Object Categorization in the Sink: Learning Behavior–Grounded Object Categories with Water

Shane Griffith, Vladimir Sukhoy, Todd Wegter, and Alexander Stoytchev

***Abstract*— This paper explores whether auditory and proprioceptive information can be used to bootstrap learning about how objects interact with water. Our results demonstrate that a robot can categorize objects into "containers" and "non-containers" based on how the objects sound like and feel like when water is flowing onto them. Using a behavior–grounded approach, the robot performed five different exploratory behaviors on the objects and captured auditory and proprioceptive data as the behaviors changed the spatial configuration between the objects and the water stream. Using this data, the robot first learned perceptual outcome classes for each behavior–modality combination. Functionally meaningful object categories were then formed based on the frequency with which different outcome classes occurred with each object.**

I. INTRODUCTION

Humanoid robots and water don't play well together. The last place researchers in this field expect to see their expensive equipment is in the sink when the water is running. Yet, water manipulation is an important domain for robots operating in human environments. Water is used for many universal activities, including cooking, cleaning, and gardening. Therefore, serious consideration is needed to address the fundamental questions associated with water manipulation.

Water manipulation tasks almost always involve the use of a container. It is not easy, however, to determine what properties of an object make it a container. Part of the difficulty stems from the fact that there are literally thousands of objects that can act as cups under the right circumstances. Another difficulty is the fact that visual information alone is not sufficient to identify all cups. For some objects, e.g., a colander, one has to pour water into them to find out.

In the context of robotics, this paper addresses this question in terms of the multimodal sensorimotor properties of the objects as the robot actively changes their spatial configuration relative to a stream of running water. This allows the robot to learn embodied representations that are extracted from the robot's own experience with water. For example, pouring water changes the weight of the object, which the robot can sense through proprioception. The robot can also detect how the pitch of the sound changes as the cup is filled up with water or as the water begins to overflow and hit the sink basin.

This work was supported in part by a National Science Foundation Graduate Research Fellowship to the first author (NSF Grant No. 0751279). Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors are with the Developmental Robotics Laboratory, Department of Electrical and Computer Engineering, Iowa State University.
{shaneg, sukhoy, twegter, alexs}@iastate.edu

Recent research in robotics that focuses on interactively learning about objects indicates that, in addition to video, audio and proprioception are major sources of information [1]. The different shapes, sizes, and materials of objects all affect how an object sounds and feels to a robot, which can be used for recognition [1] and categorization tasks [2]. It is unclear, however, whether these modalities would also be useful during tasks that involve water.

This is one of the first papers that tests the hypothesis that a robot can learn meaningful object categories using audio and proprioception when the interaction tasks involve water. A humanoid robot performed five different exploratory behaviors on 15 different objects in a sink with running water. The robot's observations for a given behavior–modality combination were first clustered to form outcome classes. The frequency with which each outcome class occurred with each object was used to form object categories. Because an object category was learned for each behavior–modality combination, the resulting categories were unified to form a single one. The results showed that sound captured size differences, proprioception captured weight differences, and when combined the unified categorization captured functional differences (e.g., container or non-container).

II. RELATED WORK

Recent research has addressed interactive object categorization, but few studies have tried to categorize objects during interactions with water. Aksoy *et al.* [3] presented an activity recognition framework that could also be used for object categorization. Activities were represented using a sequence of spatial–temporal features extracted from videos. Objects were categorized by the type of activities a human performed with them. During one of the activities a human poured dark–colored tea from one cup into another one. Cups were distinguished from other objects because they were the only objects used during pouring behaviors.

Several studies have addressed interactive object categorization using sensory modalities other than vision. Sinapov *et al.* [4] showed that a robot could identify the different shape, size, and material categories of objects using acoustic object recognition models. In the work of Nakamura *et al.* [5], a robot learned a multi-modal object categorization using perceptual features extracted from the specific behavior–modality combinations *rotate–vision*, *shake–audio*, and *squeeze–tactile*. Vision and audio were used in the multimodal object categorization framework [2], in which a robot learned from the functional properties of objects extracted using many different exploratory behaviors.

Object manipulation is an important area of research in the area of underwater robotics [6][7]. Many aspects of operating a robotic manipulator change when it is under water, including control algorithms and communication problems. Applying the correct amount of force to an object during underwater manipulation tasks is also challenging. Liang *et al.* [8] built a fingertip force sensor to address this problem. A manipulator using their sensor can receive feedback about the amount of force it applies to an object during underwater manipulation tasks.

