

Interactions in Augmented and Mixed Reality: An Overview

Theofilos Papadopoulos ^{1,*}, Konstantinos Evangelidis ², Theodore H. Kaskalis ¹, Georgios Evangelidis ¹
and Stella Sylaiou ³

¹ Department of Applied Informatics, School of Information Sciences, University of Macedonia, 54636 Thessaloniki, Greece; kaskalis@uom.edu.gr (T.H.K.); gevan@uom.gr (G.E.)

² Department of Surveying & Geoinformatics Engineering, International Hellenic University, 57001 Thessaloniki, Greece; kevan@ihu.gr

³ School of Social Sciences, Hellenic Open University, 26335 Patra, Greece; sylaiou@gmail.com

* Correspondence: priestont@gmail.com

Abstract: “Interaction” represents a critical term in the augmented and mixed reality ecosystem. Today, in mixed reality environments and applications, interaction occupies the joint space between any combination of humans, physical environment, and computers. Although interaction methods and techniques have been extensively examined in recent decades in the field of human-computer interaction, they still should be reidentified in the context of immersive realities. The latest technological advancements in sensors, processing power and technologies, including the internet of things and the 5G GSM network, led to innovative and advanced input methods and enforced computer environmental perception. For example, ubiquitous sensors under a high-speed GSM network may enhance mobile users’ interactions with physical or virtual objects. As technological advancements emerge, researchers create umbrella terms to define their work, such as multimodal, tangible, and collaborative interactions. However, although they serve their purpose, various naming trends overlap in terminology, diverge in definitions, and lack modality and conceptual framework classifications. This paper presents a modality-based interaction-oriented diagram for researchers to position their work and defines taxonomy ground rules to expand and adjust this diagram when novel interaction approaches emerge.

Keywords: augmented reality; mixed reality; human-computer interactions; taxonomy; modalities; modality-based classification



Citation: Papadopoulos, T.; Evangelidis, K.; Kaskalis, T.H.; Evangelidis, G.; Sylaiou, S. Interactions in Augmented and Mixed Reality: An Overview. *Appl. Sci.* **2021**, *11*, 8752. <https://doi.org/10.3390/app11188752>

Academic Editor: João M. F. Rodrigues

Received: 2 August 2021

Accepted: 14 September 2021

Published: 20 September 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Significant efforts have been spent, in both basic and applied research, to highlight the importance of human–computer interaction (HCI) on the end-user experience in augmented and mixed reality (AR)(MR) environments [1,2]. To a large extent, research focuses on the user capability to perform tasks and interact with the virtual world, assisted by various functions and control systems. User-centered system design (UCSD), first described by Kling [3] and later by Norman [4], generally focuses on the user’s understanding of a system. It examines what the user expects to happen, and how to perform a task or recover from an error, presenting HCI as a communicative and collaborative process between humans and machines. Exploiting immersive realities and the UCSD radically changed the way humans perform everyday tasks or perceive historical and cultural information. Mixed and augmented reality finally occupy significant space in our daily routine. The achievement of several historical milestones, from routing [5] to entertainment [6], and from social media [7] to engineering and remote collaboration [8], showcases the promising future of AR and MR.

Technological achievements in AR and MR environments have made possible interactive visualizations of previously unexplored virtual and real-world combinations. According to Milgram et al. [9], the MR environment is where the real and the virtual

coexist. Coutrix et al. [10] described a mixed object as a real object with a virtual equivalent. Recently, Evangelidis et al. [11] defined the MR ecosystem, strictly separating it from AR, by introducing geospatial modalities and implementing the concept of mixed objects, thus achieving spatial and context awareness among realities. Over the past 25 years, approximately since the introduction of the well-known reality–virtuality continuum [9], published research work and applications have profoundly changed the way humans perceive and interact with historical [12–16], future [17–21], and imaginary [22–24] reality scenarios. However, the latest research findings and innovations in review papers regarding interaction methods are not classified under a well-defined framework, thus leading to misconceptions and ambiguities. For example, having the taste, smell, and haptic modality enclosed by the sensor-based modality, a system that utilizes all of them would still be, by definition, unimodal. Bunt [25] and Quek [26] both stated that the interaction between this world and humans is naturally multimodal [27]; therefore, overviewing AR and MR HCI in the light of how humans perceive reality, through sensations, might improve creative thinking and provoke novel interactions. Categorizations commonly taking place are based on the field of application (tourism, architecture, medicine) the device of application (mobile, desktop) [28], or umbrella terms (multimodal, tangible, collaborative) [29] and without focusing on the modality or the context of interaction. As a result, an inaccurate representation of available interaction methods can affect the creative thinking of future researchers and act counterproductively concerning their efforts in the field of HCI. Previous attempts of listing or categorizing the components of HCI for augmented and mixed reality [30–39] reveal that a clear, in-depth taxonomy of interactions either does not exist or is not widely known to the scientific community.

Mixed reality is an ever-evolving field, and novel approaches and innovative applications could delineate new interaction methods. HCI is associated with established theories, such as the theory of action described in Norman's book *The Design of Everyday Things* [40], the theory of communication [41], the theory of modalities [42], and the theory of perception [43]. Interestingly, although the theory of modalities positions haptic together with audio and visual modalities, the categorizations mentioned above include it within the sensor-based modality, together with taste and smell. The taxonomy proposed here-with is much more of an overview and a first attempt to expose all modalities with their interactions in the first level. We expect that the taxonomy we propose will better organize existing interaction methods, present a complete view of what has been accomplished so far and define a set of ground rules regarding naming conventions. Pamparau and Vatavu, in a recent position paper [44], stressed several issues to the community related to user experience (UX) and HCI in AR and MR environments, one of which was to structure design knowledge for the UX of interactions. A well-defined classification framework needs to exist for this to happen, and the interaction challenges need to be known to test UX [45]. In our understanding, interactions in immersive environments reveal three fundamental challenges. First, users need to naturally interact with machines to perform main interaction tasks, such as selection, manipulation, navigation, and system control. The interaction method should be as intuitive as possible to produce natural interactions, as any disturbance of the user's attention may detract from the immersive experience [46]. Secondly, current technological limitations in positional accuracy in such hybrid environments may cause spatial misalignments [47,48] or dislocations [49]. Accurately determining the end user's position [50] is crucial for successfully visualizing an MR environment, and technical challenges regarding coverage emerge. Finally, for interactions to be as "real" as possible, there should be a semantic context connection among involved realities. Based on the abovementioned, this paper aims to analyze research in HCI for immersive realities and mobile environments and to give an overview of what has been done so far by presenting a classified representation as part of a modality-based and interaction-oriented diagram of the reviewed work. We present a new approach for classifying HCI for immersive realities by interrelating modalities (audio-based, visual-based, haptic-based, and sensor-based) with their context and methods. The main scope of this study is to present and organize

the distinct interaction methods and organize them in a well-defined and structured classification that provides more depth and accuracy concerning how modalities are being used, in what context, and with what method. This innovative classification model is the outcome of a thorough study of pertinent research, as well as the result of a methodological investigation of the optimal way to structure the categories so that the approaches employed in the surveyed papers would be presented in a consistent, precise, and more meaningful way.

This paper is structured as follows: Section 2 presents a detailed explanation of the taxonomy ground rules and highlights the categories with their definitions based on which this paper organizes research findings. Then, Section 3 introduces a brief review of the state-of-the-art interactions and an in-depth review of the visual, audio, haptic, and sensor-based modalities. At this point, it is worth mentioning that, although we recognize the taste and smell modalities, they are not part of the current review. Therefore, a detailed review for these modalities remains a research gap as far as our categorization approach is concerned. Finally, the last section concludes the paper.

2. Conceptual Framework Definition

A basic rule adopted is that there should exist one modality for each of the five human senses. That being established, the visual-based, audio-based, haptic-based, taste-based, and smell-based modalities have been created. Any other interaction unrelated to the modalities mentioned above is included in the sensor-based modality (Section 3.4). Each modality contains groups of contexts (Section 2.3), and ideally, the context categories of a single modality should not overlap. Although this is a debatable issue, the decision taken is that whenever an overlap occurs, researchers should analyze a new category semantically to justify its creation [51]. The taxonomy should always be expandable and adjustable when a new context group is identified. The context groups should be simple and broad enough so that someone can raise specific questions regarding common problems or testing methodologies. Some of the issues that should be avoided when categorizing, are the lack of focus, clarity, inspiration or creativity, redundant ideas, and the inability of locating the ideal case, identifying challenges, and inducing lateral thinking.

Each context should define its own set of tests to identify the efficacy of the interaction methods it includes (Section 2.4). For example, new users of an interaction method that involves equipment may be satisfied with the overall experience when completing a group of tasks for a specific period. At the same time, those having utilized the equipment for years find it difficult to operate [52]. Therefore, long-term usage of equipment should be a prerequisite when testing such interactions. In addition, the keywords for naming methods should be checked for adaptability by the research community. For example, the keyword *gaze detection* used in the previous classification [30] is replaced with *eye gaze detection* as it is more informative. Finally, some methods might be able to be positioned in more than one context group. For example, *location-aware sound effects* can be placed in both *location-based* and *sound-based* contexts. In this case, the researchers will have to define which group better characterizes their work, and if this is not possible, they should position their work in all relative context groups. A representation of the taxonomy levels is presented in the Figure 1 diagram. Future researchers may expand this diagram to include the “techniques” level, determining all techniques used to utilize a specific interaction method.



Figure 1. Displaying the basic categorization levels of our modality-based diagram for interactions in augmented and mixed reality.

2.1. Interaction Tasks

Mixed reality environments contain four basic interaction tasks. As also stated in Bachmann et al.'s work in 2018 [53], these are selection, manipulation, system control, and navigation.

- Selection: Refers to the task of selecting an object to perform actions, such as retrieving information or storing it as an argument for another action [54].
- Manipulation: Provides to the user the capability of changing any of the object's attributes, e.g., scale, position etc. [55].
- Navigation: Provides to the user the capability of navigating in an immersive environment by changing position or orientation [56].
- System control: Refers to the user capability of performing changes in the system state, such as menu-based changes [57].

2.2. Modality

A user interface [57] is based on information inputs and outputs (IO) via bidirectional human–computer communication channels. As input or output, we consider any human actions that convey meaning for interaction to a computer or any intentional augmentation or alteration of the human perceptual modalities. Every independent channel is called a modality, and every system that uses only one of these channels for IO is called a unimodal system. Systems that incorporate more than one of the modalities above are called multimodal. We define six modalities that allow information IO: visual-based, audio-based, haptic-based, taste-based, smell-based, and sensor-based modalities. As the taste and smell-based modalities are not reviewed in this paper, we do not provide definitions for them. Therefore, we define the visual, audio, haptic, and sensor-based modalities as follows:

- Visual-based: The visual-based modality includes any state changes that a camera sensor can capture, convey meaning and can be used to describe the user's intention to interact or present visual feedback.
- Audio-based: The audio-based modality contains all actions and feedback that include sound perception and sound stimuli.
- Haptic-based: The haptic-based modality defines all interactions that can be perceived through the sense of touch or executed through graspable-tangible objects.
- Sensor-based: Finally, the sensor-based modality includes all interactions requiring any sensor to capture information regarding an action or transmit feedback back to the user, besides visual, auditory, haptic, taste, and smell inputs/outputs. An example of this modality includes the pressure detection method.

2.3. Context

Context defines the conceptual framework through keywords used by researchers to describe their work in publications. A context is a subcategory of a modality that abstractly expresses specific interactions without fully explaining them. The usually adopted pattern comprises a noun followed by the word "based" to describe the base context, such as gesture-based, speech-based, or touch-based. At this point, it is worth noticing that the word "based" follows both a modality and a context, resulting in a repetitive naming convention for an interaction method (covered in the following subsection). We have decided to keep it this way in the taxonomical table, as the research community widely adopts these terms, and in most cases, these are used separately. With that being said, "a method of interaction that belongs in the marker-based context of the visual-based modality" is a sentence that conveys meaning and immediately positions the method in our proposed table.

