Active Bayesian Perception for Simultaneous Object Localization and Identification

Nathan F. Lepora, Uriel Martinez-Hernandez and Tony J. Prescott
Dapartment of Psychology and Sheffield Center for Robotics (SCentRo), University of Sheffield, U.K.
Email: n.lepora@sheffield.ac.uk

***Abstract*—In this paper, we propose that active Bayesian perception has a general role for Simultaneous Object Localization and IDentification (SOLID), or deciding where and what. We test this claim using a biomimetic fingertip to perceive object identity via surface shape at uncertain contact locations. Our method for active Bayesian perception combines decision making by threshold crossing of the posterior belief with a sensorimotor loop that actively controls sensor location based on those beliefs. Our findings include: (i) active perception with a fixation control strategy gives an order-of-magnitude improvement in acuity over passive perception without sensorimotor feedback; (ii) perceptual acuity improves as the active control requires less belief to make a relocation decision; and (iii) relocation noise further improves acuity. The best method has aspects that resemble animal perception, supporting wide applicability of these findings.**

I. Introduction

We do not only see, we look. We not only hear, we listen. We do not just touch, we feel [1]. Our senses are not merely passive receivers of information, but actively select and refine sensations according to our present goals and perceptions [5]. Our bodies are not external from the world, but direct actions within it to access the information that we need [2, 15]. Thus, sensation, perception and action cannot be considered simply as a forward process, but instead form a closed 'active perception' loop with a task-dependent motor control strategy.

The main goal of this work is to advance our understanding of the role of active perception in determining the 'where' and 'what' properties of objects. In this sense, we interpret the purpose of finding 'what' as finding the action possibilities that an object class affords. The purpose of finding 'where' is to enable the agent to evaluate and enact the action possibilities for achieving those affordances. We refer to the computational task as Simultaneous Object Localization and IDentification (SOLID), to emphasize a similarity with SLAM of having two interdependent task aims, in that knowledge of location aids computation of identity (mapping) and identification (mapping) aids localization. Another similarity is to use a recursive Bayesian update for processing imperfect sensory evidence under uncertainty. Algorithmically, however, we use an approach for SOLID with historical roots in sequential analysis and optimal decision making (rather than Kalman filtering). Specifically, we introduce an algorithm called active Bayesian perception that combines 'where' and 'what' decision making with a sensorimotor feedback loop to control sensor location.

To explore these ideas, we examine a simple but illustrative task of perceiving 2D position (localization) and object curvature (identification) with tapping motions of a biomimetic fingertip. Although this study does not explicitly utilize the object affordances, an example of how it could is if shape were to define the object's utility for a task (*e.g.* fitting onto another object) and position prescribed where to get it.

Our main finding is that active Bayesian perception with a 'fixation point' control strategy gives a far more efficient and accurate way of solving this localization and identification task than passive Bayesian perception with no sensorimotor loop. Moreover, not all active perception strategies are equal. We show that the 'where' and 'what' perceptual acuities improve as relocation decisions require less belief and also improve with some relocation noise. These results demonstrate how to best apply active perception to a robot touch task, and in addition we expect that the findings are representative of other 'where' and 'what' tasks and other sensory modalities.

II. Related work

In a landmark paper, Bajcsy defined active sensing as [1]: 'purposefully changing the sensor's state parameters according to sensing strategies ... [that] depend on the current state of the data interpretation and the goal or the task.' She also emphasized that feedback should regulate sensor movement via ongoing decisions of where to move the sensor. While Bajcsy emphasized online optimization of active control, another possibility is that useful control strategies could be predefined for specific tasks (or in biological systems selected by learning or evolution). Thus, when provided with appropriate feedback signals, they could operate to boost the acquisition of useful information without the need for complex computation of the optimal strategy on a moment-by-moment basis [15].

The present work is further motivated by recent progress in the neuroscience of human and animal perception over imperfect sensor information. Leading computational accounts involve the sequential accumulation of evidence to threshold, consistent with numerous psychological and electrophysiological experiments [6]. An important aspect of the accumulation to threshold mechanism is its formal relation to sequential analysis methods for optimal decision making, leading to an optimal tradeoff between costs of delaying decisions and making mistakes [18]. Work in computational neuroscience also indicates these principles may relate to the macro-architecture of the brain, in particular the basal ganglia and cortex [10].

Sequential analysis methods for optimal decision making have been applied recently to robot perception, focussing on

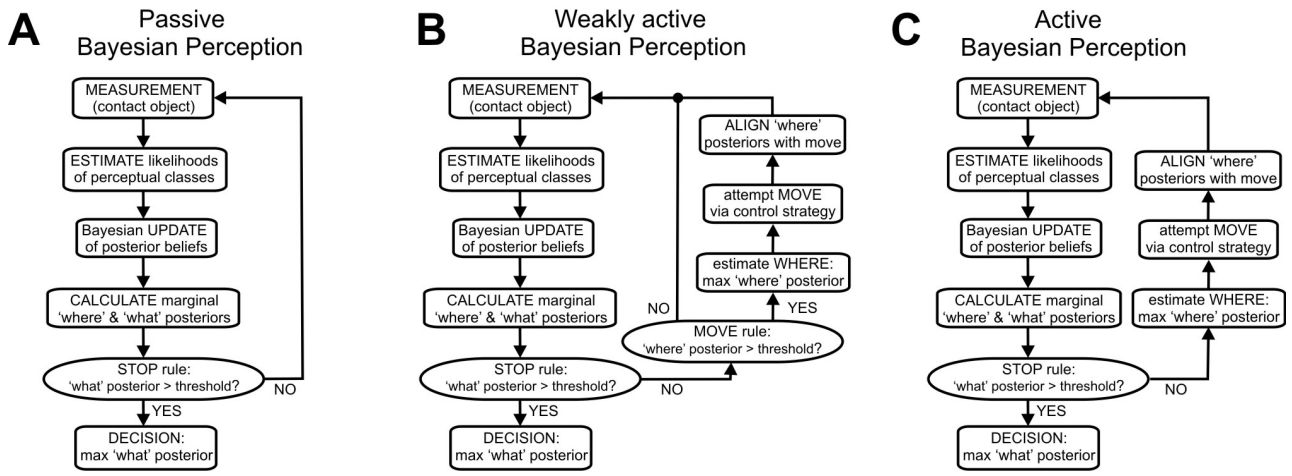

Fig. 1: Active and passive Bayesian perception for simultaneous object localization and identification. All algorithms share a recursive Bayesian update to find the marginal 'where' and 'what' beliefs, with decision termination at sufficient 'what' belief. They differ in their use of the 'where' belief for controlling sensor location: (A) passive perception has no active control; (B) weakly active perception requires the 'where' belief passes a threshold; and (C) active perception always tries to relocate. Sensor relocations follow a belief-based active control strategy, with the 'where' belief component re-aligned upon each move.

