Augmented, Virtual and Mixed Reality Simulation, Vol. 118, 2023, 30–39
https://doi.org/10.54941/ahfe1004438

Toward a New Definition of Augmented Reality

Anton Nijholt

University of Twente, Faculty EEMCS, Human Media Interaction, Enschede, The Netherlands

ABSTRACT

In 1997 Ronald T. Azuma introduced a definition for augmented reality. The definition can be considered slightly outdated because of developments in augmented reality and ubiquitous computing. Extended reality environments do not only allow interactive virtual objects superimposed on reality and aligned with reality, but also static, dynamic, and autonomous virtual content that is not under the control of the user of the environment. One aim of AR research is to superimpose (multisensorial) virtual objects on reality that cannot necessarily be distinguished from real objects that are perceived and experienced by the inhabitants of the environment. In this paper, we take it a step further. Especially if we are no longer able to distinguish between virtual and real objects, shouldn't we look for a definition of AR that is more based on experiencing (not necessarily technology-enhanced) reality than on technology? We do this by focusing on multisensorial experiences that augment our world, rather than on the technology, present or not, that enables these experiences and distinguishes our experiences from those of others. That such a viewpoint has not taken shape before is mainly due to the vision-biased view of what AR research should entail.

Keywords: Augmented reality, Extended reality, Multisensorial augmented reality, Vision, Scent, Sound, Alignment, Shared augmented reality, Ever-present augmented reality

INTRODUCTION

It is useful to start with Ronald T. Azuma's 1997 definition of augmented reality (AR): (1) AR combines real and virtual objects in a real environment, (2) registers (aligns) real and virtual objects, and (3) runs interactively, in three dimensions, and in real-time (Azuma, 1997). In (Azuma, 2019) we find the prediction that AR technology, in particular with optical see-through glasses, will be the dominant platform and interface, supplanting the smartphone, for accessing digital information.

Since that date, there have been many developments in AR, Virtual Reality (VR), and ubiquitous computing. That is, nowadays AR should be considered in the context of ubiquitous computing, integrated with ubiquitous computing and therefore being able to communicate with sensors in the environment rather than being limited to intelligence that is only available in a user's smartphone or AR interaction tablet. Moreover, although Azuma mentioned that the definition was technology-independent, it was biased toward digital technology and vision-oriented AR. Nowadays we can still take Azuma's

definition as a useful starting point, but we need to be aware that (1) whatever we introduce in AR can be enhanced by sensors and actuators that are present in the AR environment, (2) rather than demand that virtual content is interactive, we should allow virtual content that is beyond the control of the AR user: dynamic content that is introduced and controlled by the system, virtual content that acts autonomously (for example, a virtual human inhabiting the AR environment (Nijholt, 2021)), or content that is controlled by other users in a shared AR environment, and (3) we must say goodbye to the vision-based interpretation of AR; our sound, scent, taste, touch, and proprioceptive senses need to be covered by an AR definition as well. In this paper, we focus on vision, scent, sound, touch, and proprioception to make clear why we need a more than vision-oriented and non-technology-based definition of AR.

To do so we distinguish between individual AR use and experience, shared AR use and experience, and others, present among these AR users, but not able to share their experiences, for whatever reason, including, among other things, not using AR technology. We will conclude that an AR definition should focus on different experiences rather than on technology.

In the next section, we have some general observations on the development of AR, multisensorial AR, and indistinguishable AR. Rendering or synthesis of visuals, sounds, touches, and scents and their alignment with reality will be the topic of the next section. Next, we differentiate between AR and non-AR users to motivate our definition of AR. We will conclude that we need to define and discuss AR in terms of sensory stimulations (and associated experiences) than in terms of technology requirements.

AUGMENTED REALITY: FROM VIRTUAL TO REAL

Researchers have mentioned that reality has always been augmented (Sicart, 2017). They mention cave drawings, traffic signs, or similar human-produced activities and products to reality. That of course invokes the question of why such products and activity should not be considered as being part of reality. Is that because that reality, for example, cave drawings or traffic signs, are only experienced by a selected group of users and are not accessible to all of us? Or at least, not always part of our daily activities? When we make music, do we augment reality, for ourselves, our audience, or those who don't hear us? When we hear music, is that augmentation of our reality, or is it part of our reality? To explore these questions we will first have some observations on how AR is currently being looked at.

Traditional AR and our views on AR are vision-based. Digitally generated images are superimposed on our view of reality, made possible by head-worn devices. But what about artificially generated scents that are added to an environment or our bodies? What about food additives or genetically engineered food? Scents and tastes can also be digitally manipulated and provide a particular user or group of users with experiences that are not available to others. Is there any reason why we should treat such users differently from AR users?

