A Tool for Organizing Key Characteristics of Virtual, Augmented, and Mixed Reality for Human–Robot Interaction Systems

Synthesizing VAM-HRI Trends and Takeaways

By Thomas R. Groechel, Michael E. Walker, Christine T. Chang, Eric Rosen, and Jessica Zosa Forde

Frameworks have begun to emerge to categorize virtual, augmented, and mixed reality (VAM) technologies that provide immersive, intuitive interfaces to facilitate human–robot interaction (HRI). These frameworks, however, fail to capture key characteristics of the growing subfield of VAM-HRI and can be difficult to consistently apply because of continuous scales. This work builds upon these prior frameworks through the creation of a tool for organizing key characteristics of VAM-HRI systems (TOKCS). The TOKCS discretizes the continuous scales used within prior works for more consistent classification and adds additional characteristics related to a robot's internal model, anchor locations, manipulability, and the system's software and hardware. To showcase the TOKCS's capability, it is applied to the 10 papers from the Fourth International Workshop on VAM-HRI and examined for key trends and takeaways. These trends highlight the expressive capability of the TOKCS while also helping frame newer trends and future work recommendations for VAM-HRI research.

Digital Object Identifier 10.1109/MRA.2021.3138383
Date of current version: 14 January 2022

Background

The need to help identify growing trends within VAM-HRI is evidenced by four consecutive years of a VAM-HRI workshop consistently spanning 60–100+ attendees. This nascent subfield of HRI addresses challenges in mixed reality (MR) interactions between humans and robots, involving applications such as remote teleoperation, mental model alignment for effective partnering, facilitating robot learning, and comparing the capabilities and perceptions of robots and virtual agents. VAM-HRI research is becoming even more accessible to the robotics community, due in part to the widespread availability of commercial virtual reality (VR), augmented reality (AR), and MR platforms and the rise of readily accessible 3D game engines for supporting virtual environment interactions.

To showcase the TOKCS's capability, it is applied to the 10 papers from the Fourth International Workshop on VAM-HRI and examined for key trends and takeaways.

To understand what challenges and solutions have been emphasized by this new community, Williams et al. [25] proposed the *reality–virtuality interaction cube* as a tool for clustering VAM-HRI research. The interaction cube is a 3D conceptual framework that captures characteristics about the design elements involved [expressivity of view (EV) and flexibility of control (FC)] as well as the virtuality they implement (from real to fully virtual). While the interaction cube provides a useful lens for roughly characterizing research involving interactive technologies within VAM-HRI, the continuous nature of the cube makes it challenging to exactly position where design elements and environments are within it.

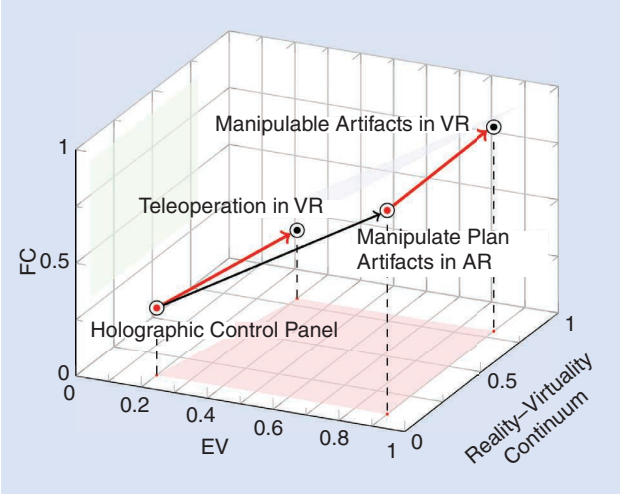

Figure 1. The reality–virtuality interaction cube used to visually categorize MRIDEs according to their FC and EV and where they lie upon the reality–virtuality continuum. Reality is indicated as 0 and virtuality as 1. FC: flexibility of control; EV: expressivity of view.

Furthermore, the interaction cube does not address other characteristics of VAM-HRI research that have recently gained attention, such as robot internal models, software, hardware, and experimental evaluation methods.

To help advance the understanding of different VAM-HRI systems, we introduce the TOKCS. The TOKCS builds off work from the interaction cube, discretizing its continuous scales and adding new key characteristics for classification. The tool is applied to the 10 workshop papers from the Fourth International Workshop on VAM-HRI to validate its usefulness within the growing subfield. These classifications help inform current and future trends found within the workshop and VAM-HRI as a whole.

The Interaction Cube Framework

The interaction cube [25] uses three dimensions to characterize VAM-HRI work: the 2D plane of interaction to represent interactive design elements and the 1D reality–virtuality continuum of Milgram et al. [15] to characterize the environment.

Interaction Design Elements: Enhancing View and Control

The first two dimensions of the interaction cube (Figure 1) are defined by the *plane of interaction*, which captures both 1) the opportunities to view the robot's internal model and 2) the degree of control the human has over the internal model. These two levels of interactivity (termed the *EV* and *FC*, respectively) are the conceptual pillars for characterizing interactivity within the interaction cube, and any components that contribute or impact either EV or FC are called *interaction design elements*. This is similar to the model–view–controller design pattern. However, in this case, the 2D placement on the interaction plane depends on a vector whose direction results from the impact a design element has on EV and the impact a design element has on FC. The magnitude of the vector is scaled by the complexity of the robot's internal model. According to Williams et al. [25], "while it is likely infeasible to explicitly determine the position of a technology on this plane, it is nevertheless instructive to consider the formal relationship between interaction design elements and the position of a technology on this plane."

