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

2 *Existing Views*

What is augmented reality? This is one of the key questions we address in this thesis. If we turn to existing answers, we can find many varying, often complementary, sometimes contradicting views on the subject. Yet, there are some notions of AR that have gained wide acceptance.[1]

[1] AR has been actively addressed from a computer engineering perspective. Many definitions and descriptions that we review have been presented in an engineering context and have no ambition to make fundamental claims about the nature of AR. We nonetheless review such descriptions because collectively, they provide an overview of how AR is commonly approached.

2.1 *Common and Complementary Views*

First, there is the widespread understanding of AR in terms of Milgram and Kishino's (1994) much-cited reality-virtuality continuum (see figure 2.1) (see also Milgram, Takemura, et al., 1994).[2] The presented continuum ranges from purely real environments to entirely virtual environments.[3] The space in between these extremes is referred to as "Mixed Reality". The field of mixed reality includes both augmented reality and augmented virtuality. Augmented reality is placed somewhat closer to the real environment, and describes an (display of an)[4] otherwise real environment that is augmented by virtual objects. Similarly, Milgram, Takemura, et al. (1994) describe augmented virtuality as a principally virtual environment that is augmented through the addition of "real (i.e. unmodelled) imaging data" (p. 285). This happens, e.g., when a user's real hand is displayed in an otherwise virtual environment.

[2] Originally, this continuum was referred to as "virtuality continuum" (Milgram and Kishino, 1994). However, by now the continuum is commonly known and referred to as the "reality-virtuality continuum".

[3] With their continuum, the authors focus on environments that are viewed via some sort of visual display.

[4] Milgram and Kishino (1994) refer to AR both as "all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects" (p. 1321) as well as "any case in which an otherwise real environment is 'augmented' by means of virtual (computer graphic) objects" (p. 1322).

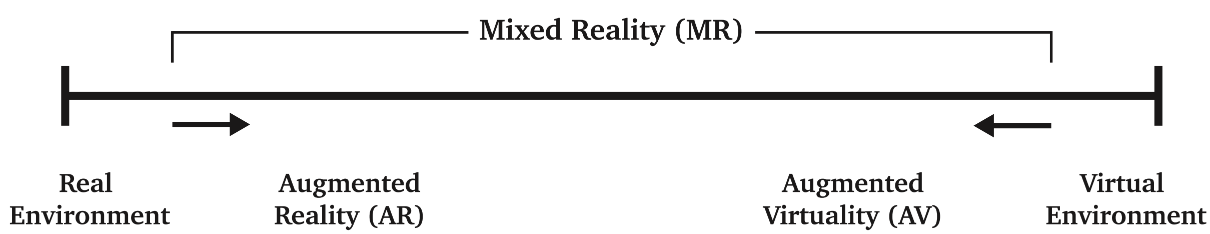

Figure 2.1: A simplified representation of the reality-virtuality continuum as shown in (Milgram, Takemura, et al., 1994).

Another view of AR that has gained an extremely wide acceptance is Azuma's (1997) much-cited definition from an early survey on AR. In this seminal survey, Azuma describes AR as a variation of virtual reality that "allows the user to see the real world, with virtual objects superimposed upon or composited with the real world" (p. 3567). Looking for a definition that does not restrict AR to a specific technology, Azuma defines AR as a system that has three decisive characteristics. It:

1. Combines real and virtual
2. Is interactive in real time
3. Is registered in three dimensions

This definition resurfaced in a somewhat more elaborate and accessible form in a follow-up survey by Azuma et al. (2001), where AR is defined in terms of systems that embody the following three characteristics. They:

1. Combine real and virtual objects in a real environment
2. Run interactively, and in realtime
3. Register (align) real and virtual objects with each other.

In addition to these two often-cited views, we can identify three prevailing ideas about the nature and characteristics of augmented reality that complement and reaffirm the ideas stated above. 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 align* this virtual content with the real world.

2.1.1 AR as a Technology

One of the most prominent understandings of AR is the idea of AR as a technology. For instance, Zhou et al. (2008), in their review of 10 years' worth of AR research presented at the primary AR conference ISMAR (International Symposium on Mixed and Augmented Reality) and its predecessor, describe AR as "a technology which allows computer generated virtual imagery to exactly overlay physical objects in real time" (p. 193). Comparably, Reiners et al. (1998) claim that "Augmented Reality is a technology that integrates pictures of virtual objects into images of the real world" (p. 31). A similar description is given by Roberts, Evans, Dodson, Denby, Cooper, Hollands, et al. (2002), who describes AR as "a technology that allows information stored digitally to be overlaid graphically on views of the real world" (p. 1) as well as by Doyle, Dodge, and Smith (1998) who describe AR as "a technology in which a user's view of the real world is enhanced or augmented with additional information generated from a computer model" (p. 147).

