A Taxonomy of Interaction Techniques for Immersive Augmented Reality based on an Iterative Literature Review

Julia Hertel*
Human-Computer Interaction
Universität Hamburg

Sukran Karaosmanoglu†
Human-Computer Interaction
Universität Hamburg

Susanne Schmidt‡
Human-Computer Interaction
Universität Hamburg

Julia Bräker§
IT Management & Consulting
Universität Hamburg

Martin Semmann¶
IT Management & Consulting
Universität Hamburg

Frank Steinicke‖
Human-Computer Interaction
Universität Hamburg

Abstract

Developers of interactive systems have a variety of interaction techniques to choose from, each with individual strengths and limitations in terms of the considered task, context, and users. While there are taxonomies for desktop, mobile, and virtual reality applications, augmented reality (AR) taxonomies have not been established yet. However, recent advances in immersive AR technology (i.e., head-worn or projection-based AR), such as the emergence of untethered headsets with integrated gesture and speech sensors, have enabled the inclusion of additional input modalities and, therefore, novel multimodal interaction methods have been introduced. To provide an overview of interaction techniques for current immersive AR systems, we conducted a literature review of publications between 2016 and 2021. Based on 44 relevant papers, we developed a comprehensive taxonomy focusing on two identified dimensions – *task* and *modality*. We further present an adaptation of an iterative taxonomy development method to the field of human-computer interaction. Finally, we discuss observed trends and implications for future work.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques—; Human-centered computing—Interaction design theory, concepts and paradigms—

1 Introduction

The research fields of augmented reality (AR) and virtual reality (VR) have their origins more than 50 years ago in the 1960s [44]. Besides the AR/VR display technologies themselves, various techniques, which allow users to interact with these systems, were proposed and investigated ever since. Previous work has established the systematization of interfaces and interaction techniques with a focus on VR. For instance, in 1999, Bowman et al. [17] formalized interaction techniques for immersive virtual environments (VEs) in the form of a taxonomy. They identified three universal tasks – (i) viewport motion control, (ii) selection, and (iii) manipulation, which they decomposed into fine-grained subtasks. For each of these subtasks, they identified several existing interaction techniques that could be used to accomplish them. The resulting tree-like taxonomy presented a broad overview of existing interaction techniques and was also used as a basis for evaluating interaction techniques [17]. While this taxonomy was focused on immersive VEs only, in 2004, Bowman et al. [18] collected 3D user interfaces that included both VR and AR. In their book, they suggest design guidelines for the development of AR and VR interaction techniques.

Since these early approaches to systematically guide interaction design, *immersive AR* has experienced enormous technical advancements. In this context, we define *immersive AR* as technology, which allows viewing AR content in an immersive way (e.g., via head-worn displays or projection-based AR) without external displays as used in mobile or hand-held systems. The recent launch of self-contained untethered head-mounted displays (HMDs), such as the Microsoft HoloLens in 2016 [34], paved the way to let consumers observe virtual content in an unobtrusive manner. Due to the integration of multisensory input systems, including voice recognition, head tracking, hand tracking, and even eye tracking (e.g., on the HoloLens 2), developers and users have access to various interaction possibilities. However, the lack of standardization in immersive AR interaction techniques as well as the variety of options often complicates the design process and might confuse AR developers and researchers.

In the last years, researchers reviewed, analyzed, and structured various aspects of the field of AR. For instance, there are reviews focusing on uses, advantages, and trends of AR in specific application fields like education [12, 27], medicine [22, 37], or maintenance [23, 24]. Further, taxonomies about subtopics of AR research like visualization [46] and context awareness [28] were developed. In 2018, Muhammad Nizam et al. [36] reviewed interaction techniques in AR, focusing on multimodal interaction. Expanding the existing body of research, we aimed to analyze interaction techniques in immersive AR systems. We did not focus on specific kinds of techniques but performed an explorative approach examining common characteristics among techniques to develop a comprehensive taxonomy. While in the field of human-computer interaction (HCI) the concept of interaction can be understood as a reciprocal exchange between a user and a system, we only focus on actions a user performs actively. Hence, we considered techniques that allow a user to provide any kind of input to an immersive AR system and did not consider aspects like visualization and feedback.

Our work aims to capture the current state of research of immersive AR interaction techniques, extract common characteristics among these techniques, and develop a taxonomy that sorts and groups them accordingly. The taxonomy is intended to provide an overview, a common ground for discussion as well as an aid to identify research gaps, emerging trends, and related interaction techniques to evaluate new ones. For this purpose, we considered research papers published within the last five years since 2016. In the year 2016 – sometimes referred to as the "year of AR" [19] – several AR milestones have been achieved such as the launch of Microsoft HoloLens or Pokémon Go [38].

In this paper, we aim to answer the following research questions:

- **RQ1:** *Which immersive AR interaction techniques are investigated in the current literature?*
- **RQ2:** *Which characteristics can be found among these techniques and how can these be systematized?*

*julia.hertel@uni-hamburg.de
†sukran.karaosmanoglu@uni-hamburg.de
‡susanne.schmidt@uni-hamburg.de
§julia.braeker@uni-hamburg.de
¶martin.semmann@uni-hamburg.de
‖frank.steinicke@uni-hamburg.de

Data Source	Query	Number of Papers
ACM *Search Date: April 23, 2021*	`"interact* AND ("augmented reality" OR "ar" OR "mixed reality" OR "mr")"` Only Abstract and Title	725
IEEE *Search Date: April 23, 2021*	`"interact* AND ("augmented reality" OR "ar" OR "mixed reality" OR "mr")"` Only Abstract and Title	1540

Table 1: Data sources, queries, dates, numbers and publication dates of results.

To answer these research questions, we conducted a literature review on the Association for Computing Machinery (ACM) [2] and Institute of Electrical and Electronics Engineers (IEEE) [9] databases. We derived a taxonomy of emerging immersive AR interaction techniques based on 44 research publications, that were identified in an iterative process [39]. In summary, our contributions are as follows:

- application and adaptation of a taxonomy creation methodology to the field of HCI,
- presentation of a taxonomy of emerging immersive AR interaction techniques focused on the task and modality features,
- report and discussion of the implications of these techniques based on the literature review.

The remainder of this paper is structured as follows. Section 2 introduces the methodology of the iterative literature review. Section 3 introduces the resulting taxonomy, supported by examples that were considered during the literature review. Section 4 discusses the findings and limitations of our approach. Section 5 concludes the paper and gives an outlook on future work.

2 METHODOLOGY

In this work, we aimed to capture the current state of research of immersive AR interaction techniques and categorize them in a comprehensible form. We decided to develop a taxonomy, which is "a system of groupings that are derived conceptually or empirically" [39]. Structured procedures to develop taxonomies are commonly used in other fields of research, like biology or information systems (IS) [39]. We introduce and adapt an established taxonomy definition and development approach from the IS field to the field of HCI. The approach was developed by Nickerson et al. [39] in 2013 and is commonly used in IS research since then [32, 45]. Nickerson et al. define a taxonomy as a set of dimensions, where each dimension consists of characteristics. The characteristics are mutually exclusive – no object can have multiple characteristics in a dimension – and collectively exhaustive – each object must have one characteristic in each dimension. We extended this definition by subcategories, so that categories can be split up if needed to reflect a more fine-grained characterization. Further, we noticed that forcing interaction techniques to be mutually exclusive could conceal valuable insights. Thus, we did not adopt the mutual exclusivity restriction to depict, for example, multimodality.

To develop such a taxonomy, Nickerson et al. designed an iterative approach, which allows researchers to include a subset of objects in each iteration until pre-defined ending conditions are met. In each iteration, the current taxonomy is refined based on newly emerged characteristics.