robot touch [12]. A strength of the formalism is that it connects closely with leading work in neuroscience, allowing insights from animal perception to be transferred to robot perception. For example, these methods have enabled the first demonstration of hyperacuity in robot touch [9], giving perceptual acuity finer than the sensor resolution, as is common in animal perception. That study also found that active perception helped the hyperacuity, although those methods do suffer from being somewhat *ad hoc* by not making best use of the 'where' and 'what' aspects of the problem.

The present work develops a principled approach to active Bayesian perception of applicability to 'where' and 'what' perceptual tasks. In this context, two recent studies of active touch in robotics bear comparison. One study proposed a biomimetic whisker control strategy inspired by rodent behavior (rapid cessation of protraction) that aided perception [17]. Although that study was neither Bayesian nor explicitly 'where' and 'what', we see their biomimetic control as analogous to the 'fixation point' strategy considered here. The other study used 'Bayesian exploration' to select between motions to best perceive surface texture with a biomimetic fingertip [4]. Their algorithm for Bayesian exploration has a similar looped architecture to that used here, but instead employs a 'disambiguation' control strategy to solve a solely 'what' task (the locations are known). The 'fixation point' strategy considered here solves 'where' and 'what' tasks, but we think an extension of their 'disambiguation' strategy would be appropriate when there is no obvious fixation point.

While active vision has been researched intensively for over 20 years [3], active touch has been neglected in comparison. However, it is becoming apparent that touch is key to solving some key problems in robotics, *e.g.* grasping under uncertainty [7]. The aim is for robots to accomplish the everyday manipulation tasks we take for granted, with radical implications for automatization in the home and industry [8, 13].

III. Active Bayesian Perception

Our algorithm for active perception is based on including a sensorimotor feedback loop in an existing method for passive Bayesian perception [9, 12]. Both methods assume that the sensor makes a discrete contact measurement (here a tap) onto an object, from which the joint likelihoods of object location and identity are used to update the posterior beliefs for those perceptual classes. In active Bayesian perception, a control strategy repositions the sensor before each contact, taking input from the beliefs and outputting the sensor move.

Because these methods are applicable to any simultaneous object localization and identification task, this section is presented in a general 'where' and 'what' notation. A general SOLID task has N_{loc} distinct 'where' location classes x_l and N_{id} distinct 'what' identity classes w_i, totalling $N = N_{\mathrm{loc}} N_{\mathrm{id}}$ joint 'where-what' classes $c_n = (x_l, w_i)$. In principle, there can be multiple location and identity dimensions, but we can still enumerate each 'where-what' class with an (l, i) index.

Each contact against a test object gives a multi-dimensional time series $z = \{s_k(j) : 1 \leq j \leq N_{\mathrm{samples}}, 1 \leq k \leq N_{\mathrm{channels}}\}$ of sensor values, with indices j, k labeling the time samples and sensor channels. The tth contact in a sequence is denoted by z_t with $z_{1:t-1} = \{z_1, \cdots, z_{t-1}\}$ its contact history.

Measurement model and likelihood estimation: The likelihoods of all perceptual classes are obtained from a measurement model of the contact data, which we find by applying a histogram method to training examples for each class [11, 12]. First, the sensor values s_k are binned into $N_{\mathrm{bins}} = 100$ intervals, with sampling distribution of each perceptual class c_n given by the normalized histogram for all data in that class

$$P(b|c_n, k) = \frac{h(b,k)}{\sum_{b=1}^{N_{\mathrm{bins}}} h(b,k)}, \tag{1}$$

where $h(b,k)$ is the histogram count for bin b ($1 \leq b \leq N_{\mathrm{bins}}$)

in channel k. Then, given a test tap z, the measurement model is constructed from the mean log likelihood over all samples

$$\log P(z|c_n) = \sum_{k=1}^{N_{\text{channels}}} \sum_{j=1}^{N_{\text{samples}}} \frac{\log P(b_k(j)|c_n, k)}{N_{\text{samples}} N_{\text{dims}}}, \quad (2)$$

where $b_k(j)$ is the bin occupied by sample $s_k(j)$. Technically, this measurement model becomes ill-defined if any histogram bin is empty, which is easily fixed by regularizing the bin counts with a small constant ($\epsilon \ll 1$), giving $h(b,k) + \epsilon$.

Bayesian update: Bayes' rule is used after each contact z_t to update the posterior beliefs with the estimated likelihoods

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

from background information given by the prior beliefs. In this recursive update, the marginal probabilities are

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

Iterating (3,4), a sequence of contacts $z_1, \cdots, z_t$ results in a sequence of posteriors $P(c_n|z_1), \cdots, P(c_n|z_{1:t})$ initialized from uniform priors $P(c_n) = P(c_n|z_0) = \frac{1}{N}$.

Marginal 'where' and 'what' posteriors: For the following methods, we will need the posterior beliefs for just location or identity, rather than the joint beliefs considered so far. These are found by marginalizing the joint beliefs for the classes $c_n = (x_l, w_i)$ over the location x_l or identity w_i component

$$P(x_l|z_{1:t}) = \sum_{i=1}^{N_{\text{id}}} P(x_l, w_i|z_{1:t}), \quad (5)$$

$$P(w_i|z_{1:t}) = \sum_{l=1}^{N_{\text{loc}}} P(x_l, w_i|z_{1:t}), \quad (6)$$

with the 'where' location beliefs given by summing over all 'what' identity classes w_i and the 'what' identity beliefs over all 'where' location classes x_l.

