favorite spots on the display, they can read additional detailed information to understand the target and make further plans regarding their movements [33, 86]. Also, some mobile AR systems have provided filtering options for users to control the visibility of information based on their preference, which contributed to reducing display clutter and fostering information comprehension [86].

3.2.2 Direct Manipulation

Manipulation is an important theme in users' interaction with computing technologies. Moving beyond traditional interaction methods based on WIMP metaphors (Windows, Icons, Menus, Pointer), a range of innovative interaction techniques are developed to assist natural and effective manipulations in mobile AR [15, 90]. The increasing level of physical participation can shape user experiences and affect their understanding of the world [62]. Indeed, the trend toward direct manipulation provides an important area to investigate the impact of mobile AR on cognitive process. Tangible interaction, direct hand interaction and multimodal interaction, each of which reveals potential for enhancing user experience in direct manipulation, are presented in this section to address cognitive issues involved in interacting with virtual information supported by mobile AR.

Tangible Interaction

Integrating tangible interaction with mobile AR interfaces has emerged as a new interaction paradigm in recent years. Rather than relying on specific input devices, users can physically manipulate traditional tools to interact with virtual information. Tangible interaction adapts to people's natural behaviours in daily lives, which in turn contributes to the intuitiveness of manipulation and reduces cognitive loads. Also, it is useful for making full use of rich human physical skills to foster input capabilities in mobile AR.

The effectiveness of using physical tools to perform different functions has a high significance in tangible interaction [88]. It is essential to enhance the richness of functionalities embedded in a single tool to manipulate virtual information. For example, a series of tiles with different functions were designed to support tangible interaction in mobile wearable AR [84]. Users could easily pick up different data tiles and arrange them on a whiteboard to link the virtual information contained by the tiles. Also, they could manipulate operation tiles to implement certain functions on virtual information, such as deleting, copying and help. Through combining input and output functions in tiles, the system made the manipulation simple and natural [84]. Virtual vouchers representing different species and relevant information were developed to support the identification of specimens in field work; users could flip the handle of the voucher and the change of position and orientation of the voucher results in displays of different characteristics of the specimen [116]. Tangible interaction is also expanded to interactively displaying spatial information in outdoor navigation. One such application enabled users to rotate a cube in different directions to browse information and target desired locations, which was more intuitive than using a paper map [78].

The physical capability of handheld devices opens up possibilities for extending tangible interaction in mobile AR. By taking advantage of the camera as an input channel, users are capable of naturally interacting with virtual objects by manipulating and altering the orientation and position of handheld devices [41, 64]. In such camera-based interaction, multiple functions are assigned to a single mobile phone. Two fundamental types of physical gestures are defined, including “static interaction primitives” and “dynamic interaction primitives” [90]. Static interaction primitives allow users to manipulate the object on the basis of different postures of mobile phone, such as pointing, rotation, tilting and adjusting distance, while dynamic interaction primitives depend on physical movement of mobile phone such as horizontal movement, vertical movement and diagonal movement. The combination and sequence of physical primitives relate to the ease and speed of manipulation. Empirical evidence has been provided to support the notion that tangible interfaces have facilitation effects on the speed of positioning virtual objects, but the advantage on rotating objects is not obvious compared with keypad input [44]. Although this type of tangible interaction could increase the speed of manipulation, the problem of interaction accuracy should not be ignored.

Direct Hand Interaction

In the human body, hands are conceived as the most natural and effective input device to perform direct human–computer interaction [27]. The ubiquitous feature of hands as input devices is able to assist the realization of AR within a wide range of environments. In recent years, there has been an endeavour to exploit the potential of users’ two hands in supporting direct manipulation of virtual objects in mobile AR [61, 82].

Direct hand interaction has been adopted as a natural input technique in outdoor wearable AR. Wearing a HMD leaves users’ two hands free, which makes it possible for users to use their hands as an intuitive input channel. By taking advantage of vision-tracked pinch gloves, interaction techniques were developed to enable users to directly manipulate virtual information through hand gestures in mobile AR [110]. Tinmith-Hand was known as a glove-based interface that allowed users to control virtual objects and create 3D models of buildings either within or out of the length of arm [82]. Users could rely on their hands to perform cursor operations such as selection, rotation, translation and scaling on 3D objects. For example, they could select certain actions in the menu by pinching their fingers and thumb, transit between different levels of the menu by pressing their finger against their palm and rotate/scale virtual objects by adjusting the distance or angle of both hands. The intuitiveness and ease of use of direct hand interaction in wearable AR have been recognized by a range of research [20, 82, 83].

Handheld devices provide unique possibilities for using hands to directly manipulate virtual objects without employing additional special equipment. When one hand holds the device, the functions of the free hand are needed to implement effective physical interaction [25]. One typical type of interaction is using the hand as a vehicle to communicate commands for manipulating 3D objects in mobile AR. Touch screen interfaces allow more intuitive and effective input modes than buttons, key pads and joysticks [52]. However, there are some limitations to touch screens when interacting with 3D objects. The small screen size makes it difficult to select objects and causes fingers to occlude the display, and patterns of 2D interaction such as pointing and clicking are not well-suited for manipulating 3D objects [52]. Hence, attempts toward designing alternative interaction techniques based on touch screen interfaces are important in order to enhance the effectiveness of spatial interaction. For example, in order to tackle the issue of occlusion, *SideSight* was designed to expand the interaction area by allowing users to manipulate virtual contents by multi-"touch" around the mobile phone [15]. It was capable of sensing the action of fingertips in the periphery around the mobile phone. Indeed, a combination of traditional touch interaction and off-screen gesture interaction enriched the interaction experience with the support of mobile phones [15]. Back-of-device interaction was proposed to enhance capabilities of touch-based inputs of handheld devices in mobile AR [61]. It empowered users to engage in interaction with virtual objects by using one hand at the back of display while the other was holding the device, which simplified the spatial interaction in addition to solving the occlusion problem. The efficiency of an interaction technique may vary from task to task and there is a need to match the technique to the action involved in activities [45, 52]. Another type of interaction has been developed to augment virtual objects on the palm or finger of the free hand. Users can directly manipulate 3D objects by changing the motion of the hand anywhere at any time [71, 100]. This approach makes it more convenient for users to realize AR contents with their hands and flexibly choose the viewpoint with which to inspect the content within the length of arm.

Bimanual interaction is characterized by using two hands to simultaneously handle one object [34]. Inspired by the conception of bimanual interaction investigated by Guiard [34], two-handed interaction has been explored in mobile AR interaction in order to increase the efficiency of manipulation through coordinating the actions of two hands [25, 96]. Using two-handed interaction shows advantages in fostering the performance of manipulating 3D objects compared to one-handed interaction [44, 46]. However, bimanual interaction cannot ensure better performance, and it is important to optimally assign functions to two hands [47]. In mobile AR supported by handheld devices, the non-dominant hand is usually used to control the viewpoint, while the dominant hand is used to manipulate virtual objects [14]. The concurrent action of two hands contributes to the reduction of shakiness of manipulation on the handheld device, which in turn increases the precision of interaction. Also, since users are empowered to merge two actions to engage in the task, the physical and cognitive efforts generated by alternating across different contexts to successively perform two actions can be reduced [35]. Guimbretière et al. [35] also posited that the smoothness of combining multiple manipulations influences the effectiveness of a hybrid interaction style. Two-handed interaction is also adopted in wearable AR. It is suggested that the accuracy of spatial interaction with objects in 3D such as rotation and scaling can be well executed through specifying the relative position and orientation between two hands [82]. By taking advantage of physical skills of using pen and notebook in daily life, personal interaction panels were created to support two-handed interaction to manipulate AR contents [89, 106]. The integration of a personal interaction panel with a HMD expanded the capability of spatial input in mobile AR by allowing users to select, rotate, drag and drop 3D objects floating in the physical world and alter the viewpoint to obtain desired information [106].

Multimodal Interaction

Multimodal interfaces engage users in interacting with virtual objects through multiple input modalities and/or output modalities in mobile AR. With the support of multimodal interfaces, users' have increased flexibility to choose interaction modes under different situations, which shows potential for increasing the efficiency of manipulation [22].

Integrating complementary multimodality is essential to the efficiency of interaction [26]. Efforts have been made to combine a class of input modalities on the basis of channels of human perception and communication [101]. This effort has been directed at complementing different modalities well and realizing natural interaction with technologies [101]. Currently, hand gestures and speech are the two main input modes of a range of multimodal interfaces in wearable AR applications [42, 65]. Incorporating speech as a means of input could augment the capability of hand gestures in directly manipulating virtual objects [54]. Gestures serve as an effective medium for carrying spatial information regarding object manipulation (location, movement manner, size), while speech supports commands that are needed to manipulate an object based on the description of its properties, which is truly important when the object is not visible to users' view. Empirical studies have demonstrated that a hybrid use of gestures and speed can positively affect the efficiency of spatial interaction in AR compared with unimodal interaction, because the former addresses the problem of ambiguity when implementing a command [54]. Gaze is also utilized in multimodal interaction to assist in the natural positioning of AR contents. Gaze input is valuable for promoting the effectiveness of hands-busy activities [5]. Gaze directions and the duration of fixation are assigned as commands to naturally position virtual objects, which can reduce cognitive loads since users do not need to engage in hand-eye coordination in hand-based manipulation [26]. The concept of multimodality is adopted in mobile AR supported by handheld devices as well. For example, alternative interaction techniques, beyond standard touch screen interaction, were designed to complement one another and promote input capabilities [15, 52]. So, it is critical to identify the strengths and weaknesses of each modality and appropriately define commands to support subtasks in an activity [26].

The combination of output modalities is also a critical aspect to support the interaction with virtual information in mobile AR. Incorporating multiple output modalities can adapt to users' preferences in different types of interaction involved in one activity. Kawsar et al. [59] adopted mobile phones and personal projectors to support manipulations of virtual contents during navigation, and the information could be displayed on both the mobile phone screen and projector. The personal projection caused the separation of input and output spaces, which then required effortful hand-eye coordination to transition attention across the two spaces in navigation. However, after the user discovering the target in navigational process, the large projection display was well-suited for him/her to manipulate objects with two hands.

3.2.3 Content Creation

User-created content in mobile AR interaction has attracted growing interest in recent years [121]. The user not only receives and manipulates AR contents, but also plays the role of author to produce virtual information in mobile AR. Authoring AR content contributes to increased availability of information and in turn enriches user experiences in mobile AR.

Annotating environments for information sharing is a main type of content creation in mobile AR. Rekimoto et al. [87] brought forward the concept of "augmentable reality" to describe mobile AR applications that empower users to generate virtual information such as textual, graphical and voice annotations and attach them to surrounding environments. Users are also able to communicate situated information with other wearable users and those using normal computers. By taking advantage of content creation, Reitmayr and Schmalstieg [86] developed a wearable AR system to support information creation and sharing among users in tourism. People could annotate their surroundings by adding predefined icons of different shapes and colours, and then share those icons with other participants. Explorations were also carried out to expand content creation in situ for ordinary users in mobile AR [68]. AR 2.0 has been discussed in recent years to highlight the importance of user-generated virtual information within the context of mobile AR [99]. The mobility and low cost of mobile phones make them a suitable choice for being AR authoring platforms [68]. Every person can be an author of AR contents in place, and then publish his/her information to the audience with the support of a mobile phone.

Locating precise 3D positions of in-place objects presents challenges for creating annotations in mobile AR. Wither et al. [120] suggested that the combination of the aerial photograph and the first person perspective view in situ allows users to create annotations in an easy way; the features of corners, edges and regions in the aerial photograph were useful for precisely annotating the scene. The switch of users' attention between the screen display and the physical site to verify the annotation point is a cognitive issue involved in creating contents, especially when labeling small objects [69]. A panoramic image of the surroundings was displayed to make users annotate the environment from the first-person perspective, which promoted the efficiency of locating target position when touching the display [69]. When the annotated object is larger than the size of display, new interaction styles are needed to aid users in identifying the target and creating annotations. A pagination mechanism implemented on mobile phones was developed to help users effectively change visualized objects and target the object to add new comments [74]. When the scene model of the environment is not available, judging the distance of the target object from the user might introduce cognitive burdens [121]. A series of pictorial depth cues were designed to help users determine the distance to the target and accurately annotate the feature [122].

The mobility of users is a concern for designing interaction to create information in mobile AR. The interaction accuracy is more likely to be increased if users are walking around. Touch screen interaction characterized with high intuitiveness and ease of use is commonly adopted in mobile AR [14, 52]. But under the condition that a user carries a handheld device in a mobile context, the unsteady status of view on the screen makes it hard to precisely interact with AR contents. In order to reduce the errors of annotation, a range of new interaction techniques have been developed [38, 70]. For example, freeze-set-go interaction allows users to freeze the real world scene first, and then add annotations once they are still and in a comfortable pose; users can then unfreeze the view when they finish authoring the content [70].

Sketch-based AR is another branch of applications designed to support content creation in an intuitive and flexible way. It allows users to create visual scenes for AR through sketching on the interface with tools such as a stylus. *Napkin sketch* was an example that assisted creative 3D image drawing on the tablet PC by taking advantage of sketch-based interaction [123]. The capability for supporting the transition to previous frames is essential for users to freely modify the content in the design process. In-place sketching is also applied to support 3D model creation in AR games [39, 53]. With in-place sketching, users are able to sketch game contents based on pre-defined sketching rules and symbols, and then play with those contents. Two users can also sketch game contents and manipulate them alternatively to engage in the game. Sketching itself can be considered playing if the aim of game is to design certain contents together.

3.3 Shared Experience in Mobile AR

In recent years, mobile AR has been applied to facilitate shared experience in multiple domains such as collaborative learning, urban planning, social gaming, and tourist guiding [37, 43, 80]. Mobile AR lets multiple people interact with virtual information while maintaining social dynamics in the real world. From the embodied perspective, the social experience is an essential aspect to influence cognitive processes of constructing meaning when people interacting with technologies [51]. It is necessary to analyze the affordance of mobile AR for supporting cognitive process in multi-user activities. Given the significance of social richness in facilitating human cognition shared experience, key components in social contexts of shared experience are discussed first in this section. Next, three fundamental issues, bodily configuration, artifact manipulation and display space, will be presented to yield insights into how they relate to the establishment of social contexts in shared experience supported by mobile AR.

3.3.1 Social Context

Mobile AR creates new opportunities for enriching collective activities. The social context in shared experience does not simply mean the co-presence of multiple people, but also includes people's social role and their awareness of shared experience in activities [63, 81]. Recently, the capacity of computing technologies to enhance social richness is emphasized when designing shared experience [18].

From this perspective, understanding the social context in shared activities is helpful for effectively applying mobile AR in multi-user experiences.

The presence of multiple social entities is a fundamental component in shared experience. With multiple users in shared activities, mobile AR supports two main types of physical presences: local co-presence and mediated co-presence. AR technology can augment shared experiences by enabling co-located users to view and manipulate virtual information in face-to-face situations indoors or outdoors [118]. Mobile AR can also expand distributed multi-user activities by establishing a virtually shared space. For example, multiple users situated in distributed outdoor contexts could solve a problem through interacting with shared virtual objects [89]. Another example allows an outdoor user to explore and annotate the environment with AR while an indoor user receives on-site information and exchanges ideas with the outdoor user through employing VR [49, 109]. Also, mobile AR enables users to share annotations created in situ with others over distance [69].

Rather than simply gather multiple users in co-located or mediated shared settings, mobile AR needs to be designed to support social roles of multiple users in shared activities [63, 124]. Mobile collaborative AR emerges as an important field in current mobile AR applications [85]. Being collaborators, users are required to continuously build mutual understanding toward a shared goal in joint activities [91]. Given the importance of building mutual understanding in collaboration, the dynamics of social interaction are identified as an indicator for assessing the effectiveness of technology in supporting collaborative activities [66]. The scale of collaboration can vary from context to context. To promote social interactivity in large-scale collaboration, Klopfer, Perry, Squire and Jan have assigned distinct roles to collaborators and delivered customized information to each role [63]. Collaborators, characterized by different roles, needed to share information with each other to jointly perform a task. The enhanced social interdependence strengthened their beliefs as a group and subsequently motivated their commitment to the social interaction. However, the degree of overlap among different roles has to be considered [63]. Too much overlap may weaken interdependence in collaboration, but too little overlap may negatively affect the amount of common ground among collaborators.

Mobile AR also possesses great capabilities to support multiple users as competitors in shared experience. Social aspects, such as types of co-presence of players, communication among players, relationship of players, are stressed in computing games in recent years [21]. It has been suggested that competition among co-located players shows potential for enhanced enjoyment compared to mediated and virtual co-play [32]. Co-located social gaming is a typical form of shared activity that can be facilitated by mobile AR technologies [124]. With respect to mobile AR, it does not only maintain the social dynamics of multi-player games in the real world, but also seamlessly integrates computer-generated contents with the physical setting. To enrich social experience, applications combining competition and collaboration in entertainment supported by mobile AR have been investigated [80]. Social interaction is perceived as a core component in the social context of this type of multiplayer games due to the fact that players in one team need to negotiate strategies to compete against the other team of players.

In addition to the role of multiple users, workspace awareness serves as a key element to address in the social context of shared activities. Workspace awareness refers to "the up-to-the-moment understanding of how other people are interacting with a shared workspace," and impacts the effectiveness of shared activities [36]. Specifically, workspace awareness is comprised of the awareness of other participants' presence, the interaction engaged in by other participants, and the happening of activities within the workspace. Given the interdependence among multiple users, it is important for computing technologies to convey workspace awareness when constructing a space for shared activities. By providing hands-on experience in a shared visual context, AR can present a situational picture and foster workspace awareness in multi-user activities [81].

As emphases are attached to social aspects of shared experience, there is a need to gain an insight into the mechanism underlying the effectiveness of constructing social contexts in shared experience. In the following sections, bodily configuration, artifact manipulation and display space in mobile AR are discussed to illustrate their roles in affecting the effectiveness of shared activities.

3.3.2 Bodily Configuration

Mobile AR facilitates the construction of shared spaces in order to engage multiple users in diverse collaborative and competitive activities in the real world. However, a shared space among multiple users does not guarantee enhancement of social richness [124]. The capacity of mobile AR to support multiple users' bodily configurations, such as location and movement, is a critical issue involved in the establishment of social contexts.

The mobility of users has been recognized as an important aspect in affecting social dynamics in activities [80]. Empirical evidence has shown that input devices enabling natural body movements can encourage communication among users in shared activities [72]. In mobile AR, the mobility around a game board has been greatly investigated [43, 113]. Multiple users sit or stand around the board and simultaneously manipulate AR contents with their own devices in the networked play. The arrangement of the game board influences users' physical movements in the shared space. Xu et al. [124] found that different game board configurations in a shared space enabled users to stimulate different physical and social behaviours during the play. Players adopted physical movements as a strategy to compete for a good position to track the board and perform the task. Also, they adjusted their locations based on the observation of the opponents' movements in the game. The involvement of social interaction and the awareness of one another's actions could be enhanced when the configuration of the game board is appropriately designed. Building a shared space where users can move independently can stimulate explorations in learning activities. For example, Kaufmann and Schmalstieg [57] contended that the physical setup of AR prompts users to walk around the 3D geometric model to obtain different viewpoints for understanding spatial relations and facilitating further construction.

In recent years, mobile AR research has increasingly examined users' mobility in a broad range of physical environments [18]. Rather than simply providing a shared space, the physical environment can serve as an integral element in multi-user activities. With the advantage of portability and mobility, handheld devices supported with AR technology introduce opportunities for users to physically explore the real world. The notion of "using the physical world as a game board" is proposed to highlight the importance of giving meaning to physical locations and movements in game contexts [13]. For example, Mulloni et al. [80] augmented physical locations with virtual characters and settings in game narratives, and players needed to move among different locations to collect and rearrange virtual information in order to solve the mission. Users got involved in social interaction with each other while carrying out physical explorations. Mobile AR introduces innovative components to collaborative field work in outdoor settings [63]. For example, collaborators with HMD were able to freely navigate the environment and exchange ideas in order to conduct investigations [89].

The physical setup of AR interfaces can have effects on organizing users' bodily configuration during activities, which relates to their engagement of social interaction to make decision or find out solutions. With the advantage of wireless connectivity, handheld devices are increasingly utilized as platforms to support collaborative activities. Morrison et al. [79] adopted handheld devices to augment a paper-based map to guide users' explorations in the physical world. The collaborators supported with a combination of AR systems and physical tools tended to gather around the device and discussed strategies together, compared to those supported with only a traditional 2D digital map. The enhanced joint attention positively affected the performance of problem solving within a group [6].

3.3.3 Artifact Manipulation

With the advance of interaction techniques in mobile AR, there are increasing numbers of endeavours to incorporate manipulation of AR contents into applications targeting multi-user activities [12, 89]. Designing physical interaction with virtual contents has become an important aspect in supporting the effectiveness of collective activities.

Cognitively, the manipulative artifact serves as a common ground for multiple users to negotiate meaning in shared activities. Direct manipulation of virtual objects, characterized by deep bodily involvement in creating and modifying shared artifacts, motivates users to jointly engage in explorations to complete the group task [58]. The hands-on experience contributes to the accumulation of common ground among multiple users in the ongoing process. Instead of artifacts being isolated from the interactional process, shared artifacts can shape social interaction patterns of users [104]. Additionally, interacting with artifacts offers an important source for promoting workspace awareness among participants [37]. Manipulative actions can publicly signal the behaviour performed by others, which in turn can exert influence on the effectiveness of group work.

Allowing users to manipulate virtual information independently is important to stimulate their participation in shared activities. Independence, characterized by users manipulating virtual objects and changing the viewpoint of objects individually, is recognized as an indispensible aspect in AR-supported collaboration [107]. Users are capable of interacting with virtual information based on their interests and knowledge, which stimulates their participation in activities. Also, enabling individual users to simultaneously interact with virtual objects shows promises for increasing the effectiveness of collaboration [111].

Some research has focused on exploiting interaction techniques to enhance shared experience in mobile AR [12, 77]. Tangible interaction is identified as an effective approach to assist in the manipulation of virtual objects in multi-user activities [12, 84]. Physical controllers with different representations provide a common ground that contributes to the establishment and sustainment of mutual understanding among multiple users. Tangible interaction characterized by high intuitiveness can minimize the distraction of action when participants are engaging in discussion [95]. Multimodal interaction has been investigated to strengthen the effectiveness of shared activities as well [5, 89]. For example, *MAGIC*, a mobile collaborative AR system for exploring archaeological sites, allows users to utilize both text and speech to post messages to their partners regarding actions or ideas; the messages can then be shared with co-located or distant users on the HMD [89]. Some researchers have examined gaze direction as a complimentary channel to coordinate collaboration with the support of AR technology [5]. They found that gaze-based interaction is especially useful for facilitating joint attentions in the interactional process to construct shared understanding in remote collaborative activities.

3.3.4 Display Space

The arrangement of output displays in mobile AR can also influence social interaction patterns and workspace awareness in multi-user activities. The manner of presenting information about users' interaction and its subsequent effects on group work should be addressed when designing shared experience supported by mobile AR [105].

Constructing a shared display space is essential to the effectiveness of multi-user activities. The shared display of information serves as a common focus for users, and has facilitation effects on problem solving by stimulating social interaction [94]. Sharing the interface among multiple users is widely applied in collaborative activities. For example, a shared on-site map was displayed on the HMD of individual users to coordinate the interaction among them in collaboration [89]. Projectors have been studied with regards to establishing a public display of information in AR-supported collaboration [11]. The small display screen on handheld devices is not a good option for presenting shared information. Recognizing projectors' advantage in expanding the public display space, some researchers have applied the projector phone to mobile AR, enabling users to flexibly view information on either their mobile phone screen or projection [40, 59]. People thought that the projection was suitable for collaborative activities because it was convenient for them to share ideas with each other around the shared display [59].

Also, as more interaction techniques are incorporated into multi-user activities in mobile AR, publicly informing the occurrence and progress of one's action to others is vital to foster situational awareness among users. Other participants' behaviors help individuals to adjust their own competitive strategies or offer elaborations for jointly solving the problem [43]. One limitation of applying AR in collaboration is that the personal information display exerts some negative effects on users' awareness of others' actions, which can make it more difficult for them to achieve mutual understanding of each other's manipulations of virtual information [114]. In order to enhance situational awareness, some research has taken advantage of multiple sensory feedbacks to indicate individual's behaviours to group members [43, 57]. Empirical evidence has revealed that audio feedback is a useful medium to raise users' awareness of actions performed by others [43]. In *Construct 3D*, colour schemes were applied to distinguish the contributions made by different users [57]. Facilitating the visibility of each other's interactions is important to distributed collaboration. The lack of physical presence can result in reduced visual cues and non-verbal interaction, which can then reduce the awareness of each other's actions and affect the construction of shared understanding. Thomas and Piekarski [109] adopted VR as a channel to provide the representation of the outdoor user's environment to enhance indoor user's awareness of his/her outdoor partners' action and context. Also, the connection of outdoor AR and indoor VR allowed users to engage in interaction over distance simultaneously.

Considerations should be also given to the personal display in AR-supported shared experience. Privacy is a critical issue in configuring spaces for multi-user activities [16]. In the context of collaboration, although people are required to work together to solve the problem, they still need to maintain individuality and engage in personal activities [36]. People sometimes expect to keep certain work or reflection private rather than sharing all information with others. With respect to social gaming of a competitive nature, participants should possess personal spaces to engage in individual actions in order to win the game [107]. Hence, it is necessary to develop hybrid interfaces to support both public and personal displays for multi-user activities in mobile AR.

So far, there have been some explorations on combining public and personal displays to enhance the effectiveness of shared activities supported by AR technology [30, 56]. For example, in *STUDIERSTUBE*, users can customize the view of scientific visualization to meet their own needs and exert self-control on whether to publicly display it to their collaborators or not [30]. A check in/out model was proposed to enable users to better perform collaborative and strategic work while collaborating remotely [56]. In this application, an augmented workspace is divided into two spaces: a public space and a private space. The user can perform actions in the private space if he/she wants to hide it from others. Additional personal interaction panels are integrated to construct a personal space for individual interaction. PDAs have been applied as platforms to allow users to make personal notes in addition to controlling the virtual object in AR [19]. In one study, people were allowed to view and manipulate virtual information privately through a personal interaction panel and choose when to display their work on a shared projection for further discussion [98]. Multi-player games have also used the personal interaction panel to keep individual player's actions invisible to their opponents [105].

4 Conclusions

Mobile AR is identified as a promising interface to support individual's direct interactions with technologies bound to physical and social environments. The close connection between computational resources and the real world creates many opportunities for users to actively explore their physical space. As mobile computing platforms advance the implementation of AR in diverse domains, designing effective mobile AR systems has become an important part in fulfilling the potential of AR technology to foster users' cognitive functioning.

Recognizing the significance of users' behavioural involvement in constructing meaning and understanding of mobile AR environments, this chapter approaches cognitive issues of mobile AR from an embodied perspective to examine how the involvement impacts cognitive functioning. Three primary cognitive issues are identified, which include information presentation, physical interaction and shared experience. A variety of issues involved in each aspect are addressed along with examples of existing mobile AR applications. Fostering a better understanding of cognitive issues is important in order to guide the design of mobile AR systems to enhance human cognitive functioning.

The above review suggests that cognitive issues are crucial to the effectiveness of mobile AR systems. Analyzing human factors from the lens of mobile AR interaction is helpful for yielding an insight into the opportunities and challenges for developing effective mobile AR systems. Furthermore, the influence of mobile AR on strengthening human cognitive process is context-dependent. Making a good match for the context should be taken into consideration when utilizing mobile AR technologies. In the future, more research is needed to evaluate the effectiveness of mobile AR systems from this embodied perspective and apply findings for improving the system design.

Acknowledgments The authors would like to thank Dr. Mark Livingston for his great help in providing comments and suggestions for modifying this work.

References

1. Aaltonen A, Lehikoinen J (2006) Exploring augmented reality visualization. In Proceedings of AVI 2006, Venezia, Italy, 23–25 May
2. Alessandro M, Dünser A, Schmalstieg D (2010) Zooming interfaces for augmented reality browsers. In Proceedings of MobileHCI 2010, Lisboa, Portugal, 7–10 September

3. Azuma R, Furmanski C (2003) Evaluating label placement for augmented reality view management. In Proceedings of ISMAR 2003, Tokyo, Japan, 7–10 October
4. Bane R, Höllerer T (2004) Interactive tools for virtual x-ray vision in mobile augmented reality. In Proceedings of ISMAR 2004, Arlington, VA, 2–5 November
5. Barakonyi I, Prendinger H, Schmalstieg D, Ishizuka M (2007) Cascading hand and eye movement for augmented reality videoconferencing. In Proceedings of 3DUI 2007, Charlotte, NC, USA, 14–17 March
6. Barron B (2003) When smart groups fail. JLS 12: 307–359. doi: 10.1207/S15327809JLS1203_1
7. Baudisch P, Good N, Stewart P (2001) Focus plus context screens: Combining display technology with visualization techniques. In Proceedings of UIST 2001, Orlando, FL, 11–14 November
8. Bell B, Feiner S, Höllerer T (2001) View management for virtual and augmented reality. In Proceedings of UIST 2001, 11–14 November
9. Bell B, Feiner S, Höllerer T (2002) Information at a glance. Computer Graphics and Applications 22: 6–9. doi: 10.1109/MCG.2002.1016691
10. Billinghurst M, Bowskill J, Dyer N, Morphett J (1998) An evaluation of wearable information spaces. In Proceedings of VRAIS 98, Atlanta, GA, 14–18 March
11. Billinghurst M, Kato H, Kiyokawa K, Belcher D, Poupyre I (2002) Experiments with face-to-face collaborative AR interfaces. Virtual Reality 6: 107–121
12. Billinghurst M, Poupyrev I, Kato H, May R (2000) Mixing realities in shared space: An augmented reality interface for collaborative computing. In Proceedings of ICME 2000, New York, NY, USA, 30 July –2 August
13. Björk S, Falk J, Hansson R, Ljungstrand P (2001) Pirates! Using the physical world as a game board. In Proceedings of Interact 2001, Tokyo, Japan, 9–13 July
14. Boring S, Baur D, Butz A, Gustafson S, Baudisch P (2010) Touch projector: Mobile interaction through video. In Proceedings of CHI 2010, Atlanta, Georgia, USA, 10–15 April
15. Butler A, Izadi S, Hodges S (2008) SideSight: Multi-"touch" interaction around small devices. In Proceedings of UIST 2008, Monterey, CA, 19–22 October
16. Butz A, Höllerer T, Beshers C, Feiner S, MacIntyre M (1999) An experimental hybrid user interface for collaboration. Technical Report CUCS-005-99, Columbia University, Department of Computer Science
17. Carpendale MST, Cowperthwaite DJ, Fracchia FD (1997) Making distortions comprehensible. In Proceedings of VL. 1997, Capri, Italy, 11–18 September
18. Cheok AD, Yang X, Ying ZZ, Billinghurst M, Kato H (2002) Touch-Space: Mixed reality game space based on ubiquitous, tangible, and social computing. Pers Ubiquit Comput 6: 430–442
19. Clark A (2004) Hybrid AR users interfaces in collaborative gaming. Honours Reports, Computer Science and Software Engineering, University of Canterbury
20. Costello A, Tang A (2007) An egocentric augmented reality interface for spatial information management in crisis response situations. In Proceedings of ICVR 2007, Beijing, China, 22–27 July
21. De Kort YAW, Ijsselsteijn WA (2008) People, places, and play: Player experience in a socio-spatial context. Computers in Entertainment 6: 1–11. doi: 10.1145/1371216.1371221
22. Dias MS, Bastos R, Fernandes J, Tavares J, Santos P (2009) Using hand gesture and speech in a multimodal augmented reality environment. Lecture Notes in Computer Science, Volume 5085/2009:175–180, doi: 10.1007/978-3-540-92865-2_18
23. Dourish P (2001) Where the action is. The MIT Press, Cambridge, MA, USA
24. Duh HBL, Ma J, Billinghurst M (2005). Human factors issues in augmented reality. In: W. Karwowski (eds.). International Encyclopedia of Ergonomics and Human Factors. London: Taylor & Francis.
25. Edge D, Blackwell A F (2009) Bimanual tangible interaction with mobile phone. In Proceedings of TEI'09, Cambridge, UK, 16–18 February
26. Elepfandt M, Sünderhauf M (2011) Multimodal, touchless interaction in spatial augmented reality environments. In Proceedings of the HCII, Orlando, FL, USA, 9–14 Jul

27. Erol A, Bebis G, Nicolescu M, Boyle RD, Twombly X (2007) Vision-based hand pose estimation: A review. Computer Vision and Image Understanding 108: 52–73. doi: 10.1016/j. cviu. 2006. 10. 012
28. Fikkert W, D'Ambros M, Bierz T, Jankun-Kelly TJ (2007) Interacting with visualizations. In Kerren et al. (eds). Human-Centered Visualization Environments (pp. 77–162), Springer-Verlag Berlin Heidelberg
29. Froehlich P, Obernberger G, Simon R, Reichl P (2008) Exploring the design space of smart horizons. In Proceedings of MobileHCI 2008, Amsterdam, the Netherlands, 2–5 September
30. Fuhrmann A, Löffelmann H, Schmalstieg D (1997) Collaborative augmented reality: Exploring dynamical systems. In Proceedings of IEEE Visualization '97, Phoenix, AZ, 19–24 October
31. Furmanski C, Azuma R, Daily M (2002). Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In Proceedings of ISMAR 2002, Darmstadt, Germany, 30 September-1 October
32. Gajadhar B, De Kort YAW, Ijsselsteijn WA (2008) Shared fun is doubled fun: Player enjoyment as a function of social setting. In: Markopoulos P, de Ruyter B, Ijsselsteijn W, Rowland D. (eds.), Fun and Games (pp. 106–117). New York: Springer.
33. Ganapathy S, Anderson GJ, Kozintsev IV (2011) Empirical evaluation of augmented presentation on small form factors-navigation assistant scenario. In Proceedings of ISVRI 2011, Singapore, 19–20 March
34. Guiard Y (1987) Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. J. Mot. Behav., 19: 486–517
35. Guimbretière F, Martin A, Winograd T (2005). Benefits of merging command selection and direct manipulation. ACM Transactions on Computer-Human Interaction 12: 460–476
36. Gutwin C, Greenberg S (1998) Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In Proceedings of CSCW 1998, Seattle, WA, USA, 14–18 November
37. Gutwin C, Greenberg S (2001) The importance of awareness for team cognition in distributed collaboration. Technical Report 2001-696-19, Dept Computer Science, University of Calgary, Alberta, Canada.
38. Güven S, Feiner S, Oda O (2006) Mobile augmented reality interaction techniques for authoring situated media on-site. In Proceedings of ISMAR 2006, Santa Barbara, CA, USA, 22–25 October
39. Hagbi N, Grasset R, Bergig O, Billinghurst M, El-Sana G (2010) In-place sketching for content authoring in augmented reality games. In Proceedings of VR' 2010, Waltham, Massachusetts, USA, 20–24 March
40. Hang A, Rukzio E, Greaves A (2008) Projector phone: A study of using mobile phones with integrated projector for interaction with maps. In Proceedings of MobileHCI 2008, Amsterdam, the Netherlands, 2–5 September
41. Harviainen T, Korkalo O, Woodward C (2009) Camera-based interactions for augmented reality. In Proceedings of ACE2009, Athens, Greece, 29–31 October
42. Heidemann G, Bax I, Bekel H (2004) Multimodal interaction in an augmented reality scenario. In Proceedings of ICMI. 2004, State College, PA, USA, 13–15 October
43. Henrysson A, Billinghurst M, Ollila M (2005a) Face to face collaborative AR on mobile phones. In Proceedings of ISMAR 2005, Vienna, Austria, 5–8 October
44. Henrysson A, Billinghurst M, Ollila M (2005b) Virtual object manipulation using a mobile phone. In Proceedings of ICAT 2005, Christchurch, New Zealand, 5–8 December
45. Henrysson A, Marshall J, Billinghurstz M (2007) Experiments in 3D interaction for mobile phone AR. In Proceedings of Graphite 2007, Perth, Western Australia, 1–4 December
46. Hinckley K, Pausch P, Goble JC, Kassell NF (1994) A survey of design issues in spatial input. In Proceedings of UIST '94, Marina del Ray, California, USA, 2–4 November
47. Hinckley K, Paysch R, Proffitt D, Kassell NF (1998) Two-handed virtual manipulation. ACM Transactions on Computer-Human Interaction 5: 260–302
48. Höllerer T, Feiner S, Hallaway D, Bell B (2001) User interface management techniques for collaborative mobile augmented reality. Computers & Graphics, 25: 799–810

49. Höllerer T, Feiner S, Terauchi T, Rashid G, Hallaway D (1999) Exploring MARS: Developing indoor and outdoor user interfaces to a mobile augmented reality system. Computers & Graphics 23: 779–785
50. Hwang S, Jo H, Ryu J.-H (2010) EXMAR: EXpanded view of mobile augmented reality. In Proceedings of ISMAR 2010, Seoul, South Korea, 13–16 October
51. Hurtienne J (2009) Cognition in HCI: An ongoing story. An Interdisciplinary Journal on Humans in ICT Environments 5: 12–28
52. Hürst W, van Wezel C (2011) Multimodal interaction concepts for mobile augmented reality applications. In Proceedings of MMM 2011, Taipei, Taiwan, 5–7 January
53. Huynh DT, Raveendran K, Xu Y, Spreen K, MacIntyre B. (2009) Art of defense: A collaborative handheld augmented reality board game. In Proceedings of ACM SIGGRAPH Symposium on Video Games, New York, NY, USA, 3–7 August
54. Irawati S, Green S, Billinghurst M, Duenser A, Ko H (2006) An evaluation of an augmented reality multimodal interface using speech and paddle gestures. In Proceedings of ICAT 2006, Hangzhou, China, 29 November-1 December
55. Kalkofen D, Mendez E, Schmalstieg D (2007) Interactive focus and context visualization for augmented reality. In Proceedings of ISMAR 2007, Nara, Japan, 13–16 November
56. Kame G, Matsuyama T, Okada K (2010) Augmentation of check in/out model for remote collaboration with mixed reality. In Proceedings of ISMAR 2010, Seoul, South Korea, 13–16 October
57. Kaufmann H, Schmalstieg D (2006) Designing immersive virtual reality for geometry education. In Proceedings of VR’ 06, Alexandria, Virginia, USA, 25–29 March
58. Kaufmann H, Dünser A (2007) Summary of usability evaluations of an educational augmented reality application. HCI 14: 660–669
59. Kawsar F, Rukzio E, Kortuem G (2010) An explorative comparison of magic lens and personal projection for interacting with smart objects. In Proceedings of MobileHCI 2010, Lisboa, Portugal, 7–10 September
60. Kim SJ, Dey AK (2009) Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In Proceedings of CHI 2009, Boston, Massachusetts, USA, 4–9 April
61. Kim SW, Treskunov A, Marti S (2011) Drive: Directly reaching into virtual environment with bare hand manipulation behind mobile display. In Proceedings of 3DUI 2011, Singapore, 19–20 March
62. Klemmer SR, Hartmann B, Takayama L (2006) How bodies matter: Five themes for interaction design. In Proceedings of DIS 2006, University of Park, Pennsylvania, USA, 26–28 June
63. Klopfer E, Perry J, Squire K, Jan M.-F (2005) Collaborative learning through augmented reality role playing. In Proceedings of CSCL 2005, Taipei, Taiwan, 30 May-4 June
64. Koh RKC, Duh HBL, Gu J (2010) An integrated design flow in user interface and interaction for enhancing mobile AR gaming experiences. In Proceedings of ISMAR 2010, Seoul, South Korea, 13–16 October
65. Kölsch M, Bane R, Höllerer T, Turk M (2006) Multimodal interaction with a wearable augmented reality system. IEEE Computer Graphics and Applications 3: 62–71
66. Kreijns K, Kirschner PA, Jochems W (2002) The sociability of computer-supported collaborative learning environments. Journal of Education Technology & Society 5: 8–25
67. Kruijff E, Swan JE, Feiner S (2010) Perceptual issues in augmented reality revisited. In Proceedings of ISMAR 2010, Seoul, South Korea, 13–16 October
68. Langlotz T, Mooslechner S, Zollmann S, Degendorfer C, Reitmayr G, Schmalstieg D (2011) Sketching up the world: in situ authoring for mobile augmented reality. Pers Ubiquit Comput 1–8. doi: 10.1007/s00779-011-0430-0
69. Langlotz T, Wagner D, Mulloni A, Schmalstieg D (2010) Online creation of panoramic Augmented reality annotations on mobile phones. IEEE Pervasive Computing 99. doi:10.1109/MPRV.2010.69
70. Lee GA, Yang U, Kim Y, Jo D, Kim K.-H, Kim JH, Choi JS (2009) Freeze-set-go Interaction method for handheld mobile augmented reality environments. In Proceedings of VRST 2009, Kyoto, Japan, 18–29 November