2.1 Meta-characteristics

The initial step in the taxonomy development process is the definition of meta-characteristics. In general, they are conceptually induced based on the researcher's knowledge and should reflect the intended purpose of the taxonomy. As meta-characteristics, we chose *people, activities, context,* and *technologies*, the four dimensions of the PACT framework developed by Benyon [14]. The framework's premise is that people use technology to undertake activities in contexts, and thus these dimensions should be considered when designing interactive systems. Here, *people* include physical, psychological, and social differences between users. For *context*, aspects like physical and social environments should be analyzed. Regarding *activities*, Benyon considers the overall purpose of the activity as most important. *Technology* includes, for example, the input and output of a system. By analyzing articles with regard to these meta-characteristics, we expect to find characteristics of interaction techniques that are valuable for researchers and designers.

2.2 Ending conditions

We also defined ending conditions before the actual iterative process started. They are both objective and subjective and determine when the taxonomy development process should terminate. We used the ending conditions proposed by Nickerson et al. which include objective ones (e.g., the taxonomy was not modified in the last iteration, the taxonomy is valid in regard to the definition of a taxonomy) and subjective ones (the taxonomy should be concise, robust, comprehensive, extendable, and explanatory).

2.3 Data Acquisition

The following steps of our taxonomy development process aimed at capturing the current state of research by extracting immersive AR interaction techniques from research articles. To obtain an initial set of articles, we followed the core steps of the PRISMA-P protocol [43]. The PRISMA protocol of our study can be found in the supplementary materials.

The PRISMA flowchart (Fig. 1) shows all steps we followed for the literature review: identification of sources ($N = 2265$), removal of duplicates and ACM articles that did not include our search terms ($N = 2100$), identification of venues depending on their number of papers (VR [7], CHI [1], TVCG [6], ISMAR [5], UIST [3], IROS [8]), iterative screening step for identified records ($N = 283$), full-text screening step for eligibility ($N = 74$). As a result of these steps, a total number of 44 papers were integrated into our taxonomy.

We performed a search using the digital libraries of IEEE and ACM, but only included venues with a CORE2020 rank [4] of "good" or higher (A*, A, or B). We chose to use these databases as they cover the best known and most relevant HCI conferences in the field related to 3D interaction techniques. The goal was to capture recent research interaction techniques with a focus on immersive AR. An initial informal analysis revealed that terms like "interaction technique" or synonyms are often not mentioned directly. Thus, we included a more general term in our search query: "interact*". Since "immersive AR" is not an established term yet, we searched for AR and MR in general and filtered for immersive AR during the subsequent screening step. We restricted our search to the document title and abstract since terms such as AR and MR often appear in conference titles and keywords. Table 1 shows the resulting search query. We included peer-reviewed research articles, excluding posters, published since 2016. The search was conducted on April 23, 2021.

The initial search resulted in 2265 research articles (ACM = 725, IEEE = 1540) from which we removed 22 duplicates based on title and year. The ACM website automatically included articles where the venue title matched our search query but the article title did not. We excluded these articles ($N = 143$). The remaining 2100 articles were grouped by venue and the venues were sorted by the number of articles in descending order. By this means, research papers

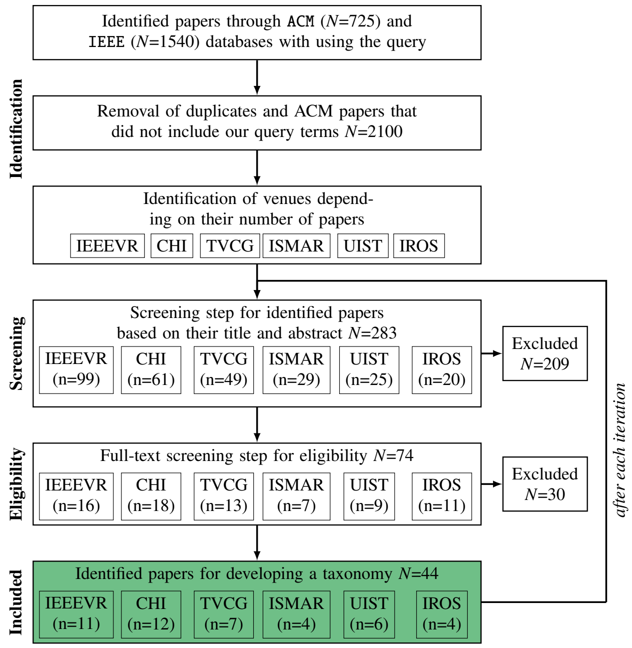

Figure 1: PRISMA flowchart, which illustrates our literature review process for creating a taxonomy with an iterative approach [39].

were divided into subsets to include them iteration-wise, with one venue per iteration. In total, we performed 6 empirical-to-conceptual iterations, resulting in the inclusion of 283 articles from 6 venues. The remaining 1817 articles were not screened.

2.4 Iterative Development Process

For each iteration, the researchers choose whether to follow an *empirical-to-conceptual* or a *conceptual-to-empirical* approach. In the former, new subsets of objects are included and common characteristics among them are identified. Then, these characteristics are grouped into dimensions and the taxonomy is created or revised. In the latter, characteristics and dimensions are conceptualized based on existing work or the researchers' knowledge. Then, objects for these characteristics and dimensions are examined and the taxonomy is created or revised.

2.4.1 Empirical-to-conceptual Iterations

In each empirical-to-conceptual iteration, we performed a literature review of the corresponding subset of research articles as follows.

Screening In total, we screened the titles and abstracts of 283 papers and excluded articles ($N = 209$) which did not match the following inclusion criteria:

- Papers should either present or evaluate one or more interaction techniques. We excluded papers with a focus on an application or experiment, where the used interaction technique was mentioned but not described in detail.
- Interaction techniques should clearly be intended for immersive AR applications. We excluded papers about interaction techniques for general or VR settings, where the authors only mention the possibility to adapt the interaction technique to AR. We also excluded interaction techniques for hand-held AR systems. We included interaction techniques that were clearly designed for AR applications but implemented or evaluated in VR settings.
- We excluded literature reviews such as meta-analyses.

Three researchers were involved in this step, two of them having a background in computer science and AR, and one of them having a cognitive systems background. Each third of the data was screened by two researchers, the third one acted as a tie-breaker. Since this step requires rigorous consensus between the coders, we conducted Cohen's Kappa test to calculate inter-rater reliability between the three researchers [26]. The findings indicated high agreement between lead and second (97.9%; $\kappa = 0.956$, $z = 9.37$, $p < .0001$), lead and third (94.7%; $\kappa = 0.846$, $z = 8.46$, $p < .0001$), and second and third (95.6%; $\kappa = 0.887$, $z = 8.52$, $p < .0001$) researchers. In this step, we excluded 209 papers based on our inclusion criteria: duplicates ($N = 7$), AR is not the main focus ($N = 6$), non-immersive AR ($N = 26$), interaction techniques are not the main focus ($N = 143$), literature review ($N = 1$), and no publication ($N = 18$).

Full-text Eligibility In this step, we considered the full text of the remaining papers ($N = 74$) for eligibility and used keyword coding as described below:

- **Keyword coding:** We searched keywords with regard to the meta-characteristics (PACT [14]), for example, *multi-user*, *selection task*, *gestures*, *HoloLens* (more examples can be seen in the supplementary material). We used keyword coding to facilitate the taxonomy development in the next step. This step was done by the same three researchers and the same division as in the previous step.

Taxonomy Development This step consists of determining the characteristics and dimensions of the taxonomy by deriving them from included papers. We considered 44 papers that resulted from the full-text eligibility step.

- **Characteristics and dimension identification:** Each keyword identified in the last step was transferred onto a digital sticky note on the digital whiteboard platform Miro [10]. In a group discussion with three researchers, these sticky notes were grouped into characteristics and dimensions by moving them around. Only characteristics and dimensions that seemed relevant to the researchers were created while keeping the purpose of the taxonomy and the main constraint (every object must have one characteristic in each dimension) in mind.
- **Tagging:** Based on the derived characteristics and dimensions, the first and third authors tagged all interaction techniques of the current batch. Subsequently, all previous interaction techniques were retagged, since the taxonomy changed iteratively. Further, the nuances from these assignments were discussed after each iteration with a group of three researchers. This step verified that the current taxonomy is valid and the main constraint is fulfilled.

