Better Teaming Through Visual Cues: How Projecting Imagery in a Workspace Can Improve Human-Robot Collaboration

Ramsundar Kalpagam Ganesan[1], Yash K. Rathore[2], Heather M. Ross[3] and Heni Ben Amor[2]

***Abstract*— In this paper, we present a communication paradigm using a context-aware mixed reality approach for instructing human workers when collaborating with robots. The main objective of this approach is to utilize the physical work environment as a canvas to communicate task-related instructions and robot intentions in the form of visual cues. A vision-based object tracking algorithm is used to precisely determine the pose and state of physical objects in and around the workspace. A projection mapping technique is used to overlay visual cues on the tracked objects and the workspace. Simultaneous tracking and projection onto objects enable the system to provide just-in-time instructions for carrying out a procedural task. Additionally, the system can also inform and warn humans about the intentions of the robot and safety of the workspace. We hypothesized that using this system for executing a human-robot collaborative task will improve the overall performance of the team and provide a positive experience to the human partner. To test this hypothesis, we conducted an experiment involving human subjects and compared the performance (both objective and subjective) of the presented system with conventional forms of communication, namely printed and mobile display instructions. We found that projecting visual cues enabled human subjects to collaborate more effectively with the robot and resulted in higher efficiency in completing the task.**

I. INTRODUCTION

The ability to quickly understand each other's intentions and goals is a critical element of successful collaboration within human teams. Efficient teaming often emerges as a result of explicit or implicit cues that are shared, recognized, and understood by the participants. Such cues act as signals that maintain trust, situational awareness, and mutual understanding among team members. The ability to communicate intentions through implicit and explicit cues is also of critical importance to fluent human-robot collaboration. As highlighted in the *Roadmap for U.S. Robotics* report, "humans must be able to read and recognize robot activities in order to interpret the robot's understanding" [10]. Especially in close-contact physical interaction scenarios that are safety critical, e.g., collaborative assembly, it is vital that the human partner quickly understand a robot's intentions. Failure to establish such a shared understanding of the situation may lead to potentially lethal accidents. Recent work towards safer human-robot interaction has focused on the generation of legible robot motion [19], as well as the verbalization of robot intentions using natural language [29].

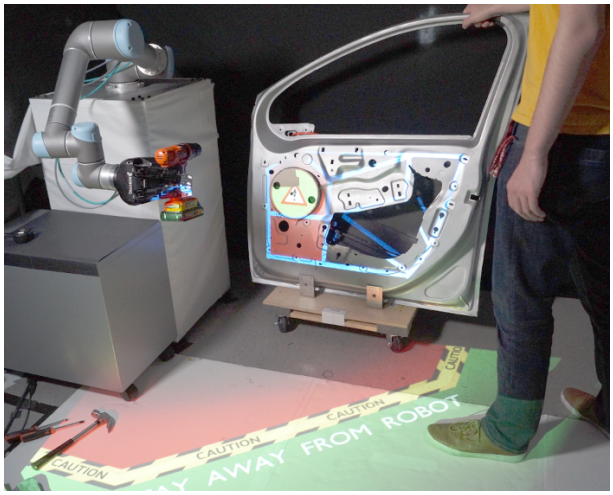


Fig. 1: Signaling during human-robot collaboration by projecting dynamic, visual cues into the environment.

In this work, we describe an alternative communication paradigm that is based on the projection of explicit visual cues. In particular, we propose a context-aware projection method that embeds visual signals within the environment, such that they can be intuitively understood and directly read by the human partner. The physical environment is used as a medium to convey information about the intended actions of the robot, the safety of the work space, or task-related instructions. To this end, a mixed reality system has been developed that combines a vision-based object tracking algorithm with a context-aware projection mapping technique. Visual cues related to the robot and the task being performed are dynamically synthesized and projected. The projection of signals is performed in a just-in-time fashion based on the current state within the joint collaboration plan. An example scenario is shown in Fig. 1.

We introduce a methodology for defining an extensible visual language that contains different categories of cues. The methodology is based on signal categories, similar to *parts of speech* in natural language, from which complex visual messages can be constructed. Following this conceptualization, we propose a domain-specific visual language that covers a reasonable fragment of visual cues related to physical collaboration tasks. Further, we describe a set of new interaction modes, that are enabled by the use of our mixed-reality system and object tracking.

[1] Ramsundar Kalpagam Ganesan is with the School of Electrical, Computer and Energy Engineering, Arizona State University.

[2] Yash Rathore and Heni Ben Amor are with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University.

[3] Heather Ross is with the School for the Future of Innovation in Society, Arizona State University, Tempe, AZ 85281, USA. Email: {ramsundar, ykrathor, hmross1, hbenamor}@asu.edu

We hypothesize that incorporating the proposed system into a complex, sequential human-robot collaborative task can improve the efficiency and effectiveness of the team, and provide satisfaction to the human co-worker in collaborating with the robot. These gains, in turn, will improve the human-robot team fluency and trust. To investigate the validity of this hypothesis, we conducted a study with 15 participants in which human subjects and a stationary manipulator jointly assembled a car door. Throughout the collaboration, human subjects received just-in-time visual signals related to the task. In addition to projecting instructions and information, the system also provided visual feedback on the effectiveness of the task currently being carried out by the human. The results of the experiments were evaluated using a mixed methods approach including quantitative and qualitative criteria to assess accuracy, efficiency, and participant satisfaction.

II. Related Work

Advances in display systems and vision technology have paved the way for incorporating real-time augmented information with physical entities. In robotics, various techniques for visually signaling commands and intentions have been proposed in the past. An early review of the use of AR for human-robot collaboration can be found in [15]. Common to many setups [17], [22], [13], [18], however, is that they display additional information by projecting onto flat surfaces in the environment, e.g., the floor. The surface becomes a replacement for the flat display screen. One of the early attempts to use projections to communicate with the robot was made by [26]. The prototype of their system, "Interactive Hand Pointer" (IHP), consisted of a LCD projector and a real-time vision algorithm to detect and track user hand gestures.

Related research studies have focused on providing a visual platform for human users to directly interact and understand the internal states of robots. [30] presented an approach to communicate navigational intentions using a projector mounted on a robotic wheelchair. The robotic wheelchair projected its future trajectory on the floor, which helped both the passenger and nearby people to navigate safely. The motion of other individuals passing by the wheelchair was significantly smoother with projected intention communication.

