The Role of Verbal and Non-verbal Communication in a Two-Person, Cooperative Manipulation Task

Sarangi P. Parikh, Joel M. Esposito, and Jeremy Searock

Abstract—**Motivated by the differences between human and robot teams, we investigated the role of verbal communication between human teammates as they work together to move a large object to a series of target locations. Only one member of the group was told the target sequence by the experimenters; while the second teammate had no target knowledge. The two experimental conditions we compared were: Haptic-Verbal (teammates are allowed to talk) and Haptic Only (no talking allowed). The team's trajectory was recorded and evaluated. In addition, participants completed a NASA TLX-style post-experimental survey which gauges workload along 6 different dimensions.**

In our initial experiment we found no significant difference in performance when verbal communication was added. In a follow up experiment, using a different manipulation task, we did find that the addition of verbal communication significantly improved performance and reduced the perceived workload. In both experiments, for the haptic-only condition, we found that a remarkable number of groups independently improvised common haptic communication protocols (CHIPs). We speculate that such protocols can be substituted for verbal communication; and that the performance difference between verbal and non-verbal communication may relate to how easy it is to distinguish the CHIPs from motions required for task completion.

I. INTRODUCTION

Human teams generally employ a variety of communication mechanisms to effectively complete cooperative manipulation tasks. Consider a scenario where two people are moving a table across a room. Intuition suggests people are capable of completing such a task using only haptic information; yet they often naturally employ voice commands to share their intentions, resolve conflicts and gauge task progress. These two types of communication, haptic and verbal, differ in many respects. Sensed forces can be considered quantitative, high-bandwidth communication signals while verbal communications are sporadic and more qualitative – often contextual – in nature (e.g. "Move a little to your left").

Cooperative manipulation is also a popular application for multi-robot teams; yet, in contrast to human teams, engineered systems rarely employ sporadic, contextual, non-haptic communication (verbal or otherwise). In one class of multi-robot coordination work, communication between the robotic partners is strictly implicit, using only the forces transmitted through the object itself [1, 2]. A similar approach is often used for human-robot teams. The human moves the object in a desired direction and the robot uses only haptic information, derived from a force sensor, to assist the human operator [3, 4, 5].

A second class of approaches to multi-robot manipulation [6] requires a lead robot to constantly stream a high-bandwidth feed-forward signal to the followers to ensure stability during the acceleration and deceleration phases. However, it is apparent that humans do not rely on quantitative feed-forward acceleration signals when working together. Instead, human partners seem to rely on qualitative (verbal) information in addition to any available haptic information.

The sharp differences between natural and engineered cooperative manipulation strategies in their use of non-haptic communication motivate the human behavioral study described in this paper. In particular we seek to quantify the impact of verbal communication on both the performance and task loading of human-human teams as they engage in manipulation tasks. While several previous studies have investigated such task, none to our knowledge explicitly treat verbal communication as an experimental factor. Understanding the role of verbal communication may help robotics researchers develop more natural human-robot interaction protocols for such tasks.

The remainder of the paper is organized as follows. Section II reviews other efforts at understanding both haptic and non-haptic communication in human-human teams. At the end of the section, specific research questions are posed. Section III describes our experimental design and procedures. Section IV describes how the performance of the team was quantified and presents a statistical analysis of the team's performance and workload assessment. Section V and VI describe a follow up experiment. Finally, Section VII provides some concluding discussion.

II. RELATED WORK

In this section we review work done on human-human cooperative manipulation establishing haptic communication as being beneficial (Sect. II-A). We note the paucity of controlled manipulation studies involving the role of verbal communication (Sect. II-B). The experimental design introduced in Sect. III was informed by the apparent ubiquity of the leader-follower architecture in both human-human and human-robot collaborative tasks (Sect. II-C).

A. Haptic communication in human-human teams

If both individuals are physically connected to an object, the obvious modality of communication is haptic [7]. There have been numerous studies on the role of haptic communication in virtual tasks [8, 9, 10, 19]. While the different studies analyzed various performance measures, the conclusions indicate that there is improved performance when haptic feedback is provided. Additionally, [11] provides further details and many references on haptic cooperation between two individuals. Groten, et. al. [19], also show in their work that intention integration benefits from haptic communication. As a result, we begin with the premise that haptic feedback is beneficial, and, as such, is available in all of our experiments.

In [12], the authors estimated the impedance and passivity of two humans moving an object. They noted that there are significant differences between the gross transport phase and the fine positioning phase of the motion. The follower actively assisted in the gross motion phase; but provided higher damping forces in the fine positioning phase. They speculated that an abrupt deceleration by the leader, prior to the fine positioning phase, might have been an attempt to signal the desired change in impedance. In our experiments, we also analyze abrupt changes in speed and direction. Our paper extends the notion that a teammate can provide haptic cues or gestures that represent an attempt to explicitly communicate with the follower. In our work we refer to them as common haptic improvised protocols – or CHIPs.

B. Verbal communication in human-human teams

The benefits of verbal communication have been illustrated for other types of tasks in which the partners do not have haptic information. For example, in tasks that do not involve manipulation [13,14] or in virtual manipulation tasks lacking force feedback [15,16,17], verbal communication can be used to signal transitions.

However, few works have specifically investigated the role of verbal communication in physical manipulation tasks. In [8], it is concluded that individuals use the haptic channel successfully to communicate with each other and integrate their intentions as they move a virtual ball on a path. In these experiments, individuals wore noise canceling headsets and were not allowed to speak with each other. They only had visual and haptic information available to them. In many studies the subjects are either not permitted to speak at all [18,19], or verbal communication is not consistently controlled or recorded [20,21]. There has been work done by Huang, et. al. [22] that specifically focuses on auditory (not verbal) cues. The auditory cues that were tested were different sounds that were coupled with specific actions. The experiment involves two individuals collaborating in a virtual environment. Half of the groups had audio, visual, and haptic information available while the other groups only had visual and haptic information. The study uses time to completion of the task as a performance measure and concludes that adding auditory cues makes task completion faster. However, in both conditions of the experiment, individuals were able to freely talk to each other. Thus, they asked clarifying questions and provided verbal guidance.