We note that we included at least two researchers (analysis triangulation) for each step of our literature review to ensure reliability and reduce potential errors [20, 33].

2.4.2 Conceptual-to-Empirical Iteration

After the third iteration of our literature review, we noticed that some identified papers use terms interchangeably, for instance, for describing gestures and tangibles. Also, taxonomy development generally requires an in-depth engagement with previous literature in order to be valid, concise, and consistent. Hence, we decided to use the conceptual-to-empirical approach after the third iteration.

We identified 8 papers to refine our taxonomy. These were Bowman et al. [17]'s taxonomy on 3D VR interaction techniques and tasks, Quek et al. [41]'s prior work on gestures and speech, Shaer et al. [42]'s paper on tangible user interfaces, Besançon et al. [16]'s work on touch/tangible spatial 3D selection tasks, Karpov et al. [30]'s work on multimodal interfaces, Buxton et al. [21]'s work on input devices, Nizam et al. [40]'s review of multimodal interaction techniques, and Augstein et al. [13]'s taxonomy of interaction modalities and devices from a human-centered perspective.

In this step, the first three authors read the above articles in detail and hold a discussion to clearly define the characteristics and dimensions of the current taxonomy at that stage. As a result, we renamed, merged, or split some characteristics and dimensions. We applied these changes to all previously considered papers. Afterwards, we continued with an empirical-to-conceptual iteration.

3 Derived Taxonomy: Immersive AR Interaction Techniques

Our review aimed to develop a taxonomy of immersive AR interaction techniques. As described in Section 2, a taxonomy is a set of dimensions, where each dimension consists of characteristics. Since the taxonomy has to be collectively exhaustive, all dimensions are required to be applicable to every interaction technique and to contain at least one characteristic suitable for each interaction technique. By analyzing the empirically derived set of interaction techniques with a focus on PACT as meta-characteristics, we finally identified two dimensions matching these constraints: *task* and *modality*. Both dimensions' characteristics can be interpreted as *categories* comprising a set of similar tasks and modalities, respectively. In the remainder of this paper, the terms characteristic and category will be used interchangeably.

To add further details, we separated these into sub-characteristics where appropriate. Fig. 2 shows the resulting taxonomy. To reflect the concept of multimodality, we chose to apply a mutual exclusivity restriction only to the task dimension: each interaction technique has exactly one task, but one or more modalities. Table 2 shows the frequency of each task-modality combination in both, uni-modal and multimodal interaction techniques. This table acts as an overview and guideline by presenting which modalities were commonly used to perform certain tasks.

3.1 Tasks

Similar to Bowman et al. [17], we identified several granular tasks. Each of them can either be used stand-alone or combined with other tasks to create more complex, high-level functionalities. Below, we will provide a brief definition of each task and present exemplary interaction techniques examined in our literature review.

3.1.1 Creation

The task category creation includes actions a user can perform in order to let virtual objects appear visually. We identified three subcategories based on what kind of object is created and how much user input is needed. An *activation* is performed when the user does not define the geometry or visual appearance of the object. That includes, for example, an existing digital object that is not rendered until a user performs an activation to let it visually appear [70], labels that are superimposed on objects [60], or details on demand in the context of data visualization [71]. Furthermore, an activation can generate a new virtual object. Here, properties like the object's position, rotation, and scale may be predefined in the application (e.g., [54,84]), set through previous actions (e.g., copying an existing object [71]), or are implicitly given by the user while performing the activation, for example, via the user's current head orientation or hand position [69,77]. In contrast, the other subcategories contain tasks that require the user to incrementally define the appearance of the created object, either by modifying textures (*2D drawing* [47,78, 87]) or by modifying the object's geometry (*3D modeling* [47,75]).

All subcategories of creation can also include the deletion of objects. However, we did not include the deletion as a separate task, since all of the reviewed deletion interaction techniques use the same modality as their activation counterparts.

3.1.2 Selection

To interact with virtual objects, various interaction techniques require the user to identify an object through a selection. We identified selection as the most common task as it was included in 32 out of 44 papers. A selection can be applied as a single task that triggers a function linked to the selected object, for example, when interacting with a button on a user interface (UI) (e.g., [68,75]) or highlighting a part of the virtual scene (e.g., [71]). Moreover, a selection can precede tasks like geometric manipulation by determining the object to which the following operations should be applied (e.g., [83]). The release of a selected object can be considered as a separate task [17]. However, based on the literature review we examined that a release is typically performed indirectly by reverting or releasing the selection activity, like stopping a gesture or releasing a button (e.g. [63,67]). Besides, a release can not be performed standalone, as it has to be preceded by a selection. Thus, our taxonomy does not consider release explicitly.

We identified two subcategories based on the arrangement of objects that can be selected: *2D selection* and *3D selection*. For 2D selection, either a single or multiple virtual objects are placed on a planar surface like a real or virtual wall or an abstract plane [66]. A common example is a UI menu, consisting of multiple virtual buttons arranged in a 2D grid structure [63,68,75]. In contrast, if selectable objects are spatially arranged, like virtual furniture placed in a room [63], a 3D selection is needed. The sub-categories differ regarding the complexity of the required interaction technique. Some of the reviewed interaction techniques used for 2D took advantage of the planar arrangement to increase precision or overall usability. For instance, a 2D menu can be displayed on the user's own arm to provide passive haptic feedback when touching buttons [87] or on the floor to be selectable via foot to prevent arm fatigue [85].

For 3D selection, we observed different kinds of methods, mainly used depending on the distance between the user and the objects. Near objects were often selected directly by letting the user's hand or a hand-held device collide with the virtual object, like grabbing or touching it [59,63]. In contrast, distant objects were often selected via *pointing*, a method to select via ray-cast, starting either from the user's head or eyes (gaze) [63], or hand (laser-pointer metaphore) [83]. Both methods typically require the user to perform a confirmation to finally select the indicated object.

Alternatively to the selection of distinct objects, a 2D area or 3D range can be selected. Such techniques are deployed for high-level use cases such as selecting a region inside a 3D data visualization for subsequent filtering operations [51], or defining a viewport [76].

3.1.3 Geometric Manipulation

Based on the results of our literature review, geometric manipulation is a commonly used group of tasks ($N = 74$). It comprises the manipulation of an object's position (*translation*), its orientation (*rotation*), and its size (*scale*). While translation and rotation are rigid transformations and can also be applied to physical objects, the possibility to scale objects is an inherent feature of virtual environments.

In our literature review, we observed that interaction techniques often combine different geometric manipulations. We identified two variants of combinations: an interaction technique can either support multiple manipulations simultaneously or sequentially. For instance, Gao et al. [57] evaluated a 6-degrees of freedom (DOF) manipulation technique that allows a user to simultaneously rotate and translate a virtual object by mapping the position and orientation of a handheld controller to the object [57]. Chaconas et al. [50] evaluated different two-handed gestures to rotate and scale objects. In their experiment, they included techniques that allow the user to simultaneously rotate and scale an object with a single hand movement as well as techniques where the user defines which manipulation they want to perform at the beginning of the gesture — either based on the hand's starting position or by the first detected moving direction after starting the gesture [50]. However, since we consider an interaction

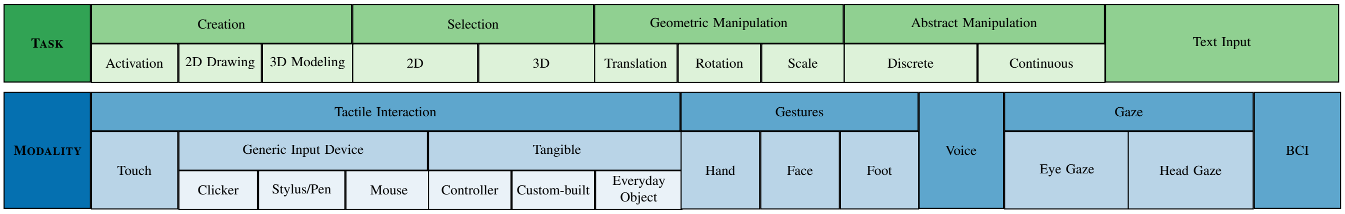

Figure 2: Our derived taxonomy to categorize immersive AR interaction techniques. We examined two dimensions - *Task* and *Modality*. Both dimensions contain a set of characteristics. For the task dimension, the characteristics are mutually exclusive.

technique as a combination of exactly one task and one or more modalities, such combinations are not reflected in our taxonomy.

