This is a repository copy of *Active Bayesian perception and reinforcement learning*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/108458/

Version: Accepted Version

Proceedings Paper:
Lepora, N.F., Martinez-Hernandez, U., Pezzulo, G. et al. (1 more author) (2013) Active Bayesian perception and reinforcement learning. In: Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, November 3-8, 2013, Tokyo, Japan. IEEE , pp. 4735-4740. ISBN 978-1-4673-6358-7

https://doi.org/10.1109/IROS.2013.6697038

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Active Bayesian perception and reinforcement learning

Nathan F. Lepora, Uriel Martinez-Hernandez, Giovanni Pezzulo, Tony J. Prescott

***Abstract*— In a series of papers, we have formalized an *active Bayesian perception* approach for robotics based on recent progress in understanding animal perception. However, an issue for applied robot perception is how to tune this method to a task, using: (i) a belief threshold that adjusts the speed-accuracy tradeoff; and (ii) an active control strategy that moves the sensor during perception. Here we propose that this tuning should be learnt by reinforcement from a reward signal evaluating the decision outcome. We test this claim with a biomimetic fingertip that senses surface curvature under uncertainty about its contact position. Appropriate formulation of the problem allows application of multi-armed bandit methods to optimize the threshold and fixation point of the active perception. In consequence, the system learns to balance speed versus accuracy while also tuning the fixation point to optimize both quantities. Although we consider one example in robot touch, we expect that the underlying principles have general applicability.**

I. INTRODUCTION

A main principle underlying animal perception is the accumulation of evidence for multiple perceptual alternatives until reaching a preset belief threshold that triggers a decision [1], [2], formally related to sequential analysis methods for optimal decision making [3]. In a series of papers [4], [5], [6], [7], [8], we have formalized a *Bayesian perception* approach for robotics based on this recent progress in understanding animal perception. Our formalism extends naturally to active perception, by moving the sensor with a control strategy based on evidence received during decision making. Benefits of active Bayesian perception include: (i) robust perception in unstructured environments [8]; (ii) an order-of-magnitude improvement in acuity over passive methods [9]; and (iii) a general framework for simultaneous object localization and identification, or 'where' and 'what' [9].

This work examines a key issue for applying active Bayesian perception to practical scenarios: how to choose the parameters for the optimal decision making and active perception strategy. Thus far, the belief threshold has been treated as a free parameter that adjusts the balance between mean errors and reaction times (*e.g.* [7, Fig. 5]). Meanwhile, the active control strategy has been hand-tuned to fixate to a region with good perceptual acuity [8], [9]. Here we propose that these free parameters should be learnt by reinforcement from a reward signal that evaluates the decision outcome, and demonstrate this method with a task in robot touch.

Past work on reinforcement learning and active perception has been confined to active vision, and was motivated initially by the *perceptual aliasing* problem for agents with limited sensory information [10], [11], [12]. Later studies shifted emphasis to optimizing perception, such as learning good viewpoints [13], [14], [15]. Just one paper has considered active (not reinforcement) learning to optimize active touch [16]. There has also been interest in applying reinforcement learning to visual attention [17], [18], [19]. We know of no work on learning an optimal decision making threshold and active control strategy, by reinforcement or otherwise.

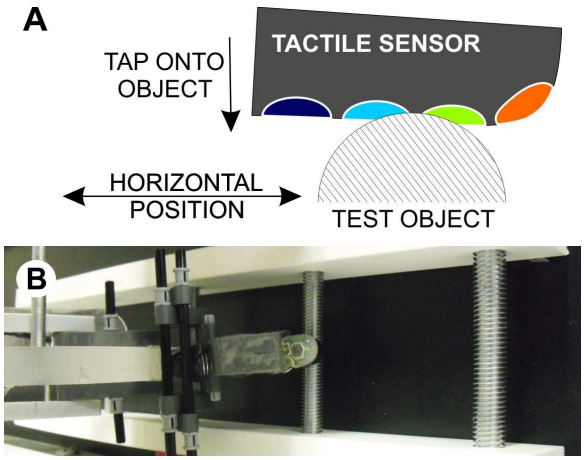

Fig. 1. Experimental setup. (A) Schematic of tactile sensor contacting a cylindrical test object. The fingertip is moved horizontally to sample object contacts from different positions. (B) Top-down view of experiment, with the fingertip mounting on the arm of the Cartesian robot visible to the left.

