Ugur, E., Nagai, Y., Oztop, E., and Asada, M. (Eds) Proceedings of
Humanoids 2012 Workshop on Developmental Robotics: Can
developmental robotics yield human-like cognitive abilities?
November 29, 2012. Osaka, Japan

The Object Pairing and Matching Task: Toward Montessori Tests for Robots

Connor Schenck and Alexander Stoytchev
Developmental Robotics Laboratory
Iowa State University
{cschenck, alexs}@iastate.edu

***Abstract*—The Montessori method is a popular approach to education that emphasizes student-directed learning in a controlled environment. Object matching is one common task that children perform in Montessori classrooms. Matching tasks also occur quite frequently on intelligence tests for humans, which suggests that intelligence correlates with the skills required to solve these tasks. This paper describes robotic experiments with four Montessori matching tasks: sound cylinders, sound boxes, weight cylinders, and pressure cylinders. The robot grounded its representation for the twelve objects in each task in terms of the auditory and proprioceptive outcomes that they produced in response to a set of ten exploratory behaviors. The results show that based on this representation, it is possible to identify task-relevant sensorimotor contexts (i.e., exploratory behavior and sensory modality combinations) that are useful for performing matching on a given set of objects. Furthermore, the results show that as the number of sensorimotor contexts used to perform matching increases, the robot's ability to match the objects also increases.**

Fig. 1. The robot and the four Montessori matching tasks that were used in the experiments. In clockwise order, the four tasks were: sound cylinders, weight cylinders, pressure cylinders, and sound boxes.

I. Introduction

The Montessori method is a 100-year-old method of schooling that was developed by Maria Montessori (1870-1952), an influential Italian educator. It is characterized by a special set of educational materials and student-directed learning activities [1] [2] [3]. One of its core principles is that of *embodied cognition*, tying movement of the body and learning together. It focuses on stimulating the development of different skill sets, including sensory development, language development, and numeracy skills. Most Montessori tasks require that the children actively touch, move, relate, and compare objects [2].

One task typical for a Montessori classroom is object matching. Children are given two sets of objects and asked to find the matches from one set to another. Sample tasks include matching colored tiles, matching 3-dimensional shapes, and matching pieces of textured cloth [4]. All these tasks are designed to stimulate a child's ability to perceive object properties and to allow the child to learn about the nature of objects and their similarities.

The skills required to perform matching are also useful for other tasks such as object grouping, category recognition, and object ordering. At a fundamental level, these skills require the ability to find differences between similar objects and similarities between different objects. Recent work in robotics has found that robots are able to recognize objects and their categories [5], [6], group objects in an unsupervised manner [7], and find the odd one out in a set of objects [8]. These studies all strongly suggest that a robot should be able to solve object pairing tasks.

This paper describes a method that allows a robot to identify and match object pairs within a set of objects based on their sensorimotor properties. To do this, the robot first interacted with the objects using a set of exploratory behaviors (grasp, lift, hold, shake, rattle, drop, tap, poke, push, and press) in order to ground the properties of the objects in the robot's behavioral repertoire. After interacting with the objects, the robot performed feature extraction on the raw sensory data to create sensory feedback sequences for each interaction. For each object, the robot recorded both proprioceptive feedback in the form of joint torques and auditory feedback in the form of an audio spectrogram. Next, the robot generated similarity scores for all possible object pairs and used these scores to match the objects. To combine information from different sensorimotor contexts (e.g., *audio-drop* and *proprioception-shake*), the robot used three different methods: uniform-weight combination, recognition accuracy based weight combination, and pairing accuracy based combination. These methods were evaluated for their ability to match standard Montessori objects.

This study used four typical Montessori matching tasks. In each task there were two groups of six objects and the goal was to find the matching pairs of objects between the two groups. The results indicate that the estimated object similarities were sufficient to adequately pair objects. The robot was able to solve the object matching task with a high degree of accuracy. Furthermore, the robot was able to identify the functionally meaningful sensorimotor contexts in which it can distinguish between objects. To the best of our knowledge, this is the first study that has applied Montessori learning techniques in a robotic setting.

II. Related Work

A. Psychology

Recent studies have found that students educated using the Montessori method often outperform students educated by traditional methods. For example, one study found that middle school students from Montessori schools had higher intrinsic motivation when it came to academic activities as compared to students from traditional schools [9]. This suggests that the Montessori method is more effective at fostering learning in young children than the traditional methods. This conclusion was supported by another study [3], which found that, by the end of kindergarten, Montessori students outperformed traditional students on standardized tests of reading and math and also showed more advanced social skills and executive control.