Interaction techniques can also limit the supported DOF of single manipulations, which can be used to prevent undesired behavior. For example, we reviewed interaction techniques to move furniture by limiting translations to the floor plane [63], and to set a clock's hand by restricting the rotation to one axis [88]. In contrast, unlimited interaction techniques were typically applied to abstract geometric shapes like cubes [64] or tetrahedrons [57].

Geometric manipulations are not limited to 3D objects. Especially translation can also be applied to interact with UI elements, for example, by scrolling or swiping through 2D content, or by moving slider handles [87, 89]. Further, interaction techniques can extend the action provided by the user by automatically applying manipulations, as Lee et al. [67] presented. Their approach enables a user to translate, rotate, and scale 2D windows. When releasing the window, it is projected (translated) to a surface while simultaneously being scaled, keeping the apparent size for the user unchanged [67].

3.1.4 Abstract manipulation

While geometric manipulations interact with virtual objects rendered in AR, abstract manipulations cover all user commands which are not directly related to visualized objects. Based on the literature review, common use cases for abstract manipulations are actions to change the internal system state, for example, commands to load content [58] or to switch between different modes [74, 77], and actions to interact with physically existing objects like internet of things (IoT) devices [72, 82] while the AR system is used to supplement the user's view with feedback or further information.

We identified two subcategories that differ regarding the provided input data type. A *discrete* command directly triggers a fixed function (e.g., switch to menu mode [74], turn on a lamp [82]). In contrast, for *continuous* control, an ongoing function is applied as long as the user performs the corresponding action, like gradually raising the volume of a speaker [82].

3.1.5 Text input

While text input is a common task for traditional desktop interfaces, our literature review only included two papers presenting corresponding techniques [68, 90]. Theoretically, text input techniques could be broken down to fundamental tasks presented in this taxonomy. For instance, a virtual keyboard could be considered as a 2D selection technique, where the user selects multiple buttons successively. However, such a classification would strongly depend on the modality and type of visualization used. Instead, we included text input as a characteristic to our taxonomy that covers all kinds of techniques that enable a user to enter characters, words, or sentences.

3.2 Modalities

Here, we will introduce the modalities that were identified during our iterative literature review. For each modality, we will provide a brief definition as well as some insights into the observed variants using the example of selected research papers.

3.2.1 Tactile interaction

Since AR allows to superimpose virtual content onto the real-world surroundings of a user, physical objects can be naturally involved in the interaction. Related analytical reviews use varying terms for such interaction techniques, as for instance *contact interaction* [30] or *physical controls* [35]. To not confuse our definition with the partially different characteristics of those related concepts, we introduce a new category of *tactile interaction* into our taxonomy, including the subcategories *touch*, *generic input devices*, and *tangibles*.

Touch For interaction techniques that require the user to contact a surface with their bare hands, we introduced the modality of touch. We identified various kinds of surfaces acting as touch recipients. Commonly known, even besides the field of AR research, are touch screen devices like smartphones [62, 74]. Besides, the deployment of various sensors allows developers to turn passive surfaces into touchable, interactive surfaces, like parts of the user's physical environment (e.g., a table [61]) or even the user's own skin [87]. Most of the interaction techniques we reviewed used single touch events as inputs for 2D selection tasks ($N = 6$), however, continuous movements along the surface were also used, for example, for 2D drawing [87] or translation [74, 89].

Generic input devices As a key differentiator between generic input devices and tangibles, we consider the form of multiplexing as suggested by Fitzmaurice [25]. According to the author's definition, all physical input devices can be classified to be either time- or space-multiplexed. Time-multiplexing denotes a dynamic assignment of functions over time and therefore allows the usage of one generic input device for a diverse set of tasks. This characterization applies to all traditional GUI input devices, such as mice and keyboards, as well as to controllers that are delivered with modern VR/AR HMDs.

The reviewed input controllers featured a differing number of DOF, including 2-DOF traditional mice and styli, 3-DOF controllers with positional tracking, 3-DOF controllers with orientational tracking, and 6-DOF controllers with a combination of both tracking types. In addition, devices with discrete input, such as the button of the HoloLens clicker, were used. While we observed such input devices in comparatively many papers, they were mostly used in multimodal input techniques or as baseline conditions for comparing input via other modalities.

Tangibles In contrast to time-multiplexed generic input devices, tangible user interfaces are space-multiplexed, meaning they involve spatially distributed objects with dedicated functions. Due to the spatial nature of such UIs, different tasks can be performed simultaneously, which can be considered as an advantage over the strictly sequential task execution in time-multiplexed UIs.

Strictly following the original definition, tangible UIs have to offer multiple input devices, each with a permanently assigned function. For creating our taxonomy, we loosen these requirements by allowing UIs with (i) single-input devices that are reasonably extendable to space-multiplexed multi-device systems, and (ii) input devices that fulfill the same task for extended periods of time but

Task \ Modality		Tactile Interaction			Gestures			Voice	Gaze		BCI	Total
		Touch	Generic Input Device	Tangible	Hand	Face	Foot		Eye Gaze	Head Gaze		
Creation	Activation	- | -	1 | -	5 | -	3 | 1	- | -	- | -	2 | 3	1 | -	- | 2	- | -	12 | 6
	2D drawing	1 | -	1 | -	1 | -	- | -	- | -	- | -	- | -	- | -	- | -	- | -	3 | -
	3D modeling	- | -	1 | 1	- | -	- | -	- | -	- | -	- | -	- | -	- | -	- | -	1 | 1
Selection	2D	6 | -	3 | 4	2 | -	6 | 5	- | -	1 | -	- | -	1 | 3	2 | 7	- | -	21 | 19
	3D	1 | 2	6 | 3	4 | 4	6 | 7	- | -	- | -	1 | 5	1 | -	1 | 5	1 | -	21 | 26
Geometric Manipulation	Translation	3 | -	6 | -	7 | -	12 | 2	- | -	- | -	2 | 3	- | -	- | 1	- | -	30 | 6
	Rotation	1 | -	6 | -	9 | -	5 | 3	- | -	- | -	1 | 3	- | -	- | 2	- | -	22 | 8
	Scale	1 | -	4 | -	1 | -	5 | 3	- | -	- | -	1 | 2	- | -	- | 1	- | -	12 | 6
Abstract Manipulation	Discrete	1 | -	- | -	3 | -	1 | -	1 | -	- | -	2 | 0	- | -	1 | -	- | -	9 | -
	Continuous	- | -	- | -	1 | -	2 | -	- | -	- | -	- | -	- | -	1 | -	- | -	4 | -
Text Input		1 | -	1 | -	- | -	- | -	- | -	- | -	- | -	- | -	- | -	- | -	2 | -
Total		15 | 2	29 | 8	33 | 4	40 | 21	1 | -	1 | -	9 | 16	5 | 18	3 | 3	1 | -	137 | 72

Table 2: From each reviewed article, we extracted one or more presented interaction techniques. This table shows the frequencies of task-modality combinations examined in these interaction techniques. For each cell, the left and right values indicate the occurrences in unimodal and multimodal interaction techniques, respectively. Hence, the total number of task-modality combinations is higher than the number of interaction techniques. In total, we reviewed 173 interaction techniques: 137 were unimodal, 35 were bimodal, and one included three modalities, resulting in 137 occurrences in unimodal techniques and 72 occurrences in multimodal techniques.

not necessarily for their entire lifespan. These adaptations are meant to take account of the prototypic nature of many research projects (e.g., [51, 81]) and the trend towards reusable everyday tangible objects (e.g., [58, 89]), respectively.

