Learning to Press Doorbell Buttons in Real Time

Vladimir Sukhoy, Jivko Sinapov, Liping Wu, and Alexander Stoytchev
Developmental Robotics Laboratory
Iowa State University
{*sukhoy, jsinapov, lipingwu, alexs*}@*iastate.edu*

***Abstract*— This paper describes an approach that a robot can use to learn to press doorbell buttons in real time. This approach combines exploratory behaviors with an active learning strategy to enable the robot to learn how and where it should press a button in order to trigger the buzzer. The framework was tested with an upper-torso humanoid robot on seven different doorbell buttons with three different learning strategies (random, stimulus-driven exploitation, and uncertainty-driven exploitation strategy). The results show that an active learning strategy can significantly speedup the robot's learning progress. Among the three strategies that were evaluated, the uncertainty-driven exploitation strategy was the most effective.**

I. Introduction

To operate in human-inhabited environments a robot must be able to press buttons. A robot can encounter thousands of different buttons, which perform a wide variety of tasks in homes, offices, and even on the streets of cities where humans live. In this work the robot learned to press seven doorbell buttons. To provide the ground truth these buttons were connected to a buzzer. The robot explored these buttons with three exploration strategies and perceived the proprioceptive, auditory, and visual feedback during these explorations. The robot localized events in time and space and used these localizations to relate an event to the button for which it was observed and to analyze temporal intervals between events in different sensory modalities.

This work focuses on different strategies that a robot can employ to learn to press buttons faster. Our robot evaluated three learning strategies: random strategy, stimulus-driven exploitation strategy, and uncertainty-driven exploitation strategy. Under the random strategy the robot did not use what it learned in the past to change the behaviors that it performed. Under the stimulus-driven exploitation strategy the robot performed a pushing behavior that was most likely to trigger the buzzer based on prior experience. Under the uncertainty-driven exploitation strategy the robot performed a pushing behavior for which it was most uncertain whether it will trigger the buzzer or not.

Different learning strategies generate events in different spatial locations around the buttons. The event distributions in visual space for both the stimulus-driven and the uncertainty-driven exploitation strategies are denser than the corresponding distributions for the random strategy. The different learning strategies also varied by their efficiency. The robot tested the performance of each of the three strategies on two learning tasks: 1) predicting if a behavior triggers the buzzer; and 2) predicting whether pushing a button in a specific location in visual space triggers it. For these two tasks it was estimated how much training is required for the robot to achieve a given performance level. The results indicate that for both learning tasks the uncertainty-driven exploitation strategy is the most efficient of the three learning strategies.

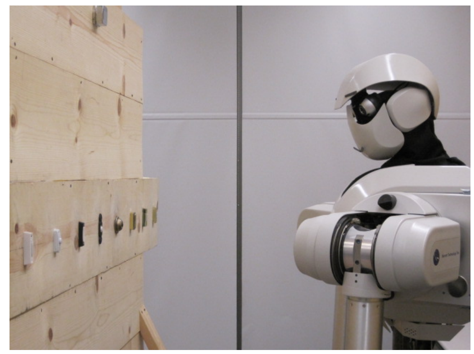


Fig. 1. The upper-torso humanoid robot used in the experiments, shown here looking at the experimental fixture with the buttons.

II. Related Work

A. Developmental Psychology

Hauf and Ascherleben [1] demonstrated that 9 months old human infants anticipate the acoustic and visual events associated with pressing different colored buttons. The broader goal of their work was to show that infants use anticipation of action outcomes to control their actions. Hutt [2] summarized the results of a number of studies of curiosity and exploration in 3 to 5 year old children and found that a child first explores a new toy and then plays with it. The first behavior category is formally termed as "specific," while the second one as "diversive" exploration. The question that specific exploration answers is "What can the object do?", while the diversive exploration is concerned with "What can I do with the object?".