Thus, our experiments involve physical manipulation of an object by two people where haptic feedback is always available and verbal communication is isolated via different experimental conditions. Additionally, we record every verbal communication that occurs between the partners.

C. Leader-Follower Architectures for Cooperative Manipulation

In addition to the studies mentioned above, there is considerable research in the psychology literature on dominance in human-human teams suggesting that a leader-follower paradigm arises quite naturally in dyads. For example the Social Relations Model (SRM) [23] proposes a three-factor model in which dominance is determined both by the inherent tendencies of the individuals as well as by components specific to the particular dyad and task. The model has been studied in the context of a manipulation task [24], and it was found that humans naturally prefer to work in a team where one partner is slightly more dominant. Similar models have been implemented in shared control architectures as well [21, 25, 26].

When designing our experiment, we use an information structure analogous to the most common human-robot cooperative manipulation paradigm. Frequently, the human plays the role of the "leader" – who dictates the direction of motion and possesses knowledge of the target position. The robot plays the role of the follower and usually has no a priori knowledge of the target location. In many of the cases, the robot is responsible for maintaining grasp and contact forces and helping support the mass of the object (see for example [27, 28, 29]). Based on the ubiquity of the leader-follower architecture in human-robot and human-human teams, we chose to design our experiments in such a way that only one member of the team was given the target information.

D. Research Questions

Based on this review of previous work, we choose to explicitly study the role of verbal communication in a two person manipulation task in order to answer the following questions.

Q1: How does the addition of verbal communication between the leader and follower change the performance of the dyad?

Q2: How does the addition of verbal communication between the leader and follower change the perceived workload of

each member of the dyad?

III. EXPERIMENT I: DESCRIPTION

We chose to study a task where two individuals moved a large object to a sequence of target locations. The task was designed to require close collaboration between the partners since the table was too large for a single person to manipulate.

A. Task Description

Figure 1 shows an overhead view the experimental setup. The experiments were done in a large open room, approximately 8 meters by 5 meters. Laser pointers were mounted on the ceiling and aimed straight down to create four different illuminated spots, representing target locations, on the floor. The targets were equidistant from each other (2.557 meters – with the exception of the path from 1 to 3 which was not used). A team of two people was asked to carry the table to a prescribed sequence of targets, as quickly as possible. The table was approximately 0.75 m by 1.2 m, and weighed 18 kg. The target was considered "reached" once the projected laser spot appeared inside the black circle mounted on the table top (0.3m in diameter).

B. Data Acquisition

The table was tagged with 14 retro-reflective, 1 cm, spherical fiducials so that its position and orientation could be recorded via an 8 camera Vicon Motion Capture System. The cameras used infrared filters and strobes to robustly localize the markers at 20 Hertz and the system calibration file reported a precision of 8 mm (95% confidence). In addition, a time synchronized video recording was made of the experiment, with the audio track providing a record of the verbal exchanges between the dyad.

C. Information Structure and Communication

One of the two individuals was randomly designated the *Leader* for the duration of the experiment. The experimenter gave the leader a sequence of 8 targets. The other individual was considered the *Follower*. The experimenter did not provide the target locations to the follower; in addition, they wore a blindfold during the trials to prevent them from seeing the lasers, their partner's facial expressions or body language. We tested two different experimental conditions.

1. Haptic Only: Neither partner was permitted to speak.
2. Haptic-Verbal: Both partners were permitted to speak.

Under both of these conditions, the follower experiences similar sensing and informational constraints as a robotic partner in a traditional human-robot manipulation architecture; and the Leader plays the role of the human partner. Haptic Only mimics the traditional set up where the human does not issue verbal commands to the robot; while, Haptic-Verbal represents what we feel would be a technologically plausible enhancement to the traditional human-robot manipulation architecture.

D. Participants, Procedure, and Methodology

24 individuals (5 female, 3 left-handed), ages 18-22, were recruited for this study. Human subject testing approval was obtained from the institutional review board.

Upon arrival, the two participants were seated in separate areas. They were instructed not to speak to their partner prior to the beginning of the experiment or during breaks. They were given a consent form, which was used to exclude individuals with injuries preventing them from safely lifting an 18 kg object. They also completed a pre-experimental questionnaire which was used to prevent the pairing of roommates or close friends (self-described).

Participants then entered the room shown in Fig. 1, and the investigator read instructions aloud from a script. They were told that their objective was to move the table to a series of targets as quickly as possible, and that the specific locations would be provided later to one of the partners.

The participant that was randomly pre-selected to be the follower was then blindfolded; the lasers were turned on; and the investigator showed the leader an annotated map of the room with a written sequence of targets (ex. "Target 3, then Target 2, then Target 4, etc."). Once the leader indicated she understood the instructions, the experiment began.

After each trial, the lasers were turned off and follower was permitted to remove his blindfold and rest while completing a post-experimental survey that assessed the individual's perception of teamwork, physical demand, frustration, stress, and their own performance (discussed in Sect. IV-B).

In this counter-balanced, within-subjects experimental design, every dyad completed a block consisting of an 8 target sequence, twice, with each of the two experimental conditions. In total, each group participated for approximately 30 minutes.
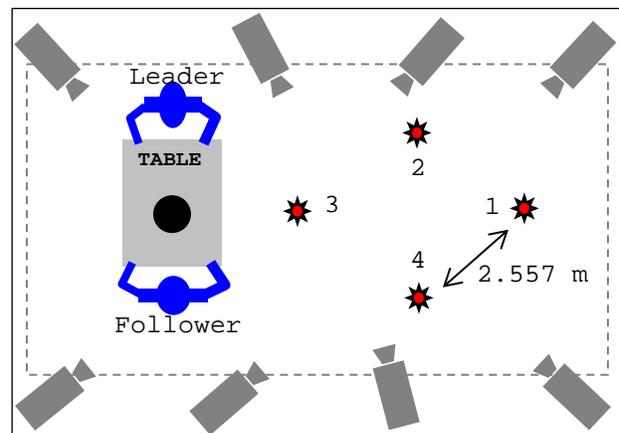

Figure 1: An overhead view of Experiment 1. The stars represent red targets projected on the floor via ceiling mounted laser pointers.