Our proposal for active Bayesian perception and reinforcement learning is tested with a simple but illustrative task of perceiving object curvature using tapping movements of a biomimetic fingertip with unknown contact location (Fig. 1). We demonstrate first that active perception with fixation point control strategy can give robust and accurate perception, but the reaction time and acuity depend strongly on the choice of fixation point and belief threshold. Next, we introduce a reward function of the decision outcome, which for illustration is taken as a linear Bayes risk of reaction time and error. Interpreting each active perception strategy (parameterized by the decision threshold and fixation point) as an action, then allows use of standard reinforcement learning methods for multi-armed bandits [20]. In consequence, the appropriate decision threshold is learnt to balance the risk of making mistakes versus the risk of reacting too slowly, while the fixation point is tuned to optimize both quantities.

Although we consider one example in robot touch, we expect that the underlying principles are sufficiently general to be applicable across a range of other percepts and modalities.

This work was supported by EU Framework projects EFAA (ICT-270490) and GOAL-LEADERS (ICT-270108) and also by CONACyT (UMH).

NL, UMH and TP are with SCentRo, University of Sheffield, UK. Email: {n.lepora, uriel.martinez, t.j.prescott}@sheffield.ac.uk

GP is with the ILC, Pisa and ISTC, Roma, Consiglio Nazionale delle Ricerche (CNR), Italy. Email: giovanni.pezzulo@cnr.it

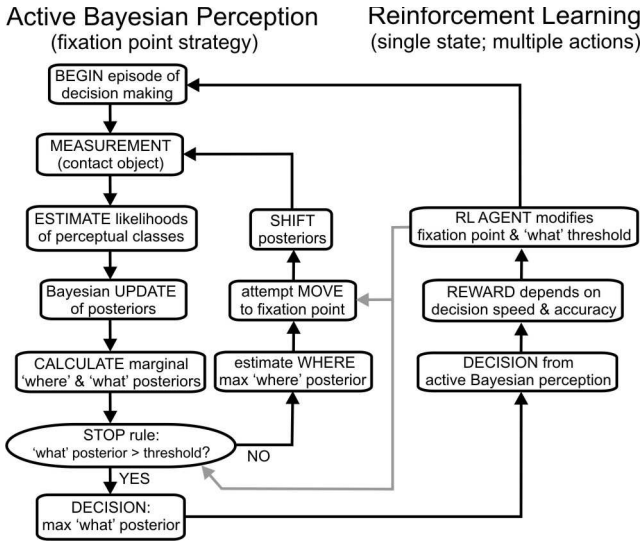

Fig. 2. Algorithm for active Bayesian perception with reinforcement learning. Active Bayesian perception (left) has a recursive Bayesian update to give the marginal 'where' and 'what' posteriors, allowing active control of the sensor position and decision termination at sufficient 'what' belief. Reinforcement learning (right) modifies the decision threshold and active control strategy based on reward information derived from the decisions.

II. METHODS

A. Active Bayesian Perception with Reinforcement Learning

Our algorithm for active perception is based on including a sensorimotor feedback loop in an optimal decision making method for passive perception derived from Bayesian sequential analysis [4]. Sequential analysis uses a free parameter, the decision threshold, to adjust the speed-accuracy tradeoff of the decisions. Our control strategy for active perception also has another free parameter, the fixation point. We thus introduce reinforcement learning to set these two free parameters according to a reward function of the speed and accuracy of the decision outcome.

Measurement model and likelihood estimation: Each tap against a test object gives a multi-dimensional time series of sensor values across the K taxels (Fig. 3). The likelihood of a perceptual class $c_n \in C$ for a test tap z_t (with samples s_j) is evaluated with a measurement model [4], [5]

$$P(z_t|c_n) = \sqrt[JK]{\prod_{j=1}^{J}\prod_{k=1}^{K} P_k(s_j|c_n)}. \quad (1)$$

The sample distribution is determined off-line from the training data using a 'bag-of-samples' histogram method

$$P_k(s|c_n) = \frac{h_k(b(s))}{\sum_b h_k(b)}, \quad (2)$$

with $h_k(b)$ the occupation number of a bin (and $b(s) \ni s$), taking 100 bins across the full range of sensor data. Here we have $K = 12$ taxels and $J = 50$ time samples in each tap.

Bayesian update: Bayes' rule is used to recursively update the beliefs $P(c_n|z_t)$ for the N perceptual classes c_n with likelihoods $P(z_t|c_n)$ of the present tap z_t

$$P(c_n|z_t) = \frac{P(z_t|c_n)P(c_n|z_{t-1})}{P(z_t|z_{t-1})}. \quad (3)$$

The likelihoods $P(z_t|c_n)$ are assumed i.i.d. over time t (so $z_{1:t-1}$ drops out). The marginal probabilities are conditioned on the preceding tap and calculated by summing

$$P(z_t|z_{t-1}) = \sum_{n=1}^{N} P(z_t|c_n)P(c_n|z_{t-1}). \quad (4)$$

Iterating the update (3,4), a sequence of taps $z_1, \cdots, z_t$ gives a sequence of posteriors $P(c_n|z_1), \cdots, P(c_n|z_t)$ initialized from uniform priors $P(c_n) = P(c_n|z_0) = 1/N$. Here we use $N = 80$ classes over 16 positions and 5 object curvatures.