Going back to the aforementioned definition of AR, we could say that AR requires that we can interact with the virtual content in real-time and that the virtual content must be aligned with the real content. In these examples, the latter is certainly the case. As for the former, that depends on what we mean by interaction. Moreover, as mentioned earlier, in AR we can have virtual content that cannot be controlled by us, although our senses can certainly be influenced by it and thus how we experience and react to it. Perhaps more implicit interaction than explicit interaction?

Looking at the alignment condition, it should be spatiotemporal but other alignment conditions, depending on the application, that is, semantically and pragmatically conditions, need to be added. Artificial, virtual, content, can be designed in such a way that there is no need, as is the case with artificially (digitally) designed visual content, to explicitly pay attention to how we experience this "virtual" content. Alignment can be implicit, interaction can be implicit.

We can go one step further. Content addresses our senses. Suppose we have an installation artwork with a real dog in it. Part of the dog can be painted, but apart from that, the 'user' or visitor of the artwork, can see, smell and hear the dog, and interact with it in real-time. There is a natural alignment of the dog's movements with the environment. It can partially disappear behind objects and reappear, its scent and its sounds come from the right direction. Moreover, our pose and gaze changes will be accompanied by perspective changes in appearance and sound and scent directions. We interact with the dog, even when the dog is not aware of it. Is there any reason we don't want to call our interaction with this dog a real-time AR interaction?

A reason may be that we think that the user has full control of the added content. Although interaction with the dog is possible and will happen, the dog in our example will also exhibit autonomous behavior. This cannot be a reason to say that this is not an AR environment. AR environments can have computer-generated 3D imagery that represents virtual humans or animals that have been given goal-oriented artificial intelligence and appear to exhibit autonomous behavior. We can also replace the real dog with Sony's Aibo dog, replacing a virtual living animal with a physical technological artifact. This does not change our observations on alignment and real-time interaction. Moreover, AR research focuses on having the ability to add content that is indistinguishable from the already present real content. That is, the user should not always be able to distinguish between the objects (visuals, sounds, flavors, touches, scents,...) in the 'original world' and objects that are added by the AR designer.

Living and material objects can satisfy the AR definition unless we say that by definition they are not virtual or digitally generated and therefore should be excluded. To define what is not virtual we should be able to define what is virtual. Moreover, in our dog example, there was no reason to worry about alignment issues such as the occlusion problems that we are familiar with in vision-oriented AR. Other content may require that alignment has to be done concerning the pose of the user or users and also require technology to realize this alignment. That is, manipulate the designed content – whether it is sound, visual imagery, flavor, touch, or scent – and control its spatiotemporal

display in the AR environment for an individual user or collaborating users where each has its own 'view' on the shared content.

VISUALS, SOUNDS, SCENTS, TOUCHES, AND PROPRIOCEPTION

To avoid our observations having a bias on our visual sense, in this section, we discuss augmenting reality with sound, scent, and touch experiences. We focus on sounds because adding sound and having real-time interaction with sound is illustrative for our view on extending reality in a way that our senses are explored by stimuli from this extended reality in whatever way these "extensions" are added, temporarily or not. For whatever reason and whatever technology, someone becomes consciously or unconsciously aware of these additions and explicitly or implicitly responds to them. To illustrate our views, we also need to look at our sense of proprioception.

Natural versus Augmented: From Sounds to Proprioception

In an environment with ambient sound or scent, the spatiotemporal alignment can be said to be 'ambient', not displayed from the perspective of a particular user. All users in the environment have the same experience and moving around does not change the experience and no explicit interaction with the scent or sound objects in the environment is possible. Although we can say that the environment, for example, a shopping mall or a particular shop, is augmented with sound and scent, such an augmentation does not differ from given a particular environment, again a shopping mall or a particular shop, a particular visual appearance. We, as users or visitors of these environments, unlike the owners, cannot make changes to the emitted sounds and scents or the environment's visual appearance. We are not in AR. Moreover, if we were familiar with the environment without the sound and scent augmentation, after a period of habituation, we do not consider the sound and scent additions as augmentations anymore, they are part of our real and familiar environment.

