applied sciences

MDPI

Review

Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review

Joo Chan Kim [1], Teemu H. Laine [2,*] and Christer Åhlund [1]

[1] Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, 93187 Skellefteå, Sweden; joo.chan.kim@ltu.se (J.C.K.); christer.ahlund@ltu.se (C.Å.)

[2] Department of Digital Media, Ajou University, Suwon 16499, Korea

* Correspondence: teemu@ubilife.net; Tel.: +82-31-219-1851

Abstract: Technology developments have expanded the diversity of interaction modalities that can be used by an agent (either a human or machine) to interact with a computer system. This expansion has created the need for more natural and user-friendly interfaces in order to achieve effective user experience and usability. More than one modality can be provided to an agent for interaction with a system to accomplish this goal, which is referred to as a multimodal interaction (MI) system. The Internet of Things (IoT) and augmented reality (AR) are popular technologies that allow interaction systems to combine the real-world context of the agent and immersive AR content. However, although MI systems have been extensively studied, there are only several studies that reviewed MI systems that used IoT and AR. Therefore, this paper presents an in-depth review of studies that proposed various MI systems utilizing IoT and AR. A total of 23 studies were identified and analyzed through a rigorous systematic literature review protocol. The results of our analysis of MI system architectures, the relationship between system components, input/output interaction modalities, and open research challenges are presented and discussed to summarize the findings and identify future research and development avenues for researchers and MI developers.

Keywords: multimodal interaction; interaction modalities; mixed/augmented reality; Internet of Things; systematic literature review; multimodal UI

check for updates

Citation: Kim, J.C.; Laine, T.H.; Åhlund, C. Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review. *Appl. Sci.* **2021**, *11*, 1738. https://doi.org/10.3390/app11041738

Academic Editor: Vassilis Charissis

Received: 22 January 2021
Accepted: 10 February 2021
Published: 16 February 2021

1. Introduction

Since the first personal computers were released in the late 1970s, computer graphics, networking, and data processing technologies have evolved to provide high accuracy and powerful performance for rich interactive applications. These developments have enabled various methods of interaction between humans and machines to provide more natural and user-friendly interfaces, thereby contributing to improve user experience and usability [1]. Increased network performance allows for communication to handle interaction modalities in real-time, which is exemplified by smart environments [2,3] and autonomous vehicles [4]. In this study, we focus on interaction methods that are implemented through one or more modalities. "Modality" is defined as a method of transferring data that can be used to interpret the sender's state [5] or intent [6].

A single modality, such as touch on a keyboard, mouse, or screen of a mobile device, is commonly used in many interactive systems. However, by combining modalities, more information can be communicated, leading to an enhanced experience. Thus, it is unsurprising that multiple modalities have started to be considered within a single system [5,6]. Furthermore, users have different interaction preferences and contextual requirements for different modalities (e.g., visual modality in loud environments). Thus, modality flexibility has become a desired feature of interaction systems that aims to meet the requirements of multiple needs, use cases, and contexts. Expanding the diversity of modalities aims not only to overcome the limitation of available modalities for the end user, but also provide a more natural interaction to enhance usability [5,6]. A system that employs multiple modalities in the input or output channel is referred to as a multimodal interaction (MI) system. Figure 1 illustrates our interpretation of a general architecture of MI systems.

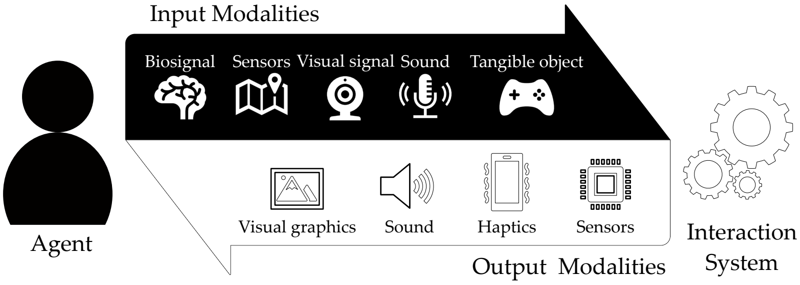


Figure 1. General architecture of multimodal interaction that uses diverse modalities for the input and output channel.

An MI system allows an agent to send input data to it by combining and organizing more than one modality. In addition, the MI system can provide output data in multiple modalities. The agent that interacts with the MI system can be either a human or machine.

Advances in technology have increased the possibility of developing new forms of MI systems that can bridge the real world and virtual world. The advancement of network and computing power allows a variety of devices to connect to the Internet with sufficient performance levels, thereby making it possible to achieve real-time rendering of virtual content that is based on the agent's context data. The Internet of Things (IoT) is the technology that realizes this objective, whereby information and communication technologies (ICT), such as sensors, are mounted on physical objects and connected to the Internet [7]. It is expected that physical items will be increasingly equipped with ICT and turned into IoT [8], thus introducing opportunities to apply interaction modalities in new ways.

While IoT mainly focuses on enabling ubiquitous connectivity, connecting physical objects to the Internet, and providing contextual data through devices, augmented reality (AR) is a type of technology that can be utilized to provide interactive interfaces in MI systems. AR is used to visualize graphical objects on a view of the real world, where a real-time camera feed on a device or head-mounted display (HMD) typically represents the view. AR objects can not only be used as simple data representations [9,10], but can also be combined with an input modality, such as touch on a mobile device [11], as part of a user interface (UI) to allow an agent to interact with an MI system [12,13]. This property distinguishes between applications that only utilize AR as a data visualization tool and those that use AR as part of their UI to interact with the underlying MI system.

For efficient interaction with the IoT, we need new means of interaction technologies. For example, when moving in a smart city environment, AR is an essential interaction technology for unobtrusively bringing context-sensitive information from IoT devices embedded in the surroundings. A practical use case is an automatic air pollution monitor that uses AR to insert visual indicators on the real-world view of the user's AR device (e.g., AR glasses) to show the air quality in the context of the user, which is measured by nearby IoT sensors in real-time [14,15]. In the scenario, when the system can automatically provide context-sensitive notifications to the user, they do not explicitly need to search for the pollution information, as it is delivered to them in a contextually relevant manner.

The combination of AR and IoT technologies makes it possible to bridge the real world, virtual world (i.e., digital content), and ambient environment (i.e., information from IoT sensors), thereby increasing the richness of information and opening possibilities for context-aware application development. For example, visualized data of an ambient environment can serve as digital content that is superimposed on the real world [10,12,13]. However, the use and development of multimodal interaction in applications that combine AR and IoT has remained a little researched area. Thus, we have identified several motivations for exploring the combination of AR and IoT from the perspective of MI systems. Firstly, the diversity in the use of AR and the affordance of IoT to provide rich data on agents' contexts form a fertile ground for developing new interaction modalities [16] and, consequently, new MI systems. Secondly, previous research has shown that a complementary combination of AR and IoT enables smart and interactive environments where the agent's interaction with IoT devices can become more intuitive [17], which is essential when designing interaction with complex artifacts and unfamiliar devices [18]. Thirdly, AR and IoT enable otherwise unperceivable information (e.g., readings from sensors, ratings of a service from other users) to be expressed through perceivable modalities (e.g., visualizations, sounds, vibrations) [16]. Finally, based on our analysis of previous literature reviews on MI systems [5,6,19,20], there is a lack of knowledge in the field on the use of AR and IoT in MI systems, as well as on the development of such systems.

Because the process of integrating different modalities is complex, it is challenging for system developers to ensure high learnability (i.e., high usability) while providing a sufficiently positive experience (i.e., Quality of Experience [QoE]). In order to achieve this, it is necessary to perform a systematic review of previous MI system development efforts, which can help shed light on the optimal design of these systems. There have been several attempts [5,6,19,20] to systematically review general research on MI systems; however, only several studies have reviewed the use of IoT and AR in MI systems [21,22].

As the contributions of this study, we analyzed the various input/output modalities and modality combinations used in MI systems. Consequently, we identified modalities and modality combinations that had not yet been explored by researchers. We also reviewed 23 systems that used IoT and AR to identify patterns of MI system architecture that developers can refer to in designing MI systems. Additionally, we analyzed the use of AR and IoT in MI systems for identifying typical, rare, and unexplored ways of using these technologies. We also attempted to identify open research challenges in the field that may attract the attention of researchers in the field of human-computer interaction (HCI) as future research avenues. The results of this study complement the previous work on MI systems from the perspectives of AR and IoT, thereby contributing to the growing knowledge base of the industrial revolution 4.0 technologies.

In Section 2, we present works that are related to the MI system and define key terms used in this study. In Section 3, we describe the methodology that is used to conduct a systematic literature survey. Section 3.1 presents the identified research questions from our survey. Subsequently, we present the result of our analysis of MI systems that use IoT and AR in Sections 4 and 5.

2. Background

2.1. Terminology

In this paper, we define modality as a way to transfer data that can be used to interpret the sender's state [5], or intent [6]. In the context of MI systems, the sender can be either the agent or interaction system (Figure 1). "State" refers to the contextual information of the sender, such as temperature, pressure, and other information. Affective computing is another way to understand the sender's context, which is a discipline that focuses on the identification of a human's emotional state via computer and sensor devices [23]. Research on affective computing in MI systems has been conducted with different data sources, such as facial expressions, sound, and body gestures [5]. "Intent" refers to the sender's planned actions or command inputs, which the MI system may be able to predict. Predicting the sender's intent by an MI system with high accuracy is an open research problem in MI system development [6].

The usability of interactive systems consists of effectiveness, efficiency, and satisfaction experienced by a user, according to the International Organization for Standardization (ISO) standard 9241-11:2018 [24]. In addition, according to the ISO/International Electrotechnical Commission technical report standard 29181-6:2013 [25] and the International Telecommunication Union standard P.10/G.100 [26], QoE is defined as a user's feeling from the use of a product, system, or service. Ease-of-use is a keyword used when a system's usability is evaluated by its users. In contrast, QoE evaluation places importance on impressions that the user receives while using a system or service. Therefore, QoE evaluation is related to subjective measurement [25], whereas usability evaluation is related to objective measurement [27].

2.2. Related Research

MI systems have been reviewed by several researchers from various perspectives. Jaimes and Sebe [5] conducted a review of MI systems that covered input modalities, system architectures, and use cases in a number of different themes (input modalities, system architectures, and applications), which were used to categorize the findings. For input modalities, they organized their findings that are based on the type of modality, such as body and hand gestures, facial expressions, and voice. Another review of MI systems was conducted by Turk [6]. Turk [6] building upon the findings of Blattner and Glinert [28], provided insights into a variety of input modalities with examples. In the list of presented by Turk [6], there are four modalities: (1) "visual", which includes facial expressions, gaze, and body gestures; (2) "auditory", which includes speech and non-speech sound; (3) "touch", which indicates pressure and actions performed by fingers; and (4) "motion capture", which records movement of the target (sender) by sensor devices. Liang et al. [20] presented a similar list of input modalities by including biosignal sensors, such as the electroencephalogram, electromyogram, and electrooculogram. Hogan and Hornecker [19] conducted a systematic review that focused on output modalities used to engage users with materials in art exhibitions. They identified visual, haptic, auditory, gustation, and thermoception modalities that were used in 154 cases. However, we noted different uses of terminology for each modality, as the classification of modalities was dependent on the perspectives of the researchers. For example, Liang et al. [20] described touch-based interaction as a "haptic-based" modality, whereas Jaimes and Sebe [5] described the same interaction as a "touch" modality. Therefore, a taxonomy of modalities is necessary, which was developed by Augstein and Neumayr [29], who presented a taxonomy of input and output modalities used in HCI. Their taxonomy, which is depicted in Figure 2, was evaluated by six HCI experts, who verified its usefulness as a tool for categorizing interaction modalities.

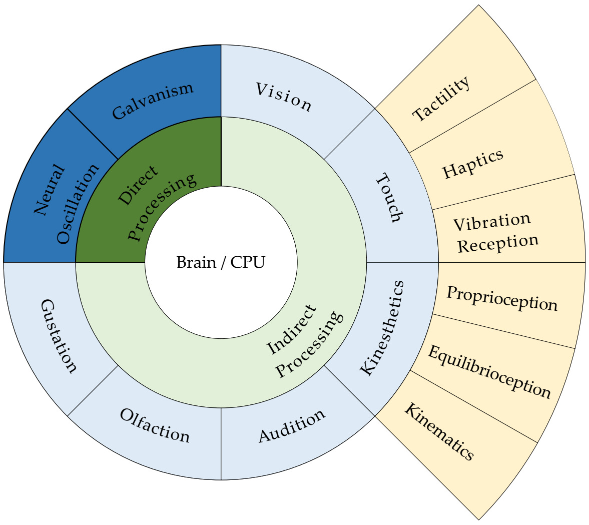


Figure 2. Taxonomy of interaction modalities proposed by Augstein and Neumayr [29].

As IoT and AR are technologies that can be used in MI systems, a number of studies using both technologies within one MI system have been published. Badouch et al. [21] explored an existing mobile application that utilized IoT and AR in the context of a smart city, while Norouzi et al. [22] analyzed 79 studies that were published from 2000 to 2018 that developed systems using IoT and AR across several platforms and different fields. The survey conducted by Norouzi et al. [22] revealed that over 60 systems utilized vision (light) as input and output modalities, while other modalities, such as sound, touch, motion, and temperature, were used in fewer than 20 systems. Furthermore, they only identified 26 studies that constructed an MI system. Nizam et al. [30] analyzed studies that were published between 2014 and 2018 that proposed MI systems using AR. From 14 studies, the researchers identified eight input modalities (gesture, speech, face, touch, adaptive, motion, tangible, and click) and reviewed the proposed MI system architectures. Furthermore, when a virtual world in AR is mapped onto a real-world geographical space, the combination, called mixed reality (MR), becomes an important topic that can attract researchers' interest as a novel technology. Mohamad Yahya Fekri and Ajune Wanis [31] presented a review of input modalities used in 10 MI systems operating in MR environments. They found that gesture and speech were used in six MI systems, while, other modalities, such as facial recognition and eye/head tracking, were used in three MI systems. In addition, they listed tangible interfaces, which are based on physical objects that are used to manipulate virtual objects, as one of the input modalities.