Transporting an open container without sloshing the liquid inside it is a particularly hard control problem that remains an active area of research. This problem is usually solved by formulating complex control algorithms [9][10][11]. Many of these algorithms, however, are designed using the precise size and shape of the container that is being controlled. As a result, they have to be redesigned when the setup changes.

As far as we know, computational modeling of water is not easy. It is difficult to avoid ad hoc methods for water detection when only vision is used (see, for example, [12][13][14][15]). Furthermore, it is hard to identify new ways to perceive water when plumbing for water is not installed in most research labs. Thus, most interaction tasks with water have been limited to pouring water from one container into another container or otherwise are imitations of that activity (see, for example, [16][17][18][19][20]), with little regard for the perception of water. Research has yet to identify a good source of information that can be used for tasks that involve interacting with water.

This paper shows that audio and proprioception can be used for learning object categories while a robot interacts with objects in a sink.

III. EXPERIMENTAL SETUP

A. Robot

The experiments described in this paper were performed using an upper-torso humanoid robot. The robot's arms are two Whole Arm Manipulators (WAMs) from Barrett Technology. Each arm is equipped with a Barrett Hand as its end effector. A microphone in the robot's head was used to capture sound at 44.1 KHz over a 16-bit mono channel.

The body of the robot was covered to protect it from water. The left hand and forearm were protected with a clear *Waterguard Cast and Skin Protector* [21], which is typically used by people with broken arms to protect their cast when they take a shower. The rest of the body, except the head, was protected with transparent rain ponchos that were held in place with clear tape.

B. Sink

Figure 1 shows the standard utility sink that was used in the experiments. The sink was assembled from the *Foremost "All in One Box Laundry Tub"* sink kit, which was purchased from Lowe's (a home improvement store). The sink fixture contains a combination faucet.

Because there were no water lines in the laboratory that hosts the robot, a five–gallon bucket was used as a water supply. A *SmartPond* water pump was used to recirculate the water between the bucket and the faucet. The pump's flow

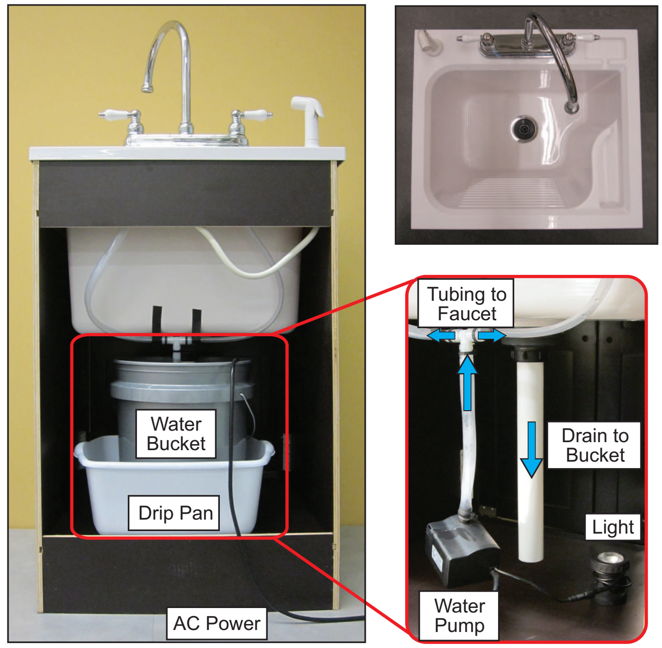

Fig. 1. The self–contained sink that was used in the robotic experiments. **(Left)** A five–gallon bucket held a reservoir of water that was pumped up to the faucet. The water was collected back into the bucket when it flowed down the drain. A plastic drip pan underneath the bucket collected excess drops of water. **(Top Right)** View of the 13-inch-deep sink from the top. **(Bottom Right)** Close–up of the water pump and the plumbing inside the bucket. Blue arrows indicate the direction of the water flow. The light was not necessary, but it could not be detached from the pump.

was directed through half-inch inner diameter vinyl tubing to both inputs simultaneously by splitting the water line with a plastic T-junction. The diverter valve for the sprayer inside the base of the faucet was removed in order to achieve the maximum amount of flow from the pump, which can move up to 300 gallons of water per hour and has a maximum vertical lift of 6.6 ft. In other words, the plumbing for the sink was entirely self-contained.