Piaget and Inhelder [3] found three stages in the development of haptic exploration skills in children: random exploration (between 3.5 and 4.5 years), active exploration (between 4.5 and 6 years) and systematically planned exploration (between 6 and 7 years). E.J. Gibson [4] showed that humans use observations obtained from active exploration as one of the key sources of knowledge about the world and, in particular, about the affordances of objects.

B. Robotics

In robotics the related work on pressing buttons has mostly focused on pressing buttons to accomplish some other task and not necessarily on learning how and where to press them. Thomaz [5] proposed to use social guidance to teach a robot to press buttons. The robot in this study pressed colored buttons that a human pointed at. The focus was on the social aspects of learning and not on the task of learning how to press the buttons. Nguyen et al. [6] proposed a "clickable world" where a human points at a button with a laser pointer and the robot presses it. Their robot used an omnidirectional camera and a pan/tilt stereo camera to estimate the location of the button from the click. Nguyen et al. [7] also proposed to use tags that combine physical, perceptual and semantic information (PPS-Tags) to help a robot locate and use various buttons and switches. For a given object, the PPS-Tag informs the robot where the object is, what it does, and how to use it by providing specific commands to the robot. Miura et al. [8] proposed that a robot can use visual search for a button template in the region of visual space to which a human points. If the robot was successful in its search for the visual template, it pressed the button using the template location. In another study with elevator buttons, Song and Wu [9] proposed to use a neural network for real time tracking of objects in video frames. The robot identified buttons on an elevator panel and pressed them using visual servoing.

Baranes and Oudeyer [10] argued that heuristics used in intrinsically-motivated learning can be viewed as a type of active learning algorithm targeted toward learning in sensorimotor spaces. Barto et al. [11] proposed an approach for learning skill hierarchies based on intrinsically motivated reinforcement learning. Marshall et al. [12] provided philosophical grounding and an algorithm to facilitate an intrinsically-motivated system to control a developmental robot. They proposed to view self-motivation as a combination of two competing goals: the need to accurately predict the environment and the need to investigate novel regions in it. Several studies [13] [14] [15] applied active learning algorithms to help a robot to learn to grasp an object.

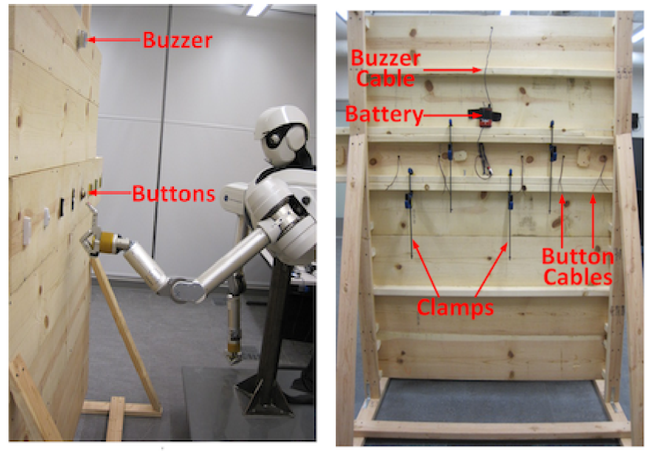
III. Experimental Setup

A. Robot

The experiments were performed with the upper-torso humanoid robot shown in Fig. 2. The robot's arms are two Barrett Whole Arm Manipulators (WAMs). The end effector of each arm is a BH8-Series Barrett Hand. A color marker was attached to the tip of the robot's finger to simplify its visual tracking (see Fig. 3).

B. Buttons and Fixture

The robot experimented with 7 doorbell buttons, which were mounted on a wooden fixture (see Fig. 2). The middle segment of the fixture can slide left to right so that a different button can be presented to the robot during various trials. Behind the wall, the currently explored button is connected to a buzzer and a battery that powers it (see Fig. 2(b)). The buttons were selected from the ones available in stock at a local Lowe's store (a home improvement store).



(a) The robot pushing a button. (b) Experimental fixture (back).

Fig. 2. The experimental setup.

