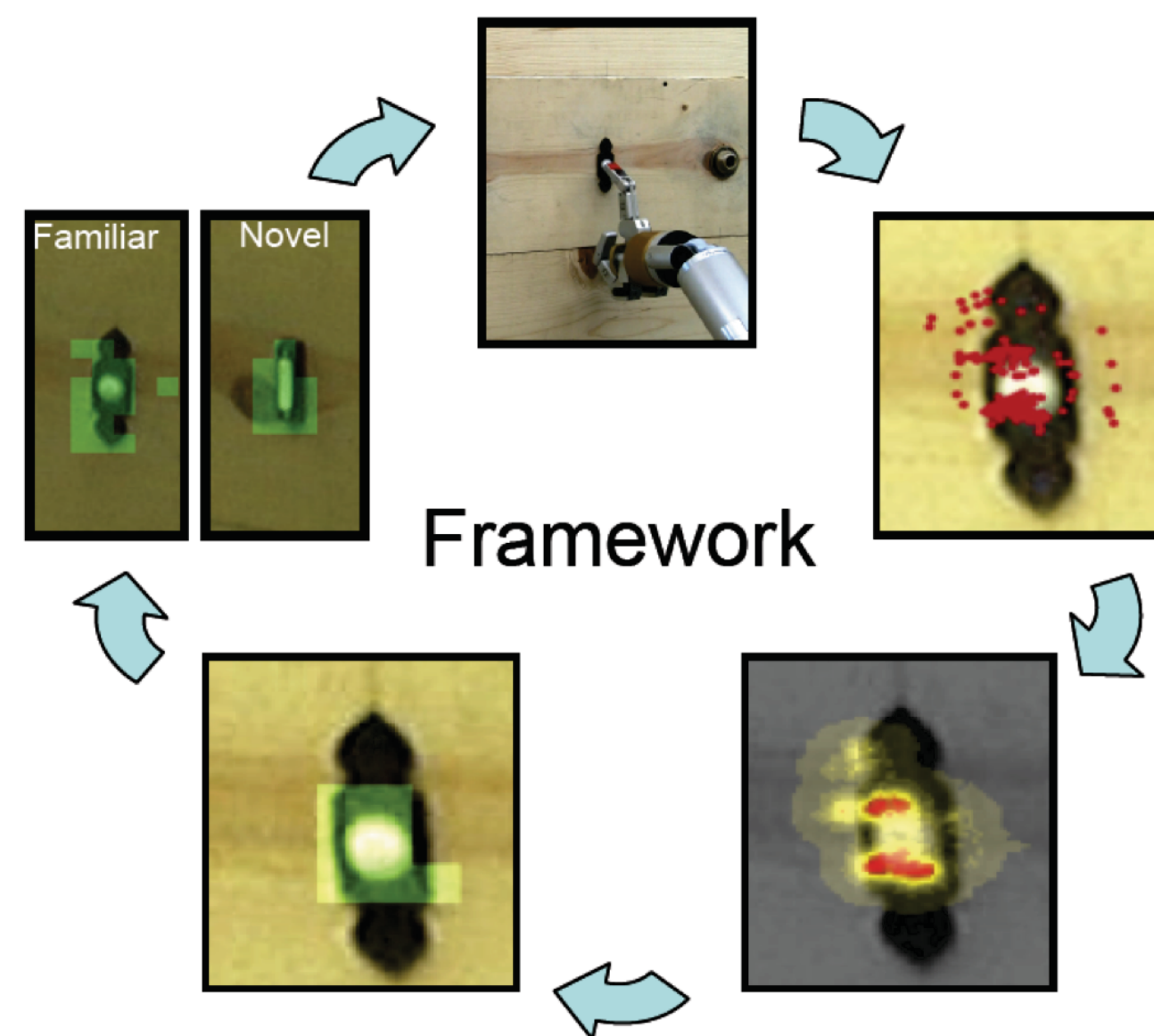