While AR is often seen as a technology, usually, these views do not limit AR to a specific hardware technology, such as head-mounted displays.[5] Rather, these views focus on what AR technology does. This brings us to the other two often mentioned characteristics of AR:

[5] In fact, researchers have been very explicit about not limiting AR to a specific hardware. For instance, Azuma et al. (2001) emphasize that they do not restrict their definition (see above) to "particular display technologies, such as a head- mounted display (HMD)" (p. 34).

virtual content is (1) visually overlaid onto our view of the world and (2) spatially registered (aligned) with real 3D space.

2.1.2 *AR as Visual Virtual Overlays*

Existing notions of AR are commonly focused on what a user or participant *sees*. Accordingly, AR is commonly understood in terms of virtual *imagery* that is overlaid onto a user's or participant's *view* of the world. This idea has already surfaced in some of the previously cited views about AR technology. In addition, this understanding is, for instance, shared by Piekarski and Thomas (2002), who describe AR as "the process of overlaying and aligning computer-generated images over a user's view of the physical world" (p. 36). Likewise, Rosenblum (2000) describes AR as "the overlaying of computer-generated imagery atop the real world using a see-through display" (p. 39). The media theorist Manovich (2006) provides a yet similar description and summaries AR as "the laying of dynamic and context-specific information over the visual field of a user" (p. 222). A focus on vision is also predominant in the research by Milgram and Kishino (1994) (see above), who discuss AR in terms of visual displays.

2.1.3 *AR as the Registration of Virtual Content in Real Space*

In addition to the view that virtual content is *overlaid* onto the real world, there is also the common belief that virtual content is spatially integrated into or aligned with the real 3D space. This spatial alignment is commonly called *registration*.

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 medical practice, for instance, registration is used to combine images obtained with different types of technologies, and/or images obtained at different points in time (L. G. Brown, 1992). For instance, two medical images of a patient that have been taken at different moments might be registered with each other to find changes (L. G. Brown, 1992).

In the case of AR, registration usually works similarly. However, here, one image contains virtual content and the other image is the participant's view of the real world. Both images are combined in a way that the virtual content appears to exist at the right position in the world. Using AR technology in the medical context, for instance, a virtual indicator might guide a surgeon in performing a surgery. This virtual indicator does not have to be aligned with a previously obtained image but has to appear at the right spot on the real patient. In this sense, virtual content is registered with the real world.[6]

In the context of AR, registration can be thought of as giving virtual content a position in the physical world. This not necessarily has to happen visually, but, for instance, could also involve aligning virtual

[6] One common form of AR that works a bit different is so-called spatial augmented reality. Here, virtual content is embedded into the real world directly (e.g., projected onto the world), and not just integrated into a participant's *view*.

sound sources with the real environment.

In AR research, the term registration is typically used to refer to *spatial* registration. However, registration also has a temporal component. For instance, Craig (2013), explains that in AR, the added information "is in both spatial and temporal registration with the physical world" (p. 20). Simply put, this refers to the fact that information has to appear at the right position at the right time. For virtual objects to appear at the intended location in space, AR systems often have to take the view of the participant into account, compute a corresponding view of the virtual object in real-time and display it with little latency. If a participant moves and there is too much latency, the virtual content will appear at the wrong position and 'lag behind'. This is why many definitions (e.g., Azuma, 1997; Azuma et al., 2001) argue that AR systems have to be run interactively and in real-time. One can argue that an accurate spatial alignment implies an accurate temporal alignment: Virtual objects are not displayed at the correct position in space if they are displayed at this position at the wrong moment.

If we are to believe existing research, registration is necessary for AR. Most importantly, the claim that AR requires registration is part of the often-cited definition of AR by Azuma (1997) and Azuma et al. (2001) (see above), which describes AR in terms of systems that, among other things, align/register virtual and real objects with each other. In his original review, Azuma (1997) illustrates the implications of this requirement and suggests that AR does not include "[t]wo-dimensional virtual overlays on top of live video" because "the overlays are not combined with the real world in 3D" (p. 356). By now, Azuma (1997)'s definition of AR is commonly accepted (cf. Zhou et al., 2008), and with it, so is the need for registration.

The notion that AR requires registration is, for instance, shared by Thomas (2009), who describes AR as "the process of a user viewing the physical world and virtual information simultaneously, whereby the virtual information is registered to the physical worldview" (p. 105). Craig (2013), too, sees AR as "a medium in which information is added to the physical world in registration with the world" (p. 15). He later argues that in AR, the added information "is in both spatial and temporal registration with the physical world" (p. 20).

According to existing research, registration is not only necessary for AR—it also distinguishes AR from other related phenomena. For instance, Piekarski and Thomas (2004) mention registration as a distinguishing factor between AR and VR: "Although AR and VR systems share some similarities, AR is unique in that it requires the registration of the physical and virtual worlds" (p.164). In a somewhat similar line of thought, Craig (2013, p.30) uses the requirement of registration in order to distinguish AR from the more general field of Mixed Reality:

> Many people use the term mixed reality interchangeably with augmented reality. However, in this book I consider mixed reality to be

> a broader interpretation that consists of anything of both the physical world and the digital world. The specific constraint of registration is relaxed.

Furthermore, Bimber and Raskar (2005), mention the lack of registration as a reason why a TV showing a cartoon or a radio playing music are no AR displays.