C. Experimental Trials

The robot used 3 different exploration strategies, which are described in Section IV-B. For each strategy the robot performed 200 experimental trials with each button. In addition to that, the robot performed 400 random trials to collect an evaluation set that was used to verify the performance of each of the three learning strategies. Thus, the robot performed $3 \times 200 + 400 = 1000$ experimental trials with each of the 7 buttons, or 7000 total trials.

Each trial consisted of 5 pushing behaviors directed at the button or the area around it. The starting position for each push was the end position of the previous push. Before each push, the robot first moved its arm away from the fixture and then pushed in the area around the button. The end point of the push was selected based on the learning strategy that was used. To randomize the starting position of each trial, the robot started with a random push that was not counted toward the 5 pushes in the trial. Each trial lasted for approximately 18-20 seconds. In other words, the robot performed one pushing behavior every 3 seconds (plus some setup time). All 1000 trials with each button were performed one after another without interruption. The data collection for each button took approximately 6 hours.

Another way to describe this dataset is: performed 5 pushes $\times$ 1000 times $\times$ 7 buttons $=$ 35000 pushes.

D. Sensory Data

During the experiments the robot recorded visual, auditory, and proprioceptive data. The data was processed in real

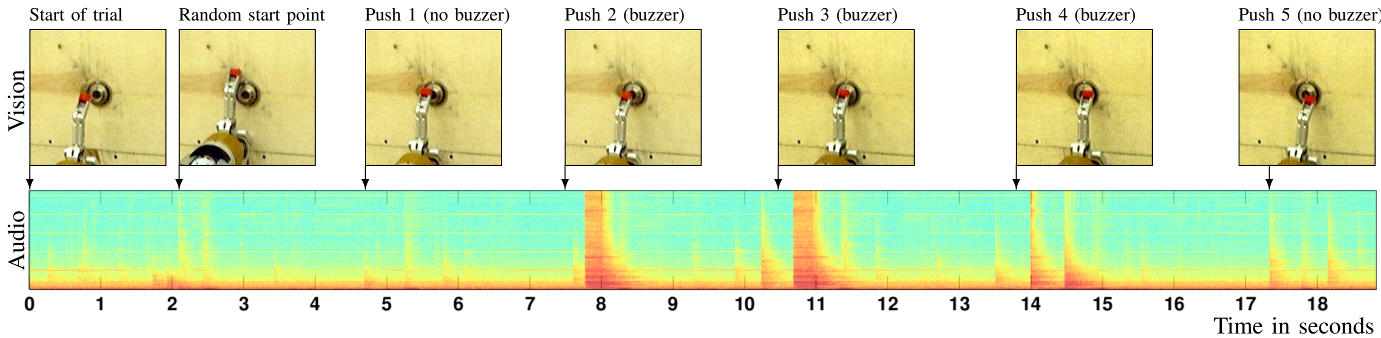


Fig. 3. One of the trials performed by the robot. The spectrogram of the sound is matched to the corresponding video frames for each of the five pushing behaviors. The robot's field of view is larger than the images shown here, which were cropped to show only the area around the button.

time but it was also stored to disk for additional offline analysis. Vision data was recorded at 10 frames per second from the robot's left eye (a Logitech QuickCam Pro 4000 web camera) at 640×480 resolution. An Audio-Technica U853AW Hanging Microphone, mounted in the robot's head was used to record audio at 44.1 kHz. Proprioceptive data, in the form of joint position and torque readings, was recorded from the left WAM arm at 500 Hz.

As the robot performed the experiments, it detected and timestamped "interesting" proprioceptive and auditory events. If the torque magnitude exceeded its predefined limit, the robot recorded a proprioceptive event, interrupted the current pushing behavior, and started the next one.

