Arguably augmented reality : relationships between the virtual and the real

Schraffenberger, H.K.

Citation

Schraffenberger, H. K. (2018, November 29). *Arguably augmented reality : relationships between the virtual and the real*. Retrieved from https://hdl.handle.net/1887/67292

Author: Shraffenberger, H.K.
Title: Arguably augmented reality : relationships between the virtual and the real
Issue Date: 2018-11-29

3 New Perspectives

The previous chapter has revealed three prevailing ideas about the nature and characteristics of augmented reality. First, AR is commonly seen as a *technology*. Secondly, AR is often understood in terms of *visual* additions that are overlaid onto our *view* of the real world. Thirdly, AR is generally considered to spatially integrate this virtual content in the real world by aligning virtual and real content with each other in 3D. All three ideas contribute to the widespread notion of AR as a technology that integrates virtual imagery into our view of the real world (see, e.g. *Augmented Reality* 2005; Reiners et al., 1998; Zhou et al., 2008). There is no doubt that such technologies play an important role in the context of augmented reality. Yet, in our opinion, such common understandings of AR are incomplete and unnecessarily limit the AR research field. In this chapter, we challenge the focus on technology, the need for registration as well as the emphasis on vision. We address shortcomings in prevailing definitions and propose alternative perspectives on AR. The proposed shifts in perspective are outlined below and subsequently discussed in detail.

The first issue with prevailing notions is their focus on AR as a *technology*. Generally speaking, technology-based definitions inform us about what an AR system does but do not reveal much about the AR environments they create and the AR experiences they evoke in the participant. Yet, the underlying purpose of AR technologies is to allow participants to experience augmented environments. Considering this, it only seems natural to also take the participant's experience into account and explore the augmented environments that they perceive. In our opinion, what a system does and whether it fits a given definition is less important than whether it evokes the intended experience. We thus believe we need to take an environment- and experience- oriented perspective. We will discuss this shift in perspective and address both the workings of typical AR systems as well as the experiences they facilitate in section 3.1.

The second issue with common notions of AR is their focus on the alignment of virtual content with the real world in three dimensions and in real-time. There is no doubt that this so-called registration process plays an important role in creating the impression of virtual objects existing in real space. However, in our opinion, there are three reasons to look beyond registration. First, virtual objects can seemingly appear to exist in real space, even if they are not aligned with the real world in 3D. Second and more fundamentally, virtual content can be part of, enhance and augment the real world even if it does not seem to exist in the physical space. For instance, an audio guide can augment our experience of an exhibition without seemingly existing in the museum space. In our opinion, this means that registration is not always necessary for creating AR experiences. Third, registration might not always be sufficient to evoke AR experience. For instance, when attempting to display a virtual ball in real space, it might matter whether this ball appears to be affected by real light sources and whether the ball moves when it is hit by a real object. It is easy to imagine that a lack of interactions between the real world and virtual objects can harm AR experiences and make virtual objects look "out of place" even when they are spatially registered with the world. We thus believe that other links between the virtual and the real aside from spatial registration need to be considered in the context of AR. We follow this line of thought in section 3.2. We propose that instead of defining AR in terms of registration between the virtual and the real on a technological level, to define it in terms of a relationship between the virtual and the real on an experiential level.

The third concern that applies to many common notions of AR is the emphasis on vision. As we have seen, many existing views approach AR in terms of visual imagery that is overlaid onto a participant's view. We see three main issues with this. First of all, AR environments are not just something the participant can see. Rather, they are environments that participants can perceive with all their senses, act in and interact with. Arguably, AR is inherently multimodal and interactive because AR environments include the multimodal and interactive real environment. A second reason to look beyond vision is that virtual content, too, can take non-visual and multimodal forms. In our opinion, there is no good reason to exclude non-visual virtual content from the domain of AR. Last but not least, a multimodal perspective is important because of the way our human perception works: Even if visual information is added to our view of the world, this information can affect how we perceive non-visual qualities of the real world. For instance, visual information can alter how a physical object feels. (This effect is called cross-modal interaction.) If we only consider a participant's view of the world, such effects will remain unnoticed. Based on these arguments, we propose to approach AR as a multimodal and interactive environment rather than as a visual phenomenon. Section 3.3 presents this move from a vision-focused view towards a multimodal perspective in detail.

We synthesize and discuss these three views in section 3.4. We propose to define AR in terms of interactive and multimodal environments where a participant experiences a relationship between virtual content and the real world. Our proposed view of AR departs from common understandings of AR in three ways: (1) it focuses on the AR environments and experiences rather than on AR technologies (2) it argues that AR is based on relationships between the virtual and the real rather than on interactive/real-time 3D registration (3) it treats AR as an interactive and multimodal rather than visual phenomenon.

Although we present these three points one by one, they are related and interdependent. For instance, our idea of AR experiences without the use of traditional AR technologies is supported by projects that make use of non-visual forms of virtual content. E.g., we can find examples of classical AR experiences that are realized with simple iPods or MP3 players in the context of sound-based AR. At the same time, the possibility of working with non-visual information, such as tastes, challenges the need for registering information with the surrounding world in 3D. After all, taste is not something we experience in three-dimensions and in the surrounding world, but something we experience in our mouth. At the same time, the move towards an experience-based view suggests that we should let go of the focus on 3D registration. In this way, the different points work together, support each other and build upon each other.

Although we challenge prevailing views, we do not mean to critique them on an individual level. For instance, the view of AR as a technology can make sense in an engineering context. Similarly, the claim that AR technology overlays virtual images onto a user's view makes sense in the context of a project that works with visual overlays. It is only natural that many authors describe AR from the perspective of their own domain and emphasize forms of AR that are relevant in their own research. Our notion of AR is meant to provide an additional, complementary perspective from which we can study and explore AR. It is not meant to replace other perspectives altogether.

3.1 *From Technologies to Experiences*

As we have seen in the previous chapter, AR is often seen as a technology or system. Most prominently, AR is considered an interactive system that combines and aligns the virtual and the real in 3D and in real-time (Azuma, 1997). But what is the point of such an AR system? What is its purpose and what is in it for the user? This section addresses these questions.

3.1.1 *The Goal of AR Technologies*

Why do AR technologies exist? What is their purpose and what goals do they serve? A look at existing research reveals some common answers: AR technologies aim at creating the illusion of virtual objects existing in the real world, and more generally, try to make it appear as if the virtual world and the real surroundings were one seamless environment. For instance, Vallino (1998) states that "[t]he goal of augmented reality systems is to combine the interactive real world with an interactive computer-generated world in such a way that they appear as one environment" (p. 1). Furthermore, Buchmann et al. (2004) propose that "[t]he goal is to blend reality and virtuality in a seamless manner" (p. 212). Billinghurst, Clark, et al. (2015), who survey almost 50 years of AR research and development, similarly state: "From early research in the 1960's until widespread availability by the 2010's there has been steady progress towards the goal of being able to seamlessly combine real and virtual worlds" (p. 73). More specifically, AR systems are commonly used to create scenarios where virtual objects appear to exist in real, physical space. E.g., Regenbrecht and Wagner (2002) state that "[t]he goal is to create the impression that the virtual objects are part of the real environment " (p. 504). Likewise, Azuma (1997, p. 356) mentions that "[i]deally, it would appear to the user that the virtual and real objects coexisted in the same space " (p.356).[1]

If we look at the AR landscape, indeed many so-called AR applications present us with virtual objects that seemingly exist in our otherwise real surroundings. To mention just a few examples: The IKEA Place app allows us to see virtual furniture in our physical environment (*IKEA Place* 2017). Likewise, the HoloLens by Microsoft (n.d.) seemingly fills our living rooms with visual virtual building blocks. Similarly, the app *Sphero* (2011) turns a robot ball into a visual virtual beaver that seemingly exists in our everyday surroundings. An example of the latter is shown in figure 3.1. This screenshot shows the little virtual beaver *Sphero* (2011), as seen through an iPad.

Figure 3.1: The virtual beaver *Sphero* (2011) is not just overlaid onto our view but integrated into our view. The picture is a screenshot showing the image displayed on the iPad. (The screenshot was taken by the author.)

As this example shows, the virtual content appears to exist in the space around us—the beaver seems to be standing on the authors living room floor, looking at the author's cat.[2]

In the following, we will discuss how AR systems achieve such effects. This look at the workings of AR technology is necessary for two

---

[1] Note that Azuma is using the word 'coexist' differently from how we use it. With coexist, we emphasize that there is no relationship between two things and that they exist independently. Azuma uses the term to refers to things that appear to exist in the same space, which implies a spatial relationship.

[2] In AR literature, we often find claims that virtual content is (a) overlaid onto our view or (b) integrated into our view. This difference can be explained with the fact, that technically speaking the content is often overlaid. At the same time, however, it is also aligned with the real world in three dimensions and appears to exist in the space. In this sense, is integrated into the view.

reasons. First, it will allow us to better understand common technological views on AR. Second, knowing how typical AR systems work allows us to show that different, alternative technologies can also be used to create AR experiences.

3.1.2 *How (Visual) AR Technologies Work*

AR systems can make it seem as if virtual objects were present in the real world. How does this work? Simply put, a typical AR system senses the participant's position in the world and consequently computes how a virtual object has to be presented so that it appears to exist in the real world. Once the virtual image is computed, it is displayed to the participant, e.g., on a head-mounted or hand-held display.

The Registration Problem

The process of giving a virtual object a position in the real space is called registration, and according to common notions (see section 2.1), characterizes AR. In his book *Understanding Augmented Reality: Concepts and Applications*, Craig (2013, p.17) explains registration like this:

> A key element to augmented reality rests with the idea of spatial registration. That is, the information has a physical space or location in the real world just like a physical counterpart to the digital information would have.

