Koenig, N., Takayama, L., & Mataric, M. (2010). Communication and knowledge sharing in human-robot interaction. Neural Networks, 23(8-9), 1104-1112

Communication and Knowledge Sharing in Human-Robot Interaction

Nathan Koenig[a], Leila Takayama[b], Maja Matarić[a]

[a] *Viterbi School of Engineering, Univ. of Southern California, Los Angeles, CA 90089*
[b] *Willow Garage, 68 Willow Road, Menlo Park, CA 94025*

Abstract

Inexpensive personal robots will soon become available to a large portion of the population. Currently, most consumer robots are simple single-purpose machines or toys. In order to be widely accepted, robots need to accomplish a wide range of tasks in diverse conditions. Learning these tasks from demonstrations offers a convenient mechanism to customize and train a robot by transferring task related knowledge from a person to a robot. This avoids the time-consuming and complex process of manual programming. The way in which a person interacts with a robot during a demonstration plays a vital role in terms of how effectively and accurately a person is able to provide the demonstration. Teaching through demonstrations is a social activity, one that requires bi-directional communication between a teacher and a student. This work studies how visual observation of a robot and audio cues affect a persons ability to teach in a social setting. Results show that audio cues provided important knowledge about a robot's internal state, while visual observation of a robot can hinder an instructor due to development incorrect mental models of the robot.

Keywords: learning from demonstration (LfD), audio, visual, teaching

Email addresses: `nkoenig@usc.edu` (Nathan Koenig), `takayama@willowgarage.com` (Leila Takayama), `maja@usc.edu` (Maja Matarić)

URL: `http://robotics.usc.edu/~nkoenig` (Nathan Koenig), `http://www.willowgarage.com` (Leila Takayama), `http://robotics.usc.edu/~maja` (Maja Matarić)

1. Introduction

Technological advancement is a never ending force, which will soon provide the public at large with personal robots capable of fulfilling a wide variety of tasks. This can been seen with advances in toys, such as the recently discontinued Sony Aibo, Woowee's Robosapian, and numerous robot kits, and consumer and research robots, such as the Nao from Aldebaran Robotics [1], PR2 from Willow Garage [2], and Sarcos humanoid [3].

As hardware makes steady strides forward, so must algorithms and software. Research generally leads implementation by roughly twenty years, and at the forefront of robotic research exists and array of learning and adaptation algorithms designed to make it possible for robots to achieve complex tasks in real-world scenarios.

Learning algorithms are not the only means to achieve robot programming. Alternative methods rely on manual generation of state machines, or other forms of control logic. This approach is very useful in many situations where the domain is simple or well known. However complex tasks are difficult to handle manually, especially if the domain has unknown variables. In these situations, it may be more efficient to utilize a form of machine learning to generate a control policy. Some algorithms are unsupervised and require no human input, while others take advantage of a human's knowledge by incorporating their input into the learning process.

The application of learning algorithms currently still requires significant experience on the part of the developer. As commercially available multi-purpose robot become available, there will be a need to customize their programming to fit different and unique situations. If specialized skills are required to program a robot, their adoption into society will be stunted.

Programming a robot can be thought of as transferring knowledge from a person to a machine. One of the first methods for improved knowledge transfer relied on a computer scientist to encode knowledge directly from a subject matter expert, such as a doctor, into a computer friendly format. The most common technique required the computer scientist to interview one or more experts and codify their answers into a knowledge base. This process is called knowledge engineering, and the result was a knowledge base that could be used by a software program, or expert system, to answer queries and resolve problems [4]. While the expert systems could respond consistently and accurately, the form of knowledge transfer was clumsy, prone to error, and in many situations the subject matter expert could not explain the logic behind their reasoning.

While expert systems can be applied to robotics [5, 6], they are not easily adaptable and are encumbered by a middle man that acts as an interpreter. Imitation learning is an alternative approach that utilizes only observations of someone or something performing a task or action to derive a control policy [7]. A robot can use this technique without the knowledge of the observed. One major stumbling block is mapping movements from the observed to the observed. If both parties share common physical attributes, then there are fewer problems. However a robot typically does not have the same characteristics as a human, and the process is therefore much more challenging.

Inspiration for imitation learning as applied to robotics comes from insights into how humans and primates learn. Our bodies are well adapted to watching and mimicking others. We can do this without much thought and concentration. Giacomo Rizzolatti has shown that our brains have special neurons, called mirror neurons, that allow us to see, process, and copy actions of other people [8]. The mirror neuron has even inspired roboticists to devise approaches to robot learning based on our physiology [9, 10].

It has also been argued that imitation can help bootstrap social cognition, in much the same way that babies imitate and learn from adults [11]. Even our ability to empathize has been tied to the mirror neuron [12]. Imbuing our natural ability to learn from imitation into a robot is still well beyond the reach of current technology. So, rather than attempting to mimic our mirror neurons, we take an approach that plays to the strengths of both man and machine.

Learning from demonstration (LfD) relies on a human teacher to actively guide a robot student through a task. Imitation learning and LfD are commonly used interchangeably, however there is a difference. LfD utilizes a deliberate and active teacher that provides task demonstration to a robot, while imitation learning relies on remote observations of a person who may not conscientiously be a teacher. LfD is a social activity that relies on communication between both the instructor and the student. The instructor should understand what is going on in the mind of the student, and the student conveys this information through verbal and non-verbal communication. Without this flow of information an instructor can only lecture, and while this method of teaching is commonly used in a classroom it's not an effective approach for teaching tasks.

Throughout our lives we both give and receive demonstrations. Parents provide their children with numerous demonstrations, such as how to swing a bat, ride a bike, and tie a neck-tie. Since this practice is so common, we can safely assume that a majority of people are capable of demonstrating tasks to another person or even a robot. The hurdle a robot must overcome is the ability to act in a social setting, and relay information back to the teacher in a meaningful manner. Through the use of LfD, the complexity associated with programming a robot has been reduced. A much larger section of the population can then theoretically “program” a robot, assuming a robot is a capable student.

Our ability to act as a student and express our internal state and understanding to an instructor comes just as naturally as our ability to act as the instructor. This is not true for a robot, which has not had the luxury of years of growing in a social setting with a complex brain and developing an appropriate theory of mind [13, 14]. Brian Scassellati has developed a framework that begins to apply the theory of mind to robots, through face detection, gaze tracking, and discriminating animate from inanimate objects [15]. These advances show that, while limited, a robot can begin to act in a social setting.

Through the use of creative hardware and software, a robot can even begin to mimic face and body expressions that humans naturally exhibit [16]. Other robots lack some of the expressive capabilities and instead focus more on utilitarian purposes [17]. With either type, it’s possible to provide feedback using some communication medium, such as audio, gestures, facial expressions, graphical interfaces, or even simple actions designed to convey intentions [18]. Picking a method of communication is sufficient only if it is easily and correctly interpreted by a human.