In a similar approach, [8] reported that using on-floor projection to visualize the intended path of a mobile robot enhanced human reaction and comfort working in a robotic environment. The subjective experiment showed that the average user rating with the projection system increased by 53% and 65% respectively for the robot moving in straight lines and for taking a sudden turn. Both studies suggest that humans find it more comfortable to interact and work with a robot when its intentions are presented directly as visual cues.

[22] demonstrated an advanced projection system, MARCPS, which augmented the physical laboratory space with real-time status and intentions of drones and ground vehicles in a cyber-physical system. Several other studies have also used projection systems to convey information to the user [27], [17]. However, these systems were confined to displaying on flat surfaces and did not consider the state of physical objects while projecting information.

In contrast to that, our earlier work [1] demonstrated an early prototype of a projection system that tracks physical objects in real-time and projects visual cues at specific spatial locations. A preliminary usability study demonstrated improved effectiveness and user satisfaction with the projection-based approach in a human-robot collaborative task. However, the proposed system at the time was limited to simple tasks like tracking, moving and rotating a single object on a flat surface and the overall collaboration was limited to an interaction of about 1-2 minutes. In contrast to that, we present in this paper an extensible visual language with 18 dynamic visual cues, which supports complex collaborations over longer periods of time in a systematic way. The extensible language features basic, task-agnostic cues that are applicable to many domains. We demonstrate the validity on an extended procedural task consisting of 12 subtasks and which is copied from a real-world automotive assembly procedure.

Besides projection-based methods, there has been substantial work on visualizing robot intent using head-mounted displays (HMD) and stereoscopic glasses. Pioneering work on this topic was conducted by Milgram and colleagues [20]. Today, modern HMD technology such as the Microsoft Hololens or Oculus Rift is used for these purposes. In [23], a system is presented that visualizes upcoming robot arm movements in AR. In a similar vein, the work in [24] uses a proprietary HMD technology to visualize robot actions in a manufacturing task. However, HMDs are typically bulky and ergonomically uncomfortable when used over long periods of time [2]. In additon, they require all participants in a collaborative task to wear a physical device at all times – a cost-intensive and technologically challenging requirement that involves synchronization among multiple devices. A low-cost and efficient approach is to use LED lights to identify the intentions of the robot [28], [4], [3]. While this simplifies the necessary technical setup needed to provide visual cues to a human partner, i.e., no expensive HMD required, it also significantly reduces the range of information that can be conveyed.

In this paper, we describe a novel system that is capable of tracking and projecting information on multiple objects in three dimensions simultaneously. We also present a rich visual language that goes beyond the display of trajectories or distances and allows for complex signaling.

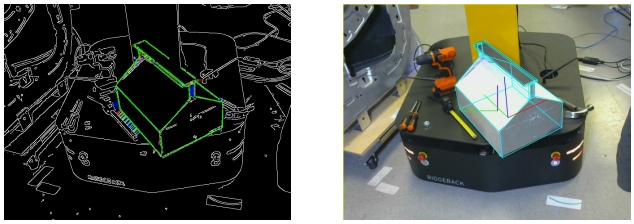
III. Visual Signaling Framework

In this section, we describe our visual communication paradigm in detail. We convey information to a human interaction partner during a human-robot collaboration task using mixed reality cues projected onto moving objects in the environment. This approach ensures that the information is communicated (a) at the right time and (b) at the right spatial location. Note that our current approach assumes information

about the environment. In particular, we assume that all objects involved in the collaboration task are available as 3D CAD models.

A. Object Tracking

Our presented system uses vision-based 3D object tracking to estimate the 6-DOF pose of objects in the environment. To this end, we use a model-based tracking algorithm inspired by [9] to estimate the pose of objects in real-time. The tracker uses polygonal mesh features from 3D CAD model to estimate the pose of a desired object. Instead of using only single low-level hypothesis for pose estimation, we handle multiple low-level hypotheses simultaneously. This enhanced approach enables robust tracking of objects even when projections are overlaid on objects. An occlusion-aware computer vision method, along with Kalman filtering is used to deal with occlusions caused by human partners. Occluded areas of an object can automatically be identified using machine learning. A detailed description of the occlusion-detection algorithm is outside the scope of this paper and can be found in our previous work [7].



(a) Sample points (green) and errors (colored lines)

(b) Estimated pose of the object

Fig. 2: Edge-based object tracking

First, an input image is captured from a monocular RGB camera and edges are extracted using the Canny edge detector. The 3D CAD model is projected onto the image and nearby Canny edges are determined using a 1-D search along the normal direction of the projected edge. Euclidean distances between sample points and their corresponding nearest edge are computed and combined together to form the distance error vector. The errors (colored lines) corresponding to the sample points (green) are shown in Fig. 2a. The pose of an object is estimated by minimizing the distance error by Iterative Re-weighted Least Square (IRLS). Fig. 2b shows the estimated pose of the object being tracked. The following section explains the mathematical model of the multiple hypotheses object tracking system, followed by evaluation of the tracker.

1) Mathematical Model of Pose Estimation: We formulate our mathematical model using the inter-frame motion. The object pose E_{t+1} at time $t+1$ can be estimated from the prior pose E_t using the inter-frame motion M.

$$E_{t+1} = E_t \; M \tag{1}$$

Motion M, in turn, can be represented using exponential map as shown below.

$$M = exp(\boldsymbol{\mu}) \tag{2}$$

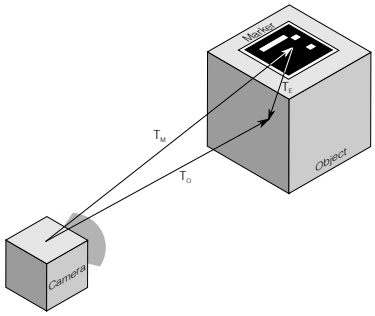


Fig. 3: Experimental setup for measuring the accuracy of the object tracker

Where $\boldsymbol{\mu} \in \mathbb{R}^6$ represents the motion velocities of 6-DOF displacement of the tracked object.