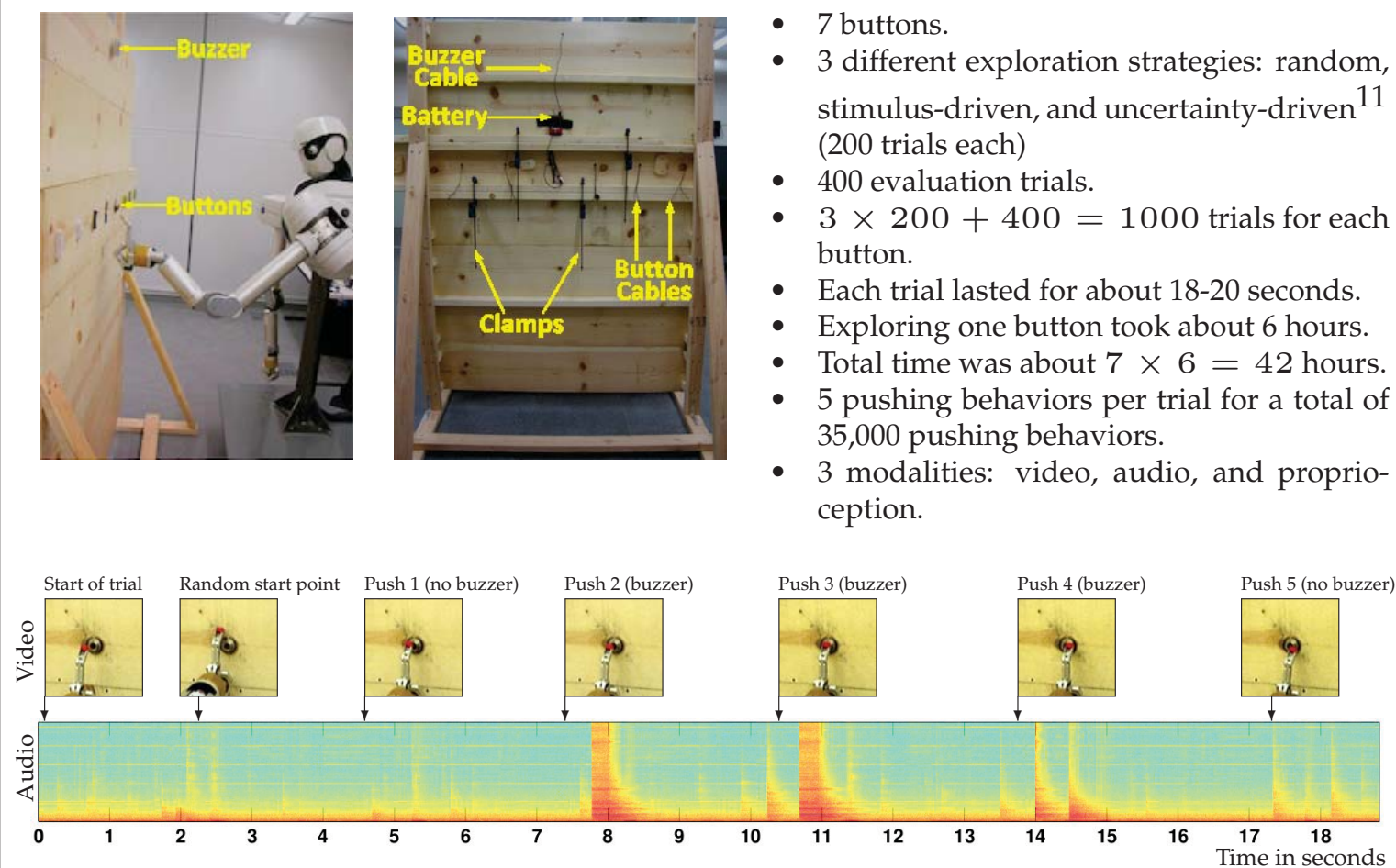