In this work we use the Willow Garage PR2 [2] as the robot student. The PR2 has a humanoid torso with stereo cameras and laser range finders on a mobile base. Since it does not have an expressive face, we rely upon audio communication and a graphical interface that can display text messages. In order to avoid gender issues and limit the robot’s portrayal of unrealistic intelligence, we use only non-verbal audio signals. Changes in pitch, tone, and frequency allow us to create audio cues that can indicate concepts such as success (rise in pitch, such as uttering “whoohoo”), failure (drop in pitch, such as “uhh-ohh”), and acknowledgment (such as “okay”).

The PR2 is capable of detecting and manipulating simple objects with two arms and navigating in cluttered indoor environments. These attributes make it the ideal platform for executing table-top manipulation tasks, such as object sorting and playing simple games. Both of these styles of tasks were used in this work in the form of sorting colored blocks into bins and a three-disk version of the Towers of Hanoi puzzle.

Environmental factors are often overlooked when teaching takes place. With high-speed Internet connections and streaming video, it's very possible to teach remotely. The instructor may not be able to see the student completely or at all. This makes it much more difficult on the part of the instructor to interpret the state of the robot student, and also more challenging for the student to communicate back to the teacher.

In this work we specifically look at what affect visual obstruction of a robot student has on an instructor when they are giving a demonstration. Performance of the instructor can be measured using the time it takes to complete a demonstration, an instructors level of frustration and effort, and the number of commands they issue to the robot. Using the Towers of Hanoi puzzle as an engaging task domain, some instructors were placed behind a curtain while some had direct visual access. Both sets of instructors used a graphical interface to teach the robot, which displayed a live video feed from the robot's point of view. Without audio cues, the results indicate that instructors provided with direct visual access had the tendency to generate improper mental models of the robot's state.

This shows that even though the PR2 is clearly not a human and does not behave as one, we still build an internal model of the robot's state. It is therefore vital that a robot communicate its state effectively, so that people interpret the state of the robot properly. In a follow-up study, we allowed the robot to emit audio cues as an additional communication mechanism to the graphical interface. These cues allowed the instructors to identify when the PR2 understood a command and if it was successful at completing it. This little bit of information allowed instructors to significantly improve their performance.

Once people start taking an active role in the learning process of robots, it's important to understand how to best achieve human-robot interaction. As a population we are relatively new to the idea of interacting with robots. While people are very adaptable, it's not realistic to force people to change habits to meet a robot's needs and limitations. Roboticists should instead incorporate social skills into a robot, allowing for natural interactions.

The rest of this paper is structured as follows. Section 2 introduces learning from demonstration, the role of the student and teacher, and. Section 4 briefly discusses the internal representation of tasks that the robot uses.

Section 3 touches upon how humans and robot can communicate, and the reasons for choosing a graphical interface. Sections 5 and 6 presents the studies involving the Towers of Hanoi puzzle and block sorting. Section 10 summaries this work and discusses the direction of future work.

2. Learning from Demonstration

Learning from demonstration is a general classification for a technique that encompasses a situation in which someone, or something, acts as an instructor that guides another entity, acting as a student, through a task. The expectation is that the student will learn how to accomplish the task following one or more demonstrations. A task can be anything from picking up an object to swinging a tennis racket. In most situations the term LfD is used when the instructor is a person and the student is a robot. This is the definition that is used for the rest of this paper.

The type of task that can be taught to a robot is limited primarily by its physical properties. Often LfD is applied to learning joint trajectories, such as how to reach and move an arm [19, 20]. However, high-level tasks such as how to bake a cake [21], or perform complex pick-and-place actions in a cluttered environment [22] can be taught as well. Generally LfD is useful when a control policy is difficult to define, or an expert (other than the system designer) is needed.

One of the primary components in the process of learning by demonstration is the interface through which a teacher communicates with the robot. Many different mechanisms have been used, including joysticks [23, 24], observation of the teacher by the robot [25], manual manipulation of the robot's physical structure [26], and graphical interfaces [27]. For this work, we have chosen to utilize a graphical interface to facilitate the transfer of knowledge from teacher to student.

A graphical interface provides a rich and customizable medium. Observation based interaction relies on the robot sensing and interpreting a human which is often difficult, subject to noise and a variety of environmental factors. Joystick control and manual manipulation requires the instructor to be trained about the robot's kinematic structure. Due the ubiquity of personal computers in our daily lives, most people are familiar with graphical interfaces. This helps reduce the learning curve for the instructor using a graphical interface and improves their comfort level when interacting with the robot.

Once one or more demonstrations have been provided, it's assumed the robot will learn a corresponding control policy. The learning algorithm used by the robot is chosen by the system designer. A common approach is to use reinforcement learning [28, 29]. Other methods utilize regression techniques [30, 31]. The approach used in this work has the robot learn decision networks [32], an extension to Bayesian networks, which incorporate the concept of choosing what action to take based on the robot's current state. In order to constrain this discussion, we will save the details of decision networks to a future article.

2.1. Role of the Student

As a student, the job of a robot is to observe itself and its environment during a demonstration. We assume that an instructor demonstrates a task to the best of their ability without introducing malicious instructions. Based on the information gathered, and all similar demonstrations, the robot should construct or adapt a representation of the task that is suitable for autonomous execution.

An instructor relies on information from the student to judge whether or not the student understands what is being taught, and how to proceed with the instructions. It is therefore a vital role of the student to accurately convey state information in a timely manner for learning to proceed smoothly.

2.2. Role of the Instructor

In a human-human teaching scenario, an instructor acts as an expert on a particular subject. It is no different when a person teaches a robot a task. We assume the instructor is knowledgeable on the subject at hand, and is capable of generating a demonstration in a non-malicious manner and as error free as possible.

Providing a demonstration is not classroom lecturing where the instructor can speak freely to an audience with little to no feedback. Rather the instructor relies on information from their student in order to guide the demonstration. The instructor must therefore be able to interpret cues from the student that offer insight into their understanding (or lack thereof) of the subject matter.

In the context of teaching robots, it's very possible that the instructor does not understand the robot's physical and expressive capabilities. It is unrealistic to assume the instructor will have sufficient time and expertise to learn about the robot prior to giving demonstrations. Instead the burden is placed on the robot. It must express itself in a manner that is easily understood by the general public. Through this approach, instructors can behave more naturally and provide better demonstrations to the robot.