2.1.4 *Composite Views*

The three ideas about AR appear in different combinations. For instance, the current definition of AR in the Oxford English Dictionary (*Augmented Reality* 2005, accessed 07-05-2016) combines the encountered notions and defines AR as "a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view".[7]

[7] We wonder whether *composite view* refers to the fact that visual additions are not simply displayed on top of what a participant sees but also spatially integrated, resulting in one coherent seamless view rather than an additional layer on top of the world.

On first sight, the various reviewed descriptions complement each another perfectly and, together, draw a clear picture of AR: Augmented reality is a technology that combines virtual content and the real world by overlaying virtual imagery onto our view and spatially registering it with the real environment. This process happens interactively and in real-time.

On second sight, however, this rather clear image of AR is somewhat simplified and generalized. In reality, ideas about AR are more varied and complex. If we take a second look at the AR research landscape, we can find many more descriptions of AR, many of which differ from or even oppose the previously stated, popular image of AR as a technology that superimposes virtual images on a user's view and provides a composite view.

2.2 *Less Common and Diverse Views*

AR research generally agrees that AR involves both the real world and some kind of additional—so-called virtual—information. Aside from this fundamental agreement, opinions about AR vary.

In particular, views differ with respect to (1) what AR is, including the question whether AR indeed is a technology, about (2) the nature of the virtual content, including the question whether *visual overlays* are actually defining for AR, about (3) the way the virtual and the real relate to one another, including the question whether AR really requires registration. Furthermore, there are many different ideas about (4) the real in AR. E.g., there are different views about the role of the user/participant in AR, and different ideas about what is actually augmented in augmented reality.

2.2.1 *The Nature of AR: A Technology?*

If we take a closer look at previously reviewed works, AR has not only been described as a *technology* (cf., e.g., Doyle et al., 1998; Reiners et al., 1998; Roberts et al., 2002; Zhou et al., 2008), but also as a mixed reality *display environment* (Milgram and Kishino, 1994), as *a system's process* of overlaying computer-generated imagery onto (the view of) the real world (cf., e.g., Piekarski and Thomas, 2002; Rosenblum, 2000) and as *a user's process* of viewing the real world and virtual information (which is registered with the real world) at the same time (Thomas, 2009). If we take a look beyond the previously reviewed papers, we can find even more views: Wikipedia's current description sees AR as "a live direct or indirect *view* [italics added] of a physical, real-world environment whose elements are 'augmented' by computer-generated or extracted real-world sensory input such as sound, video, graphics or GPS data" (Augmented reality, n.d.). Furthermore, Spence and Youssef (2015) describe AR as "an *experience* [italics added] of a physical, real-world environment whose elements have been augmented, or supplemented, by computer-generated sensory input" (p. 1). In addition, Klopfer and Squire (2008) describe AR as "a *situation* [italics added] in which a real-world context is dynamically overlaid with coherent location or context sensitive virtual information" (p. 205) and Graham et al. (2013) refer to AR as "the material/virtual nexus mediated through technology, information and code, and enacted in specific and individualised space/time configurations" (p. 222).

While these views simply present different ideas about the nature of AR, Craig (2013) explicitly opposes the trend of approaching AR as a technology. In his book "Understanding Augmented Reality" (2013), Craig writes:

> Throughout the entirety of this book, I consider augmented reality to be a medium, as opposed to a technology. By medium, I mean that it mediates ideas between humans and computers, humans and humans, and computers and humans. [...] By taking the stance that augmented reality is a medium, it will become much clearer how the technologies involved can be used to create compelling applications for a variety of purposes instead of as a mere technological novelty. (p. 1)

2.2.2 *The Virtual in AR: Beyond Visual Overlays*

While researchers generally refer to *virtual* content that is added to the real world, few researchers take the trouble of explicitly defining what they mean with "virtual". Notable exceptions are Milgram and Kishino (1994), who explicitly discuss the differences between the virtual and the real on three dimensions. They distinguish between (1) real and virtual objects, (2) direct and non-direct viewing and (3) real and virtual images. With respect to the first distinction, they consider real objects to "have an actual objective existence" (p. 1324) and virtual objects to "exist in essence or effect, but not formally or actually"

(p. 1324). Their second distinction becomes clear when we consider the viewing of real objects. Real objects can be viewed directly (e.g., through the air) or they can be sampled (e.g. filmed with a camera) and then reconstructed via a display (e.g., played back on a monitor). According to their definition, virtual objects cannot be sampled directly and thus always have to be synthesized. The final distinction takes into account whether an image has luminosity at the location where it appears to be located in the space. In contrast to a real image, a virtual image has no luminosity at the spot where it appears. Virtual images can not only exist as contents of a digital display—other common examples are mirror images and holograms.

If we consider AR literature in its entirety, the term 'virtual' is generally used to refer to the intangible or non-physical. In contrast, the real stands for materiality and physical existence. More specifically, the term 'virtual' is typically used to refer to *computer-generated* content (see, e.g., Azuma et al., 2001; van Krevelen and Poelman, 2010). Aside from these general trends, we find many descriptions of what forms this virtual content can take.