Our analysis revealed that MI systems have been studied from various perspectives. However, the combination of MI systems, AR, and IoT has rarely been selected as the survey subject. Norouzi et al. [22] conducted a review on systems using IoT and AR. However, they focused on trends in the modalities in the reviewed systems. Therefore, a detailed analysis of MI systems in conjunction with AR and IoT was not performed.

In this study, we used a taxonomy and terminology that was proposed by Augstein and Neumayr [29] in interaction modality analysis. We identified modality usage depending on the environment, and proposed patterns of system architecture for developing MI systems that utilize IoT and AR. Furthermore, we sought to identify the remaining research challenges from previous studies that may attract the attention of researchers in the field of HCI.

3. Methodology

We used the systematic literature review guidelines presented in Kitchenham and Charters [32] to conduct our systematic literature review. Sections 3.1 and 3.2 explained the literature review methodology applied in this study, which describe how the review was planned and conducted, respectively.

3.1. Planning the Review

Based on the related work that was discussed in Section 2.2, we determined the various uses of MI systems and were able to identify the shortage of information regarding MI systems that used IoT and AR. Consequently, we created research questions derived from the limitations identified in Section 2.2. We addressed the following five research questions:

3.1.1. RQ 1. Which Modalities Are Used in MI Systems with IoT and AR?

One of the limitations we identified is a lack of insight into the modalities used in MI systems with IoT and AR, as discussed in Section 2.2. Norouzi et al. [22] identified the trends of using modalities in systems that are based on IoT and AR. However, only 26 of the 60 studies constructed MI systems, and analyses of those MI systems were not provided. Therefore, it is difficult to clarify which modalities were used in the reviewed MI systems. Furthermore, Norouzi et al. [22] reviewed studies that were published between 2010 and 2018, whereas our study included studies from 2014 to 2020. We also used the taxonomy of input and output modalities that was proposed by Augstein and Neumayr [29] to avoid inconsistencies in the terminology used by different researchers. For example, "touch" modality has been referred to as both a "haptic-based" modality [20] and "touch" modality [5].

3.1.2. RQ 2. Which Modalities Are Used in MI Systems with IoT and AR?

Because MI systems require multiple modalities, a question that arises from the list of used modalities (RQ 1) is which modality combinations are used to provide interactions between agents and MI systems. In 2016, Hogan and Hornecker [19] summarized the usage ratios of output modality combinations from 154 use cases while excluding input modalities. Therefore, it is of interest to determine which modality combinations are employed as input by agents to MI systems. In addition, this study provides an update to the results of Hogan and Hornecker [19] by covering recent research.

3.1.3. RQ 3. Which Modalities Are Used in MI Systems with IoT and AR?

IoT and AR are used differently in MI systems, depending on the intended purpose. For example, Dodevska et al. [10] utilized AR as a data visualization tool by displaying data from IoT sensors that were attached to a centrifuge, whereas Sun et al. [12] created an AR interface that could be manipulated by an agent's hand gestures to control IoT devices. The agent was able to send commands to control an IoT device in both studies. In contrast, IoT devices were only used to send collected data to agents in the study by Seitz et al. [33]. There is more than one way to use AR and IoT in MI systems and analyzing different uses of IoT and AR can provide insights into how to use each technology in a given environment, as observed from several studies.

3.1.4. RQ 4. Which Modalities Are Used in MI systems with IoT and AR?

When IoT and AR are designed for different uses, the architecture of the underlying MI system should be adjusted to employ suitable input and output modalities. Analysis of system architectures for data integration [5,6] and analysis of models to design the interaction process between agents and MI systems have been performed, according to our analysis of related research [5]. Furthermore, Nizam et al. [30] presented four architecture frameworks that were used in MI systems with AR. However, an architecture analysis of MI systems consisting of IoT and AR is lacking. Considering these aspects, this research question can provide insights into the architectures upon which MI systems are built.

3.1.5. RQ 5. Which Modalities Are Used in MI Systems with IoT and AR?

In 2014, Turk [6] reported the remaining challenges in MI systems regarding architectures, usability, and security. Furthermore, although the article provides a useful introduction to the topic of MI systems, it does not qualify as a systematic literature review. Since 2014, technologies that are related to AR, IoT, and interaction modalities have improved, and we believe that it is time to identify new research challenges that have emerged. Thus, we seek to identify open research challenges that are related to MI systems that use IoT and AR. By answering this research question, we aim to identify the gaps between unsolved challenges and the results that were provided by existing studies.

We selected a number of keywords to find suitable studies to obtain answers to the aforementioned research questions. The selected keywords were terms relevant to MI systems, IoT, and AR, such as "multimodal interaction", "Internet of Things", "augmented reality", "mixed reality", "multimodality", "architecture", and "framework". We combined two or more keywords with Boolean logic for our search and also used abbreviations and short forms of the selected keywords (e.g., "IoT", "AR", "MR", "multimodal", "interaction", "augmented") to extend the coverage. When searching with keywords, we included singular and plural forms (e.g., "multimodalities", "interactions") to include as many studies as possible. Table 1 presents the keyword combinations that we used in our search.

We conducted a search with the aforementioned keywords in the following scientific libraries and journal databases that supported searches with Boolean operators:

1. ACM Digital Library
2. Google Scholar
3. IEEE Xplore
4. ScienceDirect
5. Scopus
6. SpringerLink
7. Taylor & Francis Online

8. Interacting with Computers
9. Journal on Multimodal User Interfaces

Table 1. Keyword combinations used in initial literature search.

ID	Keywords
1	multimodal interaction & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented) & (framework ‖ architecture)
2	multimodal interaction & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented)
3	multimodal interaction & (Internet of Things ‖ IoT) & (mixed reality ‖ MR) & (framework ‖ architecture)
4	multimodal interaction & (Internet of Things ‖ IoT) & (mixed reality ‖ MR)
5	multimodality & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented) & (framework ‖ architecture)
6	multimodality & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented)
7	multimodality & (Internet of Things ‖ IoT) & (mixed reality ‖ MR) & (framework ‖ architecture)
8	multimodality & (Internet of Things ‖ IoT) & (mixed reality ‖ MR)
9	multimodal & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented) & (framework ‖ architecture)
10	multimodal & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented)
11	multimodal & (Internet of Things ‖ IoT) & (mixed reality ‖ MR) & (framework ‖ architecture)
12	multimodal & (Internet of Things ‖ IoT) & (mixed reality ‖ MR)
13	interaction & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented) & (framework ‖ architecture)
14	interaction & (Internet of Things ‖ IoT) & (augmented reality ‖ AR ‖ augmented)
15	interaction & (Internet of Things ‖ IoT) & (mixed reality ‖ MR) & (framework ‖ architecture)
16	interaction & (Internet of Things ‖ IoT) & (mixed reality ‖ MR)

We selected two journal databases, because their studies were relevant to our research domain. Both of the journals supported an advanced search engine that could accept Boolean operators to search studies within their database. We were aware that there were other high quality journals; however, some journals were unable to search their databases with complex keyword combinations with Boolean operators; thus, we used digital libraries that covered these journals' databases, such as the International Journal of Human-Computer Interaction and the Human-Computer Interaction that wascovered by Taylor & Francis Online.

Our strategy for searching and for selecting studies consisted of three steps: (1) initial literature search, (2) first pass, and (3) second pass. In the initial literature search, we collected studies that were written in English and published between 2014 and 2020. Because of the low relevance between search terms and results after reaching a certain number of findings from the literature databases, we only downloaded the first 50 studies that were sorted by relevance to the search terms for each keyword combination.

In the first pass, the results of the initial literature search were assessed by their titles and abstracts in roder to determine whether to use them in the second pass. We skimmed the body of the paper if the title or abstract did not provide enough information to identify whether the study was about an MI system with IoT and AR. Studies that presented a system consisting of IoT, multiple modalities, and AR/MR were included in the list of studies to be analyzed, whereas duplicate and irrelevant studies were excluded. We only selected research papers that were published in a scientific peer-reviewed journal, conference, or magazine. Therefore, we excluded patents, books, and reports.

After the first pass, we thoroughly read the body of the remaining studies, including the discussion and conclusion, and then examined them based on the following criteria as part of the second pass:

1. We included studies that used MR instead of AR, because both technologies overlay virtual content on a real-world view.
2. We excluded studies that did not have a practical implementation, as we only considered systems whose feasibility was validated by implementation.
3. We excluded studies that used virtual reality (VR) instead of AR/MR, because VR separates users from the real-world, thus providing a different experience than AR and MR. As a result, the interaction methods that are based on dedicated modalities, such as motion controllers for a VR HMD, are not compatible with any modalities in current AR or MR environments.
4. We excluded studies that did not provide sufficient implementation details of their MI systems for proper analysis.

3.2. Conducting the Review

We conducted a systematic literature review that was followed by the aforementioned steps. The review was conducted in three steps that are described in the following subsections: (1) identification of studies, (2) selection of primary studies and quality assessment, and (3) data extraction and synthesis. Figure 3 illustrates the process of the literature search and filtering, displaying the number of findings in each step.

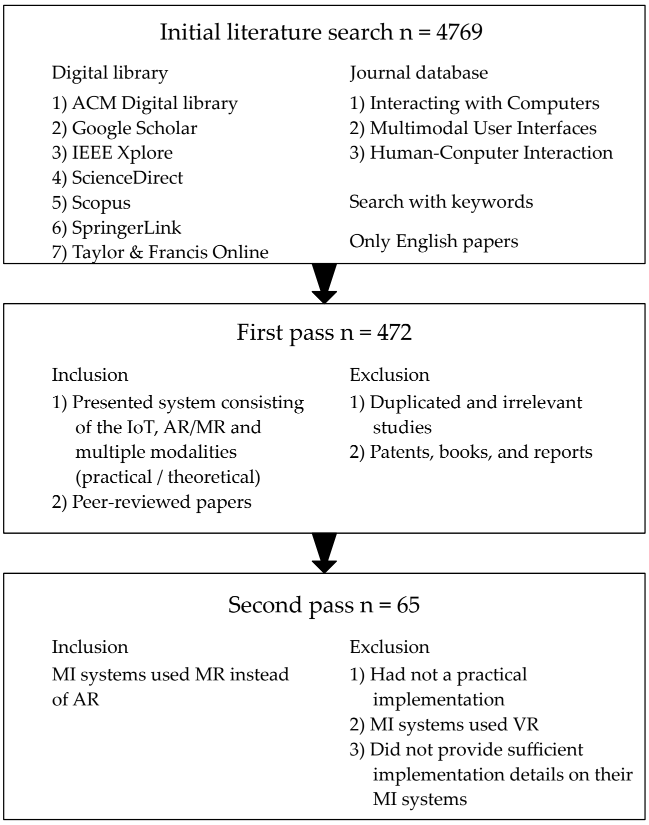

Initial literature search n = 4769
Digital library
1) ACM Digital library
2) Google Scholar
3) IEEE Xplore
4) ScienceDirect
5) Scopus
6) SpringerLink
7) Taylor & Francis Online
Journal database
1) Interacting with Computers
2) Multimodal User Interfaces
3) Human-Conputer Interaction
Search with keywords
Only English papers
First pass n = 472
Inclusion
1) Presented system consisting of the IoT, AR/MR and multiple modalities (practical / theoretical)
2) Peer-reviewed papers
Exclusion
1) Duplicated and irrelevant studies
2) Patents, books, and reports
Second pass n = 65
Inclusion
MI systems used MR instead of AR
Exclusion
1) Had not a practical implementation
2) MI systems used VR
3) Did not provide sufficient implementation details on their MI systems

Figure 3. The process of literature search and filtering with the number of findings in each step.

3.2.1. Identification of Studies: Initial Literature Search and First Pass

In this step, we conducted the initial literature search and first pass to identify relevant research papers from the scientific libraries and journal databases that are listed in Section 3.1 in April 2020. The initial literature search was conducted with the keywords defined in Section 3.1. We searched for studies that were written in English and were published between 2014 and 2020. We sorted the search results by relevance to the search terms by assigning them a grade of A, B, or C from most relevant to least relevant. When the title or abstract of a study explicitly stated a connection with an MI system that used AR/MR and IoT, we graded it with an A, whereas a grade of B was given to studies that frequently mentioned AR/MR and IoT in conjunction with the proposed MI system. A grade of C was given to the least relevant studies, which infrequently used the aforementioned words in their bodies and whose title and abstract provided no statements regarding the use of AR/MR and IoT in the MI system. We only reviewed the first 50 studies to exclude less relevant studies. Consequently, our initial literature search on the target databases resulted in 4769 studies.

We used inclusion and exclusion criteria for the next filtering step: the first pass. During the first pass, we assessed the titles and abstracts of the collected studies. We examined the titles and abstracts in order to identify whether the studies presented practical or theoretical MI systems that consisted of IoT and AR/MR. As mentioned earlier, studies that utilized VR were excluded due to the incompatibility between the input/output modalities in AR and VR. We aimed to cover as many MI studies as possible; thus, studies that employed AR and MR were included in the first pass. We also reviewed the bodies of the studies when we failed to extract sufficient information from their titles and abstracts. In order to ensure adequate quality of the selected studies, we only selected peer-reviewed scientific papers, while patents and non-peer-reviewed papers, such as books and reports, were excluded. Moreover, we narrowed the result set by excluding duplicate and irrelevant studies. Consequently, 472 out of 4769 studies from the literature search passed the first pass.

3.2.2. Selection of Primary Studies: Second Pass and Quality Assessment

During the second pass, we reduced the number of studies from the first pass by excluding studies that lacked a prototype of a proposed design of an application. The practical implementation of MI systems was important, as we were interested in modalities and architectures of feasible MI systems. We also included studies that used MR instead of AR to acquire as many adequate studies as possible, due to the similarity between the two technologies. During the second pass, studies that did not provide sufficient information about their MI systems were excluded. In addition, several VR studies that were not filtered out in the first pass were also excluded by reading the body of the papers. A total of 407 studies were excluded in the second pass; thus, 65 studies remained as the final selection.

