

TOKCS: Tool for Organizing Key Characteristics of VAM-HRI Systems

Thomas R. Groechel ^{*1}, Michael E. Walker ^{*2}, Christine T. Chang², Eric Rosen³,
and Jessica Zosa Forde³

¹U. of Southern California , groechel@usc.edu

²U. of Colorado Boulder {michael.walker-1,christine.chang}@colorado.edu

³Brown University {eric_rosen,jessica_forde}@brown.edu

Abstract

Frameworks have begun to emerge to categorize Virtual, Augmented, and Mixed Reality (VAM) technologies that provide immersive, intuitive interfaces to facilitate Human-Robot Interaction. These frameworks, however, fail to capture key characteristics of the growing subfield of VAM-HRI and can be difficult to consistently apply due to continuous scales. This work builds upon these prior frameworks through the creation of a Tool for Organizing Key Characteristics of VAM-HRI Systems (TOKCS). 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 tool's capability, TOKCS is applied to the ten papers from the fourth VAM-HRI workshop and examined for key trends and takeaways. These trends highlight the expressive capability of TOKCS while also helping frame newer trends and future work recommendations for VAM-HRI research.

1 Introduction

The need to help identify growing trends within Virtual, Augmented, and Mixed Reality for Human Robot Interaction (VAM-HRI) is evidenced by four consecutive years of a VAM-HRI workshop consistently spanning 60-100+ attendees. This nascent sub-field of HRI addresses challenges in mixed reality 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 wide-spread availability of commercial virtual reality (VR), augmented reality (AR), and mixed reality (MR) platforms and the rise of readily-accessible 3D game engines for supporting virtual environment interactions.

To understand what challenges and solutions have been focused on 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 three-dimensional conceptual framework that captures characteristics about the design elements involved (expressivity of the view and flexibility of control) 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 the cube. Furthermore, the Interaction cube does not address

*equal contribution

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 a **Tool for Organizing Key Characteristics of VAM-HRI Systems (TOKCS)**. 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 4th 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.

2 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 from Milgram [15] to characterize the environment.

2.1 Interaction Design Elements: Enhancing View and Control

The first two dimensions of the Interaction Cube (Fig. 1) are defined by the *Plane of Interaction*, which captures both (1) the opportunities to view into 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 **expressivity of view (EV)** and **flexibility of controller (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.”

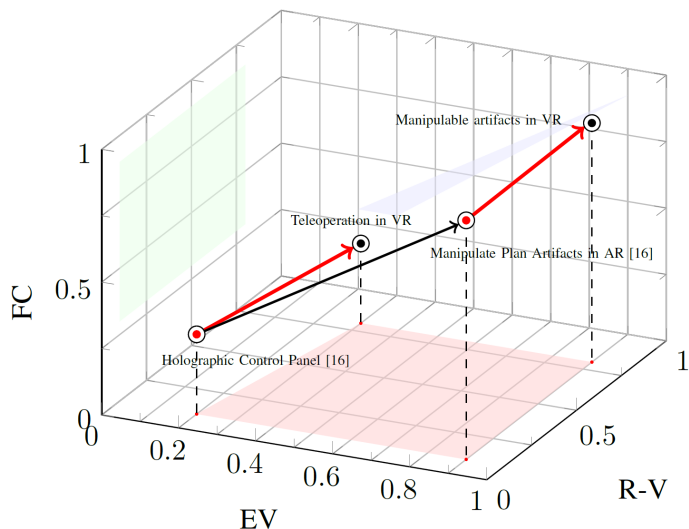


Figure 1: The Reality-Virtuality Interaction Cube used to visually categorize MRIDEs according to their Flexibility of Control (FC), Expressivity of View (EV), and where they lie upon the Reality-Virtuality Continuum (RV). Reality is indicated as 0 and Virtuality as 1.

2.2 Mixed-Reality Interaction Design Elements: Anchoring and Artifacts

The Interaction Cube categorizes the study of VAM virtual objects as MRIDEs (mixed-reality interaction design elements), which can fall into one of three categories:

- *User-Anchored Interface Elements*: Objects attached to user view. This is 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*: Objects anchored to the environment or 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*: 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.

2.3 The Reality-Virtuality Continuum & 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 virtual reality, 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 mixed reality, which captures all interfaces that incorporate a portion of both reality and virtuality in their design. There are two sub-classes of mixed reality: (1) augmented reality where virtual objects are integrated into the real world; and (2) augmented virtuality where real objects are inserted within virtual environments.