Commonly, the virtual is considered a *visual* overlay. However, at the same time, many AR researchers point out that AR is *not* limited to just visual additions, and hence not constrained to virtual visual overlays. For instance, Azuma (1997), in his widespread review of AR (see above), mentions that "Augmented Reality might apply to all senses, not just sight" (p. 361) and suggests that "AR could be extended to include sound" (p. 361) by sensing the world with microphones and adding synthetic 3D sound. Azuma et al. (2001) reaffirm this and point out that they do not limit their definition to the sense of sight. They emphasize that "AR can potentially apply to all senses, including hearing, touch, and smell" (p. 34). (In line with this, the definition by Azuma (1997) and Azuma et al. (2001) does not refer to the overlay of visual content but more generally, the combination of the virtual and the real.)

Milgram and Kishino (1994), in their popular paper on Mixed Reality Visual Displays, also briefly refer to the possibilities of mixing and spatially aligning computer-generated spatial sounds with natural sounds in the environment, as well as mention the possibilities of haptic and vestibular AR, both of which provide non-visual additional stimuli.[8] Aside from these explicit statements, many other researchers indirectly suggest that non-visual and multimodal content can play a role in AR, for instance, by providing broader definitions. One example is Craig (2013), who speaks of "digital information" that is overlaid on the physical world rather than of visual content and our view of the world.

[8] With haptic AR, Milgram and Kishino (1994) refer to haptic displays where synthetic haptic information is superimposed on existing haptic sensations. With vestibular AR, they refer to synthesized information about the participant's acceleration that contends with existing gravitational forces.

Like Craig (2013) (see above), many researchers see AR as a field that deals with *digital* additions to the real world. The possibility of virtual content taking on non-digital forms is usually not considered.

An exception is the article "Pre-Digital Augmented Reality" by Lamers (2013) which discusses the *Pepper's Ghost* effect as an early (and non-digital) form of AR. This effect makes use of a second hidden room, glass (or comparable materials) and special lighting in order to let virtual objects appear or disappear in a room, to change their transparency or to morph different objects into one another (Pepper's ghost, n.d.).[9]

Although most researchers seem to associate virtual content with digital content, researchers generally have different ideas about what *exactly* is added to the real world in AR. To mention just a few examples: when speaking of AR, various researchers refer to the addition of virtual *objects* (e.g. Azuma, 1997; Milgram and Kishino, 1994). Others refer to the overlay of *computer-generated images/imagery* (e.g. Piekarski and Thomas, 2002; Rosenblum, 2000) or to *synthetic sensory* information (Vallino, 1998).[10] Often, the differences lie in the details: Whereas Azuma (1997) summarizes AR as a field "in which *3D virtual objects* [italics added] are integrated into a 3D real environment in real time" (p. 355), others refer to information that is not necessarily *3D* nor necessarily considered an *object*. For instance, Kounavis et al. (2012, p.1) refer to superimposed "computer-generated data, such as text, video, graphics, GPS data and other multimedia formats [...]".[11]

Furthermore, *some*—but certainly not all—authors share ideas about the behavior of virtual content. For instance, Craig (2013) claims that AR not only allows us to *perceive* virtual content, but also, that we can interact with it in the same way as we interact with physical objects (e.g., Craig, 2013), or in other words: that the virtual content is *interactive* content. Furthermore, it has been suggested that virtual objects should behave like real objects: Herling and Broll (2011) state "[i]deally, the virtual content would behave exactly like real objects" (p. 255) and (S. Kim et al., 2011) write "[i]n 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). At the same time, AR and VR pioneer Sutherland (1965), points out that virtual objects can behave differently from real objects. In his vision of future computer displays from 1965, he claims that "[t]here is no reason why the objects displayed by a computer have to follow the ordinary rules of physical reality with which we are familiar" (p.2).[12]

In general, we can notice a quest for realism on the one hand and a pursuit of objects that are unlike real objects on the other hand. For instance, Azuma (1997) mentions that virtual images "do not necessarily have to be realistically rendered in order to serve the purposes of the application" but that "[i]deally, photorealistic graphic objects would be seamlessly merged with the real environment" (p. 366). According to him, "the ultimate goal will be to generate virtual objects that are so realistic that they are virtually indistinguishable from the real

[9] The *Pepper's Ghost* effect can be traced back to 1584, when Porta first described an illusion called "How we may see in a Chamber things that are not" as part of his 20 volume book "Magia Naturalis" (Pepper's ghost, n.d.; Porta, 1658). However, the phenomenon has been popularized by John Pepper in the 1860s and is nowadays known as the Pepper's Ghost effect.

[10] Of course, these opinions are not necessarily exclusive.

[11] However, Azuma (1997), too mentions the overlay of text in combination with 3D wireframes.