MR Interaction Design Elements: Anchoring and Artifacts

The interaction cube categorizes the study of VAM virtual objects as MR interaction design elements (MRIDEs), which can fall into one of three categories:

- *User-anchored interface elements*: These are objects attached to a user view, similar to traditional GUI elements that are anchored to the user's camera coordinate frame and do not change along with the user's field of view. These elements may also be referred to as part of a user's heads-up display as popularized by video games and movies.
- *Environment-anchored interface elements*: These objects are anchored to the environment or a robot, for example, virtual arms that can be anchored to a robot [7] or virtual objects that can be anchored to the physical environment.

- *Virtual artifacts*: These are objects that can be manipulated by humans or robots or may move "under their own ostensible volition" [25]. For example, virtual indicators of robot position, such as arrows, can move on their own within the environment.

The Reality–Virtuality Continuum and VAM-HRI

The third axis of the reality–virtuality interaction cube illustrates where an MRIDE falls on the reality–virtuality continuum [15]. This continuum classifies environments and interfaces with respect to how much virtual and/or real content they contain. On one end of the spectrum lies reality, which is any interface that does not use any virtual content and makes use of only real objects and imagery. The opposite end of the spectrum is VR, which would be an interface that consists of pure virtual content without any integration of the real world (for example, a simulated world presented in VR). Between these two extremes is MR, which captures all interfaces that incorporate a portion of both reality and virtuality in their design. There are two subclasses of MR: 1) AR, where virtual objects are integrated into the real world, and 2) augmented virtuality (AV), where real objects are inserted within virtual environments.

AR interfaces in VAM-HRI often communicate the state and/or intentions of a real robot. For example, the battery levels of a robot can be displayed with a virtual object that hovers over a real robot, or a robot's planned trajectory can be drawn on the floor with a virtual line to indicate its future movement intentions.

VR interfaces are often used to provide simulated environments where human users can interact with virtual robots. In these virtual settings, user interactions with robots can be monitored and evaluated without risk of physical harm for either robot or human. Additionally, the virtual robot models can be easily and quickly altered to allow for rapid prototyping of both robot and interface design. Without the need for physical hardware, robots can be added to any virtual scene without the typical costs associated with real robots.

Virtual environments can also be used to teleoperate and/or supervise real robots in the physical world. In cases like these, 3D data collected by the real robot about its surrounding environment are integrated within virtual settings to create AV interfaces. Cyberphysical interfaces and virtual control rooms are two common VAM-HRI AV methods of enhancing a remote robot operator's ability by increasing situational awareness of the robot's state and location while mitigating the limitations of virtual interfaces, such as cybersickness [13].

The TOKCS Classification Framework

The key insight of this work is the addition of key characteristics of VAM-HRI not covered by the interaction cube to create a TOKCS. These include VAM-HRI system hardware, research that seeks to increase the robot's model of the world around it, and additional granularity to MRIDEs. The characteristics are part of the TOKCS, which is then applied to the fourth VAM-HRI workshop papers in the section "Paper Classifications of the Fourth VAM-HRI Workshop." The application informs the insights and future work recommendations outlined in the section "Current Trends and the Future of VAM-HRI."

Hardware

While hardware used for VAM can vary widely, there are certain types of hardware that are commonly used in VAM-HRI. Here we outline the most common, which enable experiences along the reality–virtuality continuum: head-mounted displays (HMDs), projectors, displays, and peripherals. Because hardware technology is making significant advances every year, labeling the specific technology (e.g., HoloLens 2) is important when classifying hardware within the TOKCS. These hardware technologies then fall under these categories.

A key property of virtual object manipulation is the user's action attribution of the manipulation (i.e., whether the user perceives that he/she moved the object, the robot moved the object, or the object moved on its own).

- *HMDs*: VR, MR, and AR all commonly use HMDs. Oculus Quest and HTC Vive both allow for a full VR experience, visually immersing the user in a completely virtual environment. HTC Vive also allows for AV, such as in Wadgaonkar et al. [22], where the user is in a virtual setting but the virtual robot being manipulated is also moving in the real world. The Microsoft HoloLens and the Magic Leap are strictly AR headsets, where virtual images are rendered on top of the real-world view of the user.
- *Projectors*: Onboard projectors can provide a way for the robot itself to display virtual objects or information. Alternately, static projectors allow an area to contain AR elements. Images might be projected onto an object, the floor, or a robot.
- *Displays*: This category of hardware ranges from handheld smartphones or tablets to room-size displays. 2D and 3D monitors fall somewhere in between this range. Some of them exist in a single location, while mobile displays can be carried by a person or moved by a robot. A cave automatic virtual environment immerses the user in VR using three to six walls to partially or fully enclose the space. An AR display might include a real-time camera with overlaid virtual graphics, while a VR display contains completely virtual graphics. Displays can be an especially effective way to conduct user studies without investing in expensive hardware, for example, by showing recorded videos to participants on Amazon Mechanical Turk (MTurk) [18].
- *Peripherals*: Peripheral devices allow for a richer interaction within VR, AR, or MR. Leap Motion hand tracking can be combined with a headset, such as the HTC Vive

(as in [14]), to provide recording and playback of motions and commands. Oculus Quest controllers are handheld and can be used individually or in tandem, giving the user a modality for both gesturing and selecting with the use of buttons on the device. Peripherals may frequently be used to enhance the FC of an MRIDE.

An important component of VAM-HRI research programs is to evaluate and benchmark new approaches by using both objective and subjective metrics.

Software

There is an assortment of software applications for facilitating 3D environments for VAM-HRI research. The most popular platforms, like Unity3D, support a wide variety of VR and MR hardware, like those outlined in the section "Hardware," and offer packages for networking with robot networks like Robotic Operating System (ROS) servers and rendering robot sensor data. ROS also offers a robot simulator, Gazebo, that directly interfaces with ROS applications and has been used for VAM-HRI research. Other additional software generally relevant to HRI research is also included here, such as tracking AR tags to detect object poses using TagUp [1]. Software is not a direct part of the interaction as hardware, but we report relevant software for a holistic understanding of the resources that the VAM-HRI community uses to develop their applications.