Suppose we sleep badly because of annoying noises at night. Or, suppose we do not want our confidential conversations to be heard. In the first case we can put a "white noise" device in our sleeping room, in the second case we can attach it to our office door. The device emits sounds, for example, the sound of rain on a tent. Does that make our sleeping room or our office environment an AR environment? The device is a technological artifact. We can interact with it, perhaps using a remote control, and through it with the sounds the device emits. We can turn the device on or off, adjust the volume, use a timer, and choose different sounds. We align, manually, the device in its semi-fixed location. Further alignment is done with the remote control. Whenever we want we can interact with the device. We use the device to interact with the sound it produces. The sound is as well aligned, perhaps in a trivial way, with the environment. Our use of the device to manipulate sounds can be compared with the use of an Optical-See-Through Head-Mounted (OST-HMD) AR device that captures our hand movements to control computer-generated imagery that is superimposed on reality in vision-oriented augmented reality. The white noise device is a technological artifact, just like an HMD. It can

be used to control and align sounds just as we expect an HMD to control computer-generated imagery that we usually refer to as virtual content. We interact with augmented reality.

In AR, we usually talk about a virtual layer that is superimposed on reality. The virtual layer has objects that activate our senses: the basic senses such as sight, sound, touch, smell, and taste, but also other senses (e.g., proprioception) and combinations of senses. Our sight or hearing senses have no way to distinguish light or sound waves coming from objects in the virtual layer from those of real objects. It is our interpretation – using semantic, pragmatic, common sense, and application-dependent knowledge – of these stimuli that decides what is native to the environment and what has been added. As mentioned above, AR research aims at making this distinction disappear, although not for all applications this is desirable. Below we will return to the issue of distinguishing between objects in the layer superimposed on reality and the objects native to the environment.

Whatever name we decide to use – virtual, fictional, non-native, added, designed – this content has been or has to be created, it should allow real-time interaction, and it should allow spatiotemporal control so that it can be aligned with objects that are native to the environment. The content can be dynamic and, as we mentioned for virtual humans or animals, show some autonomy. The alignment is from the perspective of the user or inhabitant of the AR environment. In the case of multiple users, each user has a pose-dependent perception of the shared content. We will provide a few examples in which we play with sounds and scents to illustrate our investigations of Azuma's AR definition.

Consider an extremely simple situation where we have a musician who brings a simple sound-producing device, for example, a mechanical or electronic metronome that is put on a table, in her rehearsal room. By moving around in the room the experience of the metronome sound by the player changes because of pose changes and room acoustics. This is not essentially different from perspective changes of computer-generated imagery in vision-oriented AR, apart from the fact that no technical means are needed. There is a natural alignment between the sounds and the musician and the room. However, when not in the immediate proximity of the metronome the musician can not interact from his current pose with the device in real-time. This leads us to the observation that in an AR situation as defined by Azuma, the interaction with the content that has been added to the physical environment has to be performed without being obliged to change pose or at least position. In general, this can be avoided by providing a user with a wearable that allows them to control (clicking real or virtual buttons, speech, gesture, or gaze commands) the non-native content.

Perhaps not so obvious, but also in the case of a non-digital metronome we can think of non-digital control (e.g., analogous voice recognition) from a distance of the options the device offers. The principle of this observation should be clear, the sounds can be produced in a non-digital way, the interaction does not need digital means, and the alignment of the produced sounds happens naturally, it does not require technological means. What prevents us

from considering this environment as being an AR environment? Because the sound is not virtual? But what does that mean, a sound not being virtual?

Can we distinguish between perceiving "real" sounds and "virtual" sounds? Is the sound that is perceived native to the real environment or has it been artificially added? There is no way to make this distinction when the sounds are created and aligned with the real world perfectly. The user can not make this distinction unless he decides that the sounds are "virtual" because they are imperfect, their alignment with the virtual and real-world objects is imperfect, or that the added content should be considered inappropriate, not fitting the situation, not fitting history, or not following physical laws. As an example, when our AR device lets us perceive a real or virtual cow and we hear her bleat like a sheep then this contradicts our common sense and we decide that the sound must have been generated artificially and is in the virtual layer superimposed on reality. As another example, an AR designer can assign a reverse Doppler effect to a virtual train or plane moving in the AR environment. In these cases, the context determines what is non-native and what belongs (see also (Nijholt, 2022)).

Let's have a look at a physical environment with several people present. A soundscape is added to this environment. It does not yet make it an AR environment. In this example, one person has been given the ability to interact with the soundscape. The person is recognized by the environment as the "user", sensors embedded in the environment or the user's wearable technology can capture her activities. Other persons in the environment are ignored by technology. The user's interaction with the soundscape (pose changes, gestures, speech,...) can be perceived by the other people present in the environment. But for them, since whatever they are doing, they cannot influence the soundscape, they are not in an AR environment. The interactant is. Her perception of the environment is different from others since she is the agent of the changes and contrary to others present in the environment, her proprioceptive sense plays an active role in controlling and perceiving changes in her AR environment. As mentioned in (Montero, 2006), "*The object of proprioceptive experience, one's own body, can be proprioceived only by oneself.*" Being able to privately interact with sound in this way distinguishes the experience of the interactant from that of the others in the environment and could be the reason for calling the sound as it appears to her virtual. Notice that with an OST HMD, the experience of its user can in a similar way be distinguished from the experience of others near her because her sight sense is involved in her AR experience while this experience is not available to the other attendees. Whether or not other senses are involved, the experience of the AR user is private, i.e., it can be distinguished from others that share the same physical environment.