The robot recorded an auditory event if it heard a buzzing sound. This was done by analyzing the spectrograms of candidate regions in the audio stream. The candidate regions were selected if the volume of the audio exceeded a predefined threshold. If a candidate region was found to contain the buzzer sound, then an auditory event was created and its timestamp was set to the timestamp of the first sample in the candidate region. If the robot found more than one candidate region with buzzing sound during a behavior, it recorded an auditory event only for the first of these regions. To limit the computational load, the robot did not consider candidate regions that were shorter than 0.03 seconds. For a candidate region, the maximum duration was set to 0.1 seconds. The Discrete Fourier Transform (DFT) was calculated for every region to obtain a spectrogram. A frequency component histogram with 20 bins was calculated for every spectrogram. To decide if a candidate region contained a buzzing sound the robot used a cascade of two classifiers on the histogram. The cascade consisted of the Naive Bayes and the K^* classifier [16]. The structure of the cascade was determined from a separate pilot study in which a training set was manually marked up and candidate regions were classified into buzzer and non-buzzer sounds.

E. Behavior Parametrization

A pushing behavior is parametrized with two unit vectors $x^{(b)}$ and $x^{(p)}$ in the robot's Cartesian space. $x^{(b)}$ is the direction of the backup movement after the finger hits the button or the wall and $x^{(p)}$ is the direction of the push that follows.

Thus, each behavior was parametrized by a vector $x \in \mathbb{R}^6$, which is a concatenation of the two vectors $x^{(b)}$ and $x^{(p)}$. The robot started each pushing behavior from the end point of a previous push. The start point for the first push behavior in any trial was generated at random by performing a push in a random direction that was not included in the dataset. This also helped exploring the space around the button more uniformly.

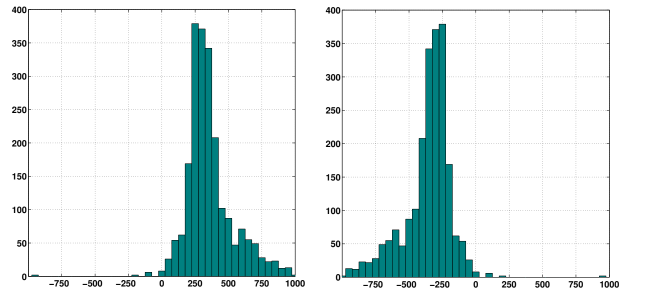
To represent the result of each behavior the robot used a set of outcomes $Y = \{buzzer, no\ buzzer\}$. The robot recorded a $buzzer$ outcome when an auditory event was observed during a behavior. Otherwise, the a $no\ buzzer$ outcome was recorded.

IV. Learning Methodology

A. Prediction

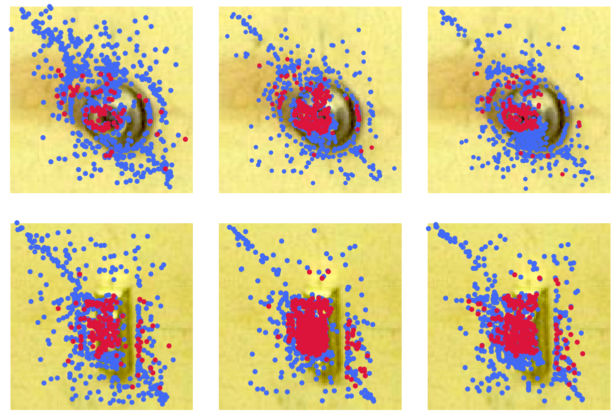
Over the course of interacting with a button, the robot's controller records labeled data points (x_i, y_i), indicating that outcome y_i was observed when performing an action parametrized by the vector x_i. Using this data, a predictive model M was incrementally trained to estimate the conditional probability $P(y|x)$ of observing outcome y given behavior parameters x.

A k-NN classifier with $k = 5$ was used to to implement the predictive model M that estimated the conditional probability distribution of outcomes given the behavior parameters. Given $x \in \mathbb{R}^6$ as input, the classifier estimated the conditional probability $\Pr(y|x)$ from the distribution of outcomes of the k nearest neighbours. For example, if 3 out of 5 nearest neighbours of x have the outcome $buzzer$, the probability estimate for $\Pr(buzzer|x)$ is $3/5$. Two key factors motivated the choice of this classifier. First, it can be updated incrementally as more training instances become available. Second, the k-NN classifier was among the best-performing classifiers for a similar task during a pilot study.