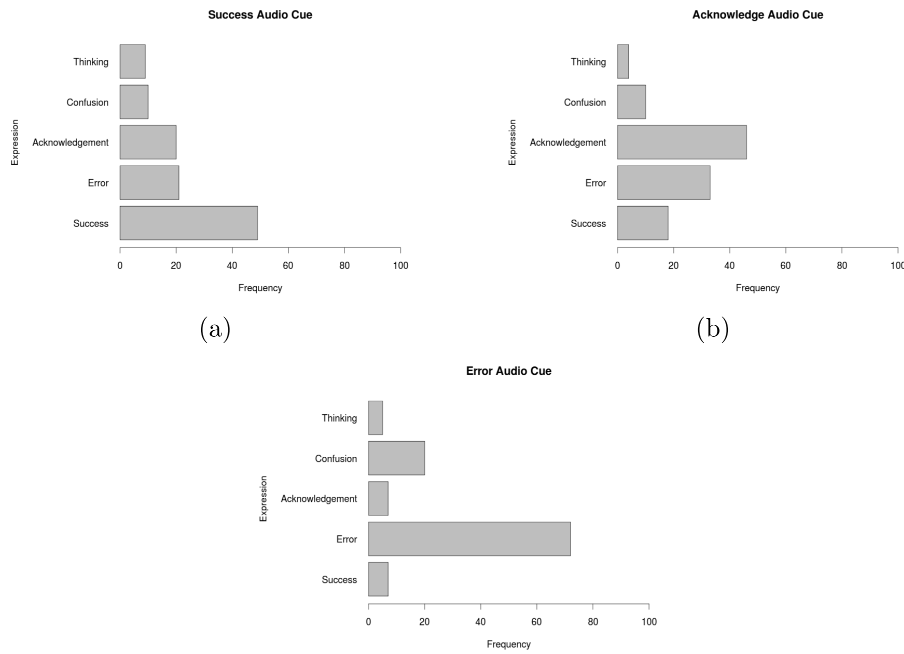
3. Human Robot Communication

The ability to communicate efficiently and naturally is a requirement for any teaching scenario. For most situations it is assumed that both parties share a common communication medium. However once a robot is introduced this assumption is no longer valid. With little day-to-day robot interaction, most people have ill defined notions of how to interact with a robot. Furthermore, any information provided to a person about a robot prior to interaction will skew their beliefs about the robot's capabilities [33].

Given that an instructor most likely has little or no knowledge about how a robot communicates, the robot should communicate through a medium that is natural and easily understood by the instructor. By following this guideline, the instructor can act more comfortably and ideally provided better demonstrations. Both input to the robot and output from the robot should be expressive enough to capture relevant information efficiently and intuitively for an instructor to use with minimal to no prior training.

We will first tackle the problem of robot input. During a demonstration, the primary job of the instructor is to pass information to the robot. Numerous options are available to the robot including speech recognition, visual observation, joysticks, direct manipulation, and a graphical interface. There are pros and cons of each option. Speech recognition and visual observation place the least burden on the instructor, but are very difficult to implement robustly. Joysticks are common tools for many people, however it's difficult to control a robot with many degrees of freedom via a joystick. Direct manipulation essentially removes the joystick by allowing an instructor to manually move the robot during a demonstration. This hands-on approach is time-consuming, and relies on the instructor having intimate knowledge of the kinematic and dynamic capabilities of the robot.

The approach used in this work utilizes a graphical interface as the primary medium through which information is sent to and from the robot. A graphical interface allows for data abstraction, such as visualization of laser scans, thereby removing unnecessary details. The proliferation of GUI's in our everyday lives, from desktop PCs to mobile phones, allows them to be almost universally accepted. Finally, a well designed GUI can be used by



(a) (b)

(c)

Figure 1: Histograms of the sound survey. Each horizontal bar is the frequency at which the expression was selected for the graph's audio cue. Graph A is the audio cue chosen to represent success, graph B for acknowledgment, and graph C for error.

many different robots, without any significant changes. Just as web browsers can render many different web pages, its realistic to assume that a GUI could interact with many different robots.

3.1. Audio Cues

Robots can also make use of natural communication methods in addition to graphical interfaces. Audio cues are one medium that is readily accessible to robots, and easy for a person to understand if properly designed.

Utilizing the speakers on a robot to emit sounds is trivial, the important aspect is choosing sounds that are meaningful and understandable to the general population. Sounds that may cause confusion or misunderstanding will only hinder the learning process, and degrade the demonstrations provided by a teacher. We have therefore endeavored to validate sounds that a robot could use through surveys in an attempt to limit any ambiguities.

We placed an additional constraint of using non-verbal communication on the robot. This decision allows the robot to remain gender neutral, prevent it from giving off a persona of intelligence beyond its capabilities, and eliminates language barriers. Non-verbal cues are also easier to hear as they are typically based upon pitch, volume, and frequency differences rather than spoken word.

Working with sound designers from Pixar Animation Studios, we developed a set of ten non-verbal audio cues. In order to create a valid mapping from sound to meaning, we designed an online survey. A participant completing this study listened to the sounds in random order. For each sound, they are asked what expression best matches the sound. The list of expressions included *success*, *error*, *acknowledgment*, *confusion*, and *thinking*. The participants are informed that the sounds are used by a robot, and a picture of a robot is accompanied with the sound. Following the survey we gathered demographic data, including age, gender, robotics experience, and whether they are a native English speaker.

This study was posted on Mechanical Turk, and we gathered information from 100 people. The results indicate that some sounds are less ambiguous than others. This has helped guide our choice of sounds for the robot, with the goal of reducing misinterpretation by a teacher.

The three best sounds, one for each of *success*, *error*, and *acknowledgment*, were chosen from the results. Figure **??** shows the frequency of selecting an expression for three best sounds. The error and success sounds had the greatest separation. Acknowledgment was more ambiguous, however, when used in practice every participant understood its meaning. This most likely stems from more contextual information.

4. Task Representation

The representation used by the robot to internally store knowledge about a task is a key component in a learning machine. This is particularly important when the robot must contend with the possibility of storing and referencing large numbers of tasks. The representation used must therefore be light weight in terms of its memory footprint, easily accessible, and adaptable. We have placed additional constraints that will allow the robot to share knowledge among other heterogeneous robots, and allow for online learning through demonstrations.

Based on these requirements, relatively few options remained viable candidates. Most supervised learning algorithms work well for small state spaces, or fail when a large amount of noise is present in the data. Reinforcement learning does not scale well to large state spaces, and requires significant number of training cycles. Gaussian processes and other regression methods require a significant amount of data which precludes teaching from demonstration. Decision trees suffer from exponential branching based on the number of variables modeled.

As a result we have settled on decision networks, also known as influence diagrams. A decision network is a generalization of Bayesian networks. They include the standard structure from Bayesian networks, as well as decision nodes and values nodes. A decision node defines a set of actions that are available for execution, and structurally they enforce a time ordering in the network. Value nodes define a utility function that is associated with a single decision node. This utility function determines which actions are most appropriate for execution by following a maximum expected utility criterion. A network can be graphically represented using unique shapes for each node type and edges that encode conditional independence.