[12] It should be noted that Sutherland (1965) did write about virtual objects in general as opposed to virtual objects that are supposed to appear in real space in. We will later raise the question whether virtual objects have to behave like real objects in order for this illusion to work (see chapter 5).

environment" (p. 380). In contrast, Craig emphasizes the possibilities of creating things that are different from real objects: "Indeed, one of the more interesting aspects of AR is that anything that can be created digitally [...], whether permutations of physical objects or objects that could not exist in the physical world" (p.18).

Finally, there are different opinions about *the role* that virtual content can play in AR. It is generally assumed that the virtual augments or extends the real world. However, there are other possibilities. More specifically, the virtual can also 'diminish' the real world and seemingly remove real elements from the perception of the participant. This is typically referred to as "diminished reality". Diminished reality is sometimes seen as its own field of research (e.g., Herling and Broll, 2010). Yet, diminished reality is also considered a subset of AR (e.g., Azuma et al., 2001).

As we have shown, there are many views about the virtual. While there is no generally agreed upon definition, AR is commonly associated with digital or computer-generated content that appears in real space. One might wonder: Why then, are advertisement screens and public information boards that display digital content in many cities, not considered a form of AR? AR researchers commonly agree that AR requires more than the mere combination of the virtual and the real in real space—that a stronger link between the virtual and the real is necessary. In particular, many researchers believe that the virtual has to be spatially integrated or registered in real 3D space. But is this really the case?

2.2.3 The Link Between the Virtual and the Real: AR Without Registration?

Judging from existing research, the registration of virtual content in real 3D space is widely accepted as a defining and necessary characteristic of AR. However, at the same time, an explicit interest in AR without registration is beginning to surface: The call for papers of the 14th edition of the International Symposium on Mixed and Augmented Reality (ISMAR 2015)—the leading conference on AR—lists *"augmented reality without 3d registration"* as one of the two emerging areas of particular interest and states 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." (ISMAR2015, n.d.).

In addition, there are views on AR that, although requiring relationships between the virtual and the real, do not require a *spatial* relationship between the two. For instance, Manovich (2006), refers to "dynamic and context-specific information" (p. 222) and claims that "a typical AR system adds information that is directly related to the user's immediate physical space" (p. 225). Similarly, Klopfer and

Squire (2008) speak of "coherent location or context sensitive virtual information" (p. 205). While these claims suggest that the information has to be *contextually* related, they do not claim that the information has to be *spatially* embedded in real 3D space.

In its totality, existing literature suggests that registration plays a key role in AR. At the same time, the different opinions about registration leave us wondering whether a spatial alignment of virtual and real objects is indeed always necessary. The above-reviewed positions show that AR can be approached more broadly and suggest that relationships between the virtual and the real in general (rather than only spatial registration in particular) play an important role in AR. We will take up the question whether registration is always necessary in chapter 3.

2.2.4 *The Real in AR*

AR researchers agree that the real world plays an important role in AR. However, opinions differ with respect to what role exactly the real world plays and what aspects of the real world are of importance.

As shown before, existing AR research is very focused on vision. In line with this, AR is often discussed in terms of virtual additions that are added to our *view* of the real world (Piekarski and Thomas, 2002; Roberts et al., 2002) or to *images* of the real world (Reiners et al., 1998). However, renowned AR researchers like Azuma (1997) and Milgram and Kishino (1994) suggest that AR not only applies to the visual sense. This view not only entails that non-visual virtual content can be added to the real world—it also means that non-visual aspects of the real world can play a role in AR. For instance, Azuma (1997) mentions the possibility of AR technology that makes use of microphones to sense the real sound environment (and that adds synthetic sounds or cancels out real sounds from this real sound environment). Furthermore, he considers the possibility of augmenting *the feel* of a real desk (rather than its visual characteristics), for example, "making it feel rough in certain spots" (p. 361). Milgram and Kishino (1994), too, mention the possibility of synthetic haptic information mixing in with real haptic sensations as well as synthetic sound sources mixing in with real auditory signals from the environment. As mentioned previously, the authors briefly consider vestibular AR, where the participants are affected by a mix of real gravitational forces as well as synthesized (vestibular) information about their bodies' acceleration.

It is a common understanding that some AR technologies allow participants to perceive the real world directly (via air or glass), while other technologies allow participants to perceive the real world in a mediated form (e.g., on an electronic screen) (see, e.g., Azuma, 1997; Milgram, Takemura, et al., 1994).[13] However, opinions differ as to whether the perceiver of the real world is merely an outside observer

[13] For instance, users can see the real world directly (via glass) with a so-called optical see-through HMD. In contrast, video see-through HMDs make use of (a) camera(s) to provide the user with a video of the real world that contains virtual elements (see, e.g., Azuma, 1997).