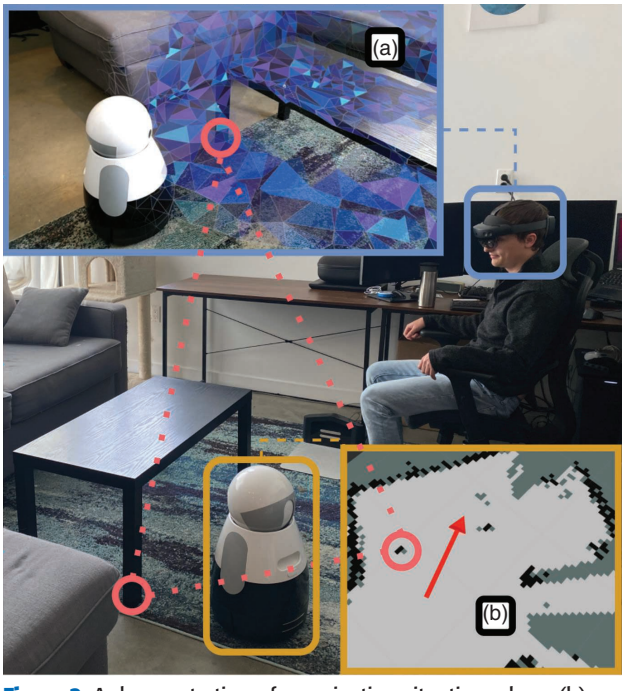

Figure 2. A demonstration of a navigation situation where (b) the robot 2D SLAM map benefits from (a) the 3D SLAM map from the ARHMD. The robot only maps the two front table legs (bottom left) as it is only equipped with a 2D lidar. The robot, however, is too tall to move past the table, so it will collide if it does not use the 3D map from the ARHMD. A combined SLAM map would be created from feature matching such as the table legs (circles).

Robot Internal Complexity of Model

The interaction cube emphasizes the increased EV and FC aspects of projected visual objects on the robot's underlying model. This fails to explore, however, the sensing capabilities and data afforded by VAM technologies [e.g., augmented reality head mounted display (ARMHD)]. The framework can be expanded by including the technologies' ability to aid the robot's internal model of the world—namely, increasing the robot's internal complexity of model (CM). The robot's internal CM benefits from data typically difficult to gather (e.g., eye gaze) as well as the technology affording data assumptions (e.g., a headset with various sensors anchored to the user's head). These data manifest in aiding a robot's model of the environment and/or model of the user.

- *Environment*: Data from the VAM technology further increase the robot's understanding of an environment. An example is provided in Figure 2. Given a mobile robot with 2D SLAM, a 3D map from an ARHMD's SLAM can be transformed into the robot's coordinate frame. The map can then be used for more accurate navigation. In another situation, a mobile phone camera can help with object recognition, both in front of and behind the robot.
- *User*: Data from VAM technology further increase the robot's understanding of the user. For example, a robot can better infer a user's intent to choose an object by using ARHMD eye gaze [20]. Data gathered from motion sensors can be used for both functional purposes (e.g., to determine where the human is in relation to the robot) as well as to infer an affective human state, such as student curiosity [6].

User-Perceived Anchor Locations and Manipulability

The MRIDE categorizations of user-anchored interface elements, environment-anchored interface elements, and virtual artifacts (described in the section "MR Interaction Design Elements: Anchoring and Artifacts") are not mutually exclusive and lack the necessary granularity. For example, a virtual artifact can be user anchored, such as a movable user-anchored element or an environment-anchored object that moves on its own. Granularity can also be added to benefit MRIDE classifications, such as distinguishing between robot- and environment-anchored objects.

To this end, two important distinctions can be added to expand the current framework. First, we apply two characteristics: *Anchor Location {User, Robot, Environment}* and *Perceived Manipulability {User, Robot, None}*. Second, we distinguish MRIDEs based on the intended user perception of the virtual object (i.e., where the user perceives the anchor to be and who can/does move a virtual object).

The first distinction allows for multiple labels within each characteristic, such as objects that are manipulable by both the

robot and the user. Visuals for path planning (see, e.g., [11]) further highlight the benefits of these granular distinctions. A planned robot pose visualized within the environment could be argued to be both robot and environment anchored since the same trajectory can be defined within the robot's local frame of reference or within a global frame of reference.

The latter distinction is important when characterizing Anchor Location as any object can be translated into the environment's coordinate frame. This translation may mathematically hold truth, but the intended perception is important to the goals of studying a virtual object's effect on the user in the interaction. For example, the granularity of Anchor Location combined with the intended user perception allows for the labeling of virtual objects intended to be perceived as part of the robot, such as adding virtual robot appendages [7], [21]. These virtual arms were specifically designed to be perceived as part of the robot to study their impact on the robot's functional and social expressivity, respectively. Therefore, labeling the study of virtual arms as anchored to the environment or user does not help when grouping and looking for trends among different research projects.

Further examining this topic, a key property of virtual object manipulation is the user's action attribution of the manipulation (i.e., whether the user perceives that he/she moved the object, the robot moved the object, or the object moved on its own). Perceived Manipulability is this action attribution, the perception the user has of the manipulation. For an object that the user manipulates (e.g., grabs), the Perceived Manipulability is the user. Virtual objects "manipulated" by the robotic system, however, are neither necessarily directly manipulated by the robot nor perceived as so. In such a case, the virtual object may be scripted to move on its own to give the illusion of robot manipulation yet may fail in its illusion. When researching social robotics, this may have significant consequences on a user's perception of the robot (e.g., the robot's social presence). Therefore, to alleviate this complication, as stated previously, the TOKCS is applied from the intended user perception of the designed system (i.e., if the system attempts an illusion of robot manipulation of a virtual object, it is classified under *Perceived Manipulability: Robot*).