This article's focus is on the process of teaching and communication between a human and robot. Details associated the decision networks, and how they are generated and used is left to future work.

5. Towers of Hanoi

In this experiment, we examine the effects of visual obstruction during human robot interaction. Based on prior work, we believed that direct visual access to a robot will improve a persons ability to understand and estimate a robot's state. By having a more complete understanding of a robot's state, an instructor will ideally provide better demonstrations.

We have chosen the Towers of Hanoi puzzle as the task to be learned by the robot. Towers of of Hanoi is a classic puzzle that consists of three poles and N disks of decreasing size. The starting state for the puzzle has all the disks on the left-most pole, and the goal is to move all the disks to the right pole. Two rules must be followed, the first states that only one disk may be moved at a time, and the second is a larger disk may not be placed on top a smaller disk.

This toy problem was selected for several reasons. For the human, the puzzle is sufficiently challenging to engage a person, without being frustratingly complex (given a small N). For the robot, it is a reasonably solvable task that is used in artificial intelligence courses because it involves a closed and highly constrained problem space. For the purposes of evaluating the human-robot interaction, the puzzle is also well suited for evaluation of a teacher's performance because there is an optimal solution, allowing for an objective comparison between teachers' performances.

The results of this study suggest directions for future work into human-robot interaction. We are specifically interested the optimal conditions in which to conduct learning by demonstration sessions, and what environmental factors may affect a teacher's actual and perceived performance.

5.1. Background

To our knowledge, this is the first study to look at the effect of visual obstructions when interacting with a robot. Prior research in human-human interaction may lend some insight into how people will respond with a robot. Most work related to interacting between two or more humans when obstacles are present look at face-to-face communication versus mediated communication. These studies examine the effects of visibility [34, 35] upon human communication and coordination.

Sociological literature has shown that face-to-face contact improves trust among humans and increases cooperation [36]. Even voice communication shows marked improvement over text based alternatives [37].

With voice and face-to-face communication, people are better able to evaluate each others state and attitude. While a robot does not have the same depth of communication as a person, we expect that the visibility of a humanoid robot will benefit a person's ability to understand the state of the robot.

It is quite possible that people in this study will extrapolate their own internal models to fit the characteristics of a robot. Work has shown that people are adept at creating accurate mental models of humanoid robots [38, 39]. These models were used to by people to determine the competencies of the robots, based on weak hypotheses. Teachers with an unobstructed view of the robot may be more likely to build an internal model of the robot. This could help the teacher choose when and which to send commands to the robot.

5.2. Towers of Hanoi GUI

The graphical interface developed for the teaching a robot the Towers of Hanoi puzzle consists of five web-based widgets as seen in Figure 5b. Each widget is self-contained entity with clear borders and a descriptive title. The amount of text and buttons has been kept to a minimum in order to reduce confusion.

Starting with top right widget and moving clockwise, the GUI contains an *Action Selector* widget that allows a teacher to control the robot. Inside this widget are two text lists. The list on the left, labeled *Actions*, contains all the available actions. These actions are movements that the robot can perform without any additional guidance from the teacher. To the right of this list is the *Objects* list. The elements in the list are all the entities in the environment that the robot can detect and interact with. Again, object recognition software has been provided to the robot without additional instruction from the teacher.

For this study, the available actions included *Move-to*, *Pick-up*, and *Put-down*. The choice of labels for these actions are design to be self-explanatory. For example. the *Move-to* action will move the robot's gripper to an object, and the *Pick-up* action will cause the robot to grasp and lift an object. The detectable objects include all the discs (*Red*, *Green* and *Blue*) and the poles (*Left*, *Middle*, and *Right*) on the game board.

By using the available actions and objects, the teacher can construct a command that the robot can easily understand. For example, by selecting *Move-to* and *Red Disk* and selecting the *Execute* button, the robot will move to the red disk. These verb-noun pairs serve to help the teacher understand the capabilities of the robot by phrasing the robot's capabilities in human-readable form.

Below the *Action Selector* widget is the *Robot Status* widget. Inside this widget are displayed status messages from the robot. This includes messages pertaining to the progress the robot is making toward completing a given command, and error messages. The position of this widget was carefully chosen to lie closest to the *Action Selector* widget. The purpose behind this decision was to promote the two-way communication that occurs via the GUI, and encourage the teachers to use the status messages during the teaching process.

The next widget is titled *Task Complete*, and contains a single button. This widget exists to provide a mechanism that indicates when a demonstration is complete.

Moving clockwise, the *Instructions* widget displays a concise list of instructions for the teacher. While not entirely necessary, providing some measure of help on the GUI prevents the teacher from looking elsewhere for technical help. Again, this was designed to ensure that the teacher stayed focused on the demonstration task.

The final widget located in the upper left is the *Camera View*. The image displayed in this widget is a video stream from robot's camera. This widget's primary purpose is to provide the teacher with context about what the robot can observe.

5.3. System Interaction

A complete system includes a teacher, a GUI displayed on a computing device, and a robot. The teacher sends commands to the robot and receives feedback from the robot through the GUI.

The teacher starts by selecting the desired action to perform and an object on which to perform the action. For instance, if the robot should pickup a red disk the action would be *Pick-up* and the object *Red Disk*. This is assuming the robot has a *Pick-up* action and an object detector for red disks.

Once a command is received by the robot, its job is to perform the action. During execution, the robot should relay useful sensor and state information back to the GUI for display. The details of how the robot actually accomplishes the action and performs object detection is left to the system designer.

This process of sending commands and observing the results is continued by the teacher until they feel the demonstration is complete. During the demonstration, the robot records all the commands and sensor data. Using this information the robot will learn a policy that matches the demonstration. The actual learning process is left to future work.

5.4. Experimental Design

In a 2-level (robot visibility: directly visible vs. visually occluded) between-participants experiment design, this controlled experiment investigated the influences of robot visibility upon human-robot interaction, specifically how human teachers would perform when demonstrating to a robot how to complete the Towers of Hanoi puzzle.

5.4.1. Hypothesis

Given our previous statement that it will be easier for a teacher to develop a mental model of the robot, given line-of-sight access, we hypothesize that such a model will be beneficial to the teacher. An intuitive understanding of the robot's state should help the teacher internally formalize a proper teaching and interaction strategy, and make them more comfortable using the robot.

5.4.2. Participants

Twenty volunteers participated in this study, with ages ranging from 22 to 59. Of the twenty instructors, six were female, and the remainder male. The six females were evenly distributed between the *visible* and *non-visible* teaching conditions.

5.4.3. Manipulation

Each participant was randomly assigned to one of two experimental conditions. Half the participants were allowed direct observation of the robot while conducting a demonstration. These participants are called the *visible instructors*. Other participants could not see the robot due to a screen behind which they were placed. These instructors could only rely on the GUI for feedback from the robot, and were termed *non-visible instructors*.

