Multimodal Augmented Reality

- The Norm Rather Than the Exception

Hanna Schraffenberger[1,2]
h.k.schraffenberger@uvt.nl

Edwin van der Heide[1]
e.f.van.der.heide@liacs.leidenuniv.nl

[1]Media Technology Research Group
Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

[2]Tilburg center for Cognition and Communication, Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands

ABSTRACT

Augmented reality (AR) is commonly seen as a technology that overlays virtual *imagery* onto a participant's *view* of the world. In line with this, most AR research is focused on what we *see*. In this paper, we challenge this focus on vision and make a case for an experience-focused and modalities-encompassing understanding of AR. We argue that multimodality in AR is the norm rather than the exception, as AR environments consist of both virtual content and our real, physical, multimodal world. We explore the role multimodal and non-visual aspects of our physical reality can play when creating AR scenarios and the possibilities and challenges that emerge when approaching AR from a modalities-encompassing perspective.

CCS Concepts

•Human-centered computing → Mixed / augmented reality; *Interaction devices; Interactive systems and tools;* **HCI theory, concepts and models; •Computing methodologies → Mixed / augmented reality; •Hardware →** Sound-based input / output; Tactile and hand-based interfaces;

Keywords

augmented reality, multimodality; interaction; multimodal interaction, multimodal augmented reality; non-visual augmented reality; mixed reality; theory; experience; perception; virtual; real; senses

1. INTRODUCTION

Can we experience augmented reality (AR) with our eyes closed? We certainly think so. We experience the real world with all our senses. Just like real objects, virtual objects can have non-visual and/or multimodal qualities. As a combination of the virtual and the real, it seems only natural that AR engages all our senses and creates multimodal experiences. Yet, existing AR research is primarily focused on vision.

The focus on vision has a long tradition in AR research. When Caudell and Mizell [8] coined the term 'augmented reality' in the early 1990s, they proposed a Heads-up display meant to "augment the worker's visual field of view" (p. 660) by overlaying virtual content onto the worker's view of the real world. Their proposed headset intends to make the life of assembly and manufacturing workers easier—for instance, by presenting virtual arrows in real space that indicate where holes have to be drilled.

In line with Caudell and Mizell's [8] proposed heads-up display, AR is commonly seen as a technology that overlays virtual *imagery* onto a user's *view* of the world. For instance, Piekarski and Thomas [27] describe AR as "the process of overlaying and aligning computer-generated images over a user's view of the physical world" (p. 36). Likewise, Reiners et al. [29] claim that "Augmented Reality is a technology that integrates pictures of virtual objects into images of the real world" (p. 31). Similarly, Doyle, Dodge and Smith [11] describe AR as "a technology in which a user's view of the real world is enhanced or augmented with additional information generated from a computer model" (p. 147). Many other sources provide similar descriptions, among which the Oxford English Dictionary [1], which defines AR as "a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view".

In this paper, we challenge this focus on vision and argue that AR is about much more than what a participant sees. Section 2 gives a short overview of existing work on multimodal AR. We review both theoretical claims as well as practical projects that make use of multimodal input and/or output. In Section 3, we propose a new, experience-focused and modalities-encompassing perspective on AR. In contrast to common views, this perspective does not focus on what an *AR system* senses or displays, but instead, considers what a *participant* can do and perceive when engaging with AR. We argue that from a participant's point of view, virtual content is experienced as part of an otherwise real, multimodal environment. Consequently, even visually augmented environments can engage all senses, simply because they consist

MVAR'16, November 16 2016, Tokyo, Japan

ISBN 978-1-4503-4559-0/16/11...$15.00

DOI: http://dx.doi.org/10.1145/3001959.3001960

of both virtual content and the multimodal real world. Section 4 explores the possibilities and challenges that emerge from approaching AR from such an experience-focused and modalities-encompassing perspective. In particular, we suggest that the multimodal properties of the real world can play a role in whether we experience virtual as belonging to the real world. We argue that when discussing multimodal interaction in AR, we should not only think about multimodal interaction between an AR system and a user but also consider the multimodal interaction between virtual content and the real world. We hypothesise that realising such interactions between the virtual and real can increase the experience of virtual content being part of the otherwise real environment. Section 5 concludes the paper. It summaries the main claims, discusses our findings and proposes directions for future research.

