Learning to Refine Behavior Using Prosodic Feedback

Elizabeth S. Kim
Yale University
51 Prospect St., New Haven, CT 06520
eliskim@cs.yale.edu

Brian Scassellati
Yale University
51 Prospect St., New Haven, CT 06520
scaz@cs.yale.edu

***Abstract* – We demonstrate the utility of speech prosody as a feedback mechanism in a machine learning system. We have constructed a reinforcement learning system for our humanoid robot Nico, which uses prosodic feedback to refine the parameters of a social waving behavior. We define a waving behavior to be an oscillation of Nico's elbow joint, parameterized by amplitude and frequency. Our system explores a space of amplitude and frequency values, using q-learning to learn the wave which optimally satisfies a human tutor. To estimate tutor feedback in real-time, we first segment speech from ambient noise using a maximum-likelihood voice-activation detector. We then use a k-Nearest Neighbors classifier, with *k*=3, over 15 prosodic features, to estimate a binary approval/disapproval feedback signal from segmented utterances. Both our voice-activation detector and prosody classifier are trained on the speech of the individual tutor. We show that our system learns the tutor's desired wave, over the course of a sequence of trial-feedback cycles. We demonstrate our learning results for a single speaker on a space of nine distinct waving behaviors.**

Index Terms* – *speech prosody, human-robot interaction, reinforcement learning, socially-guided machine learning.

I. Introduction

A. *Socially-Guided Machine Learning*

As robots increasingly appear in human environments, we must equip them with the power to adapt in response to new information about these environments. Humans modify their behaviors in response to sensing a host of cues, including social cues from other humans. If a robot cannot appropriately respond to human social cues, it will tend to disrupt a human environment, in conflict with its presumably assistive purpose in being among humans.

One important response to a social communication is to adjust one's behavior. This process can be viewed as learning in response to feedback. Besides being important for adapting to new information in the environment (e.g., avoiding danger, keeping a secret upon a stranger's entrance to the room), learning in response to human social cues is important for human-robot cooperative tasks.

A recent exploration into human-guided machine learning has revealed that a simulated robot can learn a simple sequential task, such as a cleaning up a virtual kitchen, given feedback from a human tutor. In Sophie's Kitchen, a tutor communicates using a mouse to scroll a feedback meter between extremes of strong approval and strong disapproval [1].

The present work extends the exploration of human-guided machine learning into the physical world, where a robot learns to modify its behavior, given a more naturally social human communication: speech prosody.

B. *Communicating Prosodic Affect to Robots and Computers*

Speech prosody is essentially "tone of voice." It is comprised of the highness or lowness, the scratchiness or smoothness, the loudness or softness, and the quickness or slowness, with which a speaker can alter their pronouncement of an utterance. Functionally, while prosody also communicates syntactical and pragmatic information, in the present work we are concerned with its function as a mode for communicating emotions and attitudes, or *affect*.

Humans modulate their tones of voice to communicate affect. We raise our voices in frustration, or comfort small children using hushed speech. We use consistent tones of voice to indicate displeasure or joy to our pets.

In the last decade, numerous studies have shown that, with varying degrees of constraint and accuracy, affect can be classified automatically from recordings of speech [2, 3, 4, 5].

Among these recent studies, Breazeal and Aryananda developed a prosodic affect recognition system on the humanoid robot Kismet, which classified praise, prohibition, attention, and comfort in the speech of a human tutor, and responded by assuming a corresponding, hard-coded expressive posture, for instances, hunching over and frowning in response to detected prohibition, or perking its ears and raising its eyebrows in response to approving speech [3]. Breazeal and Aryananda suggested that speech prosody serve as a training signal for a robotic learning system, but they stopped short of implementing a learning system. Kismet's hard-coded expressive posture was displayed to provide the tutor with feedback on the robot's classification of the tutor's prosody, but the classification was not used to drive a learning system.

In the present work, in response to affective prosody, we extend beyond hard-coded expressive postures to using prosodic affect recognition to drive a system which learns to refine the social behavior of waving.