5.4.4. Materials

Our version of the Towers of Hanoi puzzle consists of three disks colored red, green, and blue with the red disk being the largest and blue the smallest. With only three disks this puzzle is the simplest form of Towers of Hanoi. The purpose of this study was not to test a teachers' performance at solving the puzzle, but rather their ability to use the GUI to instruct the robot.

The robot used in these studies is the PR2 mobile manipulator designed by Willow Garage.This robot consists of a wheeled base with two 7 DOF arms, and a pan-tilt head that carries two stereo camera pairs. The PR2 is capable of navigating around typical office environments, and detecting and interacting with simple objects.

Data from the stereo head on the PR2 was used to detect the the location of the disks and the poles. A color blob tracker working in conjunction with the stereo data identified the location of the discs. The poles were identified based on their height using the point cloud data from the stereo head.

Communication from the instructor to robot utilizes a handheld mobile device. With the proliferation of smart phones, mp3 players, and other various gadgets we have made the assumption that most people would be familiar, if not comfortable, using a handheld device. The Nokia N810 Internet tabletwas chosen based on its high-speed wireless, web-browsing capabilities, touch screen, and open-source operating system.

5.4.5. Procedure

Each participant in this study was given a short written instruction sheet that described the Towers of Hanoi puzzle, the robot, and the GUI. Before starting the demonstration the participant completed a set of general survey questions aimed at assessing the individuals computer experience and current level of frustration [40].

5.4.6. Measures

The set of measurements gathered included the duration of a demonstration, number of valid commands sent to the robot, and the total number of commands. Demonstration duration was measured from the time the first command was issued to the time the teacher selected the *Finished* button on the GUI. The number of valid commands is a count of all the commands that caused the robot to execute a valid action. The total number of commands is the tally of all commands sent to the robot by the teacher. It is possible to send many extraneous commands to the robot while its in the middle of executing a command. These extra commands cannot be handled by the robot, and are a useful measure to the degree of frustration felt by the teacher, and how well the teacher understands the state of the robot.

Also recorded were demographic information including age, gender, education, and employment. We also collected data concerning computer usage and experience, and willingness to solve computer issues. This information was collected prior to completing the Towers of Hanoi puzzle.

Following the completion of a demonstration, the participants were asked about their perceived mental demand, physical demand, temporal demand, performance, effort, and frustration. These statistics were drawn from the NASA Task Load Index as a measure of cognitive load [41].

5.4.7. Data Analysis

All quantitative measures were analyzed using analysis of variance (ANOVA) with the experiment manipulation of robot visibility as the primary independent variable.

One participant's data for number of commands was an outlier (more than two standard deviations above the mean value for the condition) so his data were replaced with his group's average value.

5.5. *Experiment Results*

5.5.1. *Number of commands*

The number of commands sent to the robot varies depending on the teacher's ability to solve the puzzle efficiently and their understanding of the state of the robot. Figure 7b is a box plot that shows statistics related to the number of commands sent to the robot by the teacher. On the left side of the box plot is the total number of commands sent split between separated and collocated teachers. On the right of this figure is the total number commands.

Teachers who could see the robot sent more valid commands to the robot (M=47.1, SD=12.3) than teachers who could not see the robot (M=36.8, SD=7.7), $F(1,18)=6.12$, $p< .03$.

Teachers who could see the robot sent more commands (in total) to the robot, (M=36.7, SD=7.7) than teachers who could not see the robot(M=29.9, SD=4.0), $F(1,18)=5.04$, $p< .04$.

The total number of commands indicates that teachers who could see the robot may get more frustrated with the robot and/or have an incorrect mental model of the robot than people who could not see the robot.

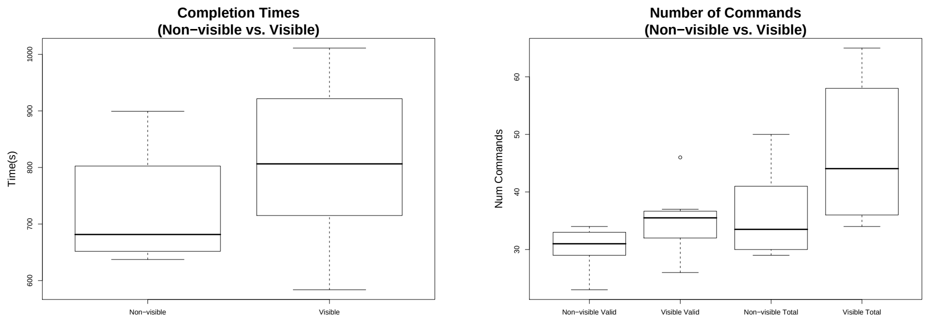
5.5.2. *Time on task*

Based on our original hypothesis, we expected those teachers who were behind the screen to require more time to complete the Towers of Hanoi demonstration. The results of this study do not support this hypothesis. As seen in Figures 7a and 7b both the time to complete the task and the number of commands issued by the teachers actually increased when the teacher had full view of the robot.

Task duration measured the amount of time it took a teacher to complete the Towers of Hanoi puzzle. Figure 7a shows the comparison of *visible* versus *non-visible* teachers. The mean duration is longer for *visible* teachers, however there is also more variance in that condition. Alternatively, those teachers who were visually blocked from the robot performed more consistently, and were able to complete the task in a slightly shorter timespan. The significance of location as an indicator of duration is not significant, $F(1,18)=2.58$, p=.13; however these data do suggest a trend.

6. Robot Kitting

The Towers of Hanoi study indicated that direct observation of a robot does not provide sufficient information to an instructor. The PR2's form has



(a) Duration times

(b) Command counts

Figure 2: Box plots of the duration times and command counts between collocated and separated teachers.

humanoid aspects, but since it does not behave like a human the instructor is unable to properly infer the robot's state. A more distinct and clear form of communication is required to help alleviate this ambiguity. We hypothesize that audio cues may serve as the needed form of communication.

In order to convey information through sound, a robot requires only a speaker and enough logic to map internal state to a sound. We have initially opted to use non-verbal sounds, in order to prevent gender identification, language difficulties, and most importantly to limit the amount of intelligence the robot projects.

Based on a previous study, see Section 3.1, we chose three sounds to cover *acknowledgment*, *error*, and *success*. These three categories were chosen due their general nature, and partly because they fit well within the context of learning from demonstrations. As the instructor gives commands to the robot, it's easy to convey *acknowledgment* of a command, an *error* if one occurs, and *success* when the command is complete.

The effect audio cues have on instructor performance was measured in a study based on a manufacturing process call *kitting*. When assembling complex objects, such as a robot, it is useful to hierarchically decompose the object into simple subcomponents stopping at raw materials (such as nuts and bolts). Each subcomponent is paired with a sheet that lists all the parts necessary for its construction. This sheet is called a *kit sheet*, and the parts required to build the subcomponent are gathered into a bin called a *kit*. The process of collecting parts into a kit is called *kitting*.