Input devices that do not conform with the extended definition of tangibles are classified as generic input devices, even if stated otherwise in the respective paper (e.g., [53, 64]). By this means, we aim to establish a consistent and transparent methodology for building our taxonomy.

For tangibles, we observed a variety of different custom-built devices ($N = 6$) as well as an emerging trend towards using everyday physical objects ($N = 7$) to interact with the AR environment. By coupling a virtual object and a tangible, the object can be naturally viewed from different angles, hence implementing an intuitive form of rotation and translation (e.g., [48, 58]). Furthermore, tangibles were used for the creation, selection, and abstract manipulation of AR content (see Table 2). Besides these basic tasks, 2 papers that utilized everyday physical objects took advantage of the object's natural affordances, for example, using a physical album to browse through photos [58]. In contrast, 5 papers introduced an artificial mapping between the tangible and its function, such as using a smartphone as a bat [52] or a marker to cycle through playing cards [89].

3.2.2 Gestures

Gestures are commonly considered as a form of nonverbal communication, in which body movements convey a message to interlocutors [31]. While some definitions focus on movements that are performed using the hands only (e.g., [41]), others do not limit the definition to the upper limbs but also include feet and facial gestures. In the reviewed papers, we found instances of all mentioned body parts and therefore use the latter, more broad definition of gestures in our taxonomy. Further, we only categorized interaction techniques to be primarily gesture-based when they are uninstrumented and, in the case of hand gestures, performed in midair. In contrast, techniques using additional input devices or touch input are classified as tactile interactions, as described in Sect. 3.2.1.

In the reviewed literature, hand gestures are the most frequently investigated modality for all selection and manipulation tasks (see Table 2). Authors reason their choice with the interaction being natural and intuitive [63, 67], allowing for direct manipulation of virtual objects that are within arm's reach [49, 84], without the need of additional user instrumentation [79, 85]. Interactions with distant targets can also leverage hand gestures by complementing them with another modality such as gaze [83] or by visually extending the arm [56]. Depending on the focused tasks, we found a preference towards unimanual gestures (selection, translation) or bimanual gestures (rotation, scale). For selection, users either performed a predefined hand sign, such as the air tap, or a pointing movement. In the literature, such gestures are also referred to as *semaphoric* or *deictic* gestures, respectively [41]. In contrast, in *manipulative* gestures, hand movements are directly controlling an object's properties, such as the pose or scale. Since the correlation between the task and the definitions of semaphoric, deictic, and manipulative gestures is very strong, we decided to not include them as subcategories of hand gestures in our taxonomy.

In contrast to the variety of reviewed papers investigating hand gestures, only one paper each was considering foot gestures [85] or facial gestures [72]. In the foot-based interaction technique [85], users can select options of a floor-projected, circular menu using directional motion. The authors report a lower overall workload and higher usability than for a reference condition using hand gestures. However, user feedback suggests that foot-based gestures are particularly useful for indoor scenarios, when actions are performed in front of familiar people, such as family members or colleagues. In comparison, interaction via facial gestures [72] was designed to be particularly subtle and, therefore, could yield a higher social acceptance in public settings.

3.2.3 Voice

Besides the non-verbal interaction methods described before, we identified human speech as another modality in our taxonomy. Voice commands were particularly used for simple tasks, such as choosing from a limited set of discrete modes [77] or for activation and deletion of a virtual object [71, 84, 86].

For more complex tasks, such as selecting from a pool of virtual targets, voice was paired with gaze or pointing gestures and only used as a confirmation [65, 83]. Whitlock et al. [83] even went a step further by suggesting a multimodal interaction technique for transforming virtual objects. In the presented study, users were able to select an object via gaze before initiating a movement or rotation using commands such as ”Move up/down/left/right”. Another voice stop command was necessary to terminate the transformation. However, when compared to hand gestures as well as controller input, voice interaction was shown to be less accurate, slower, and the least enjoyable condition. Another disadvantage of voice commands was illustrated in a Wizard of Oz study by Williams et al. [84]. Participants of their study were asked to manipulate a displayed virtual cube using speech, however, without predetermining a correct command. An analysis of the collected responses showed that there was much disagreement on the most appropriate utterances, showing that voice interfaces have to be learned to a certain extent.

3.2.4 Gaze

While gaze could be considered as a deictic gesture in the broader sense [41], we observed a high number of explicit mentions in the literature and thus introduced a separate category in our taxonomy.

23 of the relevant interaction techniques were considering interaction via head gaze, while only 6 described an eye gaze technique. One reason for this imbalance could be the advantage of head-based input of being a “more affordable surrogate” [55] of eye gaze. While the paper’s authors found that most study participants tended to use the eyes rather than head movements to track objects, they also measured an improved accuracy of head gaze when participants were explicitly asked to follow the targets with their heads.

Only four papers used gaze in a unimodal setting [49, 55, 60, 71], primarily because an additional mechanism is required to not only hover over a scene object but for triggering a selection. This can be either done by using time-based methods, such as dwell [49] or smooth pursuit [55], or by adding a second modality. In our literature review, we found combinations of gaze with hand gestures [50, 63, 66, 69, 71, 83, 85], generic input devices [66], and voice [65, 77, 83].

3.2.5 BCI

Brain-computer interfaces (BCI) convert measured activity of the central nervous system to control commands of an external device [29]. The mapping of neural signals to a meaningful output is challenging since each natural task that is controlled by the brain usually requires the involvement of multiple cerebral regions.

This complexity is also reflected in the fact that only one of our reviewed papers considered this modality for the use in AR interfaces [80]. In the paper, the authors present an operational prototype for mobile robot control using the combination of an EEG headset for input and the HoloLens for augmented output. The EEG signals are scanned for specific patterns that are associated with visual stimulation at certain frequencies. By assigning different frequencies to multiple visual targets, researchers are able to discriminate which target the user is focusing, therefore supporting a 3D selection task.

4 Discussion

Below, we will reflect on the results of the literature review as well as the methodology itself.

4.1 Reflection on Empirical Research

During the review and the concurrent taxonomy design, we observed some trends and possible future directions of research.

Real-world interactions One principal difference between AR and VR is the role of the user’s real environment. While VR aims at a full replacement of all sensory information provided by the real world, AR still allows users to directly involve their physical surroundings. This feature of AR particularly manifested in the excessive use of techniques involving tactile interaction – a modality that was not considered in the well-known taxonomy on 3D UIs by Bowman [18]. 54 of the overall 173 interaction techniques that were identified in our literature review were based on touch or tangibles. Authors praised the intuitive, comfortable, and accurate use of tactile interactions in general (e.g., [53, 73, 87, 89]) as well as the potential of collaborative use of tangibles in particular (e.g., [51, 54, 58]).

Besides physical objects being part of the input mechanism, we also observed a range of tasks that were directly aiming at manipulating the real environment of the user, for example, in the context of smart home systems. AR interactions were used to control inherent properties of real-world objects, such as the brightness of a lamp [82], height of a desk [56], or volume of a speaker [87]. To represent these AR-specific tasks, we introduced the category of *abstract manipulations* in our taxonomy.

Multimodality 36 of the 173 identified interaction techniques involved at least two different modalities. As can be seen in Fig. 3, hand gestures were paired with all different modalities except BCI. Furthermore, head gaze and voice were particularly used in multimodal settings (cf. last row of Table 2). On their own, they have limited power to support complex, fine-grained tasks such as geometric manipulations. However, in combination with other modalities, they act as a natural support for pointing or confirmation actions – two subtasks that are involved in almost every identified interaction. Besides the simultaneous use in pointing-confirmation tasks, multiple modalities can also be used sequentially. For example, voice commands can specify whether a following hand gesture should be interpreted as a rotation or scaling operation [71].