Marginal 'where' and 'what' posteriors: The perceptual classes have L 'where' (position) and M 'what' (curvature) components, with each class c_n an (x_l, w_m) 'where-what' pair (*i.e.* $C = X \times W$). Then the beliefs over the individual 'where' and 'what' classes are found by marginalizing

$$P(x_l|z_t) = \sum_{m=1}^{M} P(x_l, w_m|z_t), \quad (5)$$

$$P(w_m|z_t) = \sum_{l=1}^{L} P(x_l, w_m|z_t), \quad (6)$$

with the 'where' beliefs summed over all 'what' classes and the 'what' beliefs over all 'where' perceptual classes. Here we use $L = 16$ position classes and $M = 5$ curvature classes.

Stopping condition on the 'what' posteriors: Following methods for passive Bayesian perception using sequential analysis [4], a threshold crossing rule on the marginal 'what' posterior triggers the final 'what' decision, given by the maximal *a posteriori* (MAP) estimate

$$\text{if any } P(w_m|z_t) > \theta_W \text{ then } w_{\text{MAP}} = \arg\max_{w_m \in W} P(W|z_t). \quad (7)$$

This belief threshold θ_W is a free parameter that adjusts the balance between decision speed and accuracy.

Active perception with the 'where' posteriors: Here we consider a control strategy with fixation point x_{fixed} that the sensor attempts to move to. Then the appropriate move Δ is

$$x \to x + \Delta\left(x_{\text{MAP}}\right), \quad \Delta(x_{\text{MAP}}) = x_{\text{fixed}} - x_{\text{MAP}}, \quad (8)$$

with x_{MAP} the 'where' decision of sensor location determined after every test tap

$$x_{\text{MAP}} = \arg\max_{x_l \in X} P(X|z_t). \quad (9)$$

The 'where' posteriors should be kept aligned with the sensor by shifting the joint 'where-what' posteriors with each move

$$P(x_l, w_m|z_t) = P(x_l - \Delta, w_m|z_t). \quad (10)$$

For simplicity, we recalculate the posteriors lying outside the original range by assuming they are uniformly distributed.

Reinforcement learning: The active perception strategy is defined by two free parameters, the decision threshold θ_W and fixation point x_{fixed}, to be learnt by reinforcement. Each learning episode i is a perceptual decision with reaction time T_i and error e_i, with an ensuing scalar reward signal $r(T, e)$ taken here as the negative Bayes risk

$$r_i = -\alpha T_i - \beta e_i, \quad (11)$$

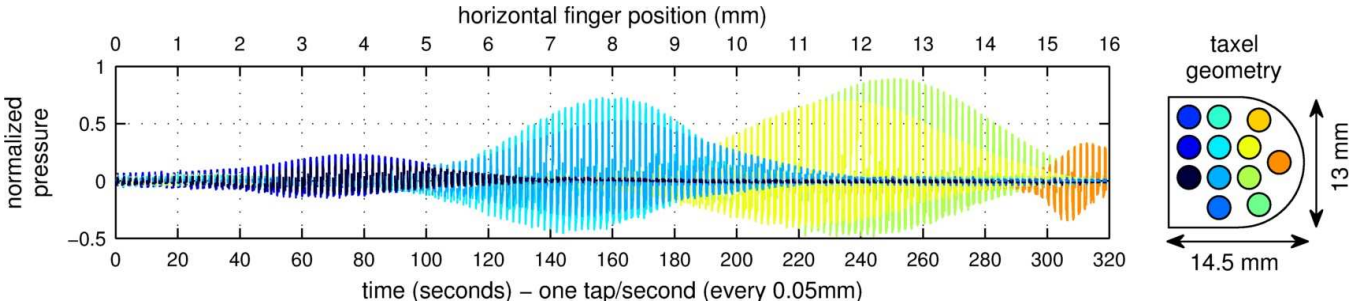

Fig. 3. Fingertip pressure data recorded as the finger taps against a test rod (diameter 4 mm) at a constant rate of 1 tap/sec. The range of finger positions spanned 16 mm over 320 s, giving 320 taps spaced every 0.05 mm. Tickmarks are shown every 1 mm displacement, or 20 taps. Data from the different taxels are represented in distinct colors depending on the taxel position shown on the diagram to the right.

where α, β are positive coefficients that parameterize the riskiness of increasing reaction times and errors. Note that only the relative value α/β is important, because we aim to learn the optimal speed-accuracy tradeoff.