Lastly, these MRIDE labels are only applied to virtual objects and are not tied to classifying VAM-HRI research under model, view, and control, as described in the sections "Interaction Design Elements: Enhancing View and Control" and "Robot Internal Complexity of Model." VAM-HRI studies a variety of modalities provided by VAM technologies. HMD data used for improving a robot's SLAM, for example, still firmly sit under increasing the robot's internal CM but are not applicable under Anchor Location or Perceived Manipulability. Thus, these MRIDE characteristics are designed for and applied only to virtual objects within VAM-HRI.

Framework Limitations

The TOKCS framework was designed to capture and classify the key characteristics of VAM-HRI systems at the time of writing. However, the framework may ultimately be incomplete as advancements in both VAM-HRI research and VAM technology capabilities lead to currently nonexistent key characteristics differentiating VAM-HRI systems of the future. As the field of VAM-HRI advances, the classification framework will likely need to grow as well.

Paper Classifications of the Fourth VAM-HRI Workshop

The TOKCS consists of characterizing VAM-HRI systems according to the following classifications:

- *Anchor Location {User, Env, Robot}*: This indicates where the intended user perception of the virtual object's coordinate frame anchor is located (see the section "User-Perceived Anchor Locations and Manipulability").
- *Perceived Manipulability {User, Robot, None}*: This is the intended user perception of "who" is able to or is currently manipulating the virtual object (see the section "User-Perceived Anchor Locations and Manipulability").
- *Increases EV {0,1}*: VAM technology is used to more explicitly show a robot's internal model, such as using virtual objects to visualize robot sensors (see the section "Interaction Design Elements: Enhancing View and Control").
- *Increases FC {0,1}*: This refers to the use of VAM technology to add control modality to a robot (see the section "Interaction Design Elements: Enhancing View and Control").
- *Increases CM {0,1}*: This is the use of VAM technology to help the robot's understanding of the environment and/or the interaction (see the section "Robot Internal Complexity of Model").
- *Milgram continuum {AR, AV, VR}*: This classifies which form of virtuality is being used (see the section "The Reality–Virtuality Continuum and VAM-HRI").
- *Hardware description*: This describes which VAM technology is used (see the section "Hardware").
- *Software description*: This describes which VAM software is used (see the section "Software").

We apply the TOKCS to papers from the Fourth International Workshop on VAM-HRI to understand the ways in which researchers have been developing new technologies that leverage VAM. The 10 papers and their categorizations within the TOKCS are summarized in Table 1.

Within these 10 papers, a variety of contributions was observed. In most cases, a given system focused its improvements on a specific dimension of the TOKCS; five of the 10 papers developed improvements within a single dimension. The two that contributed expansions along all three axes leveraged AR/VR in a domain that had previously not utilized AR/VR. Higgins et al. [9] developed a method for training grounded-language models in VR, instead of with real-world robots. Ikeda and Szafir [10] leveraged AR headsets for robotic debugging, where previous methods had used 2D screens. Four papers of the 10 increased EV, four increased FC, and three improved upon the robot internal CM. Of these papers, half can be described as VR, three are AV, and two are AR. The majority of methods are anchored at the environment

Table 1. Summary of the TOKCS.

Paper	Anchor Location	Perceived Manipulability	EV	FC	CM	Milgram Continuum [15]	Software	Hardware
Boateng and Zhang [2]	Robot, Env		↑			AR	Unity	HoloLens video recordings via MTurk
Ikeda and Szafir [10]	Env	User	↑	↑	↑	AR	Unity	HoloLens
LeMasurier et al. [11]	Env, Robot	User		↑		AV	Unity, ROSNET, ROS	HTC Vive
Puljiz et al. [19]					↑	AV	Unity	HoloLens
Wadgaonkar et al. [22]	Env, Robot		↑			AV	Unity	HTC Vive
Barentine et al. [1]	Env			↑		VR	Unity, TagUp	Oculus Quest VR headset and controllers
Higgins et al. [9]	User	User	↑	↑	↑	VR	Unity, ROS#, ROS, Gazebo	SteamVR headset
Mara et al. [14]	Env	Robot, User				VR	Unity	HTC Vive Pro Eye and Leap Motion
Mimnaugh et al. [17]						VR	Unity	Oculus Rift S
Mott et al. [18]	Env, User		↑			VR	Unity	MTurk Web Video of VR

Up arrow symbols (↑) indicate that the work increases the functionality within this aspect of the TOKCS. Blank entries indicate that the contributions of the paper for this aspect are on par with prior work. Env: environment

level. The anchors of two methods are located at the robot, and two are located at the user. If a perceived manipulable is available, it is typically available at the user level.

We also observe a broad range of utilized hardware and software. Unity was overwhelmingly popular among papers as the 3D game engine of choice; nine of the 10 papers explicitly mention Unity3D. The most popular HMD mentioned was the HoloLens, which was used in three of the papers. Oculus Quest, HTC Vive, and MTurk are each used in two of the 10 papers.

Table 2. A description of objective and subjective metrics in the fourth VAM-HRI workshop papers.

Paper	Objective Metrics	Subjective Metrics
Boateng and Zhang [2]		NASA TLX; identification of robot position, orientation, and movement
Ikeda and Szafir [10]		SUS; think-out-loud process
Wadgaonkar et al. [22]		Post-experiment interviews; custom survey questions
Higgins et al. [9]	Task accuracy; amount of training data	Custom survey questions
Mara et al. [14]	Task completion time; task completion rate	Custom survey questions
Mimnaugh et al. [17]		Custom survey questions
Mott et al. [18]		Custom survey questions

Blank spaces indicate a lack of metric of that type for that paper. Papers omitted from the table did not report metrics. TLX: task load index; SUS: System Usability Scale.