In contrast to applied and technology-driven research, our paper takes a fundamental and conceptual approach. We aim at advancing AR through a better understanding of what it is and potentially can be. Technological implementations of our ideas as well as practical experiments with participants fall out of the scope of this paper.

2. MULTIMODALITY IN AR

AR is commonly understood as a technology that overlays virtual imagery onto a participant's view of the world. Yet, the realisation that AR can engage all of our senses is not new. In fact, many researchers point out that AR technologies can also blend non-visual virtual and real stimuli. What is more, we can find various practical AR projects that prove them right.

To mention but a few examples, Azuma [2], in his widespread review of AR mentions that "Augmented Reality might apply to all senses, not just sight" (p. 361). Among other, he suggests that "AR could be extended to include sound" (p. 361) and make use of headphones and microphones to add synthetic directional 3D sound and potentially also cancel out real sounds. Azuma et al. [3] reinforce this image of AR and likewise emphasise that "AR can potentially apply to all senses, including hearing, touch, and smell" (p. 34). (In line with this, their definition [2, 3] does not refer to the overlay of virtual imagery but more generally, to the combination and alignment of the virtual and the real.) Moreover, Milgram and Kishino [21], in their popular paper on Mixed Reality Visual Displays, briefly refer to the possibilities of mixing computer-generated spatial sounds with natural sounds in the environment. Furthermore, they also note the possibilities of haptic and vestibular AR, both of which provide non-visual additional stimuli.

Unfortunately, these possibilities for multimodal AR often accompany vision-focused AR research and are only mentioned in passing. A notable exception is the work by Lindeman and Noma [19]. The authors propose a framework that classifies AR technologies for the five traditionally recognised senses. The classification is based on where on the pathway from the real environment to the brain real stimuli and computer-generated stimuli are mixed. Furthermore, their paper presents existing and envisioned technologies that (might) allow us to merge real and virtual stimuli directed at the five primary senses. Like the research presented in this paper, their paper considers the effects that the real environment can have on virtual content. For instance, they propose that "the voice of a virtual character should also be influenced by environmental objects, such as occluders or reflectors" (p. 175).

Clearly, the AR research community is aware that AR can be more than meets the eye. If we look at the AR landscape, we can find various practical examples that put this knowledge into practice. In particular, there are various AR projects that present a participant with multimodal or non-visual virtual content.

One such project that makes use of non-visual additions is the *REVEL* device. This tactile technology allows participants to feel virtual tactile textures when touching real physical objects [4]. Furthermore, there is the *Gravity Grabber* [22], which allows participants to perceive the ruffle of the water in a glass, although they actually are holding an empty glass. Furthermore, there are various sound-based projects, which often allow us to hear virtual content in real 3D space (e.g., [30, 13, 9]). In addition, researchers have also been working with our senses of smell and taste. In particular, there have been several attempts to create pseudo-gustatory displays (e.g., [25, 17]). These typically make use of cross-modal effects, and aim at changing the experiences of real food or drinks without changing their underlying chemical composition—merely by changing accompanying information, such as its visual appearance or smell. An example of such a pseudo-gustatory device is the *MetaCookie* headset [23, 24]. This project is based on the idea that seeing and smelling a certain type of cookie when eating a plain cookie might affect its perceived flavour. Consequently, their setup changes the visual appearance of a plain cookie and, for instance makes it look like a chocolate, almond or cheese cookie. At the same time, it also features an olfactory display with scents that match the visual choices. Reportedly, this additional information often alters the taste of the real plain cookie.

In addition to projects that display multimodal content, there are also some projects that incorporate multimodal input. Typically, these projects not only allow a participant to navigate through the environment and perceive virtual additions from different perspectives, but instead, allow the participant to interact with the AR scenario in a multimodal way.

An example of such a project has been presented by Irawati et al. [15]. Their multimodal AR interface allows users to arrange visual virtual furniture through both speech input as well as gestures made with a physical paddle. Furthermore, Olwal, Benko and Feiner [26] have created a multimodal AR system that supports interaction with virtual objects through speech and hand-gestures. In their paper, they propose a set of statistical tools designed to identify what object a user refers to when using such gestures. Although these projects present solely visual content to the participant, they make use of a multimodal *interface* in order to facilitate interaction between the participant and virtual content.