Augmented reality 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 the robot’s future movement intentions.

Virtual reality 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 is integrated within virtual settings to create **augmented virtuality** interfaces. Cyber-physical interfaces and virtual control rooms are two common VAM-HRI augmented virtuality methods of enhancing remote robot operators ability by increasing situational awareness of their robot’s state and location while mitigating the limitations of virtual interfaces such as cyber sickness [13].

3 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 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 *mixed-reality* interaction design elements (MRIDEs). The characteristics are part of TOKCS which is then applied to the 4th VAM-HRI workshop’s papers in Sec. 4. The application informs the insights and future work recommendations outlined in Sec. 5.

3.1 Hardware

While hardware used for virtual, augmented, and mixed reality 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 TOKCS. These hardware technologies then fall under these categories.

HMDs. Virtual, mixed, and augmented reality all commonly use head-mounted displays. The Oculus Quest and HTC Vive both allow for a full virtual reality experience, visually immersing the user in a completely virtual environment. The HTC Vive also allows for augmented virtuality, 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 augmented reality 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 augmented reality elements. Images might be projected onto an object, on the floor, or onto a robot.

Displays. This category of hardware ranges from handheld smartphones or tablets to room-size displays. Two-dimensional and three-dimensional monitors fall somewhere in between this range. Some of these exist in a single location, while mobile displays can be carried by a person or moved by a robot. A cave automated virtual environment (or CAVE) immerses the user in virtual reality using 3 to 6 walls to partially or fully enclose the space. An augmented reality display might include a realtime camera with overlaid virtual graphics, while a virtual reality 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 [18].

Peripherals. Peripheral devices allow for a richer interaction within virtual, augmented, or mixed reality. 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 might frequently be used to enhance the Flexibility of Control (FC) of a MRIDE.

3.2 Software

There are a variety 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 Section 3.1, and offer packages for networking with robot networks like ROS servers and rendering robot sensor data. ROS also offers a robot simulator, Gazebo, that directly interfaces with ROS applications and which 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 what resources the VAM-HRI community uses to develop their applications.

3.3 Robot Internal Complexity of Model

The Interaction Cube emphasizes the increased expressivity of view and flexibility of controller aspects of projected visual objects having on the robot’s underlying model. This fails to explore, however, the sensing capabilities and data afforded by VAM technologies (e.g., ARHMD). 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 being 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 increases the robot's understanding of an environment. An example is provided in Fig. 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 and behind the robot.

User - Data from VAM technology further increases 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 both for functional purposes (e.g., where is the human in relation to the robot) as well as used to infer affective human state such as student curiosity [6].