Summary



Experimental Setup

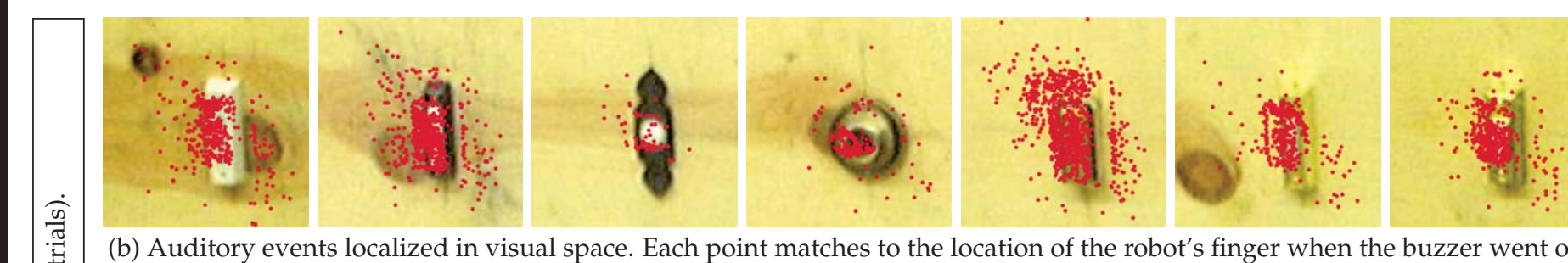


A sample trial performed by the robot. The robot's field of view is larger than the images shown here, which were cropped to show only the area around the button.

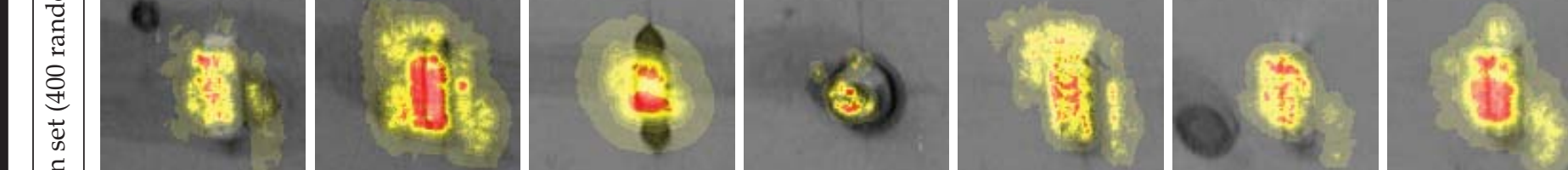
Results



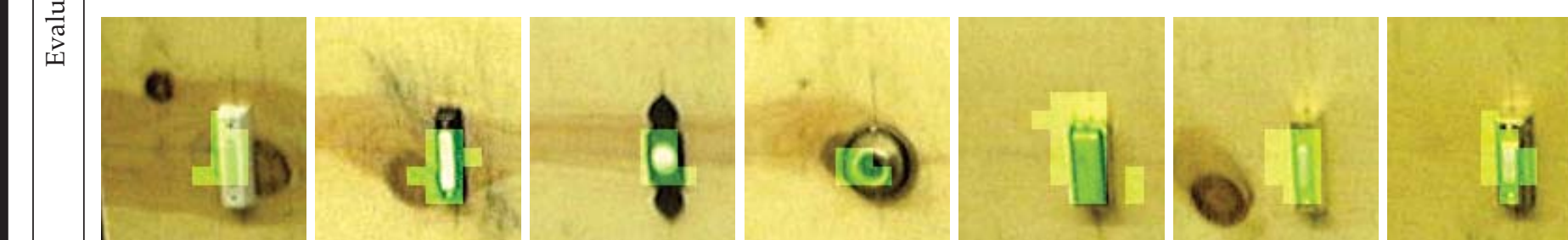
(a) The seven doorbell buttons explored by the robot.