(a) Auditory-proprioceptive (b) Proprioceptive-auditory

Fig. 4. Histograms of the temporal intervals between auditory events and proprioceptive events (a) and vice versa (b). These were calculated over all 5000 pushing behaviors that the robot performed with button 1. The bin size for each histogram is 50 ms. The histograms show that the robot first hears the buzzer and then feels that the button is fully pressed.



(a) Random (b) Stimulus-Driven (c) Uncertainty-Driven

Fig. 5. Auditory (red) and proprioceptive (blue) events localized in space for each of the three learning strategies for buttons 3 and 7.

B. Exploration

Since the predictive model M is updated incrementally, the robot's controller can use it when choosing the parameters for the next pushing behavior. To do that, after each behavior execution, a candidate set $x_1, \ldots, x_N \in \mathbb{R}^6$ was generated, which represents a set of possible behaviors that the robot can perform. Three different exploration strategies that select the next behavior $x_j \in \{x_1, \ldots, x_N\}$ were evaluated.

1) Random Exploration: Under this exploration strategy, the robot always picks the first candidate in the set $\{x_1, \ldots, x_N\}$, i.e., $j = 1$. Since the candidate set of behaviors is generated randomly, this strategy results in a random exploration.

2) Stimulus-Driven Exploration: The second exploration strategy selected the behavior from the candidate set that maximizes the expected likelihood of observing a buzzing event, as estimated by the robot's predictive model M. In other words, a candidate behavior parametrization that is likely to push the button and cause a buzzing sound is favored over one that is not. Formally, this strategy selected the candidate behavior $x_j \in \{x_1, \ldots, x_N\}$ such that:

$$j = \underset{i=1,\ldots,N}{\operatorname{argmax}} \Pr(y_i = \mathit{buzzer} | x_i).$$

3) Uncertainty-Driven Exploration: The last exploration strategy selects the behaviors, for which the robot's model M is most uncertain regarding its outcome. For each candidate behavior x_i, the uncertainty is quantified using the entropy of the conditional distribution over the set of outcomes in Y. Formally, the uncertainty-driven strategy selects behavior x_j according to:

$$j = \underset{i=1,\ldots,N}{\operatorname{argmax}} \sum_{y \in Y} - \Pr(y_i = y | x_i) \log(\Pr(y_i = y | x_i)).$$

V. Results

A. Temporal Intervals between Modalities

For the trials in which the buzzer went off, the robot measured temporal intervals between the proprioceptive and the auditory events. Fig. 4 shows a histogram of these temporal delays. In general, the robot first heard the buzzer and then detected that it could not press the button any further. The average temporal delay was consistently around 250ms. This was true for all buttons as shown in Fig. 7(d). Some buttons were harder to press, e.g., buttons 3 and 4 for which the surface area of the functional part of the button was smaller than that for the other buttons. For these buttons there were fewer entries in the bins of their histograms.

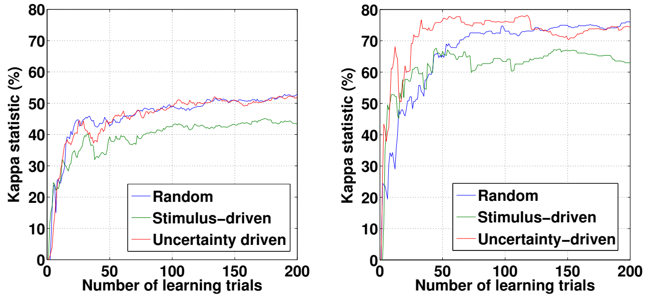
Fig. 4(b) is a mirror copy of Fig. 4(a), which means that an auditory event is almost always followed by a proprioceptive event. The opposite is not true (see Fig. 5).

B. Localizing Events in Space

The robot combined the timestamps of the video frames with the timestamps associated with auditory and proprioceptive events. In this way the tip of the robot's finger could be localized in visual space during interesting events. The spatial distribution of these events for all buttons is shown in Fig. 7(b) and 7(c) for the best performing exploration strategy. To highlight the difference in spatial distributions of events for different learning strategies, these distributions are shown for two of the buttons for all three strategies in Fig. 5. Both active learning strategies produced more auditory events than the random strategy for the same number of pushing behaviors.