Standard techniques from reinforcement learning can be used to learn the active perception strategy that maximizes reward. If each strategy $(\theta_W, x_{\text{fixed}})$ is considered an action, then the problem is equivalent to a multi-armed bandit. Discretizing the decision threshold $\theta_W \in \{\theta(1), \cdots, \theta(D)\}$ and noting the L 'where' classes are already discrete, allows the use of standard methods for balancing reward exploration versus exploitation (see *e.g.* [20, ch. 2]). Here we have $D = 13$ and $L = 16$, giving 208 distinct actions. For simplicity, we keep a running average of the rewards for each action $a = (\theta(d), x_l)$, using an incremental update

$$Q_a \leftarrow Q_a + \frac{1}{i_a + 1}\left(r_i - Q_a\right), \tag{12}$$

with i_a the number of times that action has been performed. Exploration is achieved with initially optimistic Q_a and exploitation by choosing the action with maximal Q_a.

B. Tactile data collection

The tactile sensors have a rounded shape that resembles a human fingertip [21], of dimensions 14.5 mm long by 13 mm wide. They consist of an inner support wrapped with a flexible printed circuit board (PCB) containing 12 conductive patches for the touch sensor 'taxels'. These are coated with PCB and silicone layers that together comprise a capacitive touch sensor to detect pressure via compression. Data was collected at 50 samples per second with 256 vales, and then normalized and high-pass filtered before analysis [21].

The present data were collected for a previous study [7], and have direct relevance to the work presented here. These experiments were designed to test the capabilities of the tactile fingertip sensor mounted on an xy-positioning robot. This robot can move the sensor in a highly controlled and repeatable manner onto various test stimuli ($\sim$50 μm accuracy), and has been used for testing various tactile sensors [22]. The fingertip was mounted at an angle appropriate for contacting axially symmetric shapes such as cylinders aligned perpendicular to the plane of movement (Fig. 2). Five smooth steel cylinders with diameters 4 mm, 6 mm, 8 mm, 10 mm and 12 mm were used as test objects: they were mounted with their centers offset to align their closest point to the fingertip in the direction of tapping.

Data were collected while having the fingertip tap periodically along the vertical y-axis onto and off each test object with contact duration 0.5 secs and 1 sec of each tap saved for analysis. The cylinder axis lay across the fingertip (down the taxels in Fig. 3). Between each tap, the fingertip was displaced 0.05 mm in the horizontal x-direction across the face of the cylinder. Altogether, 320 horizontal displacements spanning 16 mm were used for each cylinder, comprising 1600 taps in total. Distinct training and test sets were collected for all 5 cylinder diameters.

III. RESULTS

A. Simultaneous object localization and identification

As observed previously [7], the pattern of taxel pressures from each tap against a test object (here a cylinder) depended on both the surface curvature and the horizontal position of the fingertip relative to the object (Fig. 3), permitting simultaneous object localization and identification. As the fingertip moved across its horizontal range, the taxels were activated initially at its base (dark blue; Fig. 3), then its middle (light blue to green) and finally its tip (red). An important aspect of the taxel activation is the broad ($\sim$8 mm) receptive fields: the overlap between these fields enables perceptual hyperacuity, whereby finger position may be localized more finely than the taxel resolution ($\sim$4 mm spacing) [7]. This hyperacuity will be apparent in the following results.

Previous work has considered passive Bayesian perception with this dataset [7]. Passive Bayesian perception accumulates belief for distinct 'where' (horizontal position) and 'what' (curvature) classes by making successive taps against the test object until at least one of the marginal 'what' posteriors crosses a belief threshold, when a 'where' (localization) and 'what' (identification) decision is made. For the present dataset, the best passive perceptual acuity was $\sim$2 mm for cylinder diameter (4-12 mm range), which is the primary decision considered in following sections. The horizontal position was localized to $\sim$0.6 mm (16 mm range), demonstrating hyperacuity at $\sim$15% of the sensor resolution.