C. Objects

The robot interacted with the 15 objects shown in Fig. 2. The objects could hold water in one orientation, but became non-containers when flipped over. The objects (cups and mugs) varied by their size (small, medium, and large) and their material type (plastic, metal, paper, ceramic, and glass). Because the plastic glove that covered the robot's hand reduced the grasp friction, objects were chosen only after we determined that the robot could securely grasp them.

D. Exploratory Behaviors

The experiments were divided into trials. Before the start of a trial, the robot moved its left hand to a location near the edge of the sink, where an experimenter placed one of the 15 objects in its hand. During each trial, the robot performed a sequence of behavioral interactions with each object in the sink while the water was flowing. The five behaviors were: *hold*, *flip*, *up and down*, *shake*, and *in and out* (see Fig. 3). Each behavior was performed twice, with a *flip* between each execution, before moving on to the next behavior in the sequence. In other words, the object was in the container orientation during the first execution of the behavior and in

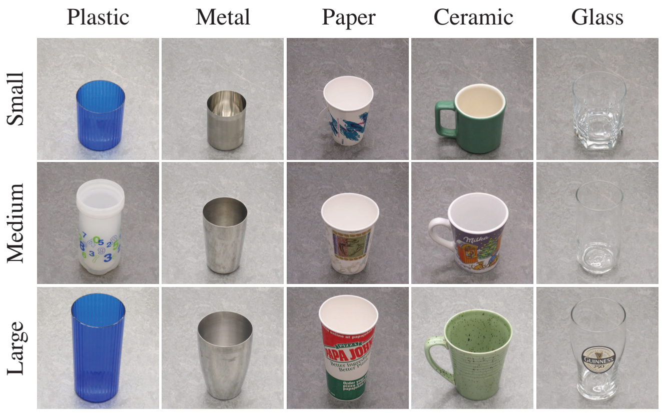

Fig. 2. The objects used in the experiments. The set included objects from five different material types, three different sizes, and various weights. The objects could hold water in one orientation, but became non-containers when flipped over.

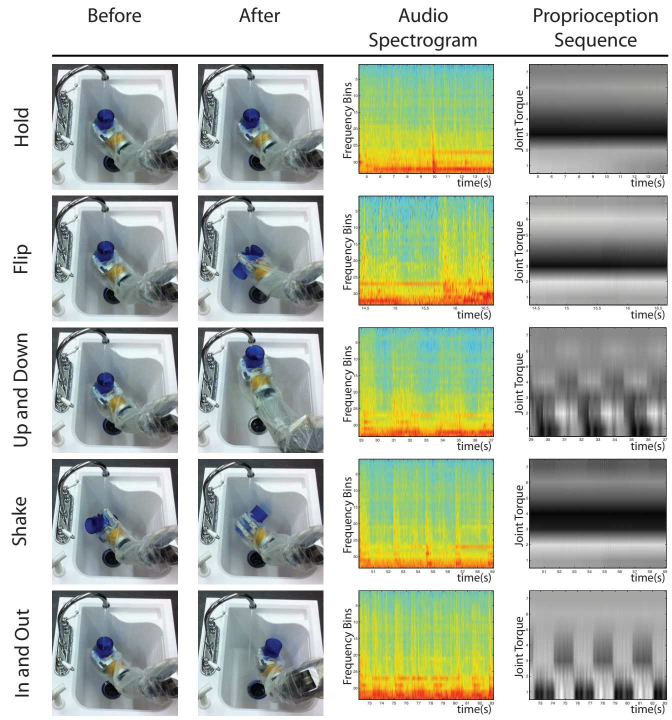

Fig. 3. The five behaviors performed by the robot with their resulting audio spectrograms and sequences of joint torque readings.

the non-container orientation during the second execution of the behavior. Altering the object pose from container to non-container and back ensured that the outcomes that occurred during one behavior were independent from the outcomes that occurred during previous behaviors, because the water was poured out between behaviors. The individual behaviors are described in detail below.

Hold: At the start of this behavior the robot moved its hand directly under the water stream while holding the object. It remained in that configuration for approximately 10 seconds. This duration was sufficient to fill up halfway the largest plastic cup (i.e., the large blue cup in Fig. 2).