71. Lee T, Höllerer T (2007) Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. In Proceedings of ISWC 2007, Busan, Korea, 11–15 November
72. Lindley S, Le Couteur J, Bianchi-Berthouze (2008) Stirring up experience through movement in game play: Effects on engagement and social behaviour. In Proceedings of CHI 2008, Florence, Italy, 5–10 April
73. Livingston M. A., Ai Z., Karsch K., Gibson G. O. (2011) User interface design for military AR application. Virtual Reality, 15: 175–184. doi: 10.1007/s10055-010-0179-1
74. Lucia A, Francese R, Passero I, Tortora G (2010) SmartBuilding: A people-to-people-to-geographical-places mobile system based on augmented reality. In Proceedings of UBICOMM 2010, Florence, Italy, 25–30 October
75. Makita K, Kanbara M, Yokoya N (2009) View management of annotations for wearable augmented reality. In Proceedings of ICME 2009, New York, NY, USA, 28 June-2 July
76. Milgram P, Kishino F (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, E77-D(12): 1321–1329
77. Mogilev D, Kiyokawa K, Billinghurst M, Pair J (2002) AR pad: An interface for face-to-face AR collaboration. In Proceedings of CHI 2002, Minneapolis, MN, USA, 20–25 April
78. Moore A, Regenbrecht H (2005) The tangible augmented street map. In Proceedings of ICAT 2005, Christchurch, New Zealand, 5–8 December
79. Morrison A, Oulasvirta A, Peltonen P, Lemmelä S, Jacucci G, Reitmayr G, Näsänen J, Juustila A (2009). Like bees around the hive: A comparative study of a mobile augmented reality map. In Proceedings of CHI 2009, Boston, MA, USA, 4–9 April
80. Mulloni A, Wagner D, Schmalstieg D (2008) Mobility and social interaction as core gameplay elements in multi-player augmented reality. In Proceedings of DIMEA 2008, Athens, Greece, 10–12 September
81. Nilsson S, Johansson B, Jönsson A (2009) Using AR to support cross-organisational collaboration in dynamic tasks. In Proceedings of ISMAR 2009, Orlando, Florida, USA, 19–22 October
82. Piekarski W, Thomas BH (2001) Tinmith-Metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In Proceedings of ISWC 2001, Zurich, Switzerland, 8–9 October
83. Piekarski W, Thomas BH (2003) ThumbsUp: Integrated command and pointer interactions for mobile outdoor augmented reality systems. In Proceedings of HCII 2003, Crete, Greece, 22–27 June
84. Poupyrev I, Tan DS, Billinghurst M, Kato H, Regenbrecht H, Tetsutani N (2002) Developing a generic augmented reality interface. Computer 35: 44–50. doi: 10.1109/2.989929
85. Reitmayr G, Schmalstieg D (2001) Mobile collaborative augmented reality. In Proceeding of ISAR 2001, NewYork, USA, 29–30 October
86. Reitmayr G, Schmalstieg D (2004) Scalable techniques for collaborative outdoor augmented reality. In Proceedings of ISMAR 2004, Arlington, VA, USA, 2–5 November
87. Rekimoto J, Ayatsuka Y, Hayashi K (1998) Augmentable reality: Situated communication through physical and digital spaces. In Proceedings of ISWC 1998, Pittsburgh, Pennsylvania, USA, 19–20 October
88. Rekimoto J, Sciammarella E (2000) ToolStone: Effective use of the physical manipulation vocabularies of input devices. In Proceedings of UIST 2000, San Diego, California, USA, 5–8 November
89. Renevier P, Nigay L (2001) Mobile collaborative augmented reality: the augmented stroll. In Proceedings of EHCI 2001, Toronto, Canada, 11–13 May
90. Rohs M, Zweifel P (2005) A conceptual framework for camera phone-based interaction techniques. In Proceeding of Pervasive 2005, Munich, Germany, 29–30 May
91. Roschelle J, Teasley SD (1995) The construction of shared knowledge in collaborative problem solving. In: O'Malley C. (ed.), Computer supported collaborative learning (pp. 67–97). Berlin, Germany: Springer
92. Sandor C, Cunningham A, Dey A, Mattila V.-V. (2002). An augmented reality X-ray system based on visual saliency. In Proceedings of ISMAR 2010, Seoul, Korea, 13–16 October

93. Sandor C, Cunningham A, Eck U, Urquhart D, Jarvis G, Dey A, Barbier S, Marner MR, Rhee S. (2009). Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In Proceedings of VR 2009, Lafayette, Louisiana, USA, 14–18 March
94. Sareika M, Schmalstieg D (2007) Urban sketcher: Mixed reality on site for urban planning and architecture. In Proceedings of ISMAR 2007, Nara, Japan, 13–16 November
95. Sareika M, Schmalstieg D (2008) Urban sketcher: Mixing realities in the urban planning and design process. In Proceedings of CHI 2008. Florence, Italy, 5–10 April
96. Sareika M, Schmalstieg D (2010) Bimanual handheld mixed reality interfaces for urban planning. In Proceedings of AVI 2010, Rome, Italy, 26–28 May
97. Schmalstieg D, Fuhrmann A, Hesina G (2000) Bridging multiple user interface dimensions with augmented reality. In Proceedings of ISAR 2000, Munich, Germany, 5–6 October
98. Schmalstieg D, Langlotz T, Billinghurst M (2011) Augmented reality 2.0. In: Coquillart et al. (eds.), Virtual Realities (pp. 13–37), Springer-Verlag/Wien
99. Schöning J, Rohs M, Kratz S, Löchtefeld M, Krüger A (2009) Map torchlight: A mobile augmented reality camera projector unit. In Proceedings of CHI 2009, Boston, MA, USA, 4–9 April
100. Siltanen S, Hakkarainen M, Korkalo O, Salonen T, Sääski J, Woodward C, Kannetis T, Perakakis M, Potamianos A (2007). Multimodal user interface for augmented assembly. In Proceedings of MMSP 2007, Chania, Crete, Greece, 1–3 October
101. Seo B-K, Choi J, Han J-H, Park H, Park J-I (2008) One-handed interaction with augmented virtual objects on mobile devices. In Proceedings of VRCAI 2008, Singapore, 8–9 December
102. Stedmon AW, Kalawsky RS, Hill K, Cook CA (1999) Old theories, new technologies: Cumulative clutter effects using augmented reality. In Proceedings of IV 1999, San Francisco, USA, 24–29 October
103. Sukan M, Feiner S (2010) SnapAR: Storing snapshots for quick viewpoint switching in handheld augmented reality. In Proceedings of ISMAR 2010, Seoul, South Korea, 13–16 October
104. Suthers DD (2006) Technology affordances for intersubjective mean making: A research agenda for CSCL. Interactional Journal of Computer-supported Collaborative Learning 1: 315–337
105. Szalavári Z, Eckstein E, Gervautz M (1998) Collaborative Gaming in Augmented Reality. In Proceedings of VRST 1998, Taipei, Taiwan, 2–5 November
106. Szalavári Z, Michael M (1997) The personal interaction panel-a two-handed interface for augmented reality. In Proceedings of EUROGRAPHICS 1997, Budapest, Hungary, 4–8 September
107. Szalavári Z, Schmalstieg S, Fuhrmann A, Gervautz M (1998) Studierstube-An environment for collaboration in augmented reality. Virtual Reality 3: 37–48. doi: 10.1007/BF01409796
108. Tönnis M, Klein L, Dey A, Klinker G (2008). Perception thresholds for augmented reality navigation schemes in large distances. In Proceedings of ISMAR 2008, Cambridge, UK, 15–18 September
109. Thomas BH, Piekarski W (2002) Glove based user interaction techniques for augmented reality in an outdoor environment. Virtual Reality 6: 167–180
110. Thomas B H, Piekarski W (2009) Through-walls collaboration. Pervasive Computing 8: 42–49.doi:10.1109/MPRV.2009.59
111. Tsuda T, Yamamoto H, Kameda Y, Ohta Y (2005) Visualization methods for outdoor see-through vision. In Proceedings of ICAT 2005, Christchurch, New Zealand, 5–8 December
112. Veas E, Mulloni A, Kruijff E, Regenbrecht H, Schmalstieg D (2010) Techniques for view transition in multi-camera outdoor environments. In Proceedings of GI 2010, Ottawa, Canada, 31 May-2 June
113. Wagner D, Pintaric T, Ledermann F, Schmalstieg D (2005) Towards massively multi-user augmented reality on handheld devices. In Proceedings of PERVASIVE 2005, Munich, Germany, 8–13 May

114. Wagner D, Schmalstieg D, Billinghurst M (2006) Handheld AR for collaborative edutainment. In Proceedings of ICAT 2006, Hangzhou, China, 29 November-1 December
115. White S, Feiner S (2009) SiteLens: Situated visualization techniques for urban site visits. In Proceedings of CHI 2009, Boston, MA, USA, 4–9 April
116. White S, Feiner S, Kopylec J (2006) Virtual vouchers: Prototyping a mobile augmented reality user interface for botanical species identification. In Proceedings of 3DUI 2006, Alexandria, VA, USA, 25–26 March
117. White S, Feng D, Feiner S (2009) Interaction and presentation techniques for shake menus in tangible augmented reality. In Proceedings of ISMAR 2009, Orlando, Florida, USA, 19–22 October
118. Wichert R (2002) A mobile augmented reality environment for collaborative learning and training. In Proceedings of E-learn 2002, Montreal, Canada, 15–19 October
119. Wilson M (2002) Six views of embodied cognition. Psychonomic Bulletin & Review, 9: 625–636
120. Wither J, Diverdi S, Höllerer T (2006) Using aerial photographs for improved mobile AR annotation. In Proceedings of ISMAR 2006, Santa Barbara, CA, USA, 22–25 October
121. Wither J, Diverdi S, Höllerer T (2009) Annotation in outdoor augmented reality. Computers & Graphics, 33: 679–689. doi: 10.1016/j.cag.2009.06.001
122. Wither J, Höllerer T (2005) Pictorial depth cues for outdoor augmented reality. In Proceedings of ISWC 2005, Galway, Ireland, 6–10 November
123. Xin M, Sharlin E, Sousa MC (2008) Napkin sketch-handheld mixed reality 3D sketching. In Proceedings of VRST 2008, Bordeaux, France, 27–29 October
124. Xu Y, Gandy M, Deen S, Schrank B, Spreen K, Gorbsky M, White T, Barba E, Radu I, Bolter J, Maclntyre B (2008) BragFish: Exploring physical and social interaction in co-located handheld augmented reality games. In Proceedings of ACE 2008, Yokohama, Japan, 3–5 December
125. Zhang F, Sun H (2005) Dynamic labeling management in virtual and augmented environments. In Proceedings of CAD/CG 2005, Hong Kong, China, 7–10 December
126. Zhou F, Duh HBL, Billinghurst M (2008) Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In Proceedings of ISMAR 2008, Cambridge, UK, 15–18 September

Part III
Design Principles and Recommendations

Chapter 6
Mobile Augmented Reality: A Design Perspective

Marco de Sá and Elizabeth F. Churchill

1 Introduction

Mobile devices have been rapidly growing in power, capabilities and features. As a result, opportunities for designers and engineers to create exciting new experiences and applications are emerging. However, while the potential for innovative service design is great, there are also challenges when it comes to designing effective, useful and usable services that offer users long-term value. Challenges derive from smaller screen sizes that offer reduced real estate for content presentation and interactive elements. Challenges also arise from the complexity inherent in most usage contexts, and the fact that users are more often than not actually in motion—that is, moving through different contexts and set-tings [1]. Acknowledgement of form-factor, interaction and context-of-use challenges has led to the establishment of a new field of study called Mobile Human–Computer Interaction (MobileHCI) or Mobile Interaction Design [2]. The MobileHCI research field started emerging in the late 1990s with the advent of the second generation of cell-phones available to mainstream users [3]. Initially focused on how users inter-acted with hardware (e.g., keypads) and on designing usable interfaces (e.g., UI for the limited interaction modalities and small screens) [4], the field has evolved rapidly to encompass consideration of use context through reports of field deployments, the design of development environments, and the creation of innovation methodologies. Mobile HCI as a field is the fastest growing in the broader human–computer interaction (HCI) community.

Our recent work in this space has focused on mobile augmented reality ("MAR"). Augmented reality offers new affordances for interaction; it has potential to enhance users' experiences through provision of digital information relevant to real world surroundings, without depriving users of their context [5–8]. Through placement of virtual objects or feedback over reality users are provided access to information

---

M. de Sá (✉) • E.F. Churchill
Yahoo! Research, Santa Clara, CA, USA
e-mail: marcodesa@acm.org; churchill@acm.org

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_6,

that is typically not available to them using their own senses [6]. Augmented reality and mixed reality are established areas of investigation [6, 7], but only recently has the emergence of mobile augmented reality been made possible owing to the advent of increasingly powerful smart-phones and smaller, more effective sensors (e.g., GPS, cameras). Mobile augmented reality thus enables users to move around and explore what is around them freely within real world settings [7, 9]. While the promise of MAR has been discussed on the research world since at least 2000 [10], we are only just seeing the promise turning into a reality. MAR applications are becoming more sophisticated, and re-cent research suggests that MAR will be responsible for amassing an enormous amount of profit in advertising, games and applications in general [11].

In this chapter we present an overview of the challenges posed by MAR to designers and researchers and discuss on the benefits and drawbacks of the most used techniques and approaches, from a design and HCI perspective. We discuss the current state of the art and most commonly used methods for the design and evaluation of MAR. In addition, we present our own design methodology, through the description of our approach to designing a MAR application. Our application focuses on allowing users to find whom of their friends are in the local vicinity by displaying icons and avatars overlaid on the current camera-rendered location, showing distance and direction in relation to the user. We utilized three different techniques to prototype our concept: (1) a low-fidelity prototype (i.e., mock-up and Wizard of OZ); (2) a mixed-fidelity prototype using video; and (3) a high-fidelity working prototype used on an actual smart phone. Using these three prototypes, we con-ducted a set of field studies, gathering feedback from users in real life scenarios. We detail the benefits and drawbacks of each approach and the settings, scenarios and phases of the design process (such as during early ideation or later stage evaluation) in which they perform most effectively.

2 Mobile Augmented Reality

Early experiments with MAR required considerable amounts of equipment in their creation [6, 8]. Use testing relied heavily on complex, often unreliable infrastructures, using laptops and heavy equipment or requiring Wi-Fi connectivity in order to function [12, 13]. Such equipment arguably interfered with the user experience and therefore potentially invalidated the user's assessments of the benefits of such applications.

Things have changed: current smart-phones can easily support this type of experience in real world settings. In addition, utilizing GPS and integrated compasses, the combination of AR and location-enabled and positioning services [5, 14] means that it is now possible provide users with ways to use AR and gain easy access information about their surroundings while on-the-go. Today cumbersome, intrusive equipment [15] is no longer required to create these experiences [16].

Thus far, the majority of literature on MAR has been strongly focused on technical challenges and constraints: issues to do with set-up and management of equipment, issues with quality of image rendering; consideration of marker- or marker-less detection; and development of detailed and reliable location-based services [10, 12, 15, 17]. Location detection and/or image recognition algorithms, information retrieval from different data sources and computational efficiency continue to be the main focus for researchers working on MAR. Recently, however, the research area has been starting to address user experience. This has been driven in large part because of the proliferation of end-user services and commercial applications using MAR (i.e., Layar,[1] Yelp[2]). Research reports increasingly relate users' experience while interacting with MAR, investigate usability issues and pose design challenges for this technology.

2.1 Design Challenges Posed by Mobile Augmented Reality

While innovations in sensor technologies, developments in interaction and presentation modalities and increasing access to data streams in real-time are inspiring services and applications that were previously only imaginable in futuristic films, these new features and capabilities are not always ready for everyday use [18]. Challenges include:

- Discoverability—many users, even savvy smart-phone aficionados, are not aware of what services and applications are available.
- Interpretability—many users are unclear about the value these services and applications offer and most services and applications are unclear in their presentation of value beyond immediate entertainment and "wow" factor.
- Usability—many users find learning to use the applications challenging and find interaction features and interaction paths cumbersome, and often context of use is not well enough considered.

Usefulness/utility/meaningfulness—once in use, many users do not find these applications offer long-term value. While it is clear there are many opportunities for deeper interaction and experience design engagement for MAR services and applications, there are still few guidelines for would-be designer-developers and few documented accounts of how to design such rich experiences [18]. Although some reports on designing MAR can be found, most refer to highly complex settings and infrastructures, defined for very specific purposes [5, 6, 14, 19] or, as we note above,

---

[1] http://www.layar.com/.

[2] http://www.yelp.com.

are strongly focused on the hardware constraints [9, 20]. Studies on design or from a user-centred perspective are scarce [18].

One reason for this scarcity of studies, we purport, is that testing and validating ideas at initial stages of the design process is hard. Assessment of the usability of specific interface or interaction design features yield well to laboratory-based evaluation of static of minimally interactive mock-ups (e.g., is the target area large enough to be selected?, is the font readable and in what lighting conditions?, is the image occluding other content?), but full service ecosystem design, such as those imagined for mobile AR applications do not yield so well to such methods. To be able to offer effective feedback, users need to be able to experience MAR services and applications as they are intended to be experienced: "in the wild" with interactive experiences and data presented in real-time. However, development costs can be high, even if all that is developed is a simple prototype. Further, the field is not mature enough yet, nor are user experiences sufficiently common-place for there to yet exist "discount methods" (e.g., *guerilla* evaluation sessions, simple heuristics) for creating informed design elaborations; typically in more mature interaction paradigms, these discount methods appear in the form of design guidelines and heuristics wherein are codified typically successful or problematic design options. But to avoid making premature commitments to interaction design choices that require costly back-end infrastructure development which, in the end, *may not* actually be useful or usable, designers and developers in the emerging field of MAR need grounded design guidelines and methods for evaluating design options.

Our research faces this paradox head on. As is often the case, designers are required to come up with new ways to convey their visions and test their concepts when designing new services [18, 21]. To prototype the envisioned, fully-fledged service is obviously costly and will result in premature, often non-retractable, commitments within the service design; such premature commitments are precisely what iterative design and evaluation are aimed at circumventing. However, to get useful feedback from users, it is necessary to create an experience that is "realistic" or provocative enough of the envisioned scenario of use such that users are able to give meaningful and actionable feedback.

We ask: how do we simulate the experience in a realistic enough way to get feedback early in design on concepts and on interaction designs without spending too much effort on building working prototypes? What are the key features a prototype needs to have to simulate the experience sufficiently for users to be able to give us meaningful feedback?

To begin addressing these questions, in this chapter we offer an introduction to relevant papers that do try to address at least some of the issues above: discoverability, interpretability, usability, usefulness. We describe projects where user-centred approaches, embodied mostly through user studies and participatory design, have been used. We follow this overview of related work with a description of our own experiments with prototyping and evaluation of MAR during the early stages of design ideation where we attempt to address the issues of interpretability and usability.

2.2 Summary

In summary, the design of MAR services and applications faces the following issues:

- Owing to novelty and relatively low adoption, the number of available MAR applications is limited. Therefore, awareness and understanding of MAR is still relatively low. We need compelling ways to "tell the story" about and illustrate MAR services and applications so that users understand the concepts well enough to offer meaningful insights.
- MAR usually requires sophisticated infrastructures to be built before users can fully experience their capabilities and features. Guidelines and design heuristics are lacking for approaching the design of MAR services and applications.
- Validation and testing are complex due to the richness of the experience, especially out of the lab.

In this chapter we ask:

> How can we create a user experience that is sufficiently "realistic" and provocative for users to envisage the final service experience and thus give meaningful and actionable to developers and designers, even at the earliest stages of design.

3 Mobile Augmented Reality Trends

Technological advances in sensor technologies (e.g., GPS, accelerometers, gyroscopes, cameras) and their increasing inclusion in consumer mobile-phone handsets mean that MAR applications can utilize a number of modalities in their content presentation: vision, sound and haptics are the main modalities.

The most commonly explored form of AR is visual-based. Indeed, the majority of existing commercial applications used visual augmentation overlays. Examples are Yelp's Monocle,[3] Layar,[4] and various research/artistic projects where visual effects are used to create artistic renderings of reality or augment monuments and art installations with colours, drawings and animations [22–24]. Visual MAR applications typically overlay additional information on top of what the smartphone's camera captures, displaying a combination of the real world with added information on top of what is around the user. Some experiments have even utilized MAR as way to present information that is concealed by physical structures or simply by the users' orientation, displaying information and directions to content and data that is located behind the user or off the screen, exploring different types of presentation [25]. Typical use-cases fall into the categories of information

---

[3] http://www.yelp.com.

[4] http://www.layar.com/.

seeking/search, navigation, push content such as on-the-spot content recommendation and advertising, entertainment, gaming, or information augmentation such as provision of historic details [15]. The majority of such applications draw geo-referenced data from a variety of services (e.g., restaurant listings, pharmacies), providing information on the user's surrounding [9, 17] and augmenting the location context with relevant data [15, 26].

Two other trends within MAR, much less explored, make use of the additional feedback channels available on most smart-phones—haptics and sound [27–29]. The most common example is the eyes-free, turn-by-turn navigation system that uses speech to provide directions to its user, like for instance the Google's Navigation application.[5] Research efforts have started to focus on combining haptics and sound to augment reality at specific locations, providing context-awareness of additional information. These techniques have previously been explored in gaming apps (e.g., treasure hunts) and for commercial or advertising applications (e.g., play a jingle/song when close to a certain shop). Other approaches also consider body movement recognition and gestures as a way to interact with augmented reality experiences [30]; we speculate that, with the advent of in-home technologies like the Xbox Kinect[6] from Microsoft, users are becoming increasingly familiar with this kind of interaction so we can expect to see a growth in this area in the upcoming years.

4 Design Approaches for Mobile Augmented Reality

Although some instances of user-centered design methodologies [48] and techniques have been applied to the design of augmented reality systems [31], reports suggest that user involvement and user studies very rarely take place for the design of augmented reality as a field in general [32]. Design methods for MAR, being a substantially younger field, are even less explored. Table 6.1 summarizes the most relevant papers in this regard from our perspective. These case studies were selected because of their exposition of a specific formative or summative design perspective and/or process. The papers also represent contributions that span a range of different domains (e.g., games, visualizations, interaction paradigms, health care, shopping). While not all these papers are explicitly concerned with design methods and processed per se, they do provide some insight on design issues when it comes to MAR applications and services. The papers can be divided into two main areas: (1) MAR design experiments that describe some of the procedures used to conceive, design and test the concepts and (2) field studies where prototypes or systems have been evaluated in context, out-of-the-lab.

In the two following sections we summarise the content of these papers and tease out what the authors have to say about the design methods and processes they used.

---

[5] http://www.google.com/mobile/navigation/.

[6] http://www.xbox.com/en-US/kinect.

Table 6.1 Summary of used methods and techniques for the case studies discussed in this chapter

	Domain/application	Design approach	Type of prototype/ system	Evaluation approach	Participants	Data gathering techniques
Nilsson and Johanson [33] (2007)	Hospital vision-based augmented instructions	Cognitive sciences engineering Usability evaluation	High-fidelity head mounted display	In situ task based User study	12	Observation and questionnaires
Xu et al. [13] (2008)	Shopping	Iterative, user centred, ethnography	High-fidelity shopping app for smart-phones	Formative field evaluation	17	Observation, survey and semi-structured interview
Nigay et al. [29] (2002)	Archaeology	Scenario-based design	High-fidelity software prototype	Task analysis	Not available	Not available
Lee et al. [14] (2009)	New interaction approach	Not available	High-fidelity prototype on hand-held device	Lab-based evaluation	8	Questionnaires and observation/task analysis
Damala et al. [20] (2008)	Art Museum Guide	Scenario-based design	High-fidelity prototype	In situ evaluation	12	Head-mounted cameras for video and audio recording
Morrison et al. [26, 16] 2009–2011)	Map-enhancement/ navigation	User centred	High-fidelity prototype on smart phones	Open field study, role playing	37	Video-recordings, logs and field notes
Avery et al. [21] (2008)	See through vision-based augmented reality	User centred	High-fidelity prototype on desktop and mobile devices (mocked videos)	Mixed between indoor and outdoor settings Task-based approach Role play	34	Questionnaires and interviews
Schinke et al. [10] (2010)	Visualization (i.e., displaying concealed objects)	User-centred	High-fidelity prototype	Field study	26	Observation and questionnaires
You et al. [22] (2008)	Game	User centred	High fidelity prototype	Iterative field study	30	Logs (e.g., GPS, time), questionnaires and interviews

In each case we highlight the tacit or explicit design questions the authors pose, the design options under consideration [and where possible the final decision/outcome of the research in terms of design choice(s) made], and the methods and techniques used.

4.1 Evaluating the Usability and Effectiveness of Mobile Augmented Reality

The significant differences we find between traditional user interfaces and augmented reality have already motivated several studies trying to assess the adequacy of traditional techniques to this domain or even experimenting with new ones. In their work, Nilsson and Johanson [33] conducted a usability study on the use of augmented reality within hospital settings, using a cognitive sciences engineering (CSE) perspective. Rather than studying the system as a combination of different parts, this approach focuses on the system as a whole, including both users and the system itself. This approach was used to assess the user experience and user acceptance of a system for hospital workers who are required to read instructions during their activities. Although the system does not use a smart phone or traditional mobile device, it is still somewhat mobile as it is composed of a head mounted, video, see-through display attached to a laptop. Users had to perform an everyday task for their work context—the assembly of a medical device, following the instructions provided by the AR system. The study took place within the working place itself, as the CSE approach suggests that conclusions about the use of a new technology should be drawn in its intended use-setting. Arguably, in situ testing is even more relevant for assessing the effectiveness of applications intended for specialists in high-risk settings. Twelve participants interacted with the system and data were collected during the evaluation sessions through observation and questionnaires. The results showed that by applying a CSE perspective and conducting the evaluation in situ, at the hospital with end-users, issues emerged that wouldn't have been noticed otherwise. This real-world setting was crucial in revealing the pros and cons of the design, considering not only the usability facet of the experience but also dimensions such as the social impact that it might have on workers and the working environment. In particular, one of the most noticeable issues for this particular case was the physical appearance and size of the system, which posed some difficulties in the intended use environment.

Xu et al. [34] describe a system that supports in-store shopping using a vision-based mobile AR application. Their study focuses particular attention on the visual attention required to interact with a mobile device while shopping. As a starting point for their design process, the authors conducted an ethnographic study, divided into diary studies and interviews with 12 participants, over the duration of 1 month, to understand the role of mobile devices (a phone in particular) during the shopping experience. Results from this study were divided into four main categories (1) communication—for instance, taking pictures and sending them to friends, calling someone or chatting while waiting in line; (2) organizational—remembering product requirements, location or number of a store and reminding the time of purchase; (3) informative—monitoring biddings, searching for prices and promotions; and (4) transactional—purchasing using the phone. Based on the results from this initial study, a set of design principles and a prototype for a vision-based shopping tool were created. The tool, Point&Find, allows users to point at objects and retrieve information related to those objects. Transactions were not contemplated in this prototype, which focused on the three first categories, as these pertain more to the user experience and interaction process that takes place during the shopping activity. The prototype consisted of a working system composed of three main components: an object recognition function, the user interface that shows recognition results with the viewfinder and connection to various Internet services. To evaluate the prototype a formative field evaluation was conducted, using actual devices. The authors' intention was to create a more realistic understanding of usability issues, with a particular focus on cognitive and interaction concerns. The authors also conducted the study in a real shopping setting with 17 shoppers who were recruited on the spot. There were significant challenges to conducting the study, with issues arising with instrumentation that affected data collection, issues controlling the circumstances of the testing environment and recruiting participants for longer periods of time. However, results from the study clearly demonstrated that user attention constantly switched between the physical and digital world. In addition, this experiment allowed the authors to detect several patterns of attention switch (i.e., browsers, frequent switchers and immersed researchers). Clearly, these patterns showed that the application interfered with or changed their shopping flow, especially at a physical level, restricting their manual interaction with objects. The kinds of results the researchers observed underscore that evaluation techniques that abstract the use of a device away from real world settings (either empirical methods like lab studies or analytic techniques like task analysis), may not be the most suitable techniques for evaluating this kind of MAR applications and prototypes; the issues observed would simply not have arisen and not been documented but for the occurrence of the environmental factors and natural behaviour in the real world setting. As such, the authors highlight the need to, in addition to conducting studies within real world locations and settings, define tasks that mimic common behaviour for the task being analyzed during the design process.

In a similar experiment, Nigay et al. discuss the use of scenarios for the design of mobile and collaborative augmented reality [35]. The motivation behind their work arises from the object oriented and real world-based approach that characterizes some MAR applications, where physical objects and constraints of the real-world play an increasing role in the design process and resulting experience. The authors argue that field studies and scenarios are especially important for MAR design processes. Field studies force us as designers and evaluators to account for the context of use, involving consideration of physical, technical and social dimensions that are seldom predictable or articulated in initial design specifications and envisionments. Scenarios offer a discursive common ground for the collaboration between the design team and users. To illustrate their points, the authors applied a scenario-based design approach that has two stages. The first stage consists of the design of scenarios based on reports from users on their work practices while the second stage involves iteration and refinements of the scenarios based on an analysis of users' activities by the researchers—users are observed and recorded at their work site. The data that result from these two stages is then used to derive a set of functional requirements, which serve as the basis for the system specification. Once these requirements are defined, they are evaluated once again against the defined scenarios, now integrated with the specified functions, which the authors call "projected scenarios". Results from this evaluation stage are then used to create the interaction techniques, which are based on the resulting functional specifications. Once the system is developed it is later evaluated in location, assessing both functionality and usability. Highly iterative, this process is reportedly also most effective: the authors report a successful application of the method in the development of the MAGIC platform (mobile, augmented reality, group interaction, in context), a component of an archaeological prospecting system. In particular, this approach highlighted some inherent limitations to the domain and common practices that had to be addressed, especially for collaborative activities (e.g., collection of data in the field, contextual evaluation of elements and remote discussion between archaeologists). The projected scenarios that emerged addressed collaboration and data gathering, and propelled the design of an interactive system that offers mixed-reality features, allowing users to move objects between the physical and digital world. This was found to greatly facilitate collaboration and information sharing between local and remote archaeologists.

Lee et al. also requested end-users to validate their own work through a series of comparative studies [14]. Instead of focusing on new services or applications, Lee et al. propose a new interaction approach for MAR. Their approach, called Freeze-set-Go addresses some of the problems that result from manipulating objects and interactive items on mobile displays. Not only does the size of the displays affect how users interact with the interface, but also the new usage paradigms that require users to interact while walking or while making use of both hands, decrease accuracy and sometimes produce poor usability results. By allowing users to freeze a scene of the mobile augmented context, creating an image of the real worldview, the system allows users to interact with the items and with content that is overlaid on top of the image more accurately/effectively. To evaluate their system's performance the authors conducted a study with end-users assessing task performance when making annotations on a MAR environment. Although the experiment took place within a lab setting, the study involved people conducting tasks while simulating real world situations and poses. To do so, participants were requested to complete the tasks under four combinations of difficult and easy tasks, defined by the height in which the objects to be manipulated were located and the use of their interaction approach (FSG—Freeze-Set-Go) or a traditional MAR approach (without freezing the scene). In their study, and although not utilizing an open approach to the evaluation, or taking it to the field, the inclusion of difficult poses, simulating real-world settings and scenarios showed that their design worked better under difficult poses for users, without sacrificing time for accuracy. In sum, given the impossibility of taking the design out of the lab, the authors simulated some aspects of the real world in order to create a life like experience. This yielded better results, allowing for the detection of issues regarding accuracy and how different behaviours were restricted or motivated by different settings.

In 2008, Damala et al. described the design process of a MAR museum guide, offering insights into their approach and the value added by taking a user-centred, scenario-based approach [20]. Their process started with the definition of a set of scenarios and a list of possible functionalities that were presented to possible stakeholders, including museum professionals and technology specialists. Developers and stakeholders were involved in a participatory, collaborative process of potential system requirements specification. This stage was followed by the creation of the actual content that would be presented to users and associated with the works of art displayed at the museum. This process took place at the lab. The final prototype consisted of a mobile device that, when pointed at the paintings, displays 2D and 3D virtual objects. These objects can be interacted with and additional information and digital documents can be easily accessed. To evaluate the system, a set of evaluation sessions took place inside the museum itself, using field observation methods to assess system effectiveness, usability and utility. Users were requested to wear a belt that included a set of additional media recording devices to capture sound and video through a head mounted camera, used to record the users' interactions with the devices and museum art. The evaluation sessions took place during the course of 2 weeks. Throughout this period 12 users visited the museum using the system and were observed, filmed and interviewed. During each test, participants were required to locate the paintings that were augmented with additional information and to freely navigate in the content according to their preferences. Each test lasted between 25 and 60 min, followed by a 15-min interview. The authors highlight some of the major findings from their user study. In particular, the sheer number of visitors to the museum during some of the sessions posed some limitations to what participants could do; clearly lone visitors would have had a very different experience. The field trial therefore exposed assumption about people's physical space allotment and the quality of their line of sight to objects of interest. The large number of visitors also made it difficult for the participants to understand the audio that was being provided by the system. Further findings demonstrated that more playful content appealed more to users and that the overall experience was considered to had benefited from the digital guide. Overall, the environmental constraints and the real-world setting in which the study took place provided insight into issues that would not have been found had the study been conducted under a more controlled, lab setting. It is also notable that, while highly valuable, the technical set-up of the study had its challenges. The authors also comment on the inherent challenges posed by the novelty of the used technology. Qualitative approaches were needed, in addition to quantitative ones (e.g., questionnaires with Likert-scales), in order to gain a deeper understanding of the pros and cons of the system in use—only with a qualitative analysis of rich interaction data could analysts distinguish between usability and utility issues that were likely to persist from issues that arose specifically from the unfamiliarity of the users with the technologies themselves. The latter would likely be extinguished through experience while the former would not; this is a crucial distinction when documenting usage difficulties in user trials.

4.2 Field Studies

The work of Morrison et al. illustrates nicely how field studies, combined with some role-playing, have been used in the evaluation of AR experiences [36, 37]. While developing an application called MapLens, a MAR map using lens over traditional paper-based maps, the authors conducted field studies with end-users in a city centre. A very high-fidelity functioning prototype of the application was used for these studies. Twenty-six participants interacted with their application in a game-like scenario while 11 participants performed the same tasks using a 2D traditional map application. Video recordings, logs and field notes, questionnaires and interviews were collected. The authors studied how users held the devices, how they used their hands to interact with the device and maps, nuances of and shifts in their body posture, the kinds of manipulations that were applied to objects, how users walked while using the devices, and forms of collaborative use. In particular they highlight how the most common behaviour was for users to stop walking and to gather around the device and map to explore the area and review detailed information. Moreover, users' interactions with the environment were documented during the field trial. The possibility that MapLens could be used effectively in conjunction with billboards or other maps was revealed during the study—evidence that field trials can lead to creative invention as well as summative evaluation. In their paper, the authors emphasize that without taking the trial into a real-world setting, the study would not have offered such rich results.

Schoning et al. also utilize a field-based approach to evaluate a map-based AR application [40]. The authors' goal with their project is to overcome one of the main issues with magic lens approaches: the attention switching between the device and the physical map. With magic lenses, dynamic information is presented on the device's display when pointed at traditional maps. Here, as a solution for this attention switching issue and the relative small screen and resolution from mobile devices, the authors present a system that utilizes mobile projectors to display additional information directly on the map—i.e., the content is projected onto a map instead of being rendered on the device's screen. To test their concept the authors created a simple prototype using a small projector connected to a smart phone. Based on the information captured by the phone's camera, the system projects the augmented content on top of the map. The major benefit from such approach is that users no longer need to switch between the physical map to gain context and the device's screen to access the extra information. Additionally, the amount of information that can be displayed because of the larger projection is substantially greater when compared to the mobile device's screen. This larger display area also affords for easier collaborative use of the information that is overlaid on top of the map. A user study was conducted to validate the initial prototype. A set of 12 participants interacted with the prototype and completed a task using the system (i.e., find five parking lots on a map and identify the cheapest one). Results from this study showed the potential that this approach had in terms of collaboration and use by small groups, overcoming some of the limitations that magic lens interfaces pose.

Moreover, using the map TorchLight users completed the task 15% faster than using the magic lens approach. However, the low light intensity from the projectors pose some issues to this system as they become difficult to read in outdoor settings, requiring alternative technologies (e.g., laser projectors). Again, these details can only be fully understood and detected when conducting the evaluation process in realistic settings.

In their work, Avery et al. present a similar study and results [38]. To test a see-through vision MAR prototype, they conducted a set of field studies and indoor studies with different users groups, with a total of 37 participants. The system aims to allow users to see through objects and buildings by overlaying videos of what is behind on the screen. Two groups of participants took part in the user study. One group of 20 participants was assigned to an indoor setting and the other group of 17 participants was assigned to an outdoor location. The indoor test was completed on a desktop computer while the outdoor tests used a mobile system. To simulate the working system, the authors created a set of videos that replaced the real-time video stream that would ordinarily be received by the system from the cameras located at the required spots. These videos were used during the evaluation sessions at the locations they were filmed. One of the hypotheses was that users would be able to understand a video more quickly and comprehend its contents if seen in situ with a see-through vision system compared to watching it remotely through a LCD display. In addition to two basic tasks that each participant had to complete (e.g., identify locations based on the videos and AR content shown through the system), a scenario-based approach was also used. This latter task was designed to simulate an emergency rescue situation. To complete this task, participants had to locate three injured people and chart the best route to reach them from an adjacent building. The application enables participant to see through walls and buildings that occluded where the injured people were located. Results indicated that outdoor participants were more efficient at completing the tasks compared to indoor participants and that the videos were significantly easier to understand on the see-through system when compared to the desktop counterpart. Moreover, the learning curve appeared to be small, as most participants completed the second task in shorter times. The success of this study again points to the value of doing evaluations "in the wild"—in more realistic settings, but also that scenario-based approaches that re-enact real life situations, provide a valuable tool to validate mobile AR systems.

A similar project 2010 project by Schinke, Henze and Boll also tackled the challenge of providing information that is concealed or beyond the screen (besides or behind the user) [22]. The goal is to replace the traditionally used 2D mini-map that is often combined with the AR view and replace it with objects and points of interest that are displayed even if off the screen. The system works by displaying arrows pointing in several directions that indicate the existence of additional points of interest, even if these are located behind the user. Such information is not usually displayed on traditional augmented reality apps. To evaluate their approach, the authors conducted a user study. The study took place at a city centre and the 26 users who participated were recruited on location. After having the concepts involved in the system explained (i.e., points of interest and augmented reality), users completed a set of two tasks (e.g., identifying and locating points of interest). A questionnaire and observations were administered. The study results showed clear differences between the two approaches, suggest that MAR interfaces could be improved by use of 3D arrows. It is also noteworthy that although the study took place in the field, the data were simulated. Still, this approached afforded a realistic situation that allowed for field studies and that yielded positive results. Nevertheless, despite the positive outcome, the authors point out that further studies are required to assess whether the quantity and quality of the simulated information being displayed did not affect the realism of the experiment.

Another interesting example of how field studies play a relevant role and how they can be used for the evaluation of MAR can be found in 2008 work by You et al. [39]. The authors conducted a field study with end users to evaluate a mixed-reality mobile game, using both virtual and real world cards placed at different locations using a mobile phone. The study was divided into three stages. The first, a pilot test, was conducted to verify that the system was working properly. This portion was conducted without end users' participation. The second stage consisted of material preparation; interview scripts, questionnaires, storyboards and additional evaluation material were prepared during this stage. Five users were requested to assist during this stage conducting a meta-evaluation, testing the procedure and the evaluation material and allowing for the adjustment of the process for the final evaluation stage. Quantitative data were collected through logs (e.g., GPS data, trail and time) and qualitative data through interviews and questionnaires. In addition to these two data collection techniques, users were also followed and observed throughout the game/evaluation session and, at points, interviewed during the game, following a method like that advocated contextual inquiry. The final stage replicated the previous one with 30 players and some minor study design adjustments. One of the adjustments was the composition of teams following suggestions offered by earlier participants; users from the second stage commented that the game would be more entertaining if played with friends. The authors report results throughout the process. Users' reflections on distances, on set up difficulties and on the game's ability to sustain engagement are reported. The field study also allowed for the assessment and impact of contextual details such as terrain difficulty and safety. Such factors have a direct impact on the experience and can shape the design of the game, but often are overlooked in more device/technology-focused studies.

4.3 Summary

Although mobile augmented reality is a relatively recent field of research and development, there is already a significant body of published work. However, much of this published work is very technology-oriented. Very little of the published work presents authors' design philosophy, design perspective or design methodologies deployed during the design and development of these systems. Discussions of user participation during early stage design and design processes to elaborate user needs are rare and use of low and high fidelity prototypes infrequent.

However as Table 6.1 illustrates, user centred design approaches and field studies are gradually becoming more popular; a more user-oriented perspective is evident from 2007 onwards. Out of the nine case studies summarized in Table 6.1, eight included users and some sort of user study while designing their system. In particular, it is noteworthy that the majority of these user studies were conducted in realistic settings, mostly in the field.

In the following sections, we present our own approach to the design of MAR applications where we have been focusing on prototyping and evaluation techniques that provide a sufficiently rich experience to enable the gathering of relevant feedback to drive design insights, inspire interaction methods and select between design alternatives.

5 Prototyping Mobile Augmented Reality

In this section we discuss the design of a social media, MAR application we have been developing in our group. A major challenge we have faced in the design and development of these applications is the creation of high-enough fidelity prototypes to conduct meaningful and effective evaluation of design concepts. By "high-enough fidelity" we mean creating prototypes whose embodiment (form factor plus appearance plus interactivity) is suggestive enough for users to be able to give an accurate estimate of their likely utility and usability. Our goal is to, with the least effort/time/design commitment, create effective props to simulate the kind of interactivity that the final application will support. Too little interaction and users are left unable to imagine usage scenarios effectively, yet truly high fidelity prototypes require too much development time—in some instances a high fidelity prototype can lead to infrastructure design decisions, once developed, remain in place simply because they are too costly effort-wise to dismantle and rebuild—irrespective of whether they are (or are not) interactionally effective and elegant. This, therefore, defeats the purpose of the formative evaluation.