Evaluations: Subjective and Objective Metrics

In addition to the TOKCS, we further evaluated measures and metrics applied to VAM-HRI research. An important component of VAM-HRI research programs is to evaluate and benchmark new approaches by using both objective and subjective metrics. *Objective* metrics are any metrics that can be directly determined through sensors or measurements and do not involve a human's subjective experience. Examples of objective metrics include task completion time, the number of successful and failed trials, and accuracy and precision of visualization alignment.

Subjective metrics are any metrics that depend on the perceived experience of the users involved. Examples of subjective metrics include the mental workload, levels of immersiveness, and perceived system usability. Both subjective and objective metrics are important and complementary benchmarks for determining how effective new VAM-HRI contributions are compared to existing approaches. A wide variety of metrics is available for these measurements, and understanding which metrics VAM-HRI researchers are using helps highlight what aspects of interaction these technologies are improving.

The most popular method of evaluating the effectiveness of a given design was conducting surveys among the study participants. Additional evaluation metrics focused on quantitative performance metrics on an evaluation task and subjective experience (see Table 2). Here we give general

definitions for the categories of metrics used in the VAM-HRI contributions and examples from the contributions on how they implemented that metric for their application.

The following four objective metrics were used in the VAM-HRI contributions:

- *Task accuracy*: This is the proportion of correct predictions to the total number of predictions (e.g., in Higgins et al. [9], task accuracy is measured by the robot's ability to correctly classify the objects referred to by the human).
- *Amount of training data*: This is the amount of training data that are collected or required for a machine learning application (e.g., in Higgins et al. [9], the amount of training data refers to the amount necessary to close the sim2real gap versus learning in reality).
- *Task completion time*: This is the amount of time between tasks or events (e.g., in [14], it is the recorded time between robot signaling and human reaction).
- *Task completion rate*: This is the proportion of successful attempts at a task to the total number of attempts at the task (e.g., in [14], it is the number of successful completions of a minigame in a VR robot game environment).

There were six subjective metrics used in the VAM-HRI contributions:

- *NASA task load index (TLX) [8]*: This is a multidimensional scale for measuring user workload during and after task execution (e.g., in Boateng and Zhang [2], it measures the user workload of situational awareness in proximal human–robot teaming with virtual shadows).
- *Perceived robot identification*: This refers to the user's perceived estimates about the robots in the environment (e.g., in Boateng and Zhang [2], users identified the position, orientation, and movement patterns of an out-of-sight robot member based on virtual shadows).
- *System Usability Scale (SUS)* [3]: This questionnaire measures a user's perceived usability of a system (fitness for purpose) on a seven-point Likert scale ranging from "strongly disagree" to "strongly agree" (e.g., in Ikeda and Szafir [10], the SUS is used to assess the AR robot debugging tool's usability).
- *Think-out-loud process*: In this technique, participants actively voice their thoughts when using an application for researchers to receive real-time feedback (e.g., in Ikeda and Szafir [10], participants talk out loud about their thought process when using the AR robot debugging tool).
- *Interviews*: Researchers ask participants to comment on specific features after using the VAM-HRI applications (e.g., in Wadgaonkar et al. [22], they request participants to comment on which robot features, such as color and texture, impact robot behavioral anthropomorphism in VR).
- *Custom survey questions*: These surveys are similar to interviews, except that users fill out specific custom survey questions that are application and task specific (e.g., in Higgins et al. [9], users are asked about what they found frustrating for training ground language models in VR with simulated robots; in Mimnaugh et al. [17], users reported on VR sickness).

Current Trends and the Future of VAM-HRI

In this article, the Fourth International VAM-HRI Workshop is used as a case study for MRIDE classification and categorization within the reality virtuality interaction cube; however, the papers submitted to this workshop can also be used to exemplify and project current and future trends in the field of VAM-HRI. This growing subfield of HRI is showing promise in enhancing all areas of HRI from robot control (e.g., teleoperation and supervision interfaces) to collaborative robotics and improving teamwork with autonomous systems. The following will cover some of the key insights gathered from this year's workshop that show how VAM-HRI is evolving and improving the field of HRI as a whole.

An Experimental Evaluation of VAM-HRI Systems

Research in HRI heavily features user studies in the evaluation of robotic systems and their interfaces. It has been an ongoing challenge to adequately record and play back human interactions with robots to answer questions such as: "Where was the user looking at X time?," "How close was the human positioned relative to the robot at Y moment?," and "What were the user's joint values when using a new interface, and how are the physical ergonomics evaluated?" As a possible solution to many of these challenges, VAM-HRI allows for unprecedented recording, playback, and analysis of user interactions with virtual or real robots and objects in an experimental setting because of the inherent ability of HMDs (and other devices like a Leap Motion) to record body/hand/head position/orientation and gaze direction from a seemingly limitless number of virtual cameras recording from different angles [24]. This is exemplified at a highly polished level in CoBot Studio [14] (see Figure 3).

The CoBot Studio project unites roboticists, psychologists, artificial intelligence experts, multimodal communication researchers, VR developers, and professionals in interaction and game design.

However, it is interesting to note that, although precise objective measures can be relatively easily gathered from VAM-HRI experiments, only two of the 10 submissions to the fourth VAM-HRI workshop gathered any objective data (see Table 2). The lack of objective measures may be due to a handful of factors, such as the work being in a preliminary stage best suited for a workshop or the research questions being more focused on social responses and subjective opinions from users. Regardless of the reason, we encourage authors of future VAM-HRI submissions to any venue to take full advantage of the objective measurements that VAM-HRI systems inherently provide as objective observations are still

useful for evaluating a multitude of social interactions (e.g., user pose for evaluating body language, user–robot proxemics, and user gaze).

