How People Talk When Teaching a Robot

Elizabeth S. Kim
Dept. of Computer Science
Yale University
51 Prospect St.
New Haven, CT
eliskim@cs.yale.edu

Katherine M. Tsui
Dept. of Computer Science
Univ. of Massachusetts,
Lowell
1 University Ave.
Lowell, MA
ktsui@cs.uml.edu

Dan Leyzberg
Dept. of Computer Science
Yale University
51 Prospect St.
New Haven, CT
dan.leyzberg@yale.edu

Brian Scassellati
Dept. of Computer Science
Yale University
51 Prospect St.
New Haven, CT
scaz@cs.yale.edu

ABSTRACT

We examine affective vocalizations provided by human teachers to robotic learners. In unscripted one-on-one interactions, participants provided vocal input to a robotic dinosaur as the robot selected toy buildings to knock down. We find that (1) people vary their vocal input depending on the learner's performance history, (2) people do not wait until a robotic learner completes an action before they provide input and (3) people naïvely and spontaneously use intensely affective vocalizations. Our findings suggest modifications may be needed to traditional machine learning models to better fit observed human tendencies. Our observations of human behavior contradict the popular assumptions made by machine learning algorithms (in particular, reinforcement learning) that the reward function is stationary and path-independent for social learning interactions.

We also propose an interaction taxonomy that describes three phases of a human-teacher's vocalizations: *direction*, spoken before an action is taken; *guidance*, spoken as the learner communicates an intended action; and *feedback*, spoken in response to a completed action.

Categories and Subject Descriptors

I.2.9 [**Computing Methodologies**]: Artificial Intelligence—*Robotics*

General Terms

Experimentation, Human Factors

Keywords

1. INTRODUCTION

As robots enter human environments as our teammates, assistants, guides, and therapeutic partners, their value will depend on their ability to adapt to users and their environments. For adults, verbal communication is both readily available and powerful, lending itself naturally as a primary input modality for robot learners. In this study, we examine how untrained people naturally teach by observing their vocalizations as they teach a robot a simple task.

While there are many technologies for automatic speech recognition (the process of transcribing spoken words into written words), interpreting the meaning of these words is beyond the current state of the art for all but limited vocabularies or under very restricted environments. Fortunately, for many utterances directed by an instructor to a learner, the meaning of the words is often mirrored by the affect being conveyed. In other words, even without knowing *what* is said, *how* we say it carries much of the speaker's intent. Domesticated animals such as dogs learn from their caretakers' approving and prohibitive vocalizations, and it has been suggested that exaggerated affective expressions in infant-directed speech (Parentese) are an important learning input for pre-verbal infants. Affective expressions have been shown to be useful as an input to artificial learners [5, 17].

Automatic affect recognition from voice and other modalities has succeeded in approaching or meeting levels of human agreement in judgments of affect, within closed application domains [1,4,17–19,23]. Currently these systems are limited by constraining users to atypical interaction patterns, or by training to recognize affect only within constrained application scenarios. While these studies offer impressive technology, they are technology-driven accounts of affective interaction; none directly provides evidence for how untrained users naturally provide input to a computational or robotic agent. Nevertheless, as these automatic classifiers approach human recognition of affect, we expect that robotic learners will be able to use automatic affect recognition to sustain

natural social interaction with people, eliciting natural affective communications from which to learn.

This paper presents an experiment that attempts to characterize how untrained people freely provide input to a robot when asked to teach it to perform a simple task. While the teaching strategies we observe are based on affective vocalizations, we expect that they may reflect teaching behaviors that people apply more generally, not only through affective vocalizations.

2. BACKGROUND

2.1 Learning from Human Interaction

Natural human social behaviors have been studied by many communities as inputs to robotic learners. Non-affective input has been used in the form of gestures for imitative learning (e.g., [21]), speech for word learning (e.g., [11, 25]), and pet-directed touch for a robotic dog learning to behave politely [25]. Natural affective expressions have also been used as input by robotic learners [5, 17].

To help focus our discussions, we will compare the findings of this paper with one popular computational technique, reinforcement learning, which has often been used to model learning from social input (for surveys of reinforcement learning, see [14, 24, 27]). While there are many other techniques available, reinforcement learning is both widely used in social learning experiments and will serve to highlight how our experimental results on feedback from a human instructor differ from the assumptions often used in machine learning.

In reinforcement learning, the learner's environment itself, or a knowledgeable instructor, provides positive and negative feedback to the learner without necessarily providing the appropriate answer. For example, if attempting to instruct an agent to play checkers, the instructor might say "good" or "no" or "bad move" but would not provide the best move to the learner. Researchers have used various forms of social cues as feedback in reinforcement learning systems, including affective facial expressions [5] and affective vocalizations [17], as well as non-affective "clicker" stimuli to robotic dogs [3, 26] and other non-affective inputs [13, 15, 28].