In a previous study, we showed that perceptual acuity was improved with an active perception strategy that moves the

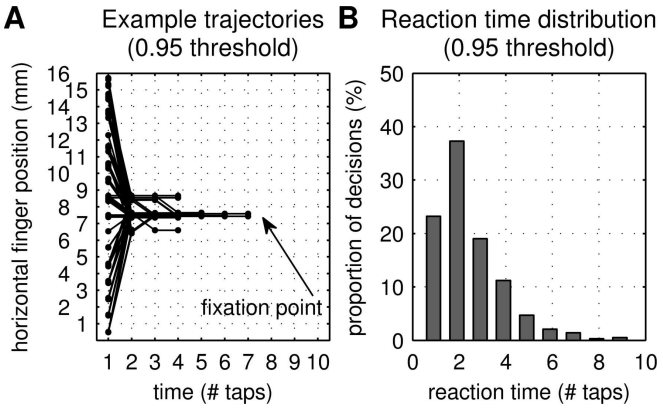

Fig. 4. Active perception behavior. (A) Trajectories converge on the fixation point (8 mm) independent of starting position. (B) Reaction times have a positively skewed distribution, as commonly seen in psychology.

sensor randomly after a fixed deadline [7]. The present study uses an improved method for active perception based on orienting the sensor to a fixation point, which has been shown to both improve acuity over passive perception and enable robust perception in unstructured environments [8], [9].

B. Active Bayesian perception

Active Bayesian perception also accumulates belief for the 'where' (horizontal position) and 'what' (cylinder diameter) perceptual classes by successively tapping against a test object until reaching a predefined 'what' belief threshold. In addition, it utilizes a sensorimotor loop to move the sensor according to the online marginal belief estimates during the perceptual process (Fig. 2; left loop). The active perception method considered here uses a 'fixation point' control strategy, such that the marginal 'where' beliefs are used to infer a best estimate for current location and thus a relative move towards a preset target position on the object.

For the present data set of several cylinder diameters over a range of horizontal positions, the typical behavior for active Bayesian perception with fixation point strategy has the sensor orienting quickly to the fixation position within a few taps independent of starting placement (Fig. 4A; example fixation point at 8 mm; decision threshold 0.95). The reaction times to reach the belief threshold have a positively skewed distribution (Fig. 4B) reminiscent of that obtained from behavioral/psychological experiments with humans and animals. Note that active Bayesian perception leads to greatly improved mean reaction times and perceptual acuity compared with passive methods for estimating cylinder diameter (*cf.* results in [7]). For the decisions shown in Fig. 4, the mean absolute error was $\sim$0.7 mm, much better than $\sim$2 mm for passive perception. Note also that active perception has an added benefit of aligning the sensor onto the same point of the object whatever the relative initial positioning, in effect compensating for an unstructured environment.

The decision accuracy and reaction times for active Bayesian perception depended strongly on both the belief threshold and fixation point (Fig. 5; threshold indicated by gray shade of plot, fixation point on x-axis). As the

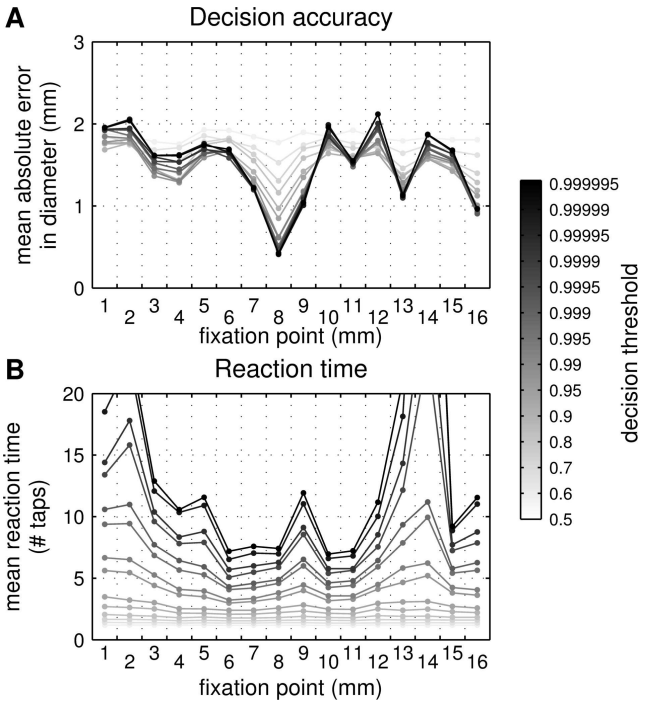

Fig. 5. Active perception results depends on the decision threshold and fixation point. The mean accuracy of identifying the cylinder (A) and the mean reaction time (B) vary with threshold (gray-shade of plot) and fixation point (x-axis). Each data point corresponds to 10000 decision episodes.

belief threshold is raised (darker gray plots), more evidence is required to make a decision, which results in lower errors of perceiving cylinder diameter and longer reaction times. The choice of fixation point is also important for perception, with the central region of the horizontal range giving lower errors and reaction times compared with those at the extreme positions. This dependence on fixation point is due to the physical properties (morphology) of the tactile sensor coupled with shape and dynamics of the perceived object. Thus, central contacts of the fingertip activate more taxels and have improved reliability, in contrast to glancing contacts at its base or tip (Fig. 3). In consequence, errors improved from $\sim$2 mm for fixation at the base or tip, down to $\lesssim$1 mm at the center (Fig. 5; decision thresholds $\gtrsim$0.95).