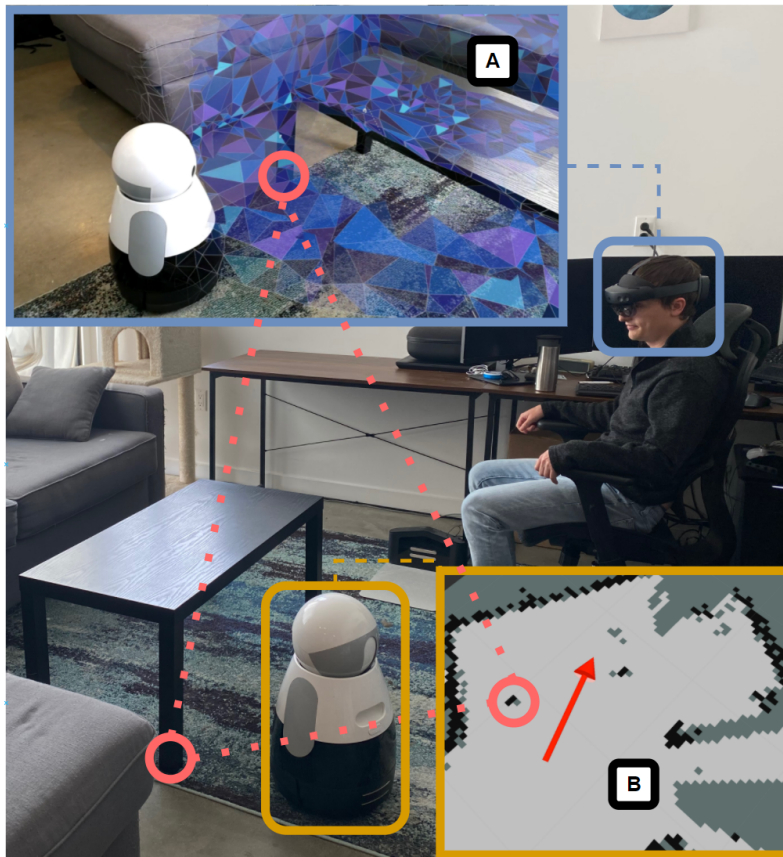


Figure 2: Demonstrates a navigation situation where the robot 2D SLAM map (B) benefits from the 3D SLAM map from the ARHMD (A). The robot only maps the two front table legs (bottom left) as it is only equipped with a 2D lidar. The robot, however, is too tall to move past the table so it will collide 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).

3.4 User Perceived Anchor Locations and Manipulability

The mixed reality interaction design element (MRIDE) categorizations of user-anchored interface elements, environment-anchored interface elements, and virtual artifacts (described in 2.2) are not mutually exclusive and lack 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 does the user perceive 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 (e.g., [11]) further highlight the benefits of these granular distinctions. A planned robot pose visualized within the environment could be argued as 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 intended user perception allows for labeling virtual objects intended to be perceived as part of the robot such as adding virtual robot appendages [21, 7]. 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 toward this idea, a key property of virtual object manipulation is the user’s action attribution of the manipulation (i.e., does the user perceive that they 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 not 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 and as stated above, 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 described in Sec.n 2.1 and 3.3. VAM-HRI studies a variety of modalities provided by VAM technologies. HMD data used for improving a robot’s SLAM, for example, still firmly sits under increasing the robot’s internal complexity of model but is not applicable under Anchor Location nor Perceived Manipulability. Thus these MRIDEs characteristics are designed for and only applied to virtual objects within VAM-HRI.

3.5 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 stand to 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 field of VAM-HRI advances, the classification framework will likely need to grow as well.

4 Paper Classifications of the 4th VAM-HRI Workshop

TOKCS consists of characterizing VAM-HRI systems with the following:

Anchor Location {User, Env, Robot} – where is the intended user perception of the virtual object’s coordinate frame anchor (Sec 3.4);

Perceived Manipulability {User, Robot, None} – the intended user perception of “who” is able to or is currently manipulating the virtual object (Sec. 3.4);

Increases Expressivity of View (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 (Sec. 2.1);

Increases Flexibility of Controller (FC) {0,1} – using VAM technology to add control modality to a robot (Sec. 2.1);

Increases Complexity of Model (CM) {0,1} – using VAM technology to help the robot’s understanding of the environment and/or the interaction (Sec. 3.3);

Milgram Continuum {AR, AV, VR} – classification of which form of virtuality is being used (Sec. 2.3);

Hardware Description – which VAM technology is used (Sec. 3.1)

Software Description – which VAM software is used (Sec. 3.2)

Table 1: Summary of TOKCS. Up arrow symbols (\uparrow) indicate that the work increases the functionality within this aspect of TOKCS. Blank entries indicate that the contributions of the paper for this aspect are on par with prior work.

Paper	Anchor Location	Perceived Manipulability	Expressivity of View	Flexibility of Controller	Complexity of Model	Milgram Continuum [15]	Software	Hardware
Boateng and Zhang [2]	Robot, Env		\uparrow			AR	Unity	Hololens video recordings via MTurk Hololens
Ikeda and Szafer [10]	Env	User	\uparrow	\uparrow	\uparrow	AR	Unity	
LeMasurier et al. [11]	Env, Robot	User		\uparrow		AV	Unity, ROSNET, ROS	HTC Vive
Puljiz et al. [19]					\uparrow	AV	Unity	Hololens
Wadgaonkar et. al [20]	Env, Robot		\uparrow			AV	Unity	HTC VIVE
Barentine et al. [1]	Env			\uparrow		VR	Unity, TagUp	Oculus Quest VR headset & controllers
Higgins et al. [9]	User	User	\uparrow	\uparrow	\uparrow	VR	Unity, ROS#, ROS, Gazebo	SteamVR headset
Mara et al. [14]	Env	Robot, User				VR	Unity	HTC VIVE Pro Eye & Leap Motion
Mimnaugh et al. [17]						VR	Unity	Oculus Rift S
Mott et al. [18]	Env, User		\uparrow			VR	Unity	MTurk Web Video of VR

We apply TOKCS to papers from the 4th International Workshop on VAM-HRI to understand the ways in which researchers have been developing new technologies that leverage virtual, augmented, and mixed reality. The ten papers and their categorization within the TOKCS are summarized in Table 1.