Natural versus Augmented: Scents and Touches

Let's shift our focus to smells and the sense of smell. If we add smells to an environment we have similar issues as in the case of sounds. We can have a scent scape added to a physical environment that can be experienced by anyone present in that environment. Changes in the scent scape are more

difficult to realize than in a soundscape because scents need time to dissipate and be replaced by another scent.

Scent can be made available from scent cartridges. It is not digitally created. Cartridges can be attached to a smartphone or other wearable, e.g. close to the nose on an HMD. They can also be distributed in the physical environment where they can be triggered to emit their smells. If we want spatiotemporal control of scents emitted from a near-nose cartridge, the HMD should allow horizontal and vertical movement of the cartridges. That is, with eyes and ears closed, we should be able to say from which direction the scents reach us. If scents need to be aligned with real or virtual objects that are perceived by our senses in the AR environment, in a near-nose scenario we can profit from cross-modal effects. A rose, whether it is physically present in the environment or computer-generated, is supposed to give off a scent. When we perceive a particular object in the AR environment that is expected to give off a scent and we have a corresponding smell sensation from a near-nose cartridge at the same time, we tend to associate them. We may conclude that with near-nose scent emissions, the AR user can experience sensations that cannot be experienced by non-AR users who are present in the same environment. The experience of the AR user is private, i.e., it can be distinguished from others that share the same physical environment.

Can we come to the same conclusion when the scent sources are not near-nose? The spatiotemporal alignment condition is less easy to meet when the sources are not nearby. Scent sources can be positioned in an environment and are not necessarily visible to those present in that environment Hence, they can be where they will be associated with a virtual object that will be put and fixed in that position and aligned with the real-world objects on site. In this non-near nose scenario, and although scents spread over an environment, an AR user can have a different experience than other persons who are present in that environment and experience the same scent. The different experience can be caused because the AR user perceives the virtual object with which the scent can be associated and assumes the scent is coming from that direction, the earlier mentioned cross-modal effect. But the different experiences can also be caused because we can direct the emissions of odors (Yanagida et al., 2004; Yanagida et al., 2019) and although the scent can be perceived by others, we can designate the person to whom the emission is directed as the AR user in this environment, whether or not such a person has access to additional information in the AR environment through other senses. Hence, as in the case of sounds, also for scents we can distinguish the experience of an AR user from that of others that share the physical environment with the AR user. The experience of the AR user is private, i.e., it can be distinguished from others that share the same physical environment.

What about our haptic sense? We can feel the texture of an object, we can explore the shape of an object by touch, a moving object can press against us, we can feel the resistance when we press against an object, and we can feel the weight of an object. When these objects are present as material objects in our AR environment we can provide them with haptic characteristics that differ from how we experience these objects in the real, non-AR, world. Computer-generated imagery that has been added and aligned with the real

objects in the AR environment can also be given haptic properties that can be sensed with an AR device such as a data glove. In the case of mid-air haptics, focused ultrasound generates tactile experiences on a user's skin. Hence, no (wearable) AR device is needed. In the case of kinaesthetic sensing, a force feedback device provides the user with a haptic experience. Others, non-AR users present in the physical environment can not have the same experience. Even if these others use an OST HMD, and can see virtual (CGI) objects, their experience is different from the user with a haptic AR device. The experience of the AR user is private, i.e., it can be distinguished from others that share the same physical environment.

DEFINING SOMEONE'S PERSONAL AUGMENTED REALITY

Based on the foregoing, we do not want to characterize an environment as augmented because of a distinction between "real" and "virtual". We can distinguish between already present objects, perceptive for everyone, those that are added digitally, but also those that are perceived and experienced differently from others in a particular environment. In our view augmentation should not be defined in terms of digital technology but in having experiences that are not shared with others. For that reason, we rather speak of an augmented experience (AE) 'user', than an AR user. What important is, not how that experience came to life, but how that experience distinguishes someone who has that experience from others.

