**Project Statement**

The fundamental goal of our research is to explore the feasibility of a system that facilitates human-robot shared space collaboration. The intent is to allow a human director to command a robot using, at a minimum, gesture and voice. In a broad sense, the team seeks to create an understanding of shared context between the human collaborator and the system. What is unique is that this shared context will be drawn from experience and should evolve with increased interactions with the human collaborator. This evolution is facilitated by communicating the current state of understanding with the human collaborator. Both a dynamic shared context and a way of expressing that shared context through feedback are the key undercurrents that will fuel much of the research for this work.

To constrain our problem, we restrict the robot’s task domain to moving blocks. This reduction allows for focus on the above research topics and not on the complexity endemic to a more generalizable problem domain. This constraint is inherited from the previous year’s work, where the team implemented both a means for integration and a rudimentary language-processing component. Their work was entirely concentrated in a 2D game, whereas our work is intended to be centered on a tactile, 3D collaboration space.

Given the previously completed work, there is a natural progression to the beginning of our work. Our goal is to implement the components that will create an MVP of a human-robot interaction system. That is, there is still need for a gesture component to recognize the visual aspects of a collaborator’s direction. There is still need for a preprocessing component that identifies the blocks in the collaboration space. Additionally, no form of feedback exists to indicate whether or not the integrated system was able to identify the correct block. These components, along with the integration between them, are what a shared context must be built upon and thus form the crux of the initial work.

Once an MVP is produced, any remaining time should be invested into how our method of integration facilitates or limits forming a sense of shared context. Indeed, our team hypothesizes that independently trained gesture and language-processing components can be integrated to provide details about the current context without extensive retraining. However, the actual verification of this hypothesis requires iterations of prototyping and reflection to evaluate whether our hypothesis is correct, iterations which form the core of our workflow. This overarching hypothesis is supported by many smaller hypothesis for relevant prototypes. Despite these smaller prototypes and hypotheses, it is the overarching goal of first developing, then evolving a sense of shared context in a gesture and language driven system that is the ultimate aim of our research.

## Requirements

**Functional Requirements**

First Iteration

1. The system shall respond to a pointing gesture from the user.
2. The system shall recognize blocks in the collaboration space.
3. The system shall locate blocks in the collaboration space.
4. The system shall inform the user of which block it has determined the user indicated.
5. The system shall provide feedback if it cannot distinguish between blocks.
6. The system shall respond to verbal instructions from the user.

Second Iteration

1. The system shall receive verbal instructions through speech.
2. The system shall solicit additional information from the user when needed.

Future Iterations

1. The system is able to move a robotic arm to complete instructed tasks.
2. The system is able to recognize non-pointing gestures.
3. The system shall be able to repeat previously completed tasks before the user is finished describing these tasks.
4. The system will provide a prompt if no objects are in the collaboration space.

**Nonfunctional Requirements**

1. A new user shall not require more than an hour of training to use the system.
2. The system shall recognize any human user.
3. The system shall be tolerant to up to 25% occlusion of blocks.
4. The system shall be tolerant to slight changes in lighting.
5. The system shall be tolerant to slight changes in sensor angle.

**Constraints**

1. The system shall consist of a flat collaboration space.
2. Instructed tasks shall be limited to moving blocks.
3. The collaboration space is distinguishable from the involved blocks.

## **User Scenarios**

### Use Cases

**Use Case:** Pointing-Only Disambiguation

**Description:** User points to one of several blocks.

**Actors:** Human Collaborator

**Pre-Conditions:**

* User is sitting in front of the system’s depth sensor. AND
* The collaboration space is in view of the system’s camera.

**Basic Flow:**

1. User points towards a block.
2. System displays vector representing the pointing gesture.

1. System notifies user of the nearest block.
2. System prompts the user to determine if the selected block is the desired one.

1. User responds that the selected block is the desired one.
2. System records that processing was correct.

**Alternate Flow:**

1. System identifies several likely matching blocks ([step 3](#h2so9xj6mytp)).
   1. System prompts user for additional information.
   2. Return to [step 1](#4va9zc13u2o).
2. User responds that the selected block is not correct ([step 5](#clf9b6hercg2)).
   1. System records that processing was incorrect.
   2. Return to [step 1](#4va9zc13u2o).

**Post-Conditions:**

* System notifies user of the correct block OR
* Returned to beginning of this use case.

**Use Case:** Pointing and Speaking Disambiguation

**Description:** User points to one of several blocks and describes it vocally

**Actors:** Human Collaborator

**Pre-Conditions:**

* User is sitting in front of the system’s depth sensor. AND
* The collaboration space is in view of the system’s camera.

**Basic Flow:**

1. User points towards a block.
2. System displays vector representing the pointing gesture.
3. User describes block
4. System notifies user of the nearest block that matches description.
5. System prompts the user to determine if the selected block is the desired one.
6. User responds that the selected block is the desired one.
7. System records that processing was correct.

**Alternate Flow:**

1. User responds that selected block is not the desired one
   1. System prompts user for additional input.
   2. Return to step 3.

**Post-Conditions:**

* System notifies user of the correct block OR
* User will repeat use case.

### Quality Attribute Scenarios

**Source:** System

**Stimulus:** Cannot distinguish between several likely matches.

**Artifact:** System - Feedback

**Environment:** Runtime

**Response:** System asks human collaborator for further clarification

**Response Measure:** Within 10 seconds of receiving collaborator instruction

**Source:** Human Collaborator

**Stimulus:** Interacts with the system for the first time.

**Artifact:** System

**Environment:** Runtime

**Response:** User uses application productively.

**Response Measure:** Within 1 hour of experimentation.

**Source:** Human Collaborator

**Stimulus:** Collaborator gives a previously completed task.

**Artifact:** System

**Environment:** Runtime

**Response:** System begins to perform task before instruction is completed.

**Response Measure:** Before the collaborator is done stating their task.

**Source:** Human Collaborator

**Stimulus:** Collaborator interacts with the system for the first time.

**Artifact:** System

**Environment:** Runtime

**Response:** System recognizes the new user.

**Response Measure:** Functions as normal (completes task within 10 seconds of receiving collaborator instruction).

**Source:** External Light

**Stimulus:** Collaboration space is brighter than typical operation environment.

**Artifact:** System

**Environment:** Runtime

**Response:** System functions as normal.

**Response Measure:** Functions as normal (completes task within 10 seconds of receiving collaborator instruction).

**Source:** System - Sensor

**Stimulus:** Sensor is at a slight angle that still captures the entire collaboration space.

**Artifact:** System

**Environment:** Runtime

**Response:** System functions as normal.

**Response Measure:** Functions as normal (completes task within 10 seconds of receiving collaborator instruction).

**Source:** System - Sensor

**Stimulus:** Block is partially occluded.

**Artifact:** System

**Environment:** Runtime

**Response:** Block is recognized as part of environment.

**Response Measure:** System functions as normal (completes task within 10 seconds of receiving collaborator instruction).