Within these ten papers, a variety of contributions were observed. In most cases, a given system focused its improvements on a specific dimension of the TOKCS; five of the ten 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] leverages AR-headsets for robotic debugging, where previous methods had used 2D screens. Four papers of the ten increased expressivity of view (EV), four increased the flexibility of the controller (FC), and three improved upon the robot internal complexity of model (CM). Of these papers, half can be described as virtual reality, three are augmented virtuality, and two are augmented reality. The majority of methods are anchored at the environment level. Two methods’ anchor is 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 ten 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 ten papers.

4.1 Evaluations: Subjective and Objective Metrics

In addition to 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 metric 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 metric that depends on the perceived experience of the users involved. Examples of subjective metrics include 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 are available for these measurements, and understanding which metrics VAM-HRI researchers are using helps highlight what aspects of interaction these technologies are improving on.

Table 2: Description of objective and subjective metrics in 4th VAM-HRI Workshop papers. Blank spaces indicate a lack of metric of that type for that paper. Papers omitted from the table did not report metrics.

Paper	Objective Metrics	Subjective Metrics
Boateng and Zhang [2]		NASA TLX; Identification of robot position, orientation, and movement
Ikeda and Szafir [10]		System Usability Scale; 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

The most popular method of evaluating effectiveness of a given design was conducting surveys of 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 give examples from the contributions on how they implemented that metric for their application.

There were four objective metrics used in the VAM-HRI contributions: **Task accuracy**: 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**: The amount of training data collected or is 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**: the amount of time between tasks or events (e.g: In [14], the recorded time between robot signalling

and human reaction), **Task completion rate**: The proportion of successful attempts at a task to the total number of attempts at the task (e.g: In [14], the number of successful completions of a minigame in a VR robot game environment).

There were 6 subjective metrics used in the VAM-HRI contributions: **NASA-task load index (NASA TLX)** [8]: a multi-dimensional scale for measuring user workload during and after task execution (e.g: In Boateng and Zhang [2], measuring user workload of situational awareness in proximal human-robot teaming with virtual shadows) , **Perceived robot identification**: user’s perceived estimates about the robots in the environment (e.g: In Boateng and Zhang [2], users identified what position, orientation and movement patterns of a out-of-sight robot member based on virtual shadows), **System Usability Scale (SUS)** [3]: a questionnaire for measuring 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 Szafer [10], SUS is used to assess the AR Robot debugging tool’s usability), **Think out loud process**: participants actively voice their thoughts when using an application for researchers to receive real-time feedback (e.g: In Ikeda and Szafer [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], asking participants to comment on which robot features like color and texture impact robot behavioral anthropomorphism in VR) **Custom survey questions**: similar to interviews, except 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, or in Minnaugh et al. [17] where users reported on VR sickness).

5 Current Trends & the Future of VAM-HRI

In this paper, the 4th 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 sub-field 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 whole.

5.1 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 playback human interactions with robot, 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?’, ‘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 due to 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).

However, it is interesting to note that although precise objective measures can be relatively easily gathered from VAM-HRI experiments only 2 of the 10 submissions to the 4th 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