The motion M can be estimated by minimizing the error between the prior pose E_t and current pose E_{t+1}. First, the 3D CAD model of the object is projected onto the Canny edge image using prior pose E_t and points are sampled along the projected edges. Next, the edges corresponding to sample points on the projected 2D edges are determined using a 1-D search from each sample point along the normal direction of the projected edge. For each sample point $\boldsymbol{p_i}$, the Euclidean distances to all the edge correspondents $\boldsymbol{p'_{ij}}$ are computed and stacked to form a distance error vector $\boldsymbol{e}$. Finally the pose is estimated by minimizing the error $\boldsymbol{e}$ using Iterative Re-weight Least Square (IRLS) and M estimator.

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{N} \|e_i\| \tag{3}$$

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{N} min\left(\left\|p_i - p'_{ij}\right\|\right) \tag{4}$$

Where $\hat{\boldsymbol{\mu}} \in \mathbb{R}^6$ is the estimated pose of the object in current frame, obtained by minimizing the distance error corresponding to N sample points. During each iteration of optimization process, only one hypothesis corresponding to each sample point that results in minimum error is taken into account.

2) Evaluation of Single versus Multiple Hypotheses Approach: Using multiple low-level hypotheses for estimating the pose resulted in more robust tracking than using single hypothesis. To test this, we conducted an experiment to quantitatively measure the accuracy of the object tracker using single and multiple hypotheses approaches. Fiducial markers were employed to measure the ground truth pose of the object. The experimental setup is shown in Fig. 3. The ground truth transformation of the object T'_O can be calculated as shown in equation 5.

$$T'_O = T_M \; T_E \tag{5}$$

Where T_M is the transformation between the camera and the marker, and T_E is the transformation between the marker and the object. T_M is obtained by tracking the marker, while T_E is manually measured and remains constant throughout the experiment.

TABLE I: Root Mean Square (RMS) errors of the tracked objects

Objects		Translational Errors in meters			Rotational Errors in degrees		
		x	y	z	roll	pitch	yaw
Box	SHT	0.00436	0.00341	0.03141	5.33171	3.23881	1.37898
	MHT	**0.00184**	**0.00288**	**0.02018**	**1.87967**	**1.74491**	**1.00182**
Car door	SHT	0.08636	0.01508	0.11473	24.90995	14.13488	40.69923
	MHT	**0.05024**	**0.01006**	**0.05210**	**9.14722**	**5.28910**	**9.29414**
Toolbox	SHT	**0.00850**	**0.00448**	0.01239	2.00256	0.64617	1.58144
	MHT	0.00877	0.00462	**0.00956**	**1.61309**	**0.59072**	**1.31593**
Circular Object	SHT	0.00445	0.00306	0.03935	3.21225	4.41108	2.17598
	MHT	**0.00286**	**0.00171**	**0.00929**	**1.58369**	**0.78398**	**0.77677**

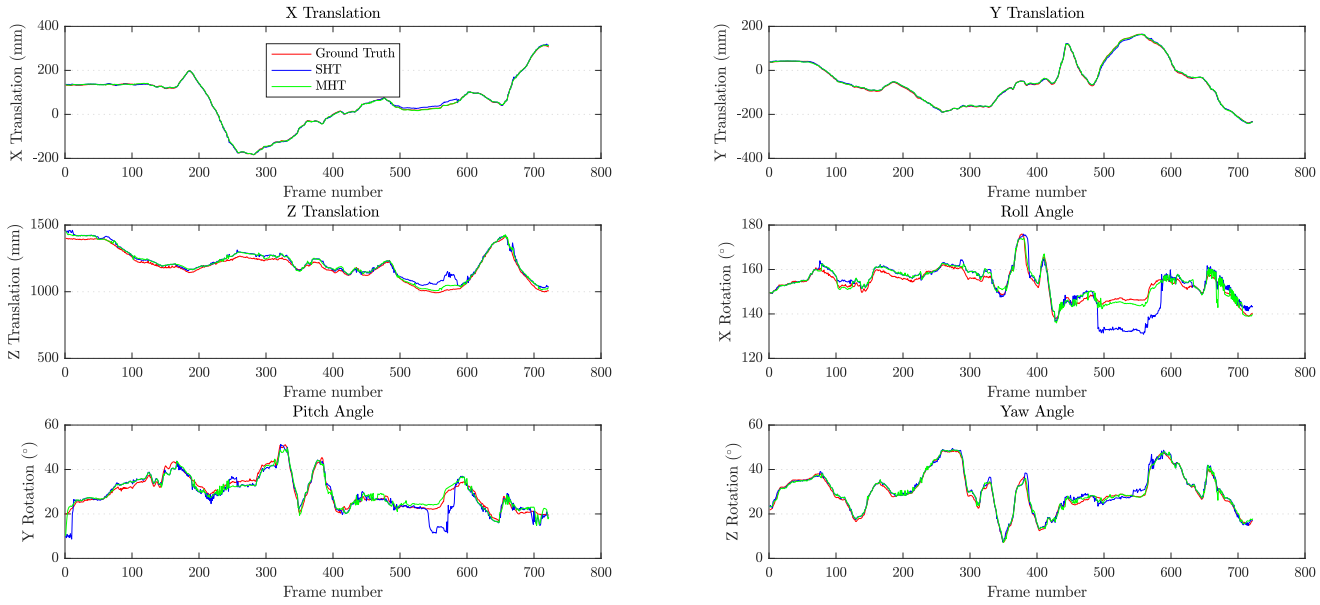


Fig. 4: 6-DOF pose plots of the box object showing the measured translation and rotation values using Single Hypothesis Tracking (SHT) and Multiple Hypothesis Tracking (MHT). Ground truth is also shown for comparison.

The experiment was conducted with four different objects: box, car door, toolbox and circular object. The objects were tracked using the single hypothesis and multiple hypotheses approaches. The 6-DOF pose data of the box measured form the experiment is shown in Fig. 4. The data in the Table I shows the Root Mean Square (RMS) errors of the tracked values in both approaches. It is evident from the Table I that multiple hypotheses tracking outperforms the single hypothesis tracking in terms of accuracy in all cases except for x and y translations of toolbox object.

It was observed from the experiment that using single hypothesis resulted in loss of tracking when there was significant occlusion, while considering multiple hypotheses enhanced the accuracy, as seen in Fig. 4 (Frame number 490–600).

B. Projection Mapping System

Given the 3D pose, we can perform projection mapping in order to display additional information on top of an object while taking into account the geometric structure. Using a projection device, the visual cues are projected into the environment in order to rapidly communicate important aspects of the tasks. The pose and shape of objects from the tracker are incorporated into the generation of visual cues, which enables the system to display only on objects-of-interest.