Although VR interfaces have the aforementioned strengths for enhancing experimental evaluation, they have their own set of unique evaluation challenges as well—one of which is the use of online studies with crowdworkers (e.g., on MTurk). HRI in general has made prolific use of online user studies (especially during the COVID-19 pandemic) that take advantage of cheap and readily available participants. However, VAM-HRI heavily draws upon 3D visualizations (as often seen with HMD-based interfaces), which cannot be properly displayed to crowdworkers who lack HMDs and/or 3D monitors.

Additionally, a strength of AR interfaces is that 3D data and visualizations can be rendered contextually in user environments and are able to be observed from any angle desired by the user. VAM-HRI studies that utilize crowdworkers to evaluate VAM interfaces, such as those performed by Mott et al. [18], are restricted to online images and videos viewed by MTurk on 2D monitors that restrict the user's viewpoint to that of prerecorded videos, which does not allow for a true VAM experience. It remains an open question whether results from crowdsourced VAM-HRI studies provide comparable results to VAM-HRI studies run in person since 3D VAM technology is inherently experienced differently than the 2D experiences found on crowdsourcing platforms. Regardless, using crowdworkers still holds value in the early prototyping phases of VAM-HRI research where the initial formulation of object and interaction designs can be evaluated quickly and inexpensively.

VAM-HRI as an Interdisciplinary Study

HRI is well known to be an interdisciplinary field, and VAM-HRI is proving to be no exception. The CoBot Studio project unites roboticists, psychologists, artificial intelligence experts, multimodal communication researchers, VR developers, and professionals in interaction and game design [14]. As the VAM-HRI field grows, it will likely become increasingly common (and needed) to see teams with varied experiences and skill sets contributing to collaborative research.

Research in multirobot systems is an underexplored inspiration for VAM-HRI research in regard to enhancing the CM. VAM technology can be formulated as another robot within a system—a robot with nondeterministic, nondirectly controllable behavior but with a data-rich sensor suite. The frameworks and techniques of the adjacent field may be able to be modified or even directly applied when treating the human user as an autonomous mobile sensor platform, akin to the person being treated as though he/she is another robot in the system. For example, spatial and semantic scene understanding are important perceptual capabilities for both active robots (to navigate their environments) and passive VAM technologies (to localize the user's field of view).

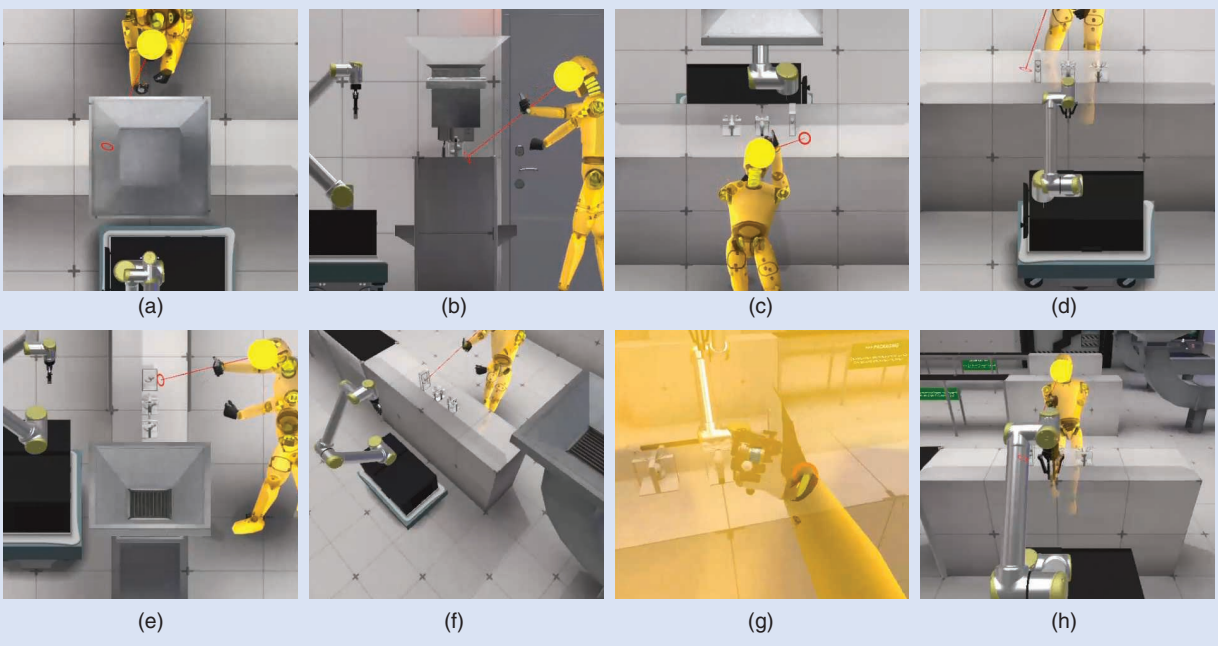

Figure 3. Advances in VAM-HRI research have enhanced the ability to precisely record, play back, and analyze human interactions with robots and other experimental stimuli in controlled user studies. This is exemplified in the CoBot Studio project of Mara et al. [14], where HRI user studies are conducted in a VR environment with numerous virtual cameras monitoring the experimental area from a multitude of angles. (a) Top, (b) side, (c) over the shoulder, (d) top angular, (e) side angular, (f) perspective, (g) user point of view, and (h) robot point of view. These cameras make use of the VR hardware to track body and head motion to record human postures and posture shifts, task-related human movements, gestures, gaze behaviors, and so on. Techniques such as these can benefit the field of HRI as a whole and allow for more complete and feature-rich data of human behavior that would otherwise be lost without VAM-HRI technology and recording techniques.