Finally, some AR projects feature systems that incorporate both multiple input modalities and multiple output modalities. For instance, Bianchi et al. [5] have presented an AR-based ping-pong game that reacts to the participant's head pose as well as haptic input and that provides both haptic and visual stimuli. The game makes use of a rather typical visual AR setup that displays a virtual ball in the real environment via an HMD. In addition, the setup also makes use of a force feedback device that allows the participant to play with the virtual ping pong ball in the real environment. This device acts both as an input and as an output unit: the user can steer a virtual bat and feel the impact of the virtual ball on his/her simulated bat. Another project that makes use of both multimodal input and output is the mobile AR game *GeoBoid* by Lindeman et al. [18]. In their game, players are surrounded by flocks of virtual geometric creatures called GeoBoids. These creatures are represented both visually as well as by means of spatialized audio. Furthermore, the underlying system reacts to the players movement, sounds and on-screen touch gestures: players move towards a swarm of GeoBoids by running to their location in the real world, they can scare the flock by whistling at a certain pitch and for a certain duration and they can capture individual creatures by pointing the device at them and swiping over the screen of their mobile device.

As these examples show, AR systems can incorporate both multimodal input or/and output. Both virtual and real elements can be more than meets the eye. In our opinion, the fact that AR can make use of non-visual input and output alone is reason enough to approach AR from a modalities-encompassing perspective. However, we believe there is another—more fundamental—reason to do so: AR aims at integrating virtual content into otherwise real environments. Typically, these real environments are not only something a participant can see, but something the participant can perceive with all their senses, act in and interact with. As a combination of the virtual and the real, AR environments, too, engage all our senses and allow for interaction. If AR meets its goal, AR technologies allow for multimodal *experiences*, even if they do not display or sense multimodal information.

3. MULTIMODALITY IN AR REVISITED

What is the purpose of AR? Judging from existing research, the goal of AR is "to create the impression that the virtual objects are part of the real environment " [28, p. 504], or more generally, "to combine the interactive real world with an interactive computer-generated world in such a way that they appear as one environment" [33, p.1]. In yet other words, AR typically aims at seamlessly blending the virtual and the real world (cf. [7, 6]).

In order to achieve this goal, an AR system usually combines the virtual and the real, aligns virtual and real elements in three dimensions and runs interactively and in real-time [2, 3]. Most commonly, this is realised visually: virtual imagery is overlaid onto and integrated into a participant's view of the world. Accordingly, AR is considered to provide a "composite view" [1] or to augment the user's view of the world.

However, if AR meets its goal of blending the virtual and the real, this is not all that happens. If AR technologies do their job well, a participant does not experience the *overlay* of virtual content on their *view*—instead, they experience the existence of the virtual content in their real environment. As Azuma [2] points out "[i]deally, it would appear to the user that the virtual and real objects coexisted in the same space" (p.356). If AR creates this illusion, there is a difference between its *technological* characteristics and its *perceptual* characteristics: From a technological perspective, AR systems integrate virtual imagery into a participant's view of the world. From a participant's point of perspective, AR integrates virtual content into their otherwise real surroundings.

Let us illustrate this difference between AR's technological and perceptual characteristics with an example: When Caudell and Mizell [8] coined the term "Augmented Reality" (see above), they presented a prototype where a virtual line guides an assembly worker with a wiring task. Technologically speaking, their proposed technology inserts the virtual line in the worker's *view* of the world. From a technological and objective perspective, there *is* a line in the worker's view, but there is no line in the actual world. In this sense, the device "is used to 'augment' the visual field of the user" [8, p. 660].

From a participant's perspective, however, virtual objects (ideally) appear to exist in the world, and not just in the user's view. In the case of Caudell and Mizell [8]'s system, the perceptual result is the following: "a red indicating line appears to protrude from the connector at the exact location where the current wire is to be inserted" (p. 665). From a participant's subjective perspective, there is a line in the world. In this sense, their AR system is not augmenting the worker's field of view but rather, augmenting their real surroundings.

Considering that AR ultimately aims at blending the virtual and real in a way that they appear as one environment, it only seems natural to take the participant's perspective into account and study the environments a participant experiences. If we take a participant's point of view, the virtual becomes part of the real world. Typically, this world is not just something the participant can *see*. Rather, it is a world they can smell, feel, touch, hear, taste, see and interact with. Consequently, if the goal of creating seamlessly mixed virtual-real environments is met, even visually augmented reality can engage all our senses, simply because it consists of virtual content and the multimodal real world. If we understand AR in terms of the *environments* a participant *experiences* rather than in terms of technologies that *enable* those experiences, multimodality is the norm rather than the exception [32].

