

Tapping into Touch

Eduardo Torres-Jara*

Lorenzo Natale**

Paul Fitzpatrick*

*MIT CSAIL
Cambridge, Massachusetts, USA

**LIRA-Lab, DIST, University of Genova
Genova, Italy

Abstract

Humans use a set of *exploratory procedures* to examine object properties through grasping and touch. Our goal is to exploit similar methods to enable developmental learning on a humanoid robot. We use a compliant robot hand to find objects without prior knowledge of their presence or location, and then tap those objects with a finger. This behavior lets the robot generate and collect samples of the contact sound produced by impact with that object. We demonstrate the feasibility of recognizing objects by their sound, and relate this to human performance under situations analogous to that of the robot.

1. Introduction

Grasping and touch offer intimate access to objects and their properties. In previous work we have shown how object contact can aid in the development of haptic and visual perception (Natale et al., 2004, Metta and Fitzpatrick, 2003). We now turn our attention to audition: developing perception of contact sounds. Hearing is complementary both to touch and vision for contact sounds. Unlike touch, hearing doesn't require the robot to be the one causing the contact event. And unlike vision, hearing doesn't require line of sight – it won't be blocked by the arm, hand, or the object itself. We are motivated by an experiment we report in this paper, where human subjects successfully grasped objects while blindfolded. Several studies have revealed the importance of somatosensory input (force and touch); for example human subjects with anesthetized fingertips have difficulty in handling small objects even with full vision (Johansson, 1991). The extensive use of vision rather than haptic feedback in robotics may be due to technological limits rather than merit. The robotic hand used in this paper was designed to overcome these limitations. It is equipped with dense touch sensors and series elastic actuators which allow passive compliancy and to measure force at the joints. Force feedback and intrinsic compliance are exploited to successfully control the interaction between robot and environment without relying on visual feedback.

Humans use exploratory procedures in their perception of the world around them. This has in-

spired work on robotics. An analog of human sensitivity to thermal diffusivity was developed by (Campos et al., 1991), allowing a robot to distinguish metal (fast diffusion) from wood (slow diffusion). A robotic apparatus for tapping objects was developed by (Richmond and Pai, 2000) to characterize sounds so as to generate more convincing contact in haptic interfaces. In (Femmam et al., 2001), a special-purpose robot listens to sounds of the surface it “walks” on.

We use a *tapping* exploratory procedure, applied to natural objects by a general purpose, compliant hand (rather than a rigid, special purpose tapping device). Repetitive contact between the fingers and the object (the tapping behavior) allows the robot to collect information about the object itself (the sound produced by the collision of the fingers and the object surface) which is used for object recognition.

The paper is organized as follows: in Section 2. we present a motivating experiment with human subjects. Section 3. describes our robotic platform. Section 4. and 5. respectively detail the tapping experiment and the results we obtained. Section 6. discusses the results and places them on a broader perspective.

2. Simulating our robot with humans

Human haptic perception is impressive, even under serious constraint (Lederman and Klatzky, 2004). To get an “upper bound” of what we could expect from our robot, we evaluated human performance when wearing thick gloves that reduced their sensitivity and dexterity to something approaching our robot. We blocked their vision, since we know our robot cannot compete with human visual perception, but let them hear.

We sat 10 subjects in front of a padded desk covered with various objects – a wooden statue, a bottle, a kitchen knit, a plastic box, a paper cup, a desktop phone, a tea bag and a business card. The subjects wore a blindfold and a thick glove which reduced their haptic sensitivity and the number of usable fingers. The glove only allows them to use their thumb, their index and middle finger.

A goal of the experiment was to determine how much and in what way humans can manipulate unknown objects in an unknown environment with capabilities reduced to something approximating our robot. Here are a summary of our observations:



Figure 1: Subjects exploring a desk while blindfolded and wearing a thick glove. Top: light objects were inevitably knocked over, but the sound of their fall alerted the subjects to their presence, location, and (often) identity. Bottom: the sound of object placement was enough to let this subject know where the cup was and suggest a good grasp to use.