2.4. Method

An interaction method is a keyword combination that describes or includes a series of coordinated procedures used to accomplish an interaction task. A pattern that researchers

and inventors usually adopt to describe their methods comprises two parts, the first one to be the base medium of interaction and the second one to be a verb or a noun that defines the action to be performed. For example, *eye gaze detection* and *body posture analysis* are simple and easy to understand. However, “optical mouse sensor attached to finger” could be renamed as a *finger motion tracking* method in the motion-based context of the sensor-based modality. The techniques used to exploit a method should be avoided as part of the naming, as the resulting method names (e.g., YOLO hand gesture recognition and R-CNN hand gesture recognition, etc.), violate several ground rules previously defined, such as redundant ideas or inability of inducing lateral thinking. However, some methods presented in the final model-based diagram (e.g., fiducial marker recognition and infrared marker recognition) overlap in concept. Nevertheless, they are included as they induce lateral thinking.

3. Research Results

Before analyzing each modality, a brief state-of-the-art review is presented to identify current trends. In a recent study, Rokhsaritalemi et al. stressed that “mixed reality is an emerging technology that deals with maximum user interaction in the real world compared to other similar technologies” [58]. The impact of augmented reality on a user’s satisfaction in numerous applications in the fields of engineering [59], archaeology [60], medicine [61], or education [62] cannot be questioned. Nevertheless, separating the physical world from the virtual has a significant impact on the user’s immersion level. This enforced the development of mixed reality environments and upgraded augmented reality interactions to become more natural and include more aspects of the physical world. Chen et al. in 2017 [63] proposed a framework to boost the user’s immersion experience in augmented reality through material-aware interactions by training a neural network for material recognition. In 2018, Chen et al. [64] mentioned that semantic understandings of the scene are necessary for mixed reality interactions to become realistic. The importance of structural information of physical objects is inextricably connected with proper augmentation and placement, but natural interactions between virtual and real objects require semantic knowledge. We have noticed a lack of publications related to material-aware MR interactions, and it seems that more research needs to be done in this field. Context-aware and material-aware interactions can lead to realistic physically-based sound propagation and rendering. In 2018 [65], Serafin et al., in their work on sonic interactions in virtual reality, concluded that the auditory outcomes of sound synthesis are not yet indistinguishable from real ones. Sonic interactions involve the techniques of sound propagation [66] and binaural rendering for binaural hearing [67] to provide immersive action and environmental sounds. An example includes the sound of the footsteps of a virtual man walking in a grass field. Through context-aware and material-aware interactions, the sound propagation algorithm would “consider” the material of the grass and the open area to generate the sound. The same material would sound differently inside a cave.

3.1. Visual-Based Modality

The visual-based modality includes any state changes that a camera sensor can capture, convey meaning, and can be used to describe the user’s intention to interact or present visual feedback. Figure 2 visualizes all the contexts and methods identified for the visual-based modality. A detailed review for each context is presented in the following subsections.

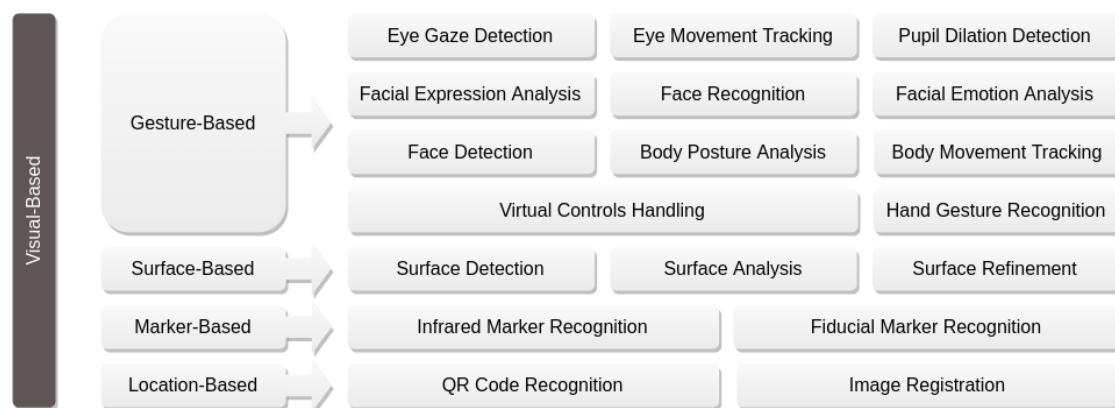


Figure 2. Displaying all categories of the visual-based modality.

There are two ways of interacting with or visually perceiving a mixed reality environment. As previously stated, an MR environment is where the real and the virtual coexist. Thus, two of the main ways of visual coexistence are [68]:

- Optical see-through systems (OST): by displaying digital objects in a semi-transparent screen where real objects can be directly perceived through the glass.
- Video see-through systems (VST): by displaying digital objects on a screen together with real objects captured by a camera sensor, commonly used by smartphones in AR.

3.1.1. Gesture-Based

When interacting through the gesture-based context, computers get visual input and recognize body language without any other sensory information. *Eye gaze detection* is included in this context of interaction. It is characterized by two major issues: (a) avoidance of unintended actions and (b) limitations related to eye movement accuracy. The gaze and dwell interaction model [69] is used for this method, as described by Microsoft [70], where basically the user needs to look at an object and retain this action by staring to select it. It has a high accessibility rate [71] as even severely constrained users can perform this interaction. In 2017, Piumsomboon et al. explored the advantages of this interaction by exploiting some of the basic functionalities of the human eye [72]. This method takes advantage of the eye inertial and the natural vestibulo-ocular reflex (the ability to lock a target regardless of head movements) and is used with head-mounted displays. The authors conclude that more research should be done analyzing the collected large-scale *eye movement tracking* data and improving user experience. Jacob [73] examined techniques and challenges related to eye movement interactions. He proposed an approach based on separating the actual data (eye movements) from noise and then estimating the user's intention of interaction. An interesting interaction method related to the eyes is *pupil dilation detection*. In [74], the authors used this method as a reliable indicator of cognitive workload.

In 2017 Samara et al. [75] performed task classification in HCI via the visual channel. The authors combined *facial expression analysis* and eye-gaze metrics for computer-based task classification. The outcome was that these two interaction methods combined resulted in higher classification accuracy. Facial-expression interactions exploit the *face detection* [76] method and can be used in controlling a user's virtual avatar facial expressions [77] in *facial emotion analysis* [78] and emotion recognition [79]. *Face recognition* [80] also exploits the *face detection* method but is usually used to identify a person by facial characteristics. In 2017 Mehta et al. [81] conducted a review on human emotion recognition which could assist in teaching social intelligence in machines. Techniques commonly used to exploit *emotion recognition* and *emotion analyses* methods include, without being limited to, the geometric feature-based processes [82] and machine learning [83].

In the context of emotion analyses and human psychology, another method interprets body language. The body movements in immersive realities and how they enhance the sense of presence is examined in-depth by Slater's and Usoh's work [84]. In 2020 Lee [85]

applied the Kinect Skeletal Tracking (KST) System in an augmented reality application to improve the social interaction of children with autism spectrum disorder (ASD). They used a Kinect camera to scan the therapist's body gestures and visualized them on a 3D virtual character. In another work [86], the authors used a webcam exploiting the user's *body movement tracking* method to interact with the AR system. The user had to do simple tasks like jumping, stretching, or boxing to "hit" the correct answers presented in the virtual world. Such methods utilizing physical exercise make the learning process more appealing to students of younger ages. Umeda et al. [87] exploited the *body posture analysis* method to locate a person's two hands perceived by a Firewire camera and superimpose artificial fire on them in real time. Algorithms of hand tracking and gesture recognition are used to detect the gestures through a camera sensor and perform functions accordingly. In 2016 Yousefi et al. [88] introduced a solution for real-time 3D *hand gesture recognition*. They used the embedded 2D camera of a mobile device, supporting 2D and 3D tracking, joint analysis and 10+ degrees of freedom. They accomplished these features by pre-processing the image to segment the hand from the background and matching the normalized binary vector outcome with gestures already stored in a database. Some of the issues to be further examined as regards the hand-gestures interaction method are (a) the efficiency of gesture recognition algorithms, (b) the efficiency on low contrast environments, (c) high consumption of computing resources, (d) lack of haptic feedback, and (e) hand occlusion with the augmented scene. However, the applications of hand gestures in real-life scenarios are unlimited due to the naturalness of the interaction. MixFab [89] is an example of how this interaction method can be helpful to non-expert users and allow them to perform tasks in a 3D environment with ease. The authors, Weichel et al., present this application prototype facilitating the manipulation of 3D objects by hand gestures and have them 3D printed without the need for any 3D modelling skills. Yang and Liao [90] utilized hand-gesture interactions to create interactive experiences for enhancing online language and cultural learning.

The *virtual controls handling* method of interaction refers to any interactivity between the user and a virtual control panel. In [91], Porter et al. demonstrated a prototype system that projects virtual controls onto any surface, using finger-tracking techniques to understand the user's intention of interaction with the virtual control panel. One of the benefits of the virtual controls interaction method is that any control component (virtual button, display, etc.) is rearrangeable. Therefore, its ergonomic design can vary to satisfy different needs and users. Besides hand gestures, they used 3D tabletop registration to place the virtual objects at any location on their tabletop.

3.1.2. Surface-Based

Surface detection is an interaction method of the surface-based interaction context in which the geometry, position, or rotation of a surface is considered in the interaction. Telepresence, the experience of distant worlds from vantage points [92], is a field where surface detection interactions can be applied to create realistic representations of the remote world to surpass the virtual–real occlusion problem. This method is commonly based on algorithms that provide solutions in the simultaneous localization and mapping (SLAM) [93] computation problem where both the point of view of an unknown environment and the user's position in it are updated and need to be tracked [94]. It is also used in marketing applications that exploit the advantages of immersive environments, such as the IKEA app [95]. In [96], the authors state that one of the well-designed features such applications should contain, among others, is "match between system and the real world", which refers to scaling and positioning virtual objects in the real world properly. In engineering, a frequently used method used for interactions is the *surface analysis* method. This method is somewhat more sophisticated than *surface detection*. It is used to analyze geometries and predict user intention of assembling different compartments [97] or interact with flexible clay landscapes to create new terrain models with *surface refinement* [98]. *Surface refinement*, another interaction method of the surface-based context, describes the techniques that

alter the geometry or the material of a natural surface. Punpongsanon et al. demonstrated SoftAR [99], an application capable of visually manipulating the haptic softness perception. The authors proposed that by virtually exaggerating the deformation effect of a material, it is possible to alter the haptic perception of the object's softness. Such interactions also have applications in realistic representations of mixed reality environments, such as a virtual object projecting shadow onto a natural surface or a virtual light source to illuminate an actual surface [100]. In [101], the authors present the interaction model and the techniques that lead to a successful application of instant indirect illumination for dynamic mixed reality scenes. For the placement of the 3D models in the MR environment, they used marker-based interactions.

3.1.3. Marker-Based

Marker-based interactions contain all interactions in immersive environments supported by marker tags, such as ARTag markers [102]. Onime et al. [103] performed a reclassification of markers for mixed reality environments based, among other things, on the level of realism, the level of immersion and visibility. However, not all markers are suitable for marker-based interaction. In [104], Mark Fiala mentions that Data Matrix, Maxicode, and Quick Response (QR) codes are ideal for conveying information when held in front of a camera and not localized. Thus, they are not suitable to be used as a fiducial marker, as is needed for applications of immersive realities. In contrast, InterSense [105], Reac-TIVision [106], Cyber-code [107], Visual Code [108], Binary Square Marker [109], Siemens Corporate Research (SCR) [110], and BinARyID [111] are several examples of fiducial marker systems that can be used in AR and MR applications. Mateos [112] proposed AprilTags3D that improves the accuracy of *fiducial marker recognition* of AprilTags in field robotics with only an RGB sensor by adding a third dimension to the marker detector. In [113], Wang et al. used the *infrared marker recognition* method and proposed an AR system for industrial equipment maintenance. As the markers are infrared and invisible to the naked eye, they do not cause any visual disturbance to the user [114]. A camera capable of capturing infrared light can detect information from infrared markers regarding position and rotation and successfully superimpose virtual objects in the real environment.