Final decision on the 'what' posteriors: Here we follow sequential analysis methods for optimal decision making that recursively update beliefs until reaching threshold [18], as used in passive Bayesian perception [12]. The update stops when the marginal 'what' identity belief passes a threshold, giving a final decision from the maximal *a posteriori* (MAP) estimate

$$\text{if any } P(w_i|z_{1:t}) > \theta_{\text{id}} \text{ then } w_{\text{id}} = \arg\max_{w_i} P(w_i|z_{1:t}). \quad (7)$$

This belief threshold θ_{id} is a free parameter that adjusts the balance between decision speed and accuracy. For $N = 2$, this speed-accuracy balance can be proved optimal [18]; optimality conditions are not known for many choices, and so we make a reasonable assumption of near optimality [12].

Move decision on the 'where' posteriors: Analogously to the stop decision, a sensor move requires a marginal 'where' location belief to cross its own decision threshold, with the MAP estimate giving the 'where' location decision

$$\text{if any } P(x_l|z_{1:t}) > \theta_{\text{loc}} \text{ then } x_{\text{loc}} = \arg\max_{x_l} P(x_l|z_{1:t}). \quad (8)$$

Here we consider three cases (Figs 1A,B,C), termed:

A. passive perception: $\theta_{\text{loc}} = 1$ (never moves)
B. weakly active perception: $0 < \theta_{\text{loc}} < 1$ (decides to move)
C. active perception: $\theta_{\text{loc}} = 0$ (always tries to move)

Active control strategy: The sensor movements are determined by the active control strategy based on the posterior beliefs. Here we consider variants of a 'fixation point' strategy: the sensor attempts to relocate to a predefined fixation point x_{fix} relative to the object assuming it is at the location x_{loc},

$$x_{\text{sensor}} \leftarrow x_{\text{sensor}} + \Delta(x_{\text{loc}}), \quad \Delta(x_{\text{loc}}) = x_{\text{fix}} - x_{\text{loc}}, \quad (9)$$

where x_{sensor} is the actual (unknown) location of the sensor. The arrow denotes that the quantity on the left is replaced with that on the right. We also consider a control strategy 'fixation point with noise' in which Gaussian noise of variance σ^2 is added to the fixation point $x_{\text{fix}} + N(0, \sigma)$ on each move, then rounded to the nearest 'where' class.

Align 'where' posteriors: Whatever control strategy, the 'where' location beliefs should be kept aligned with the sensor by shifting the posterior 'where-what' beliefs upon each move

$$P(x_l, w_i|z_{1:t}) \leftarrow P(x_l - \Delta(x_{\text{loc}}), w_i|z_{1:t}), \quad (10)$$

where we recalculate the beliefs outside the original range by assuming they are uniform and the shifted beliefs sum to unity.

IV. Tactile Data Collection

The aim of our data collection is to set up a 'virtual environment' in which methods for perception can be compared off-line on identical data. This is achieved by measuring contact

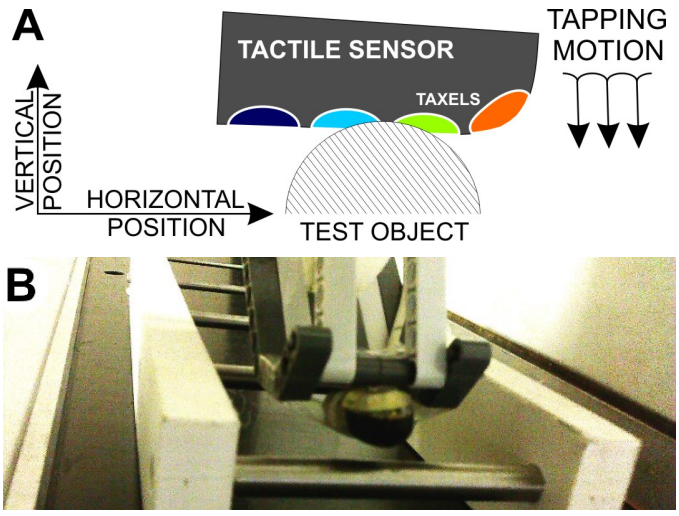

Fig. 2: Experimental setup. (A) Schematic of tactile sensor tapping against a cylindrical test object: the fingertip taps down and then back up again to press its pressure-sensitive taxels (colored) against the test object; each tap is then followed by a small horizontal move to systematically vary contact location. (B) Forward view of the experiment showing the fingertip mounted on the arm of the Cartesian robot. This experimental setup is ideal for systematic data collection to characterize the properties of the sensor interacting with its environment.

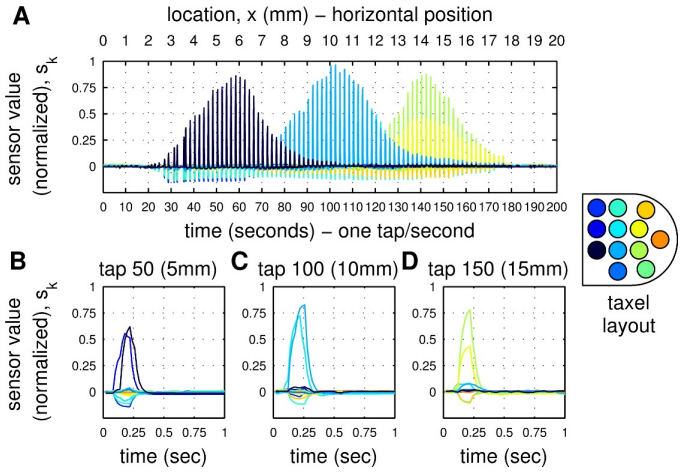

Fig. 3: Tactile data at fixed vertical position (4 mm test rod). (A) Horizontal sweep (20 mm; 200 taps) at the lowest vertical position. (B-D) Individual tap data taken from panel (A). Taxels are colored according to their layout on the fingerip.