Our aim in this chapter is draw on insights from previous work, outlined above, by following field trials but to offer a cost–benefit analysis of different type of prototype from low- to high-fidelity for effective user-driven design elaboration and evaluation in this space. Below, we describe the design probes we developed in the early stages of developing our Friend Radar application.

5.1 The Design of Friend Radar

The Friend Radar application merges social networks, messaging tools and location-based services, and makes use of AR. Information is presented in a way that allows users access without losing the current, local, physical context around them. The Friend Radar Application draws data from existing technologies such as social

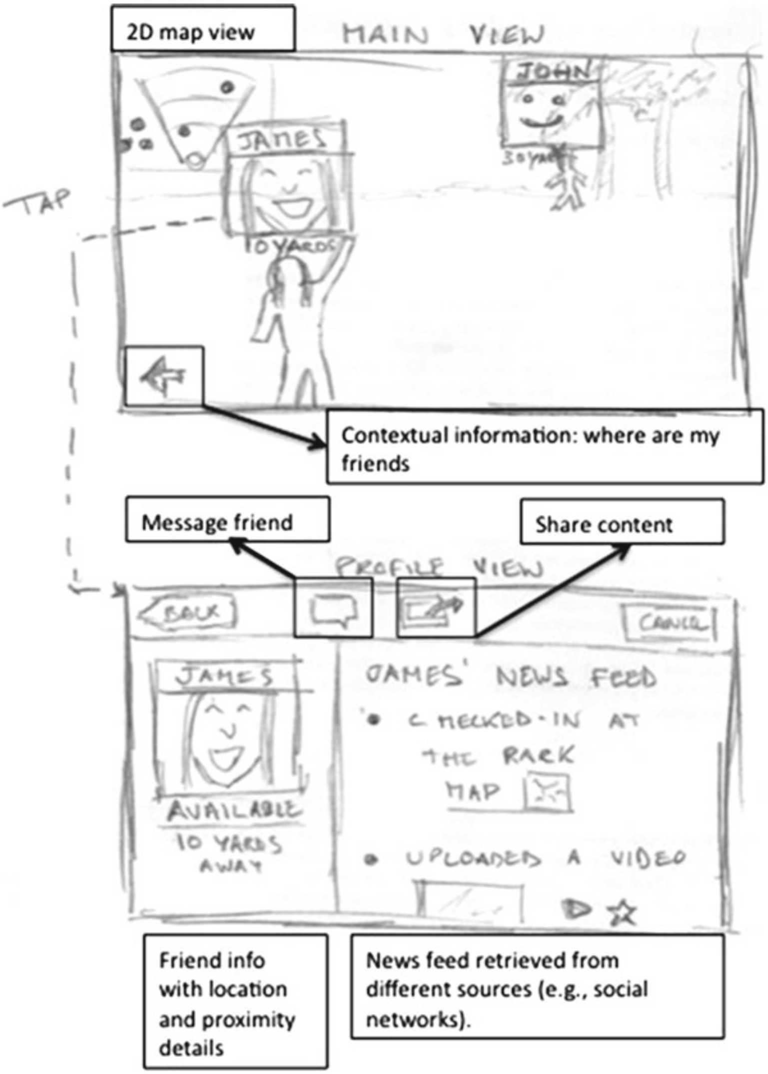


Fig. 6.1 Initial sketches for the Friend Radar prototype. The *top figure* shows the main view where the camera feed is displayed and the various friends are identified and their avatars overlaid on top of the real world view. *Below*, the user profile is shown with information about distance, availability, and recent activity retrieved from different sources and sharing options

networks like Facebook and Google+, location based apps/services such as Foursquare and similar check-in tools and previous experiments designed specifically for mobile devices like DodgeBall Social [38]. However, the Friend Radar enhances the experience by providing an enriched visual display of friends overlaid on the users' surroundings (Fig. 6.1). In particular it provides added affordances that allow users to see where friends are situated in relation to him/her, and their distance from

Fig. 6.2 (*Left*) Low-fidelity mock. The see-through hole allows users to see what is behind the device, simulating the camera. (*Right*) Low fidelity mock and used icons (*right*) next to an actual working device with the hi-fidelity prototype (*left*)

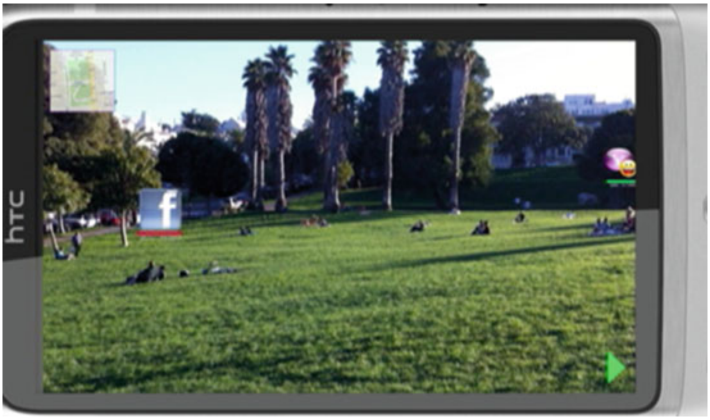


Fig. 6.3 Video of the mixed-fidelity being played on a smart-phone

the user. It also presents details such as preferred friends, directions where groups of friends are located and their availability. Unlike map-based radars, this AR based concept does not require users to translate two-dimensional views into their surroundings [41].

Our goal with this application, in this initial stage, is to show friends' positions in relation to the user, including details such as proximity and availability. Friends' location and status can be retrieved from various sources (e.g., Messenger apps, Social networks) and their avatars personalized with different images or pictures. User avatars are overlaid on the camera view of the location at which the user is pointing his/her camera/device (Figs. 6.1, 6.2 and 6.3). In addition, tapping on a friend's icon displays his/her profile with information from the service they are using at that moment and allows for some types of communication (e.g., sending a message or alert). The application thus merges social network data (both the contact's information but also the contact's relevance—such as, emailed you yesterday—for the viewer) with the real world, enhancing it with additional, contextually relevant, information. This in-formation contains both the user's location, proximity as well as data such as mutual interests, recent check-ins, multimedia capture within the context of both users' location and other information that has the potential to trigger a physical connection or some form of collaborative activity.

5.1.1 Design Approach

Our aim is to understand the best way to prototype for MAR. To achieve this goal, we began by utilizing some techniques based on previous experiments for mobile interaction design [1]. As suggested by recent literature, better results can be achieved while designing mobile interaction when immersing users within realistic scenarios, including prototypes and outdoors tests [1, 42, 43]. In order to do so for MAR, our goal was to approach the design process by experimenting with different prototyping techniques and exploring the benefits and drawbacks of each.

5.1.2 Prototypes

Three different mocks/prototypes for the Friend Radar concept were created. To build these we used three significantly different approaches but following the same philosophy: to create a prototype as close to the real experience as possible in terms of form factor and weight. Each is described below.

Low-Fidelity Mocks

The lowest fidelity prototype used for this study was built using a dummy/non-functional device (e.g., a product design mockups that was created typically to illustrate a form factor design). The dummy phone mimicked a common Android device with a 3.5-in. screen (Fig. 6.2). In order to simulate the camera feed, the screen was removed and a hole was cut on the back cover of the phone (the inside of the dummy phone is hollow). A transparent screen was placed on top of this hole to simulate the device's screen (e.g., reflection) and to allow users to easily use it as a touch input device, while maintaining the ability to see through it (see Fig. 6.2). In addition, this screen was also used to allow for an easier use of the Wizard-of-Oz technique, where small icons were glued to the screen to simulate the augmented reality. For the moving avatars longer pieces of paper were used. This facilitated their movement by the "wizard" simulating the actual location-based interaction.

The building process lasted approximately 1 h, including the creation of the attachable icons (e.g., map, avatars).

Mixed-Fidelity Videos

For the second prototype, a mixed-fidelity prototyping approach was used, combining aspects from both low and high fidelity prototypes. In particular, for this case, and following the categories defined by McCurdy et al.'s [44], the degree of visual and aesthetic fidelity was high while the interactivity remained low. Our hypothesis was that video would be adequate for simulation of a realistic field-based experience at during early formative design, as movement can be simulated and additional content can be placed over previously captured footage and displayed directly on the device.

To create the prototype, two different locations were selected. Videos were shot at the selected locations. For the first location, a public park with a few people sitting down and walking around was selected (Fig. 6.3). The second set of videos was shot in a busy square with shops, buildings and people walking and standing in different areas. Both locations were selected because they represent places where friends usually meet up, seek encounters and congregate. Each video had an approximate duration of 30 s and included light panning and some jitter to emulate a realistic usage scenario (i.e., scanning the area for friends). Once the videos were captured, they were edited and the friends' icons and avatars were overlaid using video editing software. This process took around 1 h for each of the videos. The videos were exported to a phone, used during the evaluation sessions.

High-Fidelity Prototype

The high-fidelity prototype was developed using the Android Development Kit. The prototype uses the camera feed, displaying it live and showing whatever the camera is capturing. On top of this feed, shown on full screen mode, a set of icons and avatars is also displayed on semi-fixed positions. Using the accelerometer and compass sensors, whenever the device is rotated, the icons and avatars will maintain their position in relation to the surrounding environment. They will be occluded when the device is not facing the icon's position. This offers an accurate rendering of MAR (Fig. 6.4). In addition to the interactive view, this prototype also supports interaction with some of the icons. Once an icon is tapped, a second screen displaying detailed information about the selected entity (e.g., a person) will be shown (see Fig. 6.1). The working prototype required approximately two working days to be fully developed. We note that the icons had already been designed for the previous prototypes.

5.2 *Evaluation and Discussion*

These three prototypes were shown to end-users at different locations and interacted with during some outdoor experiments. Three groups of eight users each

Fig. 6.4 High-fidelity prototype working on an Android smart-phone

experimented the prototypes—only one type of prototype per group. A summary of the results organized according to the following design stages and goals can be seen in Table 6.2:

- Probing—triggering users' imagination in a provocative way and using the prototypes to explore new applications, concepts and usages for the technology being studied [45, 46].
- Concept validation—addressing the concept in general, the overall idea of the application and its goals, by presenting it to users and requesting their feedback.
- Feature validation—validating the different features and functionalities that compose the application in more detail.
- Usability testing—addressing interface usability issues and breakdowns and assessing efficiency and ease of use [47].
- User experience evaluation—understanding users' feelings, opinions, expectations, acceptance, pleasure and deeper emotions regarding the experience of use.

As expected, each type of prototype has its benefits and drawbacks. However, while all have positive aspects and are adequate to particular stages of the design process, two types of prototypes stand out for the best and worst reasons. On the one hand, low-fidelity prototypes, which have shown to yield good results from mobile design , were not favoured by participants and provided poor results during the experiments. Major complaints pointed out the distracting process of simulating the movement of the different icons (i.e., Wizard of Oz) and the lack of interactivity of the prototype. Even considering the easy to update and adjust on the fly approach that these prototypes afford, they did not work well for usability testing and functionality validation. More importantly, the cumbersome nature of the experiment and Wizard of Oz approach with this type of prototype also affects they way in which these prototypes can be used to probe users and experiment with different ideas very quickly.

Table 6.2 Summary of results and adequacy of each prototyping approach for design stage

	Probing	Concept validation	Feature validation	Usability testing	User experience evaluation
Low-fidelity (Mock-ups and WOz) Figure 6.2	Can be good for probing as it allows to easily add features but cumbersome to experiment	Not very good to demonstrate the concept or interaction flow	Difficult to demonstrate features but easy to include new ones in situ	Not great to test usability issues, but enough to validate icon size and readability	Not adequate to evaluate more complex dimensions such as excitement, aesthetics, etc.
Mixed-fidelity (video) Figure 6.3	Not great for probing as it is non-interactive and non-flexible	Very good for concept validation as it shows features and interaction flow very easily	Good for simple features. Difficult to demonstrate more complex features as it is non-interactive	Not ideal for usability testing but still allows for the detection of some issues	Good to assess some aspects of user experience (e.g., aesthetics, flow)
High-fidelity (functional prototype) Figure 6.4	Supports probing to some extent but does not allow for in situ add-ons	Good for concept validation mainly because it is interactive	Good for feature validation as it allows users to explore them in detail	Very good for usability testing as it supports interaction and functionality	Good for experience evaluation, if care is taken to polish the interactivity

Mixed-fidelity prototypes, on the other hand, were very easy to understand by participants, provided a great way to discuss features and brainstorm over the concept. Moreover, even with limited interactivity, the mixed-fidelity prototypes also allowed for the detection of some usability issues (e.g., avatar size, amount of information displayed at the same time, labels and even the application's layout). Considering the time spent to create each of the video prototypes, these showed the best trade-off in terms of cost-effectiveness, being very easy to build and providing great results at both early and later stages of design.

Finally, when it came to the high-fidelity prototype, the results fell shorter than we initially expected. The prototype's interactivity, functionality and the fact that users were actually using a real device raised expectations to such a point that every feature was faced and interpreted as final and working. Although it provided some room for probing and brainstorming, the few glitches and minor bugs distracted users from the concept being tested and detracted from the exploratory nature of the experiment. The observation that prototypes which are too polished can result in user disappointment, more critical assessments and less creative feedback has been observed elsewhere [48]; often more sketch-like prototypes lead testers to creative insights as they "fill in the gaps" [cite Buxton's book here]. The tension is to support a close-enough experience while allowing room for creative feedback. Of course, higher-fidelity prototypes are likely to provide good results for functionality validation and usability testing later in the design phase. However, considering the time and effort required to build this type of prototype, these are not always adequate for early design stages or as props for ideation and scenario-based experiments.

6 Conclusion and Future Works

MAR is a fast growing and increasingly relevant field; researchers and commercial concerns alike are focused on building the next generation of innovative products in this space. A wide variety of services and applications are taking advantage of the benefits that augmented reality provides. These are especially interesting when used on mobile devices where users are free to interact and see the world augmented by information that would not be available otherwise.

However, despite the appeal and the growing number of services and applications, very few guidelines, design techniques and evaluation methods have been presented in the existing literature. In this chapter, we posed the question:

> How can we create a user experience that is sufficiently "realistic" and provocative for users to envisage the final service experience and thus give meaningful and actionable to developers and designers, even at the earliest stages of design.

We provided an overview of methods and techniques that have been reported in the literature for the design of a variety of MAR experiences, ranging from maps, shopping tools, games and even museum guides. Different modalities and interaction paradigms were discussed; our summary focused on design process and in particular the way in which the design process proceeded, the experience was conceived and prototyped and how those prototypes were validated with end users. A common trend observed was the use of field-based evaluations, experimenting mostly with high-fidelity prototypes and end-users within the context where the services and applications will most likely be used.

We noted that, to prototype the envisioned, fully-fledged service is obviously costly. More problematically, development without design evaluation can result in premature, often non-retractable, commitments within the service design; under-examined, premature commitments are precisely what iterative design and evaluation are aimed at circumventing. However, we noted the tension: first, to get useful feedback from users, it is necessary to create an experience that is "realistic" or provocative of the envisioned scenario of use to enable users to give meaningful and actionable feedback, but, second, to make the system realistic enough requires commitment to engineering resources, development costs and design commitments, that, due to limited resources, end up being reified into the system design whether or not they are in fact the most efficient, effective or engaging design options.

Building on this prior work, we evaluated the use of in-context evaluation prototype probes that ranged from low to high fidelity. We presented experiments to assess the prototypes and their potential for revealing design insights at different phases in the design cycle. We highlighted the benefits of different prototyping approaches and discussed the trade-offs in terms of effort and time costs for each of these approaches at various stages of design.

Overall, our results indicate that low-fi prototypes are of little value when used to validate or probe MAR concepts—they do not provide the necessary affordances nor the interactivity required to gather valuable feedback from participants, especially those who are not familiar with AR. At the other end of the scale, high-fidelity prototypes which we expected to yield the best results, were surprisingly ineffective. They provided a relatively realistic experience for users, but raised expectations that led to disappointment and focused negative critique with little creative engagement in dialogue about opportunities for improvement. In the final analysis, the video prototypes proved to be the best option for rapid prototyping. They led to engaged user participation, actionable feedback and creative insights for effective MAR design for location-based social networking. Although interactivity was limited and location/setting/scenario requirements were constrained, from a cost–benefit standpoint, these were the most effective prototypes: rapid generation with low-effort development, coupled with sufficient realism to support scenario engagement whilst retaining the feel of a mutable prototype. Combined, these factors offered participants the best experience of the concept under development while giving them the space to offer constructive critique.

References

1. Marco de Sá, Judd Antin, David Shamma, and Elizabeth F. Churchill. 2011. Mobile augmented reality: video prototyping. In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems (CHI EA '11). ACM, New York, NY, USA, 1897–1902.

2. Matt Jones and Gary Marsden. 2006. Mobile Interaction Design. John Wiley & Sons.
3. Scott Weiss. 2002. Handheld Usability. John Wiley & Sons, Inc., New York, NY, USA.
4. Stephen Brewster. 2002. Overcoming the Lack of Screen Space on Mobile Computers. Personal Ubiquitous Computing . 6, 3, pp. 188–205.
5. Johannes Karlsson, Shafiq ur Rehman, and Haibo Li. 2010. Augmented reality to enhance visitors experience in a digital zoo. In Proceedings of MUM '10. ACM, 4 pages.
6. Peter Barrie, Andreas Komninos, and Oleksii Mandrychenko. 2009. A pervasive gesture-driven augmented reality prototype using wireless sensor body area networks. In Proceedings of the 6th International Conference on Mobile Technology, Application \&\#38; Systems (Mobility '09). ACM, New York, NY, USA, Article 61, 4 pages.
7. Susanna Nilsson and Bjorn Johansson. 2007. Fun and usable: augmented reality instructions in a hospital setting. In Proceedings OZCHI'07, ACM, pp. 123–30.
8. Wayne Piekarski, et al. 2004. Designing Backpacks for High Fidelity Mobile Outdoor Augmented Reality. In Proceedings of ISMAR '04. IEEE, Washington, USA, pp. 280–281.
9. Andreas Dunser, Mark Billinghurst, James Wen, Ville Lehtinen, Antti Nurminen (2011) Handheld AR for Outdoor Navigation, Proceedings of 1st Workshop on Mobile AR, MobileHCI'11, Stockholm, Sweden, 2011.
10. Torben Schinke, Niels Henze, and Susanne Boll. 2010. Visualization of off-screen objects in mobile augmented reality. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10). ACM, New York, NY, USA, pp. 313–316.
11. Holden, W. Augmented Reality on the Mobile to Generate $732 million by 2014(2009) Juniper Research, http://juniperresearch.com/viewpressrelease.php?pr=166
12. Patricia P. Wang, Tao Wang, Dayong Ding, Yimin Zhang, Wenyuan Bi, and Yingze Bao. 2009. Mirror world navigation for mobile users based on augmented reality. In Proceedings of the 17th ACM international conference on Multimedia (MM '09). ACM, New York, NY, USA, pp. 1025–1026.
13. Yan Xu, Mirjana Spasojevic, Jiang Gao, and Matthias Jacob. 2008. Designing a vision-based mobile interface for in-store shopping. In Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges (NordiCHI '08). ACM, New York, NY, USA, pp. 393–402.
14. Gun A. Lee, Ungyeon Yang, Yongwan Kim, Dongsik Jo, Ki-Hong Kim, Jae Ha Kim, and Jin Sung Choi. 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 (VRST '09), Steven N. Spencer (Ed.). ACM, New York, NY, USA, pp. 143–146.
15. Yoshio Ishiguro and Jun Rekimoto. 2011. Peripheral vision annotation: noninterference information presentation method for mobile augmented reality. In Proceedings of the 2nd Augmented Human International Conference (AH '11). ACM, New York, NY, USA, Article 8, 5 pages.
16. Ann Morrison, Alessandro Mulloni, Saija Lemmelä, Antti Oulasvirta, Giulio Jacucci, Peter Peltonen, Dieter Schmalstieg, and Holger Regenbrecht. 2011. Mobile Augmented Reality: Collaborative use of mobile augmented reality with paper maps. Comput. Graph. 35, 4 (August 2011), pp. 789–799.
17. Reitmayr, G., Schmalstieg, D., (2003), Location based Applications for Mobile Augmented Reality, In Proceedings. of the Fourth Australasian User Interface Conference (AUIC2003), Adelaide, Australia. Conferences in Research and Practice in Information Technology, Vol. 18, 2003.
18. Marco de Sá and Luís Carriço. 2008. Lessons from early stages design of mobile applications. In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (MobileHCI '08). ACM, New York, NY, USA, pp. 127–136.
19. Selim Balcisoy et al. 2000. A framework for rapid evaluation of prototypes with augmented reality. In Proceedings. of VRST '00. ACM, pp. 61–66.
20. Areti Damala, Pierre Cubaud, Anne Bationo, Pascal Houlier, and Isabelle Marchal. 2008. Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit. In Proceedings of the 3rd international conference

on Digital Interactive Media in Entertainment and Arts (DIMEA '08). ACM, New York, NY, USA, pp. 120–127.

21. Benjamin Avery, Bruce H. Thomas, Wayne Piekarski (2008) User Evaluation of See-Through Vision for Mobile Outdoor Augmented Reality. In Proceedings of IEEE International Symposium on Mixed and Augmented Reality 2008 15–18 September, Cambridge, UK, pp. 69–72.
22. Yilun You, Tat Jun Chin, Joo Hwee Lim, Jean-Pierre Chevallet, Céline Coutrix, and Laurence Nigay. 2008. Deploying and evaluating a mixed reality mobile treasure hunt: Snap2Play. In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (MobileHCI '08). ACM, New York, NY, USA, pp. 335–338.
23. Tim Gleue and Patrick D\&\#228;hne. 2001. Design and implementation of a mobile device for outdoor augmented reality in the archeoguide project. In Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage (VAST '01). ACM, New York, NY, USA, 161–168.
24. Eduardo E. Veas and Ernst Kruijff. 2010. Handheld devices for mobile augmented reality. In Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (MUM '10). ACM, New York, NY, USA, Article 3, 10 pages.
25. Olsson T., Ihamaki, P., Lagerstam, E., Venta-Olkkonen, L., Vaananen-Vainio-Mattila, K. (2009) n: Norros, L. et al. (eds.). ECCE 2009 - European Conference on Cognitive Ergonomics. Designing beyond the Product - Understanding Activity and User Experience in Ubiquitous Environments, 30.9.-2.10.2009, Helsinki, Finland. VTT Symposium 258 pp. 177–184.
26. Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, J., and Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. In Proceedings of the 27th international conference on Human factors in computing systems (CHI '09). ACM, New York, NY, USA, pp. 1889–1898.
27. Florian Heller, Jan Borchers. (2011) Corona: Audio Augmented Reality in Historic Sites, Proceedings of 1st Workshop on Mobile AR, MobileHCI'11, Stockholm, Sweden, 2011.
28. Yan Xu and Blair MacIntyre, (2011) Theory-driven Design for Social Play with Mobile AR Games, Proceedings of 1st Workshop on Mobile AR, MobileHCI'11, 2011.
29. Laurence Nigay, P. Salembier, T. Marchand, Philippe Renevier, and Laurence Pasqualetti. 2002. Mobile and Collaborative Augmented Reality: A Scenario Based Design Approach. In Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction (Mobile HCI '02), Fabio Paterno; (Ed.). Springer-Verlag, London, UK, pp. 241–255.
30. Jaewon Ha, Kyusung Cho, and H. S. Yang. 2010. Scalable recognition and tracking for mobile augmented reality. In Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '10). ACM, New York, NY, USA, pp. 155–160.
31. Selim Balcisoy, Marcelo Kallmann, Pascal Fua, and Daniel Thalmann. 2000. A framework for rapid evaluation of prototypes with augmented reality. In Proceedings of the ACM symposium on Virtual reality software and technology (VRST '00). ACM, New York, NY, USA, pp. 61–66.
32. Joe L. Gabbard and J. E. Swan II. 2008. Usability Engineering for Augmented Reality: Employing User-Based Studies to Inform Design. IEEE Transactions on Visualization and Computer Graphics 14, 3 (May 2008), pp. 513–525.
33. Christian Jacquemin, Wai Kit Chan, and Mathieu Courgeon. Bateau ivre: an artistic markerless outdoor mobile augmented reality installation on a riverboat. In Proceedings of the international conference on Multimedia (MM '10). ACM, New York, NY, USA, pp. 1353–1362.
34. Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. interactions 6, 1 (January 1999), pp. 32–42.
35. Mary Beth Rosson and John M. Carroll. 2001. Usability Engineering: Scenario-Based Development of Human-Computer Interaction. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
36. Anders Henrysson and Mark Ollila. 2004. UMAR: Ubiquitous Mobile Augmented Reality. In Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia (MUM '04). ACM, New York, NY, USA, pp. 41–45.

37. Robert Albrecht, Tapio Lokki, and Lauri Savioja. 2011. A mobile augmented reality audio system with binaural microphones. In Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications (IwS '11). ACM, New York, NY, USA, pp. 7–11.
38. Humphreys, L. (2007). Mobile social networks and social practice: A case study of Dodgeball. Journal of Computer-Mediated Communication, 13(1), article 17.
39. Johannes Karlsson, Shafiq ur Réhman, and Haibo Li. 2010. Augmented reality to enhance visitors experience in a digital zoo. In Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (MUM '10). ACM, New York, NY, USA, Article 7, 4 pages.
40. Johannes Schöning, Michael Rohs, Sven Kratz, Markus Löchtefeld, and Antonio Krüger. 2009. Map torchlight: a mobile augmented reality camera projector unit. In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA '09). ACM, New York, NY, USA, pp. 3841–3846.
41. Barrie, P., Komninos, A., Mandrychenko, O., (2009) A Pervasive Gesture-Driven Augmented Reality Prototype using Wireless Sensor Body Area Networks, In Proceedings of Mobility 2009, September 2–4, Nice, France, ACM Press.
42. Christian Monrad Nielsen, Michael Overgaard, Michael Bach Pedersen, Jan Stage, and Sigge Stenild. 2006. It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles (NordiCHI '06), Anders M\&\#248;rch, Konrad Morgan, Tone Bratteteig, Gautam Ghosh, and Dag Svanaes (Eds.). ACM, New York, NY, USA, pp. 272–280.
43. Y. Rogers et al., "Why it's worth the hassle: The value of in-situ studies when designing ubicomp," in Ubicomp 2007, Innsbruck, Austria: ACM, September 2007.
44. Michael McCurdy et al. 2006. Breaking the fidelity barrier. In Proceedings CHI '06,ACM, pp. 1233–1242.
45. Sami Hulkko, et al. 2004. Mobile probes. In Proceedings of NordiCHI '04. ACM, NY, USA, pp. 43–51.
46. S. O'Brien and F. Mueller. 2006. Holding hands over a distance: technology probes in an intimate, mobile context. In Proceedings. OZCHI '06. ACM, USA, pp. 293–296.
47. Nielsen, J. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. Cost-Justify in Usability. Academic Press Professional, 1994.
48. Lars Erik Holmquist. 2005. Prototyping: generating ideas or cargo cult designs?. interactions 12, 2 (March 2005), pp. 48–54.

Chapter 7
Design Guidelines for Mobile Augmented Reality: User Experience

Subhashini Ganapathy

1 Introduction: Mobile Augmented Reality

Mobile augmented reality (MAR) enabled devices have the capability to present a large amount of information in real time, based on sensors that determine proximity, visual reference, maps, and detailed information on the environment. Location and proximity technologies combined with detailed mapping allow effective navigation. Visual analysis software and growing image databases enable object recognition. Advanced graphics capabilities bring sophisticated presentation of the user interface. These capabilities together allow for real-time melding of the physical and the virtual worlds and can be used for information overlay of the user's environment for various purposes such as entertainment, tourist assistance, navigation assistance, and education [1].

In designing for MAR applications it is very important to understand the context in which the information has to be presented. Past research on information presentation on small form factor computing has highlighted the importance of presenting the right information in the right way to effectively engage the user [2–4]. The screen space that is available on a small form factor is limited, and having augmented information presented as an overlay poses very interesting challenges.

MAR usages involve devices that are able to perceive the context of the user based on the location and other sensor based information. In their paper on "Context-Aware Pervasive Systems: Architectures for a New Breed of Applications", Loke [5],

---

S. Ganapathy (✉)
Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, 207 Russ Engineering Center, 3640 Col. Glen Hwy, Dayton, OH 45431, USA
e-mail: subhashini.ganapathy@wright.edu

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_7,

talk about two approaches to context-aware systems based on where the information resides—self supported context awareness and infrastructure supported context awareness. In the case of self-supported context awareness the device has the ability to identify the context and alter its behavior based on the context. This is enabled through the hardware, software and other sensors present on the device. In the case of infrastructure supported context awareness, the devices acquire context-awareness by using the infrastructure external to the device.

Context aware computing has been applied in different contexts. Some examples include—the TEA project—mobile phone that detects situations of the phone such as "in hand," "on table," "in pocket," and "outdoors" [6]. Hakansson et al. [7] talk about cameras that can sense sound, pollution in the air, temperature and smell, and create visual effects in photographs given its context. Augmented reality has also been discussed in the context of clothing SensVest [8], worn during different sporting activities has sensors to measure physiological data such as movement, energy expenditure, heart rate, and temperature.

When designing for a MAR application it is important to understand the following segments:

- Usage scenario
  - Shopping assistant, tourism/vacation related usage scenarios, social networking related usage scenarios, and gaming
- Interaction modalities
  - Voice/audio, touch-based, tactile feedback
- Device form factor
  - Ultra mobile device, MID, camera, smartphones

The next couple of sections will go into the details of each of these segments. This chapter outlines the design guidelines related to mobile augmented reality based on the context in which the information has to be presented. The first section gives an overview of the different usage areas where mobile augmented reality is most relevant. The second section discusses the different input methods that are used for accessing the information. The third section talks about the different form factors that are in the space of mobile computing that needs to be understood to design effectively. The last section summarizes the different design aspects in developing persuasive usage model interactions for a great user experience on mobile devices.

2 Usage Scenarios

In the context of mobile computing, some of the key application areas include—shopping usage scenarios, tourism/vacation related usage scenarios, social networking related usage scenarios, and gaming. Understanding the context or the usage

Mir Creek Vineyards
Merlot Wine 2005
The fine Merlo grape with lots of fruit. It's medium bodied with lingering fruit and a silky tannic finish. It's the peoples choice.

Fig. 7.1 MAR shopping assistant usage scenario. Bottle selected to capture in an image. Wine information appears

scenario is very important when designing MAR applications as that provides the right user experience. By detailing the usage scenario the user requirements can be gathered which helps set the design of the product. The usage scenario and user requirements can be developed by understanding the users' likes, dislikes, perceptions, desirability, and expectations and their interaction with the product(s). The following sub sections discuss the different usage scenarios.

2.1 Shopping Usage Scenario

Shopping usage scenario typically involves a certain object of interest. The type of augmented information can be based on a particular object. It can be product information, reviews, or pricing comparisons based on the photo of an item or barcode. For example, as shown in Fig. 7.1, by taking a picture of a wine bottle, the MAR application on the phone would match the image of the bottle with a database listing and present related information, such as competitive pricing and reviews. The application can also act like a shopping assistant by developing a profile of your buying patterns, likes/dislikes, style, and guide you to selections that might appeal to you or match items you have or remind you if already have something similar. It can act like a virtual dressing room where you can superimpose items over an image of yourself to virtually try on items scanned.

2.2 Navigation Assistant: Tourism Related

Navigation assistant scenario provides a rich context for MAR. There are a variety of information that can be augmented such as restaurants, buildings, points of interest, attraction spots, public transport stops, and so on. It is also a very practical implementation that is now more commonly available on phones such as G1 Android and iPhone® [9]. Wikitude® is an AR application that is available on mobile phones that allows the user to view the environment with overlay information on points of interest and a brief description to it. Augmented GeoTravel application for the iPhone uses sensors to show the Augmented Reality view of a point of interest. The Layar Reality Browser is another application that shows the user augmented information of the environment.

Device can retrieve publicly available data and develop personal tour based on user interests/preferences. The user can actively request access to location relevant shared content from others such as text, photos, audio and videos regarding their experiences at that location at other points in time. The device can augment live video feed as seen through the device with overlays identifying landmarks, waypoints, etc. Some of the other compelling usages include:

- *Personal translator*: take snapshot of written information and receive real-time transcription. Speak into device (or type/select phrases) and receive real-time translation.
- *Virtual guide*: tour of an area pointing out things of interest and providing additional reference information. Can be visual/auditory.
- *Voyeur vacation*: view photos/videos, listen to voice commentary, or read blogs from others sharing their experiences at your location at other points in time. Video overlay of historical re-enactments (battles or see what people/places looked at a point in time).

2.3 Gaming

Gaming is another rich context in which augmented reality has been explored a lot. Specifically in the case of mobile augmented reality there are opportunities to create games that let you do a visual recognition of the landscape and place avatars in the space. It can be used for position tracking to hunt virtual treasure. The augmented information can be a display of other players' avatars or simulated scenes. As show in Fig. 7.2, it can be a simulated image of a treasure box when the person holds the mobile device in front of the scene as they are searching for treasure. Figure 7.3 shows an avatar of a character when a person holds a camera in front.

2.4 Social Networking Usage Scenarios

In the case of social media and networking scenario the user can provide information on friends with similar preference based on dynamic extraction of social

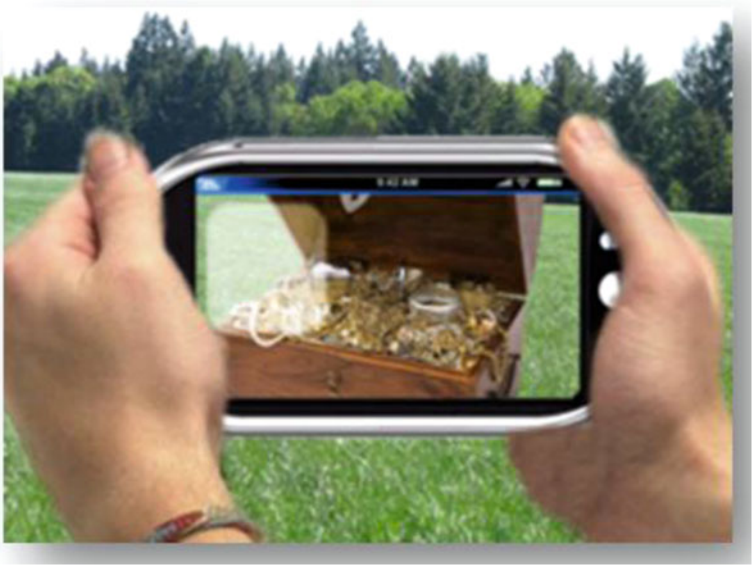


Fig. 7.2 Gaming augmented information

Fig. 7.3 MAR gaming: avatar

network. Based on your location, the device can automatically alert other users information (movies, restaurant). It can also provide real-time media recommendations and sharing based on user preference. The device can act like “People Minder” to provide reminders or details about an individual of interest (basic info, or tap into user database like last meeting, interests, etc.). The user can manage their location broadcast settings (on/off, who can see). User can enter additional information in individual’s profile (type, scan biz cards, hit database, etc.). User can also ping others with information about their location/activities/plans for opportunistic socialization.

The next section gives an overview of the different input modalities that are important to consider in designing a mobile augmented reality system or application.

3 Input Modalities

With the human–computer interaction model moving from traditional input methods to more natural, ubiquitous input techniques there is a need for us to understand and increase the richness of the user experience with higher integration and functionality of these types of technologies with augmented information. There have been many recent developments in the field of multi-modal user interaction. Reactable™ is a multi-touch input based electronic music instrument that allows performers to simultaneously share complete control of the instruments by moving and rotating physical objects on a table surface. Microsoft’s multi-touch, Surface™, allows users to manipulate digital content by the use of natural motions such as hand gestures. There has also been a significant improvement in speech recognition algorithms and software sophistication combined with computation power that allows natural language inputs to be useable.

Each of the input modalities will enable different and important experiences. Voice input allows easy navigation and searching without keypad. In the case of dirty hands or if your hands are occupied and you want to search of a specific information it is useful to use voice input. In that case the system should be able to identify the context in which the user is requesting information and present the right information. Voice can also be used to add precision to the manipulation of information.

Figure 7.4 shows an example of multi-modal input interaction on a mobile device. In the case of gesture input you can move your hand in a specific way to send command to your device. Gesture interaction can be very useful for browsing, scrolling and zooming. It is easy to use it for flipping through a lot of information and for 3D manipulation. For example, the current smartphone implement applications that can be easy moved through hand gestures instead of physically touching the device. You can also use a hand-held device, equipped with accelerometers and positional sensing, to wave in the air and create natural gestures that can interact with your laptop or desktop PC. Both voice and gesture are subject to involuntary/accidental input.

Fig. 7.4 Multi modal input output interaction model

When designing for augmented reality it is important to understand the strengths and challenges of these different modalities and the human interaction through these modalities.

4 Form Factor

In developing applications for new device classes such as tablets, it is important to pay early attention to the physical characteristics of the device. These include: ergonomics (feel, grip, balance and weight, hold-position for different hand-sizes); surface texture and screen vs. surround proportions. These design features quickly constrain other aspects of the solution, including: electronics, technology components feasibility, and so on. Figure 7.5 shows an example of UI on different screen size. As seen in Fig. 7.5 the UI designed shows better on a larger screen vs. a small screen size. The UI doesn't allow the user to see details of the different building in a small screen size. Hence it is important to make sure that the information that is presented is scalable across different form factors.

Also based on the orientation of the device the augmented information needs to be adjusted whether it is in the landscape mode or portrait mode. When a person is holding the device in the portrait mode for a mobile like that is of ~5″ screen size there is not a lot of information you can show on the image. If it is in the horizontal orientation then more amount of augmented information can be presented. Figure 7.6 shows an example of different screen size for mobile form factor.

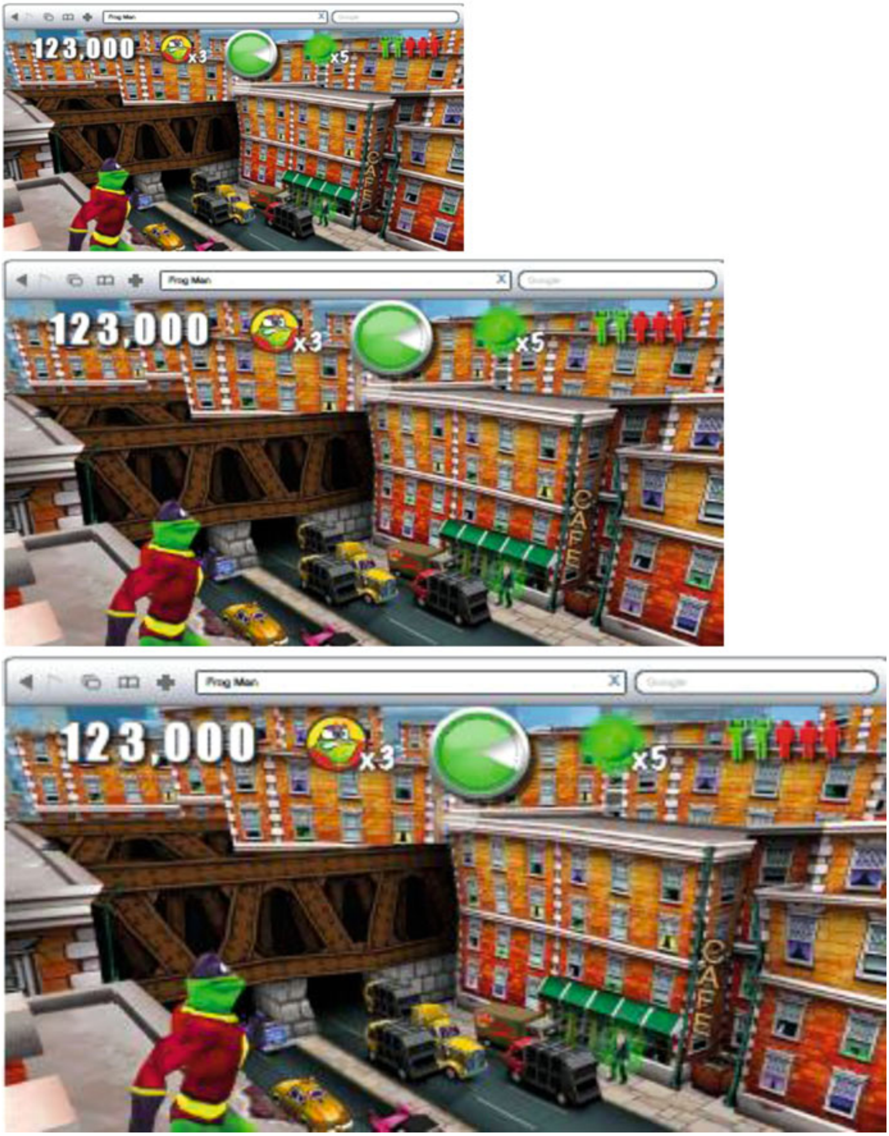


Fig. 7.5 User interface on different screen sizes

Baus et al. [10] discusses the restriction that the user has in terms of the time available to look at the augmented display information and the amount the time it takes for the user to understand and place that information in the spatial context of their environment. There can also be other parallel tasks that the user performs that can affect the viewing time. Figure 7.7, shows an example of different density of

Fig. 7.6 Example of different form factors

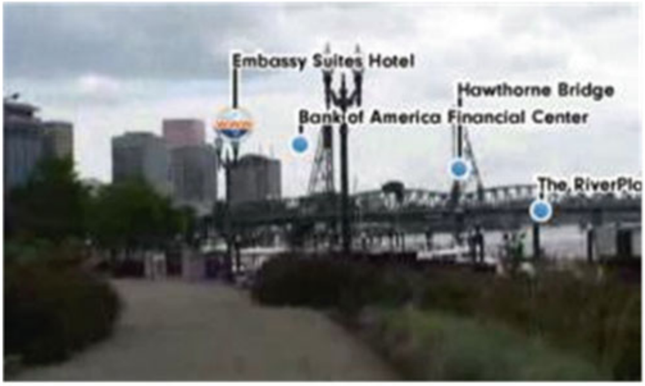
information that is presented to the user. Based on the context and the device the user may prefer to see a low level of density of information to a high density level of information. If it is a larger screen like 9.5″ then it will be easy to read the information with high density.