Several steps were taken to mitigate learning effects. First, a dyad was never asked to execute a particular target sequence more than once. Second, counterbalancing was used to control for order effects.

IV. EXPERIMENT I: ANALYSIS

A. Performance

While there are a number of performance metrics that could be studied, we used time to completion. Recall, subjects were instructed to move as quickly as possible between the target locations. The time-stamped motion capture data made it easy to compute this measure.

The mean (and standard deviation) time to completion for each experimental treatment was: Haptic Only 88.94 s (19.34 s) and Haptic-Verbal 102.28 s (40.91 s). The Haptic Only condition resulted in better performance than the Haptic-Verbal condition. Also, note that Haptic-Verbal has a larger coefficient of variation (standard deviation divided by mean) of 0.4 vs 0.22 –implying that performance varied considerably under that condition. A repeated measures analysis of variance (ANOVA) test showed no significant difference between Haptic-Verbal and Haptic Only at the 0.05 level.

B. Subject's Perception of Performance and Workload

At the conclusion of each block, both participants were separated, and asked to complete a modified version of the NASA TLX survey [30]. The survey is used to rate their experiences along the dimensions presented in Table 1. Figure 2 depicts the results graphically. With the exception of the Leader's response to Question 7 (Overall Satisfaction), the responses follow the same pattern. Participants rate Haptic-Verbal more favorably than the Haptic Only condition. None of the comparisons are statistically significant according to an ANOVA after applying a Bonferroni correction.
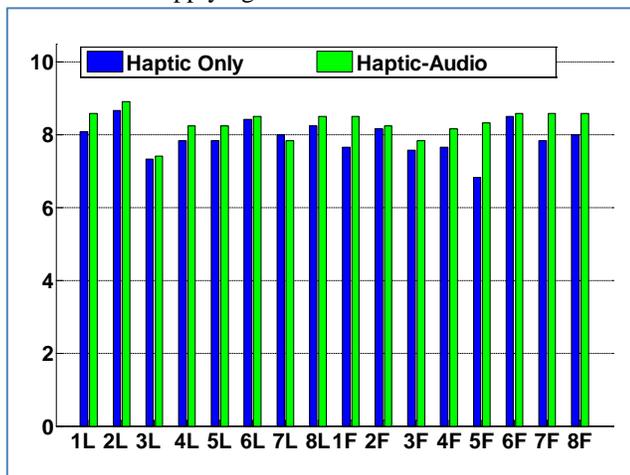

Figure 2: The mean responses of Leaders (L) and Followers (F) to the 8 questions in the modified NASA Task Load Index (TLX) survey in Table 1.

Table 1: The modified NASA TLX post-experimental survey.

Question	0 (worst)	10 (best)
1 Working with partner	Difficult	Easy
2 Physical demand	Demanding	Easy
3 Temporal demand	Frantic	Leisurely
4 Frustration	Stressful	Relaxing
5 Overall effort	Max	Min
6 Performance appraisal	Discontent	Satisfied
7 Overall satisfaction	Frustrated	Satisfied
8 Overall difficulty	Difficult	Easy

C. Additional Experimental Observations

There were several unanticipated phenomena worth discussing. First, upon reviewing and transcribing the audio recording, we realized the communication was very structured across all 12 groups. We were able to partition these exchanges into the following categories.

- Direct Commands (e.g. "Stop"): 314 occurrences
- High Level task Descriptions (e.g. "Targets are in a triangular pattern"): 16 occurrences
- Confirmations (e.g. "Am I doing OK?"): 9 occurrences

Regarding the direct commands, the most common interactions were single word instructions. The most frequent being "Stop", followed in descending frequency by "back", "slow", and "left" or "right". Only 4 of the 12 groups gave high level instructions like "we are going to move back and forth between these two targets 8 times". Groups that gave instructions like this usually did so once at the beginning of the trial, while standing stationary. This may help explain why the Haptic-Verbal has a greater time to completion and a higher standard deviation as compared to Haptic Only. Participants did not always walk and talk simultaneously.

Finally, since the follower is permitted to speak during the Haptic-Verbal condition, we expected that some would occasionally request instructions or feedback. Yet only one of the 12 did.

Upon watching the video recording, we noted two interesting, and possibly related behaviors. The most obvious was that 9 of the 12 teams developed an impromptu non-verbal communication protocol. Rather than hold the table parallel to the floor, the leader tilted the table downward in the direction that he or she wished to travel. We found it remarkable there was no *a priori* verbal agreement between the leader and follower about what this meant (recall that once participants signed the disclosure, they were not allowed to speak with each other). Furthermore, it is interesting all 9 teams developed this concept seemingly independent of one another – the disclosure form also asks them not to discuss the experiment with those who have not yet participated. We have termed this type of behavior as a Common Haptic Improvised Protocol (CHIP).

A second observation is that the teams appear to avoid rotating the table about its yaw axis (i.e. zero yaw rate), but do not seem to exhibit a consistent preference on yaw angle. Clearly the task does not require controlling the yaw angle, but for some reason 10 groups (all 9 groups discussed above, and one additional group), actively stabilized the orientation in this fashion. One might expect groups would prefer certain ergonomic orientations (e.g. so that neither member has to walk backward, or so that the leader always points toward the target, etc.) However, there was no obvious pattern to how groups selected the final orientation as seen in Fig. 3. An

ANOVA showed there was no significant effect on the absolute heading difference. One possible explanation is that it is difficult for the follower to haptically distinguish lateral motion from rotational motion [2]. Since, rotations are not necessary to complete the task, the team may have been artificially reducing the number of degrees of freedom by stabilizing the yaw motion. Indeed several human-robot cooperative manipulation protocols employ alternate artificial constraints for the same purposes [2]. This seems to be an implicit type of Common Haptic Improvised Protocol (CHIP).