C. Learning to Predict Auditory Events

The robot used the collected data to predict if a pushing behavior would trigger an auditory event. For each of the three learning strategies, and for each of the buttons, the number of available learning trials was varied to find out how



(a) Proprioceptive Space (b) Visual Space

Fig. 6. Learning curves for predicting if a an auditory event will be heard given: (a) the kinematic parameters of a pushing behavior; and (b) the location of the last proprioceptive event occurred in visual space. The results are for button 3.

much training is required to achieve a specific performance level. The predictions were evaluated on the 400 trials from the testing set.

1) Prediction in Proprioceptive Space: The robot trained the same model based on the k-NN classifier that was used by the learning strategies (see Section IV-B) to estimate the conditional probability distribution of outcomes given the behavior parameter vector $x \in \mathbb{R}^6$. The results of this evaluation are shown for button 3 in Fig. 6(a). The results for all buttons in Fig. 7(e). For all buttons the model trained for the uncertainty-driven exploitation strategy was either the top performer or one of the top performers. For button 4, each of the active learning strategies outperformed the random strategy.

2) Prediction in Visual Space: Fig. 4 shows that auditory events are shortly followed by proprioceptive ones. Therefore, the question arises: How much training does the robot need to learn to tell if the buzzer will sound if the button is pressed at a specific location in visual space?

To answer this question, a model based on the k-NN classifier with $k = 5$, similar to the predictive model used in Section IV-B and Section V-C.1, was trained to estimate the conditional probability distribution of behavioral outcomes given the 2D position of the robot's finger in visual space. The behaviors which produced no proprioceptive events were not considered for learning or evaluation.

The evaluation results are shown for one button in Fig. 6(b) and for all buttons in Fig. 7(f). For all buttons the uncertainty-driven active learning strategy performed the best, and for some of the buttons it was the best by a wide margin. Under the uncertainty-driven active learning strategy 100 trials were sufficient to achieve the best performance for each of the buttons.

VI. Conclusion and Future Work

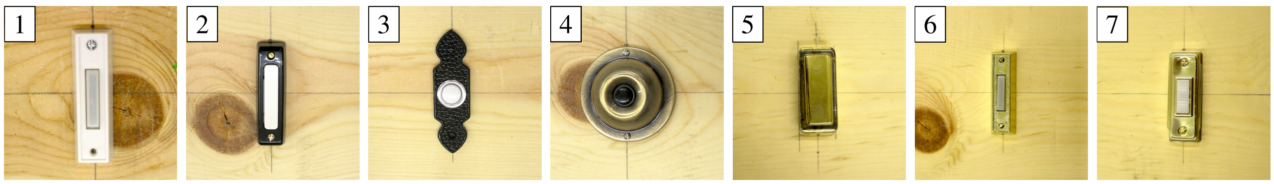
This paper evaluated several strategies that a robot can use to learn how to press buttons in real time. Three learning strategies were evaluated: random strategy, stimulus-driven exploitation strategy, and uncertainty-driven exploitation strategy. The random strategy always selected a random push. The stimulus-driven exploitation strategy selected a behavior that was most likely to press the button. The uncertainty-driven exploitation strategy selected the behavior with the most uncertain outcome.

The robot evaluated each of the three learning strategies on two tasks: 1) learning if a behavior would successfully press a button; and 2) learning if pressing the button at a specific spatial location would cause the buzzer to go off. The uncertainty-driven exploitation strategy performs better than each of the alternatives for both tasks.