Since the goal of this study was not to assemble an actual kit, usually consisting of small parts that a robot would find difficult to grasp, we used colored boxes as substitutes for the parts. The study was also conducted in simulation, providing a consistent environment in which to teach that mimics the real world. The interactions between the human and robot were of primary concern. The task of kitting acts just as a means to engage a person with a meaningful and realistic task.

Instructors used the graphical interface that was developed for the Towers of Hanoi study, see section 5.2. The only difference between the two interfaces are the labeled objects, which now consist of a set of six colored boxes instead of three colored disks.

6.1. Experimental Design

In a 2x2 controlled study (audio cues: enabled vs. disabled, robot visibility: directly visible vs. no visualization), audio cues were varied between subjects and visibility varied within subjects. The goal is to investigate the effects audio cues and visibility have on interaction in a teaching context. Performance measures include time on task, number of commands issued by the instructor, and the number of errors that occur. It should be noted that simulated robot is able to perform each command perfectly. As a result, any error is due to an improper command sent by the instructor.

6.1.1. Hypothesis

Since communication plays a vital role in human-human interactions, it is safe to assume that it will also play a significant role in human-robot interactions. With the robot's ability to express state through audio cues, the instructor should have a more informed understanding of the robot. The result, we predict, is a reduction in the number of erroneous commands issued by the instructor, and a reduction in the time on task.

6.1.2. Participants

Twenty volunteers participated in this study, with ages ranging from 22 to 60. Of the twenty instructors, ten were male and ten were female. The men and women were evenly distributed among the test conditions, and the order in which the test conditions were run was randomized.

6.1.3. Manipulation

Each participant completed two demonstrations, each with a different kit sheet, one in which they could not see the robot and one in which they could. The order of the kit sheet and robot visibility were varied between subjects. Half of the participants received audio cues from the robot, while the other half did not. This resulted in four test cases based on the two conditions: visibility, and sound. Test cases were randomly assigned and evenly distributed among men and women.

6.1.4. Materials

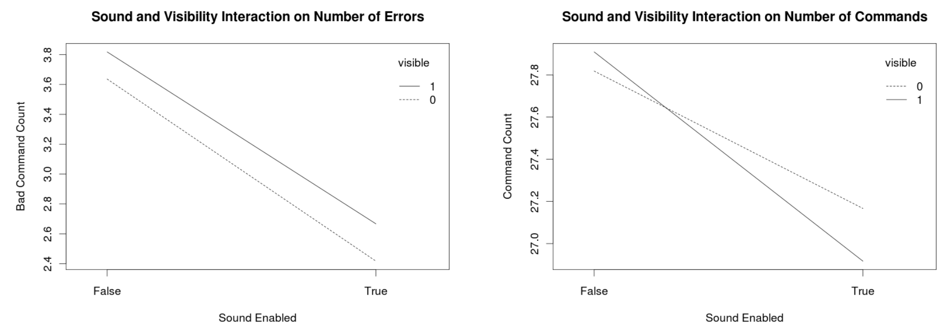
The kitting task was completely simulated in Gazebo [42], an open-source 3D robot simulator. Physical interactions, lighting conditions, and materials are all simulated within Gazebo. The result is a well defined and immersive environment in which a simulated robot can operate.

A simulated version of the PR2 robot was used in this study. Six colored blocks (red, green, blue, yellow, purple, turquoise), sat in a row on a table in front of the robot. To either side of the PR2 were two bins, labeled A and B. The blocks and bins were well within the reach of the PR2's arms, which made movement of the base unnecessary. Participants therefore were only allowed to controlled arm and gripper movements.

The graphical interface, used to communicate with the robot, ran in a web browser on the desktop computer along side Gazebo. Changes between this graphical interface and that used in the Towers of Hanoi study included only different objects with which the robot could interact. A mouse was the only physical device required to used the interface.

6.1.5. Procedure

A short instruction sheet was provided to each participant that described the process of kitting, their role as instructor, and how to use the graphical interface. Once they read through the instructions, the participant was given a kit sheet and asked to complete a demonstration. Following the demonstration, the participant completed a short questionnaire designed to assess their level of frustration. A second kit sheet with different bin assignments for the colored blocks was then completed by the participant followed by the same questionnaire and a general survey that gathered demographic information.



(a) (b)

Figure 3: Interaction plots of sounds and robot visibility on (a) the number of bad commands, and (b) the total number of commands

6.1.6. Measures and Data Analysis

The set of measures collected matched those from the Towers of Hanoi study, see Section 5.4.6. Two sets of measurements for cognitive load were gathered, one after each demonstration.

Similarly, analysis of the data followed the procedure used in the Tower of Hanoi study, see Section 5.4.7. Three participant's data for the number of command and number of bad commands exceeded two standard deviations above the mean. These values were replaced with the group's mean.

6.2. Experiment Results

6.2.1. Number of Bad Commands

The number of bad command was marginally predicted by the presence of auditory feedback from the robot. People who heard auditory feedback gave fewer bad commands to the robot (M=1.9, SD=1.5) than people who did not have any auditory feedback (M=3.3, SD=2.3), $F(1,40)=4.01$, $p=.056$. While this effect would be significant if we chose a significance cut-off value of $p=.10$, it is only approaching significance because we had originally chosen out significance cut-off level to be $p=.05$.

Figure **??** depicts the interaction sound and visibility has on the number of bad commands. When sounds are enabled, bad commands reduce in both cases. When the robot is not directly visible, the number of bad commands are also slightly reduced.

6.2.2. Total Number of Commands

The total number of commands issued to the robot was influenced by visual access to the robot. When people had visual access to the robot, they gave slightly more commands to the robot in total (M=27.1, SD=2.1) than when they did not have visual access to the robot (M=26.4, SD=2.2), $F(1,40)=3.99$, $p=.01$. This effect is significant at the $p=.05$ level.

Figure **??** shows the interaction sound and visibility has on the total number of commands. Audio cues decrease the total number of commands in both cases, with a slightly more dramatic effect when the robot is visible.

6.2.3. Time on Task

Neither auditory nor visual feedback from the robot was found to be a significant predictor of time on task so we ran a regression analysis of robot experience as a predictor of time on task. The most significant predictor of the time that a person spent on the task was the amount of experience that s/he had with robots. The more experience that the person had with robots, the less time s/he spent on the task, beta=-6.30, p¡.01.

6.2.4. Perceptions

After completing each demonstration, the instructor was asked to complete a survey designed to measure cognitive load. Data from this survey indicates that instructors who received auditory feedback from the robot user interface perceived that they exerted less effort on the task (M=2.1, SD=1.5) than people who did not received auditory feedback from the robot user interface (M=3.4, SD=2.1), $F(1,34)=4.60$, p¡.05.