D. *Discussion*

We found it unexpected that verbal communication had no significant effect on completion time or workload. We offer several possible explanations. First, perhaps verbal exchanges are simply a social expectation and truly do not impact performance. Such expectations are not readily captured in the TLX survey. Secondly, perhaps the task is too easy. We performed initial follow up experiments with target sizes as small as 2 centimeters in diameter but did not observe any significant effect. Third, perhaps there is something in the nature of this task that obviates verbal exchanges. For example, maybe the use of the blindfold is a limiting factor in the participants walking speed. Also, in follow up interviews, a few leaders told us that a follower's active cooperation was only required in the gross motion phase. Once they stopped walking, all they needed the follower to do was support the weight of the table while the leader engaged in the fine positioning. Fourth, perhaps we were not considering the correct performance measures. Several related performance measures were investigated but failed to yield additional significant differences such as walking speed and deviation from a straight line trajectory. Finally, perhaps the CHIPs were a satisfactory proxy for verbal communication. However, with 9-10 of the groups employing them there is no way to control for their effect post hoc.

V. EXPERIMENT II: DESCRIPTION

We developed a follow up experiment, intended to be more difficult. We eliminated the gross motion phase and the blindfold, and made orientation a component of the task to reduce the dyad's ability to use it to transmit CHIPs.

A. *Task Description*

Figure 4 shows the revised experimental setup. The participants must move a 2.5 kg board, measuring 0.1 m by 0.03 m by 1.2 m. Handles are mounted at either end, to promote uniformity across participants in their choice of grasping locations. Our intention was to reduce the likelihood that roll angle could be used as a CHIP, by making the object thin and forcing a one handed grip.

The task requires them to move the board in the vertical plane to a "target" – a specified height between 0.50 and 1.5 meters measured from the ground and a specified pitch
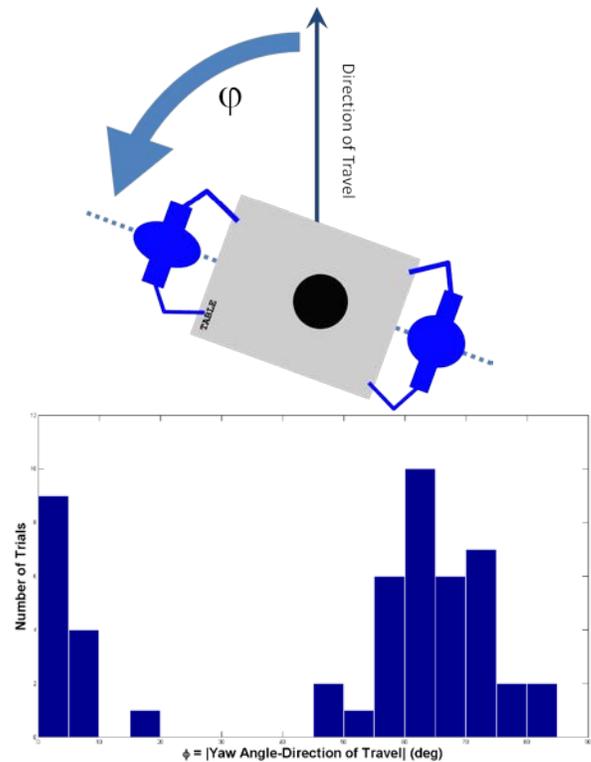

Figure 3: (Top) Illustration of the absolute heading difference between direction of travel and table's longitudinal axis (overhead view). (Bottom) A histogram of the steady state mean heading difference in degrees for all trials which exhibited a near constant steady state orientation.

between plus and minus 20 degrees. A sequence of 10 randomly generated targets was created, with the constraint that any subsequent target be at least 10 cm apart. This same sequence was used for all dyads to facilitate comparison.

Figure 4: Experiment 2. Two individuals move a board vertically. The cloth in the middle hangs from the ceiling, preventing the subjects from seeing each other. The dashed red line is a visualization of an example target position and orientation.

A target is considered "reached" when the board stays within plus or minus 0.005 m and 1 degree for at least 0.2 sec. The duration requirement was added to prevent groups from

"swinging" the board through the target. Requiring both a height and an orientation means that it is difficult to use the pitch angle as a CHIP. We also hoped that it would require the follower to take a more active role in the task.

B. Data Acquisition

The height and orientation of the board was respectively determined via a Hokuyo URG scanning laser range finder and a VectorNav inertial measurement unit (IMU) mounted at the center of the board. The laser reports a precision of 2 mm (95% confidence), while the IMU's orientation solution is precise to plus or minus 0.5 deg (95% confidence). The sensor package updates at 15 Hz. In addition, a video recording was made of the experiment, with the audio track providing a record of the verbal exchanges between the dyad.
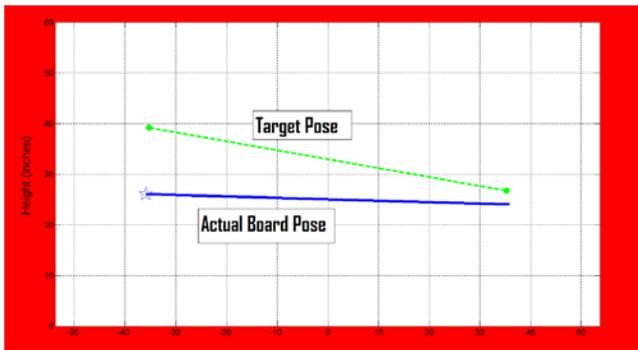

Figure 5: A screen shot of the realtime display see by the "leader", depicting the current position of the board as well as the target position.