- ▷ Exploration strategies vary. Some subjects face their palm in the direction of motion, others towards the desk. The speed at which people swing their arm is generally slow and cautious, with occasional contact with the table.
- ▷ Very light objects were consistently knocked over.
- ▷ Subjects quickly reorient their hand and arm for grasping if either their hand or their wrist makes contact with an object.
- ▷ Subjects exhibited a short-term but powerful memory for object location.
- ▷ Sounds produced by objects and surfaces were used to identify them, compensating partially for the reduction in tactile sensitivity (see Figure 1). This was occasionally misleading: one subject unwittingly dragged a teabag over the desk, and thought from the sound that the surface was covered in paper.

Inspired by the last observation, in this paper we focus on exploiting the information carried by sound in combination with tactile and force sensing.

3. The robot Obrero

The humanoid robot used in this work, Obrero, consists of a hand, arm and head, shown in Figure 2. Obrero was designed to approach manipulation not as a task mainly guided by a vision system, but as one guided by the feedback from tactile and force sensing – which we call *sensitive manipulation*. We use the robot’s limb as a sensing/exploring device as opposed to a pure acting device. This is a convenient approach to operate in unstructured environments, on natural unmodeled objects. Obrero’s limb is sensor rich and safe – it is designed to reduce the risk of damages upon contact with objects.

The arm used in Obrero is a clone of a force-controlled, series-elastic arm developed for the robot

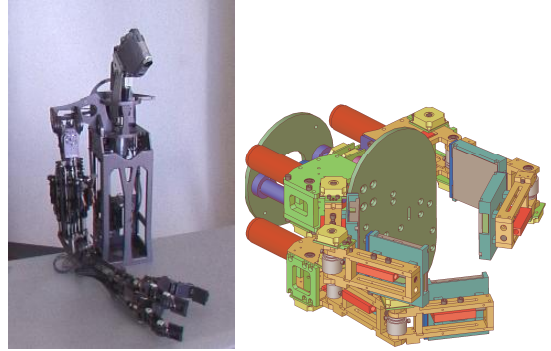


Figure 2: The robot Obrero (left) has a highly sensitive and force controlled hand, a single force controlled arm and a camcorder as a head (used simply as a microphone in this paper). Obrero’s hand (right) has three fingers, 8 DOF, 5 motors, 8 force sensors, 10 position sensors and 7 tactile sensors.

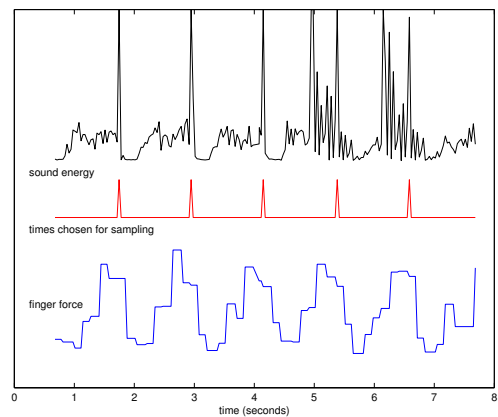


Figure 3: Force readings from the fingers (bottom) reveal when tapping may occur. Swings in the force are compared against sound intensity (top), looking for synchronized sounds. Peaks within one fifth of a period from a force swing are accepted. This process lets the robot filter out environmental sounds that occur when the arm is not moving, and even during tapping. In this example, the first three peaks of sound are clean, but the last two are corrupted by a phone ringing (see Figure 4).

Domo (Edsinger-Gonzales and Weber, 2004). The hand consists of three fingers and a palm. Each one of the finger has two links that can be opened and closed. Two of the fingers can also rotate. Each one of the joints of the hand is controlled using an optimized design for a series elastic actuator (Torres-Jara and Banks, 2004). Series elastic actuators reduce their mechanical impedance and provide force sensing (Williamson, 1995). Summary information about the hand is given in Figure 2.