II. Refining Behavior Using Prosodic Feedback

We have implemented our prosody-driven learning system on our humanoid robot Nico, within Nico's lab environment. Our learning system is trained using an interaction loop, shown in Fig. 1. For each iteration of the interaction loop, Nico performs a waving behavior, after

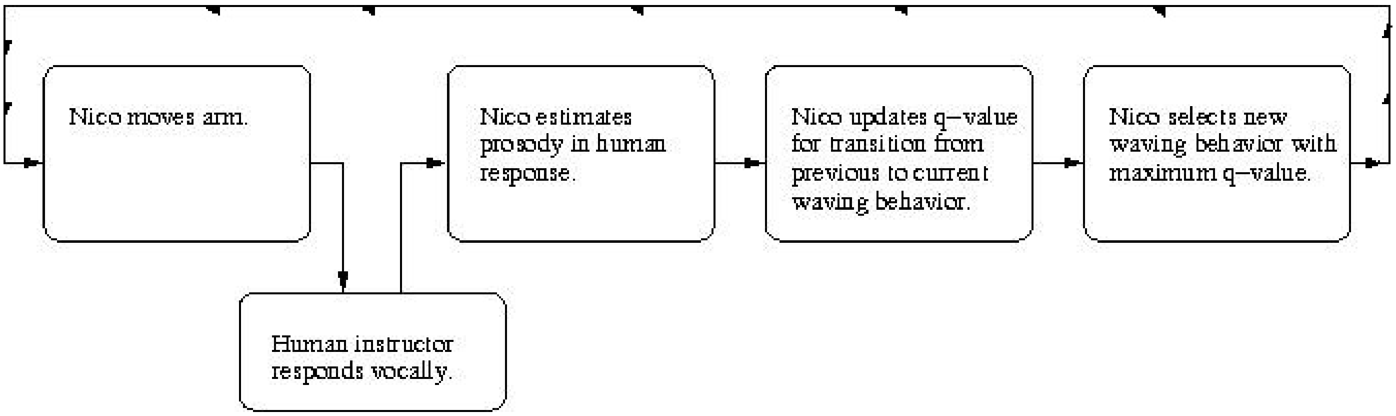

Figure 1. Interaction loop flow for prosody-driven learning. This loop iterates until Nico selects the same waving behavior a pre-determined number of cycles in a row, at which point it declares that behavior of fixation to be the goal behavior. Nico's estimate of prosodic affect takes the form of a binary approval/not-approval signal.

which it waits a pre-determined amount of time for a possible utterance of feedback. If the tutor utters a response, the affect of the utterance is calculated, producing a binary approving/not-approving result. This binary approval signal is the feedback which drives the q-learning system. Nico iterates through the interactive loop until the q-learner fixates for some pre-selected number of cycles on a single waving behavior, which Nico estimates to be the goal behavior.

A. Robotic Embodiment

Nico, shown in Fig. 2, is an upper-torso robot, built in the proportions of a one-year-old infant. Nico is equipped with a seven degree-of-freedom neck and head assembly, and a six degree-of-freedom arm. Nico wears a fixed smile and infant clothing, encouraging humans to interact with it socially.

We make a fundamental assumption regarding human interaction with Nico: we assume that people will interact with Nico as though it is a small child or an infant, speaking to it using exaggerated prosody. Whereas even human listeners struggle to identify prosodic affect in utterances spoken to adults, it is easier to classify affect in the prosody of infant-directed speech [6]. Psychologists have observed that speakers tend to use exaggerated prosody with infants, in a speaking style called *"Motherese"* [7]. Breazeal and Aryananda observed that people tend to extend their use of Motherese to their robot Kismet, a humanoid with facial features designed to appear childlike [3].

Breazeal and Aryananda, and Robinson-Mosher and Scassellati's prosody classification results are based on robot-directed speech and assume that humans tend to speak to infantile or childlike humanoid robots, using Motherese [2, 3].

B. Interaction Environment and Audio Capture

Nico's tutor's utterances are recorded in a real-time interaction loop, coupled with Nico's actions, within our lab environment. Acoustically, the lab environment is extremely noisy, given the unavoidable proximity of a rack of computers controlling Nico's motor and visual systems.

The tutor's speech is recorded using a mono-input microphone clipped to the tutor's clothing, within six inches of the tutor's mouth, to increase signal energy, given high environmental noise.