Flip: The robot rotated its wrist by 180 degrees, flipping the orientation of the object. Although this behavior was performed several times during the interaction sequence, only the *flip* following the *hold* behavior was used for learning. During this execution of the *flip* behavior, many of the containers were full with water due to the duration and the nature of the *hold* behavior. The *flip* behavior lasted approximately 2 seconds.

Up and Down: The robot moved the object up and down under the water stream during this behavior. The object started near the bottom of the sink and ended just below the tip of the faucet. This movement was repeated four times, ending at the starting position. The entire behavior lasted roughly 8 seconds.

Shake: This behavior was performed by shaking the object back and forth under the water. The back and forth motion was repeated three times during this behavior. The whole behavior took about 10 seconds to perform.

In and Out: This behavior consisted of moving the object in and out of the water flow four times. It lasted 10 seconds.

IV. METHODOLOGY

A. Data Collection

Multiple sequences of audio and proprioception data were collected during each trial. The robot collected one data sequence per modality for each of the five exploratory behaviors (*Hold*, *Flip*, *Up and Down*, *Shake*, *In and Out*). The robot performed the 5 behaviors 10 times on each of the 15 objects when they were in each of the 2 object poses. In other words, a total of $5 \times 10 \times 15 \times 2 = 1500$ behavioral interactions were performed. The outcome sequences varied in length according to the duration of each behavior as described in the previous section. It took 6 hours to collect this data. The plastic glove protecting the robot's hand was replaced halfway through the data collection process because a few micro-holes developed in the plastic, which allowed a very small amount of water to get through.

During the i^{th} behavioral interaction, the robot acquired a data point of the form (B_i, O_i, W_i, T_i), where $B_i \in \mathcal{B}$ was one of the five exploratory behaviors, $O_i \in \mathcal{O}$ was one of the 30 objects (15 objects $\times$ 2 orientations), W_i was the audio waveform, and T_i was the sequence of joint torque readings. The audio sequence, W^i, was sampled at 16-bit/44.1 KHz in mono and stored as a wave file. The sequence of joint torque readings, T^i, consisted of joint torque readings for each of the seven joints of the robot's left arm, recorded at 500 Hz and stored as a text file. Another way to look at this data set is as a collection of 300 wave files and 300 joint torque sequences for each behavior.

B. Auditory Feature Extraction

We used the auditory feature extraction pipeline and the publicly available source code that are described in [1]. They are briefly summarized below. The three stage process includes: 1) Transforming each 44.1 KHz, 16-bit single channel wave file into a spectrogram $S_i = s_1^i s_2^i, \ldots s_{li}^i$, where $s_j^i \in \mathbb{R}^{33}$, using the Discrete Fourier Transform (with a window length of 25.625 ms and an overlap of 10 ms); 2) Training a 6×6 SOM using a subset of the column vectors, s_j^i, from all of the spectrograms for the behavior.

In this case, 5% of the total number of column vectors were used during the training procedure; 3) Converting each spectrogram $S_i = s_1^i s_2^i, \ldots, s_{l^i}^i$ into a state sequence $A_i = a_1^i a_2^i, \ldots a_{l^i}^i$ by mapping the column vector, s_j^i, to the most highly activated node, a_j^i, in the SOM. Audio was converted into state sequences using a separate SOM for each behavior.

C. Extracting Proprioceptive Features

The proprioception feature extraction process is similar to the one used for audio. A sequence of joint torques is represented as a state sequence of the most highly activated nodes in a Self–Organizing Map (SOM). During this conversion, the proprioception data is reduced from seven–dimensional numeric column vectors to two–dimensional nominal states. Because the proprioception values are already in a column vector format, the feature extraction process has only two stages in this case: 1) Training a 6×6 SOM using 5% of the collected column vectors that have the form $t_j^i \in \mathbb{R}^7$; 2) Mapping column vectors to the most highly activated node of the trained SOM to extract a state sequence, $P_i = p_1^i p_2^i, \ldots, p_{n^i}^i$, from the joint torque sequence $T_i = t_1^i t_2^i, \ldots, t_{n^i}^i$. One SOM was trained per behavior.

D. Outcome Class Learning

Object properties are indirectly captured in the outcome state sequences, which can be clustered to form meaningful outcome classes. For example, information about the shape, the height, and the material of the small metal container is detectable in the audio recording of the robot holding the object under the faucet. Water falling on metal makes a unique sound. The sound of water filling up a container and then spilling over the top can also be heard. Similarly, information about the weight of the same object is present in the proprioception sequences in which the robot flipped it over. The joint torques change as water is poured out because the small metal object is fairly light to begin with.