Mobile optical see-through AR In the literature review, we intentionally focused on the period of 2016 to 2021, thus starting with the year of the Microsoft HoloLens launch. This milestone is also reflected in the reviewed literature, to the effect that 29 papers were using an optical-see through (OST) display while only 9 papers draw on video-see through (VST) technology. With 3 mentions, projection-based systems were the least used output technology in the considered research papers (the remaining 3 papers were using VR to simulate AR). Through further technological improvements, this prevalence of OST displays may be even strengthened as some authors mentioned the currently inferior field of view and render quality as the main reasons to decide in favor of a VST device [75].

Another aspect of modern self-contained devices is their suitability for mobile applications since they do not have to be tethered to a powerful workstation. As such, involvement of the current environment of the user is especially preferable, to eliminate the necessity for dedicated input devices [72, 89].

Subtle interaction Besides affecting the technical aspects of interaction techniques, advancements in AR technology also allow developers to have a stronger focus on user experience. We observed attempts to replace expansive movements with more subtle alternatives to increase perceived comfort in public or other social settings. Examples include discreet finger and facial gestures [49, 72, 87] as well as eye gaze [60, 70]. While the easier trackable alternatives, such as hand gestures and head gaze, were still prevalent in the reviewed literature, it can be speculated that we will experience a gradual shift towards the subtle variants with improving technology such as integrated eye trackers.

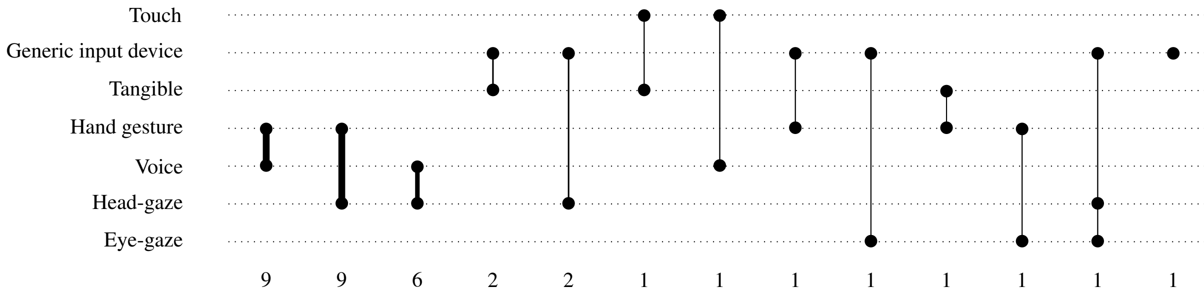

Figure 3: Observed combinations of multiple modalities, each represented by a dot, with the absolute numbers of occurrences (references can be found in the supplementary materials). The last dot represents an interaction technique combining mouse and stylus.

Brain-computer interfaces In the literature review, we encountered one paper using neural signals to select 3D virtual objects in an AR environment [80]. Considering current developments in BCI research, such as the emergence of noninvasive systems [29] and open-source platforms [11], it is reasonable to assume that BCIs become more widespread beyond clinical usage and that more complex AR tasks will be performed based on neural signals.

Application fields Throughout the literature review, we came across a variety of application fields, which the authors envisioned for their developed interaction techniques. The problem statement that was addressed most frequently in the reviewed literature was the interactive visualization of (abstract) data (e.g, [48,51,54,71]). Authors suggested combining the spatial view of 3D data with the precision of tangible devices to focus on different aspects. Another trend was the usage in smart homes (e.g., [55,56,72,82]), which is strongly related to the support of real-world interactions as stated earlier in this section. Finally, we would like to mention interactions with robots [65,77,80,88] and drones [69] that were facilitated by using AR technology to preview future actions of the robot or to provide a direct interface to the system's current state. In summary, we observed a strong connection between the suggested interaction techniques and physical tasks, either in the form of directly controlling properties of the user's real surroundings or by using tactile interactions for working with abstract data.

4.2 Reflection on the Taxonomy Development Method

In this work, we applied Nickerson's [39] iterative methodology, which is widely used in the IS literature, to create a valid and extendable taxonomy. This methodology offers many advantages. For example, a novice researcher can easily use this method and obtain comprehensive information about the literature by following the steps we mentioned above (e.g., the empirical-to-conceptual approach). Due to its iterative nature, it offers a time-efficient solution. Therefore, researchers can get easily familiar with the literature in each iteration. Finally, it provides a valid result: a taxonomy. Even if some of the articles could have been missed in the literature review, the taxonomy itself still serves its purpose and can easily be extended by future work.

Despite Nickerson et al. [39]'s method being our basis for generating a taxonomy, we encountered aspects that are not applicable to our case: (i) limitation of subcategory creation and (ii) mutual exclusivity. We observed that many modalities differ fundamentally and should be categorized in detail (e.g., stylus/pen and clicker) to create a useful taxonomy, leading us to extend the original methodology by subcategories. Moreover, we noticed that several modalities are used as parts of the same interaction technique in the literature (e.g., voice and gestures). Hence, we chose not to use the mutual exclusivity feature of this method to avoid restricting the use of multimodality.

We also note that taxonomies within the field of HCI are often presented without a thorough description of the applied development method. To support transparency, we reported our methodology in detail and, therefore, hope to provide guidance for future reviews.

4.3 Limitations

This work considered 44 articles from the 2016-2021 time frame. Although we kept our search term broad, used an iterative approach, and covered traditional techniques (e.g., the baseline condition of several articles), this may be considered a limitation.

Due to the broadness of our search terms, we limited the included venues by focusing on ACM and IEEE and restricting the scope using the CORE rank. While this ensures that only high-quality venues were included, it does not guarantee the inclusion of all relevant sources.

Regarding our query, we observed many articles mentioning "interact*" in their abstracts but focusing on either feedback or visualization in their full paper. It is likely that some papers did not use this term and, therefore, were not considered in our review. During the process, we noticed that "input" could be used as one of the key terms, e.g., to specifically capture text input articles. Although this category is part of our taxonomy, we speculate that frequencies might be higher when including the term "input" in the query.

At least two researchers were involved in each step [20,33] and the inter-rater reliability was high, yet the researchers could introduce bias because of their perspectives and backgrounds. Following the recommendations of qualitative research practices [15], we reported the backgrounds of the three main researchers in Sect. 2.4.1.

5 Conclusion

In this paper, we introduced a taxonomy for interaction techniques for immersive AR technologies based on an iterative literature review including 44 research papers from the period of 2016 to 2021. Our taxonomy is based on emerging and immersive AR interaction techniques and focused on two identified dimensions – *task* and *modality*. In the process of taxonomy formulation, we put emphasis on the identification of distinctive characteristics for each category, in order to create consistent terms and according definitions. This holistic view on AR interaction techniques is of particular value since we observed partially conflicting definitions of concepts in previous reviews that only focused on specific sub-areas of AR interaction, such as gestures or tangibles.

Our taxonomy can provide important guidelines for the development of future AR interaction techniques. It could serve as a basis for (i) developers to determine common uni- and multimodal methods to perform specific tasks in AR, and (ii) researchers to identify research gaps and, accordingly, potential for future investigations.

Acknowledgments

The authors acknowledge the support of the German Research Foundation, European Union's Horizon 2020, the German Federal Ministry of Education and Research, and the Federal Ministry for Economic Affairs and Energy.

REFERENCES

[1] ACM Conference on Human Factors in Computing (CHI). https://dl.acm.org/conference/chi (Accessed: 28.05.2021).

[2] ACM Guide to Computing Literature. https://dl.acm.org/ (Accessed: 28.05.2021).

[3] ACM Symposium on User Interface Software and Technology (UIST). https://uist.acm.org/ (Accessed: 28.05.2021).

[4] CORE Rankings Portal. https://www.core.edu.au/conference-portal (Accessed: 28.05.2021).

[5] IEEE International Symposium on Mixed and Augmented Reality (ISMAR). https://www.ismar.net/ (Accessed: 28.05.2021).

[6] IEEE Transactions on Visualization and Computer Graphics (TVCG). https://www.computer.org/csdl/journal/tg (Accessed: 28.05.2021).

[7] IEEE Virtual Reality Conference (VR). https://ieeevr.org/ (Accessed: 28.05.2021).