3.1.4. Location-Based

In a recent paper, Evangelidis et al. [11] demonstrated an application prototype that constitutes a continuation of the research proposal development called Mergin' Mode [115]. Their demonstration used the *QR code recognition* method to locate a user in the real world and serve virtual objects presented in an MR environment. In addition, QR codes contain information regarding the user's orientation and which content should be delivered from several predefined virtual worlds created to promote cultural heritage. Each observation point (the stations where the QR codes are located) exposed specific interactions in selecting a virtual agent to obtain information. In [116], Stricker et al. used the image registration method to determine the user's position and serve virtual content. The technique they describe needs a sufficient number of calibrated reference images stored in a database and requires a continuous internet connection. However, this method can provide the user's position accurately, thus improving the interaction experience.

3.2. Audio-Based Modality

The audio-based modality contains all actions and feedback, which are included in sound perception and sound stimuli. Auditory stimuli are essential in human understanding of the environment as contextual information provides situational awareness. Through auditory inputs, one can obtain information beyond visual boundaries (places out of sight, behind walls, etc.). Thus audio-based interactions can improve the immersive experience, making it "feel natural" and closer to the human way of experiencing reality. Figure 3 presents all the contexts and methods identified for the audio-based modality. Afterwards, a detailed analysis follows for each context.

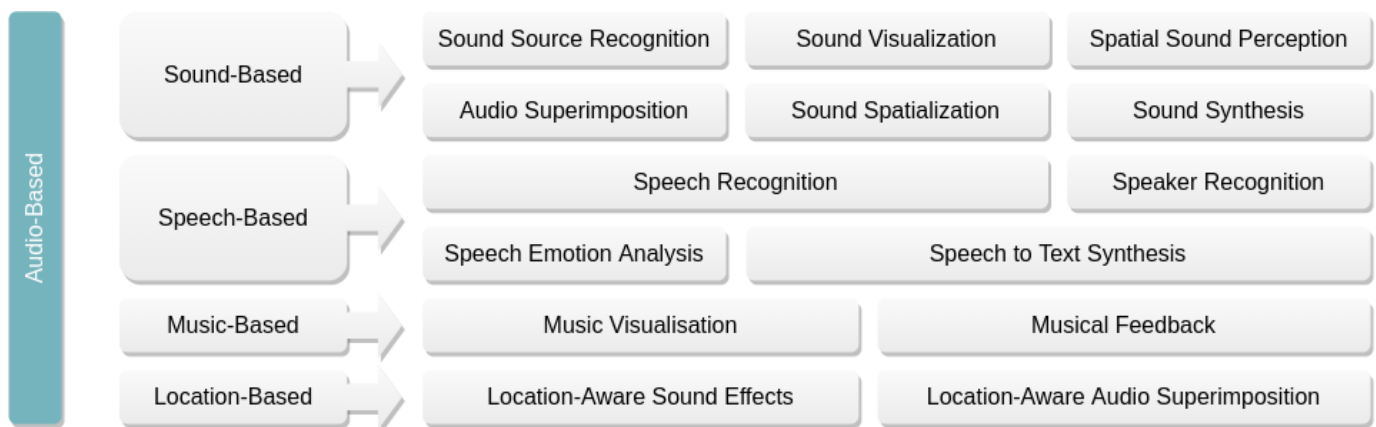


Figure 3. Displaying all categories of the audio-based modality.

3.2.1. Sound-Based

Sound source recognition and *sound visualization* are two methods Shen et al. used in [117] to augment human vision in identifying sound sources for people with hearing disabilities. The system can successfully perform identification based on recognition algorithms that exploit the microphone's capabilities and AR tags placed on variant objects to position a virtual object that indicates the sound source. Rajguru et al. [118] reviewed research papers to determine challenges and opportunities in spatial soundscapes. Spatial soundscapes exploit the *spatial sound perception* method to enhance situational awareness. *Audio superimposition* is another method that is commonly used in audio augmented reality. In [119], the authors Harma et al. combined virtual sounds with real sounds captured by a microphone and reproduced them with stereo earphones. However, the users that utilized this method expressed difficulties in separating the virtual from the real sounds.

Sound synthesis, often used in real time for immersive realities, includes the techniques and algorithms that estimate the ground reaction force based on physical models exploited by a sound synthesis engine. Nordahl et al. [120] proposed a system that affords real-time sound synthesis of footsteps on different materials. To feed the sound synthesis engine, they used inputs regarding the surface material. For solid surfaces (metal and wood), they exploited the impact and friction model to simulate the act of walking and the sound of creaking wood accordingly. For aggregate surfaces (gravel, sand, snow, etc.), they further enhanced the previous models by using some reverberation by convolving in real time the footstep sounds with the impulse response recorded in different indoor environments. In addition, *sound synthesis* combined with *sound spatialization* can create realistic environmental sounds. Verron et al. [121] presented a 3D immersive synthesizer dedicated to environmental sounds. Based on Gaver's work [122], three basic categories concerning environmental sounds stood out: liquid sounds, vibrating objects, and aerodynamic sounds. The proposed system reduces the computational cost per sound source compared with implementations and constitutes a reliable tool for creating multiple sound sources in a 3D space.

3.2.2. Speech-Based

Besides engineering, sound-based interactions can produce benefits in education by exploiting *speech recognition*. Hashim et al. [123] developed an AR application to enhance vocabulary learning in early education. The authors integrated visual scripts for orthography and audio for phonology. As a result, they concluded that such applications provide high levels of user satisfaction and can significantly affect pronunciation as students repeat words and phrases until they get correct feedback. In Billingham et al.'s [124] work, the authors modified the VOMAR application [125]. They added the Ariadne [126] spoken dialogue system using Microsoft's Speech API [127] as a speech recognition engine to allow people to rapidly put together interior designs by arranging virtual furniture in

empty rooms. *Speech recognition* is commonly used for computers to understand auditory commands and execute tasks. Among others, the authors used in their application the “select”, “move”, and “place” commands. Having an answer to the question “what are they saying”, the next step is to answer the question “who is speaking”. *Speaker recognition* is the method of interaction where machines can produce several pieces of information about the person who is speaking or even identify the person when the information provided can support that. In Hanifa et al.’s [128] exceptional review, several research papers regarding the *speaker recognition* method were included. They further classified this method in identification, verification, detection, segmentation, clustering, and diarization and proposed issues regarding variability, insufficient data, background noise, and adversarial attacks. Chollet et al. [129] reviewed technological developments in augmented reality worlds, emphasizing speech and gesture interfaces. They stressed that speaker verification could be used for authentication purposes before starting a dialogue, for example, regarding a bank transfer in a virtual world.

Lin et al. [130] designed and built an emotion recognition system in smart glasses to help improve human-to-human communication utilizing *speech emotion analysis*. The target group for such applications that could benefit the most includes doctors and autistic adults. The authors collected speech sentiments and speech intonation data using Microsoft’s Azure API and their intonation model to analyze the collected data. The system communicates the detected emotions to the users via audio, visual, or haptic feedback in real time. *Text to speech synthesis* is an interaction method that synthesizes auditory outputs based on texts. Mirzaei et al. [131] combined *speech recognition*, *speech to text synthesis*, and *text to speech synthesis* methods to support deaf people. The system they presented captures the narrator’s speech, converting it into texts, and displays it to the user’s AR display. Additionally, the system converts typed text into speech to talk back to the narrator if the user wants to respond.

3.2.3. Music-Based

Musical interactions refer to all methods related to sounds arranged in time that composite the melody, harmony, rhythm, or timbre elements. In Altosaar et al.’s [132] paper, the authors present a technique of interacting with virtual objects to produce musical pre-recorded sounds. As the musician manoeuvres VR controllers and collides with the virtual objects, *musical feedback* is played. In addition, the musician can interact with full-body movements to produce music, a method that can enhance musical expression. Bauer and Bouchara [133], exploiting the *music visualization* method, presented a work in progress where a user can visualize the parts of a music clip (intro, middle and outro) and individual audio components, such as kick, synthesizer or violins. Then, the user can manipulate these components, altering the final music outcome.

3.2.4. Location-Based

Bederson [134] proposed a technique that exploits the *location-aware audio superimposition* method to enhance the social aspects of museum visits. An infrared transmitter detects the visitor’s location, which signals a computer to play or stop playing a pre-recorded description audio message. Paterson et al. [135] stressed that immersion, besides interaction, involves the “creation of presence”, which is the feeling of being in a particular place. They have created a location-aware augmented reality game where the user navigates a field looking for ghosts. Based on the user location, sound effects were triggered indicating the paranormal activity level, thus exploiting the *location-aware sound effects* method. Lyons et al. [136] presented an RF-based location system to play digital sounds corresponding to the user’s location. They have created a fantasy AR game where the user enters a game space (convention hall, gallery, etc.) with RF beacons set up. Through superimposed environmental sounds, sound effects, and narrator’s guidance, the story of the game evolves. In another research paper [137], the authors described the technical aspects of a smartphone application that helps blind people experience their surrounding

environment. They utilized *location-aware sound effects* and non-spatialized sound details, such as the full name of their location. It is worth mentioning that navigation accuracy posed an important challenge in both previous papers.

3.3. Haptic-Based Modality

The haptic-based modality defines all interactions that can be perceived through the sense of touch or executed through graspable-tangible objects. Figure 4 visualizes the contexts and methods that we identified for the haptic-based modality. In the following subsections, we explain each individual context.



Figure 4. Displaying all categories of the haptic-based modality.

3.3.1. Touch-Based

Yang et al. [138] presented an authoring tool system for AR environments using *single-touch* and *multi-touch* interaction methods of the touch-based context of interaction. *Single-touch* refers to any touch interaction that requires only one finger, while *multi-touch* benefits from more than one finger. These methods make the ‘select’ and ‘manipulate’ tasks available for virtual objects through the touch and drag interaction model. Jung et al. [139] proposed another on-site AR authoring tool for manipulating virtual objects, utilizing the *multi-touch* method for mobile devices. They presented two interaction tasks, selection and manipulation (select, translate, and rotate), which can be executed simultaneously, and stated that this method is convenient for non-expert users. Kasahara et al. [140] extended the touch interaction by manipulating (moving and rotating) the device. Later, Yannier et al. [141] examined the effect of mixed reality on learning experiences. In [142], the authors created a multi-touch input solution for head-mounted mixed reality systems, making any surface capable of interaction similar to touchscreens. The problems they provide solutions to are relative to the constant user’s motion that constrain the use of a simple static background model. This approach also provides a solution to the haptic feedback problem of many of the other HMD available interactions. Zhang et al. [143] examined the haptic touch-based interactions that can take place on the skin. They proposed ActiTouch, a novel touch segmentation technique that uses the human body as a radio frequency waveguide that enables touchscreen-like interactions.

3.3.2. Marker-Based

Jiawei et al. [144] suggested an interactive pen tool that transforms the actions in real space into the virtual three-dimensional CAD system space. To achieve that, they used *fiducial marker recognition* for markers placed on top of a simple pen to capture its position and orientation. Then, the system could locate any movements and draw virtual lines and shapes into the CAD system space. In [145], Yue et al. presented WireDraw, a 3D wire object drawing using a 3D pen extruder. They also used *fiducial marker recognition* for identification purposes in the mixed environment and superimposed virtual objects as indicators to help the user design high-quality 3D objects. Yet another pen-based interaction is demonstrated by Yun and Woo [146], this time using a space-occupancy-check algorithm. First, they used a depth map to capture the 3D information of the natural world, and then the algorithm checked if any point in a depth map collides with the geometry of the virtual object. Haptic-based modality also utilizes the *RFID marker recognition* method.

For example, Back et al. [147] described an interaction that can augment the experience of reading a book with sound effects. To achieve that, they used RFID tags embedded into each page and an electric field sensor located in the bookbinding to sense the proximity of the reader's hands and control audio parameters.