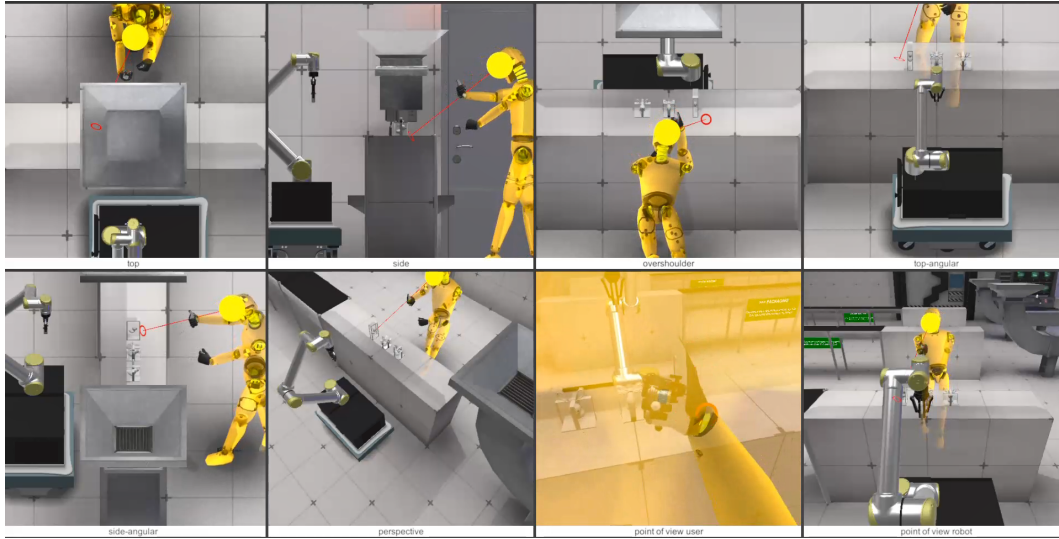


Figure 3: Advances in VAM-HRI research have enhanced the ability to precisely record, playback, and analyze human interactions with robots and other experimental stimuli in controlled user studies. This is exemplified in Mara et. al’s [14] CoBot Studio project where HRI user studies are conducted in a VR environment with numerous virtual cameras monitoring the experimental area from a multitude of angles. 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, and gaze behaviors, etc. Techniques such as this 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.

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, user gaze).

Although virtual reality interfaces have the aforementioned strengths for enhancing experimental evaluation, they have their own set of unique evaluation challenges as well, one of which being use of online studies with crowdworkers (e.g., on Amazon Mechanical Turk). 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 in 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’s 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 Mechanical Turks on 2D monitors that restrict the user’s viewpoint to that of pre-recorded videos which does not allow for a true VAM experience. It remains an open question if 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.

5.2 VAM-HRI as an Interdisciplinary Study

HRI is well known to be an interdisciplinary field and VAM-HRI is showing to be no exception. The CoBot Studio project brings together roboticists, psychologists, AI experts, multi-modal communication researchers, VR developers, and professionals in interaction design 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 multi-robot systems is an underexplored inspiration for VAM-HRI research in regard to enhancing the complexity of model (CM). VAM technology can be formulated as another robot within a system – a robot with non-deterministic, non-directly controllable behavior but one 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 human being treated as though they are another robot in the system. For example, spatial and semantic scene understanding are important perceptual capabilities for active robots (to navigate their environment) and passive VAM technologies (to localize the user’s field of view).

Additionally, experimentation techniques seen in the field of general Virtual Reality may aid in the administering of questionnaires and gathering participant feedback. 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 poses as 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 Virtual Reality 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 mid-task survey); and (2) retrospective surveys provided well after exposure to experimental stimulus [14].

The cross-disciplinary trends and ideas from the field of virtual reality are not unidirectional however; 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 pose 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.

5.3 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 are added to a real robot, such as limbs [7], or form 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 that of the end-user, a lesser studied category of interaction exists, which is that between robots and their developers and designers. Debugging robots often proves to be a challenging and tedious task with robot faults and unexpected behavior being 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 '21 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.

Finally, VAM-HRI interfaces have been a popular topic of study within HRI for many years now, and many standard methods of interacting with robots through MR or VR have emerged (e.g., AR waypoints for navigation or AR lines for displaying robot trajectory [23]). However, novel methods of interacting with robots are still being designed today, an example of which being persistent virtual shadows, aimed at tackling the issue of knowing a robot's location when 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 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 sub-field 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 (NSF) under award 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.

References

- [1] Christian Michael Barentine, Andrew McNay, Ryan Pfaffenbichler, Addyson Smith, Eric Rosen, and Elizabeth Phillips. 2021. Manipulation Assist for Teleoperation in VR. In *Proc. of the 3rd International Workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interactions (VAM-HRI)*.
- [2] Andrew Boateng and Yu Zhang. 2021. Virtual Shadow Rendering for Maintaining Situation Awareness in Proximal Human-Robot Teaming. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (Boulder, CO, USA) (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 494–498.
- [3] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
- [4] Toby Hartnoll Joshua Collett and Bruce Alexander Macdonald. 2010. An augmented reality debugging system for mobile robot software engineers. (2010).