Examining Fig. 5 by eye, reveals that an active perception strategy with central fixation point gives the finest perceptual acuity and quickest reaction times. However, the plots in Fig. 5 were obtained by 'brute force' over millions of validation episodes. This raises the question of how optimal active perception should be determined in practice from a manageably small number of decision episodes.

C. Active Bayesian perception with reinforcement learning

The main theme of this paper is that the parameters controlling active perception (here the decision threshold and fixation point) should be learnt by reinforcement using a reward function that evaluates the decision outcome (Fig. 2).

For simplicity, we use an example reward function given by (minus) the linear Bayes risk of reaction time and absolute decision error (Eq. 11). Although the proposed approach

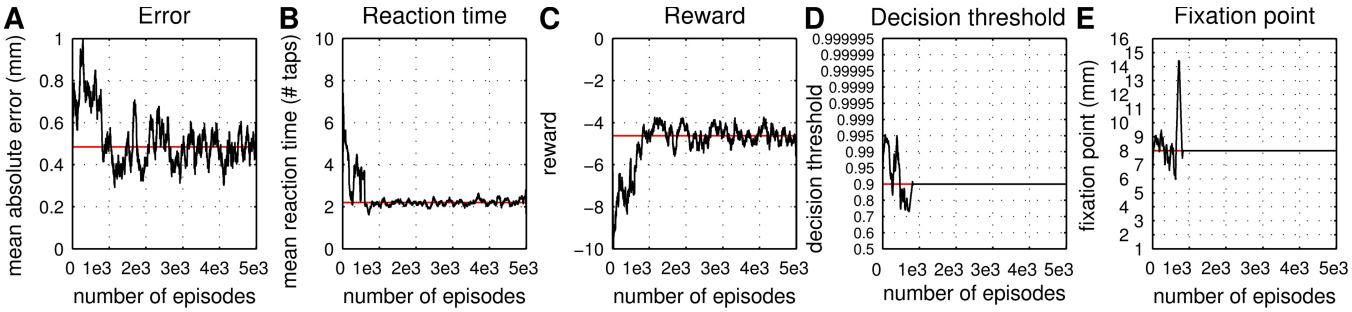

Fig. 6. Example run of reinforcement learning to optimize the active perception strategy. Change in decision error (A) and reaction time (B) as the belief threshold (D) and fixation point (E) are learnt to optimize mean reward (C). Target values from brute-force optimization of the reward function are shown in red. All plots are smoothed over 100 episodes. Results are for risk parameter $\alpha/\beta = 0.2$ and initial optimistic reward estimates of 100.

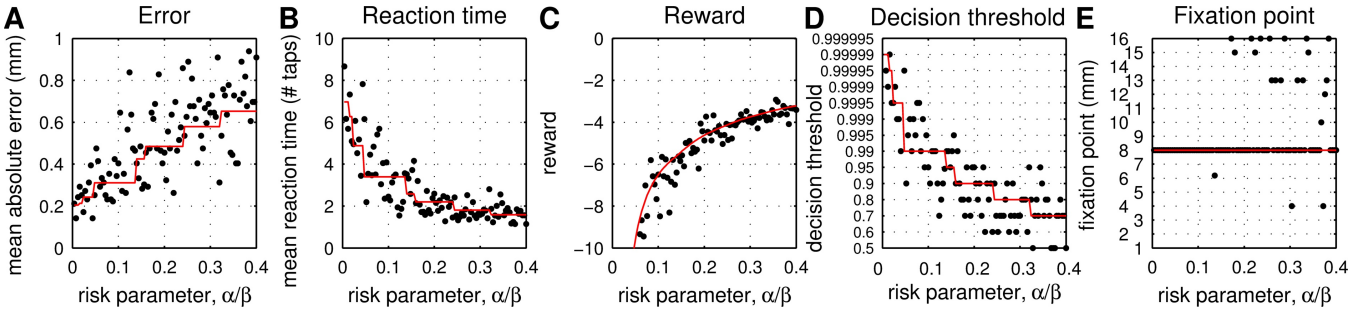

Fig. 7. Dependence of optimal active perception strategy on Bayes risk parameter. The final error (A), reaction time (B), reward (C), decision threshold (D) and fixation point (E) are shown after 5000 reinforcement learning episodes, for 100 risk parameters ranging from 0 to 0.4. The risk parameter describes the relative reward benefits of improving speed versus accuracy (Eq. 11). Target values from brute-force optimization of the reward function are in red.