[8] IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). http://www.iros.org/ (Accessed: 28.05.2021).

[9] Institute of Electrical and Electronics Engineers. https://ieeexplore.ieee.org/ (Accessed: 28.05.2021).

[10] Miro. https://miro.com/ (Accessed: 28.05.2021).

[11] OpenBCI Documentation. https://docs.openbci.com/ (Accessed: 28.05.2021).

[12] M. Akçayır and G. Akçayır. Advantages and Challenges Associated With Augmented Reality for Education: A Systematic Review of the Literature. *Educational Research Review*, 20:1–11, 2017. doi: 10.1016/j.edurev.2016.11.002

[13] M. Augstein and T. Neumayr. A Human-Centered Taxonomy of Interaction Modalities and Devices. *Interacting with Computers*, 31(1):27–58, 02 2019. doi: 10.1093/iwc/iwz003

[14] D. Benyon. *Designing Interactive Systems: A Comprehensive Guide to HCI, UX and Interaction Design*. Pearson Edinburgh, 2014.

[15] R. Berger. Now I See It, Now I Don't: Researcher's Position and Reflexivity in Qualitative Research. *Qualitative Research*, 15(2):219–234, 2015. doi: 10.1177/1468794112468475

[16] L. Besançon, M. Sereno, L. Yu, M. Ammi, and T. Isenberg. Hybrid Touch/Tangible Spatial 3D Data Selection. In *Computer Graphics Forum*, vol. 38, pp. 553–567. Wiley Online Library, 2019. doi: 10.1111/cgf.13710

[17] D. A. Bowman and L. F. Hodges. Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages & Computing*, 10(1):37–53, 1999. doi: 10.1006/jvlc.1998.0111

[18] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3d User Interfaces: Theory and Practice*. Addison-Wesley, 2004.

[19] L. Boyajian. 2016: The Year of Augmented Reality, December 2016. https://www.networkworld.com/article/3152394/2016-the-year-of-augmented-reality.html (Accessed: 28.05.2021).

[20] N. Buscemi, L. Hartling, B. Vandermeer, L. Tjosvold, and T. P. Klassen. Single Data Extraction Generated More Errors Than Double Data Extraction in Systematic Reviews. *Journal of Clinical Epidemiology*, 59(7):697–703, 2006. doi: 10.1016/j.jclinepi.2005.11.010

[21] W. Buxton. Lexical and Pragmatic Considerations of Input Structures. *SIGGRAPH Computer Graphics*, 17(1):31–37, 1983.

[22] M. Eckert, J. S. Volmerg, and C. M. Friedrich. Augmented Reality in Medicine: Systematic and Bibliographic Review. *JMIR mHealth and uHealth*, 7(4), 2019. doi: 10.2196/10967

[23] J. Egger and T. Masood. Augmented Reality in Support of Intelligent Manufacturing – A Systematic Literature Review. *Computers & Industrial Engineering*, 140, 2020. doi: 10.1016/j.cie.2019.106195

[24] I. Fernández del Amo, J. A. Erkoyuncu, R. Roy, R. Palmarini, and D. Onoufriou. A Systematic Review of Augmented Reality Content-Related Techniques for Knowledge Transfer in Maintenance Applications. *Computers in Industry*, 103:47–71, 2018. doi: 10.1016/j.compind.2018.08.007

[25] G. W. Fitzmaurice et al. *Graspable User Interfaces*. Citeseer, 1997.

[26] M. Gamer, J. Lemon, M. M. Gamer, A. Robinson, and W. Kendall's. Package 'irr'. *Various Coefficients of Interrater Reliability and Agreement*, 2012.

[27] J. Garzón, J. Pavón, and S. Baldiris. Systematic Review and Meta-Analysis of Augmented Reality in Educational Settings. *Virtual Reality*, 23(4):447–459, 2019. doi: 10.1007/s10055-019-00379-9

[28] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *Transactions on Visualization and Computer Graphics (TVCG)*, 2017.

[29] B. He, H. Yuan, J. Meng, and S. Gao. Brain–Computer Interfaces. In *Neural engineering*, pp. 131–183. Springer, 2020.

[30] A. Karpov and R. Yusupov. Multimodal Interfaces of Human–Computer Interaction. *Herald of the Russian Academy of Sciences*, 88(1):67–74, 2018. doi: 10.1134/S1019331618010094

[31] A. Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004.

[32] L. Land, S. Smith, V. Pang, et al. Building a Taxonomy for Cybercrimes. In *Pacific Asia Conference on Information Systems (PACIS)*, p. 109, 2013.

[33] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)*, 3(CSCW), 2019. doi: 10.1145/3359174

[34] Microsoft. Microsoft HoloLens, March 2016. https://www.microsoft.com/en-us/hololens.

[35] M. R. Mine. Virtual Environment Interaction Techniques. *UNC Chapel Hill CS Dept*, 1995.

[36] S. S. Muhammad Nizam, R. Zainal Abidin, N. Che Hashim, M. C. Lam, H. Arshad, and N. A. Abd Majid. A Review of Multimodal Interaction Technique in Augmented Reality Environment. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4-2):1460, 2018. doi: 10.18517/ijaseit.8.4-2.6824

[37] B. W. Munzer, M. M. Khan, B. Shipman, and P. Mahajan. Augmented Reality in Emergency Medicine: A Scoping Review. *Journal of Medical Internet Research*, 21(4), 2019. doi: 10.2196/12368

[38] Niantic and Nintendo and The Pokémon Company. *Pokémon Go*. Game [AR], July 2016. https://www.pokemongo.com/en-us/.

[39] R. C. Nickerson, U. Varshney, and J. Muntermann. A Method for Taxonomy Development and Its Application in Information Systems. *European Journal of Information Systems*, 22(3):336–359, 2013. doi: 10.1057/ejis.2012.26

[40] S. S. M. Nizam, R. Z. Abidin, N. C. Hashim, M. C. Lam, H. Arshad, and N. Majid. A Review of Multimodal Interaction Technique in Augmented Reality Environment. *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, 8(4-2):8–4, 2018.

[41] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari. Multimodal Human Discourse: Gesture and Speech. *Transactions on Computer-Human Interaction (TOCHI)*, 9(3):171–193, 2002. doi: 10.1145/568513.568514

[42] O. Shaer and E. Hornecker. *Tangible User Interfaces: Past, Present, and Future Directions*. Now Publishers Inc, 2010.

[43] L. Shamseer, D. Moher, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: Elaboration and Explanation. *Bmj*, 349, 2015. doi: 10.1136/bmj.g7647

[44] I. E. Sutherland. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, pp. 757–764, 1968.

[45] P. Vogel, C. Grotherr, C. Kurtz, and T. Böhmann. Conceptualizing Design Parameters of Online Neighborhood Social Networks. In *International Conference on Wirtschaftsinformatik*, 2020. doi: 10.30844/wi_2020_o5-vogel

[46] S. Zollmann, R. Grasset, T. Langlotz, W. H. Lo, S. Mori, and H. Regenbrecht. Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns. *Transactions on Visualization and Computer Graphics (TVCG)*, 2020.