- [5] Francesco De Pace, Federico Manuri, Andrea Sanna, and Davide Zappia. 2018. An augmented interface to display industrial robot faults. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 403–421.
- [6] Thomas Groechel, Roxanna Pakkar, Roddur Dasgupta, Chloe Kuo, Haemin Lee, Julia Cordero, Kartik Mahajan, and Maja J Matarić. 2021. Kinesthetic Curiosity: Towards Personalized Embodied Learning with a Robot Tutor Teaching Programming in Mixed Reality. In *Experimental Robotics: The 17th International Symposium*, Vol. 19. Springer Nature, 245.
- [7] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J Matarić. 2019. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–8.
- [8] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
- [9] Padraig Higgins, Gaoussou Youssouf Kebe, Adam Berlier, Kasra Darvish, Don Engel, Francis Ferraro, and Cynthia Matuszek. 2021. Towards Making Virtual Human-Robot Interaction a Reality. In *Proc. of the 3rd International Workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interactions (VAM-HRI)*.
- [10] Bryce Ikeda and Daniel Szafr. 2021. An AR Debugging Tool for Robotics Programmers. (2021).
- [11] Gregory LeMasurier, Jordan Allspaw, and Holly A Yanco. 2021. Semi-Autonomous Planning and Visualization in Virtual Reality. (2021).
- [12] Lorraine Lin, Aline Normoyle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 510–518.
- [13] Jeffrey I Lipton, Aidan J Fay, and Daniela Rus. 2017. Baxter’s homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters* 3, 1 (2017), 179–186.
- [14] Martina Mara, Kathrin Meyer, Michael Heiml, Horst Pichler, Roland Haring, Brigitte Krenn, Stephanie Gross, Bernhard Reiterer, and Thomas Layer-Wagner. 2021. CoBot Studio VR: A Virtual Reality Game Environment for Transdisciplinary Research on Interpretability and Trust in Human-Robot Collaboration. (2021).
- [15] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. International Society for Optics and Photonics, 282–292.
- [16] Alan G Millard, Richard Redpath, Alistair M Jewers, Charlotte Arndt, Russell Joyce, James A Hilder, Liam J McDaid, and David M Halliday. 2018. ARDebug: an augmented reality tool for analysing and debugging swarm robotic systems. *Frontiers in Robotics and AI* 5 (2018), 87.
- [17] Katherine J Mimnaugh, Markku Suomalainen, Israel Becerra, Eliezer Lozano, Rafael Murrieta, and Steven LaValle. 2021. Defining Preferred and Natural Robot Motions in Immersive Telepresence from a First-Person Perspective. In *Proc. of the 3rd International Workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interactions (VAM-HRI)*.

- [18] Terran Mott, Thomas Williams, Hao Zhang, and Christopher Reardon. 2021. You Have Time to Explore Over Here!: Augmented Reality for Enhanced Situation Awareness in Human-Robot Collaborative Exploration. (2021).
- [19] David Puljiz, Bowen Zhou, Ke Ma, and Björn Hein. 2021. HAIR: Head-mounted AR Intention Recognition. In *Proc. of the 3rd International Workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interactions (VAM-HRI)*.
- [20] Eric Rosen, David Whitney, Michael Fishman, Daniel Ullman, and Stefanie Tellex. 2020. Mixed reality as a bidirectional communication interface for human-robot interaction. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11431–11438.
- [21] Nhan Tran, Kai Mizuno, Trevor Grant, Thao Phung, Leanne Hirshfield, and Tom Williams. 2020. Exploring mixed reality robot communication under different types of mental workload. In *International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction*, Vol. 3.
- [22] Chinmay P Wadgaonkar, Johannes Freischuetz, Akshaya Agrawal, and Heather Knight. 2021. Exploring Behavioral Anthropomorphism With Robots in Virtual Reality. (2021).
- [23] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafr. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.
- [24] Tom Williams, Leanne Hirshfield, Nhan Tran, Trevor Grant, and Nicholas Woodward. 2020. Using augmented reality to better study human-robot interaction. In *International Conference on Human-Computer Interaction*. Springer, 643–654.
- [25] Tom Williams, Daniel Szafr, and Tathagata Chakraborti. 2019. The reality-virtuality interaction cube. In *Proceedings of the 2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI*.