Each study was evaluated based on four quality criteria listed in Table 2 to validate the quality of the studies selected in the second pass. We assessed each study on a dichotomous scale (“yes” or “no”) and only excluded a study only when it did not satisfy the first and second criteria. The first criterion verified whether the proposed MI systems actually used two or more different modalities in input or output channels. We used the taxonomy of modalities proposed by Augstein and Neumayr [29] to identify the number of modalities used in the proposed MI systems because our study focused on MI systems. The second criterion ensured that the selected studies were relevant to our research questions (RQ 1, RQ 2, RQ 3, RQ 4). While using this criterion, any uninformative studies that passed the second pass were identified and excluded. The third criterion was a basis for identifying research challenges that were established from user studies. In particular, when feedback from a number of different users is collected over a long period of time, the probability of obtaining valuable comments and ideas that can result in new research directions increases. The fourth criterion evaluated the clarity of open research challenges to identify valid challenges that were explicitly or implicitly presented by the reviewed studies. In relation to RQ 5 shown in Section 3.1, we attempted to identify open research challenges when reading the papers.

Table 2. Quality validation criteria.

ID	Criteria
1	Did the proposed MI system actually use two or more modalities in input or output channels?
2	Is there a detailed description of the proposed MI system regarding its input/output modalities and architectures?
3	Was an application created with the proposed MI system evaluated by users?
4	Was there a statement about open research challenges?

Consequently, 42 papers were excluded, as they had neither regarding modalities in input/output channels nor sufficient information about their MI systems. However, 23 out of 65 papers were determined to be of good quality, satisfying all four quality criteria.

3.2.3. Data Extraction and Synthesis

During the data extraction step, the 23 studies that were selected in the quality validation step were systematically analyzed in order to collect the necessary details for answering the research questions. We formed data extraction categories that covered the date of data extraction, title, publication year, authors, study aim (e.g., purpose of MI system), system design (e.g., input/output modalities, system architecture, interactivity between agent and interaction device, IoT devices, and other entities), experimental procedure (e.g., participants, experimental setup, user evaluation, data collection, and data analysis), and open research challenges (e.g., research hypothesis, findings, limitations, conclusions). When recording data during the review process, we copied the original text from the selected studies to minimize the impact of revision on the original information.

After data extraction, we performed data synthesis to create a list summarizing the results of data extraction in order to obtain answers to our research questions. When we performed data synthesis of the modalities, all extracted data regarding input/output modalities and modality combinations were organized in quantitative summaries. We used the taxonomy of modalities that was proposed by Augstein and Neumayr [29] to organize the extracted data in tabular form. The reason for using this taxonomy was that not all selected studies used the same terminology when describing the used modalities. When we discovered a novel or unidentified modality that was not present in the taxonomy, we categorized it based on the interaction perspective (e.g., user, technology), as suggested by Augstein and Neumayr [29]. For example, we classified robot movements as “vision”, because the results of a robot’s actions are conveyed as visual signals to the agent. For system architectures, descriptions of each architecture were analyzed to classify architectures with similar patterns according to their structure. During data synthesis for system architectures, we mainly focused on the methods of using IoT and AR in addition to input/output modalities. Furthermore, we produced quantitative summaries of system architecture patterns to identify commonly used system architectures. The identified open research challenges were tabulated and grouped together into higher-level research challenges to provide a comprehensive overview of open research challenges.

4. Results

This section presents the results of data analysis of the 23 reviewed studies. Specifically, Section 4.1 provides bibliometric information as well as statistics on the types of devices that were used in the reviewed studies, while Section 4.2 provides an analysis of the identified MI system architectures. Section 4.3 presents the analysis results of the interactions between an agent and other entities (i.e., interaction device, IoT device, other entity). Section 4.4 describes the types of input and output modalities and modality combinations that were identified in the reviewed studies, and Section 4.5 describes each open challenge discussed in the reviewed studies. Finally, Section 4.6 provides a summary of the reviewed studies.

4.1. Statistics

4.1.1. Database Distribution

Table 3 displays the distribution of the reviewed studies across the databases that were used in our literature search. Seven studies were found on Google Scholar, which was the highest number of studies for a single database. The IEEE Xplore database provided the second highest number of studies with six results, followed by SpringerLink with four studies, and two different databases, ACM Digital Library and ScienceDirect, with three studies each. We did not select any studies from two journal databases due to their failure to be approved in the second pass and quality validation step, although nearly 30 studies were obtained from each database in the first pass.

Table 3. Distribution of reviewed studies across scientific literature databases.

Databases	Number of Studies
ACM Digital Library	3
Google Scholar	7
IEEE Xplore	6
ScienceDirect	3
Scopus	0
SpringerLink	4
Taylor & Francis Online	0
Interacting with Computers	0
Journal on Multimodal User Interfaces	0

4.1.2. Yearly Distribution

We also determined the number of studies published per year to identify any trends. We discovered that the number of published studies gradually increased, with 2019 featuring the highest number of publications thus far.

4.1.3. Output Device Distribution

From the extracted data, we analyzed the distribution of devices that were used to present the output (e.g., AR and/or sensor data from IoT) to agents in the MI systems that were proposed in the reviewed studies. The number of studies in Table 4 is larger than the number of reviewed studies in Table 3 because some of the studies used multiple devices to present the output. For example, the MI system that was proposed by Sahinel et al. [34] presented data from IoT devices on a wearable application, such as a smartwatch. Meanwhile, AR was presented through a mobile device (e.g., tablet PC) to be used for indoor navigation and to control a machine that could be identified by using object recognition. Thus, a smartwatch and handheld mobile device were used within one MI system, and the devices are separately counted in Table 4. Overall, we observed that the reviewed studies mostly employed handheld mobile devices, such as a smartphone or tablet PC. An HMD was the second most used output device for MI systems in the reviewed studies.

Table 4. Output device distribution of proposed multimodal interaction systems from reviewed studies.

Output Devices	Number of Studies
Handheld mobile device (e.g., smartphone, tablet PC)	16
HMD (e.g., HoloLens, wearable glasses)	5
Smartwatch	1
Monitor (e.g., PC, TV)	3
Projector	1

4.2. MI System Architectures

We identified four patterns consisting of an agent, an interaction device (device that presents AR content), a server, an IoT device (including IoT sensors), and another entity (including real-world objects and environments) based on our analysis of the proposed MI systems' architectures. The patterns were established based on the intended purpose of AR and the type of identifier used for the AR visualization. The identifier could be either an AR marker that was a printed image, a physical object that could be identified using computer vision techniques, or an environment that presented spatial data information. Figure 4 illustrates the four identified MI system architecture patterns.

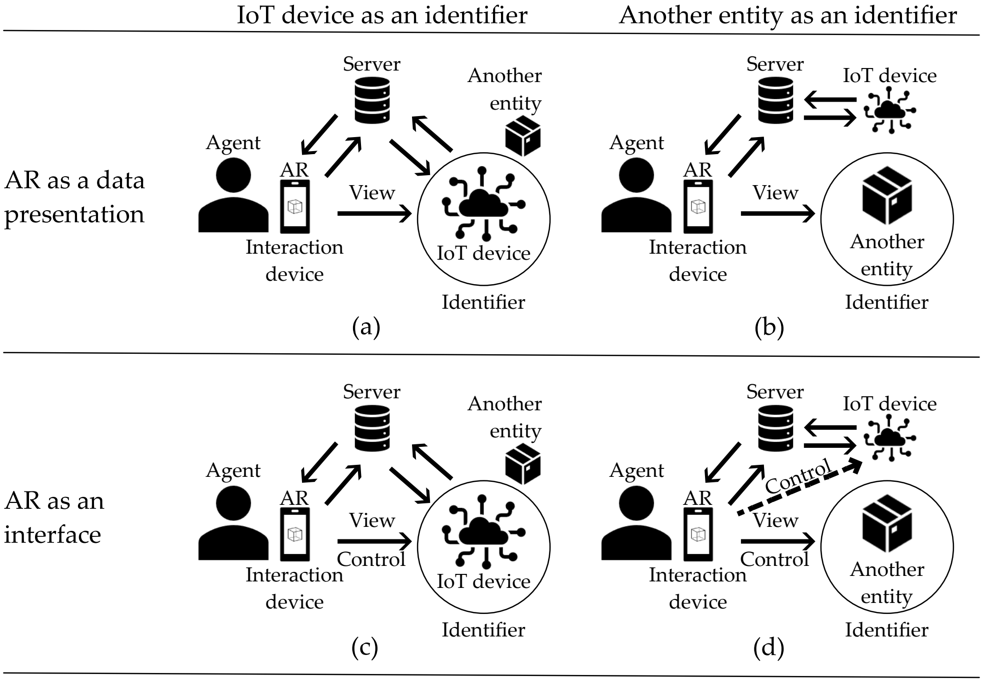


Figure 4. Four identified patterns of multimodal interaction system architecture from 23 reviewed studies. Each identified pattern is categorized by two bases: (1) whether an agent is able to control the IoT device through AR (**c**,**d**) or not (**a**,**b**) and (2) whether an IoT device (**a**,**c**) or another entity (**b**,**d**) is used as an identifier for AR.

4.2.1. AR for Data Presentation

Data presentation using AR is one of the applications of AR in MI systems. When an interaction device recognizes an identifier by looking at an IoT device (Figure 4a) or another entity (Figure 4b), the MI system sends the data collected from the IoT device to the interaction device. Data that are sent by the IoT device are presented by two-dimensional (2D) or three-dimensional (3D) graphical AR content [35,36] and are used for signaling a sound cue [37,38]. The data displayed on an interaction device is changed when an agent manipulates either an IoT device (Figure 4a) or another entity (Figure 4b).

As an example of the pattern that is illustrated in Figure 4a, Leppanen et al. [35] proposed a mobile AR application that can identify a coffee maker using an AR marker that is attached to it. The IoT device acquires the status of the coffee maker and sends data to the mobile AR application. When the agent manipulates the coffee maker, the state of the coffee maker in the mobile AR application is updated.

Pokric et al. [39] developed a mobile educational AR application that is an example of Figure 4b. The application requests the agent to estimate the temperature and CO_2 level of the location, where an IoT device is installed. When the application is started, a 3D character is placed on an AR marker or a plane that is recognized by the smartphone based on clusters of feature points in the real-world. The 3D character indicates the temperature and CO_2 levels by changing its clothes. The agent is asked to guess the current temperature and CO_2 level. The closer the guessed value is to the actual value that is sent by the IoT device, the more points are awarded.

4.2.2. AR as an Interface

As an interface, AR is used not only to visualize data for an agent, but also to allow an agent to provide input to an MI system. When AR is used as an interface, the agent can manipulate one or more IoT devices by interacting through the AR interface on an interaction device, such as a smartphone, tablet PC, or HMD (Figure 4c). Likewise, an agent can control IoT devices (as indicated by a dotted arrow in Figure 4d) while watching AR content from another entity (Figure 4d). The difference between the two patterns is what is used an identifier: the pattern that is depicted in Figure 4c uses an IoT device as an identifier, whereas the pattern depicted in Figure 4d uses another entity as an identifier.

The MI system that was developed by Dodevska et al. [10] is a case that employs the architecture displayed in Figure 4c. The AR mobile application connected to their MI system can recognize an AR marker placed on a miniature centrifuge. By scanning the AR marker, the agent can monitor the data that are sent by sensors installed on the motor of the centrifuge, and can control the motor by touching a button on the mobile screen.

As an example of the pattern that is depicted in Figure 4d, Mylonas et al. [40] developed a companion application used in educational lab kit activities that was tested in elementary, middle, and high school. The companion application helps students perform physical activities in an AR environment that require assembling circuits from a printed floor plan. The printed floor plan contains AR markers in order to visualize AR content on it. Some of the AR content is data collected by IoT devices that are installed within the school building, such as the temperature, humidity, and energy consumption rate. To receive the data, the agent must scan an AR marker that is attached to IoT devices using the companion application. Once the IoT devices are scanned, the agent can send commands to them to control light-emitting diode (LED) light.

4.2.3. Markerless (Without Identifier) AR

Variations in the patterns that are displayed in Figure 4 can be achieved by omitting the AR marker, which is a type of identifier. Our analysis of previous MI system architectures suggests that markerless (i.e., without an identifier) tracking approaches have been used instead of AR markers to identify an IoT device or another entity. In this subsection, we discuss one variation of each pattern that is displayed in Figure 4.

The MI system that was proposed by Seitz et al. [33] is a variation of Figure 4a that uses another AR target tracking technique instead of AR markers. Seitz et al. [33] developed an AR mobile application for workers in a factory to monitor the temperature of machines. IoT devices are attached to each machine to stream the temperature data in real-time. The AR mobile application uses object recognition instead of AR markers to identify the IoT devices. Furthermore, in this MI system, all of the IoT devices have the same appearance; thus, LED colors on IoT devices are used to distinguish devices by the AR mobile application.

As a variation of the architecture presented in Figure 4b, Simões et al. [41] developed a system for a work space that can recognize an electric cabinet placed on a specific location on a table. When the electric cabinet is located within the camera view, the system presents the assembly instructions for the electric cabinet. The assembly instructions are displayed on a Microsoft HoloLens device using 3D models that are drawn over the electric cabinet. Cameras with a depth sensor installed in the work space are used to identify the position of the agent's hands, and the built-in HoloLens camera is used to detect the position of virtual content (e.g., virtual buttons, virtual interfaces) to allow agents to interact with the system while using their hands. Agents can follow the assembly instructions in the real-world and they can navigate the instructions and other information by manipulating the virtual content with their hands. In the work space that was described by Simões et al. [41], RGB-D cameras are used in combination with the HoloLens camera to increase the accuracy of hand gesture detection. There are two projectors that visualize information regarding the assembly instructions and temperature of the work space on the table.