We conclude that an AE user is provided with perception and interaction capabilities that are not shared with others in the same environment. The AE user's experience of the environment is different from that of others and as a consequence, the AE user's behavior is different. We can talk about shared AE environments, but also in that case each user has a private view of the environment that can not be shared with non-AE users present in the underlying physical environment. An AE environment can be shared with other AE users. In that case, it should be clear which senses are involved and which sensory experiences are shared, taking into account each user's perspective. Although an AE user can share the same physical environment with others, and even share a particular sense-oriented AE environment with others, an AE user can have a sense-oriented perspective of the environment that is not shared with others that are present in the environment.

We should also comment on the (real-time) interaction issue in Azuma's definition. We can distinguish different levels of interaction. A change of pose (position, head, and gaze orientation) concerning the stimuli sources leads to experiencing a realignment of the stimuli following this new perspective, whether this is done in a physical, or in a digital way.

In the case of optical see-through AR, the computer-generated imagery needs to be realigned with the real objects in the perceived AR environment. In our previous examples of scent and sound-augmented environments, a change in a user's pose should affect their perception, either by manipulating their reception or their generation. Also for these modalities, changing a pose can be considered a low-level interaction with the environment. In a smart environment full of sensors and actuators a change of pose also requires

realigning of the smart technology that is present. We can say this technology is present in an 'augmentation layer' and changes this can affect the physical objects in the environment, for example by the triggering of an actuator.

Whether or not wearing and using AR devices, any environment, digitally enhanced or not, can provide its inhabitants with augmented multisensorial views that are not necessarily shared with others that are present in the same environment. This is true for non-digitally enhanced environments where an inhabitant experiences certain stimuli different from others, for example, have its proprioceptive sense stimulated by the music of a street musician in such a way that he starts dancing, while this stimulus does not necessarily affect others in the same environment. And it is true for smart environments where smartness provides each user with a personalized view of the environment that is not necessarily shared by other users. These environments also assume real-time interaction with the embedded technology, whether the interaction is implicit (monitoring a user's behavior and adapting the environment's smartness to this behavior), or explicit (i.e., consciously performed interactions with the environment.

CONCLUSION

From the previous sections, we decide that AR experiences should not be defined in terms of technology. This decision goes a step further than (Azuma, 1997) who in his definition relied on digital technology. We focus on experiences that may or may not be enabled by digital technology, but which primarily make our experiences different from those of others.

With real-time interaction and alignment presupposed, we can now mention what characterizes a user that experiences an augmentation of reality.

- One or more senses of the user perceive stimuli (percepts) – whether or not through technology – from changes in the environment that are not perceived by the senses of others present in the same physical, perhaps digitally enhanced, environment.
- Changes in the environment are from the 'sense' viewpoint of the user; the changes are aligned with the already present content (from the user's sense viewpoint).
- The user can interact, requiring the environment to adapt to his 'view' on alignment, with the content that provides him with sensorial stimuli, and a sense-oriented view of the interaction, that is not available to other users present in the same physical environment.

Obviously, the conclusion and definition are open for discussion.

REFERENCES

Azuma, Ronald T. (1997). A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6, 4, 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Azuma, Ronald T. (2019). The road to ubiquitous consumer augmented reality systems. Hum Behav & Emerg Tech. 1:26–32.

Montero, Barbara. (2006). Proprioception as an Aesthetic Sense. Journal of Aesthetics and Art Criticism 64:2, 231–242. DOI: 10.1111/j.0021–8529.2006.00244.x.

Nijholt, A. (2021). Augmented Reality Humans: Towards Multisensorial Awareness. In Proceedings Digital Economy: Emerging Technologies and Business Innovation. 6th International Conference on Digital Economy, ICDEc 2021, Lecture Notes in Business Information Processing, Volume 431, Rim Jallouli, Mohamed Anis Bach Tobji, Hamid Mcheick, and Gunnar Piho (eds.), Springer, Cham, Switzerland, 237–250.

Nijholt, Anton. (2022). Weaving Augmented Reality into the Fabric of Everyday Life. Symposium on Electronic Art, ISEA2022: Possibles. - Barcelona, Spain, 471–478.

Sicart, Miguel. (2017). Reality has always been augmented: Play and the promises of Pokémon GO. Mobile Media & Communication, Sage Journals 5(1), 30–33.

Yanagida, Y., Kawato, S., Noma, H., Tomono, A. and Tesutani, N. (2004). Projection based olfactory display with nose tracking, IEEE Virtual Reality 2004, 43–50, DOI: 10.1109/VR.2004.1310054.

Yanagida, Y., Nakano, T. and Watanabe, K. (2019). Towards precise spatio-temporal control of scents and air for olfactory augmented reality, 2019 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Fukuoka, Japan, 1–4, DOI: 10.1109/ISOEN.2019.8823180.