Since rendering of visualizations is performed within the reference frame of the projector, transforming the tracked object pose from the camera to projector frame of reference is required. To this end, projector-camera calibration is performed between the two reference frames [21]. Our system setup consists of a low-cost, monocular RGB camera (Logitech C920 Pro Webcam) which is rigidly attached to a LCD projector and pointed in direction of the scene. All algorithms are implemented in C++ and run on a single desktop PC. Our system can simultaneously track, render, and project on multiple objects in real-time at a frame rate of 20–30 Hz.

C. Extensible Visual Language

In this section, we introduce a conceptualization for dynamic visual messaging using projected mixed-reality cues.

In particular, we propose an extensible visual language to explicitly convey information to a human collaborator through visual signals. A set of patterns, analogous to *parts of speech*, are used to form a visual language from which visual messages can be formed. The language includes a reasonable fragment of patterns for human-robot interaction tasks, but can be further extended according to the application domain. Since the visual processing system in humans is very fast, visual messages can rapidly be processed without additional cognitive effort.

The basic fragment of visual cues proposed here includes patterns for designating and targeting objects (substantives), indicating positions, relations, and orientations (prepositions), basic movement instructions (verbs), success and failure (affirmation), hazards and visualizing the robot work area, as can be seen in Table II. Basic cues can be composed to generate a sequence of instructions or a visual equivalent of a phrase. These, in turn, are translated into a visual message by generating appropriate mixed-reality signals.

TABLE II: Subset of Proposed Visual Cues

Substantives	`highlight_object(X)`
	`highlight_object_part(X,Y)`
Verbs	`move_to(X,Y)`
	`remove(X)`
	`join(X,Y)`
	`align(X,Y)`
Prepositions	`in_front_of(X)`
	`left_of(X)`
	`right_of(X)`
	`at_position(X,Y)`
	`relative_to(X,Y,Z)`
Affirmation	`success()`
	`failure()`
Safety and Hazard	`stop(X)`
	`caution(X)`
	`robot_workarea()`
Text	`text(X)`
	`text_flash(X)`

D. Visual Plan Signaling

Given the conceptualization of an extensible visual language in Sec. III-C, we demonstrate a domain-specific visual language for collaborative manufacturing tasks, such as a human and a robot jointly performing manipulations on a car door prototype. This is an example of a generic language applied to a specific domain.

Fig. 5 shows a collection of visual cues and interaction metaphors that can be used to signal the state of the collaboration, next tasks, etc. For example, the robot can (a) project the boundaries of its work area, (b) communicate information about the success of the current subtask, (c) highlight specific objects, or (d) highlight a particular object part. Similarly, the user may be instructed to (e) move the object to a specified location. In this case, a slider metaphor is used in order to dynamically indicate the remaining amount of translation needed. The robot may also (f) indicate a safe position for the human partner or instruct the user to (g) join specific components. Finally, as can be seen in (h), the mixed-reality approach also allows us to visualize hidden objects, e.g., the contents of a box. This is particularly helpful in domains where information about content can be derived from bar codes or other types of input that are not human-readable. In our implementation, all visual cues are generated through a procedural approach: specific patterns are produced in real-time by modifying the available 3D CAD model, e.g., coloring the model, or overlaying textures. Hence, the approach can easily be applied to different environments and object sets as long as the corresponding 3D models are available. This is, however, typically the case in manufacturing environments. We used an open source 3D creation suite, Blender, for creating the 3D CAD models and developing the visual elements.

The above signals can, in turn, be chained into sequences and incorporated into a robot plan. This can be implemented as follows:

- `highlight(CARDOOR)`
- `move(CARDOOR, right_of(ROBOT))`
- `align(CARDOOR, relative_to(ROBOT, [1.2m, 0.3m], -35°))`

In the above example, the human is instructed to move the car door to a location near the robot, see Fig. 5e. The distance to the goal position is projected onto the work floor, which provides real-time feedback to the human. Finally, the system projects the current (green) and desired (white) position and orientation of the car door, as shown in Fig. 6. As the human tries to align the car door, the current position and orientation are displayed in real-time as a circle and a line.

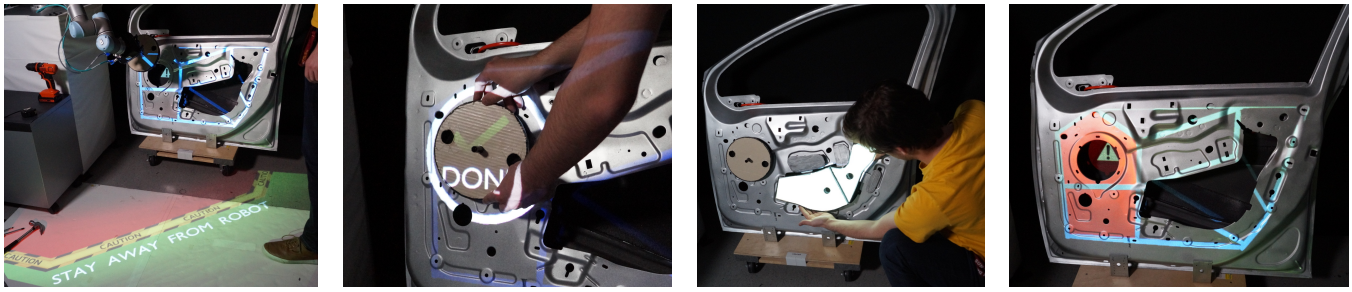
IV. Experimental Objective

A human subject experiment was conducted to compare the performance and usability of the proposed system using real-time projected cues in the workspace with a conventional method using static printed instructions. The aim of the experiment was to collect objective and subjective measurements from human subjects to analyze and evaluate the efficiency, effectiveness and satisfaction of collaborating with a robot teammate.