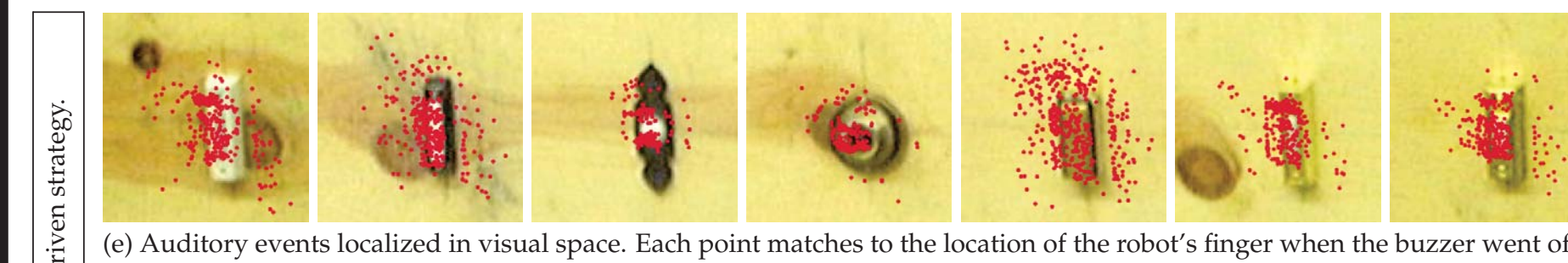
(b) Auditory events localized in visual space. Each point matches to the location of the robot's finger when the buzzer went off.



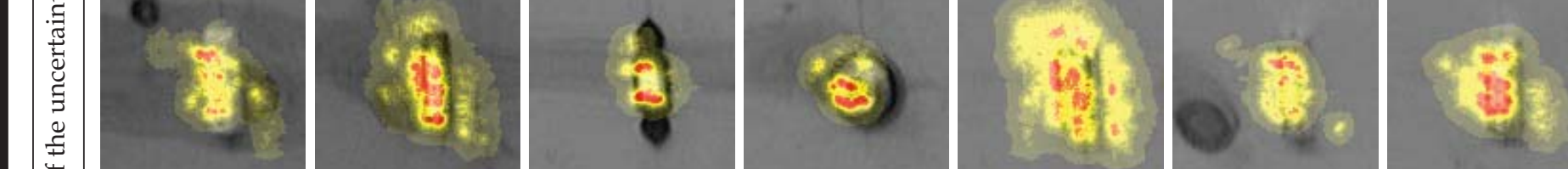
(c) Densities for the auditory events in visual space shown in (b), estimated using k-NN with $k = 5$.



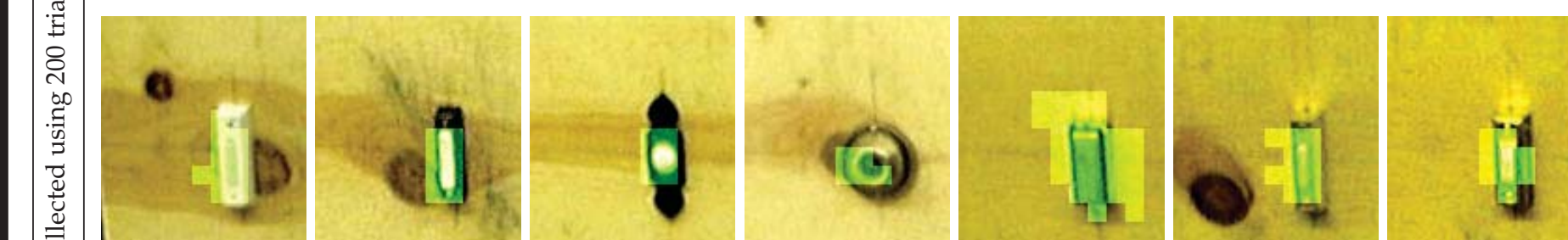
(d) "Ground truth" about the visual positions of the functional components extracted by thresholding the densities shown in (c).