As discussed in subsection 2.1.3, registration is a common process in image processing, where it refers to the process of "transforming different sets of data into one coordinate system" (Rani and Sharma, 2013, p. 288). In the context of AR, registration typically refers to the alignment of virtual and real content. Strictly speaking, descriptions of this process vary slightly. For instance, Drascic and Milgram (1996) use registration to refer to the alignment of "the coordinate system of the virtual world with that of the real world" (p. 129). In addition, registration is also understood as aligning virtual and real objects with respect to each other (Azuma et al., 2001). However, in the end, registration always refers to a process that makes sure that virtual content has a position in the real world.

The challenge of properly aligning the virtual and the real is commonly referred to as "the registration problem" (e.g., Azuma, 1997), and regarded one of the key issues in AR research (e.g., Azuma, 1997; Bimber and Raskar, 2005; You and Neumann, 2001). Accurate alignment is considered important because improper registration of virtual and real objects can cause virtual objects to appear as if they existed separately from the real world, rather than in the real world. In other words, improper registration can compromise or break the illusion of virtual objects existing in real space (cf., e.g., Azuma, 1997; Bajura and Neumann, 1995; Vallino and C. Brown, 1999). In addition to breaking the illusion of virtual objects existing in real space altogether, inaccurate alignment by an AR system might cause virtual objects to appear at a *wrong* position in real space. For instance, (Azuma, 1997) suggests that inaccurate registration could cause a virtual pointer to appear at an incorrect position: "[...] many applications *demand* accurate registration. For example, recall the needle biopsy application. If the virtual object is not where the real tumor is, the surgeon will miss the tumor and the biopsy will fail." (p. 367, italics in original). Likewise, Bajura and Neumann (1995) explain that "[i]f accurate registration is not maintained, the computer-generated objects appear to float around in the user's natural environment without having a specific 3D spatial position" (p. 52).

The effect of improper registration can, for instance, be seen when playing the game Pokémon GO. Here, virtual creatures often appear in unrealistic positions in the environment or look like an independent overlay that floats on top of the camera feed, rather than as part of the environment. Screenshots of such moments are presented in figure 3.2.

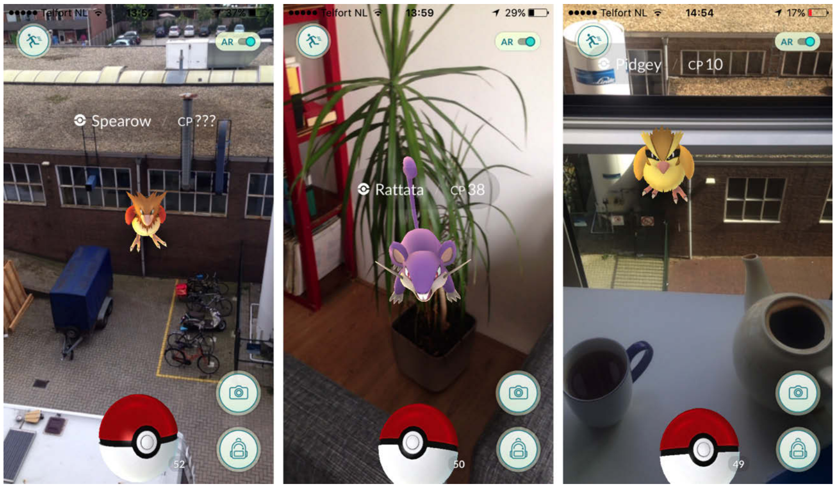


Figure 3.2: Due to inaccurate registration, virtual Pokémon creatures can appear at unrealistic positions or overlaid onto the live view, rather than as part of the real environment. (In order to amplify this effect, the author has manually moved the phone in space. However, Pokémon quite regularly appear 'detached' from the real world without trying to achieve this.)

Proper registration is particularly difficult because participants can move through AR environments and experience the world from different perspectives. For instance, we do not want a virtual cup of coffee to move in space, simply because we are moving our head. Also, if we stand up and look at the cup from above, we expect to see it from this particular perspective and, e.g., expect to see the cup's contents. Simply put: the virtual information has to dynamically adapt to our movement and perspective, in order to continuously appear correctly positioned in the real world.[3] What is more, for a virtual object to appear on top of, inside of, behind or otherwise related to a real object, the AR system needs to know the position of such real objects (Azuma et al., 2001).

---

[3] The possible movement of the participant also explains why many definitions (e.g., Azuma, 1997; Azuma et al., 2001) not only require registration but also point out that the AR system has to work interactively and in real-time.

TRACKING

In order to accurately respond to changes in the participant's location and orientation, a variety of so-called tracking technologies are used to keep track of the participant's position in the real world (or of the position of a mobile device, through which the participant perceives the augmented environment). Often, computer-vision-based systems are used to determine the position and orientation of a participant (or of the intermediate device) (Craig, 2013). These systems make use of cameras in order to sense the world. Based on what the camera 'sees', software determines where the camera must be located and how it must be oriented in order to obtain this view of the world. For this to work, the environment must contain some cues that the software can recognize. These cues can take many forms. In the early days of AR, the cues typical took the form of so-called "fiducial markers" (see figure 3.3), which were physically integrated into the environment and specifically designed so that computers could easily recognize them. Currently, however, many efforts are put into markerless tracking and into using natural features of the environment, such as buildings and objects, as cues.[4]

Computer-vision-based tracking has the advantage that it is rather precise. Furthermore, the software can not only keep track of the location of the participant but also recognize and track objects of interest. As a result, virtual content can be positioned relative to real objects. For instance, a computer-vision-based system might be able to recognize a vase and display a flower in it. (Because of this, the stem of the flower can be hidden by the real vase. Furthermore, the flower can remain in the vase, even if the vase moves in space.)

Figure 3.3: Three typical fiducial markers that can be recognized by AR software, such as the popular open-source ARToolKit tracking library. (The displayed markers are part of the download of the *ARToolKit SDK* (1999).)

Another common approach to tracking (and ultimately, registration) is the use of positioning systems, such as GPS (Global Positioning System) in order to obtain the location of the participant (or the location of the used device) in 3D, in combination with a compass, gyroscope and accelerometer to ultimately determine all six degrees of freedom of the participant.[5] This approach has the advantage that the required technologies are currently widely available, and integrated in many smartphones. Unfortunately, such smartphone-based solutions often also have several disadvantages. First of all, they can suffer from poor

---

[4] AR without markers is also referred to as markerless AR. The use of natural features for tracking is referred to as natural feature tracking (NFT). This concept can also be used to recognize magazine pages, photographs, posters or products and ultimately display virtual information on top of them. In these cases, the line between marker-based and markerless AR is blurry. For instance, a photograph can act both as an object that is augmented, as well as serve as a marker that is added to a scene to allow for tracking. Hence, NFT and markerless tracking overlap, but are not the same.

[5] Six degrees of freedom (6DoF) refers to the six ways a rigid body can move in three-dimensional space. The possible movements include three ways of changing the *location*: (1) surging (moving forward and backward on the X-axis), (2) swaying (moving left and right on the Y-axis) and (3) heaving (moving up and down on the Z-axis) and three ways of changing the *orientation*: (4) rolling (tilting side to side on the X-axis), pitching (tilting forward and backward on the Y-axis), and yawing (turning left and right on the Z-axis) (Six degrees of freedom, n.d.).

accuracy. For instance, Blum et al. (2012) compared the accuracy of the orientation and location of iPhone 4, iPhone 4s and Samsung Galaxy Nexus phones. They found mean location errors of 10-30 meter as well as mean compass errors around 10-30°, both with high standard deviations that, according to the authors, render them unreliable in many settings. Also, GPS is especially unreliable indoors and in urban areas, where GPS signals can be blocked by high buildings (Cui and Ge, 2003). Furthermore, because the camera image is not analyzed, the application has no knowledge about the spatial structure of the physical environment and cannot track other objects of interest. As a result, they cannot be used to align virtual content with respect to a real object. In other words, GPS-based solutions are fine for displaying a virtual bird in the real sky, but not for showing a virtual flower in a physical vase (the accuracy would be too low), especially if this vase can be moved around (the system would not be able to recognize the vase and track its movement).

In addition to computer-vision and GPS-based approaches to tracking, other possibilities exist. For instance, the AR system by Feiner, Macintyre, et al. (1993), which helps with the maintenance of an office printer, make use of ultrasonic transmitters and receivers mounted on both the participant's head and on the printer in order to determine the spatial relationship between the two. Furthermore, many applications combine several different methods and sensors in order to obtain better (more accurate) results. E.g., Persa (2006) use a firewire webcam and a GPS receiver in combination with a radio data receiver to obtain position and orientation information. Similarly, the PhD thesis by Caarls (2009) focuses on fusing information from various sensors with different accuracies, update rates, and delays to address the challenge of real-time pose estimation of a user's eyes.

Computing Virtual Output

Once positioning data is obtained, the AR system typically uses this information to compute (or, in the case of images "render") the corresponding virtual output. If you are looking at my desk with an AR device, the information can, e.g., be used to compute a believable image of a virtual cup of coffee on my desk. If you are looking at the desk straight from above, the rim of the computed cup will have a circular shape. If you change your perspective slightly, the rim will have an elliptical shape. Ideally, the appearance of a virtual object changes depending on the participant's perspective, just like the appearance of real objects varies when one changes one's point of view. Commonly, this computed content takes a visual form.[6]

Display

As soon as the corresponding virtual output is computed, it is pre-

---

[6] However, as we emphasize throughout this thesis, virtual content can also take non-visual forms. For instance, the song of a bird could be synthesized in a way that it becomes louder if the participant gets closer and in a way that the song appears to originate from the same tree, even when the participant turns around and changes their orientation.

sented to the user. Most often, AR systems present virtual content in real space by means of a head-worn or hand-held display. However, other possibilities exist. Virtual content can, e.g., also be embedded into the world directly with projectors or flat panel displays. For instance, Benko et al. (2014), use three projectors to allow two participants to see virtual content in the real environment, and, for instance, toss a virtual (projected) ball back and forth through the space between them (see figure 3.4) Such forms of AR where virtual content is directly embedded into the real world is typically referred to as spatially augmented reality (Raskar, Welch, and Fuchs, 1998) or spatial augmented reality (Bimber and Raskar, 2005). Furthermore, in addition to visual displays, also other types of stimuli are sometimes used to convey the presence of virtual objects in real space. E.g., the Sound-Pacman game by Chatzidimitris et al. (2016) makes use of synthesized 3D sound played back on headphones in order to give virtual ghosts a position in the real physical environment and communicate their location to the player.