One task commonly used in the Montessori style of teaching for younger children is the matching task [4]. In this task, a child is given a set of objects (sometimes split into two subsets and sometimes not) and asked to pair the objects. A variant of that task was used by Daehler *et al.* [10] who used both objects and pictures of objects in their experiments. They found that children around the age of two are able to correctly match objects from both pictures and objects to sets of pictures or objects. One interesting result of this experiment was that the children performed significantly better on tasks where they were asked to match an object to a set of objects, versus picture to object, object to picture, or picture to picture matching. They suggested that this was due to the ability of the children to perceive the objects from multiple angles, thus giving them more reliable information about the objects than they could extract from the pictures.

Other studies have shown infants' ability to identify object pairs and group objects into categories. A study by Leslie *et al.* [11] demonstrated that eleven-month-old infants can individuate pairs of objects only when there is a large amount of physical similarity between objects in the same pair (in this study they used identical objects) and a large physical difference between objects of different pairs. Younger [12] showed that ten-month-old infants can form object categories and determine the variants and invariants of the objects within a category and based on that information they can determine the inclusion of a novel object in the given category. These studies show that even at an early age, humans are able to identify object properties and use them to compare objects, which suggests that this is a fundamental part of intelligence.

Another experiment by McPherson and Holcomb [13] examined event-related brain potentials. Participants were shown a picture of an object, then a picture of an object from one of three categories: related, moderately related, or unrelated. The electroencephalogram (EEG) results showed that across all participants, there was a large negative spike in the N400 family of potentials in the participants' brain shortly after being shown the second picture. The study found that the magnitude of the spike was related to the similarity between the two objects in the pictures. This suggests that, at least at some level, the brain makes a quantitative measure of how similar the two objects are.

B. Robotics

Several studies have demonstrated that robots can measure perceptual as well as functional object similarities for a variety of tasks [14], [15], [16], [17], [18], [8]. The ability to measure the similarity between two objects is extremely useful for tasks such as category recognition and object grouping. Several studies [16], [5] have used unsupervised approaches for object categorization, in which objects were categorized by the similarity of their perceptual features. Their results showed that when the robot was allowed to use all of its sensory modalities, its object categorizations closely resembled the human-provided ones. This suggests that allowing robots to perceive more features about objects can improve their ability to detect similarities between the objects.

Sinapov and Stoytchev [8] showed how a robot can solve the odd-one-out task. The robot picked the object in the group that was least similar to the rest and resulted in the rest of the objects being maximally similar. In this paper we use a similar method to generate similarity scores between objects. We then use this similarity measure to perform object matching rather than solving the odd-one-out task, though they are fundamentally related problems.

III. Experimental Platform

A. Robot and Sensors

The experiments in this study were performed with the upper-torso humanoid robot shown in Fig. 1. The robot has as its actuators two 7-DOF Barrett Whole Arm Manipulators (WAMs), each with an attached Barrett Hand. Each WAM has built-in sensors that measure joint angles and torques at 500 Hz. An Audio-Technica U853AW cardioid microphone mounted in the robot's head was used to capture auditory feedback at the standard 16-bit/44.1 kHz over a single channel.

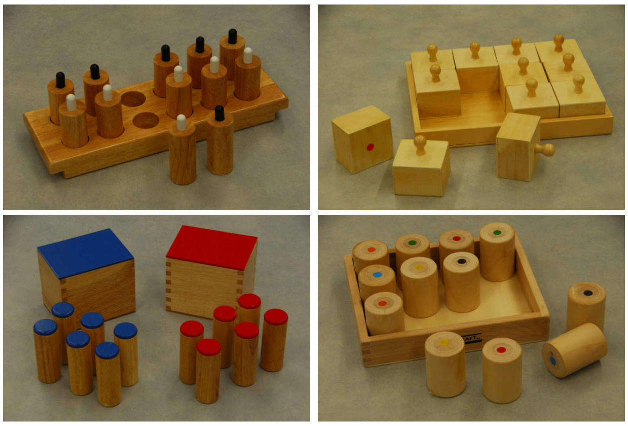

Fig. 2. The four sets of Montessori objects used in the experiments. From left to right and top to bottom the object sets are: *pressure cylinders*, *sound boxes*, *sound cylinders*, and *weight cylinders*. All the objects are marked with colored dots on the bottom to indicate the correct matches; other than that, the objects in each set are all visually identical (except for the *pressure cylinders* and the *sound cylinders*, which also have different colors for the tops to indicate the two sets of six objects).