Additionally, experimentation techniques seen in the field of general VR may aid in administering questionnaires and gathering participant feedback. The typical questionnaires administered by VAM-HRI researchers can be quite jarring for participants who experience extreme context shifts between virtual worlds (where the study took place) and the real world (where the feedback is gathered). This represents a potential confounding factor for participants who no longer visually reference what they are evaluating and may romanticize or incorrectly remember experimental stimuli they can no longer see.

The field of VR has similar challenges, and some studies have started to provide in situ evaluations where questionnaires are posed to users within the virtual environments [12]. We are beginning to see this trend of in situ surveys in VAM-HRI as well. In the CoBot Studio project, surveys are administered within the experiment's virtual setting, removing the confounding factors of 1) reality–virtuality context shifts (having to leave the immersive virtual environment by taking off an HMD to take a midtask survey) and 2) retrospective surveys provided well after exposure to experimental stimuli [14].

However, the cross-disciplinary trends and ideas from the field of VR are not unidirectional; VAM-HRI is currently posed to inform and improve the field of VR in return. Enhancing immersion has always been a primary goal of the field of VR since its inception many decades ago. With the rise of mass-produced consumer-grade HMDs, visual immersion has reached new heights for users around the world. However, the challenge of providing physical immersion through the use of haptics has largely remained an open question: How can a user reach out and touch a dynamic character in a virtual world? Research in VAM-HRI has proposed a potential solution for dynamic haptics, where robots mimic the poses and movements of virtual dynamic objects. Work by Wadgaonkar et al. [22] exemplifies the notion of VAM-HRI supporting the field of VR with robots acting as dynamic haptic devices and allowing users to touch characters in virtual worlds and further enhance immersion in VR settings.

Advancements in VAM-HRI

A strength of VAM-HRI is the ability to alter a robot's morphology with virtual imagery. This technique can take the form of body extensions, where virtual appendages, such as limbs, are added to a real robot [7] or the formation of transformations where the robot's entire morphology is altered, such as transforming a drone into a floating eye [23]. Recent VAM-HRI developments have further expanded upon this idea of changing a real robot's appearance through the aforementioned morphological alterations to include superficial alterations as well, where virtual imagery can be used to change a robot's cosmetic traits. Prior work has demonstrated that robot cosmetic alterations can communicate robot internal states (e.g., robotic system faults) [5]; however, to our knowledge, this is the first time such superficial alterations have been used to manipulate social interactions between human and robot [22].

Although the interactions studied in HRI are typically focused on those of the end user, a lesser studied category of interaction exists: that between robots and their developers and designers. Debugging robots often proves to be a challenging and tedious task; robot faults and unexpected behavior can be hard to understand or explain without parsing through command lines and error logs. To address this issue, prior work in VAM-HRI has used AR interfaces to enhance debugging capabilities [4], [16]. Work by Ikeda and Szafir [10] in VAM-HRI 2021 has built upon these concepts by providing in situ AR visualizations of robot state and intentions, allowing users to better compare robots' plans with their actions when debugging autonomous robots. As AR hardware becomes increasingly intertwined with robotic systems, debugging tools such as these will likely become more commonplace to increase the efficiency and enjoyment of robot design.

VAM technology can be formulated as another robot within a system—a robot with nondeterministic, nondirectly controllable behavior but with a data-rich sensor suite.

Finally, VAM-HRI interfaces have been a popular topic of study within HRI for many years now, and numerous standard methods of interacting with robots through MR or VR have emerged (e.g., AR waypoints for navigation or AR lines for displaying robot trajectories [23]). However, novel methods of interacting with robots are still being designed today, an example of which is persistent virtual shadows, aimed at tackling the issue of knowing a robot's location when it is out of the user's line of sight. Whereas prior solutions have tried using 2D top-down radars for showing robot locations [23], issues remain as interfaces such as these require that repeated context shifts be performed by the user to look at the physical surroundings and then to the radar. Solutions such as persistent virtual shadows circumvent this limitation by embedding robot location data into the user's environment, providing a natural method of displaying a robot's location. This is a location cue that humans have learned to interpret almost subconsciously throughout the course of their lives. Creative advances such as these will continue to emerge in this relatively nascent subfield of HRI, presenting an exciting new future for both VAM-HRI and the field of HRI as a whole.

Acknowledgments

This work was supported by the National Science Foundation under awards IIS-1764092 and IIS-1925083. This work was also supported by the Draper Scholar Program. Any opinions,

findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Draper. Thomas R. Groechel and Michael E. Walker contributed equally to this work.

References

[1] C. Michael Barentine, A. McNay, R. Pfaffenbichler, A. Smith, E. Rosen, and E. Phillips, "Manipulation assist for teleoperation in VR," in *Proc. 3rd Int. Workshop Virtual, Augmented, Mixed-Reality Human-Robot Interact. (VAM-HRI)*, 2021, p. 5, doi: 10.1145/3434074.3447210.

[2] A. Boateng and Y. Zhang, "Virtual shadow rendering for maintaining situation awareness in proximal human-robot teaming," in *Proc. Companion 2021 ACM/IEEE Int. Conf. Human-Robot Interact. (Boulder, CO, USA) (HRI '21 Companion)*. New York, NY, USA: Association for Computing Machinery, pp. 494–498, doi: 10.1145/3434074.3447221.

[3] J. Brooke, "SUS: A 'quick and dirty' usability scale" in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland, Eds. London: Taylor & Francis, 1986, pp. 189-194.

[4] T. Hartnoll, J. Collett, and B. A. Macdonald, "An augmented reality debugging system for mobile robot software engineers," *J. Softw. Eng. Robot.*, vol. 1, no. 1, pp. 18–32, 2010.