This section illustrates how this new classification can be used to classify existing applications. It also shows how it could guide a developer to place his application in the proper level in order to find out the resources needed as well as the density of competitors existing in that level.

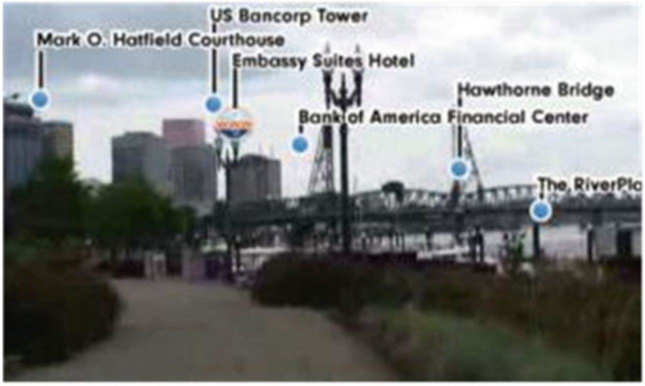
5 Stages of MAR Evolution

Current methods of augmented information presentation on mobile devices are very crowded. This can increase the cognitive overload and frustration for the user. As shown in Fig. 7.8, there is no clear textual information, representation of image overlay. The dials can be confusing.

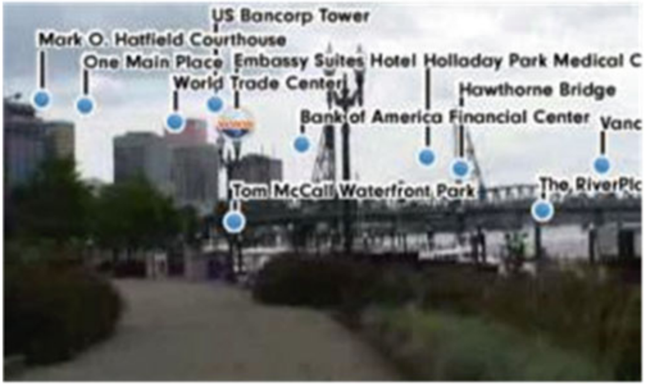
Hence there is a need to systematically design the MAR application to be most effective. The context or usage scenario in which the application is used must be understood that leads to the usage requirements which can then translate to the design requirements of the application. MAR usages can be mapped against the



Low (~3-4 elements)



Medium (~6-7 elements)



High (~10-11 elements)

Fig. 7.7 Three levels of density

Fig. 7.8 Example of MAR system

different stages of MAR evolution from sensory data to visual search to data overlay and then to video overlay. As shown in Table 7.1, the different usages discussed in the previous section can be broken down into the different scenarios in order to understand the requirements.

Based on these usage scenarios it is important to create usage requirements. Below is an example UI for a MAR usage (shopping scenario), shown in Fig. 7.9, in order to demonstrate best practices and ideate on the potential interactions and requirements.

Example usage scenario: the device detects that the user is in a grocery store and begins to prefetch grocery related information for a potential search. The user pulls out the handheld device and select the application. The screen comes up with a viewfinder window for the device phone enabled and ready and providing video feed of whatever device is pointed at. The user points the device down toward the Eggplant on the refrigerated rack and the device snaps a photo. The user can then tap the highlighted region of the Eggplant and the device prompts for “limit search to grocery items?” The device prompts the user if he or she would like to add the ingredients to a shopping list and they can tap the case to select the Yes (default option), As the user picks up items she takes a photo of the barcode and the device checks it off her list and maintains a running total for the cost of the items that she is purchasing.

Table 7.1 Usage vs. technology matrix

Usage areas	Sensory data	Visual search	Data overlay	Video overlay
Tourism/travel	ID places in scene, (low data density) Use GPS directions, and orientation	Search based on recognition Geo tagging	ID places in scene (high data density) Self-guided tours identification of written language	Past/future representation of scene—historical recreations
Shopping assistant	Provide directions within store	Search based on recognition Use of profile	Identification of information on items in scene Comparative pricing	Live virtual dressing room Incorporating multiple items/accessories/cosmetics
Social network/people finder	Location reporting of self/others ID places in scene Data push vs. pull	Recognition of faces (poor)	People recognition Social tagging	Social tagging Virtual meetings/collaboration
Gaming/entertainment	Integration with sensors for I/O Peripheral detection Translation of angular positioning/tilt/acceleration/motion	Role playing games, social interaction games	Treasure hunt, social interaction games	Integration of virtual game items into real scene Natural tracking algorithms

Fig. 7.9 MAR usage scenario—shopping assistant

6 MAR User Interface Guidelines

This section details the design guidelines when developing the user interface for mobile augmented reality. It is critical that MAR enabled device provides feedback to the user when performing the appropriate mode of activity to ensure that the device is in the right state to perform the task. It is also essential that when multiple objects are being displayed on the MAR enabled device, the user has a way of selecting a subset of the objects, or a subset of the screen for the MAR enabled device to process. Figure 7.10, shows example of using the design guidelines. The design guidelines to be considered are

6.1 Clear Textual Information

You need to make sure that the textual information presented is clear. The type of font used should be such that it is easy to read.

6.2 Contrast

Text/background visibility—visual overlay should be such that there is sufficient contrast between the overlay text and background. The text that is presented should

Fig. 7.10 Information presentation using the design guidelines

be visible across different background. It is important to understand the type of background before determining the text color. In certain cases it is useful to use a textbox to define the background for the text. The disadvantage of that is the background will be hidden.

6.3 Grouping

Organization of information is very important when presenting overlay of information.

6.4 Placement

It is important that the information that is presented should not obscure the item of interest.

6.5 Alert/Attention Sensitivity

Specifically calling out information that may need action. Especially in the case of training or medical overlay the information presented should be able to identify the areas which are important and need immediate attention.

6.6 Interaction Methods

Based on the context, the user should be able to switch to different interaction methods. The user may want to pause the information on the screen so that they can hold the smartphone more comfortably or may want to share the information with their friends/family. They could then toggle different labelling options to appear on the screen. For example, they could show all restaurants, and then switch to a type of restaurant, then switch to restaurants within 5 miles.

6.7 Distinct Icons

Augmented information, using icons, must make sure that the items are labeled with varied icons for easy (arsing of information). For example, all social media updates can be of a particular icon, listing of restaurants can be of a different icon, and so on. The user should be able to identify the items' category at a glance without reading the text label.

6.8 Visibility and Distance

In augmenting information, especially in an outdoor scenario, it is important to provide filter based on distance and visibility [11]. It would be helpful if the icon can indicate whether the labelled item is visible.

7 Conclusions

In this book chapter we have focused on addressing the factors that influence the design of mobile augmented reality applications. Future research can focus on integrating the findings from this research as a model-based decision support system that can be integrated into the computing algorithm to achieve high joint performance in the context modeled. The design variables can be dynamic and can support the decision support system for different usage scenarios. For example, a rule-based model can be developed to effectively present information on objects visually occluded without overloading the computing power based on user selection of how far to look for objects, placement of the information (on top of the screen), and so on.

Acknowledgements Special thanks to Delbert Marsh, Glen J. Anderson, Ashley McCorckle, and Igor V. Kozintsev who worked on the project.

References

1. S. Ganapathy, G.J. Anderson, I.V. Kozintsev. MAR Shopping Assistant Usage: Delay, Error, and Utility. In *Virtual Reality Conference (VR),* pp. 207–208, 2011.
2. B. Karstens, R. Rosenbaum, and H. Schumann. Information presentation on mobile handhelds. 1997.
3. D.A. Norman. Cognitive engineering. In *User centered system design: New perspectives on human computer interaction*, pp. 31–61, 1986.
4. A.W. Stedmon, R.S. Kalawsky, K. Hill and C.A. Cook. Old Theories, New Technologies: Cumulative Clutter Effects Using Augmented Reality. In *IEEE International Conference on Information Visualization,* 1999.
5. S. Loke. *Context-Aware Pervasive Systems: Architectures for a New Breed of Applications*. 2007.
6. H.W. Gellersen, A. Schmidt, and M. Beigl. Multi-Sensor Context-Awareness in Mobile Devices and Smart Artifacts. In *Mobile Networks and Applications (MONET),* 2002.
7. M. Hakansson, S. Ljungblad, and L.E. Holmquist. Capturing the invisible: Designing Context Aware Photography. In *Proceedings of DUX - Designing for User Experience*, 2003.
8. J.F. Knight, A. Schwirtz, F. Psomadelis, C. Baber, H.W. Bristow, and T.N. Arvanitis. The Design of the SensVest, In *Personal and Ubiquitous Computing*, volume 9, no 6, 2005.
9. Augmented Reality [Online]. Available: http://en.wikipedia.org/wiki/Augmented_reality. [Accessed on 05/07/2011].
10. J. Baus, A. Kruger, and C. Stahl. Resource-Adaptive Personal Navigation In *Multimodal Intelligent Information Presentation*, pp. 71–93, 2005.
11. T. H¨ollerer, S. Feiner, T. Terauchi, G. Rashid and D. Hallaway. Exploring MARS: Developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers and Graphics*, volume 23, no. 6, pp. 779–785, 1999.

Chapter 8
Usability Recommendations for Mixed Interactive Systems: Extraction and Integration in a Design Process

Emmanuel Dubois, Dominique L. Scapin, Syrine Charfi, and Christophe Bortolaso

1 Introduction

In a recent past, the multiplicity of sensors and effectors, the computing systems size decrease, the computing capabilities increase, and the expansion of communication supports have lead to the emerging of new forms of interactions that better take advantage of the physical environment and artifacts surrounding the user. Augmented and mixed reality, tangible user interfaces, ubiquitous systems, etc. are different forms of such advanced interactive systems: we hereafter use the term mixed interactive systems (MIS) to denote all such forms of interactive techniques. Initially targeting specific, very demanding and highly constrained application domains, these interaction forms are now used in arts, knowledge transfer, education, communication, marketing, etc. [35]. They also support interaction in new spaces such as public spaces, museum, theatre, concerts, etc.: the traditional desktop context is no longer the only available interactive space. Similarly new usages

---

E. Dubois (✉)
University of Toulouse, IRIT – Elipse, 118 route de Narbonne, 31062 Toulouse Cedex 9, France
e-mail: Emmanuel.Dubois@irit.fr

D.L. Scapin
National Research Institute in Computer Science and Control (INRIA),
Domaine de Voluceau – Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France
e-mail: dominique.scapin@inria.fr

S. Charfi
Ecole Nationale d’Ingénieurs de Tunis, BP 37, Le Belvedere 1002, Tunis, Tunisia
e-mail: syrine.charfi@hotmail.fr

C. Bortolaso
University of Toulouse, IRIT – Elipse, 118 route de Narbonne, 31062
Toulouse Cedex 9, France
e-mail: Christophe.Bortolaso@irit.fr

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_8,

are offered: performance and accuracy are not the sole criteria for assessing the quality of MIS, satisfaction and comfort also play a very important part. As a result, designing MIS can easily turn out to be a real challenge. Indeed, Shaer and Jacob [36] underline that "designing and building a TUI requires cross-disciplinary knowledge" and argue that due to the intrinsically complex nature of these systems, "designers and developers face too many conceptual, methodological and technical difficulties". To face those difficulties, dedicated models and frameworks have been developed to describe relevant facets of MIS and support their design at different abstraction levels [8, 12, 20]. Chapter 6 of this book also reflects on different forms of prototyping, especially in the field of mobile augmented reality systems. However, developing interactive systems necessarily include usability evaluations, for which these new interactive situations also raise new questions. Unfortunately, for now only few considerations have been paid to the evaluation of MIS [14]. Indeed due to the complex nature of MIS, and especially the amount of entities involved, their variety, their interconnection, their place in the physical environment, etc. the evaluation of mixed interactive systems turns out to be a very complex issue. Obstacles to the evaluation of mixed interactive systems have been identified in [3], more recently complemented with a list of open issues and advances about the overall user evaluation methods [34] and more specific issues such as mobile interactions [28, 37]. Chapter 7 of this book also proposes high level design guidelines dedicated to mobile augmented reality applications. Nevertheless, no established and validated support for guiding the evaluation of MIS is available yet.

In this context, the goal of this chapter is to improve the design of MIS from a user-centered perspective, in order to increase the attention paid to the evaluation and more generally to the usability of MIS. Indeed, an "ill-designed" MIS can imply physiological problems like dizziness or stiffness due to physical objects weight or physical movements specific to an interaction technique; a high cognitive workload can result from an "ill-designed" MIS because users are not used to associate everyday objects with system data or functions; etc.

Improving the design from a user-centered perspective first requires understanding and then systematically collecting usability issues and results explored in the literature. Then, a structured representation of this usability knowledge is required in order to foster the use of these usability results during the development of a MIS. Therefore this chapter synthesizes in Sect. 2, evaluation methods used in user's experiments involving a MIS and the recurrent criteria addressed in such experiments. Section 3 then presents a template for expressing the results presented in the literature and thus elaborating a repository of unified usability recommendations. To facilitate the use and retrieval of these recommendations, a classification scheme is proposed. Section 4 further refines the repository of usability recommendations through the identification of bindings with the development process: this section anchors the usability recommendations to the design process of a MIS and in the design resources used all along the development process. Finally Sect. 5 briefly introduces a first version of an interactive tool supporting the use of the recommendations and the associated classifications.

2 Related Work

This first section summarizes the evaluation approaches used in the context of mixed interactive systems (MIS). We first describe the main evaluation methods for assessing the usability of MIS. We then extract from these numerous attempts the most relevant and recurrent criteria considered when evaluating a MIS. This is based on an extensive survey of 185 studies published in the literature on MIS evaluation. These studies have been extracted from scientific portals (ACM, IEEE and SpringerLink), as suggested in Dünser [11]. Additional searches have been performed using web engines such as Hcibib, Google and GoogleScholar.

2.1 Usability Methods Used for the Design and Evaluation of MIS

So far, there is no usability method or standards dedicated to MIS. The most approaching standard is ISO 14915-1 standard [18]: it "establishes design principles for multimedia user interfaces and provides a framework for handling the different considerations involved in their design". However it also specifies that this standard applies to professional contexts and does not target other activities such as entertainment and others mentioned before in which MIS are emerging. All the experiments reported in the literature are based on approaches initially dedicated to interactive software in general. Different assessment techniques exist and can be classified in three main categories of evaluation methods [33, 34]: collecting and modeling methods, usability evaluation methods and creativity methods. We use this distinction in three classes to present the methods used in the literature for assessing MIS.

The contribution of collecting and modeling methods to the development of interactive systems is based on acquired data related to users, tasks and context of use. Collecting methods are used to extract design requirements and to support inspection of user's activity in situations of use. Questionnaires, interviews, think-aloud technique, users observations and critical incident technique are examples of such collecting methods. For developing MIS, such methods have mainly been used to support evaluation rather than the design process itself. Modeling methods are used to provide the designer with descriptions of the system according to different points of view on the system. It includes ergonomic data relevant to describe the user's activity. Task, domain, and user's models are example of such methods. For developing MIS, to our knowledge very few evaluation work rely on such approaches.

The role of usability evaluation methods (UEM) is to detect usability issues, i.e. the ability of the system to support the user in reaching his goal and its ease of use. A first set of such methods is based on pre-existing knowledge or models. These methods enable usability inspections without any user's involvement. Such methods may rely on experts, documents such as ergonomic recommendations, models and computer assisted evaluation tools [7]. Given the youth of MIS domain, very little experts can be identified and repository of usability recommendations do not yet exist: therefore such method are not often used in the context of MIS; some examples are limited to the use of restricted ergonomic properties identified as relevant for MIS, such as the continuity property [9]. Alternatively, empirical methods constitute the second set of UEM. They are based on data collection, either quantitative to assess measurements such as performance, or qualitative, to assess subjective aspects in the system. Measures include times to perform the activity, error rate, amount of actions performed, etc. As opposed to the first set of UEM, users are actively involved in such approaches; problems raised during a true activity are therefore more prone to be efficiently detected in such contexts. This form of evaluation method is currently the most used for assessing MIS. User-tests represent 94% of the evaluation studies retrieved from our analysis of the literature. In these evaluations, quantitative and qualitative data are often combined.

Finally, creativity methods allow a form of participatory design and may involve different profile of users. They particularly contribute to develop the communication between stakeholders and they are particularly well suited to new and emerging forms of interaction. Examples include focus groups, brainstorming and divergent thinking. However, they are still rarely used during the design of MIS. Attempts have been reported in [26] for the design of an AR based application for teaching activities. More recently [5] introduced a structured approach for leading such design approach in MIS; it is however not clearly related yet to evaluation steps of the development process.

Despite the variety of available usability methods relevant to the design and evaluation of MIS, user-test is outstandingly the most used technique. In fact, one benefit of this approach is its flexibility: hypotheses, i.e. the focus of the analysis, and dependant variables, i.e. the selected metrics, can be adapted to each study. The following section synthesizes usability aspects considered in the studies we analyzed in our literature survey.

2.2 Usability Aspects Considered in MIS Evaluation

Analyzing published work related to the evaluation of MIS led us to highlight relevant MIS characteristics considered in such context, for the purpose of usability knowledge structuring. Six major topics of interest were identified:

Interaction forms. Evaluations focus on interaction devices (e.g. HMD [30]) display format (e.g. perspective view [39]) or interaction languages (e.g. gesture [19]). Chapter 2 for example synthesizes results related to the use of Audio Augmented Reality. This is particularly relevant for MIS because various techniques can be used for the fusion of physical and digital worlds.

Environment. Evaluations revealed that it is important to take into account some aspects of the environment such as luminosity. This has a major impact on the tracking accuracy [30], on screens or projections used. For instance, pucks on which digital information is displayed can be tracked [21]. Environment in this case is important to consider in order to avoid an occlusion effect.

Quality of display devices. Evaluations revealed that viewing areas remain limited and that the distinction between objects is not easy. Furthermore it requires more effort from users. Chapter 3 of the present book specifically focuses on issues related to Head Worn Augmented Reality displays. This aspect is very important in MIS because most of the feedback is displayed; in addition, different locations may require user's attention: viewing is thus distributed over different interaction spaces and must be carefully considered.

Physiology. Evaluations revealed problems related to stiffness or dizziness due to portable devices weight or to the need to keep a stable motion because of the low accuracy technique. Physical ergonomics provides recommendations in this area [17].

Influence of technology on social interaction. Evaluations revealed that technology can facilitate communication, but it can also disrupt interaction because of cumbersome equipment.

Cognitive constraints. Evaluations revealed that some MIS require a significant mental effort to understand the system [10] or cause cognitive and perceptual discontinuity due to the metaphors or information localization [29]. Chapter 5 further discusses cognitive issues in the field of Mobile Augmented Reality.

2.3 Outcomes

In this survey, many MIS are developed, in different area and thanks to a wide variety of technologies. However, only a limited part of these studies also pay attention to usability aspects for design and/or evaluation. Furthermore, when performed, such usability studies cover very different aspects: qualitative and quantitative aspects, cognitive and physical consideration, technical support and environmental settings, etc.

The youth of the domain is definitely the main cause of such a lack of interest in this area. As a result, capitalized and established results are not yet available to support document based inspection, expertise, or to provide guidance during design process, unlike other domains such as the web [24, 25], graphical environments [38] and virtual environments [1, 13, 23], where such support exists. Specificities of MIS require a dedicated set of recommendations and criteria: some aspects of the usability studies, methods and criteria extracted here should be considered as a solid starting point for structuring a reusable set of usability results and recommendations, relevant for the development of MIS.

And worse, methods reported to collect these results are not always the same, and not even always comparable. As a result, outcomes of these usability studies are not following the same syntax and not expressing the same elements. It is not even expressed at the same level of precision. There is thus a need for a structured organization of usability knowledge, in order to tend towards a uniform expression of usability considerations for the development of MIS.

Based on this analysis, we first propose a structure for expressing (and therefore capitalizing) usability knowledge acquired along MIS design and evaluation (Sect. 3) and, in order to support the use of this usability knowledge, we also introduce anchoring mechanisms of this repository into steps of the development process of MIS (Sect. 4).

3 Extracting and Formatting ER

In order to build a set of usability recommendations dedicated to MIS, we extracted recommendations on the basis of evaluation results of MIS reported in the literature. In addition we identified a set of existing recommendations, proposed in a different context (virtual environments) but relevant or adaptable to MIS. Finally we classified this set of recommendations in order to facilitate search and retrieval of usability results.

3.1 Synthesis of Usability Recommendations for MIS

3.1.1 Selection of User's Evaluation Results

As mentioned above, the expansion of MIS lead to a multiplicity of implementations, technological proofs of feasibility and user tests, in order to assess different parameters. From this plethoric set of experimental results, we focused on studies dealing with specific aspects of MIS that may potentially guide design choices. In addition, only results collected from the use of a valid experimental protocol have been considered. An evaluation procedure has been considered as valid if and only if the following elements were explicitly expressed: method used (e.g. interviews, user test), participants details (e.g. level of expertise, number), steps of the protocol (e.g. task to perform, timing), measuring conditions (e.g. tools, metrics, log files, questionnaires), clarity and validity of the results (e.g. units, significance, existing control group). This set of characteristics of the evaluation procedure is required to ensure a reproducibility of the experiments and therefore to ensure the validity of the results. Any studies that do not fully comply with these requirements have been excluded from our literature review. 185 publications have been considered and only 20 were fully complying with these constraints (list of selected works available on the RESIM Web site [32]), thus demonstrating the lack of maturity in terms of usability knowledge established for MIS.

For example, an experiment has been conducted to compare the use of different interaction techniques for selecting targets in 3D [27]. Participants are wearing a head-mounted display and manipulate a wii-mote to successively control the position of a virtual hand, a ray casting and a magnifying glass. The evaluation procedure is described in a form which is conforming to our specifications. Extracted result of this experiment establish that in such context, users prefer and perform better with the magnifying glass metaphor that with the ray-casting and virtual hand.

The set of selected user's evaluation results has then been complemented with existing usability recommendations proposed in other domains.

3.1.2 Extracting Existing Usability Recommendations

Design and usability recommendations have been proposed for AR games, tangible user interfaces (TUI) and virtual environments (VE).

AR games recommendations [31] concern both design and evaluation. Design recommendations specifically deal with elements of the design process and games design. In the first set, recommendations are general. However, some of the design recommendations match points of interest for MIS, such as the technological requirements (quality of screen and resolution) and physical or social constraints. Other recommendations concern time modeling, collaborative interfaces, tracking and graphical issues. Finally, evaluation recommendations advocate the consideration of collaboration and environment.

In the field of TUI, existing design principles provide guidelines about objects' affordance and forms, the ability to support parallel activity, physically based interaction techniques and multi-handed interaction [22]. Moreover, according to users observations and testing results, a second set of TUI design recommendations [19], particularly on multimodal systems, was extracted. It focuses matching modalities and interaction forms, feedback, the amount of commands to learn and the interaction context.

In addition, a set of 170 recommendations were extracted from previous work dedicated to virtual environments [2]. These recommendations focus on elements influencing the user's interaction with a 3D virtual environment. A first available classification of these recommendations is structured according to a pre-defined set of elements involved in the interaction with 3D virtual environments: user profile, represented objects, actions, spatial organization, decor, frontier, autonomous elements and behaviors. A second form of classifications has been elaborated to fit with ergonomic criteria for graphical user interface [4]. For example, it is expressed that "when a pointing device is used, it is required to also provide information about the actions that can be applied once the selection is operated". Out of these recommendations, some are generic to any kind of interactive systems: they were thus not interesting for our study. Others are recommendations expressed for 3D virtual environments and establishing results already mentioned in one or several selected users' evaluation results (see Sect. 3.1.1). As a result, only 88 recommendations initially dedicated to 3D virtual environments were selected for their usefulness and applicability for the design and evaluation of MIS.

Out of these selected evaluation results and extracted existing usability recommendations, a total set of 151 usability recommendations has been created: 47 corresponding to selected evaluation results, 104 extracted from repositories of recommendations initially dedicated to other application domains. Due to the diversity of origin of these results, the expression was not at all similar from one to another. Therefore, based on the requirements we enounced for establishing that the procedure is valid, and based on the specificities that make up a MIS, we elaborated a unified template for expressing such recommendations.

3.2 Adopting a Unified Format

The unified format adopted is used to express usability recommendations, valid for use in the context of MIS design or evaluation. It therefore includes elements related to the evaluation procedure and elements related to the characteristic of the mixed situation considered. A recommendation is thus structured as follow:

1. The *context* identifies the application domain of the system (e.g. game, medical application) and the particular type of the system (e.g. memory game, surgical training).
2. The *system category* denotes the specific category of interactive system (e.g. collaborative, mixed, tangible, mobile, etc.).
3. The *situation* identifies a set of elements from which the result has been extracted. It contributes to the presentation of the evaluation procedure:
   (a) The *task type* (e.g. selection).
   (b) The *object of the task*, i.e. the object really affected or causing the task to be performed by the user. For example when switching from one slide to another during the oral presentation of a paper, the object of the task is the file containing the slideshow supporting the talk.
   (c) *Targeted users* and their specificities (e.g.: male/female, driving hand).
   (d) *Environmental settings* (e.g.: luminosity, temperature).
4. The *interaction form* specifies details about the interaction spaces and required entities. It includes:
   (a) The *nature of each object* involved (e.g.: digital, physical) and its characteristics (e.g. location, simultaneous availability).
   (b) The *set of devices* used and their characteristics such as input/output, location, operating area.
   (c) The *representation* adopted for each perceivable data (e.g.: virtual hand, magnifying glasses) and performable actions (e.g. motions, commands).

   (d) The *point of view* adopted for each rendering or action to perform.
   (e) The *language* used to encode an information (e.g. speech, gesture, 3D).
5. The *collected data type* specifies whether the results are based on subjective data (questionnaires, etc.) or objective data (performance results, log files, etc.).
6. The *text* of the recommendation.

On one hand, collected data type, targeted users and task type ensure that the elements making an evaluation procedure valid are well identified (method used, participants details, steps of the protocol, measuring conditions, clarity and validity of the results). On the other hand, the interaction form express specificities of the mixed situation considered (physical and digital entities, devices and information exchanged).

Finally, each usability recommendation is identified via a couple (i,j) where i corresponds to the reference number of the published paper from which the result has been extracted, and j is used to differentiate recommendations extracted from the same paper: (7.1) and (7.3) for instance, are both extracted from [7] in our list of selected papers.

All selected evaluation results and extracted existing recommendations have been reformulated according to this template. The complete list is available on the RESIM web site [32]. Although they all have been judged as valid, they do not all provide every piece of information required in this template: all the parameters are thus not necessarily instantiated. As a result, the more parameters are instantiated, the more the recommendation is specific to an interactive context. For example, let us consider the usability recommendation (5.2.1). The text of this recommendation is "In a game application, and particularly war game, considering users' preference, it is preferable to use a computer or a board game rather than an augmented reality game" [30]. This recommendation describes:

- The *context*: the application domain is game and the particular type is war game.
- The *system category*: computer or board game.
- The *collected data type*: users' preference.

Other attributes have not been defined in this experiment. Therefore this recommendation can be applied to any kind of tasks involved in the context of a war game. However, this may appear to be a wrong recommendation in different context, such as in a museum for example.

Beyond the unified structure offered to express usability recommendations, this template also adds some flexibility in the use of collected usability knowledge. This collection of recommendations can be used in different ways. First, an expert may use them to inspect a MIS. Alternatively a designer can also browse this set of recommendations to establish design decisions. To better support these activities, we further classified this collection of recommendation according to the target of the recommendation, i.e. the element affected by the recommendation.

3.3 Usability Recommendation Classification According to the Target

This classification identifies the attribute on which a recommendation focuses (the target of the recommendation). Seven classes of recommendations thus emerged:

- *Components*. This class is made of three sub classes (physical entities, digital entities and adaptors) and groups recommendations related to the objects involved (part 4.1 of the template). These three classes correspond to concepts frequently expressed in design models for MIS such as TAC [36], MIM [8] or ASUR [15]. These can be considered as intrinsic components of a MIS: each time a designer or a usability expert manipulates a constituent, the associated set of recommendations might appear to help the designer fine tune the constituent or its integration in the system.
- *Representation*. This class concerns general aspects of the representation such as language, point of view, dimension and localization (parts 4.2, 4.3, 4.4 of the template). Such properties are required to characterize the form of a data exchange. In a MIS, different entities are coexisting and exchanging data. This is therefore a second relevant entry point for a designer or usability expert developing or assessing a MIS.
- *Spatial association*. This class relates to recommendations dealing with the spatial association of physical entities (part 4.1 of the template). Such considerations are crucial when designing or evaluating MIS because the interaction space is often divided among different places. Ensuring the adequacy of this split over space is thus useful.
- *Interaction*. This class expresses global recommendations about aspects of the interaction forms (part 4 of the template), such as the use of a specific metaphor. This group can thus provide indications useful to fix early an overall metaphor to apply to the rest of the design.
- *Synchronization*. This class of recommendations affects feedback synchronization (part 3 of the template). It is very similar to recommendations included in "spatial association" but it deals with none topological considerations: latency, coherence among information exchange, etc.
- *Interactive system choice*. This class provides recommendations for choosing the type of interactive system to develop (part 2 of the template). It is especially useful to validate the adequacy of MIS in specified application domains or activities.
- *Task*. This class is made of recommendations that have an impact on the tasks structure sequence (e.g. adding a calibration step).

In the case of the recommendation (5.2.1) considered above, the use of computer or board game is recommended rather than an AR game: the target is related to the "interactive system choice". Overall, the identification of the recommendation target provides an overview, and thus a quick access to the recommendations. However, it is entirely dependent on the element constituting the MIS. A complementary classification approach to better take advantage of these recommendations consists in following the designer's activity and sticking to the different steps of his/her activity. This is the goal of the two alternate classifications we present in the following section.

4 Structuring Usability Recommendations Repository

These additional classifications help towards the definition of a method incorporating usability recommendations for MIS design. The first one offers a view on these recommendations according to the development process, while the second one deals with the concepts of different models for the design of MIS.

4.1 Positioning Usability Recommendations in a Development Process

Given the multidisciplinary teams necessarily involved in the design of a MIS (e.g. physical and digital spaces, application domain, end-users, ergonomists, etc.) and the limited set of on-the-shelf "widgets" for MIS, most of the MIS developments follow an iterative and incremental cycle. Four steps structure this process: analysis, design, implementation and evaluation. In these steps the most relevant and specific design considerations for MIS include requirements, task analysis and interaction design. The other parts of the design cycle are similar to any other one for interactive systems. The recommendations for MIS dealt with here aim at supporting these three steps only: requirement (r), task (t) and interaction (i). Each one of them can make use of one or more design models (Fig. 8.1).

Section 3.3 identified a set of targets of the recommendations. Similarly, each recommendation also include source data i.e. information that constitute the hypotheses of the recommendation. For example one mentions that "when a task of target localization need to optimize the performance, it is better to use a bird-eye view than a perspective view", the *target* of the recommendation is the representation and the *source* of the recommendation is the task to perform (localization task). Source and target can therefore be related to the same step of the development process or to different ones. We therefore structure our recommendation repository in three main classes:

- The class *Connection*. This class aggregates recommendations that contribute to the linking of two design steps: the design of the source and target data are considered in two different steps of the development process. Given that we focus on three main steps of the design process (*r: requirement, t: task, i: interaction*), there are three subclasses (Fig. 8.2, left):
  - *Crt* and *Ctr* group recommendations connecting requirement and task steps.

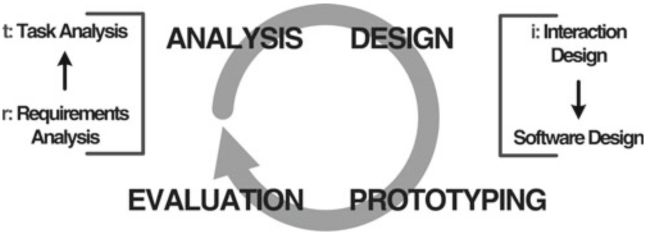


Fig. 8.1 Main design considerations for a MIS, along an iterative development cycle

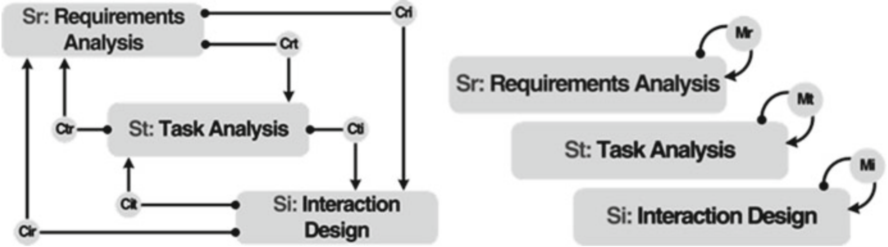


Fig. 8.2 Six subclasses of the "Connection" class and three subclasses of the "Models" class

  - *Cri* and *Cir* group recommendations connecting requirement and interaction steps. The following recommendation is such an example: "when developing museographic application (requirement consideration), it is recommended to use Augmented or Virtual Reality (interaction consideration)".
  - *Cti* and *Cit* group recommendations connecting task and interaction steps.

- The class *Models*. This class is made of recommendations that contribute to the modeling of the system in one step of the development process: source and target data are related to the same development step and lead to the refinement of the model, such as the definition of a new attribute. The three steps of the design process we consider (*r: requirement, t: task, i: interaction*) lead to the identification of three subclasses (Fig. 8.2, right):

  - *Mr* groups recommendations related to the definition of requirements.
  - *Mt* groups recommendations related to the task modeling.
  - *Mi* groups recommendations related to the modeling of the interaction. The following recommendation illustrates this subclass: "The form of the physical entities (element supporting the interaction) must trigger and support spatial manipulations (attribute of interaction elements)".

- The class *Connection and Models*. This class contains recommendations that simultaneously contribute to the linking of two design steps and the refinement of one of the models used in these two steps. Source data of the recommendations are related to different steps of the design process and have an impact on one of these steps (*r: requirement, t: task, i: interaction*). Similarly to the class "Connection", six subclasses have been identified: *CMrt* and *CMtr*, *CMri* and *CMir*, *CMti* and *CMit*.

For example, the recommendation stating that "it is preferable for selectable (task consideration) physical entities (interaction consideration) to be grouped in the same place (interaction consideration)" illustrates the subclass *CMti*: source information are related to task and interaction consideration and the target of the recommendation guide the design of the interaction.

A final distinction can be drawn between each recommendation. Indeed, the classes presented are tightly coupled to one or two steps of the design process. Each step may rely on one or more different design models. Therefore, if the source and target of the recommendations explicitly correspond to concepts described and expressed by the model, the recommendation is a *prescriptive recommendation*. Otherwise it is just a *document aiding the design* of a MIS.

4.2 Anchoring Usability Recommendations Within MIS Design Models

To further exploit such prescriptive recommendations, source and target of the recommendations have been more precisely associated with concepts of existing design models specific to MIS. Based on this characterization, an additional classification of the recommendations repository is available and allows a designer to straightforwardly access usability recommendations related to the concept of the model s/he is considering.

Practically, this approach has been applied to the ASUR model [16], a model dedicated to the description of the user's interaction with a MIS. This model contains six major concepts: user, physical entities and their attributes, digital entities and their attributes, adapters bridging the two worlds, information channels carrying data between two concepts, and grouping mechanisms expressing links between physical and digital concept or properties of part of the system. With this model, each prescriptive recommendation is thus deeply anchored into the ASUR design model. For example, one recommendation states that "When tangible interactions are used on a projection plane, the projecting device must be appropriately positioned to prevent occlusion from the user": the target of the recommendation is the device; it is related to the interaction design step of the development process; more precisely, in terms of ASUR it is constraining the attribute "location" of an adapter, i.e. the attribute expressing where the ASUR entity is positioned.

In addition, a second instance of this approach has been applied to the set of recommendations that have an impact on the link between task design considerations and interaction design consideration, i.e. the class *Cti*. Indeed, translation rules have been built and formalized to transform a task model expressed with KMAD into an initial, and partial, ASUR model describing the user's interaction with the mixed environment [6]. Four sets of rules support the translation of KMAD model into ASUR interaction channel and their attributes, adaptors, users and physical or digital entities. Each translation rule is thus now associated with a set of usability recommendations that guides the application of the rule or complement the result. For example, a recommendation states that “When classification and grouping data, it is preferable to use tangible interaction rather than a WIMP-based user interface”; it guides the designer in the selection of the most appropriate type of entity (physical in this case); the translation rule from KMAD to ASUR stating that object expressed in the KMAD model must be transformed into ASUR entity is thus influenced by this recommendation: the application of the rule combined with the recommendation will result in the generation of ASUR physical entities (and no ASUR digital entities).

In order to fully take advantage of these diverse classifications and select the most appropriate one according to the design activity, static lists are no longer sufficient: a flexible and interactive support is required. We present a first version of such a tool in the following section.

5 A Tool for Exploring a Usability Recommendations Repository

In order to more efficiently support the use of this repository of recommendations and to better take advantage of the classification of recommendations we developed the tool RESIM. For exploring the repository, four different accesses are supported. We describe and illustrate them in the following section before detailing further various technical aspects of RESIM.

5.1 Access Modes and Their Role

A first mode corresponds to simple browsing among the overall set of recommendations: in a text field, the user enters one or several words. These words will be searched over all the recommendations and in every parts of the template (see Sect. 3.2). Each time the target words are found, the recommendation identifier, text and reference are displayed. For example, one designer may want to consult all recommendations dealing with “pointing”: currently RESIM would retrieve four results. This mode is intended to be used in early design situations, in which designers seeks to identify ideas or general orientations for the overall process.

The second mode takes particularly advantage of the recommendations target (see Fig. 8.3). In this mode, the user selects one target among the list of targets mentioned in Sect. 3.3. Then, the associated list of recommendations is dynamically displayed and may be further structured by sub-targets. For example, when

[1] Cheok, Adrian D., Xubo Yang, Zhou Zhi Ying, Mark Billinghurst, et Hirokazu Kato. 2002. Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing. Personal and Ubiquitous Computing 6:430-442.

[2] Mathews, Moffat et coll. 2007. Evaluation of spatial abilities through tabletop AR. Pp. 17-24 dans Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI. Hamilton, New Zealand: ACM.

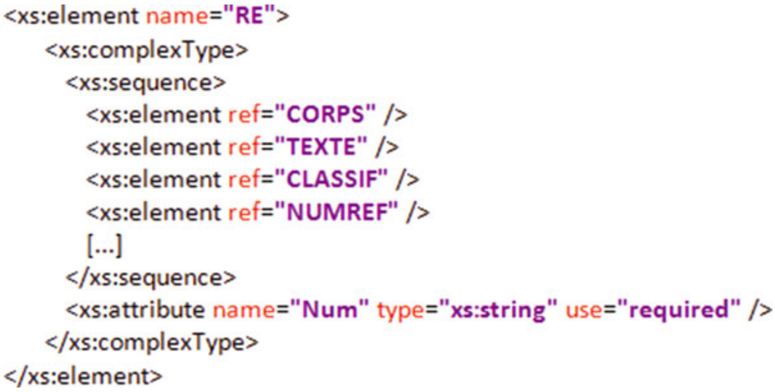
[3] Chastine, Jeffrey W., Kristine Nagel, Ying Zhu, et Luca Yearsovich. 2007. Understanding the

Fig. 8.3 Illustration of the use of RESIM when accessing to the recommendation through the list of targets (second mode). Visualization of the references used to extract the recommendations. Web site: http://irit.fr/recherches/ELIPSE/resim/AccueilEng.html

selecting the “representation” target, a long list of 56 recommendations appears, structured into five sub-sets, respectively, corresponding to different attributes of the representation (language, localization, point of view, etc.). This mode is intended to be used when a designer or a usability expert needs to consider a specific aspect of the MIS: RESIM will provide them with a set of established results that should be considered and may guide the design or even help to diagnose usability problems.

The third mode is very similar to the previous one: instead of supporting an access via the target of the recommendations, it identifies the recommendations related to the different steps of the design process. Selecting one step (articulation, model or both) displays, as previously, the list of recommendations related to this step. The resulting list is potentially refined through the selection of a sub-step of the process. For example, when selecting the step “model”, the user may browse three sub-sets, respectively, dedicated to recommendations related to models for analysis, interaction or task. This mode is intended to be used during the design of a MIS in order to suggest future optimization or options to consider. According to the current design step, RESIM retrieves a list of recommendations that highlight potential complementary issues.