Clearly, a robot may observe many different types of outcomes as it interacts with the objects. Presumably, it can directly identify the different types of outcomes that it observes by clustering them into outcome classes. In other words, given acoustic state sequences $\{A_i\}_{i=1}^{300}$ or proprioception state sequences $\{P_i\}_{i=1}^{300}$ for a given behavior, the task of the robot is to identify outcome classes $C_1, \ldots, C_k$. For this task we use the Spectral Clustering hierarchical clustering algorithmm [22]. The algorithm takes as input a similarity matrix, determines the correct number of clusters k, and outputs the outcome classes $C_1, \ldots, C_k$ that it finds. The similarity matrix is created by measuring the similarity between all pairs of state sequences using the Needleman–Wunsch Global String Alignment Algorithm [23], one of several ways to compute the similarity between two strings. In this case, the penalty for mismatched tokens during the string alignment process is specified by the Euclidean distance between their corresponding nodes in the SOM.

The spectral clustering algorithm recursively bi-partitions the similarity matrix using the Shi–Malik Normalized Cut objective function [24]. The recursion stops when a termination criterion is reached. The leaf nodes of the tree formed by the algorithm are the outcome classes $C_1, \ldots, C_k$.

E. Object Categorization

Some outcome classes occur more often with certain object categories compared to others, and this difference is used to form object categories. For example, the sound of pouring can be heard more often during the *flip* behavior when holding a cup than when holding a non-container. Similarly, the change in weight can be detected more often during this behavior with a cup compared to a non-container. In other words, the distribution of outcome classes that occur with a cup is probably different from those that occur with a non-container. The robot uses these differences in the outcome class distribution of objects in order to cluster the objects into object categories.

Given a set of outcome classes, $C_1, \ldots, C_k$, the robot acquired feature vectors $Z_1, \ldots, Z_{30}$, where each $Z_i = z_1^i, \ldots, z_k^i$ and z_j^i is the frequency with which outcome class C_j occurred with object O_i. Thus, the feature vector Z_i estimates a probability distribution of how likely each outcome class is to occur with object O_i. A negligible amount of noise ϵ was added to z_j^i in cases when it was polarized to 0 or 1. The feature vectors $Z_1, \ldots, Z_{30}$ were passed to the X-means [25] unsupervised clustering algorithm (with the desired number of clusters, k, varying from 2 to 10) in order to form object categories. X-means extends the K-means clustering algorithm by automatically estimating the correct number of clusters. The result of this process is an object category labeling $\lambda^{(u)} = l_1, \ldots, l_{30}$ of the objects for the u^{th} behavior–modality combination.

F. Unified Categorization

Because different behaviors and sensory modalities capture very different information about the objects, the categorizations formed using them can be very different as well. Some of the resulting categorizations are more meaningful than others. Presumably, however, there is a consensus among the individual categorizations about a good way to categorize the objects. The unified categorization step identifies this consensus clustering.

The overall goal is to identify a unified categorization, $\hat{\lambda}$, that is representative of the multiple input categorizations $\lambda^{(1)}, \ldots, \lambda^{(m)}$. The best unified clustering of the input clusterings maximizes the Normalized Mutual Information, i.e., $\mathrm{argmax}_{\hat{\lambda}} \sum_{u=1}^{m} \o^{NMI}(\hat{\lambda}, \lambda^{(u)})$ [26]. Solving this system for the best unified clustering is NP-Hard. It is possible, however, to search for a good approximation of the best clustering. The Strehl and Ghosh Ensemble Clustering Algorithm [26] was used for this task. The algorithm takes as input the object categorizations $\lambda^{(1)}, \ldots, \lambda^{(m)}$ and a desired number of clusters k (which was varied between 2 and 10), and then returns the best found ensemble clustering, $\hat{\lambda}$.