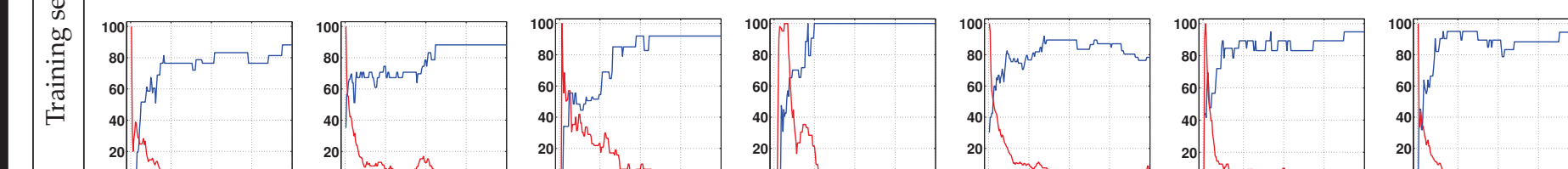
(e) Auditory events localized in visual space. Each point matches to the location of the robot's finger when the buzzer went off.



(f) Densities for the auditory events in visual space shown in (e), estimated using k-NN with $k = 5$.



(g) Functional components for each button learned after 200 trials performed with the uncertainty-driven strategy.

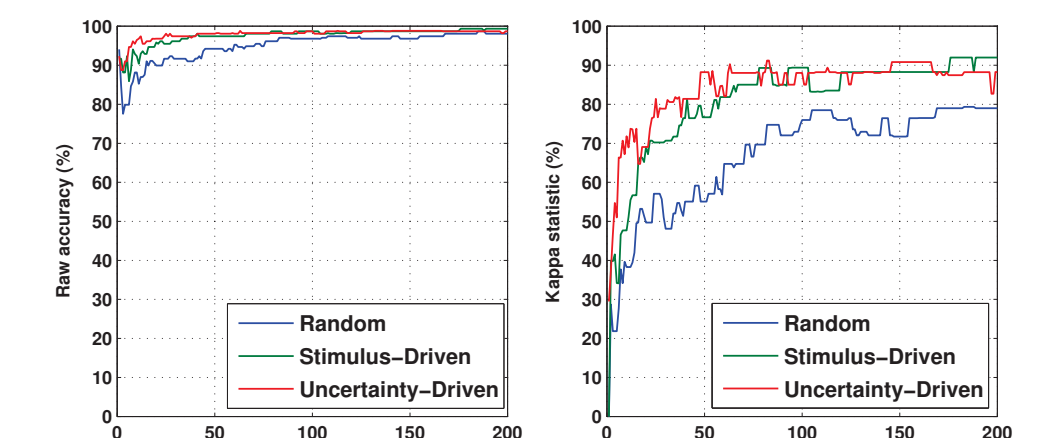


(h) Two measures of learning progress. The predictions after 200 trials are shown in (g). The "ground truth" is shown in (d). The kappa statistic (blue line, %) and the normalized smoothed rate of change (red line, %) are shown as functions of the number of trials.