The fourth mode corresponds to an advanced search mode, intended to be used by experts in the design of MIS and of usability recommendations. It combines the use of a text field to enter a set of keywords (as in the first mode) with a form in which the user can select a combination of attributes among the predefined targets (design steps and the design models elements used in MIS). Currently, only two design



```
<xs:element name="RE">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="CORPS" />
        <xs:element ref="TEXTE" />
        <xs:element ref="CLASSIF" />
        <xs:element ref="NUMREF" />
        [...]
      </xs:sequence>
      <xs:attribute name="Num" type="xs:string" use="required" />
    </xs:complexType>
</xs:element>
```

Fig. 8.4 Partial XSD scheme used to encode the recommendation in the XML files

models for MIS have been used to complement the encoding of each recommendation: KMAD and ASUR. As illustrated in the next section, RESIM has been thought to easily include additional keywords matching other classifications schemes.

Finally another mode has been developed to support the easy insertion and encoding of additional usability recommendations. However its access is restricted so far to RESIM authors.

5.2 *Further Technical Details*

RESIM has been developed as a dynamic web page. In order to facilitate its use in different contexts, XML technologies have been used, rather than databases, to store the recommendations and their classification. Indeed, a designer working in his office may have access to internet and so to a distant database server, while in situ inspections will hardly allow being online all the time. In addition, it is really easy to connect XML files to different programming languages. Regarding the rendering of RESIM, it is based on a combination of HTML, CSS and JQuery javascript framework.

Each recommendation is conform to an XSD schema. This schema includes five mains tags (in French) as illustrated in the Fig. 8.4:

- “CORPS” is refined into subtags that corresponds to each category forming the recommendation template introduced in Sect. 3.2.
- “TEXTE” includes the textual version of the recommendation, as displayed by RESIM as a result of a query.
- “CLASSIF” is including a set of subtags for each classification scheme considered in RESIM. For example, a subtag is used to mention the target of the recommendation, a set describe the design steps for which the recommendation

may be relevant, another set of subtags characterizes the recommendation in terms of an interaction model, etc.
- "NUMREF" is the identifier of the reference from which the recommendation has been extracted.
- "NUM" is unique identifier of the recommendation.

6 Conclusion

Developing a repository of usability recommendations is a recurring preoccupation and has already been accomplished in different domains (web, graphical UI, etc.). To fully describe this process, Vanderdonckt [40] identified five phases:

1. Collecting a set of relevant recommendations
2. Expressing in a unique format and then classifying the recommendations
3. Integrating the recommendations in a method of use
4. Providing a tool to exploit the set of recommendation and tend toward an automatic use of it
5. Refining the method of use on the basis of lessons learnt

In this chapter, we covered the four first phases by offering a structured usability piece of knowledge related to mixed interactive systems. Recommendations have been created on the basis of evaluation results and extracted from existing lists of recommendations. A template has been elaborated to express all these recommendations in a unified manner; the set of recommendations has been sorted according to several targets. Each recommendation has also been associated to specific steps of the design process and even more precisely to elements of design models potentially used during the development process: this constitutes the basis of a method for using this set of recommendations. Finally, an interactive tool, RESIM, has been developed as a web application [32] to support the retrieval of recommendations according to a multi-criteria search.

This contribution in the field of usability studies for mixed interactive systems is useful in the context of design, i.e. to guide, suggest or complement design choices, but also as a resource for leading usability inspection. Furthermore, given the links explicitly established between the usability recommendations and design models relevant at different stage of the development process (Sect. 4), usability issues corresponding to one of recommendations is easily connected to the design resource used in the development process: this connection tends to facilitate the identification of the place where design modifications must be adopted.

These recommendations, classifications and tools have already been successfully used to design, implement and deploy MIS in a museographic context [15]. Nevertheless a deeper usability study of the tool itself is required for assessment and potential improvements.

The web application presented here is only a first version of the tool, useful to demonstrate the benefits of such a usability recommendations repository for MIS.

However we strongly believe that it is required to be able to integrate such a support directly in the tools support used to manipulate the different design models involved in the development process and the classifications we presented. In this perspective, we are considering the development of an Eclipse Plugin for the manipulation of this repository and its different classifications. Transformations defined and supported by Model Driven Engineering Approach would thus be available to make concrete the link between the repository and the computer assisted environments for the manipulation of design resources such as ASUR and KMAD. Such a link would therefore be in line with second part of Vanderdonckt's phase 4, automatic use of recommendations: adding an element to a model would automatically raise a window displaying the appropriate recommendations.

References

1. Bach, C., Scapin, D.L.. Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions. In Int. Conf. Interact'03, Zurich, IFIP, pp. 880–883, 2003.
2. Bach, Cedric. 2004. (In French) "Elaboration et validation de Critères Ergonomiques pour les Interactions Homme-Environnements Virtuels.", PhD of the University of Metz, France.
3. Bach, C., Scapin, D.L. (2004) 'Obstacles and Perspectives for evaluating Mixed Reality systems usability', IUI-CADUI MIXER, Island of Madeira.
4. Bastien, J.M.C., Scapin, D.L. 1993. Ergonomic criteria for the evaluation of human-computer interfaces, Technical report, INRIA.
5. Bortolaso, C., Bach, C., Dubois, E. A combination of a Formal Mixed Interaction Model with an Informal Creative Session. ACM SIGCHI conf. EICS 2011, Italy, 2011, pp. 63–72.
6. Charfi, S., Dubois, E., Bastide, R. Articulating Interaction and Task Models for the Design of Advanced Interactive Systems. in TAMODIA'07, Vol. 4849, Springer, LNCS, pp. 70–83, 2007.
7. Cohen, A., Vanderdonckt, J., Crow, D., Dilli, I., Gorny, P., Hoffman, H.-J., Iannella, R., et al. (1995). Tools for working with guidelines. ACM SIGCHI Bulletin, 27(2), pp. 30–32. doi:10.1145/202511.202517
8. Coutrix, C. and Nigay, L. Balancing physical and digital properties in mixed objects. Proc. of AVI'08, ACM (2008), pp. 305–308.
9. Dubois, E., Nigay, L., Troccaz, J., "Consistency in Augmented Reality Systems", EHCI'2001, Canada, 2001, pp. 111–122.
10. Dubois, E., Truillet, P. and Bach. C. 2007. Evaluating Advanced Interaction Techniques for Navigating Google Earth. University of Lancaster, UK.
11. Dünser, A, Grasset, R., Billinghurst, M. 2008. "A survey of evaluation techniques used in augmented reality studies." ACM SIGGRAPH ASIA 2008 courses, pp. 1–27.
12. Fishkin, K.P. A taxonomy for and analysis of tangible interfaces. Personal Ubiquitous Comput. 8, 5 (2004), pp. 347–358.
13. Gabbard, J.L., Hix, D.. 1997. "A taxonomy of usability characteristics in virtual environments." Master thesis, Virginia Tech, Blacksburg.
14. Gabbard, J.L. and J, E.S.I. Usability Engineering for Augmented Reality : Employing User-based Studies to Inform Design. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, (2011).
15. Gauffre, G., Charfi, S., Bortolaso, C., Bach, C., Dubois, E. Developing Mixed Interactive Systems: a Model Based Process for Generating and Managing Design Solutions. in The Engineering of Mixed Reality Systems. Springer-Verlag, 10, pp. 183–208, Vol. 14, HCI Series, 2010.

16. Gauffre, G., Dubois, E. Taking Advantage of Model-Driven Engineering Foundations for Mixed Interaction Design. in Model Driven Development of Advanced User Interfaces. Springer-Verlag, 4.1.3, pp. 219–240, Studies in Computational Intelligence, Vol. 340, 1, 2011.
17. International Standards Organisation ISO 9241–5, Ergonomic requirements for office work with visual display terminals - Part 5: Workstation layout and postural requirements, (1998).
18. International Standards Organisation. ISO 14915, Software ergonomics for multimedia user interfaces -- Part 1: Design principles and framework, (2002).
19. Irawati, S., Green, S., Billinghurst, M., Duenser, A., Ko, H. 2006. An Evaluation of an AR Multimodal Interface Using Speech and Paddle Gestures, ICAT, pp. 272–283.
20. Jacob, R.J., Girouard, A., Hirshfield, L.M., and al. Reality-based interaction: a framework for post-WIMP interfaces. Proc. of CHI'08, ACM (2008), 201–210.
21. Jacob, R.J.K., Ishii, H., Pangaron, G., Patten, J. 2002. A tangible interface for organizing information using a grid, CHI, USA, pp. 339–346.
22. Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K. 2000. Virtual object manipulation on a table-top AR environment, ISAR, pp. 111–119.
23. Kaur, K. 1998. "Designing virtual environments for usability." PhD Thesis, London City University, Londres, UK.
24. Leavitt, M O., Shneiderman, B. 2006. Research-Based Web Design and Usability Guidelines. Washington, DC: U.S. Department of Health and Human Services.
25. Leulier, C, Bastien, C., Scapin, D.L.. 1998. Compilation of ergonomic guidelines for the design and evaluation of Web sites. Rocquencourt, France, (INRIA).
26. Liu, Wei, Adrian David Cheok, Charissa Lim Mei-Ling, et Yin-Leng Theng. 2007. "Mixed reality classroom: learning from entertainment.", DIMEA'07., Australia: ACM, pp. 65–72.
27. Looser, Julian, Mark Billinghurst, Raphaël Grasset, et Andy Cockburn. 2007. "An evaluation of virtual lenses for object selection in augmented reality." Int. conf on Computer graphics and interactive techniques in Australia and Southeast Asia. Australia: ACM, pp. 203–210.
28. Lumsden, J. eds (2009) Handbook of research on User interface design and evaluation for mobile technology. Information Science Reference, Hershey, PA, USA.
29. Mansoux, Benoit, Laurence Nigay, et Jocelyne Troccaz. 2005. "The Mini-Screen: An Innovative Device for Computer Assisted Surgery Systems." pp. 314–320 dans Medicine Meets Virtual Reality 13: The Magical Next Becomes the Medical Now, vol. 111/2005, Studies in Health Technology and Informatics. IOS Press.
30. Nilsen, T. 2005. Tankwar: AR games at GenCon Indy, ACM-ICAT, NZ, pp. 243–244.
31. Nilsen, T. 2006. Guidelines for the Design of Augmented Reality Strategy Games, Master thesis, Canterbury University.
32. RESIM web site: full list of usability recommendations adapted to Mixed Interactive System; http://irit.fr/recherches/ELIPSE/resim/AccueilEng.html, last access November 2011.
33. Scapin, Dominique L. 2006. "Exigences ergonomiques, méthodes, et normes pour la conception de Systèmes d'Information centrés humain." dans Encyclopédie de l'informatique et des systèmes d'information. VUIBERT.
34. Scapin, D.L. and Law, E.L. (2007). Review, Report and Refine Usability Evaluation Methods (R3UEMs), COST 294-MAUSE 3rd International Workshop, Greece.
35. Shaer, O. Tangible User Interfaces: Past, Present, and Future Directions. Foundations and Trends® in Human–Computer Interaction 3, 1–2 (2009), pp. 1–137.
36. Shaer, O. and Jacob, R.J. A specification paradigm for the design and implementation of tangible user interfaces. ACM Trans. Comput.-Hum. Interact. 16, 4 (2009), pp. 1–39.
37. Siek, K.A., Neely, S., Stevenson, G., Kray, C. and Mulder, I. (2009). Advances in evaluating mobile and ubiquitous systems. IJMHCI vol. 1 (2), pp. 5–14.
38. Smith, S.L., Mosier, J.N.. 1986. Guidelines for designing user interface software. The MITRE Corporation Technical Report (ESD-TR-86-278).
39. Tonnis, M., Sandor, C., Langen C., Bubb, H. 2005. Experimental Evaluation of an Augmented Reality Visualization for Directing a Car Driver's Attention, ISMAR, pp. 56–59.
40. Vanderdonckt, Jean. 1999. "Development Milestones Towards a Tool for Working With Guidelines." Interacting with Computers 12:81–118.

Part IV
User Experience

Chapter 9
Concepts and Subjective Measures for Evaluating User Experience of Mobile Augmented Reality Services

Thomas Olsson

1 Background

Mobile AR makes it possible to explore and interact with the world in a totally new way. As discussed also in previous chapters, future use of mobile augmented reality services can be envisioned to cover a wide range of sectors of life, such as tourism, shopping, navigation, social interaction, entertainment, learning and education, and work related activities. By making the world itself a user interface to related digital information provides a highly natural paradigm that might revolutionize the way of accessing and presenting information in the future user interfaces [38, 60].

The AR technology is little by little entering such a level of maturity that true mobile *services* that utilize and demonstrate AR can be developed [25, 59]. The first publicly available and extensively adopted applications utilizing AR, such as Layar [36], Junaio [28] and Google Goggles [19], have gained great interest in early adopters. Despite being only the first generation of consumer-targeted applications that utilize AR, the current applications demonstrate several features related to AR and computer vision: browsing the location-bound content as visually superimposed on the real-world view, identifying physical objects and visual markers to acquire further information, and providing interactive content. Overall, most current applications focus on either accessing digital content with the help of augmenting the mobile device's camera view to the physical world or using the same for gaming and entertainment.

Nevertheless, the survey by Olsson and Salo [43] showed that, despite the high number of downloads, the current publicly available mobile AR applications show rather little practical or pleasure value as they are largely used for their novelty and curiosity values. The current applications were concluded to be rather far away

---

T. Olsson (✉)
Unit of Human-Centered Technology, Tampere University of Technology,
Korkeakoulunkatu 6, 33720 Tampere, Finland
e-mail: thomas.olsson@tut.fi

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_9,

from the full potential of mobile AR as a platform for truly useful and pleasurable services.

Therefore, in this chapter we take a perspective on the *potential* that mobile AR services shows for the future. With an *AR service* we refer to the comprehensive technological entity that the users might perceive as the service: including the technical system and application features and functionalities, the AR information content, as well as the interaction with the required technological artifacts. With this holistic perspective to mobile AR, there is much room to invent and develop services that take the full advantage of AR and are appropriate with regard to their target user group and the purpose and context of use.

Considering both the great potential and risks involved in AR [46], the user's experience of a MAR service can become a complex whole: cognitively and emotionally highly varying situations and contexts of use, a novel paradigm for accessing and interacting with information in situ, objects and locations being sources of information and search keys, the risks related to various sources of information etc. Therefore, we set off with a statement: in order to create successful MAR services for consumers, it is critical to understand what the user experience of MAR can be like and what elements it is based on.

1.1 User Experience and Evaluation

Offering a stimulating and pleasurable *user experience* (UX) is becoming a central goal and design strategy in design of digital artifacts and services (e.g. mobile devices, web sites, digital games, digital content). Their success is agreed to be positively influenced by the extent to which they promote a high-quality experience in users [34]. UX is often described as a personal, subjective phenomenon that is related to the use of technology [35, 48]. Law et al. [35] emphasize that user experience is the individual and personal phenomena (feelings and experiences) that emerge within users when they interact with a product, service or system. The recent ISO standard [26] defines UX comprehensively as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service".

When compared to the theoretical foundation of *usability*, UX moves towards a more emotionally appealing relationship between the user and the product. As consumers' expectations of technological products constantly grow because of improvement and development activities and shorter product lifecycles, a product has to not only provide useful (utilitarian) functionalities and a usable interface but also create satisfying experiences, such as stimulation, beauty and pleasure [27]. Hassenzahl and Ullrich [23] point out that UX focuses on the positive outcomes of interaction. Usability theories focus on pragmatic aspects of product use that furthermore are relatively persistent and at least partly objectively definable; for example task completion efficiency, effectiveness and ease-of-use. User experience broadens the view to cover also aspects like stimulation, identification with the product, appeal, aesthetics, and trust and privacy, which all are highly subjective and inherently dynamic [34]. This allows broadening the theoretical base to consider also the experiential and emotional aspects that take place in interaction with technology: the design focus moves from removing negative factors and deficiencies in the user interface to creating positive experiences that surpass the users' expectations.

UX is associated with vague, dynamic and hard-to-quantify concepts, such as "experience", "perception", "pleasure", and "emotions". Additionally, UX depends on very dissimilar constructs that all encompass an endless area of research: that is, the users themselves, the technical systems and functionalities, product or service, and the contextual factors like social setting, cultural layers, and the users' other activities. The concept of UX is therefore challenging to define and be evaluated with regard to its characteristics. However, reaching a common definition can offer great advantages in developing products that truly appeal to the target users groups. The potential especially from the combination of AR and the experiential perspective create great business and service possibilities.

The main challenges related to UX can be divided to (1) designing a user experience that is pleasurable, engaging and stimulating, and appropriate in the user's context, and (2) evaluating the UX and overall acceptability of the applications. This chapter focuses on the latter by considering what aspects of the UX to evaluate and with what kind of measures to perform it.

Before continuing, certain terms need to be clarified and contextualized to this work. With *evaluation* we mean the systematic acquisition and assessment of information that provides useful feedback about the service in question. We agree with the broad consensus that the major goal of evaluation should be to influence decision-making (here, regarding design) through the provision of empirically-driven feedback [55]. With a *metric* we mean a way of measuring or evaluating a phenomenon or an object quantitatively. More specifically, it is based on (1) a set of individual *measures*, (2) references to a definition of the phenomenon, (3) derivation from models based on empirical evidence and systematic observations, and (4) reliable procedure agreed upon a community of practice. Furthermore, with a *measure* we mean a single artifact (often a part of a *metric*) used for measuring a specific aspect of a phenomenon, e.g. one subjective statement to be answered with a numeric Likert agreement *scale* from one to five. Here, *scale* refers to the level of measurement to obtain a discrete value of the measured phenomenon (e.g. ordinal scale, interval scale) [51]—in contrast to psychology where the term *scale* is often used to describe *a collection of questions* to measure a specific topic.

Vermeeren et al. [57] pointed out that, to date, there are few—if any—widely accepted standard methods with which to assess UX in general. What is more, there are no specific metrics for evaluating AR applications in specific or the effects of AR on the UX (for example the usefulness of the novel functionalities, emotional aspects arisen, intuitiveness and engagement in the novel interaction, and the power of AR to inspire and empower the user). However, the necessity of specific measures for UX in AR—or UX in general—is apparent: measures of UX allow benchmarking and selecting and iterating the most appropriate design solutions [34].

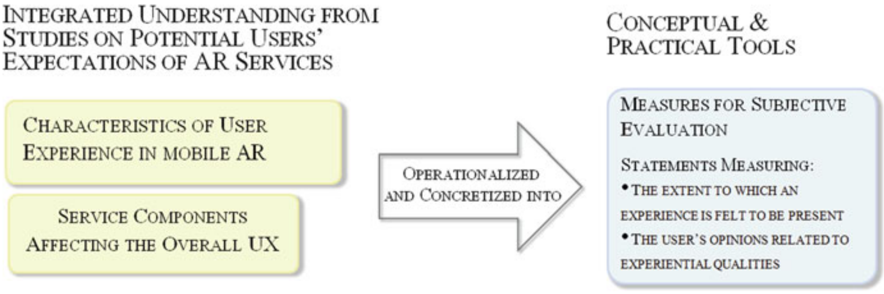


Fig. 9.1 Illustration of the research process: from theoretical understanding to practical tools for evaluation

1.2 Research Setting and Objectives

Our general objective is to understand the potential end users' central expectations of mobile AR services, especially from an *experiential point of view*. We want to elucidate what specific characteristics of UX are such that mobile AR is expected to allow or catalyze, as well as how different components of future AR services would influence the user experience. For this target we have conducted user research with various qualitative and quantitative methods, ranging from contextual interviews to online surveys and long-term diary studies [45–47, 56]. The focus has been on consumer-level mobile AR services in various daily situations, covering areas like exploration of the nearby environment, shopping and product comparison, and acquiring location-related and user-generated information. The summary of the characteristics of UX is presented in Sect. 3 and is based on an earlier synthesis presented in [44].

Consequently, this chapter discusses mostly aspects that have an origin in the research of user experience—instead of the "traditional" usability related issues and heuristics, such as system feedback, sense of control, consistency, or efficiency of the UI. This is by no means to say that usability issues would be irrelevant. We simply want to highlight the experiential aspects in the use of AR systems, and thus identify aspects that the previous AR-related user research has not much covered. Furthermore, there are other important aspects that affect the UX and especially user acceptance (i.e., decision of taking a system into use) of any system, but that have been purposely ignored here: for example pricing, brands, availability of related services and content, social acceptance of using new devices, as noted by, for example, Curtis et al. [7].

In addition to the descriptive understanding of UX and users' expectations, we have operationalized a set of user evaluation measures as concrete tools for AR practitioners in their service evaluation endeavors (Fig. 9.1). We claim that generic and technology-agnostic UX and usability evaluation metrics like AttrakDiff [1], PrEmo [11], PANAS [54] and self-assessment manikin (SAM) [33] do not usually provide concrete and specific enough results that would help in remedying the experiential issues in AR services. After all, user evaluations should be conducted to assess the quality of the target system and/or to help the developers to design, refine, or determine requirements for an application [49]. To provide a sound basis for measuring UX the measures need to be established with desirable properties of, for example, reliability and validity, which requires much research understanding of the related phenomena. On the other hand, the well-established and validated usability related metrics like nasa-TLX [20] and SUS [4], and metrics related to technology acceptance, e.g. [8, 24], do not focus on the experiential and emotional aspects.

Despite the great extent of methods that could be used in UX evaluation [57], there's a general lack of evaluation measures and metrics specific to certain technology domains that allow novel activities, a novel interaction paradigm and novel types of information content. It is not the novelty of technology per se that requires specific measures but the activities and interaction it allows and the experiences that these create. It is the interaction and the experiences that matter and are explicit to human—not the underlying technology. Carter and Mankoff [6] have also concluded that ubiquitous computing technologies, which AR undeniably demostrates, can require different metrics of success as they demonstrate so different interactions and use cases for the user when compared to "desktop" technologies.

The above-mentioned contextualized to AR, we can say that there is a need for methods and measures to scrutinize aspects like the experience of the augmented perception of one's environment, holistically interacting with the augmented environment, implications of location- and object-awareness, users creating AR content, and generally the new AR features the user is provided with. Drawing inspiration from the work of Scholtz and Consolvo [49] on ubiquitous computing in general, our premise is that identification of areas for evaluation (types of experiences in AR) accompanied with reasonable evaluation measures for mobile AR would advance the field. Our aim of bridging the AR-specific understanding of UX to relevant design and evaluation practices is addressed with (1) a conceptual assessment of the UX perspective in AR and (2) quantifiable measures for subjective evaluation of UX in mobile AR.

2 Related Work

In this section we present relevant related work to identify gaps in current research understanding and methods, hence further justifying our approach. We describe general UX frameworks and theories, and methods for evaluating the subjective UX. Finally, we summarize our prior research on user expectations, on which this work is grounding.

There exists a few focal usability and human factors related publications in the field of AR, focusing especially on perception issues, display technologies and using AR in collaboration and communication (e.g. [3, 18, 31, 37, 50]). Nevertheless, based on an extensive literature survey of papers between 1992 and 2007, Dünser et al. [12] estimate that only approximately 10% of AR-related papers have included some sort of user evaluation. Findings with similar notions of lack of user research in the AR domain have also been published by Swan and Gabbard [52]. What is more, when it comes to earlier research on UX aspects of mobile AR, there is even less to look into [45].

Bach and Scapin [2] have identified various challenges in evaluating the usability and ergonomics of mixed reality systems; for example, the testing should focus on aspects that are specific to the novelties of MR systems instead of generic, well-known, usability problems, and a lack of common knowledge from the various user tests (their methods and results) performed in different laboratories. Most importantly, they highlight a lack of, for example, common testing platforms and benchmarks in helping to compare and share the results. The evaluation methodologies have mostly utilized objective measurements like task completion times and error rates, which reveal very little of the user's subjective experience. Studies employing subjective user ratings or judgments or qualitative analysis have fallen in the minority. Dünser et al. [12] furthermore discuss that one reason for the lack of user research (esp. user experience aspects) could be lack of knowledge on how to conduct the evaluations and what kind of methods and metrics to use. We concur with this and have set our overall approach on this very gap in methodological knowledge.

Various authors have proposed well-justified methods and processes for user studies in the fields of AR and mixed reality. Nilsson [41] suggests, for example, to involve real end users in the design of the system and its evaluation and to make sure the tasks or scenarios are realistic, which more or less reflect the general principles of user-centered design. Gabbard et al. [17] propose a process for user-centered design and evaluation of virtual environments, including, e.g., user task analysis, formative user-centered evaluation and summative comparative evaluation. Furthermore, Gabbard and Swan [16] suggest that user experiments should employ tasks that are representative in the application or domain, the equipment and environments in the studies should be such that are likely used in the actual use as well, and that one should not forget iterative design. Similarly, there are some papers suggesting metrics for ubicomp-related systems in general (e.g. [49, 53]), however with most focus on usability-related issues like task completion and efficiency of the user interface. The framework of Scholtz and Consolvo [49] considers, for example, user attention, awareness, invisibility of the technology, conceptual models and intelligibility. The concepts of trust, appeal and behavioral and social impacts represent perhaps the most experience-oriented aspects in this particular framework.

All in all, the abovementioned methods and lessons learned place little focus on the experiential and emotional aspects. Despite the existence of a few seminal papers about how to utilize user-centered design and evaluation approach in the field of AR, there seems to be an acute need for having appropriate measures with which to assess the experiential aspects of the developed AR systems.

2.1 Assessing and Measuring UX

Recently, user experience as a general phenomenon in human-computer interaction has received rather much research effort. Various theories and frameworks demarcate and discuss: (1) the different types, categories or characteristics of experience that can take place in interaction with technology, (2) product features, contextual aspects or intrinsic human values and needs that create or contribute to the different experiences, and (3) practices and methods for envisioning, representing, and evaluating the UX [48].

According to the different frameworks that describe the *types of experience*, UX can be realized as, for example, pragmatic, symbolic, social or visceral experiences [5, 21, 26, 27, 48]. The following frameworks discuss the matter from slightly different standpoints.

Desmet and Hekkert [10] distinguish the characteristics (or “components”) of experience as (1) aesthetic experience: delighting one or more of our sensory modalities, (2) experience of meaning: personal or symbolic significance of products, and (3) emotional experience: emotions, such as joy and anger, elicited by the appraised relational meaning of the product. In another framework, Desmet [9] discusses the same from a perspective on emotional reactions, identifying five categories: (1) surprise emotions and amazement, (2) instrumental emotions (e.g. disappointment, satisfaction), (3) aesthetic emotions related to intrinsic pleasantness (e.g. disgust, attracted to), (4) social emotions (e.g. indignation, admiration), and (5) interest emotions (boredom, fascination) arising from, e.g., challenge. Furthermore, Jordan [27] has identified different types of *pleasure* in use of products: (1) physio-pleasure: to do with the body and senses, (2) psycho-pleasure: to do with the mind and the emotions, (3) socio-pleasure: to do with relationships and status, and (4) ideo-pleasure: to do with values and attitudes.

Buccini and Padovani [5] summarize earlier UX research by presenting a consolidated model of product experience with six categories: (1) experiences related to the senses: instinctive, with low cognitive performance, e.g. touch and appearance, (2) experiences related to feelings: subjective emotional reactions originated from the use of a product, (3) social experiences: happen between individuals but are intermediated by products, (4) cognitive experiences: related to the thought and interpretation by the user, e.g. based on semantic and symbolic product features, (5) use experiences: usability and functionality of the products, and (6) motivational experiences: when the use of a product is responsible for a certain behavior of the user.

Hassenzahl [21] has distinguished between two main perceptions of product quality: pragmatic and hedonic. Pragmatic quality refers to the product’s ability to support achievement of behavioral goals (e.g. product’s usefulness, usability, or appropriateness). Hedonic quality is divided into three dimensions: stimulation (e.g. enabling personal growth), identification (e.g. expressing and building one’s identity through the product) and evocation (provoking memories and emotions).

Furthermore, the experiences vary with regard to their *temporal granularity* or *time span* [34, 48]. The experience can be a *momentary* (refers to a specific change of a short-term feeling during interaction), *episodic* (appraisal of a specific episode of use or activity), and *cumulative* (overall views on a system as a whole, after having used it for a while; contrasted also with other products and through social processes).

All in all, these offer insight into the general characteristics of different experiences, their categorizations grounding on slightly different disciplines and theories. What is relevant for our work is the fact that these are general frameworks that do not focus on specific technologies, contexts of use or application areas. The question of *what UX is in general* has been rather well addressed but there is little knowledge of how the general level experience characteristics are manifested in specific domains or technologies, such as mobile AR.

With regard to the UX evaluation methods and metrics, the survey by Vermeeren et al. [57] managed to gather 96 methods for evaluating user experience. The methods vary in terms of what is the sources of the research data, what kind of data they allow to be gathered, in which phase of development they can be utilized, and what dimensions of the overall UX can be covered. Despite this diversity, they conclude a need for validated methods for specific experience focuses and domains, and for deeper understanding of UX to base the evaluation methods on.

Overall, UX evaluation is in need of well-defined and quantifiable metrics to determine: (1) the users' expectations before using the service and (2) the users' perceptions and experiences after having interacted with the service (see e.g. [34]).

Stemming from pedagogy and later also from usability, there are two commonly used approaches to user-based evaluation. Contextualized to information technology, *formative* evaluations are conducted to help strengthen or improve a service, product or interface as part of an iterative design process. They examine the service (or product) and its features, and the quality of their implementation from various points of view. *Summative* evaluations, in contrast, aim at assessing the overall quality (often referred as "goodness") of the service, product or interface, the effects or outcomes of it, and its ability to do what it was designed to do (i.e. *potency* or *efficacy*) [40, 55].

There is a continuous discussion about the necessity and utility of numeric measurements, some strongly advocating this and some being ambivalent about the role of numerical values in understanding of the complex interaction between human and technology [34]. Either way, measures can add structure to the design and evaluation process, provide easily understandable and comparable information to decision makers (e.g. ROI of a design decision), and, to some extent, verify improvements in the design.

2.2 *Our Prior Research on Expected UX of Mobile AR*

Our own earlier work has incorporated various user experience theories by analyzing the nature and dimensions of UX in the specific domain of AR. Several qualitative interviewing studies [45, 47, 56] have shed light on the characteristics of user experience that people expect of future mobile AR services (see Table 9.1 and

Table 9.1 Summary of the methods and participants in the earlier studies that are consolidated and further analyzed in this chapter

Paper	Method	Number, nationalities and type of participants
[45]	5 focus groups with user scenarios about AR in tourism and in day-to-day life as stimuli	23 Finnish, mostly early adopters: groups of travelers, students, tech-savvy, wellness-oriented, and people with green values
[47]	16 contextual interviews in shopping centers; individual, pair and group interviewing	28 Finnish, mostly early adopters
[56]	2-week personal diary to gather momentary needs + group interviewing and surveys	9 Finnish, early adopters: active in content creation and using location-based services
[46]	Online survey with both open questions and statements about five mobile AR scenarios	182 in the Finnish and 80 in the English version, mostly early adopters

Fig. 9.2 Examples of methods used in our studies. *Top left*: an illustration used with a textual scenario [45]. *Top Right*: paper diary to elicit needs for AR information in everyday life [56]. *Bottom*: a participatory design session (not published)

Fig. 9.2 for summary and illustrations of the used methods). These studies have pointed that, for example, mobile AR services are expected to empower people with novel context-sensitive and proactive functionalities, make their activities more efficient, raise awareness of the information related to their surroundings with a very intuitive interface, and provide personally and contextually relevant information that is reliable and up-to-date. Additionally, expectations of offering stimulating and pleasant experiences, such as playfulness, inspiration, liveliness, captivation, and surprise were identified.

Another study [46] evaluated the user acceptance of five different futuristic AR scenarios with an online survey approach. Mobile AR was seen to make possible novel interactions with the surrounding environments and provide information that has not been easily accessible before. As negative aspects the respondents highlighted, for example, fears of information flood, users' loss of autonomy, and virtual experiences and information replacing the real.

Overall, the studies have focused on the expected experience of future mobile AR services. This understanding allows both to anticipate what the UX can be like and to transfer this knowledge to practices of evaluating the episodic and cumulative user experiences created in actual interaction with mobile AR. The findings of the studies are consolidated in Sect. 3 with a perspective to the characteristics of expected experience. The consolidation is based on a qualitative cross analysis of the four research cases and aim to describe the *diversity of experiences* that have been identified.

3 Consolidating the Nature of Desirable UX in Mobile AR

This section consolidates our earlier research of potential users' expectations of mobile AR in terms of *what kinds of desirable experiences are expected to arise in the use of mobile AR services*. We describe 16 categories of experience that have been identified in our studies and thus are expected to be salient in the users' interaction with future mobile AR services. In addition to expanding the experience and its characteristics, we discuss how mobile AR and its different components influence or contribute to it.

The described experiences here can all be considered as positive and satisfying experiences. This stems from the participants mainly expressing expectations that they look forward to and consider desirable. Naturally, negative experiences and risks were also brought out to some extent but these have intentionally been omitted from this synthesis. Consequently, the following experience descriptions can serve for AR developers as inspiration and targets for design, as well as a theoretical baseline against which to compare and assess design solutions and developed prototypes. In this chapter they serve mostly as a theoretical basis for operationalizing appropriate measures for UX evaluation (Sect. 4).

The 16 categories of experience are roughly classified into six classes that represent abstract level types of UX: (1) instrumental experiences, (2) cognitive and epistemic experiences, (3) emotional experiences, (4) sensory experiences, (5) motivational experiences, and (6) social experiences. This classification is based on the work by Buccini and Padovani [5] discussed earlier. However, the descriptions of the classes have been slightly fashioned to better cover our categories of experience as abstract headings. For example, *use experiences* are termed *instrumental* experiences. The categories of experience have been placed in the classes based on where we see them to *primarily* belong to. That is, as the classes are not totally exclusionary and the experiences categories contain multifaceted types of experiences, some experiences might be related to other classes as well.

3.1 Instrumental Experiences

First, several experiences have been identified to relate to *instrumental aspects* in product or service use. Such pragmatic experiences demonstrate and originate from, for example, utility, user's accomplishment, product performance, and support for the user's activities.

1. *Empowerment* relates to the feeling of being provided with novel possibilities, instruments and ways of accessing, creating, and utilizing information. Such augmentation of human perception and activities catalyzes feelings of powerfulness and achievement. An AR service would serve as a practical tool to gain new knowledge, perform truly novel activities with information technology and the surrounding world (e.g., cause instant reactions in the physical world), and pursue goals that have been unavailable before. The AR features allow the user to access location- and object-related information by simply browsing the target visually, perceive things in the environment that cannot be perceived otherwise (e.g. things behind physical obstacles), and create virtual content and instantly link it to relevant real-world counterparts.
2. *Efficiency* describes the feeling of being able to perform everyday tasks and activities with less effort, time and other resources. Although it has its roots in usability theories [26] the perspective here is on the user's perceived and experienced efficiency instead of the objective efficiency (performance). As a result of easily noticeable and comprehensible interface and enabling location- and object-based interaction, AR services can provide a very powerful way to access information in situ. Thus, it facilitates the user's feeling of accomplishment and, for example, supports in consumer decision making. For example, less effort is needed in specifying the points of interest and information one is trying to acquire. Oftentimes an efficient performance creates experiences of satisfaction, and allows evocation of emotional experiences like delight and surprise.
3. *Meaningfulness* relates to the AR service appearing personally meaningful, appropriate and relevant in the user's current context and the activity one is engaged in. AR was expected to show only the content that corresponds to the surrounding real world, thus making it feel relevant in the current location. In addition, expectations of content being personalized based on the user's current needs and behavior were often brought out in the studies. In mobile use, AR is a very appropriate way to access personalized information in situ and visualize it for personal use—in contrast to e.g. mobile web browsing or stationary public displays. Furthermore, a central requirement was that the content should be trustworthy, especially when accessed via AR in important day-to-day activities like navigation or browsing content with a specific purpose in mind.

3.2 Cognitive and Epistemic Experiences

The *cognitive and epistemic* experiences relate to thoughts, human information processing and rationality. Such experiences stem from the product's or service's semantic features and abilities to arouse curiosity and satisfy a desire for knowledge. Two categories of experience have been identified to belong mainly to this class.

4. *Awareness* describes the increased insight into one's surroundings and the related digital elements (i.e. the "digital landscape"). With AR, the embedded, latent, digital information in the environment becomes perceivable and explicit. Awareness can be manifested as becoming aware of, realizing something about or gaining a new viewpoint to one's immediate surroundings (e.g. locations and objects). It encompasses both the momentary awareness of the current surroundings and the increased overall understanding of a place or object over time (e.g. its history, meanings and memories attached). Highly related concepts are discovery and surprise, especially in regard to finding novel aspects of already familiar environments. The merging of realities and AR proactively and holistically illuminating the virtual aspects of a physical environment play great roles in creating this experience.
5. *Intuitiveness* relates to the feeling of naturalness and human-likeness in interacting with the AR information. It can be argued to be highly natural that the ontology of information and interaction possibilities are related to certain locations or physical objects. Furthermore, the way of aligning AR content to the real world is intuitive as the AR content relates only to the visible (or otherwise perceivable) things. This reduces ambiguity and helps in balancing between too exiguous information and information overflow. AR allows a rich and multimodal interaction experience that mimics the already familiar interactions in real world: the interaction is instant, continuous, uninterrupted and real-time, it allows direct manipulation, and it is spatial in 3D (or 2.5D). Considering the work by Scholtz and Consolvo [49], this partly relates to *conceptual model* and *predictability* of application behavior.

3.3 Emotional Experiences

Emotional experiences relate to the subjective emotional (visceral) reactions originated from the use of a product: for example, pleasure, entertainment, evoking memories and facilitating positively valued feelings overall. In our studies we have identified four specific experiences related to this class.

6. *Amazement* relates to the feeling of having experienced or achieved something extraordinary or novel, hence often represented as "wow"-effect. In AR, amazement can often be emphasized in the first time of use, thus attributed especially to the charm of novelty in the interaction paradigm or service functionalities. However, it can also be considered as the emotional element in the experiences

of *awareness*, *empowerment*, and *inspiration*, which open up possibilities in creating positive amazement after the charm of novelty.

7. *Surprise* is due to receiving contextually relevant, extraordinary, and useful information (e.g. being positively astonished of the user-generated content), and surpassed expectations in general (e.g. the application performing much better than expected). Surprise has also elements of *cognitive* experiences (e.g. the potential of AR to arouse further interest towards the service and the mixed reality, and proactively please the desire for knowledge) but was put in this class because of its fundamentally emotional nature. Proactive features and context-awareness mainly initiate this experience but it is also greatly facilitated by the pervasiveness of AR as a visualization paradigm.
8. *Playfulness* refers to feelings of joy, amusement and playfulness. Playful AR can be manifested both in services that aim to entertain (e.g. games, augmenting the appearance of physical things) and in pragmatic services, in which the pervasiveness of AR and the accessed content can evoke playful and amusing feelings. In addition, the novel and unparalleled way of interacting with AR can feel playful (e.g. direct manipulation of the augmented objects). Playful experiences could arise from elements of, for example, positive challenges, competition, thrill, humour or fantasy [30]. All of these could be applied to AR, both regarding the content and the way of interacting with it. Gaming-like and playful aspects can catalyze other experiences like inspiration, surprise and amazement.
9. *Liveliness* relates to the feeling of continuous change and accumulation of the service. An AR environment with various types of content can seem vivid and dynamic, thus evoke positive feelings of vivacity, revive pleasing memories and catalyze content sharing between service users. This derives not only from socially constructed AR content (e.g. aggregated from other services and domains), but also from the comprehensiveness of AR as an interface. A certain information or media might change radically or become unavailable over time, which can be perceived with interest or, on the other hand, can also make the environment seem unpredictable. In addition, the AR view to a certain physical place might look different for different users as the content might be personalized and based on the social connections with other users.

3.4 Sensory Experiences

In the classification by Buccini and Padovani [5] *sensory experiences* relate to instinctive, non-cognitive sense related experiences, such as pleasure from touch. Here, the following experiences are sensory by origin but contain also cognitive aspects and can be conscious. These originate from a product’s or service’s ability to arouse sensory and physical pleasure, immersion and captivation, and visual, tactile, and auditory aesthetics.

10. *Captivation* describes the feeling of being immersed and engaged in the interaction with the environment enriched with AR content. It relates both to the

user's sensory-perceptual impacts and the spatial engagement and enjoyment of the imaginary world created through the system [13]. AR as a technology has potential in captivating the attention of the user, increasing the sense of *presence* in the environment, engaging or orienting the user towards further interaction with the mixed reality, and, at best, leading to a feeling of *flow* in one's activities. Considering the above-mentioned, captivation belongs also to the classes of cognitive and emotional experiences. With regard to design, multimodal way of interaction or representation of information is not only the most efficient but also the most engaging. On the other hand, AR should not disturb the understanding of what is real/virtual or the enjoying of the real world, or incur physical risks as a result of being used in mobile, congested contexts.

11. *Tangibility and transparency* describes the senses of concreteness and coherence of environment-related content and the resulted augmented environment. Because of the pervasiveness and visual comprehensiveness of AR, the augmented content seems a tangible and integral part of the environment. This, as well as captivation, can further lead to feelings of presence and unity with the surroundings. The devices are transparent in the interaction and allow the users to concentrate on the augmented environment itself. This is related to the concept of *attention* [49], which refers to the easiness of observing the overall picture of the AR environment without needing to change the focus of attention from a device or interaction to another.

3.5 Social Experiences

Social experiences relate to and originate from human to human interactions and are intermediated by technology. These originate from features that allow building or communicating one's identity or status, provide a channel for self-expression, or otherwise support social user values, such as feeling of relatedness. Our studies revealed two categories of experience that fall into this class.