Admittedly, the fact that the real world plays an important role in AR is no revelation. Many researchers point out that AR allows the participant to act in the real environment. For instance, Mackay [20] emphasises this: "The most innovative aspect of augmented reality is not the technology: it is the objective. Instead of replacing physical objects with a computer, we create systems that allow people to interact with the real world in natural ways and at the same time, benefit from enhanced capabilities from the computer" (p. 20). Likewise, Hugues, Fuchs and Nannipieri [14] point out that AR is not only something a participant can *see*, but an environment that allows for action: "we define AR by its purpose, i.e. to enable someone to create sensory-motor and cognitive activities in a new space combining the real environment and a virtual environment" (p. 47). Furthermore, the fact that AR allows the participant to act in the world has been key to AR's technological development. One of the most prominent topics in AR research are tracking techniques that make sure a virtual object's visual appearance matches the participant's current viewing perspective even if the participant moves around in the space (cf. [34]).

However, while the possibilities of interacting with the world physically are clearly acknowledged, the fact that we experience the world with all senses is often overlooked. If we approach AR from a participant's point of view, AR does not have to feature synthetic 3D sounds, in order for the resulting AR experience to include sound (cf. [2]). Also, AR not only "*might* [italics added] apply to all senses" [2, p. 361]. Rather, as combination of virtual content and the real world, even visually augmented reality *already* includes sound and applies to all our senses. When dealing with visually augmented reality, we have to remind ourselves that AR not only "allows the user to see the real world, with virtual objects superimposed upon or composited with the real world " [2, p. 356] but that it also allows the user to hear, smell, taste and feel the environment. We believe this is particularly important because what we hear, feel, smell or taste might play a role in whether we believe that a virtual object is indeed part of the real environment. We will discuss this in the following section.

4. INTERACTIVITY IN AR REVISITED

We have argued that AR is inherently multimodal, even if we are dealing with solely visual additions. So what does this mean for AR? Can we take multimodality for granted and just focus on what a participant sees? We do not think so. We believe that for virtual and real objects to appear as if they existed in the same space, it is not only important whether they coexist visually—it also matters whether they interact with each other and react to each other's visual and non-visual qualities. We expect that the multimodal qualities of the real world both pose challenges as well as open up new possibilities when it comes to blending the virtual and the real.

First of all, we believe that making virtual objects react to the multimodal properties of the real world can help to convince us that they are part of this world. Imagine, for instance, a virtual pet that gets scared when there is a sudden sound in the surroundings, a virtual objects that moves to the song playing on the radio, or a virtual character that puts on different clothes, according to the current temperature. We expect that if virtual content matches the multimodal properties of the real world, the virtual might blend in with the real world more seamlessly, ultimately enabling more holistic experiences.[1]

Furthermore, we believe that making virtual objects react to the multimodal properties of the real world would furthermore create opportunities for intuitive interactions between participants and the virtual content. If, for instance, a virtual creature senses the world, we might lure it closer with certain sounds, we might change their appearance by placing them in a colder environment or by turning on the heat, or change their behaviour by putting on a different song or by shedding light on them with a torch. Such possibilities for interacting with virtual objects might increase the participant's believe that virtual content indeed is present in *the same space* as the participant (and thus, ultimately contribute to the illusion of virtual objects existing in real environment).

What is more, virtual content can also 'sense' and react to properties of the real world that we humans can not perceive directly, such as radiation or magnetism [32]. In this way, virtual objects might help us to perceive normally unperceivable aspects of the real multimodal world. Presumably, the fact that virtual content can 'sense' the world (even in ways we can not sense it), would contribute to the feeling that it is present in this world.

At the same time, we hypothesise that neglecting the multimodal qualities of the real world may also compromise the illusion of virtual objects existing in this world. For instance, we might expect a virtual mouse to be frightened (or at least react) when it hears a miaow (or when a cat approaches it). Likewise, we would expect a virtual tree to get wet when it rains and expect virtual leaves to move if we feel that the wind is blowing [32].