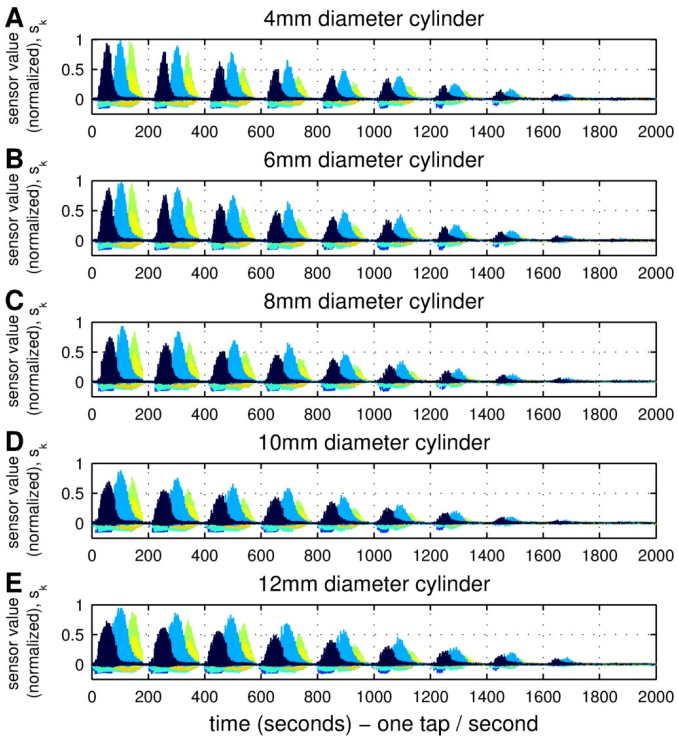

Fig. 4: Complete tactile data set. Each panel (A-E) is for one of the $N_{\rm id}$ =5 test rods (diameters 4,6,8,10,12 mm). Tickmarks show the limits of each 20 mm horizontal sweep, after which the vertical position is increased by 0.4 mm to span a 4 mm range. Taxel colors are the same as Fig. 3.

signals against multiple objects (for 'what') over an exhaustive range of contact locations (for 'where'). Our experimental situation has a tactile fingertip tap against $N_{\rm id} = 5$ smooth rods over a 2D range of locations (Fig. 2), which we partition into $N_{\rm loc} = 200$ location classes (20 horizontal by 10 vertical).

The tactile sensor has a rounded shape that resembles a human fingertip [16], with dimensions 14.5 mm long by 13 mm wide. It consists of an inner support wrapped with a flexible printed circuit board containing 12 conductive patches for the touch sensor 'taxels', about 4 mm apart. This is coated with non-conductive foam and conductive silicone layers, together comprising a capacitive touch sensor that detects pressure via compression. Data were collected at 8-bit resolution and 50 cycles/sec then high-pass filtered and normalized [16].

For measuring contact data over an exhaustive set of locations, we mounted the fingertip on a Cartesian robot capable of precise positioning in a horizontal/vertical plane ($\sim$20 μm accuracy). The fingertip was mounted at an orientation appropriate for contacting axially symmetric shapes such as cylinders aligned along an axis perpendicular to the plane of movement (Fig. 2), initially contacting at its base and finishing on its tip. $N_{\rm id} = 5$ smooth steel rods with diameters 4,6,8,10,12 mm were used as test objects, mounted with their centers offset vertically (by 4,3,2,1,0 mm) to align their closest point of contact with the fingertip in the direction of tapping.

Touch data were collected while the fingertip tapped vertically onto and off each test object, followed by a horizontal move $\Delta x = 0.1$ mm across its closest face (Fig. 2A). Every 200 taps, the vertical position was increased by $\Delta y = 0.4$ mm and the horizontal position decreased by 20 mm to reset. A horizontal x-range of 20 mm and vertical y-range of 4 mm was used, giving 2000 taps for each of the $N_{\rm id} = 5$ objects, or 10000 taps in total. From each tap of the fingertip against the object, a 1 sec time series of pressure readings ($N_{\rm samples} = 50$) was extracted for all $N_{\rm taxels} = 12$ taxels (Fig. 3). A 4 sec pause was taken between brief (~ 0.1 sec) contacts to ensure transients decayed; no noticeable hysteresis then occurred. All data were collected twice to give distinct training and test sets.

For analysis, the data were separated into $N_{\rm loc} = 200$ distinct location classes x_l by collecting groups of 10 taps each spanning 1 mm$\times$0.4 mm of the 20 mm$\times$4 mm range (Figs 3,4). In total, there were thus $N = N_{\rm loc} N_{id} = 1000$ distinct 'where-what' perceptual classes (x_l, w_i). These were used to set up a 'virtual environment' to compare the methods from Sec. III on identical data. A Monte Carlo validation ensured good statistics, by averaging perceptual acuities over many test runs with taps drawn randomly from the classes (5000 runs per point in the following plots). Perceptual acuities $e_{\rm loc}$, $e_{\rm id}$ were quantified using the mean absolute error (MAE) for each dimension between the actual $x_{\rm test}$, $w_{\rm test}$ and classified values $x_{\rm loc}$, $w_{\rm id}$ of object location and identity over the test runs.

Inspecting the data, taps typically took $\sim$0.1 sec to reach peak amplitude, followed by rapid decay to baseline (Figs 3B-D); note that some amplitudes were negative because of outwards deformation of the fingertip surface. The most obvious effect of varying horizontal position was a change in taxel identity and amplitude (Fig. 3A); changing vertical position also changed amplitude (Fig. 4). For each contact, the pattern of taxel pressures depended on both the curvature of the surface and the 2D position of the fingertip relative to the rod, permitting simultaneous object localization and identification.