C. Information Structure and Communication

The two participants were unable to see each other, thanks to the hanging curtain shown in Fig. 4, preventing them from seeing the facial cues or gestures of their partner; but still allowing them to have general spatial awareness and see the board's movements.

Each participant could see one of two computer monitors which displayed a schematic depicting the position and orientation of the board in realtime. One of the two individuals was randomly designated the *Leader* for the duration of the experiment. The leader's display (shown in Fig. 5) depicted the target position and orientation of the board (dashed line). The follower's display only depicted the current position of the board, but did not contain information about the target (no dashed line).

The same two conditions that were tested.
1. <u>Haptic Only</u>: Neither partner was permitted to speak.
2. <u>Haptic-Verbal</u>: Both partners were permitted to speak.

D. Participants, Procedure and Methodology

24 individuals (5 female, 6 left-handed), ages 18-22, were recruited for this study. Human subject testing approval was obtained from the institutional review board. Upon arrival, the two participants were seated in separate areas. They were given a consent form, which was used to exclude individuals with injuries preventing them from safely lifting a 2.5 kg object. A questionnaire was used to prevent the pairing of roommates or self-described close friends. They were instructed not to speak to their partner prior to the beginning of the experiment or during.

Before beginning the experiments, the subjects were asked to stand on different sides of the dividing curtain which hung from the ceiling (Fig. 4). Then, the investigator read instructions aloud from a script and performed a demonstration of the experiment. During the demonstration, the subjects were able to observe their respective computer screens and received printed, annotated screen shots which allowed them to understand the display. The subjects were told that their task was to move the object to a series of heights and orientations. They were also told that the specific locations would be provided on the computer screen to only one of the participants and that their goal was to complete the task as quickly as possible. Once both participants indicated that they understood the instructions, the experiment began immediately.

Each block of 10 targets took approximately 5 minutes. After each block, the participants were permitted to rest while completing a post-experimental survey that assessed the individual's perception of teamwork, physical demand, frustration, stress, and their own performance (matching the one in Sect. IV-B). In this counter-balanced, within-subjects experimental design, every dyad completed a block consisting of a 10 target sequence with each of the 2 experimental conditions. In total, each group moved the object to 20 targets during the 20 minute experiment.

Several steps were taken to mitigate learning effects. First, the target arrangement was randomly designed so that there was no obvious pattern (e.g. up-down-up-down…). Second, counterbalancing was used to control order effects. Finally, performance on the first 5 targets was considered "training" and is not used to compute a dyad's completion time.

VI. EXPERIMENT II: RESULTS AND ANALYSIS

A. Performance Analysis

The mean completion time (and standard deviation) for each experimental treatment was: Haptic-Verbal 6.64 s (2.56 s), and Haptic Only 11.25 s (6.00 s). As expected the Haptic-Verbal Condition resulted in a faster completion time. A repeated measures analysis of variance (ANOVA) test showed a significant effect ($p = 0.0294$) for communication type.

B. Subject's Perception of Performance and Workload

At the conclusion of each block, both participants were separated, and asked to complete a modified version of the NASA TLX survey [30]. The survey is used to rate their experiences along the dimensions presented in Table 1. Figure 6 depicts the mean responses graphically. On average, along all dimensions, both participants prefer the Haptic-Verbal communication type. For the leader, the difference between communication types along the 7L (Overall Satisfaction) and 8L (Overall Difficulty) dimensions are significant at the 0.05 level, after applying a Bonferroni correction. For the Followers *all* comparisons are significant at the 0.05 level *except* for 3F (Temporal Demand) and 6F

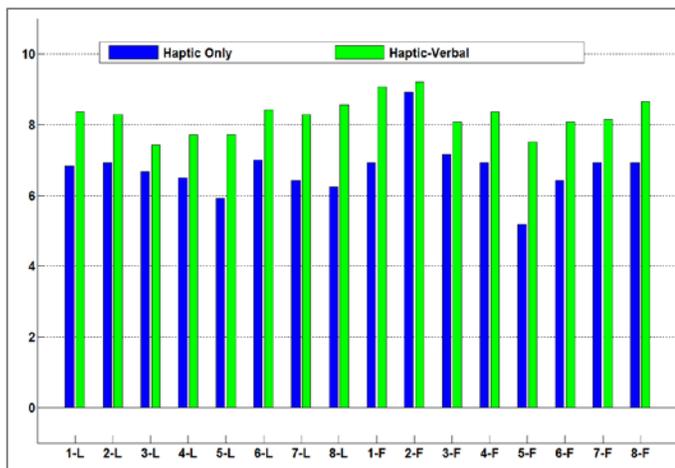(Performance Appraisal), after applying a Bonferroni correction.

Figure 6: The mean responses of Leaders (L) and Followers (F) to the 8 questions in the modified NASA Task Load Index (TLX) survey in Table 1.
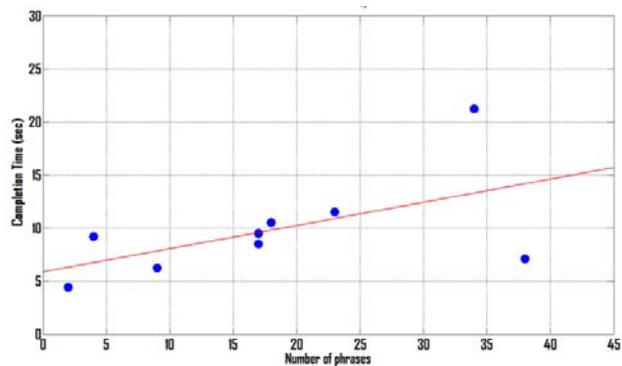
C. Additional Experimental Observations

Figure 7: Completion time vs number of verbal phrases exchanged.

While it was clear that the addition of verbal communication improved performance, the type or number of verbal exchanges has a less clear relationship to performance.