should be independent of reward function, considering the Bayes risk gives a simple interpretation of the problem in terms of minimizing the relative risks of taking too long to reach a decision versus making errors. Then the resulting speed-accuracy tradeoff depends only upon the ratio α/β of the two coefficients in the Bayes risk, with a smaller ratio placing more risk on the decision error and a larger ratio on the reaction time. Maximizing reward minimizes this risk.

In this work, each combination of the threshold and fixation point define a distinct active perception strategy, with the decision thresholds taking the discrete values in Fig. 5. If the optimal strategy is to be learnt by reinforcement over many episodes, each active perception strategy may be considered a distinct action, and there is just one state. The overall situation therefore reduces to a standard multi-armed bandit problem. In consequence, the optimal active perception strategy can be learnt efficiently using standard methods for balancing exploration versus exploitation (*e.g.* those from [20, ch. 2]). In practice, all such methods that we tried converged well for appropriate learning parameters, hence we simplify our explanation by considering only a greedy method with incrementally updated reward estimates from optimistic initial values (Eq. 12).

For a typical instance of reinforcement learning and active perception, the active control strategy converged to nearly optimal perception over $\sim 10^3$ decision episodes (Fig. 6; $\alpha/\beta = 0.2$). In particular, the decision threshold and fixation point converged close to their optimal values (Figs 6D,E; red lines) found with brute force optimization of the reward function (validated over $\sim 10^7$ episodes). The fixation point converged to the center of the range, consistent with the brute-force results in Fig. 5, while the decision threshold converged to a suitable value to balance mean reaction times and errors. Accordingly, the mean decision error and reaction time approached their optimal values with noise due to the stochastic decision making (Figs 6A,B), while reward also increased stochastically to around its optimal value (Fig. 6C).

For many instances of reinforcement learning and active Bayesian perception, the active control strategy converged to nearly optimal perception over a range of risk parameters (Fig. 7; $0 < \alpha/\beta < 0.4$). This risk parameter represents the relative risk of delaying the decision (α) versus making an error (β). All parameters, including the decision threshold, fixation point, rewards, decision error and reaction time reached values near to optimal after 5000 episodes (Fig. 7; red plots, validation with $\sim 10^7$ episodes) over range of risk parameters giving a broad span of speed-accuracy tradeoffs.

Therefore, reinforcement learning and active perception combine naturally to give a robust method for achieving optimal perception. The converged parameters values controlling active perception depend on the relative risk of speed versus accuracy. Shifting the balance of risk towards accuracy (smaller α/β), results in larger decision thresholds and longer reaction times, while the converse occurs with placing the risk in speed (larger α/β). Concurrently, the fixation point is tuned to optimize both quantities, and converges to the central position apart from very brief decisions when the active perception strategy becomes irrelevant (for large α/β).

IV. Discussion

In this paper, we proposed an algorithm that combines active Bayesian perception with reinforcement learning and tested the method with a task in robot touch, which was to perceive object curvature using tapping motions of a biomimetic fingertip from unknown initial contact location. Active perception with fixation point control strategy gave robust and accurate perception, although the reaction time and acuity depended strongly on the choice of fixation point and belief threshold. Introducing a reward function based on the Bayes risk of the decision outcome and considering each combination of threshold and fixation point as an action, allowed use of standard reinforcement learning methods for multi-armed bandits. The system could then learn the appropriate belief threshold to balance the risk of making mistakes versus the risk of reacting too slowly, while tuning the fixation point to optimize both quantities.

These results demonstrate that optimal robot behavior for a perceptual task can be tuned by appropriate choice of reward function. Here we used a linear Bayes risk of decision error and reaction time to give a simple demonstration over a range of robot behaviors, parameterized just by the relative risk of speed versus accuracy. The system then learned to make quick but inaccurate decisions when reaction time was risky compared with errors, and accurate but slow decisions when errors were risky compared with reaction times. We emphasize that the general approach does not depend on the specifics of the reward function, with the actual choice representing the task aims and goals. Imagine, for example, a production line of objects passing a picker that must remove one class of object: if the robot takes too long, then objects pass it by, and if it makes mistakes, then it picks the wrong objects; both of these outcomes can be evaluated and used to reward or penalize the robot to optimize its behavior.