or an active part of this world: Several researchers simply claim that AR allows a user to *see* the real world (or a live video thereof), while others claim that the user/participant is also present in the environment, and hence, can not only see it but also act in and interact with this world (and, as we would like to add, perceive it with all their senses rather than just *see* it). For instance, R. Silva et al. (2003) consider the technology responsible for the television weather report AR because the real image of the news reporter (who actually stands behind a blue screen in the studio) is augmented with a virtual map. Similarly, many researchers (e.g., Van Krevelen and Poelman, 2010) mention sports broadcasting systems that embed virtual "first down" lines in the live broadcasts of football matches as examples of AR technology.[14] These systems allow the viewer at home to passively see a real environment, with virtual content superimposed and integrated into the environment. However, the viewer is not part of this depicted space. In contrast, Craig (2013) emphasizes that AR allows participants to "engage in an activity in the same physical world that [they] engage with whether augmented reality is involved or not" (p. 1) and clearly state that "[a]ugmented reality is interactive, so it doesn't make sense to watch it or listen to it" (p. 2). Likewise, Azuma (1997) points out that (in contrast to VR) "AR allows the user to see the real world" (p. 356), but also that "AR requires that the user actually be at the place where the task is to take place" (p.366).[15]

[14] In football, the virtual first down line marks how far the offense has to advance to gain a so-called 'first down'.

[15] Azuma assumes that the virtual content helps a user to perform a real-world task.

Finally, while researches generally agree that AR combines the virtual and the real, there is surprisingly little consensus on what is actually augmented by this virtual content. Many argue that it is the *perception* of reality that is augmented (e.g., Normand et al., 2012; Ross, 2005). Furthermore, there is the notion that in AR, the *physical world* (Craig, 2013) or our real physical *environment* (Milgram and Kishino, 1994) is augmented. At the same time, Milgram and Kishino (1994) also refer to the augmentation of the *display* of an otherwise real environment. In addition, there is also the notion of augmented *space* (Manovich, 2006). Wikipedia's current definition of Augmented reality (n.d.), provides yet another different perspective and describes AR as "a live direct or indirect view of a physical, real-world environment whose *elements* [italics added] are 'augmented' [...] ". Furthermore, Mackay (1996) also approaches this question in another way and considers the carrier of the physical equipment as augmented (e.g., the user is augmented when he/she carries a helmet and an object is augmented when sensors are embedded in it). Consequently, she distinguishes between an augmentation of the *user*, an augmentation of the *physical object* and an augmentation of the *environment* surrounding the user/object.

2.3 Conclusion

We have reviewed existing AR literature with the goal of learning more about the nature of AR, asking the question "What is augmented reality". Did we find an answer? Yes and no! The review has provided an answer insofar as it has revealed several widely accepted views and common notions about AR. At the same time, our review has also revealed opposing and diverging views that make us doubt these widely accepted views. In many ways, our review mirrors the results of our first look at AR examples (see chapter 1): On the one hand, AR seems to involve technologies that integrate virtual images into our view of the real world. On the other hand, AR seems to take many other forms as well.

Existing research largely agrees that AR combines virtual content and the real world. More specifically, our review has revealed three common notions of AR. First, AR is generally considered a technology or system.[16,17] Second, AR is understood in terms of visual virtual overlays that are placed over a user's view. Third, AR is considered to align virtual and real objects in physical 3D space. These views are not at odds but complement each other well. Together, they reveal a rather clear image of AR as a technology that integrates virtual imagery into our view of the world.

[16] Whereas 'system' is a broader term than 'technology', we here use technology and system as interchangeable. This is because here, both refer to technological systems consisting of both hardware and software components

[17] Yet, AR is usually not limited to a specific hardware or device. Instead, AR is characterized by what the system *does*.

This concept of AR covers many of the actual AR examples encountered in the previous chapter. A typical example of a system that embodies all of these aspects is the see-through head-mounted display technology that Caudell and Mizell (1992) proposed when they coined the term *augmented reality* (see chapter 1).[18] As discussed, their proposed system helps manufacturing and assembly workers by overlaying virtual guides and instructions onto their view of the real world. At the same time, their system also registers/aligns the virtual content with the real world. As a result, the virtual information appears to have a location in the real, physical world. For instance, a virtual arrow might indicate where to drill a hole (see figure 1.1).

[18] Caudell and Mizell (1992) refer to so-called "HUDsets", which refers to heads-up (see-thru) display head set. We use the term HMD (head-mounted display) as it is more commonly used for the type of display depicted by the authors and for consistency reasons.

If we consider both our first look at AR works from the previous chapter as well as existing notions of AR reviewed in this chapter, there is no doubt that technologies that integrate virtual imagery into our view play an important role in AR. However, we also have encountered views that diverge from this notion and that suggest that AR can take different forms as well. First of all, not everyone agrees that AR is a technology. Among other things, AR has also been described as an environment (Milgram and Kishino, 1994), situation (Klopfer and Squire, 2008) and experience (Spence and Youssef, 2015). Secondly, several definitions convey a more encompassing idea about the virtual and suggest that AR technology can work with non-visual content and engage all senses. In contrast to definitions that focus on visual overlays, modalities-encompassing definitions capture a broader variety of

AR works. For instance, they include works that augment real objects with virtual tactile textures (e.g., Bau and Poupyrev, 2012). Third, not all views require 3D registration. Several descriptions and definitions of AR simply call for a relationship between the virtual and the real. Such definitions also describe a broader field. If we do not demand 3D registration, AR can, e.g., include audio guides that automatically inform us about art pieces in a museum (e.g., Bederson, 1995).