Following acknowledgement that Nico has finished performing its waving behavior, three seconds of audio are recorded, within which time the tutor has presumably responded to Nico's movement.

Figure 2. Our humanoid robot Nico, waving. Nico is built with the proportions of a one-year-old human infant.

C. *Overview of Prosody Estimation*

We estimate prosodic affect within each audio response clip as follows:

1. We first cut the response clip into overlapping, short-time windows, each 25ms long. The start times of neighboring windows are separated by 10ms. Short-time windowing is necessary for spectral analysis of auditory data, in order to employ notions of stationarity in frequency for any temporal segment. These short-time windowing values are standard in speech recognition [8].
2. We perform voice-activation detection (VAD), checking each short-time window for speech. We then concatenate consecutive windows to form continuous speech segments, smoothing over brief inconsistencies in VAD output.
3. We estimate the prosodic affect in each speech segment, and send this estimate to the waving behavior learner.

D. *Speech Segmentation*

We use a VAD to segment this three-second response clip to isolate short-time (10 ms-separated, 25 ms-long) windows containing speech. For each window, our voice-activation detector conducts maximum-likelihood detection over three features calculated over a short-time window of the acoustic signal $x_I[n]$:

1. total energy over the window

$$energy_I = \sum_{1}^{N} (x_I[n])^2 \tag{1}$$

2. variance of the log-magnitude-spectrum

$$vlms_I = \mathrm{var}(\log(abs(X(j\omega))) \tag{2}$$

3. variance of the log-spectral-energy.

$$vle_I = \mathrm{var}(\log(X(j\omega)^2) \tag{3}$$

The VAD is trained on auditory data recorded from the tutor's voice before the interaction loop begins. For our learning system, we trained on 15 seconds of continuous, ambient noise, and 8 seconds of continuous, uninterrupted speech from the speaker. Voice-activation detection results are presented under Experimental Results.

E. *Classification of Prosody by k-Nearest Neighbors*

Our prosody classifier decides whether or not an utterance indicates *approval*. We choose to map utterances to a simple, binary approving/not-approving signal because such a binary signal can apply to various machine learning contexts, and to simplify affect classification.

Previously we described prosody as "tone of voice." Others have called it the "melody and rhythm" of speech. More precisely, prosody is described psychoacoustically in terms of pitch, volume, timbre or voice quality, as well as temporal features such as frequency of consonants. In this paper, we are concerned only with pitch and volume, as we presume that for the purposes of providing approving and disapproving feedback, the tutor will tend to produce consistently short utterances in a consistent tone of voice. Physically, *pitch* and *volume* have correlates in measurements of *fundamental frequency—f0—(*for periodic signals) and *acoustic energy* [8].

We have designed our classification features based on those used by Robinson-Mosher and Scassellati in the same noisy lab environment. Our 15 features are comprised of statistics derived from estimates of from pitch, energy, and energy-weighted pitch. Each of these measurements is estimated for each short-time window in the speech segment.

We estimate f0 using a Noll's cepstral method [9, 8]. We post process f0 estimates by applying a temporal smoothing filter, which averages each window's f0 estimate with those of its two immediate neighbors.

We estimate energy for each speech segment window according to Eqn. 1. Finally, we derive a new measurement, for each short-time window, of energy-weighted pitch by taking the product of the pitch and energy estimates.

From these three measurements of pitch, energy, and energy-weighted pitch, we calculate the mean, variance, non-zero minimum, maximum, and range (or maximum-minimum) values over the speech segment. This gives us our 15 classification features.

We presume that our binary classes of approval and not-approval will separate well and cluster within each class. Therefore, we use *k*-Nearest Neighbors, to classify novel utterances. High accuracy in preliminary trials led us to select *k=3*.

The prosody classifier's training data is acquired from the individual tutor, in an interaction loop similar to the final learning interaction loop. To generate training examples of approving and not-approving prosodic affect, the tutor is given a simple, interactive training game, in a similar style to the interaction sequence used to train Nico. This training game is designed to elicit prosody similar to that elicited during Nico's waving training, and to provide automatic labeling for the prosody classifier's training data.