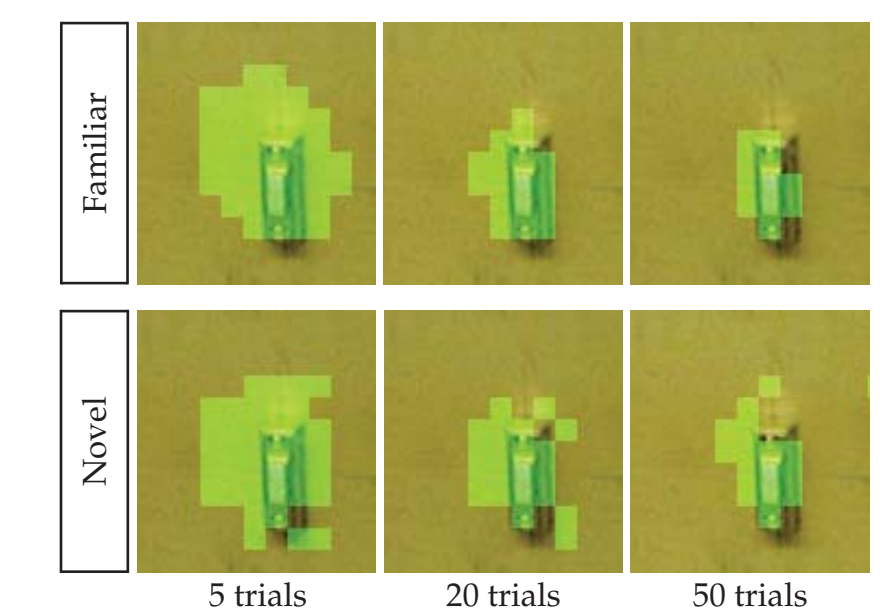
$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Raw accuracy \rightarrow $\Pr(a)$
Probability of correct prediction by chance \rightarrow $\Pr(e)$

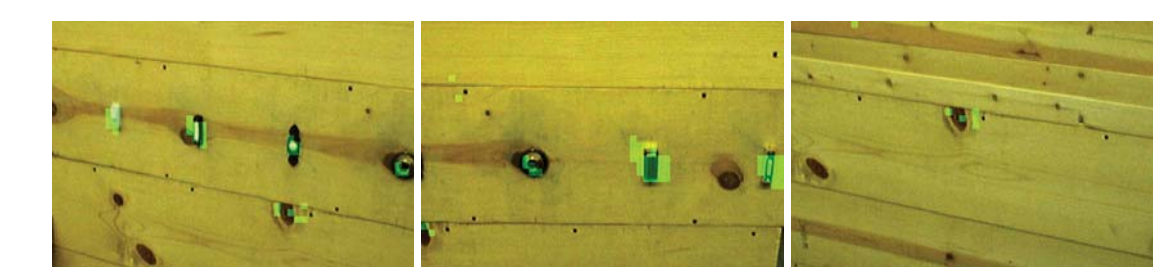
Performance was measured using the Cohen's kappa statistic⁹.



Results for predicting if a 10×10 pixel patch in an image belongs to the functional component of a familiar button as a function of the # of trials (average over all 7 buttons).



Predictions for the visual locations of functional components after different amounts of training using the uncertainty-driven strategy. For the novel button, predictions are made using the data collected with each of the remaining 6 buttons.



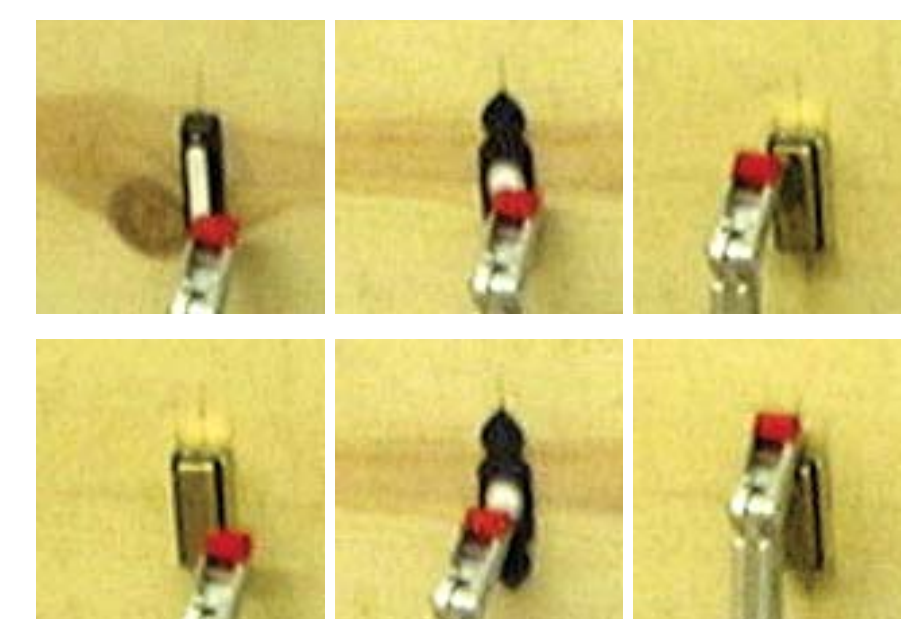
After the visual model was trained with all 7 buttons, the robot was tested with images of the experimental fixture that it had never seen before.