3.3.3. Controls-Based

Leitner et al. [148] presented IncreTable, a mixed reality tabletop game that exploits the *digital controls handling* interaction method. The users can perform interactions by utilizing digital pens with an embedded camera that tracks Anoto patterns printed on a special backlit foil. Hashimoto et al. [149] employed a joystick-based interaction to help the user drive a remote robot. The virtual 3D axes (x,y,z) were color-displayed on top of the real robot, which the user observed from a camera. Through the color indications of the robot's local axes system, the user could decide how to move the joystick to complete a task. Chuah et al. [150] used MR applications to train students in medical examinations of vulnerable patients (e.g., children with speaking disabilities). They have created two interfaces for interaction; a natural interface that utilizes a tablet for drawing shapes and a mouse-based interface for the same purpose. A final survey showed that the participants preferred the natural interface from the mouse-based interaction. Finally, Jiang et al. [151] proposed HiKeyb, a typing system with a style of mixed reality. They used a depth camera, a head-mounted display (HDM) and a QWERTY keyboard to enhance the immersive experience of typing in MR.

A practical method that allows haptic interactions is *capacitive sensing*. This method uses the measurable distortion that an object with electrical characteristics (such as capacitance) creates within an electric field oscillating at low frequency [152]. Poupyrev et al. [153] proposed Touché; a novel swept frequency capacitive sensing technique that can detect a touch event and recognize complex human hands and body configurations with extremely high accuracy. Kienzle et al. [154] proposed ElectroRing, a 3D-printed wearable ring similar to an active capacitive sensing system that can detect touch and pinch actions. Finally, HydroRing [155] was presented by Han et al., which is yet another wearable device that can provide tactile sensations of pressure, vibration, and temperature on the fingertip, enabling mixed-reality haptic interactions using liquid travelling through a thin, flexible latex tube.

3.3.4. Feedback-Based

Besides input functionality, the haptic-based modality also includes a feedback-based context of interaction. An exceptional review was published by Shepherd et al. [156]. The authors present a study on elastomeric haptic devices and stress that in haptics we have seen much slower technological advances than visual or auditory technologies because the skin is densely packed with mechanoreceptors distributed over a large area with complex topography. Their review includes haptic wearables, haptic input devices (dataglove, VR tracker, VR controller), haptic feedback output methods (direct force feedback, force substitution-skin deformation, shape morphing-surface, virtual shape rendering with lateral force control), and haptic perception. Talasaz et al. [157] explored the effect of force feedback or *direct force feedback* on the performance of knot tightening in robotics-assisted surgery. They stressed that presenting haptic force feedback to the user performing such tasks increased the effectiveness, although this is a debatable subject [158]. Another example of the use of the *direct force feedback* method is PneumoVolley [159]. The authors, Schon et al., presented a wearable cap prototype that is capable of providing pressure feedback to simulate the interaction between a virtual ball and the user's head through pneumatic actuation. Schorr and Okamura [160] presented a device wearable on the fingertip, capable of transmitting haptic stimuli exploiting the *skin deformation* method as a force substitution. In a greater depth, a review of wearable haptics and their application in AR is presented by Meli et al. [161]. Yang et al. [162] used the *shape morphing* method. They mentioned several shape-morphing surfaces, such as (a) the shape display inFORM,

(b) a bio-inspired pneumatic shape-morphing device based on mesostructured polymeric elastomer plates capable of fast and complex transformations, and (c) rheological test results of MR fluid affected by a magnetic field. They stated that morphing devices are not yet mature, but they have explosive potential.

3.4. Sensor-Based Modality

The sensor-based modality includes all interactions requiring any type of sensor to capture information regarding an action or transmit feedback to the user, besides visual, auditory, haptic, taste, and smell inputs/outputs. Figure 5 presents all the contexts and methods recognized for the sensor-based modality. A detailed review for each context is presented in the following paragraphs.



Figure 5. Displaying all categories of the sensor-based modality.

3.4.1. Pressure-Based

Kim and Cooperstock [163] used the *pressure detection* method and proposed a wearable mobile foot-based surface simulator whose haptic feedback varies as a function of the applied foot pressure. Their purpose was to simulate the surface of a frozen pond and include “crack” sound effects based on the applied foot pressure. Qian et al. [164], exploiting the *center of pressure trajectory* method, combined floor pressure data for both feet to improve recognition of visually ambiguous gestures. The outcome was a system reliable in recognizing gestures from similar body shapes but with different floor pressure distributions.

3.4.2. Motion-Based

Minh et al. [165] designed a low-cost smart glove equipped with a gyroscope/accelerometer and vibration motors. Using *finger motion tracking* and *hand motion tracking*, hand and finger motions can be detected through the glove, including angular movements of the arm and joints. Using the same concept, Zhu et al. [166] proposed another smart glove capable of multi-dimensional motion recognition of gestures to achieve haptic feedback. The glove is equipped with piezoelectric mechanical stimulators that provide feedback when interacting with a virtual object. In another research paper [167], the authors used Microsoft’s HoloLens HMD, exploiting the *head movement tracking* and *head gaze detection* methods. They introduced a novel mixed reality system for nondestructive evaluation (NDE) training, for which, after a user study, they concluded that such systems are preferred for NDE training. In [168], Gul et al. presented a Kalman filter for head-motion prediction for cloud-based volumetric video streaming. Practically, server-side rendering systems, although they can provide high-resolution 3D content in any device with an acceptable internet connection speed, suffer from interaction latency. Although the research results were promising, the authors stressed that more research is necessary to examine several shortcomings, including predicting spherical quantities and more accurate predictions of head orientations.

3.4.3. Location-Based

Another context in the sensor-based modality that provides interaction for AR and MR environments is the location-based context. Radio frequency identification (RFID) uses passive, active, or semi-passive electronic tags that store a small amount of data, usually an ID or a link, and need readers to obtain the data from the tags. The communication of readers with the tags is made through RF [169]. Tags are deployed across the environment, and readers are carried on or attached to the positioning subjects. Benyon et al. [170], in their work “Presence and digital tourism”, mention that using GPS/GNSS coordinates and near-field communication (NFC) can create location-based triggered events for user interaction. Schier [171] designed a novel model for evaluating educational AR games. Using the *location-aware triggered events* method, the participants can interact with virtual historical figures and items, which GPS triggers to appear on their personal digital assistant (PDA). Other ways of acquiring the user’s location to achieve interaction regarding navigation or present information include, without being limited, dead reckoning and WiFi [172,173], visible light communication and cellular networks (GSM, LTE, 5G) [174,175].

3.4.4. IoT-Based

Since the term IoT was first used in the late 1990s by Ashton [176], exceptional research work has been published. Through IoT-based interactions, it is easier to provide humans with situational and context awareness, enhance decision-making in everyday tasks, and control any type of system and offer novel interactions to disabled individuals. Atsali et al. [177] used open-source web technologies, such as X3DoM, to integrate 3D graphics on the web. Their paper described a methodology that connects IoT data with the virtual world and the benefits of using web-based human–machine interfaces. They used the Autodesk 3ds Max Design software to develop a 3D scene, including a four-apartment building. An autonomous, self-sufficient IoT mixed reality system was installed to exploit the *infrastructure control* and *data monitoring* methods for the water management infrastructure. In another work, Natephra and Motamedi [178] installed markers and an IoT system in an apartment for monitoring indoor comfort-related parameters, such as air temperature, humidity, light level, and thermal comfort. As the mobile device scans a marker, it can acquire real-time data transmitted by the sensors and visualize them as augmented reality content. Phupattanasilp et al. [179] introduced AR-IoT, a system that superimposes IoT data directly onto real-world objects and enhances object interaction. Their purpose was to achieve novel approaches to monitoring crop-related data, such as coordinates, plant color, soil moisture content, and nutrient information. Other applications of IoT-based data monitoring methods include fuel cell monitoring [180], campus maintenance [181], and environmental data monitoring utilizing serious gaming [182]. Using the IoT-based context, we can also interact with virtual agents to “humanize” interactions with objects or living organisms. Morris et al. [183] developed an avatar to express a plant’s “emotion states” for plant-related data monitoring. The IoT system calculates arousal and valence based on soil, light, and moisture levels. By turn, the virtual avatar is designed to express these states on behalf of the plant. For example, the avatar will grow large and angry if the plant is left unwatered or be happy when the plant’s soil moisture is at the proper levels.

4. Discussion

This paper describes a modality-based interaction-oriented taxonomy (Figure 6) aiming to organize existing interaction methods and present a complete view of the heretofore accomplishments after a thorough review of more than 200 relevant papers. A significant challenge this venture undertakes to address is the lack of a well-defined and structured schema representing human–computer interaction in the context of mixed and augmented reality environments. For example, representations commonly included in research studies classify the visual-based modality by method (e.g., gesture recognition) and the sensor-based modality by the device (e.g., pen-based or mouse-based) [33]. Other representations, generally accepted by the research community, organize interactions by research areas,

thus avoiding defining an in-depth classification [30]. These research areas are described using methods (e.g., facial expression analysis), umbrella terms (e.g., musical interaction), devices (e.g., mouse and keyboard), and keywords that arbitrarily define a research area such as “taste/smell sensors”. Our classification approach is based on several established theories, such as the theory of modalities and the theory of perception, following basic taxonomy rules and aspiring to eliminate inconsistencies in previous classifications.

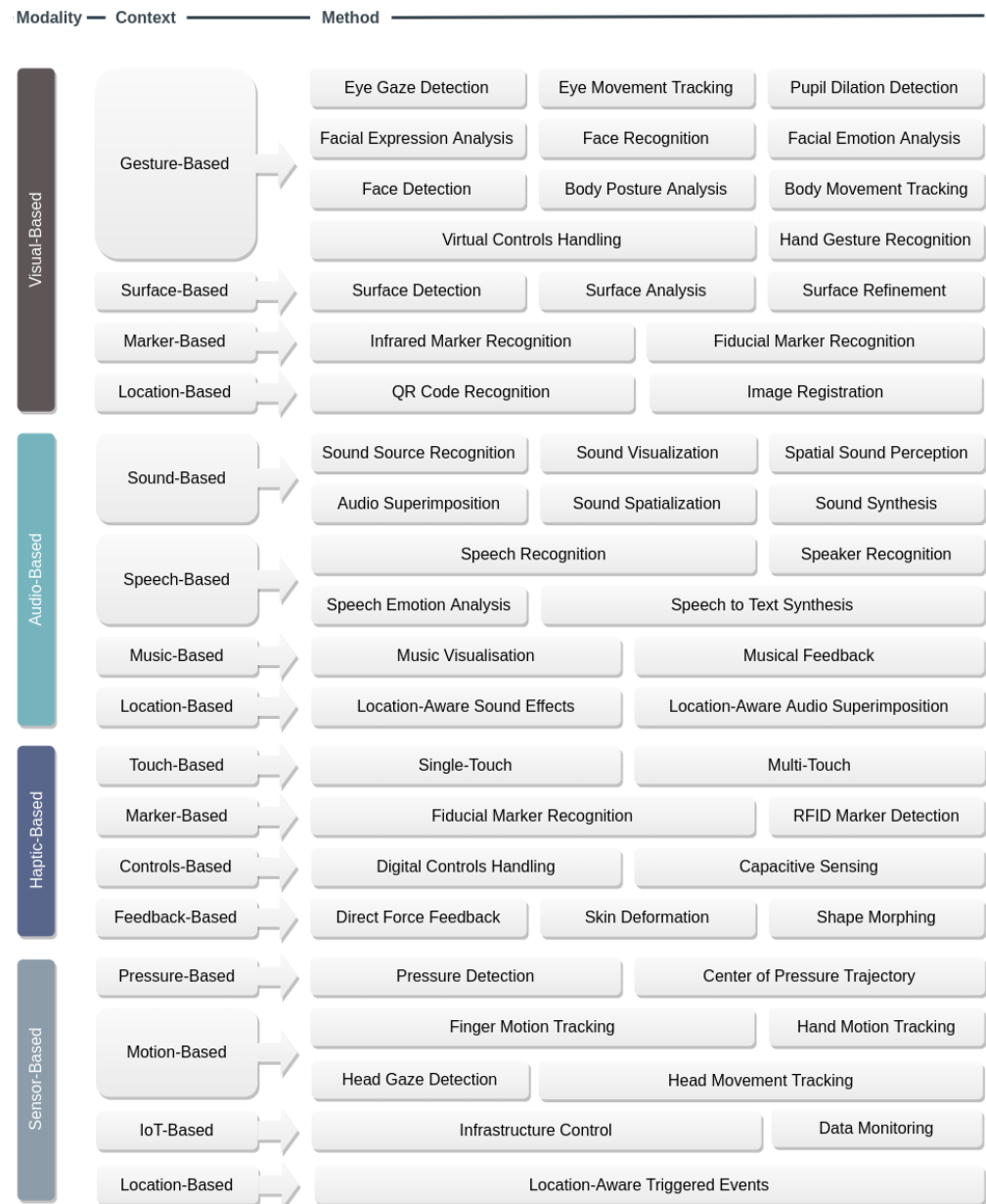


Figure 6. A complete representation of our proposed classification using a modality-based interaction-oriented diagram that visualizes all the identified modalities, contexts, and methods.