In the prosody classifier's training game, the tutor is told to train a remote robot on how far it must travel from a hazard to reach safety. The tutor is given the threshold of safe distance from a practice hazard. The tutor is allowed only to provide the remote robot with information via tone of voice. Training involves presentation to the tutor of a sequence of distances traveled by the remote robot. In response to each distance reported, the tutor must give the robot prosodic feedback. These feedback utterances form the corpus of training examples to the prosody classifier. Because the robot's performance and the threshold are known before the tutor produces each training example, the examples are easily, automatically labeled.

F. *Reinforcement Learning of Waving Behavior Parameters*

We demonstrate prosody as a feedback mechanism for the problem of refining Nico's social waving behavior.

We define a waving behavior to be an oscillatory motion at Nico's elbow joint, around a fixed raised arm and hand position. A waving behavior can be parameterized by the amplitude (measured in joint angle degrees) and frequency of oscillation.

In this paper, Nico is presented with a space of nine waving behaviors, combining three amplitudes ranging from small to large, and three frequencies ranging from slow to fast. The space of waving behaviors is organized as shown in Fig. 3. Each box in the figure represents a single waving behavior. During each trial-feedback cycle, Nico can transition to a new waving behavior if it shares an edge with its most recent waving behavior's box, or Nico can choose to repeat the same waving behavior.

Before beginning an interactive tutorial with Nico, the tutor chooses a goal for what kind of waving behavior she would like Nico to perform. Nico initiates the tutorial by arbitrarily selecting a waving behavior and performing it.

Nico uses q-learning to discover the tutor's desired wave state. Therefore, Nico maintains an internal estimate of the utilities or q-values, *Q(s,a)*, for the transitions between each waving behavior. Here, *s* is a waving behavior *s*tate, and *a* is an *a*ction or transition from *s* to Nico's next waving behavior state *s'* [10].

Following each transition and its demonstration of its newly selected waving behavior, Nico's q-learner receives prosodic feedback, $R(s,a) \in \{0,1\}$ which it treats as a q-learning reward signal and uses to update the q-value for its most recent transition between waving behaviors:

$$Q(s,a) \Leftarrow (1-\alpha)Q(s,a) + \alpha(R(s,a) + \gamma \max_{a'} Q(s',a')) \quad (4)$$

where *s'* is the next waving behavior, which transition *a* leads to, from previous waving behavior *s*, and *a'* is a transition leading from waving behavior *s'* [10].

The q-learning parameters α and γ influence the sensitivity of q-values to changes in the q-values of successor states and transitions, and the number of predecessor states and transitions, whose values will be affected by updates to q-values, respectively [10].

As for Nico's choice for its next action, we have selected the following action policy: from its current waving behavior, with some probability (1-*p)* Nico selects the transition with the highest q-value. However, with probability *p*, Nico instead uniformly randomly selects a transition.

Random exploration of the state space is important for two reasons. First, in the case where Nico misclassifies the prosodic affect, it can update its q-values to incorrectly prefer an undesirable transition. In such a case, random exploration can give Nico a new opportunity to correct its error or return to an optimal path to goal. Secondly, if the state space of waving behavior should contain a local maximum, and if Nico finds itself performing the locally optimal waving behavior, a random transition away from the local maximum can give Nico an opportunity to seek the global optimum.

In order to allow Nico to finally converge on some waving behavior, we allow the probability *p* of random exploration to decrease geometrically with each trial-feedback cycle:

$$p \Leftarrow p * \xi^n \quad (5)$$

where *n* is the number of cycles

For our waving behavior refinement problem, we select α = .5; γ = .8; *p* = .7; ξ = .95. We arbitrarily choose to let Nico start with a small, slow waving behavior, and we assign the tutor to prefer a big, fast waving motion.

III. Experimental Results

A. Experimental Parameters

The following sections present the results of a wave behavior learning demonstration, which uses prosodic feedback to drive a q-learning algorithm. Results are presented for a single tutor, the first author of this paper.

Voice-activation detection was trained on 15s of background noise and 8s of continuous speech.

Prosody classification was trained on 19 approving and disapproving feedback utterances, captured and labeled using the remote robot safety scenario.