12. *Collectivity and Connectedness* relate to the feelings of participating into a user community, having novel ways for social interaction and communication, and being aware of other people using the AR service (e.g. sharing AR information, enriching AR information created by other users, peer-to-peer entertainment). A sense of community can result from collectively producing and contributing to the AR content [42], and awareness of other users creating new information and utilizing the existing. This category could be further divided into smaller entities but is discussed here as one because AR as such contributes rather little to creating such social experiences. AR mainly *catalyzes* the experiences and provides an interface to the location- and object-based socially constructed content.
13. *Privacy* relates here both to the sense of privacy resulting from how much and what kind of information about the user is logged by the service and publicly available, and the sense of social awkwardness that results from the obtrusive way of interacting with mobile AR. Privacy is a "hygiene factor" experience

that, when realized well (e.g. unobtrusive data glasses), is not even perceived or bothered by the user, but, when realized badly, it disturbs the user and might disable positive experiences. As stated by Scholtz and Consolvo [49], the more information is shared, the more the user's awareness can be increased but often at a cost in privacy. Sense of privacy is mostly a result of other service elements than AR as such. However, AR can create unwanted privacy issues because of its easy-to-use—and possibly perceivable by others—interface can also be used to access privacy-sensitive personal information.

3.6 Motivational and Behavioral Experiences

Finally, *motivational and behavioral* experiences are created when the use or owning a product or service causes a certain behavior in the users: for example, inspires or motivates them to do something or pursue a goal with the help of technology. Our studies revealed three categories of experience that relate to such aspects.

14. *Inspiration* relates to feelings of being cognitively stimulated (relates also to the class of cognitive and epistemic), curious and eager to try new things or appropriate the AR services for new purposes. The AR interaction paradigm especially demonstrates elements that are able to inspire and stimulate people: mixing realities, services, and types of information (e.g. information from a web site visualized in AR), exposing the immaterial values related to objects and locations, comprehensibleness in application areas and contexts where AR could be utilized, and the AR content being partly socially constructed. For example, the physical surrounding could be used as a visual platform for AR games, art, or entertainment. Additionally, over time certain needs for activities lose significance for the user and are changed to other ones, which further create new possibilities for appropriation, i.e., assigning new purposes to the service.
15. *Motivation* relates to the feeling of being encouraged and motivated to participate in the service community and contribute to its content, or simply to do tedious mundane tasks with the help of information technology. Hence, mobile AR could also be considered as persuasive technology by facilitating behavioral change [15]. This experience is highly attributed to the novelty values of mobile AR (e.g. way of acquiring information and proactive service features), and partially to playfulness of the interaction paradigm and visualization of AR information. From content authoring perspective, the immediacy and pervasiveness of AR interaction lets the content created by the user to be easily enjoyed or utilized by other users.
16. *Creativity* represents self-expressive and artistic feelings in users creating AR content and in mixing the digital with the real world in previously unimaginable ways. AR has great potential to trigger imagination and serves as a fruitful interface to demonstrate artistic creativity, for example by utilizing the real in setting interesting physical contexts and frames for digital media like video,

audio, and imagery. Especially in our online survey study [46] the aspect of the user sharing one's creativity through AR was much discussed, for example in the form of virtual graffiti.

The aforementioned 16 categories represent diverse desirable experiences that were identified in our research of potential users' expectations of mobile AR. However, theoretically describing a complex concept like user experience is challenging and inherently limited with regard to comprehensiveness. Therefore, this is not meant to be a comprehensive list incorporating the entire extent of experiences that can take place with mobile AR services. Such experiences that the potential users have not been able to envision or identify further broaden the diversity of possible experiences (e.g. those evoked in actual interaction and in long-term use). Examples of such could be enchantment [39], experiences of space and deeper immersion, experiences of art (see e.g. [35]), instinctive sensory experiences like visual aesthetics, and various social and symbolic experiences, such as social admiration due to owning novel technology. Such aspects, however, have not appeared in our studies so far, and assessing their relevance in future mobile AR services requires further research.

As for the relevance of the categories to AR, we could argue that not all of these are that specific to AR as an interaction paradigm. However, the categories summarize what potential users expect from mobile AR services *as entity*. Not all experiences mainly result from AR per se but they are nevertheless relevant in the service entity. In addition, it is challenging to forecast which categories will be emphasized over time in the actual use. These considerations lead us to describe the whole extent of identified expected experiences instead of omitting some experiences from this paper based on a conjectural dichotomy of what is AR-specific and what not. Furthermore, it is worth remarking the continual convergence of technologies and services: for example AR with embodied and multimodal interaction technologies, sensor technologies, content from the Internet, as well as other mobile technologies. In the future, the aspects that are relevant in mobile AR services might be largely based on the experiences that can be considered more specific to other technologies or domains.

The different experiences are naturally slightly different with regard to how frequent they can be in the overall use and what is the level of abstraction in relation to the other categories. However, with such a priori study approach it is nonsensical to approximate the frequency or extent of each experience—at this phase of research it is more valuable to understand the diversity of experiences that can take place in the first place.

3.7 Service Components Contributing to the Experiences

In addition to the insight into experience characteristics, it is important to understand how the overall UX in mobile AR is constructed and from what elements the experiences can originate. According to Hassenzahl [22], the experiences can

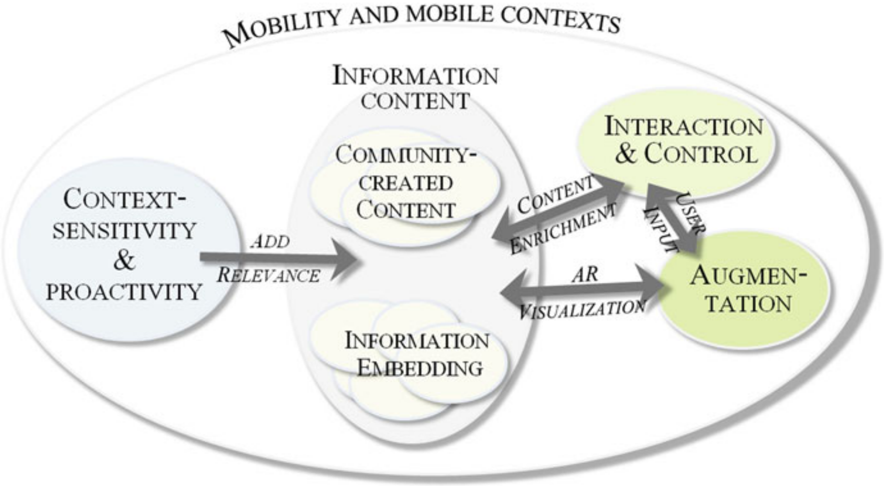


Fig. 9.3 Service components of AR services from a user experience perspective

originate from different features of the technology use: (1) the information content that is accessed, created and used, as well as its characteristics, such as origin and perceived quality and relevance, (2) functionalities that the service provides, (3) presentation and interface to the content and technology, and (4) the interaction through which the service is controlled and reacted to.

Figure 9.3 summarizes the different components that we have identified as something that people attributed their expectations to in the user studies or expected related features to be present in AR services. These can be said to play a role in initiating, creating, or catalyzing the abovementioned experience categories, thus affecting the overall UX of mobile AR. In other words, each experience category can result from one or several of these components. As the components vary with regard to how specific to AR vs. generic they are in nature, different types of user requirements and expectations are directed to them: for example, general user requirements related to mobile technologies are directed to the *mobility* and *context-sensitivity* elements, whereas mostly the components of *augmentation* and *interaction and control* are burdened with novel expectations that are more specific to AR. This further consolidates that the potential users' expectations of AR services are often directed to other technological layers than AR per se.

- Augmentation: characteristics of augmented reality as the system output, that is, an egocentric view, 3D and realistic spatial alignment and rendering (visualization) with appropriate occlusions, lighting, shadows and reflections, and visualizing digital interactivity affordances in the environment. Based on potential users' expectations it is too early to specify exactly which experiences and how this service element contributes to. However, to surmise and create hypothesis, this can be expected to be the principal origin of the experiences of amazement, captivation and tangibility and transparency. Furthermore, this can be seen as important contributor to the experiences of empowerment, efficiency, increased awareness, intuitiveness and privacy, and as a facilitator of experiences like liveliness.

- Interaction and control: the way of controlling and providing input to the service and required mobile devices, and interacting with the AR content (e.g. browsing the augmented reality to access content and creating and enriching the content in the service). This component can be seen to contribute especially to the behavioral experiences of creativity and inspiration, playfulness, captivation, efficiency, and indirectly also to liveliness and collectivity and connectedness.
- Information embedding: real-world objects and locations embedded with or linked to additional digital content that is accessed with AR and computer vision (i.e. “internet of things” or “physical browsing”). This component contributes especially to the cognitive and epistemic experiences like intuitiveness and awareness, and to tangibility and transparency as sensory experiences.
- Community-created content: user-created content, crowd sourcing and collaboration in content authoring and the content being modifiable and increasable by the service users. This is a recently well-established phenomenon in, for example, the Internet and map-based services. This can be expected to affect especially the experiences of liveliness, surprise and collectivity and connectedness, and partly also to meaningfulness, playfulness, creativity and privacy.
- Context-sensitivity and proactivity: service functionalities and content being determined by and adaptive to the user’s context, such as location and social surroundings, and proactively initiating interaction with the user (e.g. location-dependent notifications). This can be seen to contribute especially to the experiences of empowerment, efficiency, surprise and of acquiring content that is relevant in the current context (meaningfulness).
- Mobility: the technology being usable in mobile contexts and activities, and allowing “anytime, anywhere” and “in situ” kind of interactions; mobile devices as interaction devices. This can be seen as an overarching component behind all of the others and thus partly contribute to all of the experiences. However, we regard it to influence most the general experiences like privacy, liveliness, efficiency and empowerment.

4 Measuring UX of Mobile AR

The diversity of characteristics of experience and the different technological components affecting the UX solidifies that user experience of mobile AR can be a complex whole. Nevertheless, the scrutiny and theorization presented above displays methodological usefulness in terms of allowing understanding the phenomenon from several standpoints, and making it possible to build both conceptual tools and evaluation measures for AR developers and UX practitioners.

4.1 Methodological Considerations for Evaluating UX

Drawing from the general UX literature, next we highlight some key points of methodological consideration for designing UX evaluations. Most of these aspects are important to consider in any domain of interactive technology but are discussed here especially with the perspective to mobile AR. Furthermore, these serve as background to the next section where we propose evaluation measures specific to AR.

First, the UX measurements in general should essentially be *self-reported* in order to cover the subjective nature of UX [34]. As the aforementioned experiences are mostly conscious (and cognitively processed), they are very challenging to be evaluated with objective means, such as psycho-physiological measures, eye-tracking, observation, or logging UI actions. Present-day objective measures might reveal that, for example, the user is focusing on a specific part of the interface, a specific type of affect has been momentarily evoked, or her cognitive activity is high. However, making conclusions of what these mean on the overall subjective experiential level or how the system should be improved is challenging. Subjective self-reported are needed from the user to make sense of the objective measurement. Given the aforementioned, subjective evaluation requires cognitive processing and reflection from the users to allow them to, first, identify the evoked experiences and, second, verbalize or otherwise report them in detail. This requires a suitable mindset and appropriate instructions for the users to ensure that they are considering the aspects that actually are researched.

Second, it is important to allow the experiences to take place in situ and in contexts where AR services would be used in actual situations—that is, field studies rather than laboratory-based studies [14]. As AR inherently rests on mixing the real world with virtual, most of the experiences discussed earlier require an authentic setting to arise. This would include the physical environment to be augmented, content specifically related to it, authentic tasks for the users, and desirably also other people using the service.

Third, triangulation of methods in data gathering allows insight from multiple perspectives and addresses the studied phenomena with greater coverage: for example, using subjective statements that provide quantifiable values together with qualitative data gathering methods like interviews or sentence completions [32]. Regarding UX of AR, *summative* measures [40, 55] can provide a numeric overview of the overall extent to which the service succeeds in creating the intended experiences. Even approximate measurements are useful in numerically assessing how successful the product is with regard to its efficacy and ability to evoke the intended experiences. Quantitative summative metrics with high number of informants (i.e. evaluator users) allow making more valid conclusions and generalizations to the entire user population. *Formative* assessment [40, 55] provides specific quantitative measures or qualitative insights into how a certain feature or the entire service could be improved from the experiential standpoint. Quantitative formative evaluation can help in identifying, for example, the most or least appreciated features or how their implementation is valued with regard to usability or user experience. On the other hand, qualitative evaluation identifies important challenges and issues that matter in the use within the application area, the findings can often be transferred to other design cases as well.

Furthermore, related to triangulation and holistic coverage of UX, the evaluation data gathering methods should preferably cover the different levels of temporal granularity of the use: (1) the expectations towards the service measured in order to set a baseline or standard, for example with remote surveys (as in [43, 46] or Wizard of Oz [29] or other prototyping methods), (2) the momentary, instant experiences and emotions measured with, e.g., mobile surveys or other quick and easy-to-use measures in order not to interrupt the activity with the service too much, (3) episodic experiences with, e.g., interviewing, sentence completions [32], comparisons to other systems that allow similar activities, or the subjective statements presented in the next section, (4) the long-term UX with, e.g., diary methods or the subjective statements presented in the next section. In addition, the measures used on each level of temporal granularity need to be tailored accordingly. For example, it can be irrelevant or unreliable to ask the user about generic emotions that one has had while using a service for some time ago (e.g. feeling attentive, determined or bored) as those are often too weak to be remembered. After a while, people usually remember the things that have created an emotion (service features) better than the subtleties of the experience itself.

Fourth, especially with summative assessments, it is beneficial to have a baseline or point of reference against which the user may assess the numerical value of the goodness of the system in each regard. AR services could be compared with for example, existing geotagged content services (e.g. www.flickr.com, map services, POI repositories), browsing and searching information in the web (e.g. search engines, Wikipedia and other publicly available information sources), and virtual worlds with mixed reality interactions (e.g. second life).

Finally, the target user group affects the usefulness of the results of various evaluation methods. Most often, qualitative methods demand more of the users in regard to their ability to observe and analyze one's actions and reactions, ability to verbalize one's thoughts, and more motivation to put effort into the service development. Using, for example, agreement statements lowers the threshold to participate and eases the user's cognitive effort. However, they do not allow formative evaluation as well as purely qualitative methods, such as interviews.

4.2 Subjective Measures for UX of Mobile AR

To concretize our prior research and to provide practical instruments for evaluation, we present two sets of subjective statements (summative and formative) related to the different categories of UX in mobile AR. These statements are meant as measures that could be used in various user questionnaires to evaluate the users' subjective insights into the user experience of the service. The statements have not yet been used in evaluation cases, and therefore should be treated as propositions rather than validated evaluation tools. Nevertheless, they provide a sound starting point for designing evaluations and further measures that focus on the *experiential aspects*.

We present measures that are meant to be (1) generic enough to fit any type of consumer-targeted mobile AR service but at the same time (2) specific to the features that AR-services have been expected to have. The statements are mainly intended for evaluating functionally full-fledged services—not early prototypes or demonstrators. Especially the summative statements in Sect. 4.2.1 require a long-term experience with the service in question. The formative statements (Sect. 4.2.2) could be used in evaluating prototypes to some extent but this requires careful selection of the most suitable and appropriate measures. For consistency, all the statements are formulated so that agreeing with them indicates that the service performs well or that the user experience is good in this respect.

The measures were devised by identifying and concretizing specific aspects in the user studies based on which the experience categories were abstracted. Originally roughly 200 statements were formulated, of which 80 were selected and refined for this paper. The elimination and iteration process took into consideration the statements' estimated relevance in various mobile AR services, usefulness as a measure and representativeness with regard to the experience category description, and included a careful inspection of the understandability and correctness of the terminology and language.

4.2.1 Summative Measures

To start with, Table 9.2 provides subjective statements for *summative* evaluation, measuring *the extent to which an experience is felt to be present*. The temporal perspective is on *cumulative experience*, that is, after having used the evaluated service for e.g. several weeks. The intended scale for these statements is an ordinal scale of frequency of occurrence: e.g. a five-step scale: (1) never, (2) rarely, (3) sometimes, (4) often, (5) always. Being summative and measuring the overall *frequency* of the experiences, the statements evaluate how well the experiential design targets have been met (the experiential efficacy of the service). For each category, we present two to three statements with slightly different perspectives.

In addition to the frequency scale, other scales, such as Likert agreement, could be used to measure, for example, to what extent does the user agree with having experienced the aforementioned in a shorter episode of use or how strong the experience has been. This, however, changes the focus of the evaluation and would require rephrasing some of the statements.

4.2.2 Formative Measures

Table 9.3 provides formative subjective statements that measure the *user's opinions related to experience-related service qualities* and goodness of the implementation with regard to UX: for example, how well specific features allow or disable

Table 9.2 Examples of summative subjective statements with regard to the *presence of the experience* category in question (to be used with a frequency scale: *never*, *rarely*, *sometimes*, *often*, *always*)

Experience category	Subjective statement
Empowerment	When using [*name of the service*] I feel powerful and competent When using [ ] I feel that my senses are enhanced
Efficiency	When using [ ] I feel that I'm efficient in my activities When using [ ] I feel satisfied with how well I perform and accomplish things
Meaningfulness	Using [ ] feels a personally meaningful way to acquire information I feel that using [ ] is appropriate considering my goals
Awareness	When using [ ] I feel that I am aware of the information related to my surroundings When using [ ] I feel to discover things
Intuitiveness	The way of interacting with the augmented environment in [ ] feels natural to me When using [ ] I feel like I am interacting directly with the real world
Amazement	When using [ ] I feel to be engaged in something extraordinary When using [ ] I enjoy experiences of amazement and fascination
Surprise	When using [ ] I enjoy positive moments of surprise When using [ ] I feel joy of finding new things
Playfulness	When using [ ] I feel myself playful When using [ ] I feel amused
Liveliness	Using [ ] feels positively vivid and lively When using [ ] I feel intrigued of the richness of content
Captivation	When using [ ] I feel that I am captivated by the augmented environment When using [ ] I feel that I am present in the surrounding environment When using [ ] I feel that I just want to carry on with it
Tangibility and transparency	When using [ ] I feel that the augmented content is an organic part of the environment When using [ ] I feel that I'm interacting with the environment itself rather than with a device
Collectivity and connectedness	When using [ ] I feel closely connected to other people using the service When using [ ] I feel that I am contributing to a meaningful community
Privacy	When using [ ] I feel comfortable about what other users can know about me When interacting with [ ] I don't feel awkward or embarrassed
Inspiration	When using [ ] I feel myself curious When using [ ] I feel that my desire for knowledge is satisfied When using [ ] I feel like I'm on a journey of exploration in my surroundings
Motivation	When using [ ] I feel encouraged to contribute to the service content When using [ ] I feel motivated and diligent
Creativity	When using [ ] I feel imaginative Enriching the physical environment with [ ] feels creative

Table 9.3 Examples of formative subjective statements with regard to the value and *overall goodness of the service* in terms of the UX category in question (to be used with a Likert scale)

Experience category	I think that
Empowerment	… [*name of the service*] allows me to pursue goals that are not supported by other technology … with [ ] I am able to acquire and utilize information that has been hard to access before
Efficiency	…with [ ] I can perform my activities with low effort …with [ ] I can efficiently make decisions in my everyday activities …with [ ] I can access information quickly while mobile
Meaningfulness	…with [ ] I can access information at the most appropriate place and moment …the content I access and use with [ ] is up-to-date and reliable …the content of [ ] makes sense in the context I use it …[ ] provides me with the most suitable amount of information
Awareness	…with [ ] I gain interesting perspectives to the surrounding world …[ ] expands my understanding of already familiar places or objects …with [ ] I can easily understand and react to my surroundings
Intuitiveness	…[ ] allows a natural way to interact with location- or object-specific digital information …it is easy to understand which things in the real world the AR content of [ ] is related to …it is easy to distinguish between the real world and the augmented content of [ ]
Amazement	…I find great pleasure in the content accessed with [ ] …is [ ] novel and unique …with [ ] I can view intriguing perspectives to my surroundings
Surprise	…with [ ] I can find unexpected or surprising information …browsing content as augmentations helps me find the most astonishing content …[ ] works surprisingly better than I expected
Playfulness	…it is fun to view the content of [ ] as augmentations …the content of [ ] is entertaining …even the everyday routine activities feel cozy with [ ]
Liveliness	…it is pleasant that the content of [ ] continually evolves …the content of [ ] evokes memories in me …[ ] provides me with something new and interesting every day
Captivation	…I have a good conception of what is real and what is augmented in [ ] …the authenticity of the augmented environment in [ ] engages me …the interaction with [ ] captivates my attention in a positive way
Tangibility and transparency	…I can easily understand what information there is related to the physical objects I view …the augmented content in [ ] seems concrete …with [ ] I'm able to perceive my surroundings comprehensively
Collectivity and connectedness	…[ ] is a suitable service for me to view and browse socially created content …I am delighted with how other users of [ ] enrich the augmented environment
Privacy	…the way of interacting with [ ] is not too obtrusive in crowded environments …the information I have created is safe in [ ]

(continued)

Table 9.3 (continued)

Experience category	I think that
Inspiration	…when using [ ] I come up with new purposes or ways to use it for …with [ ] I can seize the day and make the most of the present moment …using [ ] further increases my enthusiasm for it
Motivation	…[ ] encourages me to produce and share information with other users …using [ ] is a good way of sharing my insights of certain places or objects …[ ] motivates me to do even the most tedious tasks I can do with it
Creativity	…using [ ] is a channel to express my artistic self …using [ ] encourages my imagination …[ ] allows me to mix the real world with the digital in novel ways

something, or stimulate or catalyze an experience. As the focus here is on the service itself—instead of the resulted experience—these can inform design by identifying problematic aspects in the service. The statements have been formulated so that they allow measuring both the *episodic experience* (e.g. quickly after a single episode of use of the service) and the *cumulative experience* (overall experience over time). These statements have been designed to be used with a Likert agreement scale with five or, preferably, seven steps (fixed extremes one: "strongly disagree" and seven: "strongly agree"). For each category, we present two to four statements with slightly different perspectives.

4.3 Methodological Discussion on the Measures

This section discusses the applicability of the measures to various evaluation cases and highlights certain advantages and limitations to be taken into consideration.

There are several statements related to each experience category, each focusing on a specific aspect. As an evaluator of a mobile AR service one should choose the statements that seem the most appropriate in the evaluation case in question—both with regard to the different statements within one category and the different categories overall. Including all the statements ensures the breadth and coverage of the evaluation but at the same time might create unnecessary intricacy in the evaluation. For example, services that do not demonstrate user-generated content or social features might not need to be evaluated with regard to the experiences of collectivity and connectedness. On the other hand, in some contexts or application areas these measures might be insufficient with regard to the specificity or even the number of measures. An alternative way to apply the proposed measures is to draw inspiration from them and further operationalize them into interview questions in qualitative studies or more specific quantitative measures to be used, for example, with specific contexts of use, types of information, or types of users. In some cases, the statements might require more specific terms and wordings. For example, the statement "the authenticity of the augmented environment…" could be replaced with "the *three-dimensionality* of the augmented environment…" The correctives naturally depend on numerous details in the specific case.

Considering a more extensive use of the measures, especially the statements in Table 9.2 allow applicability to other technology domains and applications areas as well. After all, the summative statements in Table 9.2 measure the presence of an experience and most of the presented experiences can be relevant in, for example, other ubiquitous computing and mobile technologies, location-based services, services with other kind of tangible and embodied interaction, or games with mixed reality elements.

Regarding the advantages, subjective statements demonstrate practical usefulness (as identified, e.g., in [58]) that is meaningful here as well and allows highlighting the following. The proposed statements allow evaluating an AR service: (1) with various numbers of users and various amounts of their usage of the target service, (2) with different scales, and (3) both in situ (in field), remotely and in-lab if seen necessary. Furthermore, (4) they support repeatability well and (5) they allow easy quantification of the results into a numerical and comparable output, which is often appreciated in order to assess the overall goodness of the design. Finally, (6) they are light-weight and cost-effective, and (6) the developers and practitioners, as well as most users, are familiar with the approach and Likert-scales. These are the main reasons why we decided to develop subjective statements instead of, for example, interview questions.

In formulation of the statements, we avoided the use of superlatives and strong expressions, such as "very" or "most". Therefore, they allow the evaluator to use the entire scale of one to five or one to seven more readily. This makes the different statements quantitatively more comparable: that is, the extreme and middle values of the scale denote more or less the same strength of opinion with different statements. Furthermore, the statements have been formulated so that they do not necessarily need a point of reference (e.g. another type of service or technology) against which to evaluate the service in question.

As for limitations, the statements have not been validated and their usefulness and relevance as measures have not been assessed through using them in practical evaluation cases. Therefore, these examples of evaluation statements are not yet meant as a comprehensive metric or an all-embracing set of measures. As a tentative set of measures, further iteration of the statements is most probably required in order to address possible sources of ambiguity or misapprehensions, as well as to eliminate too overlapping statements and thus decrease the total amount of measures.

Furthermore, many of the presented experiences become possible and relevant only after the service under development has reached a certain degree of readiness or certain features have been implemented. It is important to delimit the extent of measurable experiences consequently: is there merely a description of the to-be-implemented concept, is there a prototype that provides a tentative experience of the interaction, to what extent the service and its functionalities have been implemented, how widely a fully-functional service has been adopted and how much content there exists in the service, etc. Therefore, the selection of statements has to be carefully considered when applying them to evaluation of prototypes or other low-maturity demonstrators.

As the statements were formulated so that agreeing with them indicates a good user experience, most of the statements are positively stated. However, ideally a set of statements would also include negatively stated ones to counterbalance the effect of evaluators giving repetitive numerical assessment without thinking over each matter individually. An example of a negatively stated statement about intuitiveness could be “I would rather receive the information otherwise than with augmentations”. This kind of balancing remains to be addressed in future work.

Considering quantitative analysis, as the measures and their inter-correlations have not yet been statistically evaluated, numerical aggregates (e.g. sums of values, index of goodness) should not be computed. Instead, the seven-step Likert is often treated as an interval-level scale, which allows arithmetic mean to be calculated of each individual measure. The frequency scale as an ordinal scale allows calculating only medians and modes, and requires non-parametric tests for more advanced statistical analysis.

With regard to number of users and sampling, the general research principles apply also here: the more evaluators representing the target group, the more reliable and extensive quantitative analysis can be conducted, resulting in more trustworthy conclusions. However, as this is a general research design issue, we leave it out of the scope of this paper and suggest leaning on general methodology guides, such as [55].

5 Final Thoughts

Basing on our prior research, this chapter presented a synthesis of the characteristics of desirable user experience that are expected to be relevant in future mobile AR services. This qualitative framework serves as design inspiration and a basis in assessing UX aspects of such services. Furthermore, it was concretized into subjective UX evaluation measures that we propose to be utilized in evaluation of future services that demonstrate elements of mobile AR. Despite being based on user studies in the context of everyday activities and consumer applications, they can exhibit usefulness and relevance in other domains of AR as well (e.g. industrial AR applications). Having a common framework of user experience and utilizing the proposed set of measures should make it possible for researchers in the field to design more appropriate UX evaluations, and compare and learn from each other’s results. This will facilitate inventing and developing more competitive mobile services that truly take advantage of the strengths of augmented reality.

Nevertheless, using the statements is not a panacea for ensuring user-centeredness or resolving the UX issues in a service. Subjective statements as such are not necessarily sufficient to measure all the aspects of UX. They provide only a numerical evaluation of the particular operationalized aspects. A simple analysis of the responses can indicate some critical aspects that the users do or do not appreciate. However, they do not pinpoint in detail where the problems specifically lie or how to remedy them. Such understanding requires more in-depth qualitative methods like interviewing with regard to that aspect. Furthermore, to holistically evaluate future mobile AR services, the "traditional" usability aspects need also to be taken into account (considering e.g. accuracy of alignment, user control, ease-of-use of the interface, service terminology).

As the operationalized statements are tentative, validating and formulating them into valid and reliable measures requires further research and multiple evaluation cases where they are utilized. Only by applying them in real evaluation cases can their usefulness, relevance and reliability be judged. Similarly, the causalities in experiences originating from specific service components require further research with an experimental approach. Future research steps could result in, for example, a more condensed list of statements without semantically overlapping items. Naturally, the more holistic, but at the same time compact and easy-to-use the set of measures is, the more effectively it can serve in its purpose. Our ultimate goal is to develop the statements into a methodological "toolbox" that can be easily used by anyone, without specific understanding about user experience, and with any kind of consumer-targeted mobile AR services. The next step towards this goal is to utilize the statements in a few evaluation cases, in which, first, the understandability and practical relevance of the measures is assessed and, second, the statistical interactions between the measures are analyzed.

In addition to developing and validating the measures, our future work aims to concretize the presented characteristics of experience and other UX findings into practical design guidelines. This would help ensuring that the user experience aspects are considered already in the early concept design phases, thus facilitating an experience-driven design approach and a high-quality user experience.

References

1. AttrakDiff. Online: www.attrakdiff.com, accessed: December 2011
2. Bach C, Scapin D (2004) Obstacles and perspectives for evaluating mixed reality systems usability. IUI-CADUI Workshop on Exploring the Design and Engineering of Mixed Reality Systems (MIXER)
3. Billinghurst M, Kato H (2002) Collaborative augmented reality. Communications of the ACM 45(7):64–70
4. Brooke J (1996) SUS: a "quick and dirty" usability scale. In Jordan PW, Thomas B, Weerdmeester BA, McClelland AL (ed) Usability Evaluation in Industry. Taylor and Francis
5. Buccini M, Padovani S (2007) Typology of experiences. Proceedings of DPPI 2007. ACM, pp 495–504
6. Carter S, Mankoff J (2004) Challenges for ubicomp evaluation. EECS Technical Reports, CSD-04-1331. University of California, Berkeley
7. Curtis D, Mizell D, Gruenbaum P, Janin A (1999) Several devils in the details: Making an AR application work in the airplane factory. Proceedings of IWAR'98. AK Peters, Massachusetts, 48p

8. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3):319–339
9. Desmet PMA (2002) Designing emotions. Dissertation, Delft University of Technology
10. Desmet P, Hekkert P (2007) Framework of product experience. International Journal of Design 1(1):57–66
11. Desmet PMA, Overbeeke CJ, Tax SJET (2001) Designing products with added emotional value; Development and application of an approach for research through design. The Design Journal 4(1):32–47
12. Dünser A, Grasset R, Billinghurst M (2008) A survey of evaluation techniques used in augmented reality studies. Proceedings of ACM SIGGRAPH 2008. ACM Press, New York
13. Ermi L, Mäyrä F (2007) Fundamental components of the gameplay experience: Analysing immersion. In: de Castell S and Jenson J (ed) Changing views: worlds in play - International perspectives on digital games research. Peter Lang, New York, pp 37–53
14. Fields B, Amaldi P, Wong W, Gill S (2007) Editorial: In-use, in-situ: Extending field research methods. International Journal of Human Computer Interaction 22(1):1–6
15. Fogg BJ (2003) Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann publishers, San Francisco, 283p
16. Gabbard JL, Swan JE (2008) Usability engineering for augmented reality: employing user-based studies to inform design. IEEE Transactions on visualization and computer graphics 14(3):513–524
17. Gabbard JL, Hix D, Swan II JE (1999) User-centered design and evaluation of virtual environments. IEEE Computer graphics and applications 19(6):51–59
18. Gandy M et al. (2010) Experience with an AR evaluation test bed: presence, performance, and physiological measurement. Proceedings of ISMAR 2010. IEEE, pp 127–136
19. Google Goggles. Online: http://www.google.com/mobile/goggles/, accessed: December 2011
20. Hart S, Staveland L (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock P, Meshkati N (Ed), Human mental workload. North Holland, Amsterdam, pp 139–183
21. Hassenzahl M (2003) The thing and I: Understanding the relationship between user and product. In: Blythe M, Monk AF, Overbeeke K, Wright P (ed) Funology: From Usability to Enjoyment. Kluwer Academic, pp 31–42
22. Hassenzahl M (2004) The interplay of beauty, goodness, and usability in interactive products. Hum.-Comput. Interact. 19(4):319–349
23. Hassenzahl M, Ullrich D (2007) To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. Interacting with Computers 19(4):429–437
24. Hsu CL, Lu HP, Hsu HH (2007) Adoption of the mobile Internet: an empirical study of multimedia message service (MMS). Omega, 35(6):715–726
25. Hollerer T, Feiner S (2004) Mobile augmented reality. In: Karimi H and Hammad A (eds) Telegeoinformatics: Location-Based Computing and Services. Taylor & Francis Books Ltd, London.
26. ISO FDIS 9241–210:2009 (2009) Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International organization for standardization, ISO
27. Jordan P (2002) Designing pleasurable products. CRC Press
28. Junaio. Online: www.junaio.com, accessed: December 2011
29. Kelley JF (1984) An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Office Information Systems 2(1):26–41
30. Korhonen H, Montola M, Arrasvuori J (2009) Understanding playful user experience through digital games. Proceedings of DPPI'09, pp 274–285
31. Kruijff E, Swan E, Feiner S (2010) Perceptual issues in augmented reality revisited. Proceedings of ISMAR 2010. IEEE, pp 3–12
32. Kujala S, Nurkka P (2009) Product symbolism in designing for user experience. Proceedings of DPPI'09. Springer, pp 176–186

33. Lang PJ (1980) Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski JB, Johnson JH, Williams TA (Ed), Technology in mental health care delivery systems. Ablex, Norwood, pp 119–137
34. Law EL, Schaik P (2010) Modelling user experience – An agenda for research and practice. Interacting with computers, 22(5):313–322
35. Law EL, Roto V, Hassenzahl M, Vermeeren AP (2009) Understanding, scoping and defining user experience: A survey approach. Proceedings of CHI'09. ACM Press, pp 719–728
36. Layar. Online: www.layar.com, accessed: December 2011
37. Looser J, Billinghurst M, Grasset R, Cockburn A (2007) An evaluation of virtual lenses for object selection in augmented reality. Proceedings of GRAPHITE'07. ACM Press, pp 203–210
38. Mackay WE (1996) Augmenting Reality: A new paradigm for interacting with computers, La Recherche, Mar. 1996
39. McCarthy J, Wright P, Wallace J, Dearden A (2006) The experience of enchantment in human–computer interaction. Personal and Ubiquitous Computing 10(6):369–378. doi: 10.1007/s00779-005-0055-2
40. Nielsen J (1993) Usability engineering. Morgan Kaufmann, San Diego, 362 p
41. Nilsson S (2010) Augmentation in the wild: user centered development and evaluation of augmented reality applications. Dissertation, Linköping university
42. Olsson T (2009) Understanding collective content: Purposes, characteristics and collaborative practices. Proceedings of Communities and Technologies 2009. ACM Press, pp 21–30
43. Olsson T, Salo M (2011) Online user survey on current mobile augmented reality applications. Proceedings of ISMAR 2011. IEEE, pp 75–84
44. Olsson T, Väänänen-Vainio-Mattila K (2011) Expected user experience with mobile augmented reality services. Workshop of Mobile Augmented Reality, MobileHCI 2011. ACM Press, New York
45. Olsson T, Ihamäki P, Lagerstam E, Ventä-Olkkonen L, Väänänen-Vainio-Mattila K (2009) User expectations for mobile mixed reality services. Proceedings of ECCE'09. ACM Press, 177–184
46. Olsson T, Kärkkäinen T, Ventä-Olkkonen L, Lagerstam E (2012) User evaluation of mobile augmented reality scenarios. Forthcoming in Journal of ambient intelligence and smart environments, IOS Press
47. Olsson T, Lagerstam E, Kärkkäinen T, Väänänen-Vainio-Mattila K (2011) Expected user experience of mobile augmented reality services: A user study in the context of shopping centers. Journal of Personal and Ubiquitous Computing, Springer, DOI: 10.1007/s00779-011-0494-x
48. Roto V, Law EL, Vermeeren AP, Hoonhout J (eds) (2010) User experience white paper: Results from Dagstuhl seminar on demarcating user experience. Available at: http://www.allaboutux.org/files/UX-WhitePaper.pdf
49. Scholtz J, Consolvo S (2004) Towards a discipline for evaluating ubiquitous computing applications. Report from National Institute of Standards and Technology, IRS-TR-04-004.
50. Stanney K (1995) Realizing the full potential of virtual reality: Human factors issues that could stand in the way. Proceedings of VRAIS'95. IEEE, pp 28–33
51. Stevens SS (1946) On the theory of scales of measurement. Science 103(2684): 677–680
52. Swan JE, Gabbard JL (2005) Survey of user-based experimentation in augmented reality. Proceedings of 1st International Conference on Virtual Reality
53. Theofanos M, Scholtz J (2005) A framework for evaluation of ubicomp applications. Workshop on Social Implications of Ubiquitous Applications, CHI'05. ACM Press, New York
54. Thompson ER (2007) Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). Journal of Cross-Cultural Psychology 38(2):227–242
55. Trochim W, Donnelly JP (2006) The research methods knowledge base. Atomic Dog, 3rd edition, 361p

56. Vaittinen T, Kärkkäinen T, Olsson T (2010) A diary study on annotating locations with mixed reality information. Proceedings of MUM 2010, Article no. 21
57. Vermeeren AP, Law EL, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K (2010) User experience evaluation methods: current state and development needs. Proceedings of NordiCHI'10. ACM Press, New York, pp 521–530
58. Väänänen-Vainio-Mattila K, Roto V, Hassenzahl M (2008) Towards practical user experience evaluation methods. Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM), pp 19–22
59. Wagner D, Schmalstieg D (2009) Making augmented reality practical on mobile phones, Part 1. IEEE Computer Graphics and Applications 29(3):12–15
60. Wellner P, Mackay W, Gold R (1993) Back to the real world. Communications of the ACM 36(7):24–26

Chapter 10
Enhancing User Role in Augmented Reality Interactive Simulations

Pier Paolo Valentini

1 Introduction

Scientific literature reports an increasing interest for the development of applications of augmented reality (AR) in many different fields [1–3]. The AR has been used in entertainment [4–8], education [9–12], medicine [13–15], military field [16, 17], implant and components maintenance [18, 19], robotics [20], engineering [21–28] and archeology [29, 30]. Some recent developments about mobile augmented reality applications have been discussed in Chaps. 6 and 7. The most of all these applications deals with the merging in the real world of objects, scenes and animations which have been modeled and simulated outside the system. It means that the user perceives a real scene augmented with pre-computed objects. For these reasons, his interaction with the augmented scene is often limited to visual and acoustic exploration.

In 1999 the International Standard Organization (ISO) provided a definition of an *interactive system* as: "An interactive system is a combination of hardware and software components that receive input from, and communicate output to, a human user in order to support his or her performance of a task". The recent improvements of both hardware and software performances fuelled the development of innovative methodologies in order to increase of the interaction between the user and the scene [31, 32]. The purpose of these enhancements is to change the user role from spectator to actor. The main idea to achieve this objective is to use innovative approaches for going beyond a mere visual or acoustical experience of pre-computed contents, including the capability of real-time modifying and updating the contents of the scene and the two-ways interaction.

P.P. Valentini (✉)
Department of Industrial Engineering, University of Rome "Tor Vergata",
via del Politecnico 1, 00133 Rome, Italy
e-mail: valentini@ing.uniroma2.it

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_10,

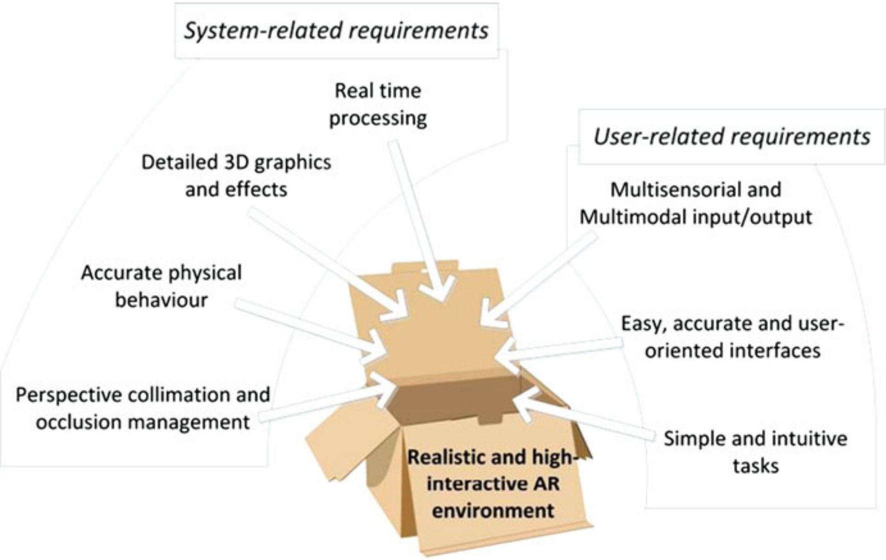


Fig. 10.1 System-related and user-related requirements for the implementation of an high interactive and realistic AR environment

Generally speaking, an augmented environment can be implemented with different levels of interaction. Interaction concerns with users tasks that can be classified according to Gabbard [33] and Esposito [34] that organized them in navigation, object selection and object manipulation, modification and querying.

An high interactive and realistic augmented reality environment needs both system-related and user-related requirements to be successfully implemented (see Fig. 10.1) [35, 36]. System-related requirements are concerned with the architecture and implementation of the processing engine (hardware and software). The user-related aspects are concerned with the way the user interact (input and output) with the environment, taking into account cognitive aspects as discussed in the Chap. 5.