Reinforcement learning can be thought of as both a problem framework and a group of techniques. In the reinforcement learning framework, the learner assigns value to problem states based on a reward signal it receives as feedback, which it receives in response to actions which transition the learner from one state to another. In contrast to the supervised learning problem framework, in which learner performance is good whenever the learner correctly labels an example (whether the label is "good" or "bad"), in the reinforcement learning framework, performance is bad if the learner acts to transition to a state having low ("bad") value. Reinforcement learning problems have been formalized as Markov Decision Processes, and classical reinforcement learning techniques have tended to be based on dynamic programming algorithms that require reward signals to be both *stationary* (the reward for a particular state and action pairing does not change over time) and *path-independent* (the reward applied depends only on the current state, not the sequence of states visited en route to that state). In this paper, our observations of human teaching behavior suggest that the reinforcement learning framework neglects useful input from people. Our observations

suggests a mismatch between human-produced rewards and the assumptions of stationarity and path-independence that classical reinforcement learning techniques make on reward signals.

2.2 Affect Recognition

Automatic vocal affect recognizers have approached or met the performance criteria of human listener agreement using a variety of methods. Breazeal and Aryananda [4], Kim and Scassellati [17], and Robinson-Mosher and Scassellati [23] have used statistical methods over acoustic-prosodic features (pitch, duration, and intensity). In addition to acoustic-prosodic features, Litman and Forbes-Riley [19], Lee and Narayanan [18], and Ang et al. [1] also used additional word features (e.g., "no," which is commonly correlated with negative affect) and other application-specific features, such as discourse cues (e.g., the words "start over" uttered during interaction with an automated phone service [18]). The affect types classified and the resulting performance have varied over different application domains, and several studies caution that performance suffers unless classifiers are domain-specific [1, 18, 19]. For instance, Litman and Forbes-Riley note that in student-tutor dialogs "turn duration was always among the most useful acoustic-prosodic predictors of emotion, and that longer durations were generally predictive of the negative emotion class" [19]. The duration of one speaker's turn in a conversation is a domain-specific feature. We can imagine, in other contexts, that longer turn durations may not correlate to negative affect. For instance, in this paper's experimental context, in which a human intermittently provides directions to a robotic task learner, turn duration may correlate with affectively neutral instruction. Furthermore, Litman and Forbes-Riley note that "the usefulness of particular acoustic-prosodic features varied across experiments; indeed across prior research more generally, the usefulness of particular acoustic-prosodic features is often domain-dependent" [19]. A recent survey of affect recognition algorithms concurs that no single "optimal feature set" has yet been agreed upon by existing voice-based affect models [31].

The addition of facial expressions and other physiological measurements to vocal features tends to improve performance in automatic affect classifiers [6, 16, 31]. However, in our post-experiment data analysis (further detailed in Section 4.3), we focus on vocal expressions only, providing our human affect annotators with only audio data, because we expect that for future applications, voice will be the most generally accessible of the multiple, redundant modalities used to express affect. Voice is in itself rich in affective expression, and audio recording is easily accessible (in contrast, facial affect recognizers tend to require specialized visual tracking devices). We expect that our findings based on voice will apply to learning systems which measure affect through other modalities as well.

We used the human judgment of a hidden Wizard of Oz operator instead of automatic affect recognition. We provided the Wizard with both live audio and video streams for judgment of voice and facial expressions. Our choice is motivated by our expectation that we could not achieve sufficient automatic affect recognition performance to sustain natural human responses during interaction with the robot without first amassing a large collection of training data from our specific task domain. At the same time, we expect that ap-

plications which seek to leverage our findings about natural variations in human affect may currently or soon be able to achieve such high-quality automatic recognition within closed application domains. We use human judgment as a baseline for automated affect recognition, which we expect to be feasible in the near future.

3. HYPOTHESES

Hypothesis 1: Naïve instructors will provide affective *guidance* and *feedback*. Explorations into interactive machine learning have observed the need to modify classical machine learning models to fit natural human teaching preferences [29]. We hypothesized that naïve speakers, given the opportunity to comment on a robotic learner's intended action, would vocalize about the intended action before the action is completed. We expected less intensity in affect expressions voiced before an action is completed then those voiced after.

Hypothesis 2: Naïve instructors will use affective vocalizations without explicit instruction. In previous studies of naïve speakers talking to robots, affective prosody has been elicited only when participants were explicitly "instructed to express each communicative intent (approval, attention, prohibition, and soothing) and signal when they felt that they had communicated it to the robot" [4]; they were instructed to act as though talking to a child or pet. Based on our own anecdotal observations of naïve people interacting with Pleo robots, we hypothesized that naïve people would use affective prosody when talking to a robot *without explicit instruction*.