A. Independent Variables

In our experiment, we manipulated a single independent variable, *mode of communication*, which can have one of the three values:

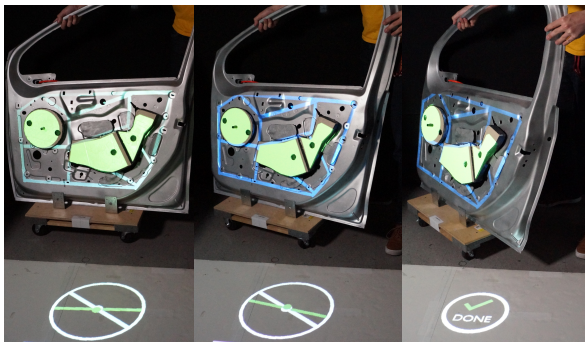
1) *Printed mode* – The subjects were provided with a printed set of instructions in the form of written descriptions and corresponding figures. The printed instructions were pasted on a wall adjacent to the workspace and were available to the subject throughout the experiment.
2) *Mobile display mode* – The subjects were provided with a tablet device consisting of instructions in the form of texts, figures, animations and videos. The device was free to be carried around while executing the task. Instructions were provided just-in-time via "forward" and "backward" buttons that allowed users to move to the next or previous tasks.



(a) Robot work area (b) Success (c) Highlight object (d) Highlight object part

(e) Move to (f) Partner at location (g) Join parts (h) Display contents

Fig. 5: A set of visual cues used to signal states of the human-robot interaction, next tasks, actions, intentions, or hidden objects during collaborative manufacturing.



(a) (b) (c)

Fig. 6: Sample use case - aligning a car door

3) *Projection mode* – The subject was provided with just-in-time instructions by augmenting (using projection mapping) the work environment with mixed reality cues.

Each participant was required to collaborate with the robot thrice (printed, mobile display and projection modes) in carrying out a procedural assembly task. The experiment used a within-subject comparison design, which enabled the participants to compare and provide subjective measures for the three methodologies. The order of conditions was varied and the order of subtasks per test condition was partially randomized on a per-subject basis to eliminate order effects.

B. Hypotheses

H1.1 *Efficiency of a human-robot collaborative team will be greater when the human subjects are provided with just-in-time instructions in the form of augmented visual cues as opposed to instructions printed on a paper or displayed using mobile device.*

H1.2 *Effectiveness of a human-robot team in accomplishing a collaborative task will be higher when the human subjects receive visual feedback as they perform and complete tasks rather than having no feedback.*

Communicating information and instructions visually and in the right place at the right time is faster, intuitive, and improves overall task performance. In contrast, instructions displayed on a mobile device or in the form of printed texts might arise ambiguities in a real-time task situation. We defined efficiency as the time taken for the human subjects to complete the task and effectiveness as the accuracy percentage of task completion.

H2 *Time taken by each human subject to understand a specific task will be constant when the instructions are in the form of just-in-time visual cues. In contrast, there will be high variation in understanding times between human subjects when the instructions are printed on a paper or displayed on a mobile device.*

We anticipate that clear and concise information in augmented visual form requires more or less the same time to understand by different human subjects. We also expect to see large variations in task understanding times between subjects in printed condition. To test this hypothesis, we measured the time taken for each subject to read or interpret a subtask in each task condition and compared the measurements.

H3 *Subjects will be more satisfied collaborating with the robot in projection mode than the other two modes.*

Additionally, explicit visual feedback will instill a positive attitude in human subjects. In contrast, subjects will feel negative or neutral when they receive no explicit feedback from the system or robot.

It is important to provide the human subjects with feedback of the robot's intention and the subject's action. This, in turn, ensures that the human collaborator will feel comfortable and satisfied working with the robot. In order to obtain the subjective measurements, human subjects completed a post-test questionnaire consisting of a series of Likert scale and free response questions.

V. Experimental Methods

We asked subjects to collaborate with a robot to carry out a well-specified assembly task in a simulated manufacturing environment. The joint assembly task involved a human subject and a stationary manipulator with six degrees of freedom (UR5 robot) performing a total of 12 manipulation steps on a car door. The assembly process required removing new components and tools from a set of toolboxes, connecting components in a specific order, and finally attaching them at different locations on the door. The car door was placed on a caster and could be moved to different locations. All experiments were reviewed and approved by the Institutional Review Board (IRB) at Arizona State University. A video demonstrating our experiment can be found at `https://youtu.be/CVY1JngYVAQ`.

A. Experiment Procedure

First, the participants were briefed on the experiment and the assembly task scenario. Participants were informed that they must collaborate with the robot in completing a procedural task consisting of 12 subtasks that must be completed successfully in sequence so that failing to complete one subtask would result in failing one or more subsequent subtasks. Nine of the 12 subtasks were assigned to the participant while the rest were assigned to the robot. The order of the subtasks was partially randomized in all three conditions (printed, mobile display and projection mode). Each participant carried out a total of three task trials under each condition. All participants were required to read and sign a consent form before beginning the experiment.

B. Experiment Task

The goal of the experimental task was to assist the robot in assembling a car door in a simulated manufacturing environment. The task involved carrying out a set of sequential subtasks $\tau = \{\tau_1, \tau_2, \ldots, \tau_{12}\}$, in a specified order. A subtask τ_i could be any one of the following:

- Pick an assembly part (interchangeable part) or tool
- Place an assembly part or tool
- Move car door to specified location inside the workspace
- Align car door with specified reference point
- Join assembly parts together
- Screw assembly parts on the car door

The instructions to execute the subtasks were framed as sequential steps and were provided to the participants as printed, mobile display, or projected instructions, depending on the test condition. The instructions also specified whether the subtask was to be completed by the human or the robot.

C. Measurement Instruments

The entire experiment was videotaped for post-hoc analysis. Efficiency and effectiveness were evaluated objectively by measuring the completion time and accuracy of each subtask. Subtask completion time, for both human and robot, was measured by recording the difference in time between start and end of the subtask. For a human subject, the subtask completion time was expressed as the total time spent on understanding the instructions and then executing it.

The percentage of task completion (fraction of successfully completed subtasks) was used as a measure to evaluate the effectiveness of the collaborative task. Additionally, accuracy of completing certain subtasks (e.g. aligning car door with a point on floor) was also measured by computing the ground truth error.

After each task trial, participants were given a post-task questionnaire consisting of seventeen 7-point Likert scale items and at the end of all the trials two free response questions, as shown in Table III. The questionnaire was designed to measure composite subjective metrics: human-robot fluency, safety and trust in robot, task execution and task load. Questionnaire items were inspired and adopted from works by [16], [14] and [11]. A few questions specific to the experiment (Questions 7-17) were added to the questionnaire.