Figure 3.4: The projection-based AR project by Benko et al. (2014) can make it seem as if virtual objects existed in real space, rather than projected onto the world. The image shows two screenshots from the YouTube video about this project (Microsoft Research, 2014).

In order to accurately register the virtual and the real even when a participant moves, these processes have to happen in real-time and with very little latency. If the registration process takes too long, the delay can cause registration errors. For instance, if you were to turn your head very fast, the virtual cup of coffee on my desk might not be able to keep up with you. In the time it would take the system to figure out your perspective and compute and display the cup of coffee, your perspective would have already changed so much that the resulting output would no longer match your view. Simply put, virtual content has to appear at the right position at the right time. This is why it is sometimes stated that an AR system has to operate interactively and *in real-time* (e.g., Azuma, 1997), and that the virtual not only has to be registered with the real world spatially but also *temporally* (e.g., Craig, 2013).

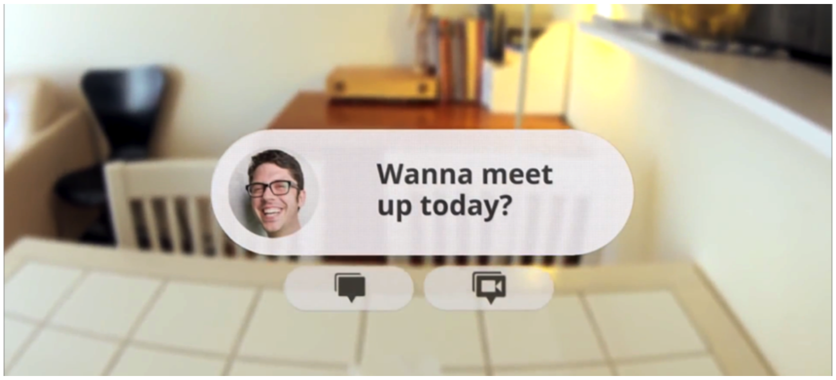
The Greatest Common Factor

As the examples above indicate, AR systems take many different forms. They can, e.g., present various forms of virtual content (e.g., visual or auditory content) and use different information displays to convey this information (e.g., screens, projectors or headphones).

Furthermore, displays can be placed in the environment statically or carried by the user (and in the latter case, can be head-mounted or hand-held). Different setups go hand in hand with different system requirements. For instance, tracking the participant's position is not necessary in cases where virtual content is projected onto the real world directly with a projector in order to change *surface* attributes of physical objects, such as their texture or color because here the rendering is independent of the viewers position (Raskar, Welch, and Chen, 1999).

Although AR systems differ, they generally make use of a computer system that registers the virtual with the real world interactively, in real-time and in three dimensions (Azuma, 1997). In the following, we refer to this type of technology as *traditional AR technology* or *traditional AR systems*.

Given the common goal of making it seem as if virtual objects existed in real space, defining AR in terms of traditional AR systems can seem like a natural choice. After all, traditional AR systems can enable this illusion. More than that, without an AR system that registers the virtual and the real, virtual content typically appears to exist independently from its real surroundings as opposed to as part of the world. Without registration, a virtual character might, e.g., appear on a screen, a voice might appear "on a sound recording", or a text might simply overlay what we see—rather than seemingly exist in the real surroundings. This happens, for instance, with the virtual overlays presented by the Google Glass device (see figure 3.5). The overlays are not registered with the real world in 3D, and appear *on top of our view*, rather than *integrated into* space.



Wanna meet up today?

Figure 3.5: A mock-up of the Google Glass concept. Virtual content is overlaid onto the view of the real world but not registered with the real space. This image is a screenshot from a video demonstrating the concept behind Google Glass (Huzaifah Bhutto, 2012). The actual realization of the overlays looks quite a bit different and can be seen in figure 3.8.

If typical AR systems create the desired illusion of virtual objects existing in real space while other types of systems do not create this illusion, why not define AR in terms of typical AR systems? In our opinion, there are two answers to this question. First, *alternative AR technologies exist*: While rare, different types of technologies can also make it seem as if virtual objects existed in real space. In other words, the assumption that other types of systems cannot create the desired AR experience is wrong. Second, *alternative AR experiences exist*: Although this illusion is commonly desired, virtual content does not have to seemingly exist in real space in order to contribute to, enhance or otherwise augment this environment. We will demonstrate the first point in the following, and pick up the second point in section 3.2.

3.1.3 Typical AR Experiences With Alternative Technologies

There is no doubt that interactive AR systems that align virtual and real content in 3D can make it look as if virtual objects were part of real space and merge virtual and real worlds. However, if we only understand AR in terms of traditional AR systems, we miss one crucial aspect: other types of technologies likewise can create the desired effect. In the following, we will discuss three examples that illustrate that we do not need a typical AR system to blend the virtual and the real and to make virtual objects appear in real space. Next to visually augmented reality, we will also consider sound-based forms of AR. We do this because different types of virtual content might blend in with the real world in different ways.

Forest Walk

Early examples of AR experiences and environments that work without the use of traditional AR systems include Janet Cardiff's audio walks (Cardiff, n.d.), such as *Forest walk*.[7] *Forest walk* can be described as a "soundtrack" to the real world, specifically recorded and mixed for a pre-determined walking route. The track includes multiple layers of recordings, such as the sounds of Cardiff walking in the forest, her footsteps, the sound of her hand brushing tree bark, the sounds of the forest, such as crows, voices and in particular, Cardiff's voice, talking about the environment, giving walking instructions and describing her surroundings. For instance, one can hear Cardiff say "Go towards the brownish green garbage can. Then there's a trail off to your right. Take the trail, it's overgrown a bit. There's an eaten-out dead tree. Looks like ants." (Cardiff, 1991), while navigating the particular environment Cardiff is talking about.

One thing that makes Cardiff's recordings special is that her virtual soundscape relates to the real environment. This relationship happens on several levels: For one, Cardiff's recordings describe the real space. Instructions such as "Ok, there's a fork in the path, take the trail to the right." refer to the real surroundings and lead the way. Furthermore, the used sounds have been recorded on the same site where they are later on experienced by the participant. Consequently, the recorded sounds are similar to the real surrounding soundscape. According to Cardiff, this similarity is important for the soundscape to mix in with the real environment. As Cardiff herself puts it: "The virtual recorded

---

[7] Cardiff is neither the only nor the first artist to work with audio walks. For instance, Celia Erens, a sound artist from the Netherlands, has realized a series of works that present pre-recorded 3D soundscapes in the real sound environment. Also, "Forest Walk" is not the only walk by Cardiff that illustrates our point. However, as it is the first in Cardiff's series of audio walks, and, unlike Erens' work, also includes spoken text and instructions, we have chosen this particular example.

soundscape has to mimic the real physical one in order to create a new world as a seamless combination of the two" (Cardiff, n.d.).

Another aspect that characterizes Cardiff's soundscape, is that the used sounds have been recorded in binaural audio. Binaural audio is a recording technique that captures the spatial characteristics of the sound in 3D and consequently provides a 3D audio experience (rather than the usual stereo distribution of the sound) when the recording is played back on headphones.[8] Binaural audio often results in a very realistic impression. To quote Cardiff: "it is almost as if the recorded events were taking place live" (Cardiff, n.d.). Cardiff mixes her main walking track with several layers of sound effects, music, and voices, in order to create "a 3D sphere of sound" (Cardiff, n.d.). Judging from Cardiff's descriptions and our own experience with binaural audio, the pre-recorded sounds appear to originate in the real environment.[9]

So what does this work have to do with AR? Little, if we take a conventional, technology-based perspective on AR. Instead of using an AR system, Cardiff's work makes use of a simple CD player (or iPod/MP3 player). There is no system that aligns or registers the virtual sound sources in real three-dimensional space.[10] Instead, the sounds are placed in the space more loosely: The participant is told where to start the walk and press play. Also, the audio mix includes instructions that tell the participant where to go and that guide their attention. Indirectly, these instructions affect the participant's position in and movement through the environment, and consequently, also roughly determine where the virtual sound sources appear in space.[11] However, although Cardiff's walk does not make use of typical AR technology, it yet shares fundamental similarities with typical AR projects: It allows us to experience the real environment, supplemented with virtual content. More than that, it makes us experience a seamless, mixed, partially virtual, partially real environment. It is such a seamless combination of the virtual and the real, which is commonly considered to be the *goal* of AR (cf. section 3.1).

MOZZIES

Another application that makes virtual objects appear in real space without a traditional AR system is the early mobile game *Mozzies*. This game was installed on the Siemens SX1 cell phone that launched in 2003 (López et al., 2014). The mobile application used to show flying mosquitos, overlaid on the live image of the environment captured by the phone's camera. Players could shoot the virtual mosquitoes by moving the phone and pressing a button when aiming correctly (Siemens SX1, n.d.). In contrast to Cardiff's work, the game makes use of an interactive system. However, the application does not make use of registration in the traditional sense, but instead, 'only' uses the camera as a motion sensor (Siemens SX1, n.d.) and applies 2D motion detection (Reimann and Paelke, 2006). Yet, judging from the images

[8] Binaural audio is based on the fact that hearing makes use of two signals: the sound pressure at each eardrum (Møller, 1992). If these two signals are recorded in the ears of a listener (or a dummy head), the exact 3D hearing experience can be reproduced by playing the signals back on a headset.

[9] This claim that sounds indeed seemingly originate in the real surrounding was confirmed by Zev Tiefenbach, the studio manager of Cardiff/Miller, who in turn confirmed this with Janet Cardiff (personal communication).

[10] If the listener turns their head, the recorded sounds will move along—they have no fixed position in real 3D space but are always relative to the position and head of the listener.