Hypothesis 3: Vocalizations will vary with respect to the history of a robotic learner's performance. Our predictions fell into two categories. (a) We expected differences in comparing a consistently successful learner with one that initially struggled. Specifically, we hypothesized that naïve speakers would produce more intensely positive prosody in response to a robotic learner's correct choice if it followed a series of wrong choices, then if it followed a series of correct choices. (b) We expected that people would speak less, and with weaker affective prosody, as a robotic learner consistently succeeded. We hypothesized that for a robot that made a series of correct choices, both the amount of vocalization and the strength of affect in prosody would fall as its successful streak continued.

4. METHODOLOGY

To investigate these hypotheses, we designed an experiment in which participants were asked to help two robot dinosaurs pick the right buildings to demolish as they walked through a model city. Unbeknownst to the participant, the dinosaur robots were being controlled by a remote operator; this experiment model is called "Wizard of Oz" (WOz) [9].

4.1 Participants

We recruited 27 participants, 9 male and 16 female, 18 years of age and above, from the Yale University and New Haven communities. Our exclusion criteria were based on English proficiency and previous research or coursework experience in artificial intelligence.

4.2 Experiment Design and Conduct

4.2.1 Interaction Protocol

A testing session lasted approximately 30 minutes. Participants gave informed consent to be recorded. The participant was brought into the room containing the two dinosaurs and the demolition training course. The participant stood at the edge of a table and clipped a lapel microphone to his/her shirt collar. Fred and Kevin, our robots, stood in front of the demolition training course, close to and facing the participant.

The participant was told the following:

These are our dinosaurs, their names are Kevin and Fred. Kevin is the one with the red hat with the "K" on it. Fred is the one wearing a bandanna. Today they're going to train to join a demolition crew. They'll be knocking over buildings with their heads. Behind them is the training course that they'll be running today. They'll go one at a time: Fred will be first and I'll take Kevin and leave the room. When Fred's done, then it'll be Kevin's turn. (The ordering of the dinosaurs varied per participant.)

You are going to help them pick the red "X"-marked buildings in the training course to demolish. In the training course, you'll see there are three pairs of colored buildings standing across from one another – the purple pair at the far end, the silver pair in the middle, and the orange pair closest to us. The robots will do the training course sequentially, starting at the purple buildings and walking towards us. For each pair, you'll see that one is marked with an "X." Kevin and Fred can see the "X"s too. For each pair of buildings it's important that they knock down the building with the "X" and that they don't knock down the unmarked building.

They already know how to knock down buildings. We want you to help them understand that they should only knock down the buildings with the red "X"s and all of the ones with the "X"s. You're going to help them by talking with them. We encourage you not to make any assumptions about how this might work. Just act naturally and do what feels comfortable. Please stay in this area [demarcated by caution tape]. *The training is complete when an orange building falls.*

The experimenter then engaged the participant by asking him/her to say hello to the dinosaurs and explain to them the task, in his/her own words. The dinosaurs returned the greetings with growls and acknowledged the receipt of instructions by looking and vocalizing at the participant in time with his/her words. The experimenter then solicited questions or provided additional clarification for the task from the participant.

Once the participant was comfortable with the task, the experimenter placed one of the dinosaurs at the start position, between the first pair of buildings, facing the participant. The experimenter left the room with the other dinosaur. Then the participant guided the first dinosaur in training. The first dinosaur gave a "Charge!" vocalization indicating the start of the trial. The dinosaur slowly (over 4 seconds) communicated his intent to topple a purple (first pair) building, by slowly turning his head towards it while vocalizing a slowly increasing growl. If the participant did not vocalize negatively towards the dinosaur, he concluded his intention communication (head-turn and increasing growl) by pushing the building over. Otherwise,

Figure 1: Participant talks to Fred as he runs the demolition training course. (Best viewed in color.)

Figure 2: The overhead view used for Wizard of Oz control of the robot's locomotion. North of this frame, a participant is standing at the end of the table. Building pairs, from bottom (beginning) to top (end) are: purple, silver, and orange. (Best viewed in color.)

the dinosaur discontinued the intention towards the current building, turned his head towards the other purple building and again vocalizing his intention to knock-down the other purple building. After the "X"-marked purple building fell, the dinosaur walked forward to the next pair of buildings. The dinosaur repeated his vocalization on knock down intention again for the silver pair of buildings and then the orange pair. The experimenter returned to the training room when either the participant indicated the end of the training or a period of time elapsed (approximately 30 seconds) after the orange building fell.