V. RESULTS

The individual categorizations formed by the robot varied in quality based on the behavior and the sensory modality used to produce them. Overall, sound captured differences primarily in size and then function. An example of the categorization formed for the *In and Out* behavior for audio is shown in Fig. 4. In this specific case, most of the small

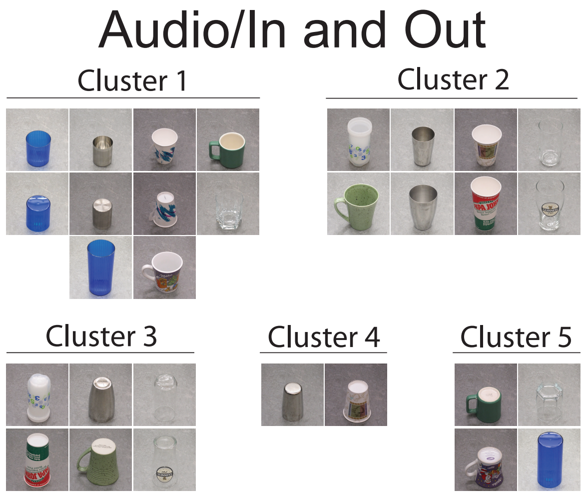

Fig. 4. The object categorization produced using the sound captured during the *In and Out* behavior. The categorization captured differences in the size (i.e., cluster 1 has mostly small objects) and the function (i.e., cluster 2 consists of the remaining containers while clusters 3-5 have only non-containers) of the objects.

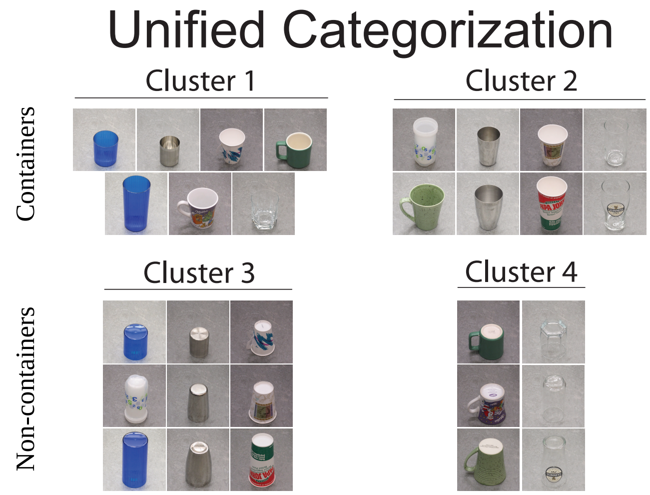

Fig. 5. The unified categorization produced using the ten individual categorizations formed by the robot. Clusters 1 and 2 represent containers. Clusters 3 and 4 represent non-containers. The categorization also captured differences in the function (all four clusters), the size (clusters 1 and 2), and the weight (clusters 3 and 4) of the objects.

objects belong to the same category. The other categories merged the medium and the large–sized objects, but captured the secondary differences in the function of the objects. Thus, sound was most suitable for discriminating between objects with different sizes.

Similarly, proprioception captured differences primarily in weight and then function. For example, the categorization formed for the *In and Out* behavior placed the heavy objects (ceramic mugs and glass cups) in their own category. The secondary differences in the functions of the objects were only captured in two other object categories.

The fact that sound captured differences in size and that proprioception captured differences in weight is observable in most of the individual categorizations that were used to form the unified clustering. Fig. 5 shows that the unified clustering also distinguished between objects of different sizes and weights. However, the functional difference between containers and non-containers is observed just as strongly. In fact, when the number of clusters was forced to two, the unified categorization perfectly separated the containers from the non-containers (i.e., the first two clusters in Fig. 5 were merged into the "cups" cluster and clusters 3 and 4 were merged into the "non cups" cluster). Thus, the unified clustering was most suitable for discriminating between objects with different functions.

To test the robustness of the unified object categorization, the framework described above was also evaluated using ten different permutations of the data collected by the robot. This procedure produced ten different unified categorizations. The information gain of each unified categorization was computed with respect to a human labeling of the objects along four different dimensions (function, size, weight, and material). For comparison with a baseline value, the average random information gain was also computed along those dimensions. The category labels from the human labeling were shuffled and then the information gain was computed with respect to the original human labeling. This procedure was performed 100 times for each of the four dimensions.

Figure 6 shows the results of the procedure. The learning framework captured a significant amount of information about the objects in terms of their function, size, and weight. The unified object categorization was most affected by differences in the container property, followed by differences in weight and then size. The size dimension had less information gain because the medium and the large–sized objects were usually categorized together. They did not overflow with water, but the small objects did. The framework captured little information about the material properties of the objects, which is why the information gain for that dimension is similar to the average random information gain.