Motivation

- **Buttons are everywhere.**
- Robots in human-inhabited environments must press buttons to be more useful.
- Buttons are designed for humans, not for robots (some buttons may be too small or too slippery for robotic fingers made of brushed aluminium).
- Different buttons produce different feedback when pressed (e.g., click, light up or ring).
- Infants learn from experience obtained through exploration⁶.
- Experience that infants obtain while exploring objects stimulates their interest in these objects⁷.
- 9 m.o. infants can predict that an interesting sound will be heard or a bright light will be seen when an experimenter presses a colored button⁸.

Methodology

- Learning from a large-scale dataset with 35000 pushing behaviors¹¹.
- Auditory events were associated with the location of the color marker on the robotic finger.



The red marker on the robot's fingertip does not have to overlap with the button to ring it.

- Visual features were used to detect familiar and novel buttons.

Related Work

1. Detecting buttons is hard but pressing them is easy.

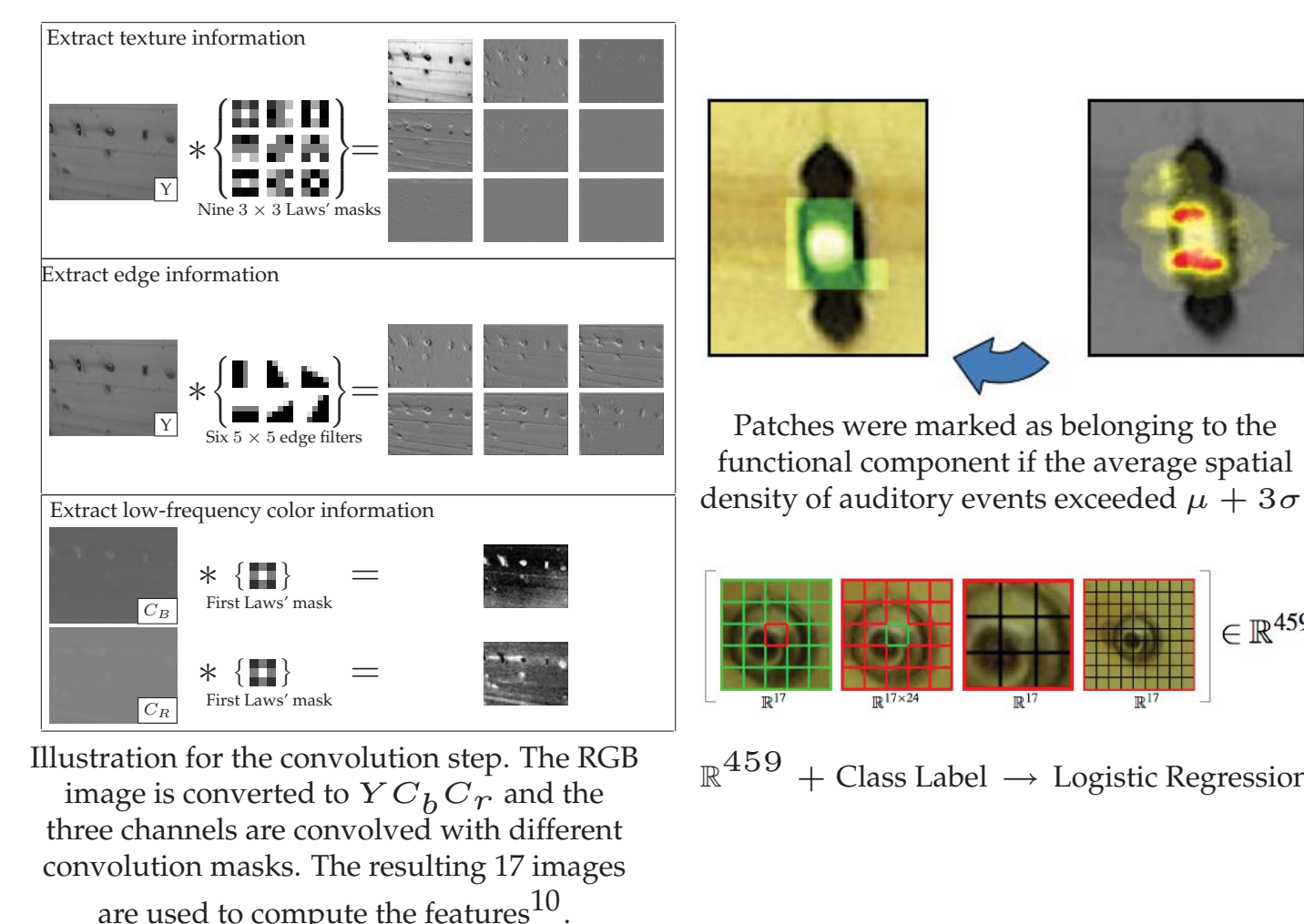


2. Both pressing and detecting buttons is hard.
3. Button pressing as a social learning task.



(Nguyen et al. 2009)

(Thomaz 2006)



Conclusions and Future Work

- The learned representations were grounded in the robot's experience with the buttons.
- **The trained visual model acted like an affordance detector, labeled patches as "pushable".**
- 50-100 trials were sufficient for convergence.
- Future work can add tactile feedback and button resistance as button properties.