When the training was complete, the participant was given a few minutes' break while the experimenter reset the demolition training course. The participant then engaged in a training session with the other dinosaur while the experimenter and the first dinosaur waited outside. The second training session proceeded the same as the first, other than the initial indicated intents of the robot. Fred always chose the correct building for every pair, whereas Kevin chose the wrong building the first and second time, but correctly in his third trial.

Once the second training session was complete, the participant completed a survey. Then the experimenter debriefed the participant by showing him/her the WOz control room, explaining the technology, explaining the purposes of the study, and answering any questions.

4.2.2 Setup

Pleo is an 8-inch tall, 21-inch long dinosaur robot, sold commercially as a toy by UGOBE Life Forms [30]. In this experiment, we named our Pleo robots Kevin and Fred, each with a distinct voice we recorded to distinguish them as social actors [20]. Fred completed the training course first in 33% of the testing sessions.

Our robots' task was set on a table about 3 feet off the ground, on which sat a "demolition training course" of 8- to 12-inch tall toy, cardboard buildings (shown in Figure 2). 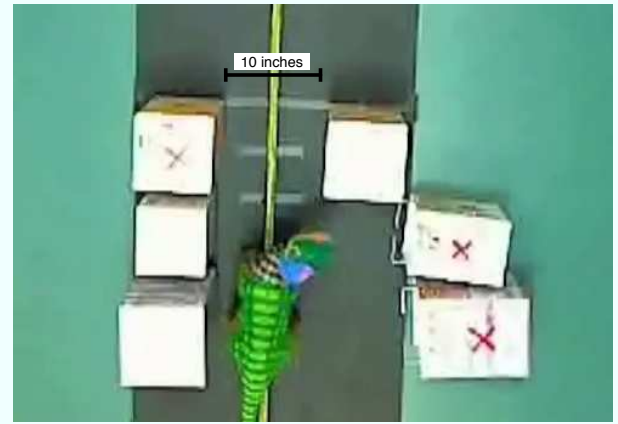The training course consisted of 3 pairs of two identically sized and painted buildings. In each pair, one of the two buildings was marked with red "X"s; the other building was unmarked. The pairs of buildings were placed on either side of a straight, yellow double-lined road: one building stood to the dinosaur's left, the other to the dinosaur's right. The three pairs of buildings were separated from each other along the road by spaces of 3 inches. From the dinosaur's perspective, the "X"-marked buildings were the right purple building, the right grey building, and the left orange building. The markings were constant for all participants. The yellow double-lines were raised, providing a track for the dinosaurs to walk along. The road was 10 inches wide by 30 inches long.

The participants stood facing the robot, looking down the course's road, about three feet from the the closer end of the course throughout their entire interactions with Kevin and Fred. The dinosaur approached the participant as it progressed through the training course.

4.2.3 The Wizard

The WOz design was necessary to ensure a real time interaction between the participant and the robot dinosaurs. Autonomous robot control may not have provided a fluid interaction because we did not construct narrowly defined expectations of participants' interactions with the robot. Thus, WOz control allowed us to simulate the application domain. Each robot was alone with the participant, so participant talked directly to the robot, and not to the person operating the robot. The deception of WOz was approved by Yale Institutional Review Board.

For WOz supervision, we used an overhead webcam for accurate estimation of the robot's range to strike buildings (shown in Figure 2) and a video camera aimed at the participant for viewing facial expression (shown in Figure 1). The wizard was also able to hear the participant through the clip-on lapel microphone.

The robot dinosaurs were controlled using infrared (IR) signals. The IR receiver was located in the dinosaur's nose.

IR signals were sent from long distance IR beacons through an IguanaIR USB-IR transceiver [12], controlled in Linux using LIRC (Linux Infrared Remote Control) software [2].

The wizard controlled the robot dinosaur's motions and vocalizations using a combination of scripted behaviors which were mapped to inputs on a USB handheld gaming pad. For example, pressing the joystick forward caused the robot dinosaur to walk forward, and pressing to the left or right caused the robot to move his head in the respective direction. These scripts were created and modified using UGOBE's software development kit and MySkit [10].

To appear autonomous and life-like, the robot dinosaurs were programmed with idling behaviors. Affective vocal and motor responses provided a heightened sense of communication. For example, the dinosaur would put his head down and make a sad "oh" sound when reprimanded. To ensure responsiveness, the walk and idle behaviors were short. Also, the intention script (dinosaur moving head towards a building and roaring) was interruptible, in the event that the participant reprimanded or corrected the robot.