[11] One potential reason why this loose alignment suffices is that the recorded sounds not necessarily have to appear at a specific position in the surrounding space. For instance, no exact 3D registration is necessary when dealing with flying elements such as crows, as it does not matter where exactly they appear in the environment.

that can be found of this (and similar) games online, it appears as if mosquitoes were flying through the space in front of the phone's lens. An impression of this can be seen in figure 3.6, which shows a similar game running on a Nokia N95.

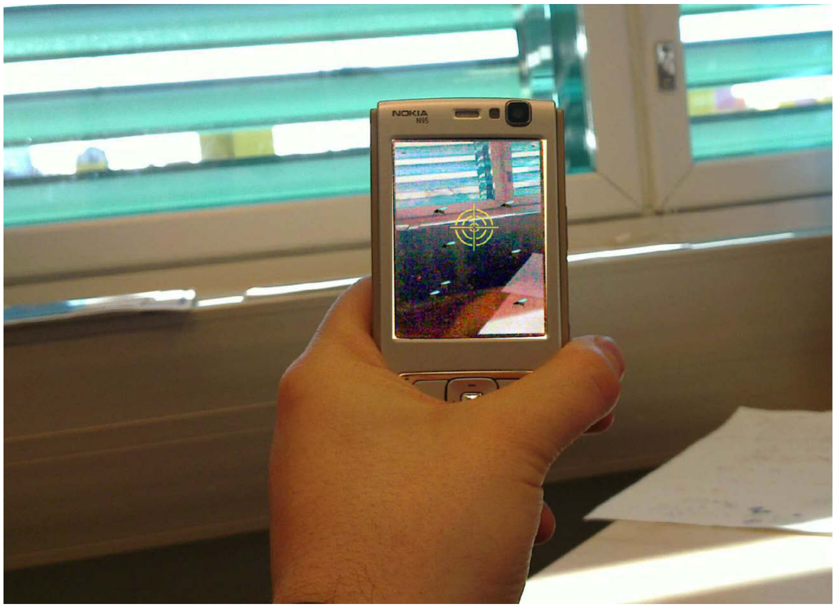


Figure 3.6: A game similar to the *Mozzies* game. Virtual mosquitoes appear to be flying in the space before the phone's lens. The image appeared in a paper by López et al. (2014), and permission to use the image in this thesis was granted by Miguel Bordallo Lopez.

Presumably, this works because mosquitoes 'only' have to appear to be flying *somewhere* in the surrounding space rather than at an exact position. To achieve this, exact registration seems not to be necessary. However, because the creatures are not registered in 3D, is not possible to walk around the virtual insects and look at them from all directions and angles. Furthermore, the virtual mosquitoes cannot disappear behind real objects. Due to the lack of first-hand experience, it remains open how one experiences these issues when quickly moving and turning the device.

NS KidsApp

A third example of AR experiences that work without typical AR systems is the NS KidsApp. This mobile application by the Dutch railway operator *Nederlandse Spoorwegen* (NS) is primarily aimed at children (and their parents) and it introduces a short story with the two characters *Oei* and *Knoei*. When starting up the application, it becomes clear that Knoei has missed the train, and that as a result, Oei and Knoei are not traveling together. It is then up to the user of the application to spend time with Knoei during the train journey.

There are several playful assignments for the player that allow them to make videos with Knoei appearing in the otherwise real environment. In these assignments, the player is asked to point their phone at a particular spot or have someone else point the phone at them and film them, while they are at a certain location. For instance, one assignment asks the players to put the camera against the window and film the outside. As a result, one can see Knoei flying next to the train in a superman kind of fashion on the phone's screen. Another assignment asks users to point the camera at the typical place-name signs that can be found on Dutch train stations. The resulting view of the scene on the phone shows Knoei swinging on the place-name sign. Yet another assignment invites the player to sit next to Knoei, while someone else is pointing the phone at them and filming. When doing so, the one filming can see Knoei hovering over a train chair, showing off his muscles to his neighbor (see figure 3.7).

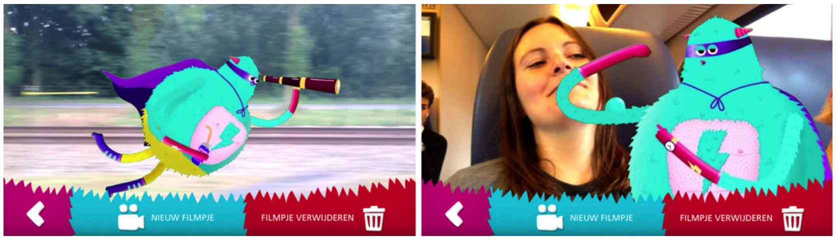


Figure 3.7: The NS Kids app shows Knoei flying next to the train (left) as well as next to the player showing off his muscles (right) on the camera feed. Screenshots by Jurriaan Rot and Hanna Schraffenberger.

This application, too, creates the illusion of virtual content existing in real space, without the use of a traditional AR system. Instead of a system, the *participant* can align the virtual and the real. Like in Cardiff's case, instructions are part of the game. These instructions make sure that what the participant sees will serve as a fitting background for the virtual overlay.

3.1.4 *AR Technologies Versus AR Experiences*

The previous sections have shed some light on the workings of traditional AR systems and the kind of experiences they aim to create. It has become clear that AR systems can make it seem as if virtual content existed in real space and as if the virtual and the real were one seamless environment. The creation of such mixed virtual-real environments and the presence of virtual content in an otherwise real environment seem to be primary goals of AR practice. However, we have seen that similar experiences and environments can also be achieved without traditional AR technologies. For instance, instead of an AR system, participants can align virtual and real content.

A question that we thus have to ask ourselves is what actually is defining for AR—the augmented environments and unique experiences that we hope to create or the technologies we develop in order to create them? Do we unnecessarily limit AR, if we only consider scenarios where an AR system registers virtual content with the real world in 3D? Do system-based definitions actually capture what we are ultimately interested in?

The answer to these questions remains a matter of opinion. Independent of one's individual position, it is clear that there are two sides to augmented reality: On the one hand, AR systems and on the other hand, the participant's experiences when using the system. If we want to truly understand and advance AR, a focus on either one alone will not suffice.

Personally, we take an experience- focused point of view. This is because AR technologies are meant to create augmented environments that a participant can experience. Arguably, the sole purpose of AR systems is for participants to use them and to experience augmented environments. Accordingly, we believe what matters most, is not what an AR system does, but what the participant experiences. If we ultimately aim at creating certain environments and experiences, why define the field in terms of the technologies that enable them rather than in terms of the environments and experience we are actually interested in? An environment- and experience- focused definition will hold, even if enabling technologies change or take unforeseen forms. We thus propose to define AR in terms of the unique environments a participant experiences, rather than in terms of certain types of systems.

So far, we have identified one key form of AR, namely otherwise real environments in which a participant experiences the presence of additional virtual objects. However, other types of AR experiences might exist as well. In fact, we suspect that virtual content can augment the real world even when it does not appear to exist in the physical space. We will address this possibility in the following section.

3.2 *From Registration to Relationships*

Registration is widely seen as a defining and necessary characteristic of AR (see, e.g., Azuma (1997); Azuma et al. (2001); Bimber and Raskar (2005)). There is no doubt that registration is important to AR. The previous section has shown that it can play a key role in making it seem as if virtual objects existed in real space. However, we believe there are three reasons to look beyond registration and to challenge the common focus on spatial alignment. First of all, making virtual content seemingly exist in real space does not always require 3D registration. The previous section has already shown that alternative approaches to placing virtual content in real space exist: For instance, Janet Cardiff's audio walks (Cardiff, n.d.) do not incorporate 3D registration, yet communicate the presence of virtual content in real space. Also, some settings require less strict forms of registration. E.g., an exact alignment might not be necessary when dealing with flying objects. Second and more fundamentally: The illusion of virtual content existing in real space, which motivates the need for registration, might not be necessary for AR in the first place. Arguably, not all forms of AR require for virtual objects to seemingly exist in real space! Simply put, other types of relationships (aside from spatial registration) between the virtual and the real are possible, potentially facilitating other forms of AR experiences. For instance, virtual content can inform us about the real world, and by doing so supplement and augment (our experience of) the real world. Third, we have to look beyond registration because registration alone might not always suffice to create the intended AR experience. For instance, it might not only be necessary to present a virtual object at the right position but also necessary to apply a realistic illumination in order for virtual objects to appear as if they existed in real space. Because the first argument has been discussed in detail (see subsetion 3.1.3), we will focus on the second and third point in the following.

3.2.1 Alternative AR Experiences

In this section, we challenge the need for registration and explore alternative forms of AR experiences that are not based on 3D registration and that do not entail the apparent existence of virtual objects in real space. In particular, we explore the idea of augmentation through content-based relationships between the virtual and the real. We present two examples that illustrate this concept and where the virtual contributes to, extends and augments our environment by informing us about it.

AUDIO GUIDES

The idea of virtual additions that inform us about the real world is common in the cultural sector. For instance, many museums provide additional information in the form of audio tours that guide the visitor through a museum, and which supplement the real world and ideally, enhance our experience of the exhibition. In our opinion, such audio tour guides can accompany a user and augment a user's experience of their real surroundings, even if they do not appear to be spatially present.