Nevertheless, it should be stressed that the scope of the presented work is not to elaborate or evaluate the interaction methods gathered in terms of comparable parameters such as efficiency, popularity, applicability, impact on human beings, nor to present challenges and limitations. Besides, such a task would require a significant effort only to identify existing implementations and applications. Rather, the proposed work attempts to identify a sufficient amount of interaction methods and organize them exploiting well-defined taxonomy rules. These rules will, in turn, place them in the field of the context

applied, such as gesture-based, marker-based, location-based, and others, and the associated modality, i.e., visual/audio/haptic/sensor-based. In other words, the scope of this review paper is only to identify distinct interaction methods and present them in a well-defined structured classification.

When defining the taxonomy rules, one of the principal decisions was that although an interaction method may be present in multiple contexts, we consider only those resulting in interactivity in an immersed environment (interaction-oriented classification). For example, in mixed reality projects, QR codes are employed to identify a user's location and therefore perform an interaction based on that location. Thus, although a QR code is usually placed in the marker-based context, in terms of interactivity, it is strictly located in the location-based context since location is what triggers the interaction and not the actual QR code image representation. Examples where a fiducial marker system actively participates in the interaction include ARTag and AprilTag systems. Such systems enable marker-based augmented reality, where 3D virtual objects are positioned over the identified markers. The rotation and angle of view are constantly acquired from the markers through perspective transformations. Therefore, the latter can be included in the marker-based context.

We faced several challenges throughout our research process, such as fragmentation of abstract umbrella terms (e.g., musical interactions) and semantically identifying new context categories (e.g., music-based). This process revealed their respective methods (e.g., musical feedback and music visualization), thus clarifying the actual contribution of previous work and highlighting research gaps. It is noteworthy that while categorizing the recognized modalities with their respective contexts in a graph, some modalities share context (Figure 7). An interesting future study may present modalities, contexts, and methods as RDF triples, representing the proposed taxonomy semantically. As this paper makes the first attempt at classifying human perceptual modalities, further research is required to establish a robust representation. Furthermore, this taxonomy may be readjusted to include a wider field of modalities, such as the smell-based and taste-based modalities not covered in this paper. The proposed contexts and methods may also be enriched, and the final taxonomy can be expanded to cover the techniques used in each method. For example, the *hand gesture recognition* method may include techniques that exploit the YOLO (you only look once) algorithm or R-CNN (region convolutional neural network). Expanding the classification to include the techniques used for each method can later connect methods with testing frameworks and methodologies. Well-defined testing frameworks can provide a schema based on which users can endorse or disapprove an interaction model and, therefore, provide ratings generated through a shared process.

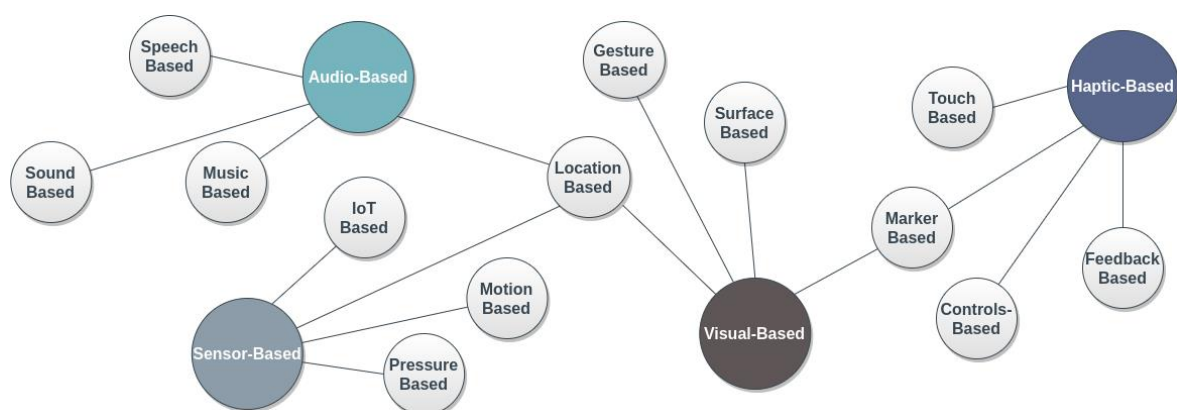


Figure 7. A graph representation of the identified modalities with their contexts.

Author Contributions: The individual contributions of the authors are (according to CRediT taxonomy): conceptualization, T.P., T.H.K., and G.E.; methodology, T.P.; validation, K.E., T.H.K., G.E., and S.S.; formal analysis, K.E., T.P., G.E., and T.H.K.; investigation, T.P.; resources, T.P.; data curation, T.P.; writing—original draft preparation, T.P. and K.E.; writing—review and editing, K.E., G.E., T.H.K., and S.S.; visualization, T.P.; supervision, K.E., T.H.K., and G.E.; project administration, G.E.; funding acquisition, G.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of a project that has received funding from the Research Committee of the University of Macedonia under the Basic Research 2020-21 funding program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Olsson, T. Concepts and Subjective Measures for Evaluating User Experience of Mobile Augmented Reality Services. In *Human Factors in Augmented Reality Environments*; Springer: New York, NY, USA, 2013; pp. 203–232.
2. Han, D.I.; Tom Dieck, M.C.; Jung, T. User experience model for augmented reality applications in urban heritage tourism. *J. Herit. Tour.* **2018**, *13*, 46–61. [[CrossRef](#)]
3. Kling, R. The organizational context of user-centered software designs. *MIS Q.* **1977**, *1*, 41–52. [[CrossRef](#)]
4. Norman, D.A.; Draper, S.W. *User Centered System Design: New Perspectives on Human-Computer Interaction*, 1st ed.; CRC Press: Boca Raton, FL, USA, 1986.
5. Etherington. Google Launches ‘Live View’ AR Walking Directions for Google Maps. Available online: tinyurl.com/y48mt75e (accessed on 28 July 2021).
6. Cordeiro, D.; Correia, N.; Jesus, R. ARZombie: A Mobile Augmented Reality Game with Multimodal Interaction. Proceedings of 2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN), Torino, Italy, 10–12 June 2015.
7. Zollhöfer, M.; Thies, J.; Garrido, P.; Bradley, D.; Beeler, T.; Pérez, P.; Stamminger, M.; Nießner, M.; Theobalt, C. State of the art on monocular 3D face reconstruction, tracking, and applications. In *Computer Graphics Forum*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2018; Volume 37, pp. 523–550.
8. Ladwig, P.; Geiger, C. A Literature Review on Collaboration in Mixed Reality. In *International Conference on Remote Engineering and Virtual Instrumentation*; Springer: New York, NY, USA, 2018; pp. 591–600.
9. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
10. Coutrix, C.; Nigay, L. Mixed Reality: A Model of Mixed Interaction. In Proceedings of the Working Conference on Advanced Visual Interfaces, Venezia, Italy, 23–26 May 2006.
11. Evangelidis, K.; Papadopoulos, T.; Sylaiou, S. Mixed Reality: A Reconsideration Based on Mixed Objects and Geospatial Modalities. *Appl. Sci.* **2021**, *11*, 2417. [[CrossRef](#)]
12. Chen, W. Historical Oslo on a handheld device—a mobile augmented reality application. *Procedia Comput. Sci.* **2014**, *35*, 979–985. [[CrossRef](#)]
13. Oleksy, T.; Wnuk, A. Augmented places: An impact of embodied historical experience on attitudes towards places. *Comput. Hum. Behav.* **2016**, *57*, 11–16. [[CrossRef](#)]
14. Phithak, T.; Kamollimsakul, S. Korat Historical Explorer: The Augmented Reality Mobile Application to Promote Historical Tourism in Korat. In Proceedings of the 2020 the 3rd International Conference on Computers in Management and Business, Tokyo, Japan, 31 January–2 February 2020; pp. 283–289.
15. Nguyen, V.T.; Jung, K.; Yoo, S.; Kim, S.; Park, S.; Currie, M. Civil War Battlefield Experience: Historical Event Simulation Using Augmented Reality Technology. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9–11 December 2019.
16. Cavallo, M.; Rhodes, G.A.; Forbes, A.G. Riverwalk: Incorporating Historical Photographs in Public Outdoor Augmented Reality Experiences. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Yucatan, Mexico, 19–23 September 2016.
17. Angelini, C.; Williams, A.S.; Kress, M.; Vieira, E.R.; D’Souza, N.; Rishe, N.D.; Ortega, F.R. City planning with augmented reality. *arXiv* **2020**, arXiv:2001.06578.
18. Sihi, D. Home sweet virtual home: The use of virtual and augmented reality technologies in high involvement purchase decisions. *J. Res. Interact. Mark.* **2018**, *12*, 398–417. [[CrossRef](#)]
19. Fu, M.; Liu, R. The Application of Virtual Reality and Augmented Reality in Dealing with Project Schedule Risks. In Proceedings of the Construction Research Congress, New Orleans, LA, USA, 2–4 April 2018; pp. 429–438.