VI. Results

In this section, we analyze and discuss our quantitative (objective and subjective) and qualitative (subjective) findings from the human-robot collaborative experiment. We also report statistically significant findings from our experiment. We used a significance level of $\alpha = .05$ for all statistical tests.

A. Participants

A total of 15 participants (aged 21–48, $M = 25.86$, $SD = 6.42$) consisting of undergraduate and graduate engineering students at a large urban research university were included in the study. All participants were recruited from the university campus via email and word-of-mouth. Of the 15 participants, 5 reported having prior experience directly interacting with a robot. Only 5 participants were native English speakers, however, all participants indicated fluency in the English language. Within-subjects design of the experiment enabled the participants to compare between the three modes of communication. To control for the learning effect, participants were told that the three task trials had different sets of subtasks, even though only the order of the subtasks was randomized. To eliminate order effects, the order of the modes (Printed, Mobile display and Projected) was also randomized for different groups of participants.

TABLE III: Subjective Measures – Post-task Questionnaire

Human-Robot fluency
1. The human-robot team worked fluently together.*
2. The robot contributed to the fluency of the interaction.*
Safety and Trust in Robot
3. I felt uncomfortable with the robot. (reverse scale)**
4. I was confident the robot will not hit me as it is moving.**
5. I felt safe working next to the robot.**
6. I trusted the robot to do the right thing at the right time.**
7. I was able to clearly understand robot's intentions and actions.*
Task execution
8. How satisfied you feel about executing the whole task?*
9. I was comfortable in interpreting the instructions. The instructions were clear and easy to understand.*
10. I feel that I accomplished the task successfully.*
11. I was able to assist the robot in completing its task successfully.*
12. The robot/system provided me with necessary feedback in order to complete the task.*
13. I would work with the robot the next time the tasks were to be completed.*
14. How was your attitude towards the task while you were performing it?*
Task load
15. The task was mentally demanding (e.g., thinking, deciding, remembering, looking, searching, etc.).***
16. The task was physically demanding. I had to put a lot of physical effort to complete the task.**
17. I never felt discouraged, irritated, stressed or frustrated at any point of time during the task execution.*
Free response questions
18. Which form of instruction (Printed or Mobile display or Projected) will you prefer if you were to collaborate with the robot on a similar task and why?
19. Explain your overall experience working on the collaborative task in all the three scenarios (Printed, Mobile display and Projected).

Note: Statistical significance found using one-way ANOVA test. $*p < .05$ favoring the projected condition, $**p = NS$ and $***p < .05$ favoring the printed and mobile display condition as more mentally demanding.

B. Objective Findings

1) Efficiency: Hypothesis **H1.1** states that the efficiency of the human-robot collaborative team will be higher in the case of the projected condition when compared to the printed and mobile display condition. Total time taken for completing all the subtasks was measured and compared between the three conditions. On comparing the measured values from printed mode and mobile display mode with the projection mode, total task completion time was found to be lower in the projection case. Fig. 7a illustrates the average task completion time in all the three test conditions.

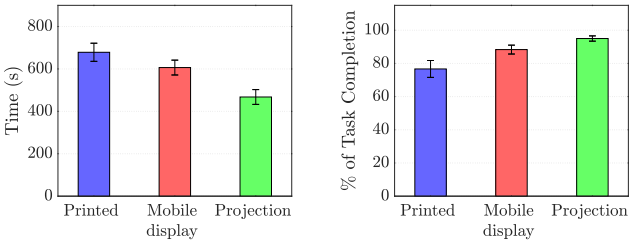


Fig. 7: Mean and standard error for (a) task completion time and (b) percentage of task completion.

An analysis of variance, using the one-way ANOVA test, showed statistically significant differences in total task completion times among the different task conditions, $F(2, 42) = 8.07$, $p < 0.01$. Task completion time in the projected condition ($M = 467.73, SD = 135.22$) was lower than the time in the printed condition ($M = 678.60, SD = 165.60$), $t(14) = 8.02, p < 0.00001$ and mobile display condition (M = 606.53, SD = 135.59) , $t(14) = 6.31, p < 0.0001$.

The statistically significant results reinforce our hypothesis that human-robot teams are more efficient with just-in-time projected instructions than with printed or displayed instructions.

2) Effectiveness: We assessed the effectiveness of the task in the three test conditions by considering the percentage and accuracy of task completion in each test scenario. The percentage of task completion by the human-robot team was computed as the fraction of successfully completed subtasks out of all given subtasks. We compared the three conditions using a one-way ANOVA test and found statistical differences in the task completion percentage as a function of the mode of communication, $F(2, 42) = 7.26$, $p < 0.01$. It can be seen from Fig. 7b that the average task completion percentage is significantly higher in the projected condition than the printed and mobile display conditions.

As a measure of accuracy, we recorded the ground truth errors for subtasks involving the alignment of the car door and objects in both task conditions. Our experiment included four error-measurable subtasks – three times the *car door alignment* and one *circular object alignment* – which involved measuring translation and rotation errors. Both translation and rotation errors were comparably smaller in the projected condition when compared to printed and mobile display condition. Analysis of variance using one-way ANOVA on the translation errors show that there is statistically significant difference between the three conditions.

In comparison, a one-way ANOVA test on the rotation errors revealed that all the tasks except *car door alignment 3* showed a significant difference between conditions, as illustrated in Fig. 8. This is acceptable because, the subtask 3 involved rotating and aligning the car door parallel ($0°$) to the robot, which is relatively easier to accomplish even without feedback when compared to other subtasks that involved rotating car door to a specified angle.

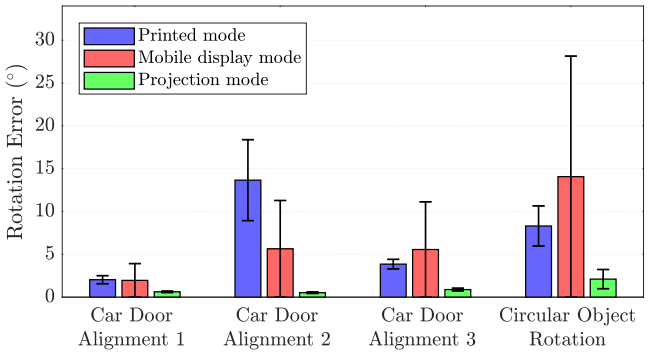


Fig. 8: Mean and standard errors of rotation errors.