The MI system that was proposed by Sahinel et al. [34] utilizes both an AR marker and object recognition. In their system, an application on the agent's mobile device uses an

AR marker to recognize an environment to initiate indoor navigation (Figure 4b), and uses object recognition to identify a machine (variation of Figure 4c). Such an MI system can be classified as both Figure 4b and Figure 4c, due to its ability to monitor and control the recognized machine by object recognition.

Cho et al. [42] developed an MI system that is a variation of Figure 4d architecture. Their system employs object recognition to recognize miniature building structures (i.e., another entity) for visualizing an AR building structure on them. When an AR building structure is displayed on a mobile device, information on energy consumption that is sent by IoT devices installed within the building and rooms is presented. Based on this information, the agent can control the light power in a building or room by touching the menu in the AR interface. The commands that were issued by the agent through the AR interface affect lights that are installed not only in the AR building structure, but also in the real-world miniature building structures.

4.2.4. Usage Ratio

Table 5 summarizes the usage ratios of the four system architectures that are identified in Figure 4. The architecture pattern presented in Figure 4c is used in 35% of the reviewed studies (i.e., eight studies), thus making it the most commonly used pattern. The pattern shown in Figure 4a is the second most used pattern, utilized in seven studies, and the depicted pattern in Figure 4b is the third most used pattern, utilized in five studies. The pattern depicted in Figure 4d was used in two studies; however, one of the studies [40] also utilized the pattern in Figure 4c in the same MI system. Finally, we found another study [34] that employed two MI system architectures (i.e., Figure 4b,c) within one system.

Table 5. Summary of usage ratios of identified multimodal interaction system architecture patterns.

Pattern	Studies	% of Total
Figure 4a	[33,35,36,43–46]	31
Figure 4b	[37–39,41,47]	22
Figure 4c	[10–12,48–52]	35
Figure 4d	[42]	4
Figure 4c,d	[40]	4
Figure 4b,c	[34]	4

4.3. Interaction with an Agent

Figure 5 summarizes the possible interactions between an agent and other objects (i.e., interaction device, IoT device, and other entity). Interactions that appeared frequently in multiple studies are listed in blue boxes, while interactions that appeared rarely are listed in red boxes. In the following subsections, we provide detailed descriptions of the interactions along with examples, and Figure 6 illustrates the interactions that are present in each reviewed study. In this figure, each entity that has interactions between an agent is categorized using different colors. Green cells indicate interactions between an interaction device and agent, yellowish cells indicate interactions between an IoT device and agent, and blue cells indicate interactions between another entity and agent. Darker cells represent the interactions that are used in the respective study.

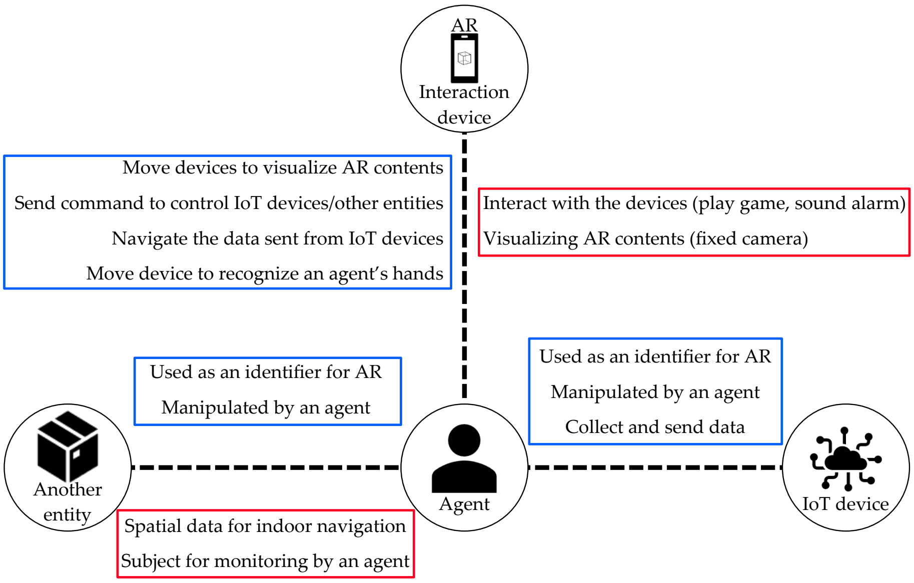
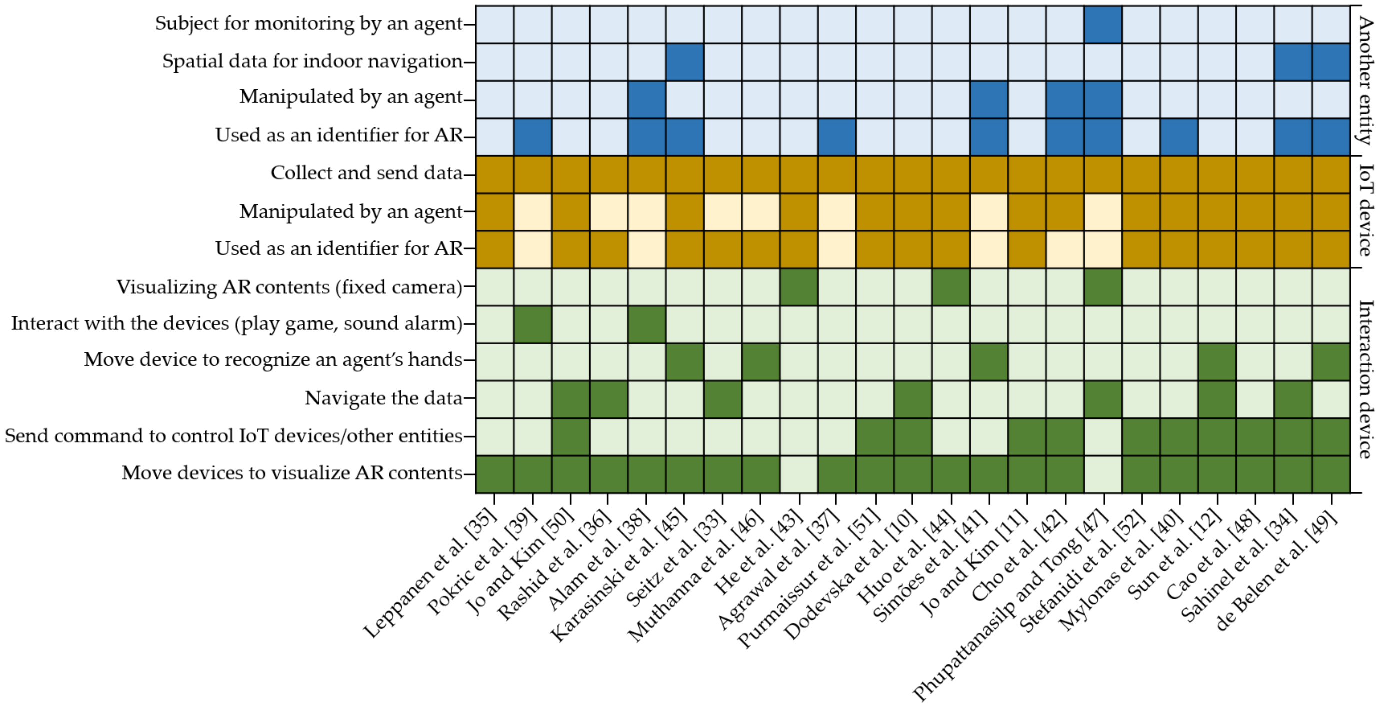

AR
Interaction device
Move devices to visualize AR contents
Send command to control IoT devices/other entities
Navigate the data sent from IoT devices
Move device to recognize an agent's hands
Interact with the devices (play game, sound alarm)
Visualizing AR contents (fixed camera)
Used as an identifier for AR
Manipulated by an agent
Used as an identifier for AR
Manipulated by an agent
Collect and send data
Another entity
Agent
IoT device
Spatial data for indoor navigation
Subject for monitoring by an agent

Figure 5. Interaction between an agent and an interaction device, IoT device, and another entity.



		Leppanen et al. [35]	Pokric et al. [39]	Jo and Kim [50]	Rashid et al. [36]	Alam et al. [38]	Karasinski et al. [45]	Seitz et al. [33]	Muthanna et al. [46]	He et al. [43]	Agrawal et al. [37]	Purmaissur et al. [51]	Dodevska et al. [10]	Huo et al. [44]	Simões et al. [41]	Jo and Kim [11]	Cho et al. [42]	Phupattanasilp and Tong [47]	Stefanidi et al. [52]	Mylonas et al. [40]	Sun et al. [12]	Cao et al. [48]	Sahinel et al. [34]	de Belen et al. [49]
Another entity	Subject for monitoring by an agent	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	■	□	□	□	□	□	□
	Spatial data for indoor navigation	□	□	□	□	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	■	■
	Manipulated by an agent	□	□	□	□	■	□	□	□	□	□	□	□	□	■	□	■	■	□	□	□	□	□	□
	Used as an identifier for AR	□	■	□	□	■	■	□	□	□	■	□	□	□	■	□	■	■	□	■	□	□	■	■
IoT device	Collect and send data	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Manipulated by an agent	■	□	■	□	□	■	□	□	■	□	■	■	■	□	■	■	□	■	■	■	■	■	■
	Used as an identifier for AR	■	□	■	■	□	■	■	■	■	□	■	■	■	□	■	□	□	■	■	■	■	■	■
Interaction device	Visualizing AR contents (fixed camera)	□	□	□	□	□	□	□	□	■	□	□	□	■	□	□	□	■	□	□	□	□	□	□
	Interact with the devices (play game, sound alarm)	□	■	□	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Move device to recognize an agent's hands	□	□	□	□	□	■	□	■	□	□	□	□	□	■	□	□	□	□	□	■	□	□	■
	Navigate the data	□	□	■	■	□	□	■	□	□	□	□	■	□	□	□	□	■	□	□	■	□	■	□
	Send command to control IoT devices/other entities	□	□	■	□	□	□	□	□	□	□	■	■	□	□	■	■	□	■	■	■	■	■	■
	Move devices to visualize AR contents	■	■	■	■	■	■	■	■	□	■	■	■	■	■	■	■	□	■	■	■	■	■	■

Figure 6. Interaction usage status for each reviewed study.

4.3.1. Interaction Devices

An interaction device is a platform that visualizes AR content to the agent and it may also receive input from the agent. The interaction device can have many forms, such as a mobile device (e.g., smartphone, tablet PC) [10,11,33–37,39,40,42,44,46,48,50–52], HMD [12,38,41,45,49], PC/TV monitor [41,43,47], and projector [41].

The most commonly implemented interactions between an agent and interaction device are as follows: (1) moving the interaction device to see AR content, (2) sending commands to control IoT devices or other entities, (3) navigating the data sent from IoT devices through a user interface on the interaction device, and (4) moving the interaction device to recognize the agent's hands. As an example of the first and second interactions, a smartphone in the MI system that was proposed by Jo and Kim [11] can identify a lamp in a shop by analyzing the smartphone camera's feed. The power of the lamp can be controlled by touching a button on the screen of a smartphone. An IoT device is attached to the lamp that sends the power status of the lamp to the smartphone while allowing the agent to control the power of the lamp. As an example of the third interaction, Seitz et al. [33] developed a mobile application that visualizes the temperature data obtained from IoT devices. The temperature data are represented by emojis that are based on the temperature level. The emojis are placed on the IoT devices, while graphs of device data that are correlated with each emoji are displayed on the mobile screen. The agent can obtain detailed information regarding the IoT devices that the application recognizes by touching the graphs. As an example of fourth interaction, four out of five studies [12,41,45,49] that used an HMD as an interaction device utilized it to identify the hand gestures of an agent. The exception was a study by Muthanna et al. [46], who used a mobile device instead of an HMD to detect the agent's hand positions to display the data that were collected from IoT devices on the interaction device when the agent's hands were placed at a specific location.

Although the aforementioned interactions are commonly used in MI systems, there also exist interactions that are used for specific purposes and in dedicated environments. For example, an agent can play a simple guessing game on a smartphone in the MI system proposed by Pokric et al. [39]. Non-AR 2D user interfaces are used to receive answers from the agent, while a 3D AR character is used to represent the current temperature and CO_2 levels while using dedicated outfits. In the MI system that was proposed by Alam et al. [38], the interaction device is an HMD that produces sound alarms when the sensor data (e.g., temperature, humidity, CO_2, O_2, heart rate) that were collected by IoT devices exceed certain levels. The purpose of the sound alarm is to alert the agent about danger. In some MI systems, interaction devices are connected to a stationary camera to monitor specific objects or stream a certain viewpoint. The MI system that was proposed by Phupattanasilp and Tong [47] utilizes a PC monitor to observe crops that were planted on a farm. In their MI system, AR content is displayed near the position at which the MI system recognizes crops using regions of interests that were detected by a computer vision technique. In the MI system that was proposed by He et al. [43], a television is the interaction device that is connected to a stationary camera to stream the space in which the agent moves smart objects on a smartboard. The location of AR content is determined by movement of the smart objects detected by light density resistors under the smartboard.