4. The robot’s behavior

The behavior of the robot is very simple. It alternates between sweeping its hand back and forth over

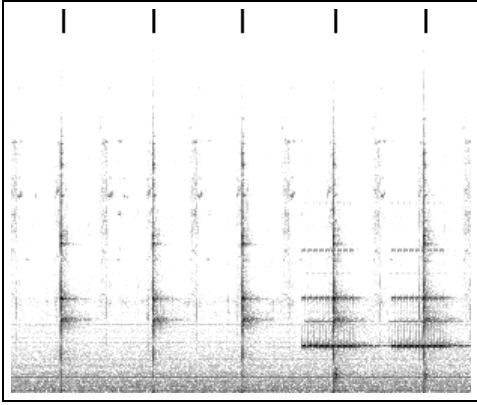


Figure 4: This is the spectrogram of the sounds in Figure 3 (time on the x-axis, increasing frequency on the y-axis, dark color corresponds to activity). The top of the spectrogram is marked to show the five sample times selected automatically. Between these times, there are patches of sound corresponding to the sound of springs in the fingers. The last two samples have the sound of a phone superimposed on them.

a table, and tapping any object it comes in contact with. Both the arm and hand are position controlled, but with the gains of the PD tuned so that, together with the springs in series with the motors, they realize a certain degree of compliance at the level of the joints. The hand in particular is extremely compliant; as a result during the “sweeping behavior” the fingers slightly bend upon contact with an object. Whenever a watchdog module detects such a collision it stops the arm and interrupts the sweeping movement. The hand then starts the “tapping behavior”; this is achieved by controlling the fingers with a periodic reference signal. The tapping lasts a few seconds, after which the arm is repositioned and the sweeping behavior reactivated.

During the experiment we recorded vision and sound from the head along with the force feedback from both the arm and hand. The visual feedback was *not used* in the robot’s behavior; it was simply recorded to aid analysis and presentation of results. All other streams were considered candidates for detecting contact. The force feedback from the hand proved the simplest to work with. Peaks in the hand force feedback were successfully employed to detect the impact of the fingers with the object during both the exploration and tapping behaviors. Force and sound were aligned as shown in Figure 3. Once the duration of a tapping episode was determined, a spectrogram for the sounds during that period was generated as shown in Figure 4. The overall contact sound was represented directly as the relative distribution of frequencies at three discrete time intervals after each tap, to capture both characteristic resonances, and decay rates. The distributions were pooled across all the taps in a single episode, and averaged. Recognition is performed by transforming these distributions

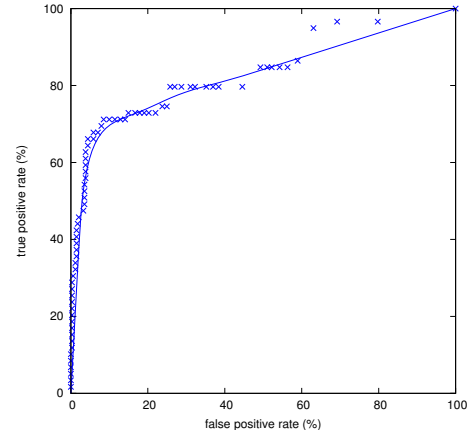


Figure 5: Receiver-operator characteristic curve. Tapping episodes from one day are matched against episodes from a previous day. Matches are ranked, then truncated based on a quality threshold. This plot shows the effect of that threshold on the trade-off between false matches and missed matches.

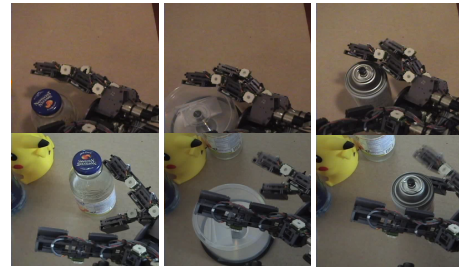


Figure 6: There were four overlapping objects in both the training and test run. Three of them were matched perfectly (using a best-match rather than threshold-based strategy) for every episode: a bottle (left), a spray-can (middle), and a CD case (right). Images on the bottom are from the test run, images on the top are from the best matching episode in the training run. These objects have quite distinctive sounds. A plastic toy (left corner of each lower image) failed to be recognized – it was quiet, and made just a dull thud.

into significance measures (how far frequency levels differ from the mean across all tapping episodes) and then using histogram comparison.

5. Results

We evaluated our work by performing an object recognition experiment. We exposed the robot one evening to a set of seven objects, and then in the morning tested its ability to recognize another set, which had an overlap of four objects with the training set.

Three of these objects were chosen (Figure 6) to represent three different materials, plastic, glass and steel (metal). The idea is that the sound produced by each object depends on its size, shape and the materi-

al with which it is made; accordingly we expected the tapping to produce three different distinct sounds. A fourth object (a plastic toy) was relatively silent.