3) Task Understanding Time: In hypothesis **H2**, we postulated that the time taken by different subjects to understand a subtask will be constant if the instructions are provided in augmented visual form. To investigate the hypothesis, we measured the understanding times of the subject for 9 subtasks that were assigned to participants and analyzed the standard errors of means. Task understanding time is defined as the time spent by the participant in reading or looking at instructions.

We observed that the standard errors for all subtasks in the projected condition were significantly lower than in the printed and mobile display condition, implying that most participants took a similar amount of time to understand a subtask. In contrast, standard errors in the printed and mobile display condition were comparatively higher, particularly for subtasks 4, 8, and 9, as shown in Fig. 9.

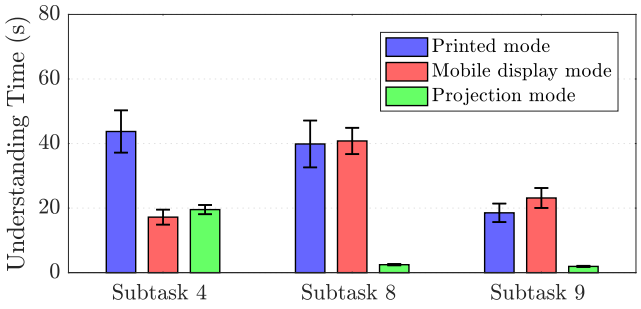


Fig. 9: Mean and standard error for task understanding time.

C. Subjective Findings

Our analysis of subjective findings was based on responses to Likert-scale and open-ended questions included in the survey. We analyzed open-ended questions using a modified Grounded Theory and content analysis approach (see [6]).

1) Questionnaire Items: We compared participant ratings for each questionnaire item between test conditions (printed vs. mobile display vs. projected) using one-way ANOVA, as shown in Table III. A post-hoc t-test using a Bonferroni correction of $\frac{\alpha}{3}$ was carried out to compare mobile the display vs. projected condition. Subjective responses significantly favored the projected condition with regard to human-robot fluency, clarity, and feedback. The t-test using Bonferroni correction supports the hypothesis that fluency is improved during the projected condition (Q1, $p = 0.0025$; Q2, $p = 0.0035$). Participants significantly favored projected and mobile display conditions compared to printed conditions for task execution, human-robot collaboration, and attitude. However, there was no statistically significant difference between scores for the projected and mobile display conditions for these items (Q8, $p = 0.13$; Q11, $p = 0.15$).

Hypothesis **H3** also states that explicit visual feedback will instill a positive attitude in participants, and that participants will feel negative or neutral when they receive no explicit feedback from the system or robot (i.e. in the printed condition). Subjective responses supported this hypothesis to some degree with the median central tendency for Q14 (How was your attitude to the task while you were performing it?) being 6 ("positive") for the projected case and mobile display case, compared to 4 ("neutral") for the printed case. There was not a significant difference between attitude scores for the projected and mobile cases.

2) Qualitative Free Response Data: All participants favored the projected condition over printed and mobile conditions. Major themes included user perceptions of their own ability (e.g. ease of performing task, ability to complete task accurately), user perceptions of robot system performance (e.g. clarity of instructions, provision of feedback, intuitiveness of the overall process, system oversight of the task series), human-robot interaction experience (including perceived safety), and overall attitude toward task condition.

Overall, free response comments were overwhelmingly positive for the projected instructions condition in contrast to more negative responses for the printed instructions condition. Responses for the mobile condition were positive, but all respondents indicated an overall preference for the projected condition. Several respondents noted that the projection system felt game-like, whereas the printed system felt like work. Respondents felt that the projection system was more intuitive, leading to more fluid and accurate task performance in contrast to the printed task, which required frequent reference to the instructions that were not always intuitive, and frequently hampered by human imprecision in the manual measuring elements of the task. Participants perceived that the projected condition yielded better accuracy with improved efficiency compared to the printed condition. However, one participant noted that in a manufacturing environment with compartmentalized worker task repetition, a worker presented with printed task instructions would most likely become fluent with the task after a few repetitions, so that the printed approach would ultimately be more efficient than the projected instruction approach. A few participants noted that the demonstration videos in the mobile condition were helpful to improve task accuracy. Several participants referred to the human-robot interaction as a team, and most participants felt that the human robot interaction was safe. Participants noted that it was a positive feature that the robotic system kept track of overall task progress in the projection system, rather than relying on human oversight.

VII. Limitations

Despite the demonstrated advantages of the proposed system, there are various limitations worth noting. The system does not take into account the human position or movements which could be of critical importance to improve the responsiveness and safety. The positions of both the projector and the camera are stationary and the human partner is occasionally seen blocking both tracking and projection. In practice, this did not affect the performance in a significant manner, but a setup using multiple cameras and projectors is possible to circumvent this issue. Regarding the above experimental design, there may be additional factors that could be incorporated. In particular, just-in-time signaling is at the moment only used for the projected and mobile

mode. By analyzing both an (1) just-in-time and (2) an all-at-once mode, deeper insights into influence of timing could be gained. The current design does not disambiguate between the two modes. Further, both the printed and projected mode were hands-free while in the mobile mode a device was carried. Since the used mobile-device was only marginally larger than a cell phone, users were mostly unobstructured during the task. However, in future work we would like to analyze the influence of carrying a device on task performance. Also, while we used a Grounded Theory approach in this paper, it would be worthwhile to create and validate a scale with established reliability. However, that would requirer multiple rounds of prospective testing, and is outside of the scope of this paper. Finally, there is likely bias in the thematic content of qualitative free responses due to conceptual priming effect [5] from administering subjective Likert scale questions on same printed form immediately before soliciting free response data.

VIII. Conclusion

In this paper, we proposed a methodology for visual signaling during human-robot collaboration and evaluated its suitability in a manufacturing domain. We introduced a mixed reality system that combines a vision-based object tracking algorithm with a context-aware projection mapping technique to communicate with human user. We introduced a conceptualization for visual languages based on signal categories, similar to *parts of speech* in natural language and also demonstrated the domain specific example.