Auditory feedback also impacted an instructor's perceived physical effort. Instructors who received auditory feedback experienced more a slightly greater physical demand during the task (M=1.7, SD=0.9) than people who did not received auditory feedback from the robot user interface (M=1.1, SD=0.3), $F(1,34)=6.38$, p¡.05. No significant differences were found for mental or temporal demand.

7. Discussion

The major finding from these studies is the effect visibility of a robot has on the performance of instructor. In both experimental setups, with a real and simulated robot, the number of commands issued to the robot increased when a teacher was allowed visual access to the robot. The time on task also tended to increase as well as the number of error commands.

An increase in the commands is interpreted to mean the instructors made more mistakes. Redundant commands and commands that do not move the robot closer to the desired goal are valid but undesirable. We can not expect instructors to give perfect demonstrations, however we can strive to make the teaching scenario more conducive to reaching that goal.

Observation of the instructor's behavior revealed a few potential causes for the differences in performance. Initial interaction with the robot caused a certain amount of fascination, which distracted the instructor from the demonstration. The robot's arm movements while manipulating objects is interesting to watch, much more so than the graphical interface. This is true even for people who work with the PR2 robot on a daily basis. As a result, visible teachers would decide when the robot was done performing an action based on their own visual cues rather than direct information from the robot via the GUI.

Teachers who were visually blocked from the robot did not have the ability to make direct visual judgments about the state of the robot. Instead, these teachers relied on information provided to them from the graphical interface and from audio cues. As a result, they issued fewer repeated commands. In effect, *visible teachers* received information from the robot that lacked context. As noted in prior work, just by increasing the amount of information provided to a person does not necessarily increase their situational awareness [43]. It would be more efficient to improve the display and throughput of information from the robot's sensors to the human in order to decrease confusion [44].

Other work has indicated that just by showing a person a humanoid robot, they automatically start building a mental model of the robot [45, 46]. By restricting some teachers from seeing the actual robot, they were limited in their ability to generate improper mental models of the robot's states.

Non-visible teachers also missed the "wow-factor" of the PR2 moving its arms. While the graphical interface showed a live video stream from the robot's perspective, the field of view was small and the frame-rate was low. This did not convey the same level of interest about the robot's arm movements.

Based on comments from the teachers, collected data, and observations from the Towers of Hanoi study, it was clear that a human teacher needs more useful feedback from the robot than just visual observations. This trend has also been noted in [47], where teachers did not wait for the robot to complete an action before providing more input.

Non-verbal audio cues were used to supplement the graphical interface in a second study based on robot kitting. The choice of audio cues were validated through an online survey of 100 people. During the course of the kitting study, we received no indication the sounds were confusing.

Results show that sounds helped to improve the performance of instructors. The extra channel of communication provided instructors with vital information about the internal state of the PR2. Comments from the instructors indicated that the audio cues properly informed them as to when the robot was ready to accept an instruction, experienced an error, and received a new instruction. Visibility of the robot still impacted instructor performance, however the audio cues did temper this effect.

8. Limitations

These experiments had several limitations, including (1) the participant sample of individuals, who were mostly engineers from the San Francisco Bay Area, (2) the use of a single robot, PR2. Future work could address these limitations by studying different sample populations and different robots, thereby testing the generalizability of the results from the current studyies.

9. Future Work

While non-verbal audio cues showed some improvement on instructor performance, they did not provide a complete solution. Other forms a communication can be leveraged to help inform instructors about the state of the robot. Gestures can be very expressive, and easily understood by people if executed properly. While more difficult to implement, they could provided a necessary form of communication between the robot student and instructor.

Both of the two studies discussed in this article have relied on a single robot, the PR2. Its size, shape, and form may affect the how an instructor perceives and interacts with the robot. Further work will utilize robots with both similar and different properties in order to see what properties induce the results discussed here.

10. Conclusions

We have shown that interaction with a robot is affected by visual and auditory signals. By allowing instructors direct visual access to the robot student, the instructors tended to construct inaccurate mental models of the robot's state. This likely stems from a preference to map a human-like model onto the capabilities of the robot. When this occurs expectations about the robot's performance and capabilities are altered to closely match what we would expect of another person. Audio cues helped to improve instructor performance by providing accurate information about the internal state of robot through a natural medium of communication.

Instructors who were not allowed direct visual access of the robot did not have the ability to build the same sort of internal representation of the robot. In this situation, their only source of information came from data displayed on a graphical interface and audio signals from the robot. The person did not see the physical structure of the robot or some of the motions that it executed.

While previous work has shown the benefits of constructing internal mental models of a robot's capabilities, this study demonstrates that there are some expectations that a robot may not meet. These expectations can alter the way in which we interact with the robot, and potentially a person's willingness to interact with the robot in the future.

11. Acknowledgments

This work was supported in part by the NSF CNS-0709296 grant, the NSF IIS-0713697 grant, and Willow Garage (`http://www.willowgarage.com`).

[1] Nao, Aldebaran robotics, http://www.aldebaran-robotics.com/en.

[2] K. Wyrobek, E. Berger, H. V. der Loos, K. Salisbury, Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot, in: Proc International Conference on Robotics and Automation, 2008.

[3] Sarcos, http://www.sarcos.com.

[4] P. Jackson, Introduction to Expert Systems, addison Wesley, 1998.

[5] C. Zixing, An expert system for robot transfer planning 3 (2) (1988) 153–160.

[6] M. L. Visinsky, J. R. Cavallaro, I. D. Walker, Expert system framework for fault detection and fault tolerance in robotics 20 (5) (1994) 421–435.

[7] A. Ramesh, M. J. Matarić, Learning movement sequences from demonstration, in: Proceedings of the International Conference on Development and Learning, MIT, Cambridge, MA, 2002, pp. 203–208.

[8] G. Rizzolatti, C. Sinigaglia, Mirrors in the Brain: How our minds share actions, emotions, and experience, Oxford University Press, 2008.

[9] Y. Demiris, G. Hayes, Imitation as a dual-route process featuring prediction and learning components: A biologically plausible computational model, MIT Press, 2002, pp. 327–362.

[10] G. Maistros, G. Hayes, An imitation mechansim for goal-directed actions, in: Proceedings of TIMR 2001 - Towards Intelligent Mobile Robots, Manchester, 2001.

[11] Cynthia Breazeal and Daphna Buchsbaum and Jesse Gray and Bruce Blumberg, Learning From and About Others: Towards Using Imitation to Bootstrap the Social Competence of Robots, Artificial Life 11 (1-2) (2005) 31–62.

[12] S. Preston, F. B. M. de Waal, Empathy: Its ultimate and proximate bases 25 (2002) 1–72.

[13] A. M. Leslie, ToMM, ToBY, and Agency: Core architecture and domain specificity, Cambridge University Press, Cambridge, 1994, p. 119148.