20. Amaguaña, F.; Collaguazo, B.; Tituaña, J.; Aguilar, W.G. Simulation System Based on Augmented Reality for Optimization of Training Tactics on Military Operations. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*; Springer: New York, NY, USA, 2018; pp. 394–403.
21. Livingston, M.A.; Rosenblum, L.J.; Brown, D.G.; Schmidt, G.S.; Julier, S.J.; Baillet, Y.; Maassel, P. Military Applications of Augmented Reality. In *Handbook of Augmented Reality*, 2011th ed.; Springer: New York, NY, USA, 2011; pp. 671–706.
22. Hagan, A. Illusion & Delusion: Living in reality when inventing imaginary worlds. *J. Animat. Annu. Creat. Aust.* **2015**, *75*, 75–80.
23. Ramos, F.; Granell, C.; Ripolles, O. An Architecture for the Intelligent Creation of Imaginary Worlds for Running. In *Intelligent Computer Graphics 2012*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 209–225.
24. Akins, H.B.; Smith, D.A. Imaging planets from imaginary worlds. *Phys. Teach.* **2018**, *56*, 486–487. [[CrossRef](#)]
25. Bunt, H. Issues in multimodal human-computer communication. In *International Conference on Cooperative Multimodal Communication*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 1–12.
26. Quek, F.; McNeill, D.; Bryll, R.; Duncan, S.; Ma, X.F.; Kirbas, C.; Ansari, R. Multimodal human discourse: Gesture and speech. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2002**, *9*, 171–193. [[CrossRef](#)]
27. Ling, J.; Peng, Z.; Yin, L.; Yuan, X. How Efficiency and Naturalness Change in Multimodal Interaction in Mobile Navigation Apps. In *International Conference on Applied Human Factors and Ergonomics*; Springer: New York, NY, USA, 2020; pp. 196–207.
28. Camba, J.; Contero, M.; Salvador-Herranz, G. Desktop vs. Mobile: A Comparative Study of Augmented Reality Systems for Engineering Visualizations in Education. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*; IEEE: Manhattan, NY, USA, 2014; pp. 1–8.
29. Bekele, M.K.; Pierdicca, R.; Frontoni, E.; Malinverni, E.S.; Gain, J. A survey of augmented, virtual, and mixed reality for cultural heritage. *J. Comput. Cult. Herit. (JOCCH)* **2018**, *11*, 1–36. [[CrossRef](#)]
30. Karray, F.; Alemzadeh, M.; Abou Saleh, J.; Arab, M.N. Human-computer interaction: Overview on state of the art. *Int. J. Smart Sens. Intell. Syst.* **2017**, *1*, 137–159. [[CrossRef](#)]
31. Nizam, S.S.M.; Abidin, R.Z.; Hashim, N.C.; Lam, M.C.; Arshad, H.; Majid, N.A.A. A review of multimodal interaction technique in augmented reality environment. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2018**, *8*, 1460–1469. [[CrossRef](#)]
32. Saroha, K.; Sharma, S.; Bhatia, G. Human computer interaction: An intellectual approach. *IJCSMS Int. J. Comput. Sci. Manag. Stud.* **2011**, *11*, 147–154.
33. Tektonidis, D.; Koumpis, A. Accessible Internet-of-Things and Internet-of-Content Services for All in the Home or on the Move. *Int. J. Interact. Mob. Technol.* **2012**, *6*, 25–33. [[CrossRef](#)]
34. Tektonidis, D.; Karagiannidis, C.; Kouroupetroglou, C.; Koumpis, A. Intuitive User Interfaces to Help Boost Adoption of Internet-of-Things and Internet-of-Content Services for All. In *Inter-Cooperative Collective Intelligence: Techniques and Applications*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 93–110.
35. Badhiti, K.R. HCI-Ubiquitous Computing and Ambient Technologies in the Universe. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2015**, *3*, 1.
36. Raymond, O.U.; Ogbonna, A.C.; Shade, K. Human Computer Interaction: Overview and Challenges. 2014. Available online: https://www.researchgate.net/publication/263254929_Human_Computer_InteractionOverview_and_Challenges (accessed on 10 September 2021).
37. Ahluwalia, S.; Pal, B.; Wason, R. Gestural Interface Interaction: A Methodical Review. *Int. J. Comput. Appl.* **2012**, *60*, 21–28.
38. Nautiyal, L.; Malik, P.; Ram, M. Computer Interfaces in Diagnostic Process of Industrial Engineering. In *Diagnostic Techniques in Industrial Engineering*; Springer: New York, NY, USA, 2018; pp. 157–170.
39. Alao, O.D.; Joshua, J.V. Human Ability Improvement with Wireless Sensors in Human Computer Interaction. *Int. J. Comput. Appl. Technol. Res.* **2019**, *8*, 331–339.
40. Norman, D. *The Design of Everyday Things: Revised and Expanded Edition*; Basic Books: New York, NY, USA, 2013.
41. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
42. McGann, M. Perceptual modalities: Modes of presentation or modes of interaction? *J. Conscious. Stud.* **2010**, *17*, 72–94.
43. Stokes, D.; Matthen, M.; Biggs, S. (Eds.) *Perception and Its Modalities*; Oxford University Press: Oxford, UK, 2015.
44. Pamparău, C.; Vatavu, R.D. A Research Agenda Is Needed for Designing for the User Experience of Augmented and Mixed Reality: A Position Paper. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia, Essen, Germany, 22–25 November 2020; pp. 323–325.
45. Ghazwani, Y.; Smith, S. Interaction in augmented reality: Challenges to enhance user experience. In Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations, Sydney, Australia, 14–16 February 2020; pp. 39–44.
46. Irshad, S.; Rambli, D.R.B.A. User Experience of Mobile Augmented Reality: A Review of Studies. In Proceedings of the 2014 3rd International Conference on User Science and Engineering (i-USER), Shah Alam, Malaysia, 2–5 September 2014; pp. 125–130.
47. Côté, S.; Trudel, P.; Desbiens, M.; Giguère, M.; Snyder, R. Live Mobile Panoramic High Accuracy Augmented Reality for Engineering and Construction. In Proceedings of the Construction Applications of Virtual Reality (CONVR), London, UK, 30–31 October 2013; pp. 1–10.
48. Azimi, E.; Qian, L.; Navab, N.; Kazanzides, P. Alignment of the virtual scene to the tracking space of a mixed reality head-mounted display. *arXiv* **2017**, arXiv:1703.05834.
49. Peng, J. Changing Spatial Boundaries. Available online: <http://www.interactivearchitecture.org/changing-spatial-boundaries.html> (accessed on 10 September 2021).

50. Bill, R.; Cap, C.; Kofahl, M.; Mundt, T. Indoor and outdoor positioning in mobile environments a review and some investigations on wlan positioning. *Geogr. Inf. Sci.* **2004**, *10*, 91–98.
51. Vijayaraghavan, G.; Kaner, C. Bug taxonomies: Use them to generate better tests. *Star East* **2003**, *2003*, 1–40.
52. Lazar, J.; Feng, J.H.; Hochheiser, H. *Research Methods in Human-Computer Interaction*; Morgan Kaufmann: Burlington, MA, USA, 2017.
53. Bachmann, D.; Weichert, F.; Rinkenauer, G. Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors* **2018**, *18*, 2194. [[CrossRef](#)] [[PubMed](#)]
54. Franco, J.; Cabral, D. Augmented object selection through smart glasses. In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, Pisa, Italy, 26–29 November 2019; pp. 1–5.
55. Mossel, A.; Venditti, B.; Kaufmann, H. 3DTouch and HOMER-S: Intuitive manipulation techniques for one-handed handheld augmented reality. In Proceedings of the Virtual Reality International Conference: Laval Virtual, Laval, France, 4–6 April 2013; pp. 1–10.
56. Narzt, W.; Pomberger, G.; Ferscha, A.; Kolb, D.; Müller, R.; Wiegardt, J.; Lindinger, C. Augmented reality navigation systems. *Univers. Access Inf. Soc.* **2006**, *4*, 177–187. [[CrossRef](#)]
57. Reeves, L.M.; Lai, J.; Larson, J.A.; Oviatt, S.; Balaji, T.S.; Buisine, S.; Wang, Q.Y. Guidelines for multimodal user interface design. *Commun. ACM* **2004**, *47*, 57–59. [[CrossRef](#)]
58. Rokhsaritalemi, S.; Sadeghi-Niaraki, A.; Choi, S.M. A review on mixed reality: Current trends, challenges and prospects. *Appl. Sci.* **2020**, *10*, 636. [[CrossRef](#)]
59. Meža, S.; Turk, Ž.; Dolenc, M. Measuring the potential of augmented reality in civil engineering. *Adv. Eng. Softw.* **2015**, *90*, 1–10. [[CrossRef](#)]
60. Ellenberger, K. Virtual and augmented reality in public archaeology teaching. *Adv. Archaeol. Pract.* **2017**, *5*, 305–309. [[CrossRef](#)]
61. Barsom, E.Z.; Graafland, M.; Schijven, M.P. Systematic review on the effectiveness of augmented reality applications in medical training. *Surg. Endosc.* **2016**, *30*, 4174–4183. [[CrossRef](#)] [[PubMed](#)]
62. Wu, H.K.; Lee, S.W.Y.; Chang, H.Y.; Liang, J.C. Current status, opportunities and challenges of augmented reality in education. *Comput. Educ.* **2013**, *62*, 41–49. [[CrossRef](#)]
63. Chen, L.; Francis, K.; Tang, W. Semantic Augmented Reality Environment with Material-Aware Physical Interactions. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 9–13 October 2017; pp. 135–136.
64. Chen, L.; Tang, W.; John, N.; Wan, T.R.; Zhang, J.J. Context-aware mixed reality: A framework for ubiquitous interaction. *arXiv* **2018**, arXiv:1803.05541.
65. Serafin, S.; Geronazzo, M.; Erkut, C.; Nilsson, N.C.; Nordahl, R. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE Comput. Graph. Appl.* **2018**, *38*, 31–43. [[CrossRef](#)] [[PubMed](#)]
66. Sporr, A.; Blank-Landeshammer, B.; Kasess, C.H.; Drexler-Schmid, G.H.; Kling, S.; Köfinger, C.; Reichl, C. Extracting boundary conditions for sound propagation calculations using augmented reality. *Elektrotechnik Inf.* **2021**, *138*, 197–205. [[CrossRef](#)]
67. Han, C.; Luo, Y.; Mesgarani, N. Real-Time Binaural Speech Separation with Preserved Spatial Cues. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 6404–6408.
68. Rolland, J.P.; Holloway, R.L.; Fuchs, H. Comparison of optical and video see-through, head-mounted displays. In Telemanipulator and Telepresence Technologies. *Int. Soc. Opt. Photonics* **1995**, *2351*, 293–307.
69. Nieters, J. Defining an Interaction Model: The Cornerstone of Application Design. 2012. Available online: <https://www.uxmatters.com/mt/archives/2012/01/defining-an-interaction-model-the-cornerstone-of-application-design.php> (accessed on 10 September 2021).
70. Sostel. Eye-Gaze and Dwell. Available online: <https://docs.microsoft.com/en-us/windows/mixed-reality/design/gaze-and-dwell-eyes> (accessed on 28 July 2021).
71. Ballantyne, M.; Jha, A.; Jacobsen, A.; Hawker, J.S.; El-Glaly, Y.N. Study of Accessibility Guidelines of Mobile Applications. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, Cairo, Egypt, 25–28 November 2018; pp. 305–315.
72. Piumsomboon, T.; Lee, G.; Lindeman, R.W.; Billingham, M. Exploring Natural Eye-Gaze-Based Interaction for Immersive Virtual Reality. In Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, 18–19 March 2017; pp. 36–39.
73. Jacob, R.J. What You Look at Is What You Get: Eye Movement-Based Interaction Techniques. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Washington, DC, USA, 1–5 April 1990; pp. 11–18.
74. Pomplun, M.; Sunkara, S. Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction. In Proceedings of the International Conference on HCI, Crete, Greece, 22–27 June 2003; Volume 273.
75. Samara, A.; Galway, L.; Bond, R.; Wang, H. Human-Computer Interaction Task Classification via Visual-Based Input Modalities. In *International Conference on Ubiquitous Computing and Ambient Intelligence*; Springer: Manhattan, NY, USA, 2017; pp. 636–642.
76. Bazarevsky, V.; Kartynnik, Y.; Vakunov, A.; Raveendran, K.; Grundmann, M. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv* **2019**, arXiv:1907.05047.