In a first and basic level of interaction, the user can only reviewed pre-computed virtual contents. Following this approach, animations and graphics are prepared outside from the system and they are projected to the user in the right moment and context. For the superimposed geometries, the level of details of the augmented scene has to be very realistic and the registration between real world and virtual contents has to be accurate in order to give the illusion of a unique real world. On the other hands, textual information has to be clearly visible and readable in the scene.

An intermediate level of interaction is concerned with the possibility of relating to virtual objects and information in the scene. With this type of integration, the user is active in the scene and can change the augmented contents by picking, pushing and moving objects and controlling the provided information. The interaction is carried out with advanced input/output devices involving different sensorial channels (sight, hear, touch, etc.) in an integrated way [37]. In particular, in order to interacts with digital information through the physical environment, the system can be provided of tangible user interfaces (TUIs) [38–42]. The TUIs are more suitable than the graphic user interfaces (GUIs) to work as communicators between the user and the augmented system because they are based on physical entities that can be grabbed, moved, pushed, etc.

With an higher level of interaction, the user can modify the contents of the scene and the virtual objects in the environment behave according to realistic physics laws (dynamic simulation, deformation, etc.). In general, the interaction can be provided by specific TUIs whose design and features are suitable for an enhanced communication with the scene.

The highest level of interaction includes the reaction of the virtual objects on the user (action–reaction, force feedback, etc.) as well. In this case, the TUIs have to be able to produce sensorial feedback and their characteristic is a two-way communication (scene ↔ user).

The design and optimization of the tangible user interfaces involve an accurate attention to the related human factors and communication requirements. Human factors are concerned with anything that affects the performance of system operators whether hardware, software, or liveware [43, 44]. They include the study and application of principles of ergonomic design to equipment and operating procedures and in the scientific selection and training of operators.

On the one hand, the interfaces have to be able to track the user in the scene with adequate precision and robustness and acquire his intent. On the other hand, they have to be light and small enough to be minimally invasive and be used without difficulties in order to achieve the best possible performance within machine design limitations. A user interface designer is challenged by choosing the most appropriate way of acquiring and presenting information by adequate media and modalities.

With reference to Fig. 10.2, the standard implementation of an interactive augmented reality can be described as follows. First of all, an image stream of the real world has to be acquired. One or two RGB camera(s) are used for acquiring a mono or stereo vision of the scene, respectively. Then, the user is able to interact with communication devices in order to participate in the scene. This role is played by tangible or multimodal interfaces which can be different depending on the type of implementation. Their design has to take into account the specific human factors and simulated tasks. A device (external monitor or head mounted display) has to be also present in order to ensure the portable projection to the user of the augmented scene.

The pieces of information coming from the video acquisition and the user interface have to be processed in order to estimate the perspective transformation of the camera point of view, interpret the intent of the user and compute all the virtual objects to be added. At the end of the computation an augmented video stream is rendered taking into account also special effects as occlusions, congruent illumination, etc. and projected back to the user. In the applications which provide the two-way interaction with the user, a feedback has to be also sent back to the user via the interface devices.

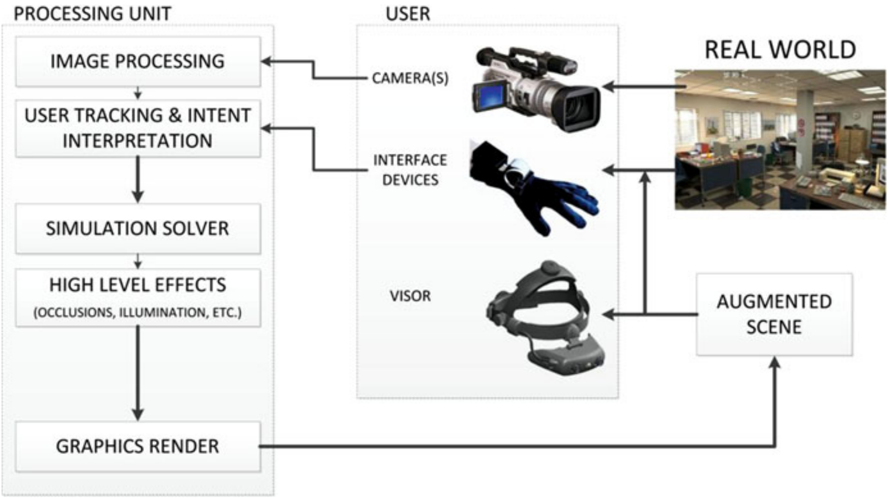


Fig. 10.2 Processing scheme for a generic augmented reality interactive implementation

Starting from this schematic representation, three main requirements for achieving an interactive augmented reality simulation can be considered.

The first one is the *realism*. The real scene and the virtual objects have to be properly integrated giving the illusion to the user to a mere real world. This means that the graphics of the objects and their illumination has to be detailed. Moreover, the AR system has to be able to manage occlusions between real objects and virtual ones, in order to avoid the perception of simple superimposition. A scene which is not able to include occlusion may produce unrealistic feeling to the user and vanish all the efforts toward the building of a realistic environment. The physical behavior of virtual objects is another important feature to achieve realism in the scene. For this purpose, the movement of all the virtual objects has to be consistent to physical laws. It means that objects cannot interpenetrate but collide, are subjected to gravity force, etc. From the user's point of view, all these features are important to increase the feeling at ease in the scene and perceive a familiar and harmonized world as an unique real environment. The presence of well-designed TUIs may surely improve this feeling.

The second requirement for an interactive AR simulation is about the *real-time processing* of the scene. It means that all the computations (image processing, user tracking, intent interpretation, physical behavior and scene updating and rendering) have to be performed in real-time (or better synchronously to the scene acquisition) in order to achieve fluidity and enhancing the illusion of a natural scene. This specification requires the development of specific simulation strategies and the use of dedicated solver and processor. The most challenging implementation is about the simulation of physical behavior of the environment including gravity, impenetrability, contact and impact dynamics, etc. The current level of hardware performances allows the use of standard computer architectures for achieving this result.

The third requirement for an interactive AR simulation is about the implementation of adequate *interaction devices and methodologies*. In order to interact with the scene, the user has to communicate to the AR system. From this point of view, the TUIs can be considered as a fundamental requirement in order to implement an interactive augmented reality environment. On the other hand, the only TUIs are not sufficient to ensure interactivity, but specific methodologies for interpreting the user's intent and his relationship with the augmented environment have to be studied and implemented. These devices and methodologies have to be integrated to the computational routines for simulating a congruent behavior of the overall system. Scientific literature reports several contributions dealing with possible methodologies achieving this interaction which are discussed in the next section of the chapter.

The chapter is organized as follows. In the first part a brief overview of the state-of-the-art methodologies for achieving interactive simulation in augmented reality environment is presented, focusing to the system-related and user-related aspects. In a second part the emerging concept of natural interface in augmented reality is introduced and discussed. In the last part of the chapter some details of implementation and examples are presented and discussed.

2 User Interaction in Augmented Reality Scenarios

Among the requirements to achieve an interactive AR simulation, the most important user-related issue is concerned with the development of devices and methodologies for achieving a robust, simple and comprehensive interface. In the very low level augmented reality implementations, the interaction is limited to graphics and (in some cases) to acoustics outputs. In basic implementations, the interaction is extended by using the mouse and the keyboard as in a standard computer application. This arrangement has the advantage that the user is already familiar to the devices and he does not need training or particular skills to use the interfaces. On the other hand, the interaction is limited to very simple operations (2D or 3D pointing and clicking).

In the intermediate-level implementations the communication between the user and the scene can be achieved using patterned makers. They are used for both computing the perspective transformation between the camera and the real world and for transferring information from the user to the scene. They are considered as communicators and the interaction is based on the computation of the their relative position and attitude with respect to the camera and the other markers reference frames. They can be rigidly mounted on real objects in order to build tangible user interfaces with 6° of freedom (three translations and three rotations). Following this approach, the advantage is that the image processing for marker detection in the acquired video stream is performed only once but it is useful for both perspective collimation and user tracking. The disadvantages of these methodologies are mainly two. First of all, their precision is low, because the position and attitude of the markers are computed by the processing of standard resolution images using segmentation and

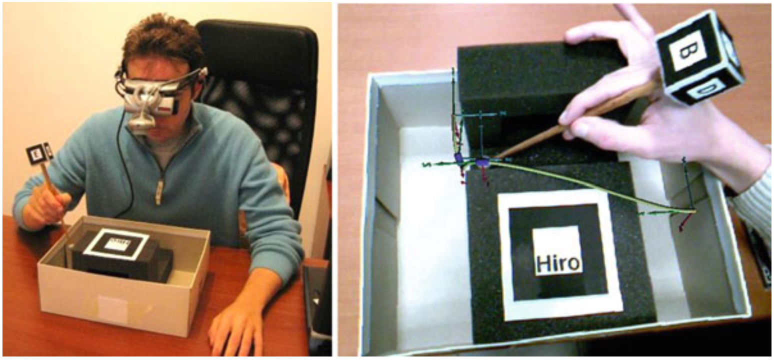


Fig. 10.3 Marker-based tracking for implementing an interactive AR procedure for supporting cable harnessing

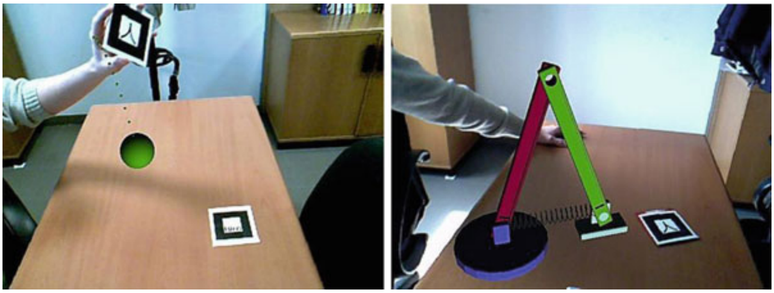


Fig. 10.4 Marker-based tracking for implementing interactive AR simulation of dynamic systems: launching a bouncing ball (on the *left*) and moving a slider-crank mechanism with a spring damper element

correlation algorithms. Secondly, in order to be tracked, the markers need to be always visible to the camera and this limits the capture volume and suffers occlusion phenomena. Figure 10.3 shows and example of this tracking methodology used for implementing an interactive procedure for supporting cable harnessing in augmented reality [24]. In this application five patterned markers are placed on a cube at the end of a stick in order to implement a traceable pen. The pen is used to sketch and modify the path of a virtual cable interactively routed in the scene. Although only one marker is sufficient to track the position of a rigid body in space (requiring 6° of freedom), redundant markers can be used to ensure a continuous visibility and more accurate tracking.

Marker-based interaction has been used also in interactive dynamic simulations. Figure 10.4 reports two examples of this implementation. In this case the markers are directly grabbed by the user in order to interactively set and modify the initial conditions of the motion simulations [45] of a collection of rigid bodies.

Fig. 10.5 Magnetic device-based tracking for implementing interactive AR geometric modeling environment

Other methodologies introduced the use of different sensors for tracking the position of the user in the scene and interpreting his intent. Some of them are concerned with the use of optical tracking systems [46]. These implementations make often use of reflective markers (usually spheres for improving visibility) or pattern of markers whose position in the scene can be recognized by photogrammetric analysis using multiple cameras which can be different from those used for real world acquisition and perspective collimation. Since the reflective markers can be rigidly placed on almost every object, they can be used to implement tangible user interfaces or for simple user main body parts tracking. These methodologies can be more precise than the previous ones because the optical tracking is performed using a dedicated system. On the other hand, the presence of several markers may be uncomfortable for the user and, as for the other optical systems, their precision is affected by the resolution of the cameras and highly suffers occlusion phenomena.

Other acquisition methodologies are based on the use of magnetic trackers [47]. In the common embodiments, these devices are comprised of an emitter and a receiver. The emitter generates a magnetic field which is captured by the receiver. The changing of the acquired signal is converted to information about the position and attitude of the receiver. Due to its small size, the receiver can be easy put on by the user or attached to a graspable stick in order to perform user tracking or build tangible user interfaces. In general, magnetic trackers are more precise than the optical ones, but their performance is tremendously influenced by electromagnetic perturbations caused by metallic parts in the scene and the capture volume is dependent on the strength of the magnetic field generated by the emitter. Figure 10.5 shows an example of the use of a magnetic tracking system for implementing a augmented reality system for interactive sketching and modeling [22]. In this application the magnetic sensor (Flock of Birds by Acension) is placed on at the end of a stick in order to implement a traceable pen and help the user in interactive operations.

In order to achieve more precise and robust tracking, mechanical (or mechatronic) devices can be used [48]. They commonly use a multi degree-of-freedom

Fig. 10.6 Mechanical device-based sketching for implementing interactive AR reverse engineering modeling tool

Fig. 10.7 Mechanical device-based tracking for implementing interactive AR engineering simulation of motion: ten pendula simulation (on the *left*) and flexible slender beam (on the *right*)

linkage to compute the position in the space of a end-effector which can be grabbed by the user. By this way the devices can be directly considered as the tangible interface. Their precision is high (<0.2 mm), they are not affected by occlusions and perturbations, but their capture volume is still limited by the dimension of the linkage. Thanks to all these advantages, they are suitable for accurate interaction involving technical and engineering aspects (interactive sketching, reverse engineering, measurement, etc.).

Figure 10.6 shows an example of a mechanical tracker (Microscribe GX2 by RevWare) used to perform an interactive reverse engineering tool in augmented reality.

Figure 10.7 shows and another example dealing with two simulations of movement performed using the same mechanical tracker for defining the boundary conditions for both rigid [48] and deformable bodies [49] simulation.

Fig. 10.8 An example of integration between a pattern-based tracking system and a data glove for hand position and gesture acquisition in an augmented reality environment

In order to increase the sensorial feedback, haptic output can be added to the tangible user interfaces. Haptics comes from a Greek word $\eta\alpha\pi\tau\varepsilon\sigma\tau\eta\alpha\iota$ meaning "grasping" or "the science of touch". In recent years, its meaning extended to the scientific study for applying tactile and force feedback sensations of humans into the computer-generated world.

All the above mentioned tracking systems can be used in addition to other specific devices in order to enhance the communication properties between the user and the augmented scene. One of the possible implementation is concerned with the use of data gloves. These wearable devices can be used for acquiring the hand gesture of the user, interpreting his intent of indexing, picking, etc. Since they provide only gesture assessment, they have to be used in addition to other tracking devices, as optical or magnetic systems. They have the advantage to enhance the possibility of interaction, interpreting an extended range of user's intent.

Figure 10.8 shows an example of integration between a pattern-based tracking system and a data glove. The system is able to acquire user's hand position and gesture and has been used for implementing a virtual assembly procedure in an augmented reality environment [50].

The described methodologies for enhancing the interaction between the user and the augmented scene can be compared in terms of system performance and user-related factors.

Concerning with the performance the main characteristic of the solutions are the cost, the precision, the capture volume and the suitable applications. Table 10.1

Table 10.1 Performance comparison among tracking devices for interactive augmented reality implementations

Interaction method	Capture volume	Precision	Cost	Preferred applications
Visual/acoustic only	n.a.	n.a.	Low	Entertainment; augmented navigation; augmented maintenance; military applications
Mouse and keyboard	n.a.	Very low	Low	Entertainment; gaming; augmented navigation
Optical patterned markers	Limited, suffer occlusion	Low	Very low	Design or model review; collaborative design; simple interactive simulation; development of simple tangible user interface for virtual prototyping
Optical reflective marker-based systems	Limited, suffer occlusion	Medium–low	Medium	Interactive and immersive simulations; interactive exploration of wide and complex scenarios; interactive sketching
Magnetic systems	Limited, suffer metallic objects	Medium	Medium	Interactive sketching and modeling; developing of efficient tangible user interface for virtual prototyping
Mechanical systems	Limited	High	Medium–high	Precise interactive sketching and modeling. Accurate interactive simulations, engineering applications
Mechanical with force feedback systems	Very limited	High	Medium–high	Precise interactive sketching and modeling. Accurate interactive simulations, engineering and surgery applications
Optical or magnetic + local sensor interfaces	Limited	Low to medium	Medium–high	Interactive simulation including hand gesture

Table 10.2 User related factors comparison among tracking devices for interactive augmented reality implementations

Interaction method	Wearability	User friendliness	Training	Invasiveness in the scene
Visual/acoustic only	Yes	Good	Very low; training required for stereoscopic projection	Low
Mouse and keyboard	No	Very good	No	Very low
Optical patterned markers	Yes	Good	No	Low
Optical reflective marker-based systems	Yes	Good	Low	Medium
Magnetic systems	Yes	Good	Low	Medium–low
Mechanical systems	No	Good	Medium	Medium
Mechanical with force feedback systems	No	Discrete	Medium–high	Medium
Optical or magnetic + local sensor interfaces	Yes	Good	Medium	Medium–low

reports a comparison among the different typologies of the interaction solutions. It can be noted that the optical system based on marker recognition are suitable for the developing of basic interactive augmented scenarios, especially for entertainment, gaming, design reviews, conceptual technical applications requiring low accuracy. The mechanical devices are very suitable for accurate interaction and for the implementation of precise engineering and surgery simulations. Due to their architecture, they are also suitable for the implementation of force-feedback sensors and immersive simulation.

Concerning with the user-related factors the main characteristic of the solutions are the wearability, the user friendliness, the necessity of dedicated training and the invasiveness in the scene. Table 10.2 reports a comparison among the different typologies of the interaction solutions. All the presented devices can be arranged to be worn by the user, except the mechanical trackers. In general, small markers (simple sphere of patterned ones) can be easily managed by the user and they are less invasive in the scene. On the other side, they need to be always visible to the cameras and require a little training to be properly used. Both optical, magnetic and mechanical systems allow the arrangement of tangible user interfaces similar to a pen, which enhances the friendliness and reveal to be familiar to the user.

3 The Concept of Natural Interface

According to the considerations in the review presented in the previous section, two different aspects have to be underlined. First of all, in order to interact with the scene, the environment has to include interfaces which are implemented by using devices and methodologies for tracking user’s position, interpreting his intent and transferring information to the simulation engine. On the other hand, the development of such interfaces may involve the use of complex and bulky devices. In too many case, they miss the point to produce a realistic scene because they are considered external, unrealistic and cumbersome by the user.

In order to overcome these problems, the idea is to avoid the use of specific interface devices and try to track the user and interpret his intent just observing the scene as it happens in real life in interpersonal communication. By this way, the user is the interface or better he uses a natural interface which is his body posture and attitude.

3.1 Implementation Details

As introduced in the first part of the chapter, most of the augmented reality applications performs the acquisition of the real world using a single camera or, for stereoscopic projection, two cameras. The role of these devices is to produce one or two RGB image(s) that can be used for both image processing and for the definition of the background image of the final augmented scene projection.

In order to implement the concept of the natural interface, tracking the user in the scene and interpreting his intent, simple cameras are not sufficient because they produce two dimensional images only. In order to have continuous three dimensional information about the acquired scene, a compound of an RGB camera, an IR projector and a IR depth camera can be used. For the specific purposes of the study, the author has tested the Microsoft Kinect Sensor which contains such arrangement in a compact bundle. The use of the Kinect Sensor allows the synchronized acquisition of an RGB image and a depth map of the same real scene. The two streams are always collimated by the fixed location in the bundle. An RGB image is a data structure containing color information of each acquired point (pixel). A depth map is a data structure containing the distance from the sensor of each pixel along a direction perpendicular to the image plane. In order to acquire and process the data coming from the Kinect sensor the Prime Sense drivers has been used. They are suitable for C++ programming language implementation and can be freely downloaded at https://github.com/PrimeSense/Sensor.

According to its architecture, there are two data streams coming from the Kinect Sensor that have to be managed. The first one, as in a traditional augmented reality application, is the RGB video stream. Each RGB frame can be processed in order to recognize the presence of patterned markers in the scene and to compute the perspective transformations between the camera and each marker. The processing of the depth map stream allows to include several enhancements useful for increasing the realism of the augmented scene and the lever of interaction. Two are the main processes involving the depth map. The first one is concerned with the computation of the environmental mesh which is a geometrical representation of the real world three-dimensional geometry. Starting from the knowledge of the 3D coordinates of each point observed by the depth camera, it is possible to build a structured polygonal mesh describing the geometry of the surrounding environment. This mesh can be textured with the color information coming from the RGB camera in order to achieve a complete 3D reconstruction of the augmented environment.

The second important use of the depth camera stream is the possibility of tracking the users in the scene and implementing the natural interface concept in a very smart way as described in the following section.

3.2 User Tracking

The processing of the depth map information allows the real-time tracking of the user. For the tested implementation involving the Kinect Sensor, the tracking is implemented using the OpenNI programming library freely downloadable at https://github.com/OpenNI/OpenNI. The OpenNI is a collection of C++ routines for direct accessing and processing data from Kinect Sensor and includes numerical procedures for achieving a robust and precise tracking of user's body main landmarks. Although the exact implementation of these algorithms is not open access, some useful information can be extracted from the related patent application [50, 51].

According to this approach, the tracking of the user is performed by processing the depth map in order to recognize the spatial position of the user's main body joints in the real scene. The OpenNi algorithm allows the real time recognition and tracking of the following 16 joints (see Fig. 10.9):

- Center of the head
- Center of the neck
- Right and left shoulder joints
- Right and left elbow joints
- Center of right and left hand
- Center of the chest
- Center of the abdomen
- Right and left hip joints
- Right and left knee joints
- Center of the right and left feet

The algorithm allows the tracking of several users at the same time.

The spatial position of the above mentioned 16 body joints are sufficient to interpret the pose of a human body. By this way, the intent of the user can be interpreted by comparing his pose to a collection of preset posture (Fig. 10.10). This assessment is very fast because requires the comparison of only a small set of 3D points.

The recognition of the body pose can be useful for activating commands and updating the scene contents.

The most important body joints to be tracker are the hands because they represent the main human interface to the physical (and virtual) world. According to the scientific literature and practical evidence almost all interaction methodologies are

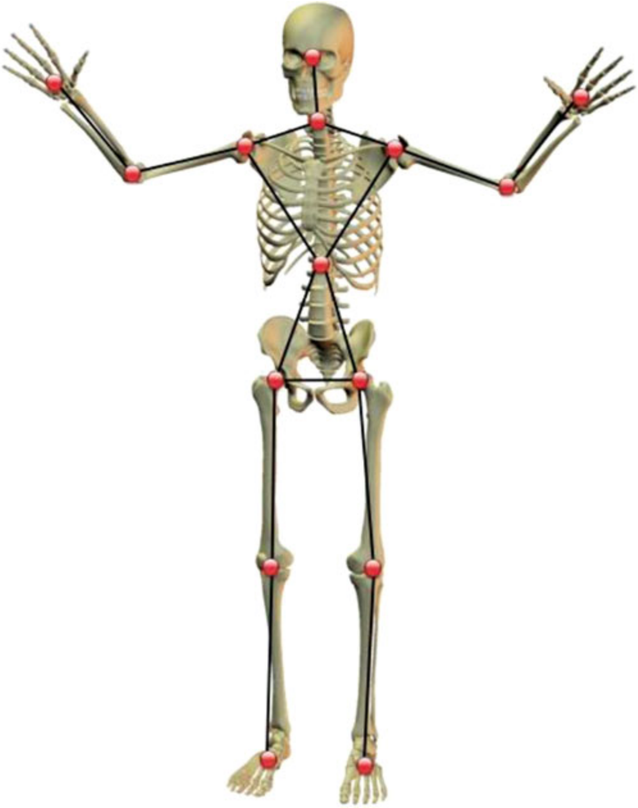


Fig. 10.9 Traceable body landmarks using numerical libraries

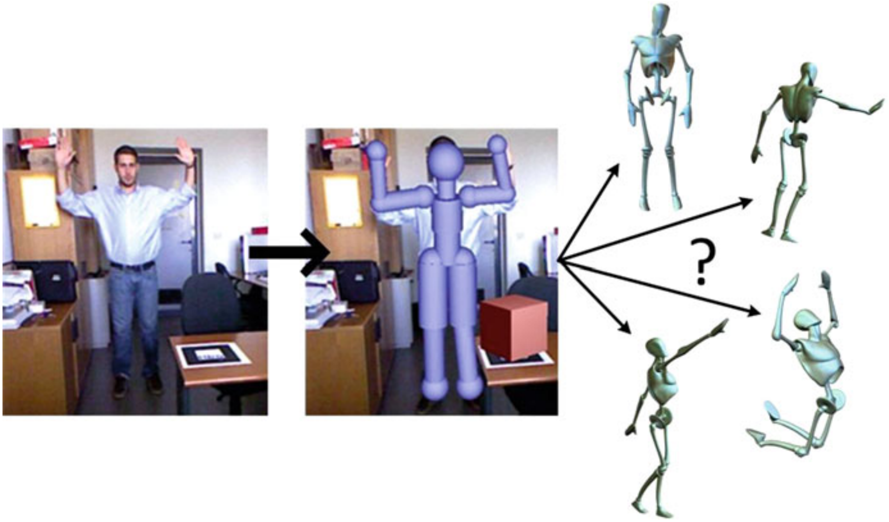


Fig. 10.10 Tracking user body and recognizing his pose

based on the tracking of the user's hands. In fact, indexing, picking, grabbing and pushing are all activities that involve the use of one or both hand. The recognition and the tracking of their position in the scene is therefore crucial.

3.3 Realism and Occlusion

As discussed in the introduction, the presence of correct occlusions between real and virtual objects in the augmented environment is a very important topic for enhancing the realism of the scene. Absent or wrong occlusion management may mine the overall quality of the environment and the user may perceive an unreal and disturbing environment.

A correct interpretation of occlusions is one of the most challenging topics in augmented reality applications [52–56]. According to some authors, the correct occlusion management is one of the most important requirement for a realistic implementation. Unfortunately, dealing with occlusions is quite complicated.

Standard augmented reality implementations usually neglect occlusions between real and virtual objects and the acquired image from the real world is considered as a simple background texture on which virtual objects and information are superimposed. On the other hand, occlusions involving only virtual objects can be easily computed by using the z-depth comparison which is a widely used technique in computer graphics. According to this approach, all the entities to be rendered are arranged in a list (called z-buffer) starting from the farthest up to the closest with respect to the point of view and along the direction normal to the image plane. Then they are rendered respecting the computed order and by this way the farther geometries are rendered after the nearer ones producing automatic occlusions.

The use of an IR projector/camera system to acquire a depth map of the environment can also help the managing of occlusion of real objects with respect to the virtual ones. As described in the subsection dealing with the implementation of the system, the Kinect Sensor produces a 3D geometric (polygonal) description of the acquired scene. Starting from this collection of data, it is possible to compute the 3D coordinates for each point of the real objects acquired by the sensor. Following a similar approach, the information coming from the depth camera can be processed in order to compute the z-coordinate (the distance from the image plane) for each pixel of the environmental mesh. The information about this mesh can be used for including the real objects in the scene in the z-depth comparison together with the other 3D virtual entities.

The processing of the environmental mesh is suitable for a real time computation and an example of application is reported in Fig. 10.11. It can be noticed that in the depicted augmented environment there are two virtual objects: a cube (placed on a real table) and a cylinder (in the right side, behind a real chair). With the proposed approach it is possible to compute the occlusion between the user body and the two objects and between the other real objects in the scene and the two virtual objects. It has to be underlined that the managing of the occlusions has some small imprecisions

Fig. 10.11 Occlusion management using depth map acquisition and computation

(the edges of the objects are often irregular) but the detail is sufficient to enhance the realism of the environment and avoid the perception an unreal and wrong (or even impossible) scene.

4 Simulating Physical Behavior

One of the other important requirements of an interactive environment is the achieving of accurate simulation of the objects behavior according to the actual physical laws. A correct mimic of the real world is very crucial for giving to the user the illusion of a consistent scene [57]. Moreover, a correct simulation respecting physical laws can be useful for implementing not only entertainment and gaming applications, but also technical and engineering scenarios in which the interpretation of the results can be used for improving product design and related performances [45].

Many augmented reality implementations make use of animation in order to transfer information to the user by using appealing moving graphics. These animations are studied and prepared outside from the running system and then are projected in the right place and context during exploration. However, performing simulations is different from simply animating. An animation concerns with the movement of the objects according to some specific predefined schemes and sequences. By this way, the animated movement can be convenient and didactical, but can be unreal and inconsistent. This solution may be an advantage for some

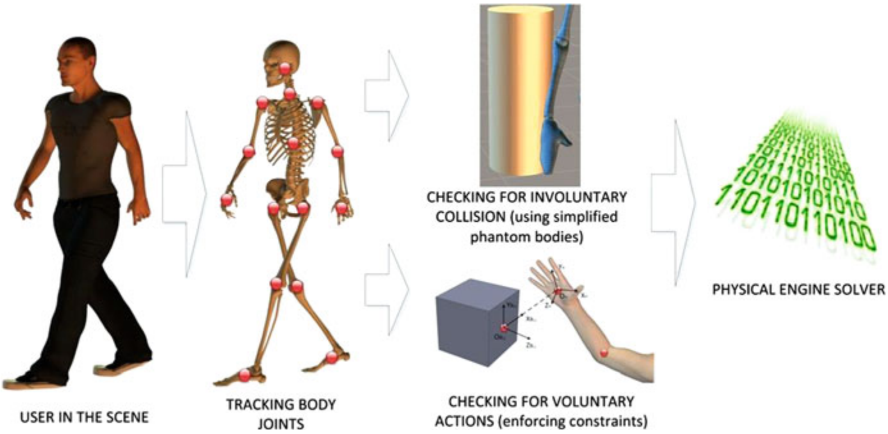


Fig. 10.12 Managing user presence for implementing accurate and realistic physical simulation

implementations, but may produce highly unreal scenarios. On the contrary, the simulation is the replication of a behavior which is consistent to the presence of physical law. A simulated virtual object behaves exactly (or quite exactly) in the same way as it would be real.

Introducing correct physical behavior in an high-interactive environment implies that the user can actively take part in the simulation. This participation has two different aspects: involuntary and voluntary ones. On the one hand, the presence of the user can affect the environment without being involved in a specific action. The collision between body limb and a virtual object is an example of this involuntary participation. On the other hand, the user can also voluntary affect the environment by picking, moving, throwing virtual objects. These two different kinds of interaction has to be taken into account in the simulation (see Fig. 10.12). And then, of course, all the virtual objects take part in the simulation and may interact among them.

The interaction between the user and the simulated environment requires the tracking of the body main joints and so it can be managed by the use of the Kinect Sensor as well. The positions of body joints are real-time computed during all the simulation. Phantom geometries (cylinders, cones and spheres) can be attached to these joints in order to check if collisions occur and manage involuntary contact between the user and the virtual objects in the scene (see Fig. 10.10). This approach can be implemented without any additional sensor to be attached to the user, respecting the purpose of a natural interface to the augmented environment.

The voluntary picking and moving of the objects can be also implemented starting from the tracking of the body main joints, but it required a more complicated approach. In particular, it is sufficient to track the position of the hands to check if the user is about to pick an object and then impose a grabbing constraint. Mathematical formulations and strategies to impose this kind of constraint goes beyond the scope of this chapter and an interested reader can find additional details in referenced papers [48, 58, 59].

The voluntary interaction can concern with both the definition of boundary conditions and initial parameters and the real-time control of the simulation.

The main difficulty in the practical implementation of the physical simulation of the virtual objects behavior is that all these computations have to be processed in real time. Three are the main problems of such processing. First of all, the scene may include many virtual objects whose dynamic behavior has to be computed, detecting and taking into account multiple collisions at the same time. Secondly, there are many different events in the simulation that require an updating of the topology of the system and a rearranging of the mathematical equations. This highly nonlinear behavior makes many standard integrators unsuitable for the purpose. Thirdly, the simulation has to be continuously performed for a long time needing robust integration and producing accurate and fluid results. For all these purposes, specific strategies have to be implemented in order to deduce and solve the equations of motion of the simulated system in a smart way.

Previous publications about the integration of dynamics simulation in augmented reality applications [48, 60–63] have revealed the interesting capability of the sequential impulse solvers. One of the most used is the *Bullet Physics Engine* which is an open source simulator with efficient real-time collision detection algorithms [64]. It is used in games, visual effects in movies and can be freely downloaded at http://bulletphysics.org/wordpress/.

The sequential impulse solver strategy allows a quick, stable and accurate simulation even in presence of all the above mentioned difficulties. According to this approach, the solution of the dynamics equations is based on the following steps. Firstly, the equations of motion are tentatively solved considering elastic and external forces but neglecting all the kinematic constraints and contact overlapping. This choice produces a solution that is only approximated. In a second step, a sequence of impulses are applied to each body in the collection in order to correct their velocities according to the limitation imposed by all the physical constraints. This second step is iterative but quite fast. It means that a series of impulse is applied to all the bodies until the constraint equations are fulfilled within a specific tolerance. Again, the detailed description of this methodology goes beyond the scope of the chapter and further details can be found in the referenced papers.

4.1 An Example of Implementation

Let us discuss some details of the implementation of the dynamic solver in the augmented reality interactive environment with an example. The scenario is about a simple interactive game in which a user can grab a ball and can throw it towards a stack of boxes. Both the ball and the boxes are virtual objects. The scene is acquired by a Kinect Sensor and processed by a DELL Precision M4400 laptop provided with an Intel Centrino 2 vPro (dual-core processor), 4 Gb RAM and a NVidia Quadro FX770M graphic card. No additional sensors have been used for tracking

Fig. 10.13 Six snapshots of the discussed example

the user in the scene. Figure 10.13 shows a sequence of six snapshots taken during the simulation.

The grabbing of the ball and the throwing are both voluntary actions. The user can freely move in the scene and the system tracks his body in real-time. When the user puts his hand near the ball the system recognizes the action which has to be confirmed by the user (snapshot A). From this moment, the movement of the ball is constrained to that of the hand. It means that the position, velocity and acceleration of the ball are dependent from those of the hand. Then, he can decide to remove the connection releasing the ball which is thrown with specific kinematic initial conditions as position and velocity (snapshot B). From this moment the ball moves subjected only to gravity till it hits the stack of boxes (snapshot C). All the collisions are managed by physics engine and the equations of motion are solved by means of the sequential impulse strategy (using Bullet Physics libraries) and continues throughout the simulation (snapshots D, E and F). Occlusions between real and virtual objects are detected during all the simulation (i.e. between the ball and the user's hand) and managed accordingly (see snapshots A, D and E). All the acquisition, computation and rendering are performed in real-time achieving a continuous and fluid output stream.

5 Discussion and Conclusion

Focusing on the user role is fundamental in order to develop augmented reality interactive environments. The user has to be considered the starting point for developing methodologies and devices. From this point of view, the interface design has to take into account both user factors as ergonomics, invasiveness, friendliness and system factors as accuracy, precision, robustness and reliability. This chapter focused on the emerging concept of natural interface. The idea is to avoid the use of additional sensors in order to implement the interface between the user and the environment. The user body is the interface as it happens in everyday communication and an intent can be expressed using posture and gesture. This approach requires the real-time tracking of the user body main joints and the interpretation of his pose. A possible solution involves the use of an infrared projector and an infrared camera which are able to produce a three dimensional depth map of the acquired scene. By the interpretation of this map, it is possible to recognize the user body limbs and track their joint positions. By this way, the tracking of the user is performed without any sensor to be mounted on the user body (like markers or magnetic transducers) and without the use of external devices.

The natural interface methodology can be integrated in complex systems including occlusion management, collision detection and physical behavior simulation. These complex scenarios enhance the realism of the scene and make the user perceived the augmented environment very close to the real one.

The discussed example and many others developed for testing different aspects of the whole methodology have underlined that the natural interface approach is suitable for real-time processing also using standard computer architectures and simulating complex scenarios. The achieved results are very promising, the tracking of the user body is very robust and the physical simulation is accurate.

Considering system-related aspects, in comparison to other standard methodologies the natural interface has a greater capture volume but a lower precision. The capture volume is influenced by the IR projector and camera properties. The precision is influenced by the approximated algorithm for estimating the position of the body main joints. On the other hand, the cost is very low.

Considering user-related aspects, the natural interface has many advantages because the sensor can be quite easily worn, it is simple to use, it requires a very short training and it is not invasive in the scene.

The methodology has also been tested on a set of 30 users of different gender and age and without any experience of augmented or virtual environments. All the 30 testers reported a very interesting experience and were surprised by the easiness in achieving the interaction with the system.

The achievement of both realistic scenario and minimally invasive interaction allow the use of this methodology for many different purposes. Moreover, the natural interface requires a reduced time for training in order to be confident with the augmented environment. The combination between augmented visualization, high interaction and simulation can be a solid base for developing specific computer-aided tools for supporting different activities from simple entertainment and gaming to technical implementations in medicine, architecture and engineering.

References

1. R.T. Azuma, Y. Baillot R. Behringer, S. Feiner, S. Julier and B. MacIntyre. Recent advances in augmented reality. IEEE Computer Graphics, vol. 21(6), pp. 34–47, 2001.
2. J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, M. Ivkovic. Augmented reality technologies, systems and applications. Multimedia Tools and Applications, vol. 51, pp. 341–377, 2011.
3. O. Bimber, R. Raskar. Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Ltd, 2005.
4. J.L. Lugrin, M. Cavazza. Towards AR Game Engines. Proceedings of SEARIS - 3rd Workshop on Software Engineering and Architecture of Realtime Interactive Systems, Waltham, MA, USA, 2010.
5. L.T. De Paolis, G. Aloisio, M. Pulimeno. A Simulation of a Billiards Game Based on Marker Detection. Proceedings of the 2nd international Conferences on Advances in Computer-Human interactions, IEEE Computer Society, Washington, DC, USA, 2009.
6. O. Oda, L.J. Lister, S. White, S. Feiner. Developing an augmented reality racing game. Proceedings of the 2nd international Conference on Intelligent Technologies For interactive Entertainment, Cancun, Mexico, 2008.
7. C. Matysczok, R. Radkowski, J. Berssenbruegge. ARBowling: Immersive and Realistic Game Play in Real Environments Using Augmented Reality. of the 3rd IEEE/ACM international Symposium on Mixed and Augmented Reality, IEEE Computer Society, Washington, DC, USA, 2004.
8. F. Liarokapis. An augmented reality interface for visualizing and interacting with virtual content. Virtual Reality, vol. 11, pp. 23–43, 2007.
9. P. Buchanan, H. Seichter, M. Billinghurst M, R. Grasset. Augmented reality and rigid body simulation for edutainment: the interesting mechanism - an AR puzzle to teach Newton physics. Proceedings of the 2008 international Conference on Advances in Computer Entertainment Technology, Yokohama, Japan, 2008.
10. S. Irawati, S. Hong, J Kim, H Ko. 3D edutainment environment: learning physics through VR/AR experiences. Proceedings of the International Conference on Advances in Computer Entertainment Technology, New York, NY, USA, 2008.
11. F. Liarokapis, P. Petridis, P.F. Lister, M. White. Multimedia augmented reality interface for E-learning (MARIE). World Transactions on Engineering and Technology Education, vol. 1(2), pp. 173–176, 2002.
12. Z. Pan, A.D. Cheok, H. Yang, J. Zhu, J. Shi. Virtual reality and mixed reality for virtual learning environments. Computers & Graphics, vol. 30, pp. 20–28, 2006.

13. E. Samset, A. Talsma, O. Elle, L. Aurdal, H. Hirschberg, E. Fosse. A virtual environment for surgical image guidance in intraoperative MRI. Computer Aided Surgery, vol. 7(4), pp. 187–196, 2002.
14. P. Edwards, D. Hawkes, D. Hill, D. Jewell, R. Spink, A. Strong, M. Gleeson. Augmented reality in the stereo operating microscope for otolaryngology and neurological guidance. Computer Assisted Surgery, vol. 1(3), pp. 172–178, 1995.
15. M. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E.D. Pisano, M. Jiroutek, K. Muller, H. Fuchs. Augmented reality guidance for needle biopsies: A randomized, controlled trial in phantoms. Lecture Notes in Computer Science: Medical Image Computing and Computer- Assisted Interventions (MICCAI), vol. 2208, pp. 240–248, 2001.
16. M.A. Livingston, L.J. Rosenblum, S.J. Julier, D. Brown, Y. Baillot, J.E. Swan, J.L. Gabbard, D. Hix. An augmented reality system for military operations in urban terrain. Proceedings of Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL, USA, 2002.
17. M.A. Livingston, J.E. Swan, J.L. Gabbard, T.H. Höllerer, D. Hix, S.J. Julier, Y. Baillot, D. Brown. Resolving multiple occluded layers in augmented reality. Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR2003), Tokyo, Japan, 2003.
18. G. Klinker, O. Creighton et al. Augmented maintenance of powerplants: A prototyping case study of a mobile AR system. Proceedings of ISAR ‘01 - The Second IEEE and ACM International Symposium on Augmented Reality, New York, NY, USA, 2001.
19. B. Schwald, J. Figue, E. Chauvineau et al. STARMATE: Using Augmented Reality Technology for Computer Guided Maintenance of Complex Mechanical Elements. E-work and Ecommerce, vol. 1, pp. 196–202, 2001.
20. M. Stilman, P. Michel, J. Chestnutt, K. Nishiwaki, S. Kagami, J.J. Kuffner. Augmented Reality for Robot Development and Experimentation. Tech. Report CMU-RI-TR-05-55, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 2005.
21. S.K. Ong, Y. Pang, A.Y.C. Nee. Augmented Reality Aided Assembly Design and Planning. Annals of the CIRP, vol. 56(1), pp. 49–52, 2007.
22. P.P. Valentini, E. Pezzuti, D. Gattamelata. Virtual engineering in Augmented Reality. Chapter in Computer Animation, Nova Science Publisher Inc., New York, 2010.
23. Y. Pang, A.Y.C. Nee, S.K. Ong, M.L. Yuan, K. Youcef-Toumi. Assembly Feature Design in an Augmented Reality Environment. Assembly Automation, vol. 26(1), pp. 34–43, 2006.
24. P.P. Valentini. Interactive cable harnessing in Augmented Reality. International Journal on Interactive Design and Manufacturing, vol. 5, pp. 45–53, 2011.
25. P.P. Valentini. Reverse Engineering Tools in Augmented Reality to Support Acquisition, Processing and Interactive Study of Cultural and Archaeological Heritage. Chapter in Virtual Reality, Nova Science Publisher Inc., New York, 2011.
26. W. Dangelmaier, M. Fischer, J. Gausemeier, M. Grafe, C. Matysczok, B. Mueck. Virtual and augmented reality support for discrete manufacturing system simulation. Computers in Industry, vol. 56, pp. 371–383, 2005.
27. S.K. Ong, A.Y.C. Nee. Virtual and Augmented Reality Applications in Manufacturing. Springer, London, United Kingdom, 2004.
28. R. Reif, D. Walch. Augmented & Virtual Reality applications in the field of logistics. Visual Computing, vol. 24, pp. 987–994, 2008.
29. S. Kirchner, P. Jablonka. Virtual Archaeology - VR based knowledge management and marketing in archaeology first results - nexts steps. Proceedings of the 2001 Conference on Virtual Reality, Archeology and Cultural Heritage Glyfada, Greece, 2001.
30. Liarokapis F, Sylaiou S et al. (2004) An interactive visualization interface for virtual museum. Proceedings of the 5th international symposium on Virtual Reality, Archaeology- Cultural Heritage, Brussels and Oudenaarde, Belgium
31. J. Vallino. Interactive augmented reality. PhD thesis, Department of Computer Science, University of Rochester, USA, 1998.
32. S. Kim, A.K. Dey. AR interfacing with prototype 3D applications based on user-centered interactivity. Computer-Aided Design, vol. 42, pp. 373–386, 2010.