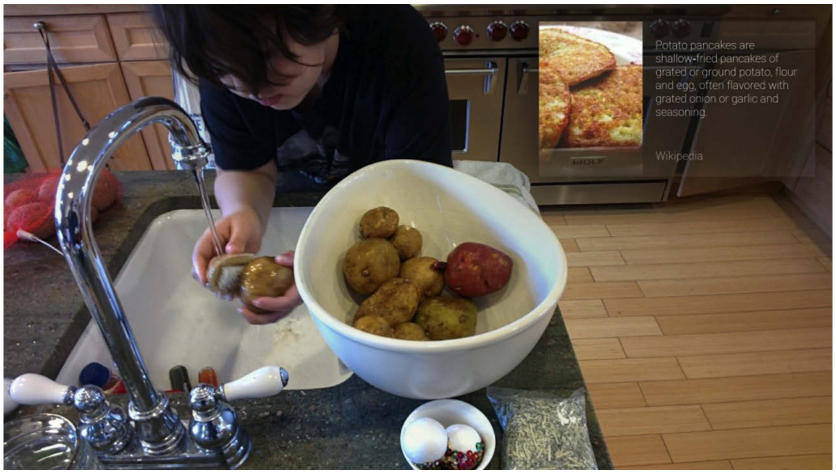
We are not alone with the opinion that audio tours and audio guides can be considered AR. For instance, Bederson (1995) argues "[o]ne place a low-tech version of augmented reality has long been in the marketplace is museums. It is quite common for museums to rent audio-tape tour guides that viewers carry around with them as they tour the exhibits" (p. 210) . Furthermore, Rozier Rozier (2000), refers to audio tours as "perhaps the earliest form of 'augmented reality"' (p. 20).

Whereas audio guides typically provide factual information about the real surroundings, other possibilities exist. An artist that takes the idea of audio tours one step further is Willem de Ridder. In 1997, de Ridder realized an audio tour in the "Stedelijk Museum" in Amsterdam that told visitors about the meaning of 'invisible' elements in the museum (*history and archive - Stedelijk Museum Amsterdam* n.d.). This shows that the virtual information can relate to the surroundings more freely. In fact, one could argue that Ridder's words also have the power to place imaginary virtual objects in a real environment and that they can create the experience of virtual objects existing in real space.[12]

GOOGLE GLASS

The concept of using a virtual layer of information to enhance our everyday lives has also been on the basis of the Google Glass project. Google Glass is essentially a head-mounted display in the shape of eyeglasses. A small display in one corner presents additional information (such as text and/or images) as an overlay on top of a user's view of the world.

The information displayed by Google Glass can be completely unrelated to a user's context (e.g., a random text message from a friend) but it can also relate to the user's real surroundings. For instance, the device can be used to translate text present in the real environment in real time, to overlay driving instructions onto a driver's view or to access relevant information in the kitchen (see figure 3.8).



Potato pancakes are shallow-fried pancakes of grated or ground potato, flour and egg, often flavored with grated onion or garlic and seasoning.

Wikipedia

Figure 3.8: Google Glass can overlay information that relates to our real surroundings and context. This image is a screenshot taken with the device, illustrating the user's view. Image created by and courtesy of Ben Collins-Sussman.

The role of Google Glass in AR is controversial. As we know, 3D registration is commonly considered necessary. This view excludes all Google Glass applications from the realm of AR. However, the 2015 call for papers of the leading AR conference ISMAR (International Symposium on Mixed and Augmented Reality) argues that "[l]ightweight eyewear such as Google Glass can be used for augmenting and supporting our daily lives even without 3D registration of virtual objects". In line with this, some researchers consider systems like Google Glass in the context of augmented reality. For instance, Liberati and Nagataki (2015) consider Google Glass an AR device, and distinguish

---

[12] However, it can be argued that these objects are imaginary rather than virtual.

among two types of current and future AR glasses: (1) AR glasses that inform the user about their surroundings and provide "informational text" to the user and (2) AR glasses that present additional objects, that are embedded in the real world and that potentially can interact with the real world as if they existed physically.[13]

If we apply Liberati and Nagataki (2015)'s distinction, Google Glass can act as an AR device, and falls in the first category of glasses, as it presents text (as well as other media) that informs us about our surroundings[14]. According to Liberati and Nagataki (2015), the information provided by such glasses modifies the objects they inform us about because the participant can change their attitude towards the objects based on the information.

We, too, believe that virtual information can modify (our perception of) real objects. Arguably, it can add to and affect our experience of the real world and in this sense become part of and augment the environment. However, we believe such augmentations are possible independently of how the virtual information is presented. In other words, information can augment our surroundings no matter whether it is, e.g., overlaid with AR glasses, displayed on a phone's screen or delivered by a recorded voice on headphones.[15,16] In our opinion, the question whether virtual content augments the real world (or vice versa) is not about the device we use, or the medium used to present such information. Instead, what matters is whether the presented content is experienced in relation to the real world. (This is likely the case when the two are inherently related on the content-level.) In line with this, the question whether Google Glass creates AR experiences depends on whether the presented information is perceived in relation to the real environment.

3.2.2 *Registration Without AR Experiences*

The previous examples have shown that spatial 3D registration is not the only link between the virtual and the real that allows us to experience virtual content as part of or in relation to the real world. Content-based relationships between the virtual information and the real environment, too, can facilitate the experience of an augmented environment. We thus believe there are different forms of augmentation aside from the apparent presence of virtual content in real space.

Another reason to look beyond registration is that registration alone might not always be sufficient in order to create AR experiences. This seems particularly relevant when it comes to the common goal of making virtual objects appear in real space. Here, many other relationships between the virtual and the real aside from spatial registration potentially contribute to the resulting experience. Among others, it can make a difference whether a virtual object appears to be affected by real light sources. For instance, Drettakis et al. (1997) claim: "Provid-

[13] In our opinion, the two categories are not exclusive. For instance, text can appear in the form of object-shaped letters that are integrated into the real environment and that seemingly interact with real objects.

[14] As we know, Google Glass can also present unrelated text. This is why we say that it can *act* as an AR device rather than that it *is* an AR device.

[15] In many ways, information defies the terms virtual and real. Arguably, information can have the same effects, no matter whether it is presented virtually or physically.

[16] In fact, we have to ask ourselves whether it actually matters whether the information is presented in a virtual form or, for instance, presented by a real person or on a physical information board. One can argue that information is never something physical, and always can affect and augment our experience of the world.

ing common illumination between the real and synthetic objects can be very beneficial, since the additional visual cues (shadows, interreflections etc.) are critical to seamless real-synthetic world integration" (p. 45). Sugano et al. (2003) go one step further and hypothesize that "[w]ithout shadows providing depth cues a virtual object may appear to float over a real surface even if it was rendered on the surface." (p. 76). In other words, registration alone might not suffice to create the desired effect. The subsequent experiment by Sugano et al. (2003) shows that presenting virtual objects with shadows as opposed to without shadows creates a stronger connection between virtual objects and the real world and increases the virtual object's presence in the world. (However, their research does not seem to support the idea that virtual objects appear completely detached from the real world due to the lack of shadows.)

In addition to optical interactions, a lack of other physical interactions and/or social interactions between real objects and virtual objects can potentially harm AR experiences and make virtual objects look "out of place" or appear as if they existed independently from the real world. For instance, Breen et al. (1996) point out: "For the new reality to be convincing, real and virtual objects must interact realistically" (p. 11). Likewise, S. Kim et al. (2011) write: "In order to make virtual objects move as if they coexisted with real objects, the virtual object should also obey the same physical laws as the real objects, and thus create natural motions while they interact with the real objects." (p. 25). Accordingly, for a virtual ball to appear as a believable part of real space, it might be necessary for it to bounce back when it hits a real wall. More than that—if we expect a realistic response, this movement might not be enough—the ball might also have to create a corresponding sound.

Furthermore, we can imagine that the presence of a virtual creature in the real environment is much more convincing if this creature seems to be able to perceive the environment and react to stimuli in the surroundings. For instance, a virtual creature might seem more present if it listens and responds to the sounds in the environment or if it sees and reacts to the participant when they are right in front of it.[17]

At the same time, the illusion of virtual elements being present in the space might be harmed if such interactions and perceptions are missing. For instance, it might disturb us if a virtual creature is not affected by real wind, if it is not reflected in real glossy surfaces or if it remains dry when it rains.

A first indication, that other factors aside from spatial registration indeed can affect the experienced presence of virtual objects in real space can be found in figure 3.9. In our opinion, the fact that the real cat does not seem to be aware of the virtual creature hurts the illusion of the virtual object actually being present in the space.

Unfortunately, a lack of empirical research makes it impossible to

[17] The idea of virtual creatures being more aware of their surroundings has been addressed by the developers of Pokémon GO with their AR+ update (Niantic, Inc., 2017). In this version, Pokémon seem to sense the player's movement. Consequently, players can scare virtual creatures away by approaching them too abruptly.

Figure 3.9: My cat shows no sign of awareness of the virtual beaver Sphero *(2011)*. According to our experience, this can harm the experience of Sphero being a part of the real environment. The picture is a screenshot showing the image displayed on the iPad. (The screenshot was taken by the author.)

conclude whether 3D registration is always sufficient to evoke AR experiences, i.e., to make participants experience virtual objects as part of or as related to the real environment. However, in our opinion, it is clear that other types of relationships also can facilitate and shape AR experiences. This should be reason enough to look beyond registration and consider relationships between the virtual and the real in general.

The notion that virtual objects should be able to sense and interact with the real world entails that we look beyond spatial registration and consider how the virtual and the real relate to one another on non-spatial levels. The idea that virtual content might have to react to non-visual aspects of the real world in order to appear as a believable part of the environment indicates that there is more to AR than what a participant sees. We will discuss this idea and in particular, the understanding of AR as a multimodal environment in section 3.3.

3.2.3 Registration Versus Relationships

In the preceding sections, we have argued that 3D registration between virtual content and the real world is only one of several ways to shape AR experiences. We believe that augmentation cannot only emerge from the *registration* of the virtual and the real but generally results from the *relationships* between the virtual and the real. In line with this, we believe that the spatial (and typically but not necessarily visual) presence and apparent existence of virtual content in the real environment is only one form in which the virtual can augment the real. Arguably, the virtual can also augment the real in different ways; e.g., by informing us about the surroundings. This, of course, raises one crucial question: If real-time registration by an interactive system in 3D is no defining factor, what then does define AR?