77. Kantonen, T.; Woodward, C.; Katz, N. Mixed Reality in Virtual World Teleconferencing. In Proceedings of the 2010 IEEE Virtual Reality Conference (VR), Waltham, MA, USA, 20–24 March 2010; pp. 179–182.
78. Shreyas, K.K.; Rajendran, R.; Wan, Q.; Panetta, K.; Agaian, S.S. TERNET: A Deep Learning Approach for Thermal Face Emotion Recognition. In Proceedings of the SPIE 10993, Mobile Multimedia/Image Processing, Security, and Applications, Baltimore, MD, USA, 4–18 April 2019; Volume 1099309.
79. Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Narayanan, S. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, USA, 14–15 October 2004; pp. 205–211.
80. Acquisti, A.; Gross, R.; Stutzman, F.D. Face recognition and privacy in the age of augmented reality. *J. Priv. Confid.* **2014**, *6*, 1. [[CrossRef](#)]
81. Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors* **2018**, *18*, 416. [[CrossRef](#)]
82. Lei, G.; Li, X.H.; Zhou, J.L.; Gong, X.G. Geometric Feature Based Facial Expression Recognition Using Multiclass Support Vector Machines. In Proceedings of the 2009 IEEE International Conference on Granular Computing, Nanchang, China, 17–19 August 2009; pp. 318–321.
83. Christou, N.; Kanojiya, N. Human Facial Expression Recognition with Convolution Neural Networks. In *Third International Congress on Information and Communication Technology*; Springer: Singapore, 2019; pp. 539–545.
84. Slater, M.; Usoh, M. Body centred interaction in immersive virtual environments. *Artif. Life Virtual Real.* **1994**, *1*, 125–148.
85. Lee, I.J. Kinect-for-windows with augmented reality in an interactive roleplay system for children with an autism spectrum disorder. *Interact. Learn. Environ.* **2021**, *29*, 688–704. [[CrossRef](#)]
86. Hsiao, K.F.; Rashvand, H.F. Body Language and Augmented Reality Learning Environment. In Proceedings of the 2011 Fifth FTRA International Conference on multimedia and ubiquitous engineering, Crete, Greece, 28–30 June 2011; pp. 246–250.
87. Umeda, T.; Correa, P.; Marques, F.; Marichal, X. A Real-Time Body Analysis for Mixed Reality Application. In Proceedings of the Korea-Japan Joint Workshop on Frontiers of Computer Vision, FCV-2004, Fukuoka, Japan, 4–6 February 2014.
88. Yousefi, S.; Kidane, M.; Delgado, Y.; Chana, J.; Reski, N. 3D Gesture-Based Interaction for Immersive Experience in Mobile VR. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2121–2126.
89. Weichel, C.; Lau, M.; Kim, D.; Villar, N.; Gellersen, H.W. MixFab: A Mixed-Reality Environment for Personal Fabrication. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 3855–3864.
90. Yang, M.T.; Liao, W.C. Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Trans. Learn. Technol.* **2014**, *7*, 107–117. [[CrossRef](#)]
91. Porter, S.R.; Marnier, M.R.; Smith, R.T.; Zucco, J.E.; Thomas, B.H. Validating Spatial Augmented Reality for Interactive Rapid Prototyping. In Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality, Seoul, Korea, 13–16 October 2010.
92. Fadzli, F.E.; Ismail, A.W.; Aladin, M.Y.F.; Othman, N.Z.S. A Review of Mixed Reality Telepresence. In *IOP Conference Series: Materials Science and Engineering*, 2nd ed.; Springer: New York, NY, USA, 2021; Volume 864, p. 012081.
93. Moares, R.; Jadhav, V.; Bagul, R.; Jacobo, R.; Rajguru, S. Inter AR: Interior Decor App Using Augmented Reality Technology. In Proceedings of the 5th International Conference on Cyber Security & Privacy in Communication Networks (ICCS), Kurukshehra, Haryana, India, 29–30 November 2019.
94. Polvi, J.; Taketomi, T.; Yamamoto, G.; Dey, A.; Sandor, C.; Kato, H. SlidAR: A 3D positioning method for SLAM-based handheld augmented reality. *Comput. Graph.* **2016**, *55*, 33–43. [[CrossRef](#)]
95. Sandu, M.; Scarlat, I.S. Augmented Reality Uses in Interior Design. *Inform. Econ.* **2018**, *22*, 5–13. [[CrossRef](#)]
96. Labrie, A.; Cheng, J. Adapting Usability Heuristics to the Context of Mobile Augmented Reality. In Proceedings of the Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, 20–23 October 2020.
97. Ong, S.K.; Wang, Z.B. Augmented assembly technologies based on 3D bare-hand interaction. *CIRP Ann.* **2011**, *60*, 1–4. [[CrossRef](#)]
98. Mitsova, H.; Mitas, L.; Ratti, C.; Ishii, H.; Alonso, J.; Harmon, R.S. Real-time landscape model interaction using a tangible geospatial modeling environment. *IEEE Comput. Graph. Appl.* **2006**, *26*, 55–63. [[CrossRef](#)] [[PubMed](#)]
99. Punpongsonan, P.; Iwai, D.; Sato, K. Softar: Visually manipulating haptic softness perception in spatial augmented reality. *IEEE Trans. Vis. Comput. Graph.* **2015**, *21*, 1279–1288. [[CrossRef](#)]
100. Barreira, J.; Bessa, M.; Barbosa, L.; Magalhães, L. A context-aware method for authentically simulating outdoors shadows for mobile augmented reality. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 1223–1231. [[CrossRef](#)]
101. Lensing, P.; Broll, W. Instant Indirect Illumination for Dynamic Mixed Reality Scenes. In Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, USA, 5–8 November 2012.
102. Fiala, M. ARTag, a Fiducial Marker System Using Digital Techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*; IEEE: Manhattan, NY, USA, June 2005; Volume 2, pp. 590–596.
103. Onime, C.; Uhomobhi, J.; Wang, H.; Santachiara, M. A reclassification of markers for mixed reality environments. *Int. J. Inf. Learn. Technol.* **2020**, *38*, 161–173. [[CrossRef](#)]

104. Fiala, M. Designing highly reliable fiducial markers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 1317–1324. [[CrossRef](#)]
105. Naimark, L.; Foxlin, E. Circular Data Matrix Fiducial System And Robust Image Processing For A Wearable Vision-Inertial Self-tracker. In *Proceedings of the International Symposium on Mixed and Augmented Reality*; IEEE: Manhattan, NY, USA, 2002; pp. 27–36.
106. Bencina, R.; Kaltenbrunner, M. The Design and Evolution of Fiducials for the Reactivision System. In *Proceedings of the Third International Conference on Generative Systems in the Electronic Arts*, 1st ed.; Monash University Publishing: Melbourne, VIC, Australia, 2005.
107. Rekimoto, J.; Ayatsuka, Y. CyberCode: Designing Augmented Reality Environments with Visual Tags. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*; AMC: New York, NY, USA, 2000; pp. 1–10.
108. Rohs, M. Visual Code Widgets for Marker-Based Interaction. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops*; IEEE: Manhattan, NY, USA, 2005; pp. 506–513.
109. Boulanger, P. Application of Augmented Reality to Industrial Tele-Training. In *Proceedings of the First Canadian Conference on Computer and Robot Vision*; IEEE: Manhattan, NY, USA, 2004; pp. 320–328.
110. Pentenrieder, K.; Meier, P.; Klinker, G. Analysis of Tracking Accuracy for Single-Camera Square-Marker-Based Tracking. In *Proceedings of the Dritter Workshop Virtuelle und Erweiterte Realitt der GIFachgruppe VR/AR*; Citeseer: Koblenz, Germany, 2006.
111. Flohr, D.; Fischer, J. A Lightweight ID-Based Extension for Marker Tracking Systems. 2007. Available online: https://www.researchgate.net/publication/228541592_A_Lightweight_ID-Based_Extension_for_Marker_Tracking_Systems (accessed on 10 September 2021).
112. Mateos, L.A. AprilTags 3D: Dynamic Fiducial Markers for Robust Pose Estimation in Highly Reflective Environments and Indirect Communication in Swarm Robotics. *arXiv* **2020**, arXiv:2001.08622.
113. Wang, T.; Liu, Y.; Wang, Y. Infrared Marker Based Augmented Reality System for Equipment Maintenance. In *Proceedings of the 2008 International Conference on Computer Science and Software Engineering*; IEEE: Manhattan, NY, USA, December 2008; Volume 5, pp. 816–819.
114. Nee, A.Y.; Ong, S.K.; Chryssolouris, G.; Mourtzis, D. Augmented reality applications in design and manufacturing. *CIRP Ann.* **2012**, *61*, 657–679. [[CrossRef](#)]
115. Evangelidis, K.; Sylaiou, S.; Papadopoulos, T. Mergin’mode: Mixed reality and geoinformatics for monument demonstration. *Appl. Sci.* **2020**, *10*, 3826. [[CrossRef](#)]
116. Stricker, D.; Karigiannis, J.; Christou, I.T.; Gleue, T.; Ioannidis, N. Augmented Reality for Visitors of Cultural Heritage Sites. In *Proceedings of the Int. Conf. on Cultural and Scientific Aspects of Experimental Media Spaces*, Bonn, Germany, 21–22 September 2001; pp. 89–93.
117. Shen, R.; Terada, T.; Tsukamoto, M. A system for visualizing sound source using augmented reality. *Int. J. Pervasive Comput. Commun.* **2013**, *9*, 227–242. [[CrossRef](#)]
118. Rajguru, C.; Obrist, M.; Memoli, G. Spatial soundscapes and virtual worlds: Challenges and opportunities. *Front. Psychol.* **2020**, *11*, 2714. [[CrossRef](#)] [[PubMed](#)]
119. Härmä, A.; Jakka, J.; Tikander, M.; Karjalainen, M.; Lokki, T.; Hiipakka, J.; Lorho, G. Augmented reality audio for mobile and wearable appliances. *J. Audio Eng. Soc.* **2004**, *52*, 618–639.
120. Nordahl, R.; Turchet, L.; Serafin, S. Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 1234–1244. [[CrossRef](#)] [[PubMed](#)]
121. Verron, C.; Aramaki, M.; Kronland-Martinet, R.; Pallone, G. A 3-D immersive synthesizer for environmental sounds. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *18*, 1550–1561. [[CrossRef](#)]
122. Gaver, W.W. What in the world do we hear?: An ecological approach to auditory event perception. *Ecol. Psychol.* **1993**, *5*, 1–29. [[CrossRef](#)]
123. Che Hashim, N.; Abd Majid, N.A.; Arshad, H.; Khalid Obeidy, W. User satisfaction for an augmented reality application to support productive vocabulary using speech recognition. *Adv. Multimed.* **2018**, *2018*, 9753979. [[CrossRef](#)]
124. Billingham, M.; Kato, H.; Myojin, S. Advanced Interaction Techniques for Augmented Reality Applications. In *International Conference on Virtual and Mixed Reality*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 13–22.
125. Kato, H.; Billingham, M.; Poupyrev, I.; Imamoto, K.; Tachibana, K. Virtual Object Manipulation on a Table-Top AR Environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*; IEEE: Manhattan, NY, USA, 2000; pp. 111–119.
126. Denecke. Ariadne Spoken Dialogue System. Available online: tinyurl.com/3wjzrds (accessed on 28 July 2021).
127. Microsoft Speech SDK (SAPI 5.0). Available online: tinyurl.com/58mf8bz5 (accessed on 28 July 2021).
128. Hanifa, R.M.; Isa, K.; Mohamad, S. A review on speaker recognition: Technology and challenges. *Comput. Electr. Eng.* **2021**, *90*, 107005. [[CrossRef](#)]
129. Chollet, G.; Esposito, A.; Gentes, A.; Horain, P.; Karam, W.; Li, Z.; Zouari, L. *Multimodal Human Machine Interactions in Virtual and Augmented Reality (v-dij-14)*; Springer: New York, NY, USA, 2008.
130. Lin, T.; Huang, L.; Hannaford, B.; Tran, C.; Raiti, J.; Zaragoza, R.; James, J. Empathics system: Application of Emotion Analysis AI through Smart Glasses. In *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*; AMC: New York, NY, USA, June 2020; pp. 1–4.