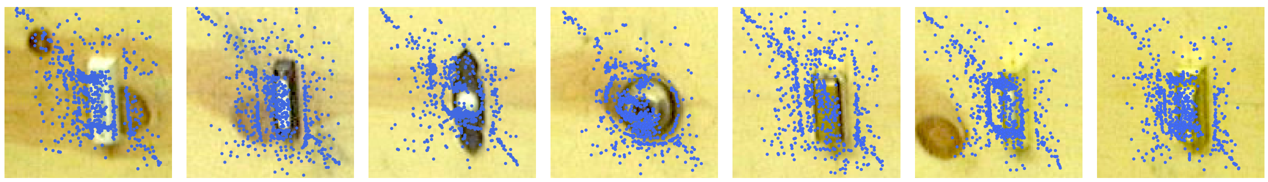
The framework already can localize auditory and proprioceptive events in visual space as the robot explores a button. Future work can establish a correlation between the visual features of the button and the spatial distributions of sensory events. The robot can use this information to segment this object into functional parts. The robot can use the extracted visual features to identify these functional parts in novel buttons and re-use the knowledge it obtained from earlier exploration. Modifying the methodology to work with light switches is yet another feasible direction for the future.

References

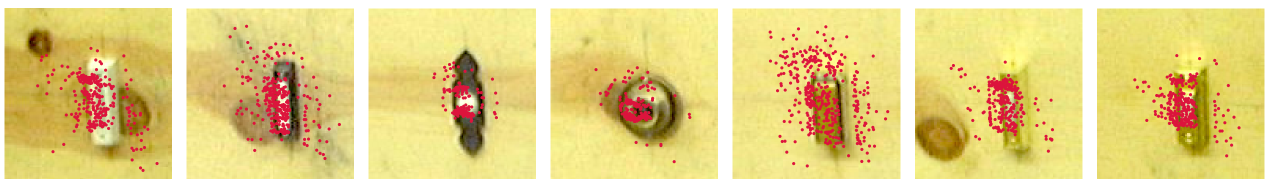
[1] P. Hauf and G. Aschersleben, "Action-effect anticipation in infant action control," *Psych. Research*, vol. 72, no. 2, pp. 203–210, 2008.
[2] C. Hutt and R. Bhavnani, "Predictions from play," *Nature*, vol. 237, no. 5351, pp. 171–172, 1972.
[3] J. Piaget and B. Inhelder, "The child's conception of space," 1956.
[4] E. Gibson, "Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge," *Annual review of psychology*, vol. 39, no. 1, pp. 1–42, 1988.
[5] L. Thomaz, "Socially guided machine learning," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.
[6] H. Nguyen, A. Jain, C. Anderson, and C. Kemp, "A clickable world: Behavior selection through pointing and context for mobile manipulation," in *Proceedings of IROS*, 2008, pp. 787 –793.
[7] H. Nguyen, T. Deyle, M. Reynolds, and C. Kemp, "PPS-tags: Physical, perceptual and semantic tags for autonomous mobile manipulation," in *Proceedings of the IROS Workshop on Semantic Perception for Mobile Manipulation*, 2009.
[8] J. Miura, K. Iwase, and Y. Shirai, "Interactive teaching of a mobile robot," in *Proceedings of ICRA*, 2005, pp. 3378–3383.
[9] K.-T. Song and T.-Z. Wu, "Visual servo control of a mobile manipulator using one-dimensional windows," in *Proceedings of Industrial Electronics Society*, vol. 2, 1999, pp. 686–691.
[10] A. Baranes and P.-Y. Oudeyer, "R-IAC: Robust intrinsically motivated active learning," in *Proceedings of ICDL*, 2009.
[11] A. G. Barto, "Intrinsically motivated learning of hierarchical collections of skills," in *Proceedings of ICDL*, 2004, pp. 112–119.
[12] J. Marshall, D. Blank, and L. Meeding, "An emergent framework for self-motivation in developmental robotics," in *Proceedings of ICDL*, 2004.
[13] M. Salganicoff, L. H. Ungar, and R. Bajcsy, *Machine Learning*, vol. 23, no. 2-3, pp. 251–278, 1996.
[14] L. Montesano and M. Lopes, "Learning grasping affordances from local visual descriptors," *Proceedings of ICDL*, pp. 1–6, 2009.
[15] A. Morales and E. Chinellato, "Active learning for robot manipulation," in *Proceedings of the European Conference on AI*, 2004, pp. 905–909.
[16] J. Cleary and L. Trigg, "K*: An instance-based learner using an entropic distance measure," in *Proceedings of the 12th International Conference on Machine learning*, 1995, pp. 108–114.