4.3 Analysis of Vocal Input

4.3.1 Three Types of Vocalizations

Each participant's interaction was both video and audio recorded. The resultant audio recordings were segmented and analyzed. We noted participants' vocalizations fell into three cycling phases based on the robot's progress in each trial. All three phases occured for each trial: *direction*, occurring before the dinosaur picked a building; *guidance*, occurring while the dinosaur swung its head to knock over a building; and *feedback*, occurring after the building fell or the dinosaur abandoned his effort. We segmented our audio data along this dimension.

For each robot, the first of the three trials began with the robot placed between the first pair of buildings, where he would indicate his readiness by vocalizing. He would then signal his intent, lasting a few seconds. Then, if he was not reprimanded or corrected, he knocked over the building he intended to. In this case, the first sentence describes the *direction* phase, the second describes *guidance* and the last is *feedback*.

In this manner, phases cycled from *direction* to *guidance* to *feedback*, then back to *direction*. Sometimes there was one cycle per trial: the robot got to the building pair, motioned towards the correct building, and knocked it down. Other times, there were two: the robot got to the building pair, motioned towards the wrong building (and received reprimand), then replied to the reprimand (this phase falls we categorized as *direction*), then motioned towards the correct building (*guidance*), and knocked it down (*feedback*).

We performed the audio segmentation according to these guidelines and exported them our coders. The segmentation was performed by recognizing the dinosaur sounds we heard on the recording that uniquely identified the phases of each trial. The only phases for which that rule did not apply were between trials: separating the last phase of one trial (*feedback*) and the first of the next trial (*direction*). We waited for a two-second pause in our participants' vocalizations, and if there was none, we divided based on the transcription of the words used such that once they stopped using disparaging words (e.g. "no," "stop"), that moment divided the trials.

4.3.2 Annotating Affect

Our coders were not informed of the content of the audio nor the experimental design. We randomized our audio files split by phase (average length approximately 20 seconds) and asked two coders separately to rate each audio clip's prosodic affect as either positive, negative, or neither. Positively affective prosody was described to the coders as sounding "encouraging," "approving," or "pleasant," whereas negative affect was described sounding "discouraging," "prohibiting," or "disappointing." We also asked the coders to rate the intensity of the affect on a differential semantic scale [22] from 0 (mild) to 2 (very strong), and their respective confidences for each judgement on a differential semantic scale from 0 (not sure) to 2 (quite sure). Word count was also extracted from the audio clips.

5. RESULTS AND DISCUSSION

Twenty-seven people participated in this experiment; however, we excluded data from three participants from post-experiment data analysis. The reasons for exclusion were technical failure of the robot and/or recording devices and gross misunderstanding of protocol by participants. None of the participants guessed, in the survey or debriefing, that the robots had secretly been controlled by a human.

For each audio clip which was segmented by phase, we analyzed word counts, prosody ratings, prosodic intensity ratings, and post-experiment survey responses. The ratings of two naïve coders of affective prosody (prosody ratings and prosody intensity ratings) showed high agreement ($\kappa = 0.84$ using Cohen's quadratically weighted, normalized test [8]). Most phase clips were short and contained few words, with 7.71 words/clip on average (1.26 words/sec) and a standard deviation of 8.43 words/clip (1.36 words/sec).

5.1 Instructors vocalize before, during, and after a learner's actions

We hypothesized that participants would speak more during *feedback* than *guidance* and more during *guidance* than *direction* (H1). Interestingly, we found that subjects provided an almost equal number of words throughout. (See box-and-whisker plots in Figure 3.) Over all phases, the frequency of words spoken was on average 1.26 words/sec, with a standard deviation of 1.36 words/sec.

Classical reinforcement learning algorithms tend only to include reward inputs which arrive as *feedback* after an action is taken. We conclude that valuable *direction* and *guidance* reward inputs are being neglected by machine learning systems.

5.2 Instructors express affect during and after a learner's actions

We hypothesized that naïve people would use affective prosody when speaking to a robot (H2). Although we did not specifically instruct participants to use prosody, they vocalized with intensely affective prosody during *guidance* (affective intensity mean = 1.28, std. dev. = 0.93) and *feedback* (affective intensity mean = 1.89, std. dev. = 0.78). (See plot in Figure 4.) Subjects showed no positive or negative affect for *direction* (affective intensity mean = 0.47, std. dev. = 0.68). The phase-to-phase differences in affective rating and affective intensity were both significant ($p < 0.001$ for both ANOVA tests, $F[2] = 58.2, 19.2$).
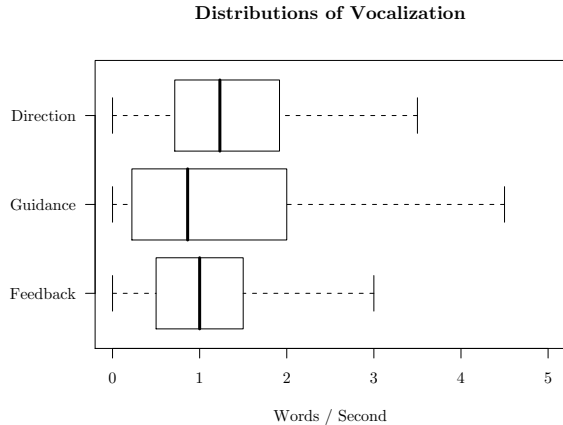
Distributions of Vocalization

Figure 3: The distributions show similar vocalizations rates across all phases. (The boxes in the plot contain the middle 50% of the observations, and the whiskers extend to the outer quartiles.)