For each run, we placed randomly selected objects on the table in front of the robot, and it was responsible for finding and tapping them. Overall the robot tapped 53 times; of these episodes 39 were successful, meaning that the sound produced by the tapping was significantly loud; in the other 14 cases the tapping did not provoke useful events either because the initial impact caused the object to fall, or the object remained too close to the hand. The high number of successful trials shows that given the mechanical design of hand and arm, haptic feedback was sufficient to control the interaction between the robot and the environment.

We evaluated the performance of our spectrum comparison method by ranking the strength of matches between episodes on the second day and episodes on the first day. Figure 5 shows what detection accuracy is possible as the acceptable false positive rate is varied. This predicts that we can on average correctly match an episode with 50% of previous episodes involving the same object if we are willing to accept 5% false matches.

6. Conclusions

We have demonstrated a compliant robot hand capable of safely coming into contact with a variety of objects without any prior knowledge of their presence or location – the safety is built into the mechanics and the low level control, rather than into careful trajectory planning and monitoring. We have shown that, once in contact with these objects, the robot can perform a useful exploratory procedure: tapping. The repetitive, redundant, cross-modal nature of tapping gives the robot an opportunity to reliably identify when the sound of contact with the object occurs, and to collect samples of that sound. We demonstrated the utility of this exploratory procedure for a simple object recognition scenario.

This work fits in with a broad theme of learning about objects through action that has motivated the authors' previous work (Fitzpatrick et al., 2003). We wish to build robots whose ability to perceive and act in the world is created through experience, and hence robust to environmental perturbation. The innate abilities we give our robots are not designed to accomplish the specific, practical, useful tasks which we (and our funders) would indeed like to see, since direct implementations of such behaviors are invariably very brittle; instead we concentrate on creating behaviors that give the robot robust opportunities for adapting and learning about its environment. Our gamble is that in the long run, we will be able to build a more stable house by building the ground floor first, rather than starting at the top.

Acknowledgements

This project makes heavy use of F/OSS software – thank you world. This work was partially funded by ABB, by DARPA DABT 63-00-C-10102, and by NTT under the NTT/MIT Collaboration Agreement. Lorenzo Natale was supported by the European Union grant RobotCub (IST-2004-004370).

References

- Campos, M., Bajcsy, R., and Kumar, V. (1991). Exploratory procedures for material properties: the temperature perception. In *5th Int. Conf. on Advanced Robotics*, volume 1, pages 205–210.
- Edsinger-Gonzales, A. and Weber, J. (2004). Domo: A force sensing humanoid robot for manipulation research. In *Proc. of the IEEE International Conf. on Humanoid Robotics*, Los Angeles.
- Femmam, S., M'Sirdi, N. K., and Ouahabi, A. (2001). Perception and characterization of materials using signal processing techniques. *Instrumentation and Measurement*, 50(5):1203–1211.
- Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. (2003). Learning about objects through action - initial steps towards artificial cognition. In *Proc. of the IEEE Internat'l Conf. on Robotics and Automation*, Taipei, Taiwan.
- Johansson, R. S. (1991). How is grasping modified by somatosensory input? In Humphrey, D. R. and Freund, H. J., (Eds.), *Motor Control: Concepts and Issues*, pages 331–355. John Wiley and Sons Ltd, Chichester.
- Lederman, S. J. and Klatzky, R. L. (2004). Haptic identification of common objects: Effects of constraining the manual exploration process. *Perception and Psychophysics*, 66:618–628.
- Metta, G. and Fitzpatrick, P. (2003). Early integration of vision and manipulation. *Adaptive Behavior*, 11(2):109–128.
- Natale, L., Metta, G., and Sandini, G. (2004). Learning haptic representation of objects. In *International Conference on Intelligent Manipulation and Grasping*, Genoa, Italy.
- Richmond, J. L. and Pai, D. K. (2000). Active measurement and modeling of contact sounds. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 2146–2152, San Francisco.
- Torres-Jara, E. and Banks, J. (2004). A simple and scalable force actuator. *35th International Symposium on Robotics*.
- Williamson, M. (1995). Series elastic actuators. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.