4.3.2. IoT Devices

There are three commonly used interactions between an agent and an IoT device: (1) IoT devices are used as identifiers for AR when printed AR markers are not used, (2) IoT devices are manipulated by the agent, and (3) IoT devices collect and send data to a server in otder to facilitate further interactions. As an example of the first and second interactions, the interaction device in the MI system proposed by Sun et al. [12] is an HMD that can identify IoT devices (i.e., smart speaker and smart light bulb) using object detection. The agent can use predefined hand gestures to trigger a number of actions on the detected IoT devices, such as playing a song, increasing the volume, selecting music, changing the color of light, and managing the light bulb power. Whereas, the MI system that was proposed by Sun et al. [12] enables remote control of IoT devices by the agent, the MI system proposed by Huo et al. [44] involves the second interaction in a way that requires the agent to have physical contact with the IoT devices. The positions of the IoT devices are spatially registered in the MI system, and a visual notification is provided when the IoT devices are moved from the original location. This visual notification conveys information on the current location of the moved IoT devices using arrows. As an example of the third interaction, Karasinski et al. [45] used Microsoft HoloLens to predefine the locations of AR content that remained in the same position, regardless of the field of view of the agent. Therefore, the AR marker is not used; instead, pre-positioned AR arrows are used to guide the agent to the destination as an application of indoor navigation. Indoor navigation is initiated when the agent looks for a tool that is placed in another room. When the agent reaches the location of the tool he or she is looking for and picks it up, the MI system determines whether the target tool is collected by the agent using IoT devices that are installed in the target tool and on the agent's wrist. If the MI system determines that the agent successfully obtained the tool, then the return path or the path to the next tool is presented on HoloLens.

4.3.3. Another Entity

Another entity can be either an object or environment that can interact with the agent. Two types of interactions are identified in a number of studies, while two other types of interactions are identified in a few studies.

The first type of commonly used interaction is that in which another entity is used as an identifier for AR. This can be implemented with either a printed AR marker [34,37,39,40], image/object recognition [38,41,42,47,49], or spatial data [45]. Another type of interaction is the manipulation of another entity by the agent. The agent can interact with the MI system while having physical contact with other entities (e.g., electric box [38] abd electrical cabinet [41]). Furthermore, the agent can control another entity through an interaction device. For example, the MI system that was proposed by Cho et al. [42] allows the agent to control lights in a miniature building, which are mapped to real lights in rooms of a building in which IoT devices are installed.

In our literature search, we found 10 studies involving interactions between agents and other entities. Of these studies, only three used spatial data of a certain place to provide an indoor navigation feature [34,45,49]. Moreover, only one study used other entities (i.e., crops in soil) as either objects to monitor conditions (i.e., the location of crops, soil moisture, temperature, water level, nutrients, and luminance) or objects that can be managed by the agent with physical contact [47].

4.4. Modalities

In this subsection, we summarize different types of input and output modalities and their combinations as used in the reviewed studies.

4.4.1. Input

Table 6 lists the input modalities that were used in the reviewed studies, and each study is classified according to the input sources that are listed within each input modality. The "Total" column in Table 6 indicates the number of studies that used each modality among the 23 reviewed studies, while the rightmost column presents the percentage of studies listed in the "Total" column that used each of the input sources.

"Vision" is a modality that every reviewed study utilized in their MI systems. There are many different input sources for using vision as an input modality. By analyzing each study, we identified four common input sources. First, 21 of 23 studies utilized vision as an input modality by using the movement of the camera view for AR. In order to visualize AR content, an agent moves the camera of an interaction device to locate an identifier within the camera view. The identifier can be either printed images used as AR markers [10,34–36,39,40,48], physical objects that can be recognized using object/image recognition [11,12,33,34,37,38,41,42,44,46,49–52], or environments for using spatial data [39,45]. The second identified input source is the camera view for capturing an agent's hands. We found that four studies used a camera feed to detect hand gestures [12,41,45,49], while one study used a camera feed to track hand position [46]. The third input source is the agents' gaze movement as they move their heads. When an MI systems employs an HMD (e.g., HoloLens) as an interaction device, in some studies [41,45,49], the agents use the gaze position, which is the center point of their view moved along their heads, to interact with AR.

Table 6. Summary of input modalities.

Modalities	Input Sources	Studies	Total	% of Total
Vision	Tracking of an AR content identifier from the camera view	[10–12,33–42,44–46,48–52]	23	91
	Capturing of an agent's hands from the camera view	[12,41,45,46,49]		22
	Gaze interaction by head movement	[41,45,49]		13
	Position of real-world objects	[43,44,47]		13
Touch: Tactility	Touching a touchscreen	[10,11,33,34,36,39,40,42,46,48,50–52]	15	87
	Mouse click	[35]		7
	Physical buttons on a machine for manipulation	[47]		7
Kinesthetics: Proprioception	Hand gestures	[12,41,45,49]	7	57
	Position of hands	[46]		14
	Position of an interaction device held by an agent	[44,48]		29
Kinesthetics: Kinematics	Detection of position change of a physical object	[43,45]	2	100
Audition	Voice command	[45,49]	2	100

The fourth input source is the position of real-world objects that is used to determine the position of AR content. For example, the MI system that was proposed by He et al. [43] uses a smartboard that can detect smart objects using light density resistors installed under the smartboard. Based on the location that was detected by the light density resistors, the positions of the smart object on the smartboard are illustrated by AR on a TV screen that is connected to a stationary camera. This camera observes the smartboard, and the agent can watch the stream that contains AR content in real-time. The MI system that was proposed by Huo et al. [44] also uses a camera feed to track the position of real-world objects when they are moved from their original positions. In addition, the MI system that was proposed by Phupattanasilp and Tong [47] uses the camera view to identify the coordinates of real-world objects (i.e., crops) on a PC screen to position the AR interface near each identified real-world object.

"Touch" is the second most used modality that we identified in 15 studies. In the taxonomy of modalities that was proposed by Augstein and Neumayr [29], tactility refers to a device's ability to sense the physical contact of an agent. A touchscreen, mouse, and machine with buttons are examples of devices that can support interaction through tactility. We found that 13 of 15 studies used a touchscreen, whereas mouse clicks and buttons on machines were used in one study.

The next modality category is "kinesthetics", which we found in eight studies. Kinesthetics can be divided into two different modalities—"proprioception" and "kinematics"—based on the taxonomy of modalities that was proposed by Augstein and Neumayr [29]. Proprioception refers to a device's ability to detect the position and movement of a body part, while kinematics refers to a device's ability to sense the movement of a physical object based on acceleration. From an agent's perspective, any activity made with his or her hands or body that is not sensed by acceleration is related to proprioception as part of the interaction, whereas any movement sensed by acceleration sensors is related to kinematics as part of the interaction. As examples of proprioception, hand gestures [12,41,45,49] and hand positions [46] can be utilized as input modalities, and the position of an agent, holding an interaction device can be used as another input modality to navigate in the AR world consisting of multiple IoT devices. As an example of kinematics, the movement of physical objects caused by an agent can be used as an input modality [43,45]. Furthermore, one of the reviewed studies that used kinematics also used proprioception in its MI system [45].

The least used input modality is "audition", which refers to sound-based interaction. In the reviewed studies, the input source of this modality was voice commands. HoloLens, which has a built-in microphone to capture the voice commands of an agent, was used as an interaction device in two studies [45,49] to allow for the agent to interact with the AR interface.

4.4.2. Output

We also identified two different output modalities that were used in the reviewed studies. Table 7 presents the output modalities with the number of output sources of the modality used in the reviewed studies. The table also lists the total number of studies that used each output modality and the ratio of each output source.

Table 7. Summary of output modalities.

Modalities	Output Sources	Studies	Total	% of Total
Vision	Presentation of 2D/3D graphics	[10–12,33–52]	23	100
	Presentation of controls to manipulate IoT devices	[10–12,34,40,42,48–52]		48
Audition	Sound cues for notification	[38,49]	4	50
	Sound effects of AR contents	[37]		25
	Sound from an external device	[12]		25

"Vision" is the first identified output modality that was used in every study that we analyzed. We identified two unique output sources for the vision modality: (1) the presentation of 2D or 3D graphics and (2) presentation of controls to manipulate IoT devices. The first output source (presentation of 2D or 3D graphics) could be found in all studies, as all of the analyzed MI systems that were used an interaction device that displayed graphical content including AR. However, the second output source could be found in 11 out of 23 studies that used vision as an output modality. When an MI system allows for an agent to manipulate IoT devices through a touchscreen [10,11,34,40,42,48–52] or by hand gestures [12], the result of IoT device manipulation is visually presented in the real-world to the agent. Thus, we categorized 11 studies that allowed the agent to control IoT devices into the vision modality category.

"Audition" is another output modality that was used in four of the reviewed studies. This output modality includes three different output sources. Two studies used sound cues to inform the agent regarding a specific event. For example, the MI system proposed by Alam et al. [38] notifies the agent by a sound alarm about an abnormal environment state based on values, such as temperature and CO_2, collected from IoT devices. The MI system that was proposed by de Belen et al. [49] plays a sound cue to notify the agent that he or she is approaching IoT devices. Another output source is the sound effects of AR content. This was exemplified by Agrawal et al. [37], who developed a mobile application that displays an animated 3D model of the human heart. This application also plays the sound of a heartbeat, which is synchronized with the real heart rate of the agent. Moreover, the beating animation of the 3D heart model is also synchronized with the real heart rate. The final output source that we identified in the audition modality category was sound from an external device. The aforementioned three studies that utilized audition as an output modality used a built-in audio device to deliver auditory stimulation to the agent. However, one study used an external device to provide sound signals to the agent. Sun et al. [12] developed an MR interface that an agent can use to control a smart speaker to select and play music. The agent can also increase and decrease the volume of the smart speaker.

4.4.3. Combinations

Table 8 summarizes the different modality combinations and the ratios of each to the total number of reviewed studies. We determined that 19 studies combined two different modalities as an input or output modality, while four studies combined three different modalities in their MI systems. Every combinationm except for Vision & Auditionm was for input modalities, and two of four studies that employed multiple output modalities (i.e., Vision & Audition) also used multiple input modalities. The most commonly used modality combination was Vision & Touch: Tactility for input, which was used in 11 studies. The second most used modality combination was Vision & Audition for output, which was used in four studies, and the third and fourth most combinations were Vision & Kinesthetics: Proprioception and Vision & Audition & Kinesthetics: Proprioception for input, which were used in three and two studies, respectively. Other modality combinations were only observed once in the reviewed studies.

Table 8. Summary of modality combinations.

Modality Combinations	Studies	% of Total
Vision & Touch: Tactility	[10,11,33–36,39,40,42,47,50–52]	57
Vision & Audition	[12,37,38,49]	17
Vision & Kinesthetics: Kinematics	[43]	4
Vision & Kinesthetics: Proprioception	[12,41,44]	13
Vision & Touch: Tactility & Kinesthetics: Proprioception	[46,48]	9
Vision & Audition & Kinesthetics: Proprioception	[49]	4
Vision & Audition & Kinesthetics: Kinematics, proprioception	[45]	4

4.5. Open Research Challenges

Through our analysis, we identified open research challenges that were mentioned in the reviewed studies. A total of 13 studies presented 19 research challenges, which we categorized based on the following keywords presented in Table 9: (1) technology, (2) standardization, (3) scalability, (4) multi-agent, and (5) multidisciplinary.

Table 9. Summary of open research challenges.

Keywords	Open Research Challenges	Studies
Technology	Narrow field of view in HoloLens	[45]
	Hardware performance problem regarding noise reduction	[37]
	High cost of technologies	[47]
	Obstacles for displaying non-visible identifiers	[33]
	Unreliable vision-based coordinate estimation in a large-scale environment	[47]
	Limited detection range for QR codes	[40]
	Requirement of specific light conditions for running AR properly	[40]
	Limited object detection range	[12]
	Low success rate of object detection on reflective surface	[12]
	Unstable vision-based feature tracking with insufficient features	[48]
Standardization	Standardization of MI system components	[50]
	Standardization of AR feature data structure and data exchange protocol	[11]
Scalability	Distributed system to reduce computational costs	[44]
	Management of diverse interfaces with different IoT devices	[44]
	Automation of IoT device registration process	[48]
Multi-agent	Individual preference for interaction devices	[36]
	Data synchronization in multi-user scenario	[44]
	Personalized UI and feedback based on agent's skill	[52]
Multidisciplinary	Understanding of various cultures for different types of agents	[33]
	Conducting user evaluations in real-world scenario	[43]

Seven studies reported research challenges regarding hardware [37,45,47] and technical problems [12,33,40,47,48], which we summarized under Technology in Table 9. Research challenges regarding hardware problems concern either an interaction device [45], IoT device [37], or both [47]. For example, Karasinski et al. [45] discovered that the narrow field of view of HoloLens makes it difficult for an agent to locate AR content placed at the edge of the HoloLens screen. For IoT devices, Agrawal et al. [37] reported a data quality problem for a heartbeat sensor due to ambient noise during data collection. Phupattanasilp and Tong [47] noted the impracticality of installing the proposed MI system in an actual site due to the price of the devices and energy consumption. Other studies faced technical challenges, especially regarding the limitations of vision-based identifier detection [12,33,40,47,48]. For example, Seitz et al. [33] reported a challenge in displaying AR content when identifiers were blocked by environmental obstacles. Because the identifier detection system was vision-based, Seitz et al. [33] stated that identifiers that are not detected by the interaction device's camera must be handled in different ways. Phupattanasilp and Tong [47] also noted a problem regarding a vision-based identifier detection system. In their proposed MI system, crops on a farm were used as identifiers, and a number of stationary cameras were used to estimate the coordinates of the crops, which were then displayed on a PC monitor. The accuracy of coordinate estimation was high; however, Phupattanasilp and Tong [47] noted a challenge involving solving the inaccuracy of coordinate estimation in a large-scale environment. In addition, Mylonas et al. [40] noted out that adequate light conditions were necessary to place AR content in the correct position, while Sun et al. [12] stated that the reflective surface of a target object caused a low success rate of object detection. Similar to Sun et al. [12], Cao et al. [48] mentioned the instability of vision-based tracking due to insufficient features of identifiers. Additionally, both Mylonas et al. [40] and Sun et al. [12] reported a short detection range for identifiers.

The research challenges in the standardization category aim to improve MI systems while using a standardized data and system structure. For example, when each IoT device contains information about the target object for object detection, the MI system can easily register new IoT devices as long as all IoT devices use the same data structure, data exchange protocol, and UI structure [11,50]. Furthermore, Jo and Kim [11] claimed that standardization is a way to achieve high scalability.