[5] F. D. Pace, F. Manuri, A. Sanna, and D. Zappia, "An augmented interface to display industrial robot faults," in *Proc. Int. Conf. Augmented Reality, Virtual Reality Comput. Graphics*, Springer-Verlag, 2018, pp. 403–421, doi: 10.1007/978-3-319-95282-6_30.

[6] T. Groechel *et al.*, "Kinesthetic curiosity: Towards personalized embodied learning with a robot tutor teaching programming in mixed reality," in *Proc. Exp. Robot., 17th Int. Symp.*, Springer Nature, 2021, vol. 19, p. 245, doi: 10.1007/978-3-030-71151-1_22.

[7] T. Groechel, Z. Shi, R. Pakkar, and M. J. Matarić, "Using socially expressive mixed reality arms for enhancing low-expressivity robots," in *Proc. 2019 28th IEEE Int. Conf. Robot Human Interactive Commun. (RO-MAN)*, pp. 1–8, doi: 10.1109/RO-MAN46459.2019.8956458.

[8] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, Sage, Los Angeles, CA, USA, 2006, vol. 50, pp. 904–908, doi: 10.1177/154193120605000909.

[9] P. Higgins *et al.*, "Towards making virtual human-robot interaction a reality," in *Proc. 3rd Int. Workshop Virtual, Augmented, Mixed-Reality Human-Robot Interact. (VAM-HRI)*, 2021, p. 5, doi: 10.1145/nnnnnnn.nnnnnnn.

[10] B. Ikeda and D. Szafir, "An AR debugging tool for robotics programmers," in *Proc. HRI 2021I Workshop VAM-HRI*, 2021, doi: 10.1145/1122445.1122456.

[11] G. LeMasurier, J. Allspaw, and H. A. Yanco, "Semi-autonomous planning and visualization in virtual reality," 2021, arXiv:2104.11827.

[12] L. Lin *et al.*, "The effect of hand size and interaction modality on the virtual hand illusion," in *Proc. 2019 IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, pp. 510–518, doi: 10.1109/VR.2019.8797787.

[13] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 179–186, 2017, doi: 10.1109/LRA.2017.2737046.

[14] M. Mara *et al.*, "CoBot studio VR: A virtual reality game environment for transdisciplinary research on interpretability and trust in human-robot collaboration," in *Proc. HRI 2021I Workshop VAM-HRI*, 2021.

[15] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Proc. Telemanipulator Telepresence Technol.*, 1995, vol. 2351, pp. 282–292, doi: 10.1117/12.197321.

[16] A. G. Millard *et al.*, "ARDebug: An augmented reality tool for analysing and debugging swarm robotic systems," *Frontiers Robot. AI*, vol. 5, no. 2018, p. 87, 2018, doi: 10.3389/frobt.2018.00087.

[17] K. J. Mimnaugh, M. Suomalainen, I. Becerra, E. Lozano, R. Murrieta, and S. LaValle, "Defining preferred and natural robot motions in immersive telepresence from a first-person perspective," in *Proc. 3rd Int. Workshop Virtual, Augmented, Mixed-Reality Human-Robot Interact. (VAM-HRI)*, 2021, p. 4.

[18] T. Mott, T. Williams, H. Zhang, and C. Reardon, "You have time to explore over here!: Augmented reality for enhanced situation awareness in human-robot collaborative exploration," in *Proc. HRI 2021I Workshop VAM-HRI*, 2021.

[19] D. Puljiz, B. Zhou, K. Ma, and B. Hein, "HAIR: Head-mounted AR intention recognition," in *Proc. 3rd Int. Workshop Virtual, Augmented, Mixed-Reality Human-Robot Interact. (VAM-HRI)*, 2021.

[20] E. Rosen, D. Whitney, M. Fishman, D. Ullman, and S. Tellex, "Mixed reality as a bidirectional communication interface for human-robot interaction," in *Proc. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 11,431–11,438, doi: 10.1109/IROS45743.2020.9340822.

[21] N. Tran, K. Mizuno, T. Grant, T. Phung, L. Hirshfield, and T. Williams, "Exploring mixed reality robot communication under different types of mental workload," in *Proc. Int. Workshop Virtual, Augmented, Mixed Reality Human-Robot Interact.*, 2020, vol. 3.

[22] C. P. Wadgaonkar, J. Freischuetz, A. Agrawal, and H. Knight, "Exploring behavioral anthropomorphism with robots in virtual reality," in *Proc. HRI 2021I Workshop VAM-HRI*, 2021.

[23] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proc. 2018 ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 316–324, doi: 10.1145/3171221.3171253.

[24] T. Williams, L. Hirshfield, N. Tran, T. Grant, and N. Woodward, "Using augmented reality to better study human-robot interaction," in *Proc. Int. Conf. Human-Comput. Interact.*, Springer-Verlag, 2020, pp. 643–654.

[25] T. Williams, D. Szafir, and T. Chakraborti, "The reality-virtuality interaction cube: A framework for conceptualizing mixed-reality interaction design elements for HRI," in *Proc. 2nd Int. Workshop Virtual, Augmented, Mixed Reality HRI*, 2019, pp. 520–521, doi: 10.1109/HRI.2019.8673071.

Thomas R. Groechel, University of Southern California, Los Angeles, California, 90007, USA. Email: groechel@usc.edu.

Michael E. Walker, University of Colorado at Boulder, Boulder, Colorado, 80309, USA. Email: michael.walker-1@colorado.edu.

Christine T. Chang, University of Colorado at Boulder, Boulder, Colorado, 80309, USA. Email: christine.chang@colorado.edu.

Eric Rosen, Brown University, Providence, Rhode Island, 02906, USA. Email: eric_rosen@brown.edu.

Jessica Zosa Forde, Brown University, Providence, Rhode Island, 02906, USA. Email: jessica_forde@brown.edu.