Q-learning was performed using parameters α = .5; γ = .8; *p* = .7; ξ = .95, where the learner's action-selection policy was to choose, with probability $1\text{-}p * \xi^n$, the behavior transition having maximal q-value, and, with probability $p * \xi^n$, to select a transition uniformly at random.

B. Voice-Activation Detector Performance

The maximum-likelihood voice-activation detector exhibited only 3.4% error, including both false positives and negatives, when tested over the VAD training corpus itself. We did not measure VAD error for any live interaction data, as we do not have a means to automatically acquire true voice-activation labels in during tutorial. Fig. 4 shows

Small, Fast (30°, 3.1Hz)	Medium, Fast (40°, 3.1Hz)	Big, Fast (70°, 3.1Hz)
Small, Medium (30°, 1.9Hz)	Medium, Medium (40°, 1.9Hz)	Big, Medium (70°, 1.9Hz)
Small, Slow (30°, 1.3Hz)	Medium, Slow (40°, 1.3Hz)	Big, Slow (70°, 1.3Hz)

Figure 3. A space of nine distinct waving behaviors. Each box represents one waving behavior state. In our experimental state space, waving behavior states are ordered from left to right with increasing amplitude, and from bottom to top with increasing frequency.

distributions of background noise and speech training samples, for energy, the decision feature exhibiting the best separation between the noise and speech distributions. The background noise distribution is shown in the top plot, and the speech distribution is shown in the bottom plot.

C. Prosody Classification

The prosody classifier was trained on a corpus of 19 utterances, including 10 of approving and 9 of disapproving affect. Leave-one-out cross-validation over training data results in false positive and miss rates both of 5.3%.

Fig. 5 shows the training utterances' distributions over f0-mean and energy-range features. Over these two features, the training data shows clear separation.

Prosody classifier performance was measured over the particular tutoring interaction sequence presented in Fig. 5. Presuming that the author's prosody was consistent with the desirability of each tried transition, the true prosodic values were estimated from Nico's sequence of transitions. For every transition which brought Nico closer to the desired waving behavior, the true prosody was estimated to be approving, and for transitions that brought Nico farther from the goal behavior, the true prosody was estimated to be not-approving. A comparison of these estimated true prosody values with the actual output of the prosody classifier showed that the prosody classifier made 0 false positive errors and missed two, or 8.3%, of all approving utterances, falsely classifying them as not-approving.

D. Learning the Tutor's Goal Behavior

Fig. 6 shows Nico's approach to the desired waving behavior, over the course of an interactive tutoring sequence. The plot shows the cumulative distance, measured in transitions, between Nico's current behavior and the goal behavior. The cumulative error curve has steep slope for trials during which Nico's waving behavior is very different from the goal behavior. On the other hand, the cumulative error curve is horizontal for those trials during which Nico is performing the goal behavior.

Fig. 6 indicates that Nico performs the goal behavior five times in a row, from the 12th-17th trials, and then transitions away from the goal behavior, exploring behaviors which are far from the goal, until finally returning back to the goal behavior.

If Nico found the goal behavior, why did it later switch to another behavior? The answer is that Nico's action policy calls for it to select the transition with optimal q-value most of the time, but with some probability to select, uniformly at random, any transition, regardless of q-value. In Fig. 7, circles mark the trials during which Nico randomly selected its next behavior. Note that from the beginning through the middle of the sequence, when Nico selects a transition at random, this results in the accumulation of error, as Nico explores far from the goal behavior. However, near the end of the sequence, Nico recovers to the goal behavior rapidly, as indicated by the horizontal cumulative error curve.