In a similar vein, Huo et al. [44] reported that reducing the computational costs using a distributed system is a way to enhance the scalability. Huo et al. [44] also reported the necessity for an MI system to handle various interfaces with different IoT devices. Furthermore, Cao et al. [48] claimed that MI systems require an automated IoT device registration process.

The next keyword, multi-agent, refers to research challenges focused on a scenario in which multiple agents use the MI system. Rashid et al. [36] found that the preference for an interaction device differed, depending on an agent's physical state. For example, when an agent had a disorder of the lower body, he/she preferred a smartphone rather than a tablet PC due to the size of the device, which became burdensome holding it to observe identifiers for AR content. Huo et al. [44] reported a data synchronization problem in the multi-agent scenario. In their MI system, the information of IoT devices could differ on each agent's interaction device due to a slow update rate. The update rate was designed for a single-agent scenario to reduce the computational costs. Therefore, Huo et al. [44] planned to use a distributed system to reduce the computational costs while supporting a fast update rate. Furthermore, Stefanidi et al. [52] identified a need for personalized UI and feedback based on an agent's skill. In order to provide a personalized experience, Stefanidi et al. [52] noted that context-awareness was necessary in an MI system.

The final keyword, multidisciplinary, refers to the necessity of collaboration with other academic disciplines. Seitz et al. [33] noted that knowledge of psychological aspects is necessary due to the different meaning of emojis, depending on the agent's culture. He et al. [43] also emphasized the necessity of a multidisciplinary approach by highlighting the limitations of their study results. They developed an MI system to evaluate the upper limb function of an agent using an IoT object on a smartboard that can sense the location of an IoT object. They conducted a user evaluation; however, the agents did not have upper limb disorders; therefore, the efficiency of their MI system was not fully evaluated. Consequently, they mentioned the need for actual patients to verify the effectiveness of their MI system. We categorized this research challenge as multidisciplinary, because it is infeasible to find real patients for a user evaluation without the help of an individual from the medical field.

4.6. Summary of Reviewed Studies

Table 10 summarizes information on the 23 reviewed studies, including the authors, publication year, input/output modalities, and output devices used in the reviewed MI systems; brief descriptions of the objectives of the studies; and patterns (PAT) of the MI system architecture, as illustrated in Figure 4. In this table, the "Output devices" column lists devices that are used to present data arriving from an MI system to an agent, while an interaction device refers to a device that can present AR content to an agent. Therefore, it is important to note that the smartwatch used in the MI system that was proposed by Sahinel et al. [34] is not an interaction device, but an output device; the researchers used a smartwatch to deliver data from IoT devices in textual form.

Table 10. Summary of reviewed studies.

Authors	Input Modalities	Output Modalities	Output Devices	Objectives	PAT
[35]	Vision (camera view for AR marker detection), Touch (tactility: machine control)	Vision (2D/3D graphics)	Tablet PC	AR application for monitoring the power state of a coffee maker.	Figure 4a
[39]	Vision (camera view for AR marker and plane detection), Touch (tactility: touchscreen)	Vision (2D/3D graphics)	Smartphone	Serious game to raise awareness of air pollution problems.	Figure 4b
[50]	Vision (camera view for object recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Smartphone	A system that presents the status and enables the manipulation of real-world objects that are recognized/tracked by data from IoT devices.	Figure 4c
[36]	Vision (camera view for AR marker detection), Touch (tactility: touchscreen)	Vision (2D/3D graphics)	Mobile device (smartphone, tablet PC)	Allows the agent to find items from a smart shelf using RFID data.	Figure 4a
[38]	Vision (camera view for image recognition)	Vision (2D/3D graphics), Audition (audio effect [alarm])	HMD (wearable glasses)	Task procedure visualization with a safety warning system (for sending alarms to agents) that can sense problems using data from IoT devices that are held by the agent.	Figure 4b
[45]	Vision (camera view for hand gestures, head movement [gaze]), Audition (voice), Kinesthetics (kinematics: measuring the distance between an agent and a selected tool while detecting the pick-up motion; proprioception: hand gesture)	Vision (2D/3D graphics)	HMD (HoloLens)	Indoor navigation to guide the agent to a specific room where the selected tool is placed. The distance between the agent and tool is measured by IoT devices to detect the pick-up motion.	Figure 4a
[33]	Vision (camera view for color and object recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics)	Smartphone	Visualization of IoT device data using emojis.	Figure 4a

Table 10. *Cont.*

Authors	Input Modalities	Output Modalities	Output Devices	Objectives	PAT
[46]	Vision (camera view for image recognition), Touch (tactility: touchscreen), Kinesthetics (proprioception: hand position [virtual button])	Vision (2D/3D graphics)	Mobile device (smartphone, tablet PC)	A prototype AR application that can visualize either the information of an art object (*The Starry Night* painting by Vincent Van Gogh) for museum visitors, or data of IoT devices in the museum for administrators.	Figure 4a
[43]	Vision (position of object by light density resistor), Kinesthetics (kinematics: movement of smart object by hand)	Vision (2D/3D graphics)	TV screen	Testing the the agent's upper limb disorder by moving a smart object on a smartboard. AR content is presented on a TV screen, while data from IoT devices are presented on the smartboard.	Figure 4a
[37]	Vision (camera view for image recognition)	Vision (2D/3D graphics), Audition (audio effect)	Smartphone	Visualizing a 3D model of a heart on the AR marker. The 3D heart model is a heartbeat animation synchronized with the agent's heart rate measured by an IoT device on the agent's fingertip.	Figure 4b
[51]	Vision (camera view for image recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Smartphone	Presenting air quality monitoring data as AR content on an IoT device and allowing the agent to control the power of the IoT device.	Figure 4c
[10]	Vision (camera view for AR marker detection), Touch (tactility: touchscreen)	Vision (2D/3D graphics), device control [smart object])	Mobile device (smartphone, tablet PC)	Allowing the agent to monitor and control a miniature centrifuge (IoT device).	Figure 4c
[44]	Vision (camera view for position of object and spatial mapping), Kinesthetics (proprioception: distance between device and IoT device)	Vision (2D/3D graphics)	Smartphone	Navigate a spatially registered room with an AR user interface that can present the position and status of eight IoT devices in the specific room.	Figure 4a
[41]	Vision (camera view for hand gesture and physical object detection, head movement [gaze]), Kinesthetics (proprioception: hand gesture)	Vision (2D/3D graphics)	HMD (HoloLens), projectors, PC screen	Presenting assembling process instructions of an electrical cabinet while cameras with depth sensors and projectors are used to track the agent's hand movements to recognize input gestures.	Figure 4b
[11]	Vision (camera view for object recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Smartphone	Manipulate merchandise in a shop; recognize and track objects using data from IoT devices.	Figure 4c

Table 10. *Cont.*

Authors	Input Modalities	Output Modalities	Output Devices	Objectives	PAT
[42]	Vision (camera view for object recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Mobile device (smartphone, tablet PC)	Energy management by AR presentation of data collected from IoT devices installed in rooms of a building.	Figure 4d
[47]	Vision (position of crops in camera view), Touch (tactility: mouse)	Vision (2D/3D graphics)	PC screen	Monitor the status of crops using data from IoT devices.	Figure 4b
[52]	Vision (camera view for object recognition), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Mobile device (smartphone, tablet PC)	The agent uses an AR interface to create rules on how IoT devices interact with each other in a smart home.	Figure 4c
[40]	Vision (camera view for AR marker detection), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Mobile device (smartphone, tablet PC)	A companion application of an educational toolkit to visualize IoT device data in AR. The agent can control IoT devices through the application, such as toggling an LED light or inserting a message on an LCD screen.	Figure 4c,d
[12]	Vision (camera view for object recognition and hand gesture [by external device]), Kinesthetics (proprioception: hand gesture)	Vision (2D/3D graphics, device control [smart object]), Audition (device control [sound from smart object])	HMD (HoloLens)	The agent can control IoT devices (speaker and smart light bulb) with hand gestures in an MR interface.	Figure 4c
[48]	Vision (camera view for AR marker detection), Kinesthetics (proprioception: position of agent), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Smartphone	The agent can scan QR codes attached to IoT devices to give commands to a robot that subsequently interacts with the respective IoT devices.	Figure 4c
[34]	Vision (camera view for object/AR marker detection and environment/movement tracking), Touch (tactility: touchscreen)	Vision (2D/3D graphics, device control [smart object])	Smartwatch, tablet PC	An AR application for a smart factory to visualize a path by indoor navigation (with AR markers) and provide information on the machines in the factory (without AR markers).	Figure 4b,c
[49]	Vision (camera view for hand gesture and image recognition, head movement [gaze]), Audition (voice)	Vision (2D/3D graphics), Audition (audio effect [sound cue])	HMD (HoloLens)	Presenting an MR interface to control IoT devices while reading data collected from them and visualizing the data on the agent's screen.	Figure 4c

5. Discussion

In this study, we used systematic literature review guidelines [32] to identify and analyze 23 studies that implemented MI systems that were composed of IoT and AR. The increase in MI systems with AR and IoT may have been affected by the expansion of the types of interaction devices due to the development of technologies, such as HMDs, projectors, and monitors (i.e., PC and TV). This is visible in Figure 7, which illustrates the yearly distribution of interaction devices across the reviewed studies. However, the smartwatch used in the MI system that was proposed by Sahinel et al. [34] is not listed, because AR content was not delivered, signifying that it did not satisfy our criterion for an interaction device (i.e., device that presents AR content) in the context of this study. The overall number of interaction devices was greater than 23, due to the fact that some studies used multiple interaction devices in their MI systems.

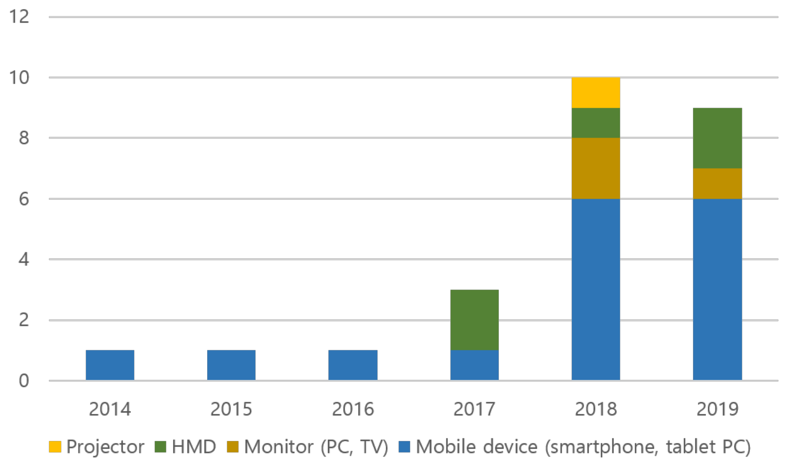


Figure 7. Yearly distribution of interaction devices.

Our analysis of the system architectures of MI systems, interaction between an agent and other entities (i.e., interaction device, IoT device, or other entity), and input/output modalities and modality combinations provided answers to the following five research questions that were initially presented in Section 3.1.

5.1. RQ 1. Which Modalities Are Used in MI Systems with IoT and AR?

We identified five unique modalities that were used as input and two unique modalities that were used as output in the reviewed studies, according to the results presented in Section 4.4. Each modality had several input sources from an agent, except for the audition. Because AR requires a view of the real world when overlaying virtual content, the vision modality was employed in every reviewed study. Moreover, vision was also used as an output modality in every reviewed study, as AR content is typically presented in either 2D or 3D graphics. Touch was the second most used input modality, possibly because the most frequently used interaction device was a mobile device (e.g., smartphone, tablet PC), which an agent primarily touches to perform interactions. While a mobile device was the most commonly used type of interaction device, an HMD was the second most used device. The third most implemented input modality was proprioception, as many interactions are performed by hands (e.g., hand gestures, position of hands). Kinematics and audition were the modalities least used as input. As an input modality, kinematics was only used in specially designed environments to detect the movement of physical objects sensed by IoT devices, whereas audition was used in two studies that utilized HMDs. Furthermore, audition was the least used output modality in the reviewed studies.

Thus, our findings suggest that audition is rarely used as either an input or output modality. Voice commands are preferred by agents over hand gestures as the input modality when using an HMD, as Karasinski et al. [45] determined through user evaluation. Therefore, the infrequency of audition as an input modality is remarkable, since both of the most commonly used interaction devices (i.e., mobile devices and HMD) support voice input from the agent. This infrequency may be caused by the type of interaction device and how familiar the agent is with the input modality. For example, mobile devices are frequently used in MI systems and they are a popular type of interaction device used by agents. They are designed to have primarily touch-based interaction on the screen; thus, agents are familiar with touch as an input modality. Moreover, environmental and social conditions may also limit the use of audition as an input modality. Thus, these findings can influence MI system developers to prioritize mobile devices as the interaction device and employ touch as the primary input modality.

The principle of familiarity can also be applied to an HMD. We only identified two studies that utilized voice commands as input [45,49]. Both of the MI systems were implemented on the HoloLens, and hand gestures could be used as the input modality along with voice commands. However, Karasinski et al. [45] found, through user evaluation of their HoloLens application, that the participants tended to use hand gestures with voice commands, although they reported preferring using only voice commands. This observation suggests that the similarity between hand gestures (e.g., finger tapping) and touch-based interaction (e.g., touching a touchscreen) provided the agents with sufficient familiarity. Therefore, even when the participants preferred using only voice commands, they continued to use both hand gestures and voice commands to interact with the MI system. This phenomenon can also be supported with one of the myths of multimodal interaction discussed by Oviatt [53], which claims that higher efficiency is the main advantage of multimodal systems. Oviatt [53] argued that the main advantages of MI systems are not about efficiency enhancement, but rather about other factors, such as the flexibility of selecting from multiple input modalities, which was demonstrated by Karasinski et al. [45]. This lack of consensus regarding efficiency improvement may be caused by missing knowledge regarding the relationship between agents' cognitive processes and each modality [5,6]. Therefore, an in-depth study on usability analysis in a number of MI systems with heterogeneous modalities is required to resolve this discord.