Regarding the types of exchanges, we transcribed the audio track and categorized phrases into one of multiple categories:

- Commands (e.g. "up", "stop"): 164 Occurrences
- High level (e.g. "the target is at eye level, tilted toward you"): 12 occurrences
- Confirmation (e.g. "yes", "OK"): 22 occurrences

Only three groups employed high level communication, with no apparent correlation to completion time. As Figure 7 shows, there is an overall positive correlation coefficient between completion time and the number of phrases exchanged. The Pearson correlation coefficient is 0.56; however, the relationship is not statistically significant (p =

0.113). This suggests while communication is helpful, more is not necessarily better.
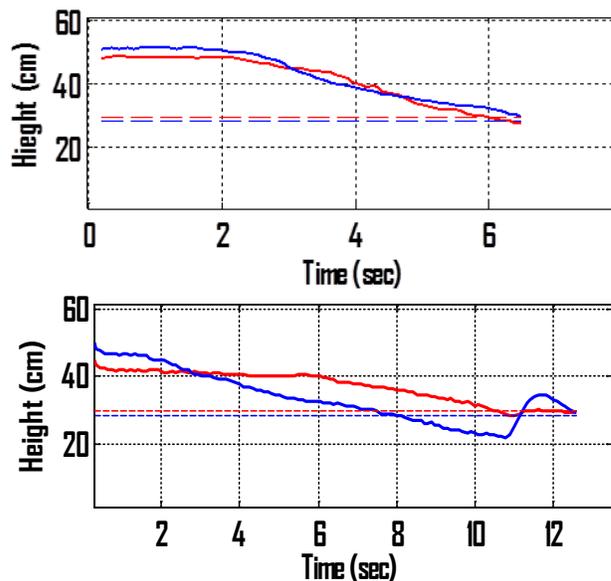

Figure 8: Sample hand heights for the Haptic Verbal (top) and Haptic (bottom) condition: Solid blue line is the Leader's hand position and solid red line is the Follower's hand position. Dashed lines represent desired hand position at target.

Upon examining the video recording, we noticed when verbal communication was prohibited 11 of the 12 groups employed CHIPs. All the observed CHIPs in this experiment were "unnecessary" motions in the vertical plane by the leader such as abrupt changes in direction, overshoot or shaking the board. No yaw or roll motions were observed. Figure 8 compares the hand motions of the leader and follower during a typical Haptic-Verbal trial with a Haptic Only trial where a CHIP was used. Both panels correspond to the same group traveling to the same target. Both partners are required to move down, but the leader must travel further to achieve the pitch specification. In the experiment depicted in the upper panel, verbal communication was permitted. The leader tells the follower ("down, down, slow, stop"). The leader never moved his hand in the opposite direction of his target position.

In the lower panel, no verbal communication was permitted. The leader moves down slightly quicker, but the follower is slower to move. The leader continues downward, past the desired hand position, in an effort to ensure the partner does the same. Once the follower is close to the desired position, the leader quickly jerks the board higher than the desired position. The partner briefly follows, but the leader quickly brings the board back down. We theorize this high velocity jerking motion is an attempt to non-verbally signal the follower to stop. It seems effective but takes them twice as

long to reach the target. Primarily because it took the follower a while to realize a CHIP was being employed.

In this task, vertical motions are required for task completion. However, for whatever reason, the participants also use motions in this direction for CHIPs. This can result in confusion which dramatically increases completion time in some situations. Consider Figure 9. The same group from Figure 8 attempts to reach a target requiring both leader and follower to hold their hands at the same height (dashed blue line). Initially the leader (blue line) moves up, but becomes frustrated by the fact that the follower is not holding the board high enough. The leader moves nearly 20 cms above the target. Finally the follower responds, reaching the correct height around 18 sec. However when the leader moves down abruptly to reach his own desired hand position, the follower fails to interpret this as a "stop" CHIP, instead mimicking his movement downward. This behavior repeats about 10 more times before the follower figures out that this gesture means "stop moving while I position my own hand". In all it too them approximately *10 times longer* to reach the target as compared to when verbal communication was permitted. It is worth noting that this trial took place after the one depicted in Figure 8. It is unclear why the CHIP was not successfully interpreted.
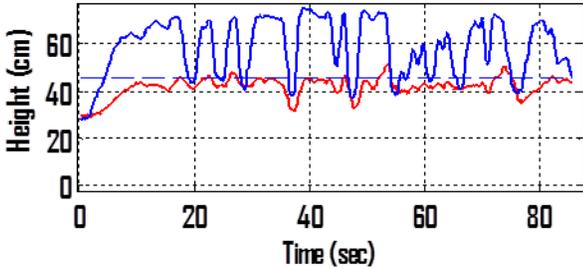

Figure 9: Sample hand heights for the Haptic condition: Solid blue line is the Leader's hand position and solid red line is the Follower's hand position. Dashed lines represent hand position at target.

VII. CONCLUSIONS AND RELATIONSHIP TO EXISTING KNOWLEDGE

In this paper we studied two person teams as they cooperatively manipulate a large object. Only one team member knows the location of the target. In one experimental condition the teammates cannot talk, while in the other they can. With regard to the research questions posed earlier, in the first experiment we found that verbal communication has a somewhat negative effect on performance, and a small positive effect on workload. However, neither result was statistically significant. In the second experiment, we found that verbal communication had a significant positive effect on both performance and some dimensions of workload.

There are several commonalities across the two experiments in how the participants chose to communicate. When they were allowed to speak, the leader did the vast majority of the talking. Most verbal exchanges were brief commands (e.g.

"stop", "up", etc.) rather than higher level plans. When the participants were not allowed to speak, we observed many teams developed common haptic improvised protocols – termed CHIPs. The two most common were: tipping the table in the desired direction of travel; and shaking the board. What was truly remarkable was that over 75% of the groups developed identical protocols without discussing the experiment in advance.