Altogether, we have encountered a variety of different views regarding the nature of AR, the qualities of virtual content, the link between the virtual and the real, the role of the real world in AR, the position of the participant and the question of what is augmented in AR. These varying descriptions and definitions do not provide a clear answer to our question of what AR is. Rather, the different reviewed views give rise to several more fundamental questions: Does AR require visual overlays? Is AR interactive? Can we experience AR remotely? What forms can virtual content take? What is the role of the participant in AR? How is the real world involved? Is AR something we experience with our eyes, or something that engages all our senses? What is actually augmented in augmented reality? Is registration really necessary?

Ultimately, our review leaves us with two options. Either (1) the widespread image of AR as a technology that overlays virtual imagery onto our view and aligns it with the world is correct. In this case, we are simply confronted with many works that have wrongfully been labeled AR, as well as many definitions and descriptions that have failed to capture the essence of AR. Or (2) this widespread picture of AR is incomplete: it captures common characteristics of AR, but also disregards many other possible manifestations of AR.

In this thesis, we consider the second option and take the position that AR goes beyond this—arguably—stereotypical image. We build on the belief that augmented reality is—or potentially can be—more than just a technology that overlays virtual imagery onto our view, thereby providing a composite view (*Augmented Reality* 2005).

Whereas the common image of AR as a technology is pretty clear, this broader conception of AR remains rather blurry. Although we have encountered a variety of concrete examples (see chapter 1), it is unclear what forms AR can take. For instance, it is unclear whether and how AR might work with virtual tastes. Is it actually necessary and possible to register "virtual tastes" in three dimensions and in real space, similarly to how visual objects are integrated into the environment? Intuitively, this seems weird. After all, taste is not something we experience in three dimensions.[19] More generally, we wonder what characterizes AR, if not a system that integrates virtual imagery into our view of the world.

In the following chapter, we will explore this question and propose an alternative understanding of AR. We will approach AR from a deliberately broad perspective that does not limit AR to a certain set of

[19] Yet, it might be possible to make a certain taste appear in a participant's mouth, whenever he or she crosses a certain spot in space or places a certain object in their mouth.

technologies, that does not require 3D *registration* of virtual and real elements and that does not limit AR to the addition of *visual* information. In line with this, we will also treat the virtual as a broad and elastic concept that encompasses stimuli that either have been synthesized or that do not directly originate from their original source. This notion of the virtual includes, but is not limited to, computer-generated simulations. For instance, it also includes the possibility of treating audio recordings or perfume as virtual stimuli.

Chasing virtual spooks, losing real weight

Augmented running and a side trip into the history of audio augmented reality

A strange voice tells me to run. My heartbeat rises as I follow the instructions without giving them a second thought. The voice's manner of speaking reminds me of my parents' TomTom. The only difference: instead of telling me to take a turn, I am instructed to accelerate, slow down, to run or—if I am lucky—to walk. I am running with my new mobile app and virtual trainer. The app tracks every move, knows when my heartbeat rises and is supposed to help me gain speed and lose weight. Today, I run to clear my head after a mentally exhausting but physically unchallenging day. However, trying to catch my breath, my thoughts return to work. More precisely, I pore over my research topic, non-visual augmented reality.

In augmented reality (AR), virtual content is added to our real environment. Most often, this happens visually. By now, probably all of us have seen some three-dimensional objects popping up upon designated markers, virtual pink bunnies above augmented cereal boxes or walking directions superimposed on real streets. However, AR does not have to be visual. Sound, in particular, has already brought forth some fascinating AR applications and artworks such as Edwin van der Heide's *Radioscape* (2000-) and Theo Watson's *Audio Space* (2005). Entering the latter, visitors can hear the sounds left by previous visitors, spatialized, as if they were actually still there. At the same time, they can leave their own audio messages at any point within a room. It is not just the fact that the physical space is augmented with the ghost-like presence of previous visitors that makes me term this work AR. Visitors can also relate their own sounds and messages to those left earlier by others; thereby establishing connections between the virtual and the real. I imagine walkers, cyclists and other runners leaving their sound-trails behind on the road, leaving it up to me to add my own sounds and follow their steps, which are spread across time and space.

My favorite mobile app, *RjDj* (Reality Jockey Ltd., 2013), can also be considered AR sound art. The app remixes the sounds of the surroundings and provides you with a soundtrack to your life that blends in, makes use of and accompanies your environment. Although it is certainly no typical AR application, the relation between the sounds of the real environment and those produced by the app is so strong that often, they seem to melt into a single soundscape.

I will have to try this app while running. I can already hear the sound of my steps on the asphalt evolving, blending into a rhythmical soundscape, slowly displaced by the wind or heavy breathing, interrupted by pitched variations of my sudden greetings whenever I meet another runner.