These distinct amounts of affect intensity are consistent with the intuition that positive and negative affect are used to provide reinforcement as feedback on an ongoing or finished behavior, whereas reinforcement is not given before a behavior begins.

5.3 Instructors say less as a learner continually succeeds

We hypothesized that for naïve speakers, both the amount of vocalization and strength of affect would fall for the robot that always chose the correct building (H3b). Indeed, people did talk less after subsequent successful trials. During all phases of interaction with Fred, the robot who thrice initially intended to topple the correct building, words/sec fell from one trial to the next ($p = 0.002$, linear regression). Examining individual phases, we find significantly fewer words/sec for *guidance* ($p = 0.018$) and for *feedback* ($p = 0.038$), but not for *direction* ($p > 0.1$). The downward trend for both *guidance* and *feedback* is shown in Figure 5.

We verified that these trends could not be explained by the dinosaur ordering. In a two-way ANOVA, we found a highly significant main effect for trial number ($p = 0.0018, F[1] = 10$) and for order ($p = 0.0004, F[1] = 13$), but not for their interaction ($p = 0.38, F[1] = 0.7$). We conclude from this that while word/sec may have dropped off after talking with the first and then second robot due to familiarity or boredom (dinosaur ordering effect), this had no bearing on the more interesting result: regardless of whether they encountered (always correct) Fred first or second, people spoke less to him with subsequent trials. A similar test for (wrong-wrong-correct) Kevin, showed no trend of decreasing word/sec over trials ($p = 0.57, F[1] = 0.38$).

5.4 Instructors say more to a previously struggling learner

We compared *direction*, *guidance*, and *feedback* phases during the third trial for Kevin against those for Fred. Recall that in the first two trials, Kevin initially communicated intent to topple the wrong buildings, while Fred only com-

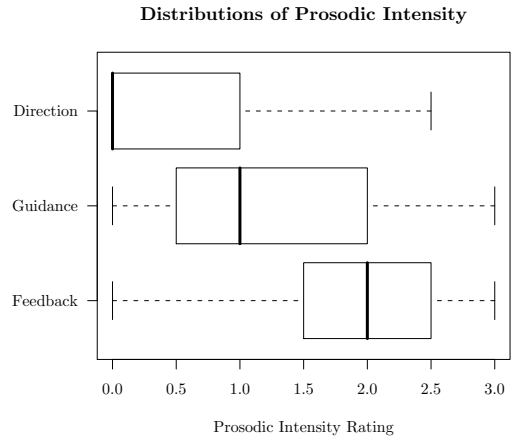**Distributions of Prosodic Intensity**

Figure 4: The distributions of the intensity of the prosody during each phase demonstrate that people use prosodic reinforcement as feedback on an ongoing or previously finished behavior.

municated intent to topple correct buildings in the first two trials. In the third trial, both Kevin and Fred initially communicated intent to topple the correct building.

We hypothesized that prosody would be more intensely positive in response to Kevin's than to Fred's third trial intent (H3a), since this would showcase the participants' relative excitement at Kevin improvement. Considering only *guidance* and *feedback* phase audio clips, we found that participants voiced marginally significantly more words/sec to Kevin than to Fred ($p = 0.089, F[1] = 3$). We found neither a main effect of dinosaur order nor an interaction between learning condition with dinosaur order. Figure 6 shows the trend for participants to give more *guidance* and *feedback* to Kevin than to Fred. We found no such difference for affect or affective intensity ratings.

We believe that even for the same state in the task, human vocal reward signals differ depending on the learning history of the robot. This disagrees with the assumption of path-independence made by standard reinforcement learning algorithms.