In our opinion, all AR scenarios have one characteristic in common: the virtual is experienced in relation to the real world and vice versa. Accordingly, we believe we are dealing with AR if one important criterion is met: the participant experiences a relationship between the virtual and the real. We thus propose that instead of defining AR in terms of registration between the virtual and the real on a technological level, to define it in terms of a relationship between the virtual and the real on an experiential level.[18]

So far, we have identified two key forms of AR. First, cases where a participant experiences the presence of virtual content in their otherwise real surroundings. Here, virtual content seemingly exists in real space, rather than, e.g., on a screen or in a separate virtual world. Second, environments where the virtual is experienced as pertinent to the environment on a content-level. The first form of AR is typically (but, as shown in subsection 3.1.3, not always) based on 3D registration of virtual content in real space. The latter form of AR, however, does not require registration. Rather, the virtual is likely experienced in relation to, as part of or as pertinent to the real environment because there is an inherent relationship between the virtual and the real in terms of content.

The question whether or not to define AR in terms of relationships between the virtual and the real remains a choice. In our opinion, AR comprises all cases where virtual content is experienced in relation to the real environment. We thus propose to broaden the view of AR and focus on the various possible relationships between the virtual and the real that facilitate such experiences. Presumably, there are many more relationships that still can be discovered. For instance, if we think about movies, a soundtrack can certainly become part of a scenery, although it is not *spatially* integrated into the movie. The virtual and the real might blend on such musical, non-spatial levels in AR as well.

Even when one disagrees with our view, it should be clear that spatial registration is not the only link between the virtual and the real that can shape AR experiences. Other relationships that play a role, e.g., include physical and social interaction between virtual and real objects.

3.3 From Visuals to Multimodal and Interactive Environments

AR is commonly understood in terms of virtual *imagery* that is overlaid onto a user's or participant's *view* of the world (see chapter 2). Accordingly, AR is thought "to 'augment' the visual field of the user" (Caudell and Mizell, 1992, p. 660) or to provide a "composite view" (cf. *Augmented Reality* 2005). In this sense, much AR research is focused on what a user or participant *sees*.

However, if we approach AR from a participant's point of view, the resulting AR environments are not just something visual. Rather, they

---

[18] Of course, a mere link between the virtual and the real does not guarantee that the participant also experiences this relationship. What is more, a participant might experience relationships that have never been created or intended. For instance, a museum visitor might listen to a virtual museum guide, but associate the information with the wrong artwork. Consequently, they might not experience the intended relationship but experience another link instead. Similarly, one and the same scenario might be experienced as AR by one person but not by another. However, we believe it is safe to assume that by carefully considering and crafting relationships between the virtual and the real, we can shape AR experiences.

are interactive, multimodal environments that potentially engage all of the participant's senses and that invite the participant to act in and interact with the space.

In our opinion, there are at least three reasons why we have to approach AR in terms of multimodal and interactive environments rather than focus on what a participant sees. First, the real world is not something visual but has multimodal and interactive qualities. Furthermore, virtual content, too, can exhibit non-visual qualities and allow for multimodal interaction. Finally, even purely visual virtual content that is superimposed onto our view can affect how we perceive non-visual stimuli. In the following sections, we develop these arguments in detail.

3.3.1 The Multimodal and Interactive Real World

If we approach AR from a participant's point of perspective, virtual content is experienced as part of or in relation to the otherwise real world. This world is not just something participants can *see*. Rather, it is a world that participants can perceive with all their senses, act in and interact with: we feel the ground beneath our feet, hear our footsteps, move over when a bike bell rings, we knock on doors and open them and engage in conversations with other people.

Although many AR systems focus on what a participant sees, non-visual qualities of the real environment often to play an important role in the overall resulting experience. For instance, in Caudell and Mizell's case of an AR system that helps assembly workers with virtual instructions (cf. section 1.1), it is crucial that the worker can touch, feel and physically interact with real objects: the worker might, for instance, drill a hole, connect wires, or place sticky fabric at the right spots. The ultimate goal of Caudell and Mizell's prototype is to support the worker with his actions in the world. Although the *virtual component* is strictly visual and intangible, the resulting augmented environment is more than what the user sees. Clearly, the system had little purpose, if the user could only see the augmented environment.

And Caudell and Mizell's project is hardly the only project where non-visual qualities matter. To mention just a few more examples: Participants need to touch and manipulate the world when they repair their printer with virtual instructions (Feiner, Macintyre, et al., 1993). A surgeon might listen to audible feedback of medical monitors, interact with colleagues, and of course, perform a surgery with the help of virtual indicators. Players of AR games like Pokémons GO walk through space, talk to our friends, and hopefully, hear a car bonk when they try to catch virtual Pokémons on real streets. Judging from these examples, AR is more than meets the eye.

If we understand AR in terms of the mixed virtual-real environments that a participant experiences rather than in terms of technologies, multimodality is the norm rather than the exception. In contrast to what widespread claims imply, AR not only "*might* [italics added] apply to all senses" or "*could* [italics added] be extended to include sound" (Azuma 1997, p. 361), but rather already applies to all our senses. AR is inherently multimodal, simply because it includes the multimodal real world. When dealing with 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." (Azuma, 1997, p. 3567) but that we can also hear, smell, touch, and taste this world.

Just like AR environments are multimodal because they entail the multimodal real environment, AR environments are also interactive simply because the real world allows for interaction. Whereas the fact that AR engages all our senses is often overlooked, the interactive qualities of AR are well known. For instance, Craig (2013) emphasizes that "[a]ugmented reality is interactive, so it doesn't make sense to watch it or listen to it" (p. 2) and argues that "that the way people engage with augmented reality is to experience it" (p. 1). Likewise, Hugues et al. (2011) 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—and hence, choose their own perspective—has been key to AR's technological development. One of the most prominent topics in AR research is tracking techniques that make sure a virtual object's visual appearance matches the participant's current viewing perspective even when the participant changes their point of view (cf. Zhou et al., 2008).

Considering that participants experience multimodal environments that allow for interaction, we believe it makes sense to approach AR as a multimodal and interactive rather than solely visual phenomenon. However, the characteristics of the real world are not the only reason to do so. Another reason to think about AR this way is the fact that virtual content, too, can take multimodal and interactive forms.

3.3.2 Multimodal and Interactive Virtual Objects

How does it feel to touch a virtual object, to run one's fingers over it? Is there a chance that we can burn our hands when doing so? How does a virtual object taste or smell, what sound does it make if we shake it and how heavy is it, if we want to carry it around?

If we look at existing AR projects, tools and technologies, the answers to these questions can be disappointing. The majority of existing projects and devices allows us to view virtual content, rather than to experience it with all our senses. Sure, quite some virtual objects also produce sounds. However, generally speaking, virtual objects in AR cannot be felt, have no smell, taste, weight, temperature or other physical properties that we know from real objects. To mention just a few examples: The *Sphero* (2011) project allows us to steer a virtual beaver through our living room, but we won't feel the beaver's fur if we try to pet it (instead we feel a robot ball). Likewise, the game Pokémon GO allows us to see virtual creatures in our everyday surroundings, but we cannot use dogs to sense their smell and chase them. Similarly, while virtual avatars can carry make-up, they typically cannot wear any perfume. Whereas touching physical artworks in a museum can get us in trouble, touching AR art (e.g., Veenhof, 2016) usually is safe but also boring: we will not feel anything if we try. Commonly, what it comes down to is that virtual objects can be *seen*—the typical virtual object is first and foremost a visual object.

While most AR projects make use of visual overlays that are superimposed onto a participant's view, there are exceptions that show that virtual content does not have to equal visual content. For instance, we have already thoroughly discussed Cardiff (1991)'s *Forrest Walk* that makes use of audio recordings. In the following, we will briefly point out a few more projects that illustrate that virtual content in AR can also take sonic, haptic, gustatory, olfactory and multimodal forms.

One of the works that use sound rather than visuals to convey the presence of virtual objects in real space is the SoundPacman game (Chatzidimitris et al., 2016). This game is an audio AR version of the traditional PacMan game. However, here all game elements are seemingly placed in the real, physical surroundings and the information about their position is provided solely by means of 3D sound. Using audio rather than visual cues clearly provides different possibilities. Among other things, the use of audio allows participants to perceive ghosts even if they are not in their direct line of sight and, for instance, positioned behind them.

An example that shows that we might perceive virtual content in real space haptically rather than visually has been realized by Bau and Poupyrev (2012). Their REVEL device injects electrical signals into a participant's body and thereby allows participants to feel virtual textures when running their fingers over real physical objects. As such, the system augments real physical objects with virtual tactile textures.

A project that aims at altering the taste rather than the tactile feel of real elements has been realized by Nakamura and Miyashita (2011). The authors propose a system that makes use of a fork or chopsticks connected to an electric circuit and thereby changes the taste of food. Similarly, they propose to change the taste of drinks by using two straws that are connected to an electric circuit. The experienced change in taste happens because the tongue is stimulated with electric current.[19]

Finally, also scents can be used in the AR context. For instance,

[19] This effect is nowadays known as "electric taste" and was discovered by Sulzer as early as 1752 (Bujas, 1971).

Lindeman and Noma (2007) suggest that under-nose displays, air-canon displays and scent emitters in the environment can be used to present the participant with a mix of computer-generated and real-world scents.[20] Also, Yamada et al. (2006) have proposed a wearable olfactory display that can present virtual odor sources in an outdoor environment. Their proposed setup takes the position of the virtual odor source as well as the position of the participant into account and varies the strength of the presented odor accordingly. This allows the system to simulate the spatial spread of the odor in the real environment. Although this project has been presented in the context of VR, it can be considered an AR project. This is because their system can simulate the existence of odor sources in an otherwise real environment.

As the previous examples illustrate, AR can also work with non-visual virtual information. In addition, AR can make use of multimodal content that combines different types of sensory information. Researchers have, among others, used force feedback devices such as the Phantom in combination with HMDs in order to create viso-haptic virtual objects. For instance, Bianchi et al. (2006) have demonstrated the use of a Phantom device in an AR-based ping-pong game. In their setup, a virtual bat is attached to the haptic device and allows players to interact with a virtual ball. The player can not only see the virtual ball via a head-mounted display but also feel its impact on the simulated bat via the haptic device. Another example of an AR project that uses multimodal virtual content is the mobile (smartphone-based) AR game *GeoBoid* by Lindeman, G. Lee, et al. (2012). 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.