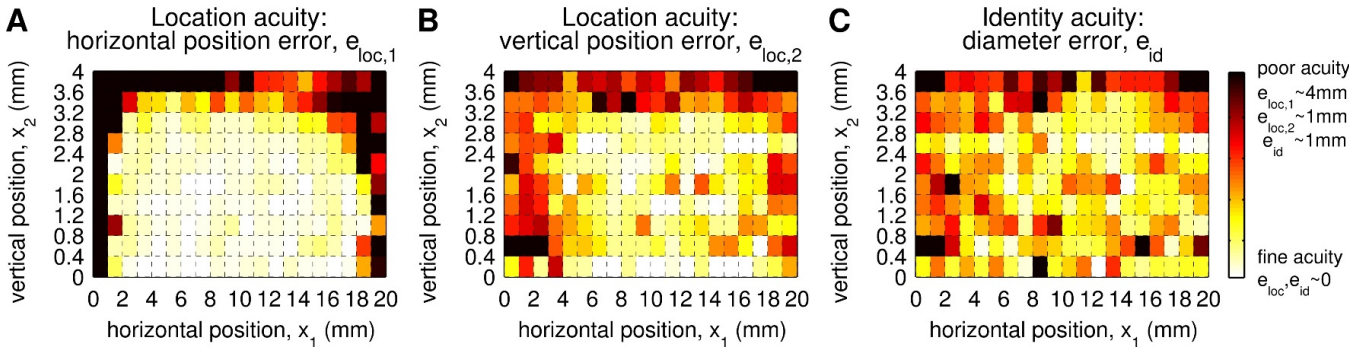

Fig. 5: Passive perception for simultaneous object localization and identification. The plots show the dependence of the 'where' and 'what' perceptual acuities on horizontal x_1 and vertical x_2 position: (A) horizontal position acuity $e_{\mathrm{loc},1}$ (range 0-4 mm); (B) vertical position acuity $e_{\mathrm{loc},1}$ (range 0-1 mm); and (C) identity acuity e_{id} (range 0-1 mm). Results are for belief threshold $\theta_{\mathrm{id}} = 0.5$. Light colors indicate finest acuities (small errors) and dark colors the poorest acuities in the range (large errors).

V. Simultaneous Object Localization and Identification

A. Passive Bayesian perception

This section examines the application of passive Bayesian perception to simultaneous object localization and identification. Bayesian perception updates posterior belief for distinct 'where' 2D location $(x_{l,1}, x_{l,2})$ and 'what' identity classes w_i, using successive taps z_t against a test object until at least one identity belief $p(w_i|z_{1:t})$ crosses a belief threshold θ_{id}. *Passive* perception means that the location class x_l does not change during perception (Fig. 1A). The results are generated using a Monte Carlo procedure over test data as a virtual environment (Sec. IV), averaging over 5000 'where' and 'what' perceptual decisions of object location and identity for each considered belief threshold $\theta_{\mathrm{id}} = \{0, 0.1, \cdots, 0.9, 0.95, 0.99\}$.

Our first observation is that perceptual acuity depends on the 2D test location class (Fig. 5). Perceptual acuities for horizontal $e_{\mathrm{loc},1}$ and vertical $e_{\mathrm{loc},2}$ object location and identity e_{id} deteriorate (increasing error) at the highest vertical test positions and both extremes of the horizontal range (Figs 5A-C; numerical ranges in legend). These findings are consistent with contacts at these extremities being weak with poor signal-to-noise ratio. For passive perception, there is no control over the location from where an object is sensed. Hence, we typify the perceptual acuities for this task as averages $\bar{e}_{\mathrm{loc},1}$, $\bar{e}_{\mathrm{loc},2}$, $\bar{e}_{\mathrm{id}}$ over all possible sensing locations (Fig. 6, red plots).

Our next observation is that the mean passive location and identity perceptual acuities $\bar{e}_{\mathrm{loc},1}$, $\bar{e}_{\mathrm{loc},2}$, $\bar{e}_{\mathrm{id}}$ improve (decreasing error) with increasing belief threshold θ_{id} (Figs 6A-C; red plots). These variations in acuity are not unexpected, because higher thresholds require greater certainty, improving accuracy but also delaying the decision (reaction) time (Fig. 6D). Optimal acuities for location and identity reach $\bar{e}_{\mathrm{loc},1} \sim 1.7$ mm and $\bar{e}_{\mathrm{loc},1} \sim 0.13$ mm and $\bar{e}_{\mathrm{id}} \sim 0.7$ mm, respectively. The better perceptual acuity for vertical compared with horizontal position relates to the touch sensors having fine sensitivity to changes in pressure, but a relatively large taxel spacing (4 mm).

B. Active Bayesian perception - fixation point strategy

This section considers active Bayesian perception applied to the same simultaneous object localization and identification task as passive perception. Active Bayesian perception also accumulates belief for location $(x_{l,1}, x_{l,2})$ and identity w_i by successively tapping against an object, while in addition trying to relocate the sensor according to these beliefs (Figs 1B,C). The first active perception method considered here adopts a 'fixation point' strategy where a best estimate of current location is used to calculate a relative move towards a fixed location on the object. We take a fixation point centered horizontally and low down the vertical range ($x_{\mathrm{loc},1} = 10$ mm, $x_{\mathrm{loc},2} = 1.0$ mm), where the passive perception had good acuity. Other details remain unchanged, apart from the test location class now represents only initial sensor position and an additional 'where' belief threshold within a range $\theta_{\mathrm{loc}} = \{0, 0.1, \cdots, 0.9, 0.95, 0.99\}$ indicates when to move.

Active and passive perception are here considered a continuum parameterized by the strength of 'where' belief θ_{loc} needed to trigger a relocation decision. One extreme is passive perception where the sensor never relocates ($\theta_{\mathrm{loc}} \geq 1$) and the other extreme is active perception where the sensor always tries to relocate ($\theta_{\mathrm{loc}} \leq 1/N_{\mathrm{loc}}$). We consider the intermediate range of 'where' belief thresholds ($1 > \theta_{\mathrm{loc}} > 1/N_{\mathrm{loc}}$) to represent *weakly active* perception where multiple taps are necessary to decide when to move. The differences in relocation behavior for passive and active perception is apparent in the location trajectories (Fig. 7A,D *vs* B,E), which for passive perception remain static whereas for active perception home in on the fixation point.