[14] A. Karmiloff-Smith, E. Klima, U. Bellugi, J. Grant, S. Baron-Cohen, Is there a social module? language, face processing, and theory of mind in individuals with williams syndrome 7 (2) (1995) 196208.

[15] B. Scassellati, Theory of mind for a humanoid robot 12 (2002) 13–24.

[16] S. Coradeschi, H. Ishiguro, M. Asada, S. Shapiro, M. Theilscher, C. Breazeal, M. Mataric, H. Ishida, Human-inspired robots 21 (4) (2006) 74–85.

[17] M. Loper, N. Koenig, S. Chernova, C. Jones, O. Jenkins, Mobile human-robot teaming with environmental tolerance, in: HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, La Jolla, California, USA, 2009, pp. 157–164.

[18] M. Nicolescu, M. J. Matarić, Task learning through imitation and human-robot interaction, in: K. Dautenhahn, C. Nehaniv (Eds.), Models and Mechanisms of Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions, Cambridge University Press, 2005.

[19] P. Pastor, H. Hoffmann, T. Asfour, S. Schaal, learning and generalization of motor skills by learning from demonstration, in: international conference on robotics and automation (icra2009), 2009.

[20] B. Petreska, A. Billard, Movement Curvature Planning through Force Field Internal Models, Biological Cybernetics 100 (5) (2009) 331–350.

[21] A. L. Thomaz, C. Breazeal, Teachable robots: Understanding human teaching behavior to build more effective robot learners, Artificial Intelligence Journal 172 (2008) 716–737.

[22] M. Nicolescu, M. J. Matarić, Natural methods for robot task learning: Instructive demonstration, generalization and practice, in: International Joint Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia, 2003.

[23] D. H. Grollman, O. C. Jenkins, Learning robot soccer skills from demonstration, in: International Conference on Development and Learning, London, UK, 2007, pp. 276–281.

[24] S. Chernova, M. Veloso, Confidence-based policy learning from demonstration using gaussian mixture models, in: International Conference on Autonomous Agents and Multiagent Systems (AAMAS'07), 2007.

[25] S. Schaal, A. Ijspeert, A. Billard, computational approaches to motor learning by imitation, no. 1431, Oxford University Press, 2004, pp. 199–218.

[26] M. Hersch, F. Guenter, S. Calinon, A. Billard, Dynamical System Modulation for Robot Learning via Kinesthetic Demonstrations, IEEE Transactions on Robotics 24 (6) (2008) 1463–1467.

[27] S. Chernova, M. Veloso, Teaching collaborative multi-robot tasks through demonstration, in: IEEE-RAS International Conference on Humanoid Robots, Daejeon, Korea, 2008.

[28] A. L. Thomaz, C. Breazeal, Reinforcement learning with human teachers: Understanding how people want to teach robots, in: roceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2006.

[29] F. Guenter, M. Hersch, S. Calinon, A. Billard, Reinforcement Learning for Imitating Constrained Reaching Movements, RSJ Advanced Robotics, Special Issue on Imitative Robots 21 (13) (2007) 1521–1544.

[30] F. Stulp, E. Oztop, P. Pastor, M. Beetz, S. Schaal, compact models of motor primitive variations for predictable reaching and obstacle avoidance, in: ieee-ras/rsj international conference on humanoid robots (humanoids 2009), 2009.

[31] D. H. Grollman, O. C. Jenkins, Sparse incremental learning for interactive robot control policy estimation, in: International Conference on Robotics and Automation, Pasadena, CA, USA, 2008, pp. 3315–3320.

[32] R. A. Howard, J. E. Matheson, Influence diagrams, in: R. A. Howard, J. E. Matheson (Eds.), The Principles and Applications of Decision Analysis, Vol. 2, Menlo Park, CA, 1981, Ch. 37, pp. 721–762.

[33] L. T. S. Paepke, Juding a bot by its cover: An experiment on expectation setting for personal robots, in: Proceedings on the 5th ACM/IEEE Internation Conference on Human-Robot Interaction, Osaka, Japan, 2010.

[34] H. H. Clark, Understanding Language, Cambridge University Press, 1996.

[35] H. H. Clark, S. A. Brennan, Grounding in communication, in: L. B. Resnick, J. M. Levine, S. D. Teasley (Eds.), Perspectives on socially shared cognition, 1991.

[36] E. Rocco, Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact, in: Proceedings of CHI '98, Los Angeles, CA, 1998, pp. 496–502.

[37] C. Jensen, S. D. Farnham, S. M. Drucker, P. Kollock, The effect of communication modality on cooperation in online environments, in: CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA, 2000, pp. 470–477. doi:http://doi.acm.org/10.1145/332040.332478.

[38] S.-L. L. Sara, S. Kiesler, I. Yee-man, L. C. yue Chiu, Human mental models of humanoid robots, in: 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 2005.

[39] S. Kiesler, J. Goetz, Mental models of robotic assistants, in: CHI '02: CHI '02 extended abstracts on Human factors in computing systems, ACM, New York, NY, USA, 2002, pp. 576–577. doi:http://doi.acm.org/10.1145/506443.506491.

[40] J. Lazar, A. Jones, K. Bessiere, I. Ceaparu, B. Shneiderman, Workplace user frustration with computers: An exploratory investigation of the causes and severity, Behaviour & Information Technology 25 (3) (2004) 239–251.

[41] H. S. G., S. L. E., Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: H. P. A., M. N. (Eds.), Human mental workload, Elsevier, Amsterdam, 1988, pp. 139–183.

[42] N. Koenig, A. Howard, Design and use paradigms for gazebo, an open-source multi-robot simulator, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, 2004, pp. 2149–2154.

[43] M. R. Endsley, Design and evaluation for situation awareness enhancement., in: Proceedings of the Human Factors Society 32nd Annual Meeting, Human Factors Society, Santa Monica, CA, 1988.

[44] K. Gold, An information pipeline model of human-robot interaction, in: HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, ACM, New York, NY, USA, 2009, pp. 85–92. doi:http://doi.acm.org/10.1145/1514095.1514113.

[45] A. Powers, S. Kiesler, The advisor robot: tracing people's mental model from a robot's physical attributes, in: HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, ACM, New York, NY, USA, 2006, pp. 218–225. doi:http://doi.acm.org/10.1145/1121241.1121280.

[46] R. Gockley, J. Forlizzi, R. Simmons, Interactions with a moody robot, in: HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, ACM, New York, NY, USA, 2006, pp. 186–193. doi:http://doi.acm.org/10.1145/1121241.1121274.

[47] E. S. Kim, D. Leyzberg, K. M. Tsui, B. Scassellati, How people talk when teaching a robot, in: HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, ACM, New York, NY, USA, 2009, pp. 23–30. doi:http://doi.acm.org/10.1145/1514095.1514102.