Also, if virtual and real objects (appear to) exist in the same space, we might expect that virtual content affects the real world: For instance, we might expect to hear sounds if a virtual ball bounces on a real wooden floor or expect a real window to break if the ball hits it.[2]

If, however, virtual content remains oblivious to its real surroundings, a virtual leaf does not move in real wind, a virtual mouse shows no reaction to a cat's miaowing and the floor remains silent when it is hit by a virtual ball, this might harm the impression that virtual content exists in the same space—even if the virtual content is perfectly registered with the world visually. We thus believe that in the context of AR, research into multimodal interaction should not only focus on interaction between a *user* and an AR *system*, but more broadly study the possibilities for interaction between the *virtual* and the *real*.

If we look at existing AR projects, *some* forms of interaction between the virtual and the physical world are already widely considered. For instance, researchers aim for scenarios where virtual objects cause occlusions and shadows in the real world (see, e.g., [12]). In many cases, real light appears to have an effect on virtual objects and virtual objects can cast shadows on the real world and thus seemingly affect the real world in return. Furthermore, scientists include physics simulations to make it look as if virtual objects adhered to physical laws (e.g., [16]). Quite commonly, the gravitational force of the real world seems to affect virtual objects and their virtual mass. After all, virtual elements usually do not float around in space, but instead, often appear to stand on or lie on real objects. In addition, quite some virtual characters seem to interact with the elements in and geometry of their real surroundings more actively. For instance, Corbett-Davies et al. [10] have presented a virtual spider that responds to the real environment and that clambers over obstacles, disappears behind real objects and can be carried around and occluded by the user's hand.

However, while many projects simulate interactions between the virtual and the real, the majority of these existing projects focus on the *visual* effects of these interactions: We can see a virtual and real objects collide realistically, but not hear their collision. Likewise, we can see a virtual

[1]On a practical note, sensing multimodal properties often seems easier than simulating them. For instance, a weather measurement of the real surroundings can easily be taken (or obtained from the internet). In contrast, making it seem as if it were hotter or colder is rather complex. If we want to make sure that the virtual blends in with its real surroundings, it thus might be easier to sense the surroundings and adjust the virtual, rather than the other way around.

[2]It is rather difficult to tell what exactly people expect when dealing with virtual objects. For instance, people might not expect to be able to sit on virtual chairs. In this sense, our expectations for virtual objects might differ from expectations we have for real objects. Yet, we are likely to dodge if a virtual ball comes at us. Here, we built on the assumption that virtual objects can evoke the same intuitive expectations as real objects.

spider walking over our hand, but do not actually feel it. We believe that if virtual objects do not react to and affect their surroundings multimodally, it might remind us of the fact that they are not really here. Hence, we believe that we have to approach interactions between the virtual and the real from a modalities-encompassing perspective (cf. [31]).

Finally, whereas many projects include some form of interaction, much existing research is concerned with imitating interactions from the real world. In our opinion, this is not necessary. In fact, we believe it would be much more interesting to explore new forms of interaction that have no equivalent in a purely physical world. We have previously realised an example where virtual objects are physically attracted to and consequently stick to real objects of similar colours [31]. Although no such influences exist between real objects, we found this color-based attraction very convincing. Likewise, we expect that virtual objects can react to non-visual qualities of the world in novel ways. We thus believe that instead of focusing on *realistic* interactions, we should more broadly focus on *believable* interactions.

5. DISCUSSION AND CONCLUSION

AR is often understood in terms of technologies that overlay virtual imagery onto a user's view of the world. However, we believe this common technology- and vision-focused understanding of AR is incomplete. Most importantly, it only describes what what a system does and displays, but disregards what the participant can do and perceive. Considering that AR creates mixed reality environments for humans to engage with, it only seems natural to take the participant's perspective into account and study the AR environments they experience.

If we approach AR from this perspective, it quickly becomes apparent that AR is about more than what a participant can see. In particular, we have identified three main reasons to approach AR from a modalities-encompassing perspective.

First, virtual content can take non-visual and multimodal forms. Our review of existing work on multimodal AR has shown that AR's underlying goal of creating mixed environments can be achieved by combining non-visual virtual content with the real world. Hence, we have to approach the virtual from a multimodal perspective. We expect that making use of multimodal stimuli rather than solely visual stimuli can improve the feeling that a virtual object actually exists—especially in cases where the the virtual represents an existing, multimodal object.