The input sources are another aspect of the study that produced noteworthy results. We determined that gaze was only used in three out of five studies that used an HMD as an interaction device. These three studies used Microsoft HoloLens, which enables gaze interaction. However, the HoloLens uses the center point of the screen as the gaze of an agent instead of detecting eye movements in real-time. As a result, the HoloLens imitates gaze interaction using the agent's head movement. This may lead to a limited experience, because an agent's gaze only moves with head direction, which is different from real gaze movements. In order to facilitate gaze interaction based on real eye movement, other computer vision techniques, such as the convolutional neural network [54], or devices (e.g., HoloLens 2) are essential.

5.2. RQ 2. Which Modality Combinations Are Used in MI Systems with IoT and AR?

According to Table 8, the most commonly used input modality combination was vision and touch. This result is consistent with the fact that a mobile device was the most commonly used interaction device, which is primarily used by touch-based interaction and it has a built-in camera for AR content. Proprioception is another input modality that was used with either vision or touch on mobile devices. Other input modalities, such as audition and kinematics, were combined with either vision or proprioception on other interaction devices (i.e., HoloLens and stationary camera connected to a PC/TV monitor). Notably, kinematics was only utilized when the interaction device was not a mobile device. This observation reveals unused input modality combinations on mobile devices. The findings also indicate the possibility of expanding the diversity of input modality combinations in future research. Other input modalities, such as neural oscillation (e.g., brainwave) [55] and galvanism (e.g., electrical activity of human body parts) [56,57], which were not found in the reviewed studies, can be used to explore new input modality combinations.

In addition, there also appears to be room for expanding output modality combinations, as we found that only the combination of vision and audition modalities was used as output. Technological developments can allow MI system developers to use various modalities that were not observed in the reviewed MI systems, such as olfaction [58], gustation [59], haptics [60], and vibration reception [61].

5.3. RQ 3. How Are IoT and AR Used in MI Systems?

Through our analysis, we determined that IoT devices are used either as identifiers for the placement of AR content or as devices that an agent can use to control the MI system. However, sensing and data collection over the network is the main function of using IoT devices. Therefore, all of the reviewed studies used IoT devices to collect data from their surroundings and then send them to the server for further analysis and interaction. The IoT devices used in the reviewed studies collected and communicated various data, such as the power state, energy consumption, motor speed, temperature, humidity, and relative position. Surprisingly, IoT devices were rarely used in the reviewed studies to collect an agent's biosignals, only appearing in one study [37]. Unlike an input source for the input modality discussed in RQ 2, biosignals can be collected by IoT devices and used to understand the state of an agent.

We found that the reviewed MI systems used AR either as a data presentation tool to deliver data to an agent or as an interface to allow an agent to control IoT devices and other devices (e.g., lights in miniature building structures [42]). AR was used not only to visualize data received from IoT devices, but also to present pre-stored data, such as in AR navigation, to a destination represented by 3D arrows that are spatially pre-positioned in a registered environment [34,45]. Of the reviewed studies, those that implemented the navigation feature were restricted to indoor environments; thus, the development of MI systems for navigation in outdoor contexts with the use of AR and IoT remains an open challenge. Exploring the effects of AR, IoT, and multimodality on the QoE of outdoor navigation may be an important aspect because outdoor navigation is not a novel research topic [62].

5.4. RQ 4. Which Architectures Are Used to Construct MI Systems with IoT and AR?

We identified four patterns of MI system architectures, which are illustrated in Figure 4. Three of the patterns (Figure 4a–c) were utilized in most of the reviewed studies; however, the pattern in Figure 4d and combinations of two patterns (Figure 4c,d; Figure 4b,c) were only used in one study. Similar to the combined patterns, Figure 4d was rarely observed in the reviewed studies. A reason for the infrequency of the pattern in Figure 4d may be its high complexity as compared to that of Figure 4c. For instance, an agent in Figure 4c can watch IoT devices while controlling them through an AR interface; however, an agent in Figure 4d cannot see the IoT devices that he/she is controlling, and the result of the control commands is displayed on an interaction device and/or another entity (e.g., lights in miniature buildings [42]). Consequently, Figure 4d requires an additional link between the server and another entity to allow the agent to control another entity, thereby increasing the complexity of the architecture. Similarly, system architectures with combined patterns also have a relatively higher complexity than system architectures with a single pattern, which may explain why only two studies used MI system architectures with combined patterns to support various features, namely, an AR interface for controlling IoT devices and AR navigation for visualizing a path to a destination [34]. This implies that the use of multiple MI system architecture patterns is a potential method for making the system more flexible, as discussed by Oviatt [53].

5.5. RQ 5. Which Open Research Challenges Remain in MI Systems with IoT and AR?

We found that 13 of the 23 reviewed studies explicitly mentioned research challenges regarding various topics. We categorized the research challenges in Table 9, based on five keywords: (1) technology, (2) standardization, (3) scalability, (4) multi-agent, and (5) multidisciplinary.

(1) Solutions to the identified research challenges that were related to hardware and technical limitations generally involve the development of technology and hardware. For example, the limited field of view in HoloLens, as noted by Karasinski et al. [45], has been solved in HoloLens 2, which has a larger field of view than the original HoloLens (i.e., 30° in HoloLens versus 52° in HoloLens 2). As another example, insta-

bility in vision-based tracking of object features caused by insufficient features can be corrected by a tracking recovery process [48]. These types of research challenges rely on the development of technology unless alternative approaches are developed.

(2) The standardization of MI systems is key in achieving high scalability [11,50]. When MI system developers begin to use unified forms of both software components (e.g., data structures, data communication protocols, and UI structures) and hardware components, MI systems will gain the ability to handle different types of IoT devices, interaction devices, and interaction modalities while being able to communicate with other MI systems. Therefore, standardized MI system development can reduce the burden on MI system developers, who must otherwise consider the diversity of IoT devices, interaction devices, and interaction modalities when building an MI system for use in a real scenario outside the testbed.

(3) Scalability is a research challenge that must be solved to establish a large-scale environment that can manage multiple types of IoT devices and agents that are easily and automatically registered. The absence of standardized data structures for IoT devices is the reason for the difficulty of this research challenge. Moreover, the increase in the number of IoT devices in MI systems may lead to significant computational costs on central servers. These challenges can occur when adding not only IoT devices, but also input/output modalities and interaction devices to MI systems. Therefore, considering the technology added to standardized MI systems is essential in achieving scalability.

(4) A multi-agent scenario is another critical case that MI system developers must take into account to achieve not only high scalability, but also high QoE. In the multi-agent scenario, the MI system must manage all interaction devices while maintaining quality of service, such as timely synchronization of data to all agents [44]. Moreover, the diversity of interaction devices and UIs can also enhance the QoE because agents have different preferences. The development of adaptive UIs is a challenge that can achieve a personalized experience for agents by providing customized interaction modalities, context-aware feedback, and even offering a specific interaction device, depending on the agent's preferences. Our findings suggested that 12 of the 23 reviewed studies conducted a user evaluation regarding usability and/or QoE. However, none of the reviewed studies evaluated their MI system in a multi-agent scenario. This indicates that an in-depth study of the on QoE of an MI system in the context of multiple agents is essential.

(5) The research challenges in the multidisciplinary category demonstrate the importance of collaboration with different disciplines. This importance can be observed, even in the development of MI systems utilizing AR without IoT [63]. The development of an MI system requires understanding the relevant technology and content to achieve a sufficient degree of QoE. For example, the use of emojis by Seitz et al. [33] to represent the state of IoT devices require psychological knowledge to avoid unintended results on various agents with different cultural backgrounds. In another study, He et al. [43] developed a smartboard to measure the state of an agent's upper limb in order to support a doctor's diagnosis. However, their MI system was not evaluated with real patients with upper limb disorders; thus, the effectiveness of their MI system must be reevaluated. This requires collaboration with an expert who is familiar with the technology and subjects that the MI system is used for.

In summary, by answering our research questions, we obtained the following key results that can be used by developers that are interested in creating MI systems and researchers curious about state-of-the-art trends in MI systems using AR and IoT.

1. We identified missing modalities and modality combinations that have not yet been tested in MI systems that use AR and IoT.
2. We proposed unexplored ways of using AR and IoT within an MI system.
3. We described patterns of MI system architectures that other developers can refer to in designing their MI systems.

4. We discussed open research challenges that can be considered in future research.

Although this study offers several contributions to the MI system research community, there are several limitations that the reader should be aware of: (1) we were unable to find a single study that was published in 2020 that satisfied our study selection criteria during the literature search. We assume that this result occurred because we did not cover the year 2020 in its entirety when we conducted the literature search in April 2020. We believe that we would have been able to find a number of studies if we conducted the literature search later in the year. (2) We used various keywords to search for studies that met our criteria. However, not all studies explicitly described their MI systems with the keywords used in our study. We expected to find studies that utilized biosignals (e.g., neural oscillation, galvanism) as input modality in an MI system based on AR and IoT. However, other keywords, such as brain-computer interface, may help to find additional studies, as we failed to identify these types of studies.

6. Conclusions

In this study, we conducted a systematic literature review of MI systems that use AR and IoT. We searched a number of studies with related keywords from various databases and selected 23 studies that met our quality criteria. We performed an in-depth analysis of each study to obtain answers to the research questions that were proposed at the early stage of this systematic literature review. In our in-depth analysis, we mainly focused on descriptive statistics of the reviewed studies, patterns of MI system architectures, types of interactions that occurred between an agent and other entities, input/output modalities and their combination, and open research challenges. We then discussed our findings to obtain answers to our research questions.

The results of this study demonstrate that the types of MI systems that utilize AR and IoT vary widely. As new technologies are developed, the number of usable input/output modalities that MI systems can utilize will increase continuously. Therefore, future research should conduct an in-depth study on the development of a unified architecture that supports all of the identified architectural patterns and modalities. Additionally, we are currently working on a unified MI framework that combines the architectural patterns that are presented in Figure 4 and connects to existing IoT platforms in order to support the development of future AR-enabled MI applications. We believe that our findings can provide insights for researchers with an interest in the field as well as important information for MI system developers.

Author Contributions: Conceptualization, J.C.K., T.H.L. and C.Å.; methodology, J.C.K., T.H.L. and C.Å.; validation, J.C.K., T.H.L. and C.Å.; formal analysis, J.C.K.; investigation, J.C.K., T.H.L. and C.Å.; resources, T.H.L. and C.Å.; data curation, J.C.K.; writing—original draft preparation, J.C.K. and T.H.L.; writing—review and editing, J.C.K., T.H.L. and C.Å.; visualization, J.C.K.; supervision, T.H.L. and C.Å.; project administration, J.C.K., T.H.L. and C.Å.; funding acquisition, T.H.L. and C.Å. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund (Grant project number 20201434 and NYPS 20204318), and this research also supported by the new faculty research fund of Ajou University (S-2020-G0001-00478).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this study.

Acknowledgments: The authors would like to thank Karan Mitra and Saguna saguna from Luleå University of Technology for helping us improve this article with their comments.

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

Abbreviations

The following abbreviations are used in this manuscript:

MI	Multimodal Interaction
IoT	Internet of Things
ICT	Information and Communication Technologies
AR	Augmented Reality
HMD	Head-mounted Display
UI	User Interface
QoE	Quality of Experience
HCI	Human-computer Interaction
ISO	International Organization for Standardization
MR	Mixed Reality
VR	Virtual Reality
2D	Two-dimensional
3D	Three-dimensional
LED	Light-emitting diode

References

1. Oviatt, S. Multimodal interfaces. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, 2nd ed.; L. Erlbaum Associates Inc.: Mahwah, NJ, USA, 2003; Volume 14, pp. 286–304.
2. Alam, M.R.; Reaz, M.B.I.; Ali, M.A.M. A Review of Smart Homes—Past, Present, and Future. *IEEE Trans. Syst. Man, Cybern. Part C (Appl. Rev.)* **2012**, *42*, 1190–1203. [CrossRef]
3. Gharaibeh, A.; Salahuddin, M.A.; Hussini, S.J.; Khreishah, A.; Khalil, I.; Guizani, M.; Al-Fuqaha, A. Smart Cities: A Survey on Data Management, Security, and Enabling Technologies. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2456–2501. [CrossRef]
4. Wang, J.; Liu, J.; Kato, N. Networking and Communications in Autonomous Driving: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 1243–1274. [CrossRef]
5. Jaimes, A.; Sebe, N. Multimodal human–computer interaction: A survey. *Comput. Vis. Image Underst.* **2007**, *108*, 116–134. [CrossRef]
6. Turk, M. Multimodal interaction: A review. *Pattern Recognit. Lett.* **2014**, *36*, 189–195. [CrossRef]
7. Patel, K.K.; Patel, S.M.; Scholar, P. Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. *Int. J. Eng. Sci. Comput.* **2016**, *6*, 10.
8. Nižetić, S.; Šolić, P.; López-de-Ipiña González-de Artaza, D.; Patrono, L. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *J. Clean. Prod.* **2020**, *274*, 122877. [CrossRef]
9. Bhargava, M.; Dhote, P.; Srivastava, A.; Kumar, A. Speech enabled integrated AR-based multimodal language translation. In Proceedings of the 2016 Conference on Advances in Signal Processing (CASP), Pune, India, 9–11 June 2016; IEEE: New York, NY, USA, 2016; pp. 226–230. [CrossRef]
10. Dodevska, Z.A.; Kvrgić, V.; Štavljanin, V. Augmented Reality and Internet of Things – Implementation in Projects by Using Simplified Robotic Models. *Eur. Proj. Manag. J.* **2018**, *8*, 27–35. [CrossRef]
11. Jo, D.; Kim, G.J. IoT + AR: Pervasive and augmented environments for "Digi-log" shopping experience. *Hum.-Centric Comput. Inf. Sci.* **2019**, *9*. [CrossRef]
12. Sun, Y.; Armengol-Urpi, A.; Reddy Kantareddy, S.N.; Siegel, J.; Sarma, S. MagicHand: Interact with IoT Devices in Augmented Reality Environment. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; IEEE: New York, NY, USA, 2019; pp. 1738–1743. [CrossRef]
13. Zhang, L.; Chen, S.; Dong, H.; El Saddik, A. Visualizing Toronto City Data with HoloLens: Using Augmented Reality for a City Model. *IEEE Consum. Electron. Mag.* **2018**, *7*, 73–80. [CrossRef]
14. Hadj Sassi, M.S.; Chaari Fourati, L. Architecture for Visualizing Indoor Air Quality Data with Augmented Reality Based Cognitive Internet of Things. In *Advanced Information Networking and Applications*; Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M., Eds.; Series: Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2020; Volume 1151, pp. 405–418. [CrossRef]
15. Mathews, N.S.; Chimalakonda, S.; Jain, S. AiR—An Augmented Reality Application for Visualizing Air Pollution. *arXiv* **2020**, arXiv: 2006.02136.
16. White, G.; Cabrera, C.; Palade, A.; Clarke, S. Augmented Reality in IoT. In *Service-Oriented Computing—ICSOC 2018 Workshops*; Liu, X., Mrissa, M., Zhang, L., Benslimane, D., Ghose, A., Wang, Z., Bucchiarone, A., Zhang, W., Zou, Y., Yu, Q., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 11434, pp. 149–160. [CrossRef]
17. Jo, D.; Kim, G.J. AR Enabled IoT for a Smart and Interactive Environment: A Survey and Future Directions. *Sensors* **2019**, *19*, 4330. [CrossRef]
18. Blackler, A.; Popovic, V.; Mahar, D. Investigating users' intuitive interaction with complex artefacts. *Appl. Ergon.* **2010**, *41*, 72–92. [CrossRef]