The non-zero error bars in Fig. 6 indicate that the unified clusterings changed when the data was shuffled. In other words, the order–dependent clustering algorithms produced a distribution of different categorizations. The ten different unified categorizations ranged from having three categories to six categories. However, the graph clearly shows that the unified clusterings still pick up meaningful results in terms of the functional properties of the objects. The information gained with respect to the functional properties of the objects is resilient to fluctuations in the ordering of the data. In fact, when forced to output only two clusters the unified clustering perfectly separates the containers from the non-containers in 9 out of 10 cases (an object was misclassified as a non-container in 1 case).

VI. CONCLUSION AND FUTURE WORK

This paper showed that sound and proprioception are important sources of information during water manipulation tasks. Sound consistently captured information about the size and the function of the objects. Proprioception consistently captured information about the weight and the function of the objects. Furthermore, when the information from the two modalities was combined it was possible to form even more accurate object categories. The unified object categorization

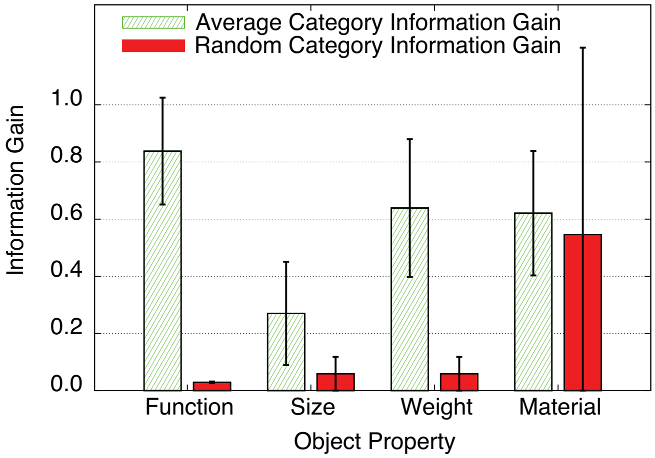

Fig. 6. The average category information gain of the unified object categorizations with respect to four different object properties (function, size, weight, and material). The information gain values were computed using a human labeling for each of the four dimensions. The average and standard deviation of the information gains were computed using the unified object categorizations from 10 different executions of this learning framework. The random category information gain is shown for comparison.

was meaningful with respect to the function, the size, and the weight of the objects.

The results begin to answer the question: "What is a cup?" For our robot, and for the current experimental setup, a cup is an object that sounds and feels in a specific way as the water flows into it. Objects that are not cups sound and feel differently. Thus, using this unsupervised approach, the robot has the ability to autonomously extract and attach symbolic labels to the clusters of objects that a human would call 'containers' or 'non-containers'.

Another key contribution of this paper is the idea that learning about water can be bootstrapped using auditory and proprioceptive data in the absence of visual and tactile information. Obviously, our robot does not know that the water feels "wet," as it does not have tactile or any sort of skin sensors the way humans do. Nevertheless, the robot "knows" how the objects from the "cups" cluster sound and feel when they are placed under the water stream in the sink. The robot also "knows" how these auditory and proprioceptive properties change as the robot actively varies the position of the object using its own behaviors. This embodied sensorimotor representation can be particularly advantageous when addressing water manipulation problems.

Future work should continue to explore interaction with liquids. Learning how objects and water interact with one another is essential for manipulating liquids more effectively. However, there are many more hurdles to overcome before robots can fully take advantage of water during object manipulation tasks. The message of this paper is that sound and proprioception might be able to bootstrap this research.

VII. ACKNOWLEDGMENT

The authors would like to acknowledge Chris Atkeson from Carnegie Mellon for specifically recommending that we "do something with water" to clearly demonstrate one advantage of our behavior–grounded learning approach.

REFERENCES

[1] J. Sinapov, T. Bergquist, C. Schenck, U. Ohiri, S. Griffith, and A. Stoytchev, "Interactive object recognition using proprioceptive and auditory feedback," *IJRR*, vol. 30, no. 10, pp. 1250–1262, Sep 2011.

[2] S. Griffith, J. Sinapov, V. Sukhoy, and A. Stoytchev, "A behavior–grounded approach to forming object categories: Separating containers from non-containers," *IEEE TAMD*, vol. 4, no. 1, pp. 54–69, 2012.