Second, AR experiences include our experience of the real multimodal world. In this sense, multimodality in AR is the norm rather than the exception. (The fact that the real world plays a crucial role in the resulting experience sets AR technologies apart from other types of interactive systems. For instance, we might use mobile devices and desktop computers to check our mail, read the news, watch a movie or play a game. Here, too, we might experience virtual content alongside our real surroundings. However, unlike AR, these systems offer a second, *independent* and digital space for the user to enter. In contrast, AR technologies explicitly aim at making the real, interactive and multimodal environment part of the experience. Consequently, our line of reasoning does not apply to any kind of interactive system. In particular, our argumentation does not apply to the related field of Virtual Reality. This is because here the virtual world tries to replace the real world rather than supplement it (cf., e.g., [2]).

Third, the illusion of virtual and real objects being part of the same space might be affected by whether virtual and real objects influence one another—both in visual and in non-visual ways. Consequently, we should broaden the traditional understanding of multimodal interaction, which focuses on the interaction between a user and a system and also focus on the multimodal influences between the virtual content and its real surroundings.

We expect that realising such multimodal influences between the virtual and the real holds many possibilities: First of all, making virtual objects react to the multimodal properties of the real world and vice versa might help to convince us that virtual objects are actually in the same space as we are. Furthermore, such influences might make the experience more entertaining and make non-visual and multimodal aspects of the experience more prominent, even without producing multimodal output. What is more, multimodal influences between the virtual and the real might enable new forms of interaction between a participant and the virtual content as well as interactions that can not exist in a purely physical world. Finally, virtual objects might react to properties of the real world we can not sense ourselves, and thus, allow us to experience more about the world.

Approaching AR from a experience-focused perspective, we come to the conclusion that even visually augmented reality might enable multimodal experiences, simply because AR includes our experience of the real world. Hence, we believe that multimodal AR is the norm rather than the exception.

So far, this conclusion is based on the fact that AR aims at making it seem as if virtual content existed in our real, everyday world and on the assumption that AR actually achieves this goal. In the future, it would be interesting to conduct experiments that address this issue from an empirical perspective and incorporate presence research that studies when and why we experience objects as 'being here'. We expect that multimodal qualities of both the virtual and the real as well as multimodal interactions between the virtual and the real play a role in this. However, we also expect that there are other important factors that have received little attention so far, such as the underlying purpose, goal and contextual reasons why virtual content is integrated into the surroundings.

In closing, let us return to the question of whether we can experience AR with closed eyes. In our opinion, this is possible but differs from application to application. Ultimately, the question boils down to whether virtual content is experienced as part of the real environment, even if we close our eyes. Creating experiences where this is the case certainly seems a worthy goal.

6. ACKNOWLEDGEMENTS

We thank the reviewers for their valuable input and feedback.

7. REFERENCES

[1] *Augmented Reality.* Oxford English Dictionary, 2005.

[2] R. Azuma. A Survey of Augmented Reality. *Presence*, 6(4):355–385, 1997.

[3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in

augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, November 2001.

[4] O. Bau and I. Poupyrev. Revel: tactile feedback technology for augmented reality. *ACM Transactions on Graphics (TOG)*, 31(4):89, 2012.

[5] G. Bianchi, B. Knoerlein, G. Szekely, and M. Harders. High precision augmented reality haptics. In *Proc. EuroHaptics*, volume 6, pages 169–178. Citeseer, 2006.

[6] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Foundations and Trends in Human-Computer Interaction*, 8(2-3):73–272, 2015.

[7] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. FingARtips: Gesture Based Direct Manipulation in Augmented Reality. In *Proc. GRAPHITE '04*, pages 212–221. ACM, 2004.

[8] T. Caudell and D. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proc. of the Twenty-Fifth Hawaii International Conference on System Sciences*, volume 2, pages 659–669. IEEE, Jan 1992.

[9] T. Chatzidimitris, D. Gavalas, and D. Michael. Soundpacman: Audio augmented reality in location-based games. In *Proc. MELECON 2016*, pages 1–6. IEEE, 2016.

[10] S. Corbett-Davies, A. Dünser, and A. Clark. An interactive augmented reality system for exposure treatment. In *Proc. ISMAR-AMH 2012*, pages 95–96. IEEE, 2012.

[11] S. Doyle, M. Dodge, and A. Smith. The potential of web-based mapping and virtual reality technologies for modelling urban environments. *Computers, Environment and Urban Systems*, 22(2):137–155, 1998.

[12] S. Gibson and A. Chalmers. Photorealistic augmented reality. In *Eurographics 2003 Tutorial*, Granada, Spain, September 2003.