19. Hogan, T.; Hornecker, E. Towards a Design Space for Multisensory Data Representation. *Interact. Comput.* **2016**. [CrossRef]
20. Liang, R.; Liang, B.; Wang, X.; Zhang, T.; Li, G.; Wang, K. A Review of Multimodal Interaction. In Proceedings of the 2016 International Conference on Education, Management, Computer and Society, Shenyang, China, 1–3 January 2016; Atlantis Press: Amsterdam, The Netherlands, 2016. [CrossRef]
21. Badouch, A.; Krit, S.D.; Kabrane, M.; Karimi, K. Augmented Reality services implemented within Smart Cities, based on an Internet of Things Infrastructure, Concepts and Challenges: An overview. In Proceedings of the Fourth International Conference on Engineering & MIS 2018—ICEMIS '18, Istanbul, Turkey, 19–21 June 2018; ACM Press: New York, NY, USA, 2018; pp. 1–4. [CrossRef]
22. Norouzi, N.; Bruder, G.; Belna, B.; Mutter, S.; Turgut, D.; Welch, G. A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things. In *Artificial Intelligence in IoT*; Al-Turjman, F., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–24. [CrossRef]
23. Picard, R.W. *Affective Computing*, 1st paperback ed.; OCLC: 247967780; The MIT Press: Cambridge, MA, USA; London, UK, 2000.
24. ISO. *Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*; ISO 9241-11:2018(en); ISO: Geneva, Switzerland, 2018.
25. ISO. *Information Technology—Future Network—Problem Statement and Requirements—Part 6: Media Transport*; ISO/IEC TR 29181-6:2013(en); ISO: Geneva, Switzerland, 2013.
26. ITU. *P.10 : Vocabulary for Performance, Quality of Service and Quality of Experience*; ITU: Geneva, Switzerland, 2017.
27. Sánchez, J.; Saenz, M.; Garrido, J.M. Usability of a Multimodal Video Game to Improve Navigation Skills for Blind Children. *ACM Trans. Access. Comput.* **2010**, *3*, 1–29. [CrossRef]
28. Blattner, M.; Glinert, E. Multimodal integration. *IEEE Multimed.* **1996**, *3*, 14–24. [CrossRef]
29. Augstein, M.; Neumayr, T. A Human-Centered Taxonomy of Interaction Modalities and Devices. *Interact. Comput.* **2019**, *31*, 27–58. [CrossRef]
30. Nizam, S.S.M.; Abidin, R.Z.; Hashim, N.C.; Chun, M.; 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*.
31. Mohamad Yahya Fekri, A.; Ajune Wanis, I. A review on multimodal interaction in Mixed Reality Environment. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *551*, 012049. [CrossRef]
32. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report; Keele University: Keele, UK, 2007.
33. Seitz, A.; Henze, D.; Nickles, J.; Sauer, M.; Bruegge, B. Augmenting the industrial Internet of Things with Emojis. In Proceedings of the 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC), Barcelona, Spain, 23–26 April 2018; IEEE: New York, NY, USA, 2018; pp. 240–245. [CrossRef]
34. Sahinel, D.; Akpolat, C.; Gorur, O.C.; Sivrikaya, F. Integration of Human Actors in IoT and CPS Landscape. In Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, 15–18 April 2019; IEEE: New York, NY, USA, 2019; pp. 485–490. [CrossRef]
35. Leppanen, T.; Heikkinen, A.; Karhu, A.; Harjula, E.; Riekki, J.; Koskela, T. Augmented Reality Web Applications with Mobile Agents in the Internet of Things. In Proceedings of the 2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies, Oxford, UK, 10–12 September 2014; IEEE: New York, NY, USA, 2014; pp. 54–59. [CrossRef]
36. Rashid, Z.; Melià-Seguí, J.; Pous, R.; Peig, E. Using Augmented Reality and Internet of Things to improve accessibility of people with motor disabilities in the context of Smart Cities. *Future Gener. Comput. Syst.* **2017**, *76*, 248–261. [CrossRef]
37. Agrawal, D.; Mane, S.B.; Pacharne, A.; Tiwari, S. IoT Based Augmented Reality System of Human Heart: An Android Application. In Proceedings of the 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–12 May 2018; IEEE: New York, NY, USA, 2018; pp. 899–902. [CrossRef]
38. Alam, M.F.; Katsikas, S.; Beltramello, O.; Hadjiefthymiades, S. Augmented and virtual reality based monitoring and safety system: A prototype IoT platform. *J. Netw. Comput. Appl.* **2017**, *89*, 109–119. [CrossRef]
39. Pokric, B.; Krco, S.; Drajic, D.; Pokric, M.; Rajs, V.; Mihajlovic, Z.; Knezevic, P.; 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.
40. Mylonas, G.; Triantafyllis, C.; Amaxilatis, D. An Augmented Reality Prototype for supporting IoT-based Educational Activities for Energy-efficient School Buildings. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 89–101. [CrossRef]
41. Simões, B.; De Amicis, R.; Barandiaran, I.; Posada, J. X-Reality System Architecture for Industry 4.0 Processes. *Multimodal Technol. Interact.* **2018**, *2*, 72. [CrossRef]
42. Cho, K.; Jang, H.; Park, L.W.; Kim, S.; Park, S. Energy Management System Based on Augmented Reality for Human-Computer Interaction in a Smart City. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 11–13 January 2019; IEEE: New York, NY, USA, 2019; pp. 1–3. [CrossRef]
43. He, Y.; Sawada, I.; Fukuda, O.; Shima, R.; Yamaguchi, N.; Okumura, H. Development of an evaluation system for upper limb function using AR technology. In Proceedings of the Genetic and Evolutionary Computation Conference Companion on—GECCO'18, Kyoto, Japan, 15–19 July 2018; ACM Press: New York, NY, USA, 2018; pp. 1835–1840. [CrossRef]

44. Huo, K.; Cao, Y.; Yoon, S.H.; Xu, Z.; Chen, G.; Ramani, K. Scenariot: Spatially Mapping Smart Things Within Augmented Reality Scenes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems—CHI'18, Montreal QC, Canada, 21–26 April 2018; ACM Press: New York, NY, USA, 2018; pp. 1–13. [CrossRef]
45. Karasinski, J.A.; Joyce, R.; Carroll, C.; Gale, J.; Hillenius, S. An Augmented Reality/Internet of Things Prototype for Just-in-time Astronaut Training. In *Virtual, Augmented and Mixed Reality*; Lackey, S., Chen, J., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10280, pp. 248–260. [CrossRef]
46. Muthanna, A.; Ateya, A.A.; Amelyanovich, A.; Shpakov, M.; Darya, P.; Makolkina, M. AR Enabled System for Cultural Heritage Monitoring and Preservation. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; Galinina, O., Andreev, S., Balandin, S., Koucheryavy, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11118, pp. 560–571. [CrossRef]
47. Phupattanasilp, P.; Tong, S.R. Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming. *Sustainability* **2019**, *11*, 2658. [CrossRef]
48. Cao, Y.; Xu, Z.; Li, F.; Zhong, W.; Huo, K.; Ramani, K. V.Ra: An In-Situ Visual Authoring System for Robot-IoT Task Planning with Augmented Reality. In Proceedings of the 2019 on Designing Interactive Systems Conference—DIS'19, San Diego, CA, USA, 23–28 June 2019; ACM Press: New York, NY, USA, 2019; pp. 1059–1070. [CrossRef]
49. de Belen, R.A.J.; Bednarz, T.; Favero, D.D. Integrating Mixed Reality and Internet of Things as an Assistive Technology for Elderly People Living in a Smart Home. In Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, Brisbane, QLD, Australia, 14–16 November 2019; ACM: New York, NY, USA, 2019; pp. 1–2. [CrossRef]
50. Jo, D.; Kim, G.J. ARIoT: Scalable augmented reality framework for interacting with Internet of Things appliances everywhere. *IEEE Trans. Consum. Electron.* **2016**, *62*, 334–340. [CrossRef]
51. Purmaissur, J.A.; Towakel, P.; Guness, S.P.; Seeam, A.; Bellekens, X.A. Augmented-Reality Computer-Vision Assisted Disaggregated Energy Monitoring and IoT Control Platform. In Proceedings of the 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC), Plaine Magnien, Mauritius, 6–7 December 2018; IEEE: New York, NY, USA, 2018; pp. 1–6. [CrossRef]
52. Stefanidi, E.; Foukarakis, M.; Arampatzis, D.; Korozi, M.; Leonidis, A.; Antona, M. ParlAmI: A Multimodal Approach for Programming Intelligent Environments. *Technologies* **2019**, *7*, 11. [CrossRef]
53. Oviatt, S. Ten myths of multimodal interaction. *Commun. ACM* **1999**, *42*, 74–81. [CrossRef]
54. Fuhl, W.; Santini, T.; Kasneci, G.; Kasneci, E. PupilNet: Convolutional Neural Networks for Robust Pupil Detection. *arXiv* **2016**. arXiv:1601.04902.
55. Gürkök, H.; Nijholt, A. Brain–Computer Interfaces for Multimodal Interaction: A Survey and Principles. *Int. J. Hum.-Comput. Interact.* **2012**, *28*, 292–307. [CrossRef]
56. Gorzkowski, S.; Sarwas, G. Exploitation of EMG Signals for Video Game Control. In Proceedings of the 2019 20th International Carpathian Control Conference (ICCC), Krakow-Wieliczka, Poland, 26–29 May 2019; IEEE: New York, NY, USA, 2019; pp. 1–6. [CrossRef]
57. Liao, S.C.; Wu, F.G.; Feng, S.H. Playing games with your mouth: Improving gaming experience with EMG supportive input device. In Proceedings of the International Association of Societies of Design Research Conference, Manchester, UK, 2–5 September 2019.
58. Risso, P.; Covarrubias Rodriguez, M.; Bordegoni, M.; Gallace, A. Development and Testing of a Small-Size Olfactometer for the Perception of Food and Beverages in Humans. *Front. Digit. Humanit.* **2018**, *5*, 7. [CrossRef]
59. Ranasinghe, N.; Do, E.Y.L. Digital Lollipop: Studying Electrical Stimulation on the Human Tongue to Simulate Taste Sensations. *ACM Trans. Multimed. Comput. Commun. Appl.* **2016**, *13*, 1–22. [CrossRef]
60. Zenner, A.; Kruger, A. Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1285–1294. [CrossRef] [PubMed]
61. Hussain, I.; Meli, L.; Pacchierotti, C.; Salvietti, G.; Prattichizzo, D. Vibrotactile haptic feedback for intuitive control of robotic extra fingers. In Proceedings of the 2015 IEEE World Haptics Conference (WHC), Evanston, IL, USA, 22–26 June 2015; IEEE: New York, NY, USA, 2015; pp. 394–399. [CrossRef]
62. Al-Jabi, M.; Sammaneh, H. Toward Mobile AR-based Interactive Smart Parking System. In Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications, IEEE 16th International Conference on Smart City, IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; IEEE: New York, NY, USA, 2018; pp. 1243–1247. [CrossRef]
63. Kim, J.C.; Lindberg, R.S.N.; Laine, T.H.; Faarinen, E.C.; Troyer, O.D.; Nygren, E. Multidisciplinary Development Process of a Story-based Mobile Augmented Reality Game for Learning Math. In Proceedings of the 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA), Smokovec, Slovakia, 21–22 November 2019; IEEE: New York, NY, USA, 2019; pp. 372–377. [CrossRef]

Short Biography of Authors

Joo Chan Kim is a computer scientist with an interest in augmented reality, Internet of Things, and human-computer interaction; he is a doctoral student at Luleå University of Technology.

Teemu H. Laine received a Ph.D. in computer science from the University of Eastern Finland in 2012. He currently holds a position of an Associate Professor at Ajou University. Laine has a strong track record in researching context-aware games and applications for education and well-being. His other research interests include augmented and virtual reality, human-computer interaction, system architectures, and artificial intelligence.

Christer Åhlund received the Ph.D. degree from the Luleå University of Technology, Skellefteå, Sweden, in 2005. He is a Chaired Professor of pervasive and mobile computing with the Luleå University of Technology. He is also the Scientific Director of excellence in research and innovation named Enabling ICT. Beyond his academic background, he has 12 years of industry experience in the ICT area. His research interests include Internet mobility, wireless access networks, Internet of Things, and cloud computing.