While *RjDj* (Reality Jockey Ltd., 2013) and successor apps like *Inception - The App* (2016) and *The app formerly known as H _ _ r* (2016) are a rather recent phenomenon, the idea of remixing the sonic environment is not new. The artist Akitsugu Maebayashi has worked with similar concepts for a long time. His portable *Sonic Interface* (Maebayashi, 1999) was built in 1999—years before mobile phones gained comparable sound-processing abilities. The custom built device consists of a laptop, headphones and microphones and uses delays, overlapping repetitions and distortions in order to recompose ambient sounds in urban space. The resulting soundscapes break the usual synchronicity between what one hears and what one sees. Unsurprisingly, Maebayashi is not the only one who has been exploring sound-based augmentations of the environment early on. In fact, audio augmentations of our environment have quite a history of their own. Unfortunately, they are less known in the context of AR and are often not even considered to be part of AR history.

"Walk!", my virtual trainer gives in to my exhaustion and I slow down. However, my thoughts keep racing. Quickly, they approach the early 1990s: Tom Caudell is believed to have coined the term *augmented reality*. It describes a head-worn display that superimposes visual information onto real objects (Caudell and Mizell, 1992). In Caudell's case, the new AR system helps workers assemble cables into an aircraft at Boeing. What usually goes unnoticed is that around the same time, Janet Cardiff started recording her so-called audio walks. Those walks are designed for a certain walking route and confront the listener with instructions such as "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 the listener navigates the space, he gets to listen to edited mixes of pre-recorded sounds, which blend in with the present sounds of the environment. Cardiff's virtual recorded soundscapes mimic the real physical one "in order to create a new world as a seamless combination of the two" (Cardiff, n.d.). By superimposing an additional virtual world onto our existing one, and thereby creating a new, mixed reality, Cardiff's sound art explores one of the key concepts of AR. And Cardiff is not alone with this idea; as early as 1987, Cilia Erens introduced sound walks, soundscapes and sound panoramas in the Netherlands. In contrast to Cardiff, she forgoes spoken content and uses largely unmixed everyday sounds. Yet, the effect is similar; they create "a new reality within existing realms, a form of 'augmented reality'." (Erens, n.d.) Clearly, the developments in non-visual AR were in no way inferior to the development of their visual counterparts. Taking slow steps, I imagine being on such a walk right now... Listening to instructions on which route to take, where to look, superimposed footsteps, sounds recorded here, on this path earlier, maybe altered with special effects. I imagine those sounds mixing in with the naturally present sounds of the river, bikes, and the occasional mopeds passing by.

"Run!", my trainer, whom I decide to call Tom, puts an abrupt end to this walk. The fact that AR sound art like Cardiff's and Erens' walks are not usually mentioned in the context of AR leaves me wondering what else we miss.

After Tom's instruction, my music fades back in. The song is intended to get me to run even faster. After my footsteps have adapted to the new rhythm it hits me: these instructions about how fast to run, the information about my heart rate, distance covered and calories burned and options such as racing against a virtual running partner in real physical space—they are just like AR.

In fact, my virtual running trainer shares most of the characteristics commonly found in AR applications. It adds another layer of content to my running. It is interactive and operates in real-time (cf. Azuma, 1997). Just like many other GPS based AR applications, it reacts to my position in the world. Most importantly, Tom fulfills my own, personal requirements for an AR experience: there is a relationship between the additional layer of content (the information I receive) and the real world (my running).

When another runner passes me slowly, my heart rate drops. I wonder whether it might be his heart rate that is mistakenly reported back to me. I am astonished, that without the sensor's help, I cannot even accurately perceive such basic and vital facts as my very own heart rate. Maybe this is farfetched, but with respect to that, the running app relates to the kind of AR applications which allow us to perceive things about the world that we normally cannot perceive, such as seeing heat, feeling magnetic fields or hearing ultra-high frequencies. (This idea is also discussed in section 4.8.) So why are virtual Tom and his colleagues not considered to be AR?

Perhaps because there are also numerous differences between running apps and typical AR applications. To begin with, this running app does not augment the environment. Rather, it augments an activity—my running. And to be honest, despite the fact that Tom follows my every move—chasing a virtual competitor or running with a virtual trainer—it still feels like they are running on my phone while I have to tackle the real road. What is more, location-based AR applications usually display content related to the user's absolute position in the world. Tom, on the other hand, is only interested in the change of my position over time.

"Stop!", apparently, my position has changed enough. My run is over. The result: more than 500 kcal burned, five miles run and the revelation that the combination of the virtual and the real encompasses much more than just adding virtual visual objects to the real physical environment. There is a whole field of augmented activities as well! I cannot wait to jam with virtual bands, to try augmented eating or to take an augmented nap. As if to approve, my heart rate makes a last excited jump. Who knows, in the future, Tom might learn from existing AR. He might then have a look at my environment and direct my turns so that I discover new routes, point out sights or, when needed, help me find a shortcut home. Considering current developments in lightweight AR glasses, I guess it cannot be long until we can also see our virtual competitor passing by, are asked to design avatars representing our personal best time in races against other runners and are challenged to chase visual virtual spooks. I would not mind that. And I bet that that is when augmented running will be truly considered to be AR.