Our first observation is that the 'where' and 'what' mean perceptual acuities $\bar{e}_{\mathrm{loc},1}$, $\bar{e}_{\mathrm{loc},2}$, $\bar{e}_{\mathrm{id}}$ improve with *decreasing* 'where' belief threshold θ_{loc} (Figs 6A-C): passive perception has poorest acuity (shown in red), improving slightly when weakly active (light gray), then steadily improving until reaching the finest acuities for fully active perception (black). In addition, the three perceptual acuities improve with *increasing* 'what' belief threshold θ_{id}, as noted for passive perception but

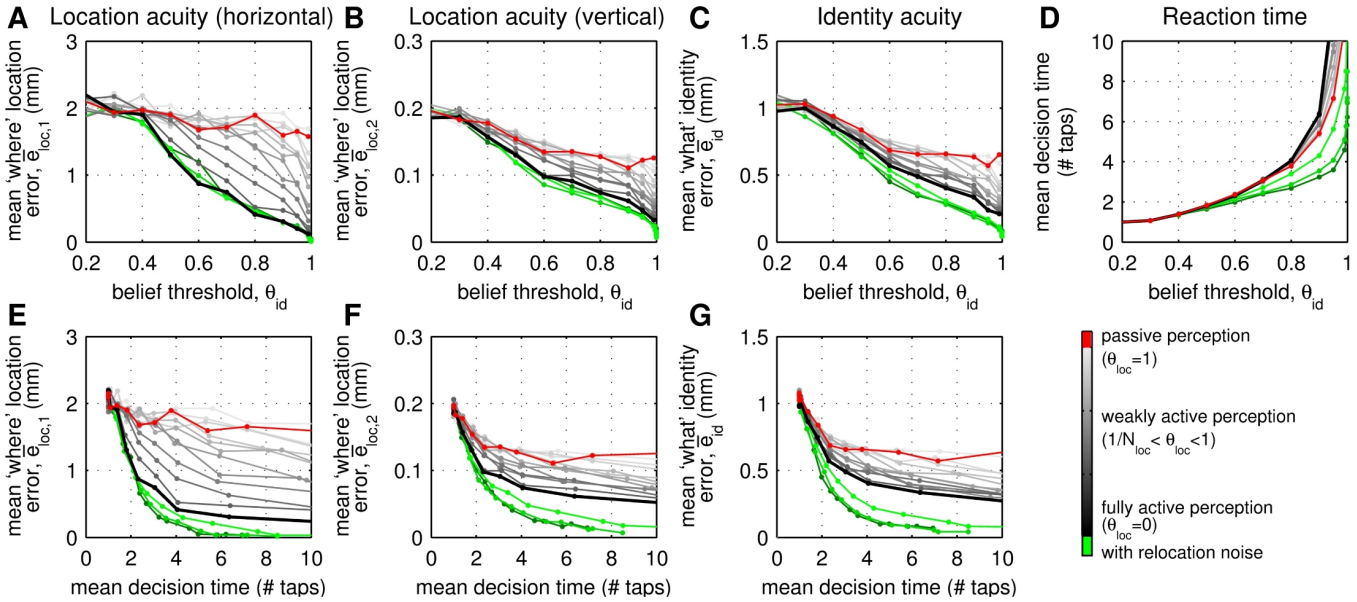

Fig. 6: Active perception for simultaneous object localization and identification. (A-C) Dependence of the perceptual acuities on 'what' belief threshold $\theta_{\rm id}$ (abscissa) and 'where' belief threshold $\theta_{\rm loc}$ (grey shade of plot), as measured by the mean horizontal and vertical location errors $(\bar{e}_{{\rm loc},1}, \bar{e}_{{\rm loc},2})$ and the mean identity error $\bar{e}_{\rm id}$, over all possible locations. Passive perception ($\theta_{\rm loc} = 1$) is shown in red, weakly active perception in gray and fully active perception ($\theta_{\rm loc} = 0$) in black. Active perception with fixation noise is shown in green, darkening in shade with increasing variance. The other panels show (D) the corresponding mean decision times and (E-G) the same plots of perceptual acuity with mean decision time as abscissa.

now apparent for both active and passive perception (Fig. 6).

Our second observation is that the mean decision (reaction) time to gather evidence for relocation increases with the 'what' belief threshold $\theta_{\rm id}$ (Fig. 6D). Just just 1 or 2 taps are required below 0.5 threshold, increasing steeply to $\sim$10 taps at $\theta_{\rm id} = 0.9$. In contrast, varying the 'where' threshold $\theta_{\rm loc}$ does not appreciably affect decision time. Overall, perceptual acuity improves steeply with mean decision time up to around 4 taps then more gradually thereafter (Figs 6E-G).

Comparing fully active perception with passive perception ($\theta_{\rm loc} = 0$ or 1), the 'where' and 'what' perceptual acuities (for 5 taps) improve from $\bar{e}_{{\rm loc},1} = 1.7$ mm to 0.4 mm for horizontal location (*cf.* 20 mm range), from $\bar{e}_{{\rm loc},2} = 0.13$ mm to 0.07 mm for vertical location (*cf.* 4 mm range) and from $\bar{e}_{\rm id} = 0.7$ mm to 0.4 mm for object identity (*cf.* 8 mm diameter range). Evidently, the finest perceptual acuity arises from fully active perception, followed by weakly active then passive methods.

C. Active Bayesian perception - fixation point with noise

A second active perception strategy is now considered with relocation movements disturbed by noise, equivalent to tap-to-tap perturbations of the fixation point. In other respects, the active perception is identical to above (Sec. V-B), which is now considered the noise-free case. Only fully active perception is examined (Fig. 1C), because that was found to give the finest 'where' and 'what' perceptual acuity for simultaneous object localization and identification.

The inclusion of relocation noise causes the trajectories to target a blurred region rather than a single point (Figs 7C,F), with greater tap-to-tap stochasticity than for noise-free active perception. Gaussian noise with standard deviation $\sigma = \{1, 2, 3, 4\}$ horizontal and vertical location classes is considered (units: 1 mm horizontal, 0.4 mm vertical), such that on each tap the two directional components of the noise are found then rounded to the nearest class label.