33. J. Gabbard, D. Hix. A Taxonomy of Usability Characteristics in Virtual Environments, Virginia Polytechnic Institute and State University, VA, USA, 1997.
34. C. Esposito. User interfaces for virtual reality systems. Human Factors in Computing Systems, CHI96 Conference Tutorial Notes, 1996.
35. F. Merienne. Human factors consideration in the interaction process with virtual environment. International Journal of Interactive Design and Manufacturing, vol. 4(2), pp. 83–86, 2010.
36. D. Bowmann, E. Kruijf, J. La Viola, I. Poupyrev. 3D User Interfaces: Theory and Practice. Addison-Wesley, 2005.
37. A. Jaimes, N. Sebe. Multimodal human-computer interaction: A survey. Computer vision and image understanding, vol. 108(1–2), pp. 116–134, 2007.
38. K. Fishkin. A taxonomy for and analysis of tangible interfaces. Personal Ubiquitous Computing, vol. 8(5), pp. 347–358, 2004.
39. B. Ullmer, H. Ishii. Emerging frameworks for tangible user interfaces. IBM Systems Journal, vol. 39(3–4), pp. 915–931, 2000.
40. M. Fiorentino, G. Monno, A.E. Uva. Tangible Interfaces for Augmented Engineering Data Management. Chapter in Augmented Reality, Intech, Croatia, 2010.
41. M. Akamatsu, I.S. MacKenzie, T. Hasbrouc. A Comparison of Tactile, Auditory, and Visual Feedback in a Pointing Task using a Mouse-type Device. Ergonomics, vol. 38, pp. 816–827, 1995.
42. H. Slay, B. Thomas, R. Vernik. Tangible user interaction using augmented reality. Proceedings of the 3rd Australasian Conference on User interfaces, Melbourne, Victoria, Australia, 2002.
43. D. Bowmann, E. Kruijf, J. La Viola, I. Poupyrev. An Introduction to 3-D User Interface Design. Presence: Teleoperators and Virtual Environments, vol. 10(1), pp. 96–108, 2001.
44. G.A. Lee, C. Nelles, M. Billinghurst, G.J. Kim. Immersive Authoring of Tangible Augmented Reality Applications. Proceedings of the 3rd IEEE/ACM international Symposium on Mixed and Augmented Reality, IEEE Computer Society, Washington, DC, USA, 2004.
45. P.P. Valentini, E. Pezzuti. Interactive Multibody Simulation in Augmented Reality. Journal of Theoretical and Applied Mechanics, vol. 48(3), pp. 733–750, 2010.
46. G. Guerra-Filho. Optical motion capture: theory and implementation. Journal of Theoretical and Applied Informatics, vol. 12(2), pp. 61–89, 2005.
47. D. Roetenberg. Inertial and magnetic sensing of human motion, Ph.D. dissertation, University of Twente, Twente, The Netherlands, 2006.
48. P.P. Valentini, L. Mariti. Improving interactive multibody simulation using precise tracking and sequential impulse solver. Proceedings of ECCOMAS Multibody Dynamics Congress, Bruxelles, Belgium, 2011.
49. P.P. Valentini, E. Pezzuti. Dynamic Splines for interactive simulation of elastic beams in Augmented Reality, Proc. of IMPROVE 2011 International congress, Venice, Italy, 2011.
50. P.P. Valentini. Interactive virtual assembling in augmented reality. International Journal on Interactive Design and Manufacturing, vol. 3, pp. 109–119, 2009.
51. T. Berliner et al. Modelling of Humanoid forms from depth maps. U.S. Patent Application Publication n. US 2010/0034457 A1, 2010.
52. D.E. Breen, R.T. Whitaker, E. Rose, M. Tuceryan. Interactive Occlusion and Automatic Object Placement for Augmented Reality. Proceedings of Computer Graphics Forum EUROGRAPHICS'96, Poitiers, France, 1996.
53. J. Fischer, D. Bartz, W. Straßer. Occlusion Handling for Medical Augmented Reality using a Volumetric Phantom Model. Proceedings of ACM Symposium on Virtual Reality Software and Technology, Hong Kong, Cina, 2004.
54. K. Hayashi, H. Kato, S. Nishida. Occlusion detection of real objects using contour based stereo matching. Proceedings of ICAT, Christchurch, New Zealand, 2005.
55. Y. Tian, T. Tao Guan, C. Wang. Real-Time Occlusion Handling in Augmented Reality Based on an Object Tracking Approach. Sensors, vol. 10, pp. 2885–2900, 2010.
56. J. Fischer, B. Huhle, A. Schilling. Using Time-of-Flight Range Data for Occlusion Handling in Augmented Reality. Proceedings of Eurographics Symposium on Virtual Environments (EGVE), Weimar, Germany, 2007.

57. C. Chae, K. Ko. Introduction of Physics Simulation in Augmented Reality. Proceedings of the International Symposium on Ubiquitous Virtual Reality, ISUVR. IEEE Computer Society, Washington, DC, USA, 2008.
58. W. Huagen, G. Shuming, P. Qunsheng. Virtual Grasping for Virtual Assembly Tasks. Proceedings of the 3rd International Conference on Image and Graphics, Hong Kong, China, 2004.
59. D.A. Bowmann, L.F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments, Proceeding of the Symposium on Interactive 3D Graphics, Providence, RI, USA, 1997.
60. B. Mac Namee, D. Beaney, Q. Dong. Motion in augmented reality games: an engine for creating plausible physical interactions in augmented reality games. Internaional Journal of Computer Games Technology, Article ID 979235, 2010.
61. B.V. Mirtich. Impulse-based dynamic simulation of rigid body systems. PhD thesis, University of California, Berkeley, CA, USA, 1996.
62. A. Schmitt, J. Bender, H. Prautzsch. On the convergence and correctness of impulse-based dynamic simulation. Internal report, n. 17, Institut für Betriebs und Dialogsysteme, Karlsruhe, Germany, 2005.
63. A. Schmitt, J. Bender. Impulse-based dynamic simulation of multibody systems: numerical comparison with standard methods. Proceedings of Automation of Discrete Production Engineering, Sozopol, Bulgaria, 2005.
64. E. Coumans. Bullet 2.74 Physics SDK Manual, 2009. [Online]. Available: http://bulletphysics.com [accessed may 2011].

Chapter 11
Interactive AR Installation: Lessons Learned in the Field of Art, Design and Cultural Heritage

Yolande Kolstee

1 Introduction

The AR Lab started as the AR + RFID Lab of the Royal Academy of Art in The Hague, The Netherlands, in which applications of augmented reality and radio frequency identification (RFID) in art and design, society and commerce have been researched since 2006.

In the past years we have seen these innovative technologies being used as a creative medium for autonomous artists, as animated prototyping tools for furniture makers and interior and urban development architects, as a virtual tool for the digital visualization of valuable cultural heritage and as geo-data visualisation tool. The applications of augmented reality are as diverse and unlimited as ones imagination [1].

It is our experience that every field of study has its own way of using new technology in its research. The creative character of Augmented Reality—creating something without matter—makes it exceedingly suited for art and design at an art academy. At the Royal Academy we have the special opportunity to experiment with high-quality technology from the Delft University of Technology, a partner in our AR Lab.

The Lab develops projects in collaboration with students and artists, on invitation of museums and/or for educational purposes. The work is presented at the involved museums, at innovation events for companies, at scientific conferences, in educational institutes and during art- or design manifestations. Furthermore, the lab regularly receives visits from students from other academies and universities, artists, architects, designers and interested companies that want to get started with the new possibilities augmented reality creates.

In this chapter we discuss three examples in the domain of cultural heritage in which we tried to enhance the interaction with work of arts through the use of innovative visualisation techniques like augmented reality. Our experiments reported

---

Y. Kolstee (✉)
AR Lab, Royal Academy of Art, The Hague, Prinsessegracht 4, 2514AN, The Netherlands
e-mail: Y.Kolstee@kabk.nl

W. Huang et al. (eds.), *Human Factors in Augmented Reality Environments*,
DOI 10.1007/978-1-4614-4205-9_11, © Springer Science+Business Media New York 2013

here- and many more—has led to a list of lessons learned that we present at the end of this chapter. The idea behind these lessons is that we and other AR artists, designers and researchers can use it as initial material to design better AR installations. A next step would be to use the UX evaluation metrics proposed in Chap. 9 to evaluate our future AR installations.

2 Pre-industrial Earthenware from 1250 to 1550; Museum Boijmans van Beuningen, Rotterdam, The Netherlands

The project "Sgraffito in 3D" has made the late medieval pottery collection of Museum Boijmans Van Beuningen accessible to a larger audience through 3D reconstruction.

Sgraffito is the term to describe earthenware pottery, in which with the aid of a sharp tool, decorations have been scratched into a thin layer of clay slip. Since the Middle Ages, this centuries-old oriental decoration technique was introduced into Western Europe of Persia and the Byzantine Empire. In the fifteenth and sixteenth centuries, potters in the Netherlands applied this technique onto simple domestic earthenware. The Van Beuningen-de Vriese collection, which is part of the Boijmans Van Beuningen Museum since 1983, contains a collection of late medieval sgraffito earthenware. These examples have been produced between 1450 and 1550 in several potteries in the Netherlands. The museum also owns a small collection of Iranian Sgraffito earthenware.

Our goal was to increase the attention for this collection [2]. Not many people are interested in old pottery. Most visitors of the museum don't feel related to old ceramics and they are not motivated by the old dishes themselves to have a closer look at its peculiarities. When the curator Historic Design Drs. Alexandra Gaba-Van Dongen asked artist Joachim Rotteveel to enhance the experience for visitors, he designed in co-operation with the AR Lab a way to make interaction possible: visitors could touch, take-up and manipulate 3D-printed replicas. A 3D-archive let visitors play with digital representations of our cultural heritage, which were also available for research and downloading online. The augmented reality installations offered visitors tactile interaction with the replicas and direct visual and aural access to background knowledge of the valuable cultural objects. Also, the website www.sgraffito-in-3d.com was developed where one could see all ceramics (not only the ones used in this experiment) and gain information. This functions as a virtual museum [3].

Seven dishes, the "protagonists" were CT-scanned (Fig. 11.1) and from the CT scans virtual copies of the dishes were made. The virtual copies were 3D printed and thus became physical copies. They were exposed very close to the original medieval ones that were stored behind safety glass. The 3D printed dishes were chained and the audience could touch and manipulate them. The augmented reality part came in with our AR installations explained below.

Fig. 11.1 CT-scanning and resulting CT scan. Mark the fractures repaired in earlier days

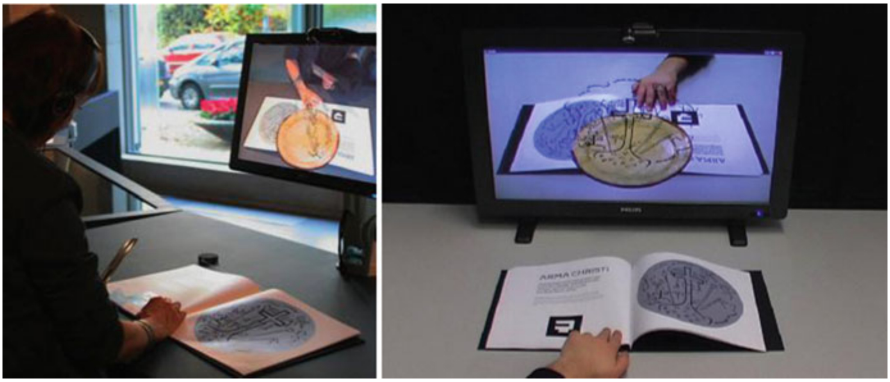


Fig. 11.2 A 3D pop-up book showed—in AR—the dish and its ascending decoration

2.1 The 3D Pop-Up Book

The 3D Pop-Up book is an AR book in which text and music is attached to the iconography of the Sgraffito decoration of the dishes. The books (English and Dutch versions) were situated in front of an LCD screen with a camera on top and next to each book were headphones available. The camera image looking to the book was displayed on the LCD screen with on top of that the (augmented) 3D graphics (Fig. 11.2).

For each of the seven protagonists, the book offered two pages next to each other. On the right page one could see the decoration in a 2D print, on the left page text

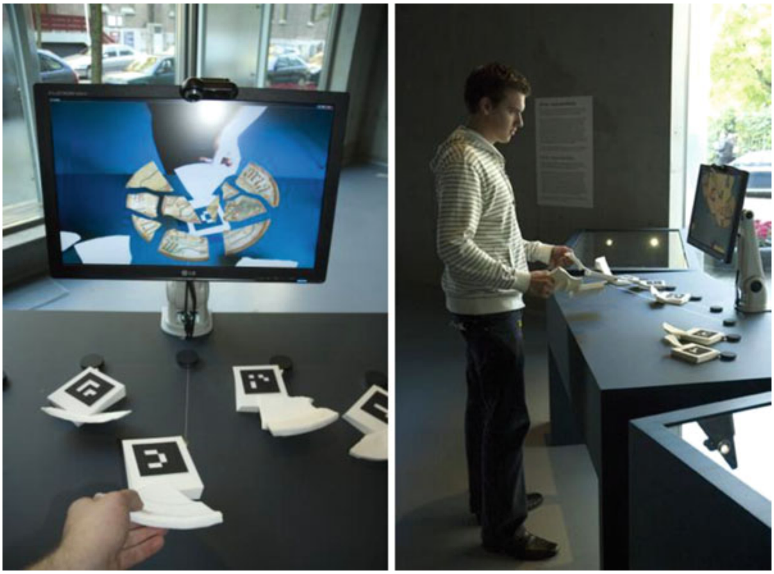


Fig. 11.3 3D printed shards create a complete virtual dish in AR on a monitor

related to the iconography of the decoration could be read. An AR marker was printed on the left page. When the camera recognized the marker three things happened:

- Music was played from the era the dish was made and related to its decoration
- The dish appeared on the LCD screen in 3D on the right page of the book
- The Sgraffito decoration was released from the dish, drifted above the dish and turned around and returned again to the dish. This animation looped

2.2 Integration of the Virtual and the 3D Printed Shards

The original lines along the dishes were once broken and restored in their time (the ceramics stem from 1250 to 1550) were clearly visible in the CT-scan. From each plate we 3D printed a shard. When holding a 3D-printed shard in front of a webcam it reveals its missing pieces on the monitor and thus forms a complete virtual dish. The 3D fragment acts as interface for the virtual object; when you move the shard in your hand you move both the physical and virtual objects simultaneously as if it were one integrated object.

The shards were chained and could be hold in front of the camera and so create a complete dish in AR on the screen (Fig. 11.3). Even children understood the concept (Fig. 11.4).

Fig. 11.4 Young children understood the interaction between shard and monitor immediately

2.3 Spatial AR "out of the screen"

At the opening of the exhibition we set a table with a cloth embroidered with black and white markers. A visitor wearing our AR headset could see the medieval earthenware on the table (Fig. 11.5). With cardboard AR markers one could add dishes and change their order [4]. We used a beamer to project the scene on the wall behind the person wearing the AR headset to show the other visitors the augmented scene.

2.4 Discussion

Our goal was to enhance the connection with "old pottery" by giving the visitors an interactive experience. We found out that:

1. Screens (monitors) with moving images generate attention; visitors tend to have a look on a screen on which action is to be seen. We saw that visitors unintentionally were attracted to our AR 3D pop-up books because the marker on a page of the open book generated the moving virtual dishes [4].
2. A book is a very well known object: it is not necessary to explain that one can turn a page. That is why visitors seemed to be "seduced" to interact with the virtual images on the screen, by turning the pages and by doing this provoked new images even without thinking. The motivation to act came from inside; intrinsic curiosity led the visitors to stand still, turn pages and after seeing the plates appear, they were persuaded to put on the available headphone.

Fig. 11.5 Laying the table with medieval pottery in spatial, immersive augmented reality

3. The delicate music heard through the headphones connected with the monitor on which the pop-up books could be seen, was for most visitors very particular. The music originating from 1250 to 1550. This single-person experience enhanced the concentration with which the visitors looked at the ascending and descending decoration of the virtual dish, meanwhile reading the text that was providing a complementary and sometimes terrifying[1] information. Thus a rather high level of immersiveness was generated: this was for most visitors an eye-opener; one could really get "involved" with a piece of old pottery [5]!
4. Since the 3D printed shards were placed in front of a monitor and each "hard copy" shard generated a sequence of centrifugal and centripetal virtual shards on the monitor (thus completing the dish and subsequently falling pieces), the relation between the "hard copy" fragment and the virtual fragments was immediately clear for each visitor. This was especially true for very young visitors.
5. From earlier experiments (Milano, see below) we learned that when using AR headsets in a crowd, the right way to attract attention is to peek the augmented scene from the headset and project this on a wall or on a very big screen [5].

---

[1] Terrifying was the text connected to the dish with a decoration of pins and the name of Agatha. She—a martyr—was tortured and her breasts were cut off with pins on order of Quitianus, governor of Julius, a Roman Emperor.

3 Showing Virtual Furniture in Salone del Mobile, Milan, Italy

Every year in April in Milano, Italy, a huge international furniture and interior design fair is held for 10 days, the so-called *Salone del Mobile*. Design academies and designers from all over the world try to attract the attention of the international press and "hot-shots" in this field. Italy and especially Milano is world famous for its innovative design, so for everybody who has an exhibition during this week there is a lot at stake.

The AR Lab wanted to research if it is possible to show virtual furniture to a large audience with spatial AR, and furthermore, how the audience receives this. Will the AR installation enhance the connection between the professional audience and the designs of our students [6]?

We rented a house in Milano and prepared the big living room for our virtual furniture show. We placed a grid of markers meticulously precise on the walls and the floor. The cameras attached to the AR headsets could identify the markers and their position in the room, making it possible to put virtual furniture at any desired place in the room.

3.1 Augmented Furnishings (Cupboards, Tables and Chairs)

We had two headsets with two backpacks available (Fig. 11.6a). They were rather heavy and not really comfortable. Figure 11.6b shows that the backpack was for a child too heavy to carry; a team member supported him. The weight was due to the heavy duty laptop, AR glasses control box and batteries.

The AR furniture could be seen from all sides, this caused some visitors to go on all four to look behind and beneath the virtual furniture (Fig. 11.7).

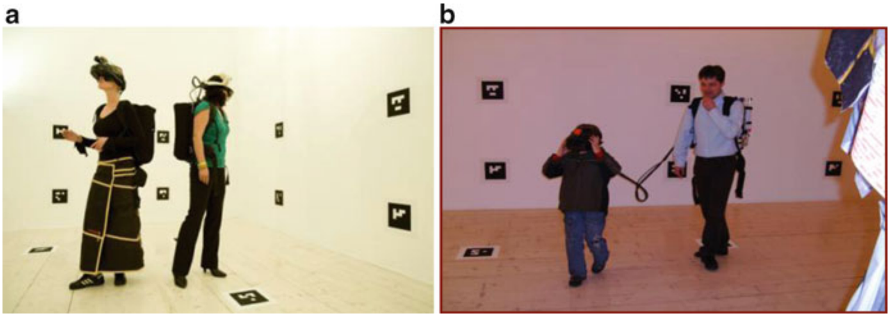


Fig. 11.6 (**a**) The two headsets with backpacks. (**b**) Sometimes a carrier was welcome

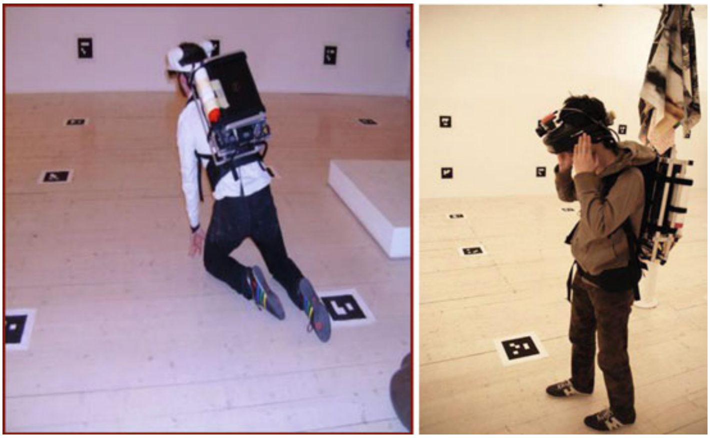


Fig. 11.7 Visitors engaged with the virtual furniture in Salone del Mobile

3.2 Animated Textiles

In the centre of the room we attached over 20 printed textiles the size of a towel at a pillar. Attached to these textiles were RFID tags. When a visitor equipped with our headset, put on a glove with an RFID-*reader* attached to the glove, he could see in augmented reality, a large banner of the designed textile from the ceiling to the ground, about 3.5 m high and 4 m wide. These augmented prints changed with the real prints he/she touched with the glove.

3.3 Discussion

First of all we have to keep in mind that this exhibition was held in a nervous Italian environment, a huge fair spread all over Milan, in which every exhibitor was in competition with the other because there was too much to see. Therefore the visitors generally want to be able to make an instant decision when entering a venue: “this is worthwhile, so I’ll spend some time”, or “I skip this and hurry to another venue”. Having to put on the heavy backpack, to adjust the headset and to give some instruction takes at least a few minutes. We had to find a way to track attention of potential visitors. Screens at the entrance showed the living room with the visitors queue and the augmented scenes from the headsets. The students at the entrance door wore fake headsets to draw the attention to AR. Hence we drew a lot of attention (Fig. 11.8a).

In this experiment we encountered the following problems:

1. An augmented reality headset is single-person equipment. This is quite a contradiction for a public space in which we want to involve as much persons as possible.

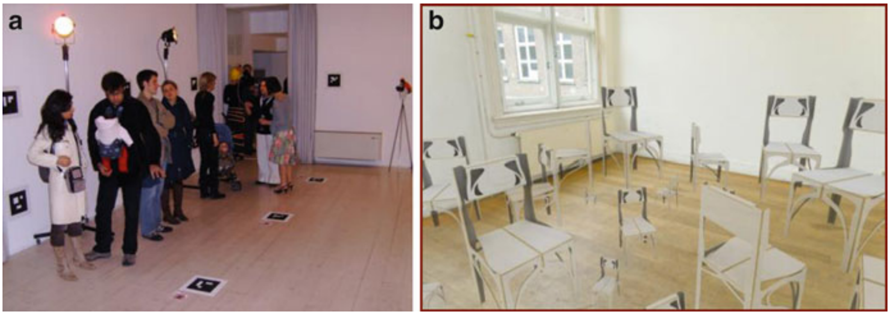


Fig. 11.8 (**a**) Visitors waiting to wear the headset. (**b**) Duplicating chairs is easy

2. Visitors do not like to wait too long for any experience; this is equally true for a quite "exotic" and immersive experience with sophisticated equipment.
3. There were difficulties with the alignment, i.e. the placing of the virtual objects on the desired place.
4. As in normal exhibitions, one needs a professional curator to do the placements of the objects to make a coherent view see (Fig. 11.8b).
5. Virtual objects or scenes might jitter, jumping a little up and down or appear and disappear, based on unpredictable human motions and camera occlusions.
6. People like to be amazed but demand an instant explanation directly following this experience. Sometimes this necessity has to become technical, e.g. explaining that when a marker is blocked virtual furniture disappears or flies away.

4 Commenting Art in the Sculpture Garden of the Kröller–Müller Museum, Otterlo, The Netherlands

The Kröller–Müller Museum named after Helene Kröller–Müller (1869–1939), is settled in the woods. The sculpture garden was opened in 1961 with works by Rodin, Moore and Hepworth, Snelson, Christo, Serra, Jean Dubuffet, and others. The works of art are spread across the entire terrain, the one clearly visible, the other secluded in a remote corner or hidden among the rhododendrons. Each piece has become inextricably linked with its place in the sculpture garden.

For our AR installations we were allotted a part of the Sculpture Garden. The Lab presented a large augmented reality installation for the "Sweet Summer Night Illusion" exhibition in 2009.

The assignment of the museum was to comment on some of the twentieth century sculptures with tools of the twenty-first century in line with the theme of the exhibition [7]. Since the theme of the exhibition was *Illusion*, we could run pretty wild. Part of our AR images were related to Hieronymus Bosch' famous triptych *the Garden of Earthly Delights*, with angels and monsters all around the scene.

Fig. 11.9 The AR walkers in the Kröller–Müller Sculpture Garden

To work in the open air with augmented reality is quite a challenge, let alone the largeness of even “our” relatively small area.

For this event, the Lab created unique 3D content and developed innovative augmented reality equipment, such as *walkers* (for elderly people) with screen-based AR systems mounted on them, and a large rotating screen showing a 360° panorama of the sculpture garden. Mounting laptops on walkers seemed us to be a humorous interface. For rainy weather we prepared an AR headset experience inside the museum.

4.1 Walkers and the Garden of Earthly Delights

With Mac books mounted on walkers, people could make a tour in “our” part of the Sculpture Garden. Some markers were placed on twigs of the trees with clothespins and other markers were attached to small sticks in the ground. The camera caught the markers and on the screens fantasy animals were seen by which we were augmenting the bushes (Fig. 11.9).

4.2 Big AR Screen Commenting on the Sculptures

From our spot in the garden we could see two sculptures, one from Mario Merz, *Igloo di Pietra* and one from E. Dodeigne, *Couple*. When preparing the AR experience, we figured that people would wonder what was inside this Igloo. When they

Fig. 11.10 Big markers can be caught by the AR camera over long distances (**a**). Igloo di Pietro, (**b**). visitors, (**c**). Couple

Fig. 11.11 A huge virtual blockhead with animated blocks could be seen through AR glasses

turned the big AR screen towards the Igloo, they could see the original sculpture and on the screen what could have been inside in augmented reality (Fig. 11.10).

Researching in Dodeignes oeuvre we found out that Dodeigne wrote in his diary he wanted to have more stone couples at that spot. We made an extra virtual couple and when turning the screen towards the stone couple, one could see two couples on the screen: one real and one added in augmented reality. The camera mounted on the large screen captured the big markers next to the real sculptures, in order to overlay the augmented content.

4.3 Spatial AR Inside the Museum

To be prepared for rainy weather we placed one installation inside the museum in an indoor garden: looking with our headset through the windows towards a cube of markers outside, visitors could see a sculpture of a head, which consisted of blocks moving forward and backwards (Fig. 11.11). The computer of the headset was also

Fig. 11.12 Inside the museum an AR headset was used with the backpack on a walker

mounted on a walker, so one could see this virtual head from different angles: aside, in front and on the back (Fig. 11.12).

4.4 Discussion

Visitors were immediately attracted to the walkers. What we hoped came out. The anachronistic interface; high tech equipment mounted on a walker for elderly people in a Sculpture Garden with sculptures from the nineteenth and twentieth century. It generated joy; young and old wanted to walk with the walkers. The images people saw on the screens were a surprise for most visitors [8]. The big screen was very easy to use: people intuitively turned it around and saw the augmented sculptures on the screen.

Inside the museum—although the weather was good—the "blockhead" was impressive.

In this experiment we encountered the following issues:

1. Visitors were amused to walk with the walkers and found their way along the markers in the trees and on the lawns. However; there was little "click" with the theme of the exhibition, unfortunately. The excitement of the new technology overshadowed the link with the theme *Illusion*.
2. Our art students related the augmented reality fantasy creatures in the bushes to the Garden of *Earthly Delights* theme, but the visitors did not notice the association. More narrative information was needed.
3. Commenting the sculptures got the highest appreciation: this really added information and created a lively discussion among the visitors.

5 Lessons Learned

From the AR installations described above we learned.

5.1 Boijmans van Beuningen

1. Screen based AR is a low cost replacement of HMD based AR and can be fruitfully used to introduce the topic at hand and the AR technology itself.
2. People tend to like manipulating virtual objects or at least have some influence on them.
3. Augmented reality can be fruitfully used to attract a broad public to displays of cultural heritage. Its inherent narrative power is huge.

5.2 Salone del Mobile

1. The augmented view can be peeked from the tracker camera and displayed on a beamer or screen to let the other visitors see through the user's eye at the same time.
2. Positioning virtual objects in the air covers up for misalignment. We met some troubles in positioning the virtual objects but we found a way to overcome this by animating the object. The visitor did not see just a virtual chair, but saw in AR an animation in which all parts of a chair (legs, backseat, elbow rests) were falling from the sky and form a complete chair on the floor. This led us to:
3. Motion of the virtual objects covers up for misalignment and jitter.
4. Manipulation of real objects can influence—through RIFD—the virtual world. This is "magic" for many people.
5. Design discussions are more vividly using HMD based AR as each user can now individually select his ("the best") viewpoint.
6. Headset based AR is at its best when a full immersive experience is required and people can walk around objects like chairs and tables, but also within larger objects; buildings, molecules, DNA.
7. The "empty" third dimension in the air around us, is very useful for information display and interaction and detaches the application from gravity, we can have art in the air. There is a lot of space up there.

5.3 Kröller–Müller

1. For outdoor AR it is necessary that the ambient light intensity and the intensity of the LCD displays on the HMD is in balance (automatic sunglasses).
2. The relationship between the real world objects and the added ones should be strong and convincing; otherwise the AR work becomes just a gadget. One should not forget the story telling or design of the augmented scene. This is more complex then in a normal situation.

5.4 All Experiments (Including the Ones Not Described Here)

1. The collaboration between researchers in the area of image processing with artists, designers and curators appeared to be very fruitful and has lead to many amazing productions and exhibitions.
2. Not only humans interact with humans in and via “cyberspace” also real, physical objects communicate with each other, with or without knowledge of people. People tend to understand this quickly.
3. Design packages such as Cinema 4D make design with animated figures possible. Most design packages like Cinema 4D don’t allow large plots. For real 3D animated films with large plots, game engines (like Unity) must be used.
4. When adding sound to virtual objects, this adds to their attention drawing and pose tracking.
5. More image processing on the tracker camera is useful, e.g. to segment the users hand and fingers to make unhandy data gloves superfluous; keyboard interfaces are useless.
6. Segmenting—with image processing techniques—moving objects such as people enables virtual objects to encircle them.
7. Standard heavy duty gaming laptops are heavy to wear but enable easy connections to new interaction devices such as the Wii. We wait for diminished equipment.
8. By applying VR design techniques, i.e. also modelling walls, floor and ceiling, virtual objects appear real and real objects appear virtual.
9. Life video streams inside the virtual world give a tele-presence awareness.
10. Augmented reality books of all experiments can be used as a way of archiving the temporarily AR installations. One can show them again at any time.

6 Conclusions

Augmented reality in the artistic and cultural domain might be used as a tool for artists (fine-art and design students) and as a medium to enhance the relationship with the collection of a museum.

For all applications the cooperation with students from the Delft University of Technology was fruitful. Partly in separately financed projects, partly financed by projects run by the Royal Academy they developed the technology that was used for our systems [9–18]. Thanks to this we were able to use cutting edge—albeit not yet consumer friendly—equipment [19, 20]. However, what turned out to be most valuable was the sincere artistic approach with which the students (with help from our Lab) were engaged to give visitors an interactive experience, which causes a special relationship with works of art [21, 22]. To our opinion researching, rethinking, refining and redefine interaction with innovative visualisation techniques will continue. AR has just begun.

Acknowledgements Below we list the persons and their role that contributed to the described AR installations.

Boijmans van Beuningen

Yolande Kolstee. AR Production, AR Lab, Royal Academy of Art The Hague (KABK)
Joachim Rotteveel. Media artist www.joachimrotteveel.com and KABK
Wim van Eck. 3D animator, www.wimeck.com and KABK
Melissa Coleman. Programmer www.dancetechnology.nl and KABK
Mit Koevoets. AR Art student, Art Science, KABK
Marina de Haas. AR Art student, Fine Arts, KABK
Margot Kalse. Specialist in Medieval Music, www.margotkalsevocaal.nl
Alexandra Gaba van Dongen. Curator Boijmans van Beuningen, www.boijmans.nl
Pieter Jonker. Professor in Robot Vision, Delft University of Technology (TUD), www.dbl.tudelft.nl

Salone del Mobile Milano

Yolande Kolstee. AR Production, AR Lab, Royal Academy of Art The Hague (KABK)
Wim van Eck. 3D animator, www.wimeck.com and KABK
Pawel Pokutycki. Interaction Designer and KABK
Melanie Luchtenveld. Student furniture and interior design, KABK
Marina de Haas. AR Art student, Fine Arts student, KABK
Iris Bijvelds. Student furniture and interior design, KABK
Barbara Kooi. Student furniture and interior design, KABK
Eva Malschaert. Student furniture and interior design, KABK
Suzanne Vruwink. Student furniture and interior design, KABK
Jan Willem Brandenburg. Researcher, Delft University of Technology
Pieter Jonker. Professor in Robot Vision, Delft University of Technology, www.dbl.tudelft.nl

Kröller–Müller

Yolande Kolstee. AR Production, AR Lab, Royal Academy of Art The Hague (KABK)
Wim van Eck. 3D animator, www.wimeck.com and KABK
Joachim Rotteveel. Media artist www.joachimrotteveel.com and KABK
Melissa Coleman. Programmer www.dancetechnology.nl and KABK
Marcel Kerkmans. Student furniture and interior design, KABK
Mit Koevoets. AR Art student, Art Science, KABK
Ferenc Molnar. AR Art student, Photography, KABK
Jing Foon Yu. AR Art student, Graphic Design, KABK
Herman Tibosch. Curator of the Kröller–Müller museum, www.kmm.nl
Guus LiquiLung. Robot Technician, Delft University of Technology, www.dbl.tudelft.nl
Pieter Jonker. Professor in Robot Vision, Delft University of Technology, www.dbl.tudelft.nl

References and Further Reading

1. Rozhen Kamal Mohammed-Amin, *Augmented Reality: A narrative layer for historic sites*, thesis of the faculty of environmental design, Calgary, Alberta, September 2010.
2. Jurjen Caarls, Pieter Jonker, Yolande Kolstee, Joachim Rotteveel, and Wim van Eck, *Augmented Reality for Art, Design and Cultural Heritage -System Design and Evaluation*, EURASIP Journal on Image and Video Processing, vol. 2009, Article ID 716160, 16 pages, 2009. doi: 10.1155/2009/716160.

3. Kadir Ulusoy, *Perspectives on using virtual museums application in teaching history subjects to open education students.* Mersin University, Faculty of Education, Mersin, Turkey. Turkish Online Journal of Distance Education, TOJDE, April 2010 ISSN 1302–6488 Volume: 11 Number: 4, Notes for Editor-4.
4. Lev Manovich, *The language of new media*, MIT Press, New edition, Feb. 2002.
5. Vassilios Vlahakis, Nikolaos Ioannidis, John Karigiannis, Manolis Tsotros, and Michael Gounaris, *Archeoguide: An Augmented Reality Guide for Archaeological Sites in Computer Graphics in Art History and Archeology.* IEEE Computer Graphics and Applications (2002) Volume: 22, Issue: 5, IEEE Computer Society Press, 52–60.
6. P. Milgram, H. Takemura, A. Utsumi and F. Kishino. *Augmented Reality: A class of displays on the reality-virtuality continuum*, In Telemanipulator and Telepresence Technologies, volume 2351, pp. 282–292. 1994.
7. S. Persa and P.P. Jonker, *On positioning for augmented reality systems*, in: H.-W. Gellersen (eds.), Handheld and Ubiquitous Computing (Proc. HUC'99, 1st Int. Symposium, Karlsruhe, Germany, September 27–29), Lecture Notes in Computer Science, vol. 1707, Springer, Berlin, 1999, pp. 327–329.
8. Christine Perey, *What will work where and when? A Mobile AR Test Suite and Lab* A Position Paper for the Mobile AR Summit @MWC 2010, Febr 9.
9. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre. *Recent advances in Augmented Reality*. In IEEE Computer Graphics and Applications, volume 21, no. 6, pp. 34–47. November/December 2001.
10. Yolande G. Kolstee, *Visualization and interaction tools for art and design*. In: Verbeek, F.J., Lenior, D and Steen, M. (Eds.) Proceedings of the 13th CHI-NL (Computer Human Interface) conference, pp. 57–60, 2009, July 11, Leiden University.
11. J. Caarls, *Pose estimation for mobile devices and Augmented Reality*, Ph.D.Thesis, Sept. 25, 2009, Delft University of Technology, Faculty of Applied Sciences, Quantitative Imaging Group, Delft, The Netherlands. ISBN: 978-90-9024585-0.
12. W. Caarls, *Automated Design of Application-Specific Smart Camera Architectures*, Ph.D.Thesis, Feb. 4, 2008, Delft University of Technology, Faculty of Applied Sciences, Quantitative Imaging Group, Delft, The Netherlands.
13. S.F. Persa, *Sensor Fusion in Head Pose Tracking for Augmented Reality*, Ph.D.Thesis, June 6, 2006, Faculty of Applied Sciences, Quantitative Imaging Group, Delft University of Technology.
14. P.P. Jonker, S. Persa, J. Caarls, F. de Jong, and R.L. Lagendijk, *Philosophies and technologies for ambient aware devices in wearable computing grids*, Computer Communications, vol. 26, no. 11 (Special Issue on Ubiquitous Computing, Edited by T. Pfeifer), 2003, pp. 1145–1158.
15. J. Caarls, P.P. Jonker, and S. Persa, *Sensor Fusion for Augmented Reality*, in: Emile Aarts, Rene Collier, Evert van Loenen, Boris de Ruyter (eds.), Ambient Intelligence, Lecture Notes in Computer Science, vol. 2875, Springer Verlag, Berlin, 2003, pp. 160–176.
16. W. Pasman, S. Persa, and F.W. Jansen, *Realistic Low-Latency Mobile AR Rendering*, in: B. Fisher, K. Dawson-Howe, C. O'Sullivan (eds.), Virtual and Augmented Architecture (VAA'01) Proc. Int. Symposium (Dublin, Ireland, June 21–22), Springer Verlag, Berlin, 2001, pp. 81–92.
17. S. Persa and P.P. Jonker, *On Positioning for Augmented Reality Systems*, Proc. Signal Processing Symposium 2000 (SPS2000), IEEE Benelux Signal Processing Chapter (Hilvarenbeek, NL, March 23–24), 2000, pp. 1–3.
18. S. Persa and P.P. Jonker, *Human-computer Interaction using Real Time 3D Hand Tracking*, in: J. Biemond (eds.), Proc. 21st Symposium on Information Theory in the Benelux (Wassenaar, NL, May 25–26), 2000, pp. 71–75.
19. Jung Yeon Ma , Jong Soo Choi, *The Virtuality and Reality of Augmented Reality*, in: Journal of multimedia, vol. 2, no. 1, February 2007.
20. Marcus Specht, Stefaan Ternier, and Wolfgang Greller, *Mobile Augmented Reality for Learning: A Case Study*, in: Journal of the Research Centre for Educational Technology, Vol 7, Issue 1, Spring 2011.

21. Viet Toan Phan and Seung Yeon Choo, *Interior Design in Augmented Reality Environment.* in: International Journal of Computer Applications (0975–8887) Volume 5 No.5, Aug. 2010.
22. Rafa Wojciechowski, Krzysztof Walczak, Martin White, Wojciech Cellary, *Building Virtual and Augmented Reality Museum Exhibitions*, Department of Information Technology, The Poznan University of Economics, Poland, Department of Informatics, University of Sussex, UK, Proceeding Web3D '04 Proceedings of the ninth international conference on 3D Web technology ACM New York, NY, USA ©2004.