6. CONCLUSIONS

We designed and conducted an experiment in which naïve teachers helped a dinosaur robot Pleo learn to topple marked buildings in a demolition training course. Our goal was to investigate how people intuitively talk *without explicit instruction* when teaching robots. We found that naïve vocalizations during human-teacher/robot-learner interaction appear to segment into three distinct phases, providing distinct input content to the learner. These three phases are *direction* (before the learner acts), *guidance* (as the learner indicates intent) and *feedback* (after the learner completes a task-action). We observed that naïve human teachers vocalize readily throughout all three phases. Our experiment showed that people are affectively expressive as they direct the robotic learner well before it approaches the learning task, as the learner communicates its intention to act (effectively querying the teacher), and in giving feedback for actions the learner has taken. Thus, we have affirmed an in-
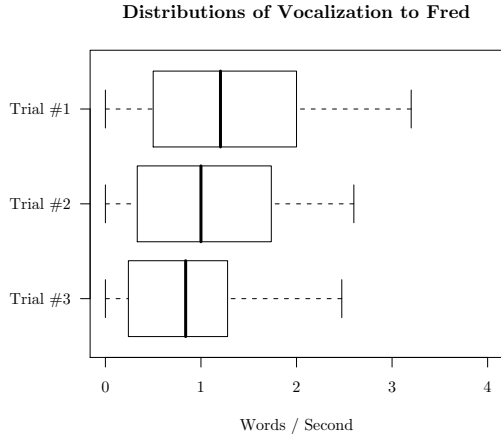
Distributions of Vocalization to Fred

Guidance Vocalization on the Third Trial

Figure 5: The rate of vocalizations decreases with successive trials for Fred, the dinosaur that doesn't make mistakes. The differences are significant for _guidance_ ($p = 0.018$, linear regression) and _feedback_ ($p = 0.038$, linear regression).

Figure 6: These are the distributions of the number of words spoken per second during the third trial's _guidance_ phase. In the first two trials, Fred has consistently intended to topple only correct buildings, while Kevin has intended to topple the wrong buildings. In this third trial, both dinosaurs initially intend to knock down the correct building. In _guidance_ during intent in the third trial, Kevin receives more utterances than Fred, with marginal significance ($p = 0.051$).

tuition held by human-robot interaction (HRI) researchers that naïve speakers do _spontaneously_ use strongly positive and negative affective prosody when talking to a robot.

We have also found that some human teaching behaviors do not fit well within classical machine learning models of interactive learning. Specifically, we found that human teachers tailor their feedback to account for the history of the learner's performance. In terms of a machine learning model, we view the affective vocalization reward signal as neither stationary nor path-independent, two assumptions made by standard algorithms. We found this to be true in two ways. First, the robotic learner that performs the correct action in a third trial will receive significantly more guidance and feedback if it previously made wrong choices then if it has been consistently correct. This shows that human feedback to a robotic learner is not path-independent. Second, for a learner who is consistently successful, guidance and feedback wane.

We suggest to HRI researchers interested in implementing machine learning from human vocalization that they model human reinforcement signals as dependent on the progress of the learner. Furthermore, we suggest that machine learning from human teaching should make use of currently neglected vocalizations giving direction to the robot before it acts as well as guidance to the robot as it indicates its intent to act. Direction has traditionally been ignored, and guidance has only recently been explored in standard machine learning models [28].

Our findings bear on reinforcement learning, which has been a popular approach to learning from human interaction [3, 5, 13, 15, 26, 28]. Reinforcement learning has been viewed in two senses, (1) as a broad framework of the problem of learning from rewards in an environment, and (2) as a collection of specific techniques [27]. Our results suggest that the reinforcement learning framework should be applied flexibly to learning from human interaction in order to take advantage of non-reward inputs arriving before learning-task-specific actions are taken. Such flexibility has
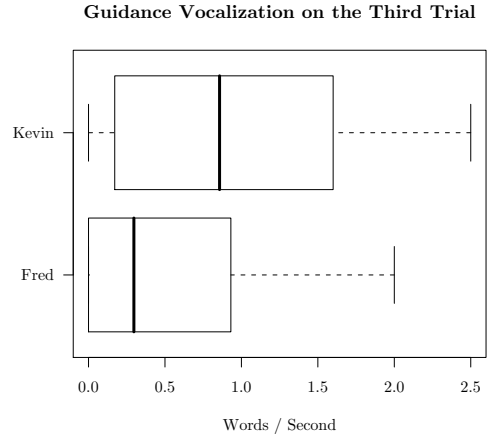
been demonstrated in the form of guided action selection, utilizing naïve people's guidance input to a learner which communicates its consideration of action options [28]. Further, we suggest a need to reconsider the characterizations of reward feedback used in classical reinforcement learning techniques based on Markov Decision Processes because we have found that human reward signals are not stationary with learning states, which stands in conflict with constraints imposed by classical algorithms [24].

Finally, our results are consistent with previous observations of human teachers' behaviors toward fellow-human learners, showing a correlation between children's improving language skills and declines in feedback from their parents [7]. As with infant learners, our findings demonstrate that people modify their inputs to a robotic learner dependent on the learner's progress.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog. In *International Conference Spoken Language Process. (ICSLP)*, volume 3, 2002.