TAXONOMY PAPERS

[47] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh. SymbiosisSketch: Combining 2D & 3D Sketching for Designing De-

tailed 3D Objects in Situ. In *Proc. CHI*, p. 1–15. ACM, 2018.
[48] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *Transactions on Visualization and Computer Graphics (TVCG)*, 24(1):457–467, 2018.
[49] E. Brasier, O. Chapuis, N. Ferey, J. Vezien, and C. Appert. ARPads: Mid-air Indirect Input for Augmented Reality. In *Proc. ISMAR*, pp. 332–343, 2020.
[50] N. Chaconas and T. Höllerer. An Evaluation of Bimanual Gestures on the Microsoft HoloLens. In *Proc. VR*, pp. 1–8, 2018.
[51] M. Cordeil, B. Bach, A. Cunningham, B. Montoya, R. T. Smith, B. H. Thomas, and T. Dwyer. Embodied Axes: Tangible, Actuated Interaction for 3D Augmented Reality Data Spaces. In *Proc. CHI*, p. 1–12. ACM, 2020.
[52] C. Eichhorn, D. A. Plecher, M. Inami, and G. Klinker. Physical Objects in AR Games–Offering a Tangible Experience. In *Proc. VR*, pp. 1801–1806, 2019.
[53] D. Englmeier, J. Dörner, A. Butz, and T. Höllerer. A Tangible Spherical Proxy for Object Manipulation in Augmented Reality. In *Proc. VR*, pp. 221–229, 2020.
[54] B. Ens, S. Goodwin, A. Prouzeau, F. Anderson, F. Y. Wang, S. Gratzl, Z. Lucarelli, B. Moyle, J. Smiley, and T. Dwyer. Uplift: A Tangible and Immersive Tabletop System for Casual Collaborative Visual Analytics. *Transactions on Visualization and Computer Graphics (TVCG)*, 27(2):1193–1203, 2021.
[55] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proc. UIST*, p. 167–178, 2017.
[56] T. Feuchtner and J. Müller. Extending the Body for Interaction with Reality. In *Proc. CHI*, p. 5145–5157. ACM, 2017.
[57] Z. Gao, H. Wang, H. Lv, M. Wang, and Y. Qi. Evaluating the Effects of Non-isomorphic Rotation on 3D Manipulation Tasks in Mixed Reality Simulation. *Transactions on Visualization and Computer Graphics (TVCG)*, 2020.
[58] A. Gupta, B. R. Lin, S. Ji, A. Patel, and D. Vogel. Replicate and Reuse: Tangible Interaction Design for Digitally-Augmented Physical Media Objects. ACM, 2020.
[59] J. Hochreiter, S. Daher, G. Bruder, and G. Welch. Cognitive and Touch Performance Effects of Mismatched 3D Physical and Visual Perceptions. In *Proc. VR*, pp. 1–386, 2018.
[60] B. Huynh, J. Orlosky, and T. Höllerer. In-Situ Labeling for Augmented Reality Language Learning. In *Proc. VR*, pp. 1606–1611, 2019.
[61] P. Jansen, F. Fischbach, J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. ShARe: Enabling Co-Located Asymmetric Multi-User Interaction for Augmented Reality Head-Mounted Displays. In *Proc. UIST*, p. 459–471, 2020.
[62] Y. Jeong, H.-J. Kim, G. Yun, and T.-J. Nam. *WIKA: A Projected Augmented Reality Workbench for Interactive Kinetic Art*, p. 999–1009. 2020.
[63] H. J. Kang, J.-h. Shin, and K. Ponto. A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality. In *Proc. VR*, pp. 275–284, 2020.
[64] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *Transactions on Visualization and Computer Graphics (TVCG)*, 24(2):1038–1048, 2018.
[65] D. Krupke, F. Steinicke, P. Lubos, Y. Jonetzko, M. Görner, and J. Zhang. Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction. In *Proc. IROS*, pp. 1–9, 2018.
[66] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. ACM, 2018.
[67] J. H. Lee, S.-G. An, Y. Kim, and S.-H. Bae. Projective Windows: Bringing Windows in Space to the Fingertip. ACM, 2018.
[68] Z. Li, M. Annett, K. Hinckley, K. Singh, and D. Wigdor. HoloDoc: Enabling Mixed Reality Workspaces That Harness Physical and Digital Content. ACM, 2019.
[69] C. Liu and S. Shen. An Augmented Reality Interaction Interface for Autonomous Drone. In *Proc. IROS*, pp. 11419–11424, 2020.
[70] F. Lu, S. Davari, L. Lisle, Y. Li, and D. A. Bowman. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *Proc. VR*, pp. 930–939, 2020.
[71] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu. Improving Information Sharing and Collaborative Analysis for Remote GeoSpatial Visualization Using Mixed Reality. In *Proc. ISMAR*, pp. 236–247, 2019.
[72] K. Masai, K. Kunze, D. Sakamoto, Y. Sugiura, and M. Sugimoto. Face Commands - User-Defined Facial Gestures for Smart Glasses. In *Proc. ISMAR*, pp. 374–386, 2020.
[73] X. Min, W. Zhang, S. Sun, N. Zhao, S. Tang, and Y. Zhuang. VPModel: High-Fidelity Product Simulation in a Virtual-Physical Environment. *Transactions on Visualization and Computer Graphics (TVCG)*, 25(11):3083–3093, 2019.
[74] E. Normand and M. J. McGuffin. Enlarging a Smartphone with AR to Create a Handheld VESAD (Virtually Extended Screen-Aligned Display). In *Proc. ISMAR*, pp. 123–133, 2018.
[75] H. Peng, J. Briggs, C.-Y. Wang, K. Guo, J. Kider, S. Mueller, P. Baudisch, and F. Guimbretière. RoMA: Interactive Fabrication with Augmented Reality and a Robotic 3D Printer. ACM, 2018.
[76] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. ACM, 2019.
[77] C. P. Quintero, S. Li, M. K. Pan, W. P. Chan, H. Machiel Van der Loos, and E. Croft. Robot Programming Through Augmented Trajectories in Augmented Reality. In *Proc. IROS*, pp. 1838–1844, 2018.
[78] J. S. Roo and M. Hachet. One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities. In *Proc. UIST*, p. 787–795, 2017.
[79] K. A. Satriadi, B. Ens, M. Cordeil, B. Jenny, T. Czauderna, and W. Willett. Augmented Reality Map Navigation With Freehand Gestures. In *Proc. VR*, pp. 593–603, 2019.
[80] H. Si-Mohammed, J. Petit, C. Jeunet, F. Argelaguet, F. Spindler, A. Évain, N. Roussel, G. Casiez, and A. Lecuyer. Towards BCI-Based Interfaces for Augmented Reality: Feasibility, Design and Evaluation. *Transactions on Visualization and Computer Graphics (TVCG)*, 26(3):1608–1621, 2020.
[81] S. Y. Ssin, J. A. Walsh, R. T. Smith, A. Cunningham, and B. H. Thomas. GeoGate: Correlating Geo-Temporal Datasets Using an Augmented Reality Space-Time Cube and Tangible Interactions. In *Proc. VR*, pp. 210–219, 2019.
[82] Y. Sun, A. Armengol-Urpi, S. N. R. Kantareddy, J. Siegel, and S. Sarma. MagicHand: Interact With Iot Devices in Augmented Reality Environment. In *Proc. VR*, pp. 1738–1743, 2019.
[83] M. Whitlock, E. Harnner, J. R. Brubaker, S. Kane, and D. A. Szafir. Interacting with Distant Objects in Augmented Reality. In *Proc. VR*, 2018.
[84] A. S. Williams, J. Garcia, and F. Ortega. Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. *Transactions on Visualization and Computer Graphics (TVCG)*, 26(12):3479–3489, 2020.
[85] W. Xu, H.-N. Liang, Y. Zhao, D. Yu, and D. Monteiro. DMove: Directional Motion-Based Interaction for Augmented Reality Head-Mounted Displays. ACM, 2019.
[86] Y.-T. Yue, Y.-L. Yang, G. Ren, and W. Wang. SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing. In *Proc. UIST*, p. 427–436, 2017.
[87] Y. Zhang, W. Kienzle, Y. Ma, S. S. Ng, H. Benko, and C. Harrison. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proc. UIST*, p. 1151–1159, 2019.
[88] Z. Zhang, D. Weng, J. Guo, Y. Liu, and Y. Wang. Toward an Efficient Hybrid Interaction Paradigm for Object Manipulation in Optical See-Through Mixed Reality. In *Proc. IROS*, pp. 3216–3221, 2019.
[89] Q. Zhou, S. Sykes, S. Fels, and K. Kin. Gripmarks: Using Hand Grips to Transform In-Hand Objects into Mixed Reality Input. ACM, 2020.
[90] F. Zhu and T. Grossman. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. ACM, 2020.