Our first observation is that relocation noise causes the decision times to decrease in comparison with noise-free active perception at the same 'what' identity threshold $\theta_{\rm id}$ (Fig. 6D; green *vs* black plots). Apparently, fixating over a region gives a greater chance of attaining sufficient evidence to reach decision threshold, compared with a point fixation where the discrimination may get stuck at a 'bad' location.

Our second observation is that relocation noise improves perceptual acuity compared with noise-free active perception at identical mean decision times (Figs 6E-G; green *vs* black plots). This result arises principally from an improved decision time, although there were further improvements in identity acuity (Fig. 6C), apparently because sensing over a region helps perceive shape (radius of curvature). In all cases, there is a moderate improvement of perceptual acuity for relocation noise up to standard deviation $\sigma = 2$ and little thereafter.

Comparing active perception with noise to the noise-free case, the 'where' and 'what' perceptual acuities (for 5 taps) improve from $\bar{e}_{{\rm loc},1} = 0.4$ mm to 0.1 mm for horizontal location (*cf.* 20 mm range), from $\bar{e}_{{\rm loc},2} = 0.07$ mm to 0.02 mm for vertical location (*cf.* 4 mm range) and $\bar{e}_{\rm id} = 0.4$ mm to 0.1 mm for object identity (*cf.* 8 mm diameter range). Evidently, considerable improvements in acuity result from including a moderate amount of noise in the active perception strategy.

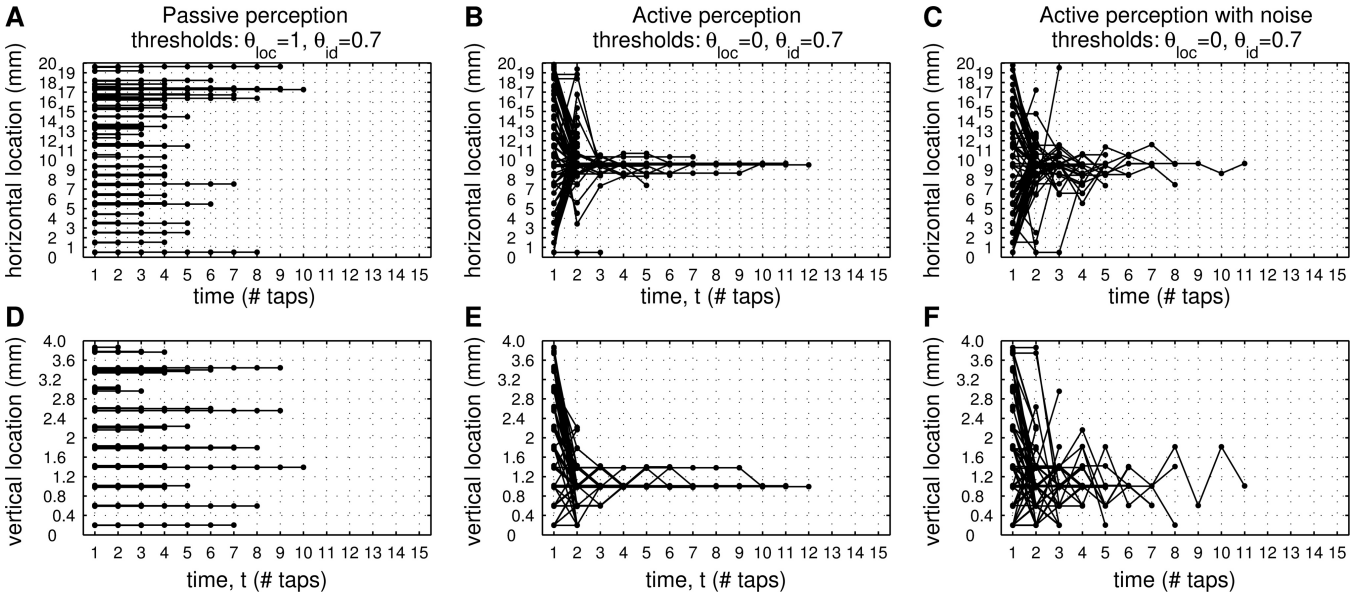

Fig. 7: Trajectories for passive and active perception. (A,D) Passive perception, with horizontal and vertical location plotted against time. (B,E) Corresponding plots for active perception, with fixation point centered horizontally (10 mm) and towards the bottom (1.0 mm) vertically. (C,F) Corresponding plots for active perception with noise (standard deviation $\sigma = 1$ class). 100 trajectories were selected randomly for each case.

VI. Discussion

A. Best strategy for robot perception

In this study, we applied Bayesian perception to simultaneous object localization and identification, or perceiving 'where' and 'what'. We compared active and passive methods on a simple but illustrative task of perceiving 2D position (localization) and object curvature (identification) with tapping movements of a biomimetic fingertip. Active perception can control changes in location of the sensor during the decision making process (Fig. 1B,C), whereas for passive perception the sensor remains at the location where it initially contacted the object (Fig. 1A). We then compared various active perception strategies to evaluate which gave the best robot perception.