An unanswered question is why verbal communication improved performance in one task but not the other. Here we hypothesize that the nature of this difference, and the role of CHIPs, can be explained through an analysis of the task as a control system [31]. Consider the schematic in Figure 10 (left), depicting Experiment 1 and the control system written in state space form:

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{\psi} \end{bmatrix} = \begin{bmatrix} 1/m & 0 & 1/m & 0 \\ 0 & 1/m & 0 & 1/m \\ 2/LJ & 0 & -2/LJ & 0 \end{bmatrix} \begin{bmatrix} F_{LX} \\ F_{LY} \\ F_{FX} \\ F_{FY} \end{bmatrix}.$$

Here m is the mass of the object, J is the rotational moment of inertia, and L is the length. There are only two output variables that need to be controlled to complete the task – the X and Y position of the table (body fixed frame). One may consider the inputs to be the leader and follower's applied forces in the X and Y direction. While there are four inputs, it is possible to control the two output variables with only two properly chosen inputs.

If the follower employs the implicit CHIP, $F_{FX} \equiv F_{LX}$, and $F_{FY} \equiv F_{LY}$, one can show that the output controllability matrix is now full rank. By making that substitution the yaw angle is no longer controllable (though it is stable); however, changes in the yaw angle are not required to complete the task. This may be an explanation for why the groups did not rotate the table. It may also explain why groups were able to complete the task without talking, when the target size was dramatically reduced.

On the other hand, consider Figure 10 (right and the state space model for Experiment 2:

$$\begin{bmatrix} \ddot{y} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 1/m & 1/m \\ -2/LJ & 2/LJ \end{bmatrix} \begin{bmatrix} F_L \\ F_F \end{bmatrix}.$$

There is no single "follower" protocol that will render the second task controllable, since controlling two independent output variables requires at least two independent inputs.
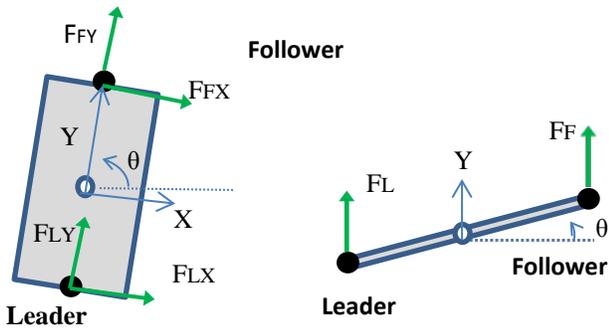
Figure 10: (Left) Table Positioning. (Right) Board Positioning.

Regarding the explicit CHIPs, going back to Experiment 1, the leader can exert forces in the verticle direction resulting in small pitch or roll motions, that have almost no impact on task completion. We speculate that a CHIP transmitted on an "unused" haptic channel may be as effective as explicit communication transmitted over a audiory channel, because they are not confused with task motions. In constrast, for whatever reason, most groups felt the most intuitive way to trasmit CHIPs in Experiment 2 was to use the verticle force channel. This is an unfortunate choice since forces along this direction are required for task completion. In the best cases, this results in inefficient motions (small overshoot) and in the worst cases this causes a cycle of confusion as illustated in Figure 9.

Regarding lessons learned for human-robot interaction, there are few noteworthy points. First, humans do appear to exhibit a preference for explicit communication (though not necessarily verbal) during cooperative manipulation tasks. Second, when the communication is of the verbal form a surprisingly small vocabulary of one word commands is sufficient for the tasks described here. Lastly, we recognize that an individual's perception of task difficulty is just as important as the actual performance when working with human-human and human-robot systems. Evidence shows that if an individual perceives a task to be easier, he or she is more likely to engage in that task.

References

[1] J.Y.S. Luh and L.F. Zheng, "Constrained relations between two coordinated industrial robots for motion control," *International Journal of Robotics Research*, vol. 6, pp. 60-70, 1987.

[2] Y. Hirata, Y. Kume, Z. Wang, and K. Kosuge, "Decentralized control of multiple mobile manipulators handling a single object in coordination," in *Proc. IEEE/RSJ conference on Intelligent Robots and Systems*, pp. 2758–2763, Oct. 2002.

[3] V. Hayward, O.R. Astley, O.R., M. Cruz-Hernandez, D. Grant, and G. Robles-De-La-Torre, "Haptic interfaces and devices," *Sensor Rev.,* vol. 24 no.1, pp. 16-29, 2004.

[4] A.M. Okamura, C. Richard, and M.R. Cutkosky, "Feeling is believing: Using a force-feedback joystick to teach dynamic systems." *ASEE Journal of Engineering Education,* vol. 92, no. 3, pp. 345-349, 2004.

[5] D. Ruspini, and O. Khatib, "Haptic display for human interaction with virtual dynamic environments," *Journal of Robotic Systems,* vol. 18, no. 12, pp. 769-783, 2001.

[6] T. Sugar and V. Kumar, "Control of cooperating mobile manipulators," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 1, pp. 94 – 103, 2002.

[7] J. Moll and E.L. Sallnas, "Communicative Functions of Haptic Feedback," *Proc. Fourth Int'l Conf. Haptic and Audio Interaction Design*, pp. 1-10, 2009.

[8] E.L. Sallnas, K. Rassmus-Grohn, an dC. Sjostrom, "Supporting Presence in Collaborative Environments by Haptic Force Feedback," *ACM Trans. Computer-Human Interaction*, vol. 7, no. 4, pp.461-476, 2000.

[9] E.L. Sallnas, "Improved Precision in Mediated Collaborative Manipulation of Objects by Haptic Force Feedback," *Proc. First Int'l Workshop Haptic Human-Computer Interaction*, pp. 69-75, 2001.

[10] S. Glynn, R. Fekieta, and R.A. Henning, "Use of Force-Feedback Joysticks to Promote Teamwork in Virtual Teleoperation," *Proc. Human Factors and Ergonomics Soc. Ann. Meeting*, 2001.

[11] K. Reed, "Cooperative Physical Human-Human and Human-Robot Interaction", A. Peer and C. Giachritsis, Editors, *Immersive Multimodal Interactive Presence, Springer Series on Touch and Haptic Systems*, 2012.