3.3.3 Multisensory Perception and Cross-Modal Effects

So far, we have argued that we have to treat AR from a multimodal perspective because both the real world and virtual content can engage all our senses. A third reason to treat AR from a multimodal perspective is the way our human perception works. When we perceive the world around us, our brain combines information from various sources. As Ernst and Bülthoff (2004, p. 162) point out:

> To perceive the external environment our brain uses multiple sources of sensory information derived from several different modalities, including vision, touch and audition. All these different sources of information have to be efficiently merged to form a coherent and robust percept.

For instance, when we sit at our desk, our arms resting on it, our fingers drumming on it, we can see, hear and feel the desk. These different sensory streams of information are integrated into our coherent perception of the desk. When information from different sensory modalities is combined, different sensory stimuli can interact with one

[20] The term 'computer-generated' can be a bit misleading when it comes to scents. We thus propose the term 'computer-controlled' scents.

another. For instance, what we hear might influence our tactile experience and what we see can influence where we perceive a sound. Such influences, where information from one sense affects how we experience information from another sense are referred to as cross-modal effects and cross-modal interactions.

A popular example of cross-modal interaction is the "Parchment-skin illusion". According to Jousmäki and Hari (1998)'s findings, the sounds that accompany hand-rubbing can influence the tactile sensation of the skin. It was found that emphasizing high frequencies can make the skin feel rougher. Another popular cross-modal illusion is the McGurk effect (McGurk and MacDonald, 1976). In this illusion, what we see can affect what we hear. When being presented a video of a person saying "ga-ga", dubbed with the sound of a voice saying "ba-ba", participants in the study by McGurk and MacDonald (1976) reported hearing "da-da". This shows that different sensory stimuli can not only complement each other but also interact with each other to create a different experience than the sum of the individual experiences.

The fact that our perception is multi-sensory also plays a role in AR. Because interactions can occur between different sensory modalities, visual virtual information might affect our perception of real non-visual characteristics of the environment. In other words, even visual virtual overlays that are superimposed on a participant's view can affect what the participant perceives with other senses. More generally, virtual sensory information can interact with and affect how we perceive 'real' sensory information (information originating from the so-called real, physical world).

As it turns out, such cross-modal interactions between virtual and real stimuli are not only a theoretical consideration. Various AR projects have already utilized the phenomenon of cross-modal interaction and explicitly used visual virtual information to transform our experience of non-visual qualities of the world. For instance, Hirano et al. (2011) and Sano et al. (2013) use an HMD to display different computer-generated deformations on an object, when it is pushed down by a participant. Their experiments show that the perceived softness can be manipulated by means of visual virtual dents, without changing the actual material: The larger the visual dent caused by pushing the object, the softer seems the object. Other projects similarly show that virtual *visual* information can alter the perceived *temperature* (Ho et al., 2014), *texture* (Iesaki et al., 2008) and *center-of-gravity* (Omosako et al., 2012) of real objects. (These examples will be discussed in more detail in section 4.7).

As these examples show, visual virtual information can affect how we perceive non-visual qualities of the world. Even more possibilities arise if we present non-visual and multimodal virtual stimuli (cf. section 3.3.2). One of the projects that combine the idea of presenting multimodal virtual information with the concept of cross-modal interaction is the MetaCookie project. Because this project supports our presented arguments for a multimodal perspective on several levels, we will discuss this project in more detail in the following.

The MetaCookie Project

The *MetaCookie+* headset[21] (Narumi, Nishizaka, et al., 2011a,b) aims at changing the flavor of a real plain cookie. The project is based on the idea that virtually changing the look and smell of a plain cookie might affect its perceived flavor. Consequently, the headset changes the visual appearance of a plain cookie and, for instance, makes it look like a chocolate, almond or cheese cookie (see figure 3.10). At the same time, it also features an olfactory display with scents that match the visual choices.

Figure 3.10: The MetaCookie+ headset. Reprinted from Narumi, Nishizaka, et al. (2011a). Reprinted under fair use.

In order to use the *MetaCookie+* system and experience the different tastes, the participant needs a real plain cookie with a special AR marker on it (see figure 3.10). (The marker makes it possible for the system to keep track of the cookies position).

When eating the cookie, the participant wears a custom head-mounted visual and olfactory display.[22] Before placing the cookie in their mouth, the participant can select a cookie of their liking from a list of options, including, for instance, chocolate, almond and cheese. After the participant has chosen their preferred cookie, an image of the selected cookie is integrated into his view at the position of the

---

[21] A first versions of this system has surfaced under the name "Meta Cookie" (Narumi, Kajinami, et al., 2010b).

[22] In the case of *Meta Cookie+*, the olfactory display consists of several air pumps, scented filters and a controller and is able to eject six types of scented air and fresh air.

real cookie, effectively making it look as if they were holding the cookie of their choice. In addition, the olfactory display dispenses a scent matching that of the selected cookie when the augmented cookie is within a range of 50 cm from the participant's nose. The strength of the scent linearly increases the closer the cookie moves to the participant's nose. When the participant is about the eat the cookie (and the cookie is in front of their mouth) the system produces the most intense version of the smell for about 30 seconds. This strong odor is produced to emulate retronasal olfaction (the stimulation of olfactory receptors via the mouth rather than nose, which typically results in a stronger sensation than stimulation via the nose). Because the smell is presented for about half a minute, the scent is assumed to be presented longer than it takes the participant to eat the entire cookie.

The authors evaluated their system (presumably very informally) with "a dozen people" in its initial version 2010. Its later incarnation was more systemically evaluated it with 15 participants (Narumi, Nishizaka, et al., 2011b) as well as 44 participants (Narumi, Nishizaka, et al., 2011a). Based on their initial trials (Narumi, Kajinami, et al., 2010b), the authors report that almost all participants perceived a change of the taste of the plain cookie. Similarly, the results from their later evaluation (Narumi, Nishizaka, et al., 2011a,b) suggest that Meta Cookie+ can change a perceived taste and allows users to experience different flavors, solely by changing the visual and olfactory information.

In our opinion, the project supports our argument for a multimodal view on AR in three ways. First of all, by augmenting the taste of a real cookie that one eats, it emphasizes that the real world is a multimodal world that we interact with and not a visual world that we look at. In other words, it emphasizes that the real component in AR entails more than what we see. Second, by displaying scents, it emphasizes that the additional information we present to participants is not limited to visual information. As such, it illustrates that the virtual component in AR, too, can be more than what we see. Finally, it builds on the concept of cross-modal interactions and thus shows that our senses do not work in isolation. This means, that we cannot simply treat what we see as independent of what we hear, smell, taste, or otherwise perceive. Naturally, this is not only true in the real world but also in AR.

3.3.4 *Visual Overlays Versus Multimodal Environments*

In the previous sections, we have presented various reasons to treat AR from a multimodal rather than vision-focused perspective.

The first reason is that AR experiences entail the real world and that this real world is a multimodal world. This means that multimodality in AR is the norm, not the exception.

The second reason is that virtual content can take non-visual and multimodal forms. In our opinion, considering these types of contents has no downside. Rather, working with other modalities opens up great opportunities: with sound, we can, for instance, also experience virtual objects that are hidden from our view. (We could, for instance, create the creepy feeling of being followed by presenting the participant with the sound of footsteps that seem to originate behind them and follow them around.) Furthermore, non-visual content can allow us to display invisible objects. This might make sense, if we, for instance, want to communicate the existence of ghosts in the environment (cf. Chatzidimitris et al., 2016).

The third reason for a multimodal perspective is that even visual additions that are integrated into a participant's view, can affect how the participant experiences non-visual aspects of the environment. Sight does not operate independently from our other senses; hence we should not treat it independently from it.

A final argument to treat AR from a multimodal perspective can be made for cases where AR technologies aim at making it appear as if virtual content existed in the real world or where it aims to imitate the real world.[23] In our opinion, multimodal and interactive properties of the real and the virtual world can play an important role in achieving these goals.

First of all, if virtual objects imitate real objects, we might expect them to display the same multimodal qualities that a real object would display. For instance, we would expect a virtual balloon to make a sound if it pops, and a virtual vase to make a sound if it breaks, and we might expect to feel something when we touch a virtual toy. Of course, our expectations are likely related to the specific object in question. However, a lack of multimodal qualities might harm the credibility of the virtual object and hurt the impression that the object is part of real space.

Furthermore, we believe that taking into account the multimodal qualities of the real world might make a virtual object's existence in the real world more believable. Imagine, for instance, a virtual pet that gets scared when there is a sudden sound in the surroundings, a virtual object that moves to the song playing on the radio, or a virtual character that puts on different clothes, according to the current outside temperature. We assume that if virtual elements react to the multimodal properties of the real world, such as its temperature and sounds, it might help convince us that virtual objects are actually in the same space as we are, and make the experience more entertaining and interesting.

At the same time, we hypothesize that the apparent presence of virtual objects in the real world may also be compromised when virtual content remains oblivious to the multimodal and interactive properties of its surroundings. For instance, we expect a virtual mouse to be

---

[23] As we have pointed out in section 3.2, we do not believe all AR projects have to create such an illusion. Because this argument only applies to cases where this illusion is desired, we present this argument as an additional reason to treat AR from a multimodal perspective rather than as a fundamental, general argument.

frightened (or at least react) when it hears a miaow.[24] 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.[25]

What is more, we might also expect that virtual content affects the real world and evokes multimodal responses: for instance, we would expect to hear sounds if a virtual ball bounces on a real wooden floor and expect a real window to break if the ball hits it.[26]

If, however, a virtual leaf does not move in real wind, a virtual mouse shows no reaction to a cat's miaowing, a virtual drum does not produce sounds when it is hit and if the floor remains silent when a virtual ball bounces on it, it might harm the impression that virtual content exists in the same space, even if this content is perfectly registered with the world visually.

To summarize, for objects to seemingly exist in the same space as real objects, it makes sense that they can interact with each other physically. Such interactions have visual and non-visual effects.