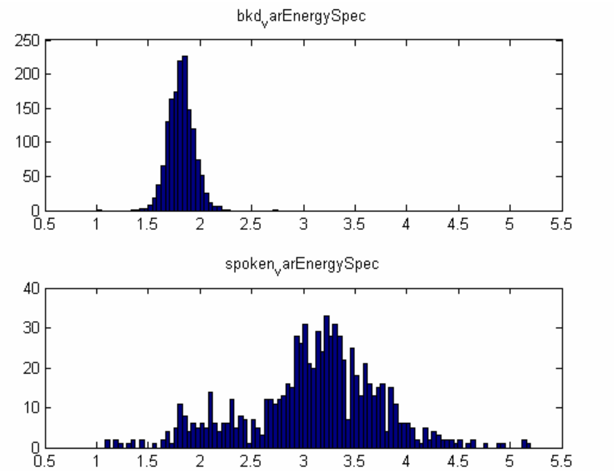

Figure 4. Training background (top) and speech (bottom) histograms over energy measurements, one of three voice-activation detection (VAD) features. The VAD derives Gaussians probability distributions from these sample distributions, and performs maximum-likelihood detection on novel short-time audio windows.

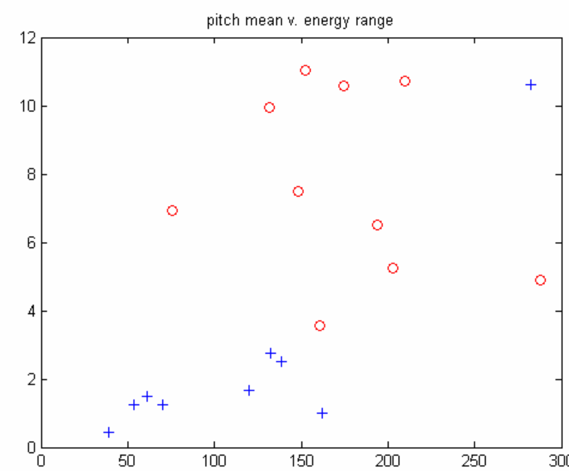

Figure 5. Prosody classifier training data distributed over f0-mean and energy-range features. Utterances featuring approving prosody are marked by "+"s and utterances featuring disapproving prosody are marked by "o"s. For these two features, the training data shows clear separation.

In general it is safe to expect that as the number of trials increases, Nico will find and stay fixated on the goal behavior. This is because the probability of randomly exploring away from the goal behavior decreases geometrically with time, and because as time passes, Nico enriches its model for the space of behaviors. For example, Nico's path indicates that during trials 25-33, Nico explored previously unvisited behavior states, causing it to learn q-values over these novel states.

This knowledge allowed Nico to recover immediately to the goal state, following random explorations away from it in trials 35 and 40. After 47 trials within this sequence, the probability of randomly choosing an action was only 6.3%.

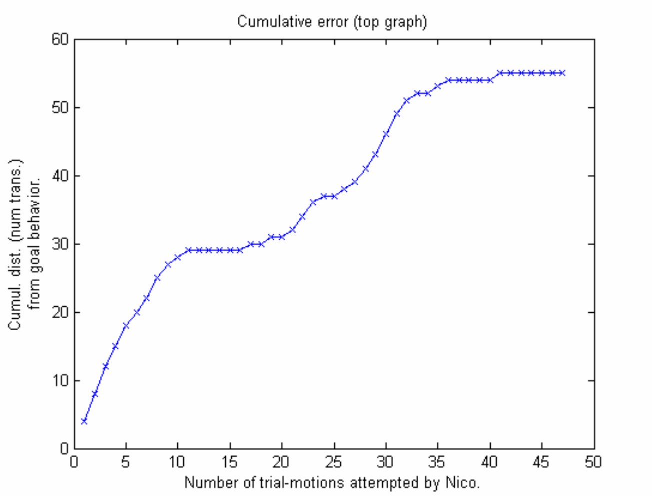

Fig. 6. Convergence of waving behavior q-learner onto desired waving behavior. This plot shows cumulative error versus number of trials in the tutorial sequence. Horizontal slope in the cumulative error curve indicates transition to the goal behavior, producing no additional error.