[12] R. Ikeura and H. Inooka, "Cooperative force control in carrying an object by two humans," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2307-2311, 1995.

[13] C. Bubb-Lewis, and M.W. Scerbo, "The effects of communication modes on performance and discourse organization with an adaptive interface," *Applied Ergonomics,* vol. 33, no. 1, pp. 15-26, 2002.

[14] L. LeBigot, E. Jamet, and J.F. Rouet, "Effects of speech- and text-based interaction modes in natural language human-computer dialogue," *Human Factors* vol. 49, no. 6, pp. 1045-53, 2007.

[15] E.L. Sallnas, and S. Zhai, "Collaboration meets Fitts' law: passing virtual objects with and without haptic force feedback," *Proceedings of INTERACT*, pp. 97-104, 2003.

[16] R. Groten, J. Holldampf, A. Peer, and M. Buss, "Predictability of a human partner in a pursuit tracking task without haptic feedback," in *Proc. IEEE Second International Conference on Advances in Computer-Human Interactions,* pp. 63-68, 2009.

[17] C. Basdogan, C.H. Ho, and M.A. Srinivasan, "An experimental study on the role of touch in shared virtual environments," *ACM Transactions on Computer-Human Interaction,* vol. 7, pp. 443-460, 2000.

[18] K.B. Reed, and M.A. Peshkin, "Physical collaboration of human-human and human-robot teams," *IEEE Transactions on Haptics,* vol. 1 no. 2, pp. 108-120, 2008.

[19] R. Groten, D. Feth, R.L. Klatzky, and A. Peer, "The Role of Haptic Feedback for the Integration of Intentions in Shared

Task Execution,' IEEE Trans on Haptics, vol. 6, no. 1, pp. 94-105, 2013.

[20] S. Gentry, R. Murray-Smith, and E. Feron, "Human-human haptic collaboration on a cyclical Fitts' task", *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3402-3407, August 2005.

[21] M.M. Rahman, R. Ikeura, and K. Mizutani, "Cooperation characteristics of two humans in moving and object", *Machine Intelligence and Robotic Control*, Vol. 4 pp. 43-48, 2002.

[22] Y.Y. Huang, J. Moll, E. Sallnas, and Y. Sundblad, "Auditory feedback in haptic collaborative interfaces," *Int. Journal of Human-Computer Studies*, vol. 70, pp. 257-270, 2012.

[23] D.A. Kenny, C.D. Mohr, and M.J. Levesque, "A social relations variance partitioning of dynamic behavior", *Psychological Bulletin*, vol. 127, no. 1, pp. 128-141, 2001.

[24] R. Groten, D. Feth, H. Goshy, A. Peer, D.A. Kenny, and M. Buss, "Experimental analysis of dominance in haptic collaboration," *IEEE International Symposium in Robot and Human Interactive Communication*, pp. 723-729, 2009.

[25] S.S. Nudeha, R. Mukherjee, and M.Ghodoussi, "A shared control approach to haptic interface design for minimally invasive telesurgical training*," IEEE Transactions on Control System Technology*, vol. 13, pp. 588-592, 2005.

[26] B. Khademian and K. Hashtrudi-Zaad, "Performance issues in collaborative haptic training", *IEEE International Conference on Robotics and Automation*, 2007.

[27] K. Kosuge, H. Yoshida and T.Fukuda, "Dynamic control for robot-human collaboration", *IEEE International Workshop on Robot and Human Communication*, pp. 398-401, 1993.

[28] R. Ikeura and H.Inooka, "Variable Impedance control of a robot for cooperation with a human," *IEEE International Conference on Robotics and Automation*, pp. 3097-3102, 1993.

[29] H. Arai, T. Takubo, Y. Hayashibara, and K. Tanie, "Human-robot cooperative manipulation using a virtual non-holonomic constraint", *IEEE International Conference on Robotics and Automation*, pp. 4063-4069, 2000.

[30] S.G. Hart and L.E. Staveland, "Development of NASA TLX (Task Load Index): Results of empirical and theoretical research", *Human Mental Workload*, pp. 139-183, 1988.

[31] K. Ogata, *Modern Control Engineering 3rd edition*, Prentice Hall, NJ, 1997.

Joel M. Esposito earned the BS degree in mechanical engineering from Rutgers University, New Brunswick, NJ, USA and the MS and PhD, in 1999 and 2002 respectively from the University of Pennsylvania, Philadelphia, PA, USA.

He is an Associate Professor in the Systems Engineering Department at the United States Naval Academy, Annapolis, MD, USA; and currently serves as the Editor in Chief of the American Society of Engineering Education's Computers in Education Journal. His research interests include multi-robot coordination, motion planning, computational tools for motion control and robotics pedagogy. Prof. Esposito is the recipient of the Naval Academy's 2010 R. A. Raouff Teaching Award.

Jeremy L. Searock was born in Nazareth, PA and attended the United States Naval Academy in Annapolis, MD graduating with a BS in Systems Engineering in 2003. He then attended the Robotics Institute at Carnegie Mellon University, Pittsburgh, PA graduating with a MS in Robotics in 2005.

After graduate school, he completed the Naval Nuclear Propulsion Training Program and reported to the USS Nevada (SSBN-733), a nuclear powered submarine, where he served as the Main Propulsion Assistant, Quality Assurance Officer, and Assistant Engineer. He currently is an engineer at Carnegie Melon University's National Robotics Engineering Center. His research interests include autonomous robot navigation, fault detection, and autonomous surface vehicle development.

Sarangi P. Parikh earned the BS degree in Physics from Fairfield University, Fairfield, CT and the MS and PhD in Mechanical Engineering, in 2001 and 2005 respectively, from the University of Pennsylvania, Philadelphia, PA.

She is an Assistant Professor in the Systems Engineering Department at the United States Naval Academy, Annapolis, MD. Her research interests include human-robot interaction, human-computer interactions, and interfacing with intelligent technologies.