131. Mirzaei, M.R.; Ghorshi, S.; Mortazavi, M. Combining Augmented Reality and Speech Technologies to Help Deaf and Hard of Hearing People. In *Proceedings of the 2012 14th Symposium on Virtual and Augmented Reality*; IEEE Computer Society: Washington, DC, USA, May 2012; pp. 174–181.
132. Altosaar, R.; Tindale, A.; Doyle, J. Physically Colliding with Music: Full-body Interactions with an Audio-only Virtual Reality Interface. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*; AMC: New York, NY, USA, March 2019; pp. 553–557.
133. Bauer, V.; Bouchara, T. First Steps Towards Augmented Reality Interactive Electronic Music Production. In *Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*; IEEE Computer Society: Washington, DC, USA, March 2021; pp. 90–93.
134. Bederson, B.B. Audio Augmented Reality: A Prototype Automated Tour Guide. In *Proceedings of the Conference companion on Human factors in computing systems*; AMC: New York, NY, USA, May 1995; pp. 210–211.
135. Paterson, N.; Naliuka, K.; Jensen, S.K.; Carrigy, T.; Haahr, M.; Conway, F. Design, Implementation and Evaluation of Audio for a Location Aware Augmented Reality Game. In *Proceedings of the 3rd International Conference on Fun and Games*; AMC: New York, NY, USA, September 2021; pp. 149–156.
136. Lyons, K.; Gandy, M.; Starner, T. Guided by voices: An audio augmented reality system. In *Proceedings of the International Conference of Auditory Display (ICAD)*; Georgia Institute of Technology: Atlanta, GA, USA, 2000; pp. 57–62.
137. Blum, J.R.; Bouchard, M.; Cooperstock, J.R. What’s around Me? Spatialized Audio Augmented Reality for Blind Users with a Smartphone. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 49–62.
138. Yang, Y.; Shim, J.; Chae, S.; Han, T.D. Mobile Augmented Reality Authoring Tool. In *Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*; IEEE: Manhattan, NY, USA, February 2016; pp. 358–361.
139. Jung, J.; Hong, J.; Park, S.; Yang, H.S. Smartphone as an Augmented Reality Authoring Tool via Multi-Touch Based 3D Interaction Method. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*; IEEE: Manhattan, NY, USA, December 2012; pp. 17–20.
140. Kasahara, S.; Niiyama, R.; Heun, V.; Ishii, H. exTouch: Spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, Barcelona, Spain, 10–13 February 2013; pp. 223–228.
141. Yannier, N.; Koedinger, K.R.; Hudson, S.E. Learning from Mixed-Reality Games: Is Shaking a Tablet as Effective as Physical Observation? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*; AMC: New York, NY, USA, April 2015; pp. 1045–1054.
142. Xiao, R.; Schwarz, J.; Throm, N.; Wilson, A.D.; Benko, H. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 1653–1660. [[CrossRef](#)] [[PubMed](#)]
143. Zhang, Y.; Kienzle, W.; Ma, Y.; Ng, S.S.; Benko, H.; Harrison, C. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*; AMC: New York, NY, USA, October 2019; pp. 1151–1159.
144. Jiawei, W.; Li, Y.; Tao, L.; Yuan, Y. Three-Dimensional Interactive pen Based on Augmented Reality. In *Proceedings of the 2010 International Conference on Image Analysis and Signal Processing*, Povo de Varzim, Portugal, 21–23 June 2010.
145. Yue, Y.T.; Zhang, X.; Yang, Y.; Ren, G.; Choi, Y.K.; Wang, W. Wiredraw: 3D Wire Sculpturing Guided with MIXED reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*; AMC: New York, NY, USA, May 2017; pp. 3693–3704.
146. Yun, K.; Woo, W. Tech-Note: Spatial Interaction Using Depth Camera for Miniature AR. In *Proceedings of the 2009 IEEE Symposium on 3D User Interfaces*; IEEE Computer Society: Washington, DC, USA, March 2009; pp. 119–122.
147. Back, M.; Cohen, J.; Gold, R.; Harrison, S.; Minneman, S. Listen reader: An Electronically Augmented Paper-Based Book. In *Proceedings of the SIGCHI conference on Human factors in computing systems*; AMC: New York, NY, USA, March 2001; pp. 23–29.
148. Leitner, J.; Haller, M.; Yun, K.; Woo, W.; Sugimoto, M.; Inami, M. IncreTable, a Mixed Reality Tabletop Game Experience. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*; AMC: New York, NY, USA, December 2008; pp. 9–16.
149. Hashimoto, S.; Ishida, A.; Inami, M.; Igarashi, T. Touchme: An Augmented Reality Based Remote Robot Manipulation. In *Proceedings of the 21st International Conference on Artificial Reality and Telexistence*; AMC: New York, NY, USA, November 2011; Volume 2.
150. Chuah, J.H.; Lok, B.; Black, E. Applying mixed reality to simulate vulnerable populations for practicing clinical communication skills. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 539–546. [[CrossRef](#)]
151. Jiang, H.; Weng, D.; Zhang, Z.; Bao, Y.; Jia, Y.; Nie, M. Hikeyb: High-Efficiency Mixed Reality System for Text Entry. In *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*; IEEE: Manhattan, NY, USA, October 2018; pp. 132–137.
152. Leydon, K. Sensing the Position and Orientation of Hand-Held Objects: An Overview of Techniques. 2001. Available online: http://ulsites.ul.ie/csis/sites/default/files/csis_sensing_the_position_and_orientation_of_hand-held_objects.pdf (accessed on 10 September 2021).
153. Poupyrev, I.; Harrison, C.; Sato, M. Touché: Touch and gesture Sensing for the Real World. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*; AMC: New York, NY, USA, September 2012; p. 536.

154. Kienzle, W.; Whitmire, E.; Rittaler, C.; Benko, H. ElectroRing: Subtle Pinch and Touch Detection with a Ring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*; AMC: New York, NY, USA, May 2021; pp. 1–12.
155. Han, T.; Anderson, F.; Irani, P.; Grossman, T. Hydroring: Supporting Mixed Reality Haptics Using Liquid Flow. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*; AMC: New York, NY, USA, October 2018; pp. 913–925.
156. Bai, H.; Li, S.; Shepherd, R.F. Elastomeric Haptic Devices for Virtual and Augmented Reality. *Adv. Funct. Mater.* **2021**, 2009364. Available online: https://www.researchgate.net/profile/Shuo-Li-38/publication/352066240_Elastomeric_Haptic_Devices_for_Virtual_and_Augmented_Reality/links/60d1f3f345851566d5809357/Elastomeric-Haptic-Devices-for-Virtual-and-Augmented-Reality.pdf (accessed on 10 September 2021). [CrossRef]
157. Talasaz, A.; Trejos, A.L.; Patel, R.V. Effect of Force Feedback on Performance of Robotics-Assisted Suturing. In *Proceedings of the 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*; IEEE: Manhattan, NY, USA, June 2012; pp. 823–828.
158. Akinbiyi, T.; Reiley, C.E.; Saha, S.; Burschka, D.; Hasser, C.J.; Yuh, D.D.; Okamura, A.M. Dynamic Augmented Reality for Sensory Substitution in Robot-Assisted Surgical Systems. In *Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society*; IEEE: Manhattan, NY, USA, September 2006; pp. 567–570.
159. Günther, S.; Schön, D.; Müller, F.; Mühlhäuser, M.; Schmitz, M. PneumoVolley: Pressure-Based Haptic Feedback on the Head through Pneumatic Actuation. In *Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*; AMC: New York, NY, USA, April 2020; pp. 1–10.
160. Schorr, S.B.; Okamura, A.M. Three-dimensional skin deformation as force substitution: Wearable device design and performance during haptic exploration of virtual environments. *IEEE Trans. Haptics* **2017**, *10*, 418–430. [CrossRef]
161. Meli, L.; Pacchierotti, C.; Salvietti, G.; Chinello, F.; Maisto, M.; De Luca, A.; Prattichizzo, D. Combining wearable finger haptics and augmented reality: User evaluation using an external camera and the microsoft hololens. *IEEE Robot. Autom. Lett.* **2018**, *3*, 4297–4304. [CrossRef]
162. Yang, T.H.; Kim, J.R.; Jin, H.; Gil, H.; Koo, J.H.; Kim, H.J. Recent Advances and Opportunities of Active Materials for Haptic Technologies in Virtual and Augmented Reality. *Adv. Funct. Mater.* **2021**, 2008831. Available online: <https://onlinelibrary.wiley.com/doi/full/10.1002/adfm.202008831> (accessed on 10 September 2021). [CrossRef]
163. Kim, T.; Cooperstock, J.R. Enhanced Pressure-Based Multimodal Immersive Experiences. In *Proceedings of the 9th Augmented Human International Conference*; AMC: New York, NY, USA, February 2018; pp. 1–3.
164. Qian, G.; Peng, B.; Zhang, J. Gesture Recognition Using Video and Floor Pressure Data. In *Proceedings of the 2012 19th IEEE International Conference on Image Processing*; AMC: New York, NY, USA, September 2012; pp. 173–176.
165. Minh, V.T.; Katushin, N.; Pumwa, J. Motion tracking glove for augmented reality and virtual reality. *Paladyn J. Behav. Robot.* **2019**, *10*, 160–166. [CrossRef]
166. Zhu, M.; Sun, Z.; Zhang, Z.; Shi, Q.; Chen, T.; Liu, H.; Lee, C. Sensory-Glove-Based Human Machine Interface for Augmented Reality (AR) Applications. In *Proceedings of the 2020 IEEE 33rd International Conference on Micro Electromechanical Systems (MEMS)*, Vancouver, BC, Canada, 18–22 January 2020; pp. 16–19.
167. Nguyen, T.V.; Kamma, S.; Adari, V.; Lesthaeghe, T.; Boehnlein, T.; Kramb, V. Mixed reality system for nondestructive evaluation training. *Virtual Real.* **2020**, *25*, 709–718. [CrossRef]
168. Gül, S.; Bosse, S.; Podborski, D.; Schierl, T.; Hellge, C. Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality. In *Proceedings of the 28th ACM International Conference on Multimedia*; AMC: New York, NY, USA, October 2020; pp. 3632–3641.
169. Brena, R.F.; García-Vázquez, J.P.; Galván-Tejada, C.E.; Muñoz-Rodríguez, D.; Vargas-Rosales, C.; Fangmeyer, J. Evolution of indoor positioning technologies: A survey. *J. Sens.* **2017**, *2017*, 2630413. [CrossRef]
170. Benyon, D.; Quigley, A.; O’keefe, B.; Riva, G. Presence and digital tourism. *AI Soc.* **2014**, *29*, 521–529. [CrossRef]
171. Schrier, K. Using Augmented Reality Games to Teach 21st Century Skills. In *ACM SIGGRAPH 2006 Educators Program*; AMC: New York, NY, USA, 2006; p. 15-es.
172. Sakpere, W.; Adeyeye-Oshin, M.; Mlitwa, N.B. A state-of-the-art survey of indoor positioning and navigation systems and technologies. *S. Afr. Comput. J.* **2017**, *29*, 145–197. [CrossRef]
173. Jiang, W.; Li, Y.; Rizos, C.; Cai, B.; Shangguan, W. Seamless indoor-outdoor navigation based on GNSS, INS and terrestrial ranging techniques. *J. Navig.* **2017**, *70*, 1183–1204. [CrossRef]
174. Zhuang, Y.; Hua, L.; Qi, L.; Yang, J.; Cao, P.; Cao, Y.; Haas, H. A survey of positioning systems using visible LED lights. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 1963–1988. [CrossRef]
175. Afzalan, M.; Jazizadeh, F. Indoor positioning based on visible light communication: A performance-based survey of real-world prototypes. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–36. [CrossRef]
176. Madakam, S.; Lake, V.; Lake, V.; Lake, V. Internet of Things (IoT): A literature review. *J. Comput. Commun.* **2015**, *3*, 164. [CrossRef]
177. Atsali, G.; Panagiotakis, S.; Markakis, E.; Mastorakis, G.; Mavromoustakis, C.X.; Pallis, E.; Malamos, A. A mixed reality 3D system for the integration of X3DoM graphics with real-time IoT data. *Multimed. Tools Appl.* **2018**, *77*, 4731–4752. [CrossRef]
178. Natephra, W.; Motamedi, A. Live Data Visualization of IoT Sensors Using Augmented Reality (AR) and BIM. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC 2019)*, Banff, AB, Canada, 21–24 May 2019; pp. 21–24.
179. Phupattanasilp, P.; Tong, S.R. Augmented reality in the integrative internet of things (AR-IoT): Application for precision farming. *Sustainability* **2019**, *11*, 2658. [CrossRef]

180. Hoppenstedt, B.; Schmid, M.; Kammerer, K.; Scholta, J.; Reichert, M.; Pryss, R. Analysis of Fuel Cells Utilizing Mixed Reality and IoT Achievements. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*; Springer: New York, NY, USA, 2019; pp. 371–378.
181. Yasmin, R.; Salminen, M.; Gilman, E.; Petäjajarvi, J.; Mikhaylov, K.; Pakanen, M.; Pouttu, A. Combining IoT Deployment and Data Visualization: Experiences within campus Maintenance USE-case. In *Proceedings of the 2018 9th International Conference on the Network of the Future (NOF)*; IEEE: Manhattan, NY, USA, November 2018; pp. 101–105.
182. Pokric, B.; Krco, S.; Drajić, D.; Pokric, M.; Rajs, V.; Mihajlovic, Z.; Jovanovic, D. Augmented Reality Enabled IoT Services for Environmental Monitoring Utilising Serious Gaming Concept. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2015**, *6*, 37–55.
183. Morris, A.; Guan, J.; Lessio, N.; Shao, Y. Toward Mixed Reality Hybrid Objects with IoT Avatar Agents. In *Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, ON, Canada, 11–14 October 2020; pp. 766–773.