Ultimately, our various observations, assumptions and arguments boil down to one fundamental point: An AR environment is more than something a participant can see. In order to understand and advance AR, it does not suffice to study visual overlays that are integrated into a participant's view. Instead, we also have to study AR in terms of multimodal and interactive environments that a participant experiences.

3.4 *Synthesis, Discussion and Conclusion*

AR is commonly considered an interactive technology that overlays virtual imagery onto our view of the real world and that aligns this virtual content with a user's view of the real world in 3D and in real-time. In this chapter, we have argued that this image of AR is incomplete and we have proposed an alternative, more encompassing view of AR. This view departs from widespread understandings of AR in three complementary ways.

First, we do not view AR as a technology. Instead, we claim that AR technology enables augmented reality. We focus on the resulting augmented reality environments and experiences rather than on the technologies that enable them.

Second, we see AR as a result of the relationships between virtual content and the so-called real world. Whereas AR is generally assumed to involve the spatial integration of virtual content in (our view of) the real world, we believe that other types of relationships between the virtual and the real are possible. We hypothesize that different and new forms of relationships will enable different and novel forms of AR.

Third, we treat AR as a multimodal and interactive phenomenon and argue that AR engages all our senses and allows for action in and

[24] If we notice that the object is virtual, we also might expect the opposite and assume that the object cannot sense the world around it and interact with it. Here, we proceed under the assumption that the virtual object appears so present and real that it evokes the same kind of expectations like a real object. However, we are very much aware that an object's virtual nature might cause us to have different expectations instead.

[25] One might consider the question whether a leaf moves in the wind a minor detail. However, we mention such details because we expect they might considerably affect the AR experience, either consciously and unconsciously.

[26] Again, our expectations about this might be different if we can see that we are dealing with virtual objects. The topic of what we expect when we are aware that we are confronted with virtual objects is an interesting topic for future research.

interaction with the environment. Instead of focusing on what a user or participant sees, we focus on non-visual and multimodal aspects of both the real world and virtual content.

We can synthesize these views and propose to define AR in terms of interactive and multimodal environments where a participant experiences a relationship between virtual content and the real environment.

This definition allows us to distinguish AR from other environments. Most importantly, it allows us to distinguish AR from solely physical environments that do not contain any virtual information. Furthermore, the definition sets AR apart from scenarios where the virtual and the real merely coexist in the same space and where both are experienced as *independent* from each other. For instance, it does not include situations where a participant listens to an audiobook and experiences this story as unrelated to their actual environment. Likewise, it sets AR apart from entirely virtual environments. For instance, our definition does not include scenarios where participants are immersed in virtual worlds and where they experience virtual elements as independent from their actual, real environment. Finally, due to its focus on the *real* environment, our definition also sets AR apart from other mixed reality environments where a participant experiences a link between the virtual and the real. For instance, it does not include situations where the participant experiences real objects (e.g. a physical toy gun) in relation to an otherwise virtual environment (we will propose to refer to such environments as "augmented virtuality" below.)[27]

Although our perspective on AR deviates from prevailing ideas, our understanding of AR is not entirely new (see chapter 2). In particular, many researchers suggest that AR can engage all senses. To mention just a few examples: Azuma et al. (2001) point out that "AR can potentially apply to all senses, including hearing, touch, and smell" (p. 34). Furthermore, Craig (2013) points out "Augmented reality can appeal to many of our senses (although currently it is primarily a visual medium)" (p. 1-2). Lindeman and Noma (2007) explicitly explore the idea of multi-sensory AR and present a classification scheme that allows for visual, auditory, haptic, olfactory and gustatory forms of AR. However, these views generally assume that AR only will engage or address other senses, if non-visual virtual content is presented to the participant. Our view sets itself apart from such ideas because it considers AR as multimodal, even when virtual content using only one modality (e.g., only visual content) is presented to the participant. This is because we consider the multimodal real world as part of the experience. As a consequence, multimodality in AR is the norm rather than the exception.

We are not the first ones to focus on the relation between the virtual and the real rather than on registration. For instance, Manovich (2006) suggests that "a typical AR system adds information that is directly

---

[27] Of course, the distinction between AR and other mixed reality environments is not clear-cut.

related to the user's immediate physical space" (p. 225). Similarly, Klopfer and Squire (2008) define AR in terms of "situation[s] in which a real world context is dynamically overlaid with coherent location or context sensitive virtual information" (p. 205). While these claims suggest that the virtual has to relate to the real, they do not claim that the information has to be *registered* in real 3D space or aligned with real objects in 3D. The main contribution of our work in this context is that it provides a detailed rationale for deviating from commonly accepted focus on registration. Furthermore, our proposed definition differs from views such as put forward by Manovich (2006) and Klopfer and Squire (2008) with its focus on the participant's *experience* of those relationships.

While scarce, some existing definitions also focus on the participant's experiences. For instance, Spence and Youssef (2015) describe AR as "an experience of a physical, real-world environment whose elements have been augmented, or supplemented, by computer-generated sensory input" (p. 1). However, if we look beyond mere definitions, there are more views that emphasize the experiential qualities of AR. For instance, Craig (2013) focuses on the experience associated with AR and writes "the way people engage with augmented reality is to experience it" (p. 1). However, our view still differs from such existing views with respect to what constitutes an AR experience: We believe AR is characterized by the experienced relationship between the virtual and the real. Furthermore, we are not aware of any existing views that explicitly argue for a shift from a technology-focused definition of AR towards an experience-focused definition of AR.[28] While the individual points are not necessarily new in isolation, our contribution is unique and new in this combination.

In our definition, the participant's experience of the *real environment* plays a key role. Existing definitions of AR generally focus on the fact that a real environment is part of AR (e.g., Azuma et al. (2001)). Most notably, Milgram and Kishino (1994) describe AR as "all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects" (p. 1321) and places AR on the side of the real environment in their much-cited reality-virtuality continuum (see section 2.1). Our definition is similar to existing views because it also focuses on the real environment. However, unlike many views, our definition does not focus on the environments that are displayed by a system but on the environments that are perceived by the participant.

In our opinion, the proposed perspective on AR advances our understanding of AR on a fundamental and theoretical level. We hope that a better and broader understanding of AR will inspire new forms of AR. Our investigation has revealed various examples of interactive applications that defy prevailing definitions of AR but yet, augment our experience of our physical surroundings. This shows that narrow

[28] However, a similar shift has been postulated in the context of Virtual Reality by Steuer (1992). In his seminal paper, Steuer criticizes that the focus of virtual reality is technological, rather than experiential. Consequently, he focuses on the participant's sense of being in an environment and defines VR as "a real or simulated environment in which a perceiver experiences telepresence" (p. 7). While we propose a similar shift in perspective, our proposal differs fundamentally from Steuer's contribution because it addresses AR rather than VR and thus, focuses on a different kind of experience.

definitions not necessarily prevent practitioners to think outside of the box and to come up with different forms of (arguably) augmented reality. Yet, we expect that a better understanding of AR and of how to create (or facilitate) it, will inspire even more and new forms of AR.

When it comes to creating AR scenarios, our view of AR suggests that we have to consider and give form to the relationships between the virtual and the real. However, we also have to keep in mind that establishing a relationship between the virtual and the real not automatically ensures that a participant also experiences this relationship. What is more, a participant might experience relationships that have never been created or intended. For instance, a museum visitor might listen to a virtual museum guide, but associate the information with the wrong artwork. Similarly, the same scenario might be experienced as AR by one person but not by another. In our opinion, the question whether a scenario should be considered AR cannot be answered based on what a system does or displays. Instead, it remains a question of personal experience.

Our investigation has revealed two main forms of AR: First, cases where a participant experiences the presence of virtual content in the real environment. We propose calling this "presence-based AR". Second, cases where the participant experiences virtual content as related to or pertinent to their surroundings on the content-level. We suggest calling this "content-based AR". In future research, it would be desirable to explore if yet different forms of AR exist. For instance, can the virtual become part of the real world similarly to how a soundtrack becomes (a non-spatial) part of a movie? Furthermore, we would like to systematically explore what factors contribute to the experience of virtual content being part of the real space. We can imagine that next to registration, aspects such as the participants' imagination and an underlying narrative can play a major role in AR.

In line with our definition, our investigation has focused on situations where the participant experiences virtual content in relation to the so-called real world. However, all of our three main considerations—the focus on (1) experience, (2) multimodality and (3) relationships between the virtual and the real – can likewise be applied to the more general field of mixed reality (cf. section 2.1). If we generalize our definition, mixed reality can be seen as any environment in which the participant experiences a relationship between the virtual and the real. In line with this, *augmentation* can be seen as the result of the perceived relationships between the virtual and the real. Those specific mixed reality environments where the participant experiences real elements in relation to their otherwise *virtual* surroundings can be described as *augmented virtuality*.

In the future, it would be interesting to further explore our arguments in the more general context of mixed reality. However, it is also necessary to further investigate what our proposed view of AR entails, and what forms AR can take if we apply our definition. This is the focus of this thesis. As a result, issues that play a role in other forms of mixed reality (i.e., augmented virtuality) fall out of the scope of this thesis. For instance, we will not discuss the experience of telepresence in virtual environments and the representation of real participants in virtual space in the form of an avatar.

In conclusion, we have presented new perspectives on AR, and arrived at an understanding that focuses on the participant's experience of the environment, the relationships between the virtual and the real and the multimodal and interactive qualities of the environment. More specifically, we have proposed that AR is characterized by the experience of virtual content in relation to an otherwise real, multimodal, interactive environment. We have already encountered several examples that fit and illustrate this broader definition, such as audio guides in a museum. However, many questions remain open: What else does augmented reality entail if we apply our definition? How can AR look, taste, smell, feel and sound like if we do not require registration? Are there yet other forms of AR, based on yet different relationships between the virtual and the real? In which ways can the virtual become part of and relate to the real environment? In the following chapters, we address these questions. We apply our proposed perspective and systematically explore the various forms AR can take.