Our three main observations about active perception were:
(i) *Active perception gave an order-of-magnitude improvement in acuity over passive perception.* A 'fixation point' active control strategy gave much better perceptual acuity for 2D location and identity than passive perception. For the best choice of parameters (see below), the mean perceptual errors improved from $\bar{e}_{\mathrm{loc},1} \sim 1.7\,\mathrm{mm}$, $\bar{e}_{\mathrm{loc},2} \sim 0.13\,\mathrm{mm}$ and $\bar{e}_{\mathrm{id}} \sim 0.7\,\mathrm{mm}$ (passive perception) to $\bar{e}_{\mathrm{loc},1} \sim 0.1\,\mathrm{mm}$, $\bar{e}_{\mathrm{loc},2} \sim 0.02\,\mathrm{mm}$ and $\bar{e}_{\mathrm{id}} \sim 0.1\,\mathrm{mm}$ (active perception), an order-of-magnitude improvement. This was due partly to attaining a good location for perception from any initial contact location. However, the results also indicated other benefits of active control (*e.g.* sampling a range of locations).
(ii) *Perceptual acuity improves with the strength of the active perception.* Just as active Bayesian perception requires an identity threshold θ_{id} on the beliefs to make a 'what' decision, there should be a location threshold θ_{loc} to make a 'where' decision to relocate the sensor. Passive perception corresponds to an unattainable threshold $\theta_{\mathrm{loc}} \geq 1$ and fully active perception to always attainable thresholds $\theta_{\mathrm{loc}} \leq 1/N_{\mathrm{loc}}$; the intermediate range $1/N_{\mathrm{loc}} < \theta_{\mathrm{loc}} < 1$ parameterizes the strength of the active perception, with this entire range referred to here as *weakly* active perception to distinguish it from fully active perception. Acuity improved as the active perception became stronger, from passive to weakly active to fully active perception. This result is not obvious in advance, and follows from greater improvements to acuity arising from rapid move decisions than the decrements from poorer relocations. Therefore, only fully active Bayesian perception need be applied for 'where' and 'what' tasks like those considered here.
(iii) *Relocation noise in active perception can improve acuity.* We also considered an active perception strategy with relocation noise. For moderate noise variance, the perceptual acuity was better than the noise-free case, giving the best overall active perception strategy 'fixation point with noise' on our tactile data. The improved perception of object identity was due partly to help from sensing curvature over a region rather than a point. Sampling from a larger region also gave greater chance of receiving 'good' evidence, which resulted in faster decision times and hence improved location acuity.

B. Similarities with animal perception

Similarly to previous work on active perception [1, 2], the inspiration for this study was from animal perception. We now describe briefly some similarities between the present approach and aspects of animal behavior and physiology.

First, the Bayesian perception method is based on leading models of perceptual decision making from neuroscience [6]. An aspect of animal perception that these models capture is to optimize the tradeoff between reaction speed and error rate.

Second, the biomimetic fingertip has taxels with broad, overlapping receptive fields (Fig. 4A), analogous to those of mechanoreceptors (touch) and photoreceptors (vision). Bayesian perception then gives perceptual hyperacuity [9], a phenomenon associated with animal perception. Given the taxel spacing is 4 mm, our results display extreme hyperacuity.

Third, active Bayesian perception employs a sensory-motor loop to move the biomimetic fingertip in response to sensory information during the perceptual process. Sensorimotor feedback during perception is a generic aspect of both animal behavior and the physiology of the vertebrate brain.

Fourth, the 'fixation point' active perception strategy used here is analogous to orienting movements common in animal perception, of which foveation in vision is an example [2].

Finally, relocation noise in the active perception strategy is similar to microsaccades that occur in vision [14]. Considering that the function of microsaccades is under debate, our results suggest that they improve acuity in active perception.

C. Generalization to other 'where' and 'what' tasks

As stated in the introduction, we expect that our formalism and results are representative of other 'where' and 'what' tasks. Our 2D example in robot touch was chosen to be simple enough to ease the presentation and analysis, but sufficiently complex to represent a realistic environment. By sensing only cylindrical stimuli from a fixed orientation, the entire environment could be mapped over a 2D range of contact positions. This situation generalizes in an obvious way to 3D position, but in addition the sensor orientation could also be under active control as part of the 'where' localization.

More generally, we interpret the 'where' location dimensions as those degrees-of-freedom that the sensing agent can actively control, like the pose of a single sensor or possibly the sensor geometry (*e.g.* the morphology of a whisker array [17]). Conversely, the 'what' identity dimensions are the perceptual degrees-of-freedom that are outside the agent's control, such as those intrinsic to the object being identified (*e.g.* curvature, as here, or surface texture, as considered in a related study [13]). Active Bayesian perception should thus apply to any SOLID problem with appropriate definition of 'where' and 'what'.

VII. Conclusion

In this paper, we propose that active perception has general role for Simultaneous Object Localization and IDentification (SOLID), or 'where' and 'what' objects are being sensed. Our active perception method combines sequential analysis for optimal decision making [18] with active control for perception [1]. Aspects of this method resemble those of animal perception, supporting wide applicability of this approach.

We believe that this work gives a step towards a long-term goal of enabling robots to perceive their environments with the remarkable perceptual capabilities of animals over limited sensor information and in demanding circumstances.

Acknowledgments: We thank Ben Mitchinson and Mat Evans for comments on earlier drafts of this paper. This work was funded by the European Commission via the FP7 project EFAA (ICT-270490); UMH was also supported by CONACyT.

References

[1] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

[2] D.H. Ballard. Animate vision. *Artificial intelligence*, 48 (1):57–86, 1991.

[3] S. Chen, Y. Li, and N. Kwok. Active vision in robotic systems: A survey of recent developments. *International Journal of Robotics Research*, 30(11):1343–1377, 2011.

[4] J. Fishel and G. Loeb. Bayesian exploration for intelligent identification of textures. *Frontiers in Neurorobotics*, 6, 2012.

[5] J.J. Gibson. Observations on active touch. *Psychological review*, 69(6):477, 1962.

[6] J. Gold and M. Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574, 2007.

[7] K. Hsiao, L. Kaelbling, and T. Lozano-Pérez. Task-driven tactile exploration. *Robotics: Science and Systems*, 2010.

[8] C.C. Kemp, A. Edsinger, and E. Torres-Jara. Challenges for robot manipulation in human environments [grand challenges of robotics]. *Robotics & Automation Magazine, IEEE*, 14(1):20–29, 2007.

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

[10] N.F. Lepora and Gurney K. The basal ganglia optimize decision making over general perceptual hypotheses. *Neural Computation*, 24(11):2924–2945, 2012.

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

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

[13] 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*, pages 3005–3010, 2013.

[14] S. Martinez-Conde, J. Otero-Millan, and S. Macknik. The impact of microsaccades on vision. *Nature Reviews Neuroscience*, 14:83–96, 2013.

[15] T. Prescott, M. Diamond, and A. Wing. Active touch sensing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1581):2989–2995, 2011.

[16] 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.

[17] J.C. Sullivan *et al*. Tactile discrimination using active whisker sensors. *IEEE Sensors*, 12(2):350–362, 2012.

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