[3] E. E. Aksoy, A. Abramov, J. Drr, K. Ning, B. Dellen, and F. Worgotter, "Learning the semantics of object–action relations by observation," *IJRR*, vol. 30, no. 10, pp. 1229–1249, Sep 2011.

[4] J. Sinapov and A. Stoytchev, "From acoustic object recognition to object categorization by a humanoid robot," in *Proc. RSS Workshop: Mobile Manipulation in Human Environments*, Seattle, WA, June 2009.

[5] T. Nakamura, T. Nagai, and N. Iwahashi, "Multimodal object categorization by a robot," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2007, pp. 2415–2420.

[6] J. Yuh, "Design and control of autonomous underwater robots: A survey," *Autonomous Robots*, vol. 8, no. 1, pp. 7–24, 2000.

[7] G. Antonelli, *Underwater Robots: Motion and Force Control of Vehicle-Manipulator Systems*, 2nd ed., B. Siciliano, O. Khatib, and F. Groen, Eds. Springer Tracts in Advanced Robotics, Springer-Verlag, Heidelberg, D, June 2006.

[8] Q. Liang, D. Zhang, Q. Song, and Y. Ge, "A potential 4-D fingertip force sensor for an underwater robot manipulator," *IEEE Control Systems Magazine*, vol. 35, no. 3, pp. 574–583, July 2010.

[9] J. Feddema, C. Dohrmann, G. Parker, R. Robinett, V. Romero, and D. Schmitt, "Control for slosh-free motion of an open container," *IEEE Control Systems Magazine*, vol. 17, no. 1, pp. 29–36, February 1997.

[10] V. Romero and M. Ingber, "A numerical model for 2-D sloshing of pseudo-viscous liquids in horizontally accelerated rectangular containers," in *Proc. of the 17th ICBE*, 1995.

[11] M. Tzamtzi, F. Koumboulis, and N. Kouvakas, "A two stage robot control for liquid transfer," in *IEEE ETFA*, Sept. 2007, pp. 1324–1333.

[12] M. Iqbal, O. Morel, and F. Meriaudeau, "A survey on outdoor water hazard detection," in *The 5th Intl. Conf. on Information and Communication Technology and Systems (ICTS)*, Indonesia, 2009.

[13] A. Rankin and L. Matthies, "Daytime water detection based on color variation," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robotics and Systems (IROS)*, Taipei, Taiwan, 2010, pp. 215–221.

[14] A. Rankin, L. Matthies, and P. Bellutta, "Daytime water detection based on sky reflections," in *IEEE ICRA*, May 2011, pp. 5329–5336.

[15] K. Okada, M. Kojima, S. Tokutsu, Y. Mori, T. Maki, and M. Inaba, "Integrating recognition and action through task-relevant knowledge for daily assistive humanoids," *Advanced Robotics*, vol. 23, no. 4, pp. 459–480, 2009.

[16] C. C. Kemp, "Wearables and robots: A shared view," *IEEE Pervasive Computing*, vol. 5, no. 3, pp. 16–20, July-September 2006.

[17] K. Kim, J. Lee, S. Kim, J. Lee, M. Jeong, C. Kim, and B. You, "Coordinated task execution by a humanoid robot," in *ICROS-SICE International Joint Conference*, Fukuoka International Congress Center, Japan, August 2009, pp. 3496–3503.

[18] M. Cakmak and A. Thomaz, "Active learning with mixed query types in learning from demonstration," in *Proc. of the ICML Workshop on New Developments in Imitation Learning*, Bellevue, WA, 2011.

[19] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in *Proc. of the IEEE ICRA*, Kobe, Japan, May 2009, pp. 763–768.

[20] W. Hwang and J. Weng, "Vision-guided robot manipulator control as learning and recall using SHOSLIF," in *Proc. of the IEEE ICRA*, Albuquerque, NM, April 1997, pp. 2862–2867.

[21] WaterGuard. 2011. http://www.waterguard.biz/.

[22] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[23] S. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, 1970.

[24] J. Shi and J. Malik, "Normalized cuts and image segmentation," *PAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[25] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proc. of the 17th ICML*, 2000, pp. 727–734.

[26] A. Strehl and J. Ghosh, "Cluster ensembes - a knowledge reuse framework for combining multiple partitions," *JMLR*, vol. 3, pp. 583–617, 2002.