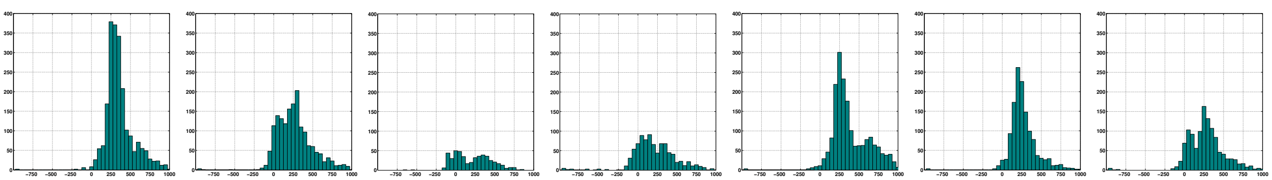
(a) The seven buttons explored by the robot.



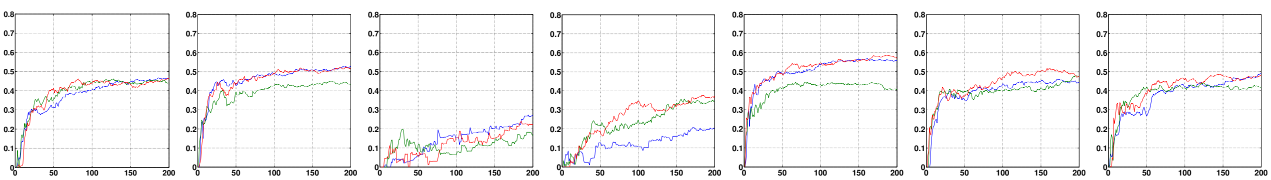
(b) Proprioceptive events localized in space for the uncertainty-driven exploitation strategy over 200 trials (1000 pushes).



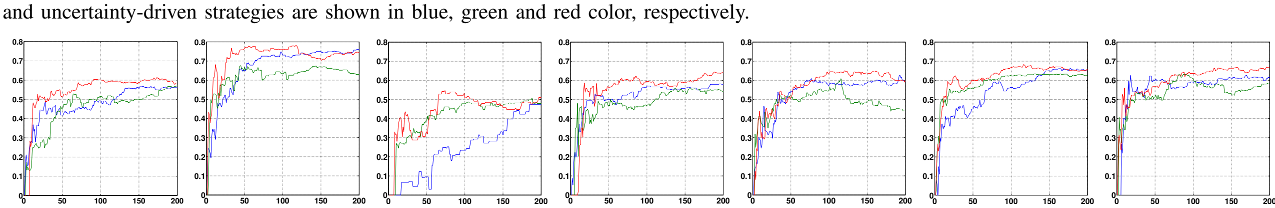
(c) Auditory events localized in space for each of the buttons using the uncertainty-driven strategy.



(d) Histograms of the temporal intervals between auditory and proprioceptive events for each button. The bin size in each histogram is 50 ms.



(e) Learning curves for predicting if a given pushing behavior triggers an auditory event as a function of the number of learning trials (x-axis). The performance measure is the kappa statistic estimated from a separate evaluation set of 400 trials. The curves for the random, stimulus-driven and uncertainty-driven strategies are shown in blue, green and red color, respectively.



(f) Learning curves for predicting if any behavior for which a proprioceptive event is detected in a specific region of visual space triggers an auditory event as a function of the number of learning trials on the (x-axis). The performance measure (y-axis) is the kappa statistic, which was estimated from a separate evaluation set of 400 trials. The curves for the random, stimulus-driven and uncertainty-driven strategy are shown in blue, green, and red color, respectively.

Fig. 7. Summary of the experimental results for all buttons.