A key step in our combination of active perception and reinforcement learning was to interpret each active perception strategy (parameterized by the threshold and fixation point) as an action. We could thus employ standard techniques for multi-armed bandits [20], which generally worked well, and for reasons of simplicity and pedagogy we used a greedy method with optimistic initial values. Although it is beyond the scope of this paper, we expect that efficient use of the reward structure could significantly reduce exploration and hence regret (reward lost while not exploiting). For example, the reward is generally convex in the decision threshold, which could be used to constrain the value estimates.

In future work, we will study scaling our method to the many degrees of freedom necessary for practical purposes in robotics. Optimal active Bayesian perception via reinforcement could then give a general approach to robust and effective robot perception.

Acknowledgements: We thank Kevin Gurney, Ashvin Shah and Alberto Testolin for discussions, and the organizers of the 2012 FIAS school on Intrinsic Motivations for hosting NL and GP while some of this work was carried out.

References

[1] J.I. Gold and M.N. Shadlen. The neural basis of decision making. *Annual Reviews Neuroscience*, 30:535–574, 2007.

[2] R. Bogacz and K. Gurney. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural computation*, 19(2):442–477, 2007.

[3] A. Wald. *Sequential analysis*. John Wiley and Sons (NY), 1947.

[4] N.F. Lepora, C.W. Fox, M.H. Evans, M.E. Diamond, K. Gurney, and T.J. Prescott. Optimal decision-making in mammals: insights from a robot study of rodent texture discrimination. *Journal of The Royal Society Interface*, 9(72):1517–1528, 2012.

[5] N.F. Lepora, M. Evans, C.W. Fox, M.E. Diamond, K. Gurney, and T.J. Prescott. Naive bayes texture classification applied to whisker data from a moving robot. *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–8, 2010.

[6] N.F. Lepora, J.C. Sullivan, B. Mitchinson, M. Pearson, K. Gurney, and T.J. Prescott. Brain-inspired bayesian perception for biomimetic robot touch. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 5111 –5116, 2012.

[7] N.F. Lepora, U. Martinez-Hernandez, H. Barron-Gonzalez, M. Evans, G. Metta, and T.J. Prescott. Embodied hyperacuity from bayesian perception: Shape and position discrimination with an icub fingertip sensor. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 4638–4643, 2012.

[8] N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active touch for robust perception under position uncertainty. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, 2013.

[9] N.F. Lepora, U. Martinez-Hernandez, and T.J. Prescott. Active bayesian perception for simultaneous object localization and identification. (under review).

[10] S.D. Whitehead and D.H. Ballard. Active perception and reinforcement learning. *Neural Computation*, 2(4):409–419, 1990.

[11] S.D. Whitehead and D.H. Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83, 1991.

[12] L. Chrisman. Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proceedings of the National Conference on Artificial Intelligence*, pages 183–188, 1992.

[13] J. Peng and B. Bhanu. Closed-loop object recognition using reinforcement learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(2):139–154, 1998.

[14] L. Paletta and A. Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1):71–86, 2000.

[15] F. Deinzer, C. Derichs, H. Niemann, and J. Denzler. A framework for actively selecting viewpoints in object recognition. *Int Journal of Pattern Recognition and Artificial Intelligence*, 23(04):765–799, 2009.

[16] H. Saal, J. Ting, and S. Vijayakumar. Active estimation of object dynamics parameters with tactile sensors. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 916–921, 2010.

[17] S. Minut and S. Mahadevan. A reinforcement learning model of selective visual attention. In *Proceedings of the fifth international conference on Autonomous agents*, pages 457–464, 2001.

[18] D. Ognibene, F. Mannella, G. Pezzulo, and G. Baldassarre. Integrating reinforcement-learning, accumulator models, and motor-primitives to study action selection and reaching in monkeys. In *7th International Conference on Cognitive Modelling-ICCM06*, pages 214–219, 2006.

[19] A. Borji, M. Ahmadabadi, B. Araabi, and M. Hamidi. Online learning of task-driven object-based visual attention control. *Image and Vision Computing*, 28(7):1130–1145, 2010.

[20] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. Cambridge University Press, 1998.

[21] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta. Methods and technologies for the implementation of large-scale robot tactile sensors. *Robotics, IEEE Transactions on*, 27(3):389 –400, 2011.

[22] M. Evans, C.W. Fox, M. Pearson, N.F. Lepora, and T.J. Prescott. Whisker-object contact speed affects radial distance estimation. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 720–725, 2010.