[13] F. Heller and J. Borchers. Corona: Audio augmented reality in historic sites. In M. de Sa, E. F. Churchill, and K. Isbister, editors, *MobileHCI 2011 Workshop on Mobile Augmented Reality: Design Issues and Opportunities*, pages 51–54, Stockholm, Sweden, August 2011.

[14] O. Hugues, P. Fuchs, and O. Nannipieri. New Augmented Reality Taxonomy: Technologies and Features of Augmented Environment. In B. Furht, editor, *Handbook of Augmented Reality*, pages 47–63. Springer New York, 2011.

[15] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko. "Move the couch where?": developing an augmented reality multimodal interface. In *Proc. ISMAR 2006*, pages 183–186. IEEE, 2006.

[16] S. Kim, Y. Kim, and S.-H. Lee. On Visual Artifacts of Physics Simulation in Augmented Reality Environment. In *Proc. ISUVR '11*, pages 25–28, Washington, DC, USA, 2011. IEEE Computer Society.

[17] N. Koizumi, H. Tanaka, Y. Uema, and M. Inami. Chewing jockey: augmented food texture by using sound based on the cross-modal effect. In *Proc. ACE 2011*, page 21. ACM, 2011.

[18] R. W. Lindeman, G. Lee, L. Beattie, H. Gamper, R. Pathinarupothi, and A. Akhilesh. Geoboids: A mobile ar application for exergaming. In *Proc. ISMAR-AMH 2012*, pages 93–94. IEEE, 2012.

[19] R. W. Lindeman and H. Noma. A classification scheme for multi-sensory augmented reality. In *Proc. VRST 2007*, pages 175–178. ACM, 2007.

[20] W. E. Mackay. Augmenting reality: A new paradigm for interacting with computers. *La recherche*, 284:1–9, March 1996.

[21] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, E77-D(12):1321–1329, 1994.

[22] K. Minamizawa, S. Fukamachi, H. Kajimoto, N. Kawakami, and S. Tachi. Gravity Grabber: Wearable Haptic Display to Present Virtual Mass Sensation. In *ACM SIGGRAPH 2007*, New York, NY, USA, 2007. ACM.

[23] T. Narumi, T. Kajinami, T. Tanikawa, and M. Hirose. Meta cookie. In *ACM SIGGRAPH 2010 Posters*, page 143. ACM, 2010.

[24] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose. Meta cookie+: an illusion-based gustatory display. In *Virtual and Mixed Reality-New Trends*, pages 260–269. Springer, 2011.

[25] T. Narumi, M. Sato, T. Tanikawa, and M. Hirose. Evaluating cross-sensory perception of superimposing virtual color onto real drink: toward realization of pseudo-gustatory displays. In *Proc. AH '10*, page 18. ACM, 2010b.

[26] A. Olwal, H. Benko, and S. Feiner. Senseshapes: Using statistical geometry for object selection in a multimodal augmented reality system. In *Proc. ISMAR 2003*, page 300. IEEE Computer Society, 2003.

[27] W. Piekarski and B. H. Thomas. ARQuake: the outdoor augmented reality gaming system. *Communications of the ACM*, 45(1):36–38, 2002.

[28] H. T. Regenbrecht and M. T. Wagner. Interaction in a collaborative augmented reality environment. In *Proc. CHI'02 Extended Abstracts*, pages 504–505. ACM, 2002.

[29] D. Reiners, D. Stricker, G. Klinker, and S. Müller. Augmented reality for construction tasks: Doorlock assembly. *Proc. IWAR '98*, 98(1):31–46, November 1998.

[30] J. M. Rozier. *Hear&there: An augmented reality system of linked audio*. PhD thesis, Massachusetts Institute of Technology, 2000.

[31] H. Schraffenberger and E. van der Heide. From Coexistence to Interaction: Influences Between the Virtual and the Real in Augmented Reality. In K. Cleland, L. Fisher, and R. Harley, editors, *Proc. ISEA 2013*, pages 1–3, Sydney, 2013a.

[32] H. Schraffenberger and E. van der Heide. Everything augmented: On the real in augmented reality. *Journal of Science and Technology of the Arts*, 6(1):17–29, 2014.

[33] J. R. Vallino. *Interactive augmented reality*. PhD thesis, University of Rochester, Rochester, NY, USA, 1998.

[34] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. In *Proc. ISMAR 2008*, pages 193–202, Washington, DC, USA, 2008. IEEE Computer Society.