[2] C. Bartelmus. LIRC - Linux Infrared Remote Control. In *http://www.lirc.org/*, September 2008.

[3] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson. Integrated Learning for Interactive Synthetic Characters. *ACM Transactions Graph.*, 21(3):417–426, 2002.

[4] C. Breazeal and L. Aryananda. Recognizing Affective Intent in Robot Directed Speech. *Autonomous Robots*, 12(1):83–104, 2002.

[5] J. Broekens. Emotion and Reinforcement: Affective Facial Expressions Facilitate Robot Learning. *Lecture Notes in Computer Science: Artificial Intelligence for Human Computing*, 4451:113–132, 2007.

[6] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 205–211. ACM, 2004.

[7] M. M. Chouinard and E. V. Clark. Adult Reformulations of Child Errors as Negative Evidence. *Journal of Child Language*, 30(03):637–669, 2003.

[8] J. Cohen. Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, 70(4):213–220, 1968.

[9] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of Oz Studies: Why and How. In *Proceedings of the 1st International Conference on Intelligent User Unterfaces*, pages 193–200. ACM, 1993.

[10] DogsBody & Ratchet Software. MySkit – Performance Editor for PLEO. In *http://www.dogsbodynet.com/myskit/index.html*, January 2009.

[11] K. Gold, M. Doniec, C. Crick, and B. Scassellati. Robotic Vocabulary Building Using Extension Inference and Implicit Contrast. *Artificial Intelligence (AIJ)*, 173:145–166, 2009.

[12] IguanaWorks. IguanaWorks. In *http://iguanaworks.net/*, September 2008.

[13] C. Isbell, C. R. Shelton, M. Kearns, S. Singh, and P. Stone. A Social Reinforcement Learning Agent. In *Proceedings of the 5th International Conference on Autonomous Agents*, pages 377–384. ACM, 2001.

[14] L. Kaelbling, M. Littman, and A. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4(2):237–285, 1996.

[15] F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklósi. Robotic Clicker Training. *Robotics and Autonomous Systems*, 38(3-4):197 – 206, 2002.

[16] A. Kapoor and R. W. Picard. Multimodal Affect Recognition in Learning Environments. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pages 677–682. ACM, 2005.

[17] E. S. Kim and B. Scassellati. Learning to Refine Behavior Using Prosodic Feedback. In *Proceedings of the International Conference on Development and Learning (ICDL)*, 2007.

[18] C. M. Lee and S. Narayanan. Toward Detecting Emotions in Spoken Dialogs. *IEEE Transactions on Speech and Audio Processing*, 13(2):293–303, March 2005.

[19] D. J. Litman and K. Forbes-Riley. Recognizing Student Emotions and Attitudes on the Basis of Utterances in Spoken Tutoring Dialogues with Both Human and Computer Tutors. *Speech Communication*, 48(5):559–590, Feb 2006.

[20] C. Nass, J. Steuer, and E. R. Tauber. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, pages 72–78, 1994.

[21] M. Nicolescu and M. Matarić. Learning and Interacting in Human-Robot Domains. *IEEE Transactions on Systems, Man and Cybernetics*, 31(5):419–430, 2001.

[22] C. Osgood, G. J. Suci, and P. H. Tannenbaum. *The Measurement of Meaning*. Univ. of Illinois Press, 1957.

[23] A. Robinson-Mosher and B. Scassellati. Prosody Recognition in Male Infant-Directed Speech. In *Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems (IROS)*, 2004.

[24] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, 2nd edition, 2003.

[25] L. Steels and F. Kaplan. AIBO's First Words: The Social Learning of Language and Meaning. *Evolution of Communication*, 4(1):3–32, 2001.

[26] A. Stern, A. Frank, and B. Resner. Virtual Petz (video session): A Hybrid Approach to Creating Autonomous, Lifelike Dogz and Catz. In *Proceedings of the 2nd International Conference on Autonomous Agents*, pages 334–335. ACM, 1998.

[27] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[28] A. L. Thomaz and C. Breazeal. Reinforcement Learning with Human Teachers: Evidence of Feedback and Guidance with Implications for Learning Performance. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006.

[29] A. L. Thomaz and C. Breazeal. Transparency and Socially Guided Machine Learning. In *Proceedings of the International Conference on Development and Learning (ICDL)*, 2006.

[30] UGOBE. PleoWorld - The Home of Pleo, the Robotic Baby Dinosaur from UGOBE Life Forms. In *http://www.pleoworld.com*, September 2008.

[31] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, pages 126–133. ACM, 2007.