A user study was performed to evaluate the introduced methodology. The objective evaluation using the task completion time and accuracy measurements corroborated our hypotheses **H1.1** & **H1.2** that using our mixed reality system would increase the efficiency and effectiveness of a human-robot team. Participants took less time to complete the task when following projected visual instructions. Our analysis also confirmed that visual instructions were intuitive and took approximately the same amount of time for different participants to understand, supporting our hypothesis **H2**.

Subjective findings from structured and free response questions supported our hypothesis that participants would experience higher satisfaction with the projected mode when compared to the printed or mobile display mode.Participants responded favorably to feedback and found the projected case to be enjoyable. Notably, multiple participants referred to the human-robot collaboration as a team, reflecting the term offered by the experimental instructions and suggesting the opportunity to explore development of qualities characterizing high-functioning teams, such as trust, in the human-robot interaction. In addition, several participants mentioned that the projected case had a game-like quality. This observation suggests the opportunity to explore further integration of game design concepts [25] to enhance the human experience and task performance.

In light of our relatively homogeneous participant cohort consisting of undergraduate and graduate engineering students at a large urban research university, we cannot generalize our findings to a broad user group. Therefore, we plan further testing with additional participant groups, including non-engineers, individuals with prior line manufacturing experience, and individuals representing a broader age range. Future plans also include usage of think-aloud protocol [12] to better understand subjects' real-time perceptions of interacting with the robot.

References

[1] Rasmus Skovgaard Andersen, Ole Madsen, Thomas B Moeslund, and Heni Ben Amor. Projecting robot intentions into human environments. In *Ro-man 2016-Proceedings of the 25th Ieee International Symposium on Robot and Human Interactive Communication*, 2016.

[2] Ronald Arkin and Thomas Collins. Skills impact study for tactical mobile robot operational units, 2005.

[3] K. Baraka, S. Rosenthal, and M. Veloso. Enhancing human understanding of a mobile robot's state and actions using expressive lights. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 652–657, Aug 2016.

[4] Kim Baraka, Ana Paiva, and Manuela Veloso. *Expressive Lights for Revealing Mobile Service Robot State*, pages 107–119. Springer International Publishing, Cham, 2016.

[5] John A Bargh, Mark Chen, and Laura Burrows. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2):230–244, 1996.

[6] H Russell Bernard. *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman & Littlefield, 2017.

[7] S. Brahmbhatt, H. Ben Amor, and H. Christensen. Occlusion-Aware Object Localization, Segmentation and Pose Estimation. *ArXiv e-prints*, July 2015.

[8] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J Lilienthal. That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *Mobile Robots (ECMR), 2015 European Conference on*, pages 1–6. IEEE, 2015.

[9] Changhyun Choi and Henrik I Christensen. Real-time 3d model-based tracking using edge and keypoint features for robotic manipulation. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4048–4055. IEEE, 2010.

[10] Henrik I Christensen, T Batzinger, K Bekris, K Bohringer, J Bordogna, G Bradski, O Brock, J Burnstein, T Fuhlbrigge, R Eastman, et al. A roadmap for us robotics: from internet to robotics. *Computing Community Consortium and Computing Research Association, Washington DC (US)*, 2009.

[11] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. Effects of robot motion on human-robot collaboration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 51–58. ACM, 2015.

[12] K Anders Ericsson and Herbert A Simon. Verbal reports as data. *Psychological review*, 87(3):215, 1980.

[13] Fabrizio Ghiringhelli, Jerome Guzzi, Gianni A Di Caro, Vincenzo Caglioti, Luca M Gambardella, and Alessandro Giusti. Interactive augmented reality for understanding and analyzing multi-robot systems. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1195–1201. IEEE, 2014.

[14] Matthew C Gombolay, Reymundo A Gutierrez, Shanelle G Clarke, Giancarlo F Sturla, and Julie A Shah. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, 39(3):293–312, 2015.

[15] S. A. Green, M. Billinghurst, X. Chen, and G. J. Chase. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, 5(1):1–18, 2008.

[16] Guy Hoffman. Evaluating fluency in human-robot collaboration. In *International conference on human-robot interaction (HRI), workshop on human robot collaboration*, volume 381, pages 1–8, 2013.

[17] Kentaro Ishii, Shengdong Zhao, Masahiko Inami, Takeo Igarashi, and Michita Imai. Designing laser gesture interface for robot control. In *IFIP Conference on Human-Computer Interaction*, pages 479–492. Springer, 2009.

[18] Florian Leutert, Christian Herrmann, and Klaus Schilling. A spatial augmented reality system for intuitive display of robotic data. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 179–180. IEEE Press, 2013.

[19] Jim Mainprice, E Akin Sisbot, Thierry Siméon, and Rachid Alami. Planning safe and legible hand-over motions for human-robot interaction. *IARP workshop on technical challenges for dependable robots in human environments*, 2(6):7, 2010.

[20] Paul Milgram, Shumin Zhai, D Drascic, and J Grodski. Applications of augmented reality for human-robot communication. pages 1467 – 1472 vol.3, 08 1993.

[21] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 464–471. IEEE, 2012.

[22] Shayegan Omidshafiei, Ali-Akbar Agha-Mohammadi, Yu Fan Chen, N Kemal Ure, Jonathan P How, John Vian, and Rajeev Surati. Mar-cps: Measurable augmented reality for prototyping cyber-physical systems. In *AIAA Infotech@ Aerospace Conference*, 2015.

[23] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. Communicating robot arm motion intent through mixed reality head-mounted displays, 2017.

[24] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. *Third Point of View Augmented Reality for Robot Intentions Visualization*, pages 471–478. Springer International Publishing, Cham, 2016.

[25] Katie Salen and Eric Zimmerman. *Rules of play: Game design fundamentals*. MIT Press, 2004.

[26] Shin Sato and Shigeyuki Sakane. A human-robot interface using an interactive hand pointer that projects a mark in the real work space. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 589–595. IEEE, 2000.

[27] Jinglin Shen, Jingfu Jin, and Nicholas Gans. A multi-view camera-projector system for object detection and robot-human feedback. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 3382–3388. IEEE, 2013.

[28] Daniel Szafir, Bilge Mutlu, and Terry Fong. Communicating directionality in flying robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 19–26, New York, NY, USA, 2015. ACM.

[29] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. Asking for help using inverse semantics. In *Proceedings of Robotics: Science and Systems*, Berkeley, USA, July 2014.

[30] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. Communicating robotic navigational intentions. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 5763–5769. IEEE, 2015.