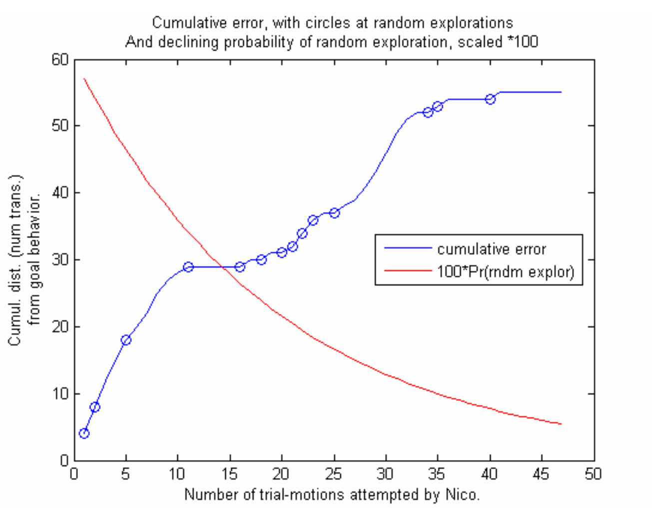

Figure 7. Another view of the learning sequence. In this plot, “o”s demarcate the trials during which Nico chose its next waving behavior uniformly at random, and also showing the declining probability (scaled by a factor of 100) of such random exploration , scaled by a factor of 100.

IV. Discussion

A. Prosody as Feedback to Drive Machine Learning

We have demonstrated that speech prosody can drive a physically embodied machine learner. We have shown that it is possible to recognize affect in prosody in a real-time, interactive loop, with a high level of accuracy that makes possible its usage in a real-time learning system.

B. Extension to Other Subjects

Initial explorations using other subjects to tutor Nico have confirmed the importance of affective response from the robot, as previously demonstrated by Breazeal and Aryananda on Kismet [3]. Even mothers of infants and highly expressive caretakers of pets, who are accustomed to speaking Motherese with their children or animals, indicate reluctance to express exaggerated affect to Nico, in the absence of affective feedback from the robot.

C. Extension to Larger Learning Problems

A pilot study into more complicated learning problems, for instance exploration to a social waving behavior within a space of 500 distinct arm-oscillation behaviors, a subset of which appear to express social waving, suggests that convergence on the goal state is highly sensitive to the balance which the action policy gives to exploration over exploitation, depending on the recent reward history.

D. Extension to More Affects

Previous work in prosody classification has successfully classified over other affects, besides approval and disapproval. Thomaz, Hoffman, and Breazeal showed that humans often prefer to give guidance, which may be viewed as attentive affect, as well as positive or negative reinforcement, which may be viewed as approving or disapproving affect [1]. It’s possible that other prosodic affects may enrich a tutoring interaction with a robot, by providing feedback other than positive and negative reinforcement.

E. Extension to Other Learning Problems

The present work should extend to other learning problems for which a human can judge whether a change in state is going “in the right direction.” Essentially, a binary approving/not-approving signal can be used to indicate positive gradient over utility.

References

[1] A. L. Thomaz, G. Hoffman, and C. Breazeal. "Reinforcement Learning with Human Teachers: Understanding how people want to teach robots." *In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2006.

[2] A. Robinson-Mosher and B. Scassellati, “Prosody recognition in male infant-directed speech,” in *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Sendai, Japan, 2004.

[3] C. Breazeal and L. Aryananda. "Recognizing affective intent in robot directed speech,” *Autonomous Robots*, vol. 12, no. 1, pp. 83-104, 2002.

[4] 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 on Spoken Language Processing (ICSLP), vol. 3, Denver, CO, 2002.

[5] M. Slaney and G. McRoberts, “Baby ears: a recognition system for affective vocalizations,**”** *Speech Communication, vol. 39, no. 3-4, pp. 367-384, Feb. 2003.*

[6] J. Liscombe, J. Venditti, and J. Hirschberg. “Classifying subject ratings of emotional speech using acoustic features,” in *Proceedings of Eurospeech 2003*, 2003.

[7] A. Femald. “Intonation and communicative intent in mothers' speech to infants: Is the melody the message?” in *Child Development*, vol. 60, pp. 1497-1510.

[8] T. F. Quatieri. Discrete-Time Speech Signal Processing, Principles and Practice. Prentice Hall, Upper Saddle River, NJ, 2002.

[9] A. M. Noll. “Cepstrum pitch determination,” *Journal of the Acoustical Society of America*, vol. *41*, pp. 293–309. Feb. 1967.

[10]S. Russell and P. Norvig. Artificial Intelligence: A Modern Approach, 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 2002.