- The exploration strategies need to take into account the predictions of the visual model.
- The framework can be extended to handle other small widgets.

References

1. K.-T. Song and T.-Z. Wu, "Visual servo control of a mobile manipulator using one-dimensional windows," in Proc. of Industrial Electronics Society, vol. 2, 1999, pp. 686-691.
2. J. Miura, K. Iwase, and Y. Shirai, "Interactive teaching of a mobile robot," in Proc. of ICRA, 2005, pp. 3378-3383.
3. E. Klingbeil, B. Carpenter, O. Russakovsky, and A. Ng, "Autonomous operation of novel elevators for robot navigation," in Proc. of ICRA, 2010, pp. 751-758.
4. H. Nguyen, T. Deyle, M. Reynolds, and C. Kemp, "PPS-tags: Physical, Perceptual and Semantic tags for autonomous mobile manipulation," in Proc. of the IROS Workshop on Semantic Perception for Mobile Manipulation, 2009.
5. A. Thomaz, "Socially guided machine learning," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.
6. 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.
7. P. Hauf, G. Aschersleben, and W. Prinz, "Baby do-baby see! How action production influences action perception in infants," Cognitive Development, vol. 22, no. 1, pp. 16-32, 2007.
8. P. Hauf and G. Aschersleben, "Action-effect anticipation in infant action control," Psych. Research, vol. 72, no. 2, pp. 203-210, 2008.
9. J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37-46, April 1960.
10. A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," The International Journal of Robotics Research, vol. 27, no. 2, pp. 157-173, 2008.
11. V. Sukhoy, J. Sinapov, L. Wu, and A. Stoytchev, "Learning to press doorbell buttons," in Proc. of ICML, 2010, pp. 132-139.