B. Objects

The robot explored four standard Montessori sets of objects: *pressure cylinders*, *sound boxes*, *sound cylinders*, and *weight cylinders* (Fig. 2). Each set is composed of six pairs of objects. The objects in each pair are functionally identical to each other. The objects in each set are designed to vary in one specific dimension and be identical in all other dimensions. The *pressure cylinders* vary in the amount of force required to depress the rod, with pairs requiring the same amount of force. The *sound boxes* vary in the sounds they make when the contents move around inside the box, with pairs making the same sounds. The *sound cylinders* vary in the same way as the *sound boxes*, but are cylindrical in shape and have different contents than the boxes. The *weight cylinders* vary by weight, going from light to heavy, with pairs having the same weight.

C. Exploratory Behaviors

The robot used ten behaviors to explore the objects: *grasp*, *lift*, *hold*, *shake*, *rattle*, *drop*, *tap*, *poke*, *push*, and *press*. All of these exploratory behaviors, except rattle, have been used in our previous work [19], i.e., they were not specifically designed for the Montessori objects used in this paper. The behaviors were performed with the robot's left arm and encoded with the Barrett WAM API as trajectories in joint-space. The default PID controller of the WAM was used to execute the trajectories. Figure 3 shows images of the robot performing each behavior on one of the sound boxes. All the behaviors were performed identically on each object, with only minor variations due to the initial placement of the objects by the experimenter.

D. Data Collection

The robot interacted with the objects by performing a series of exploration trials. During each trial, an object was placed at a marked location on the table by the experimenter and the robot performed all ten of its exploratory behaviors on the object. The experimenter then picked another object and the robot repeated this process. This was done until each object had been explored ten times. During each interaction, the robot recorded proprioceptive information in the form of joint torques applied to the arm and auditory data captured by the microphone. The robot also recorded visual data, but it was not used in this experiment. In the end, the robot performed all ten behaviors ten times on each of the twelve objects in the four sets, resulting in $10 \times 10 \times 12 \times 4 = 4800$ behavior executions. This resulted in 18 GB of data, which was stored for off-line analysis. It took approximately 20 hours to collect this dataset.

IV. Feature Extraction

We used the method and the publicly available source code for proprioceptive and auditory feature extraction that is described in [5]. It is briefly summarized below. Proprioceptive data was recorded as joint torques over time resulting in a $7 \times m$ matrix, in which each column represents one set of torque readings for all joints and m is the number of readings. To reduce noise, a moving-average filter was applied over each row in the matrix, which corresponds to the torques from one joint. Audio data was recorded as wave files, one for each interaction. A log-normalized Discrete Fourier Transform was performed on each audio file using $2^5+1 = 33$ frequency bins resulting in a $33 \times n$ matrix, where each column represents the activation values for different frequencies at a given point in time and n is the number of samples in the interaction. The Growing Hierarchical Self-Organizing Map (SOM) toolbox [20] was used to map each column to a single state. Two 6×6 SOMs were trained (one for audio and one for proprioception) using 5% of the columns that were randomly selected from all the joint torque and auditory data recorded by the robot. Each joint torque and auditory record was then mapped to a discrete sequence of states, where each column in the record was represented by the most highly activated SOM state for that column. For more details see [5].

V. Experimental Methodology

A. Estimating Similarity

Given a set of objects $\mathcal{O}$ the robot must be able to estimate the pairwise similarity for any two objects $i, j \in \mathcal{O}$ in a given sensorimotor context (i.e., exploratory behavior and sensory modality combination). Let $\mathcal{X}_c^i = [X_1, ..., X_D]$ be the set of sensory feedback sequences detected while interacting with object $i \in \mathcal{O}$ in sensorimotor context $c \in \mathcal{C}$ (where $\mathcal{C}$ is the set of all contexts) and let $sim(X_a, X_b)$ be the similarity between two sequences X_a and X_b. The similarity between objects i and j can be approximated with the expected pairwise similarity of the sequences in $\mathcal{X}_c^i$ and $\mathcal{X}_c^j$:

$$s_{ij}^c = \mathbf{E}[sim(X_a, X_b) | X_a \in \mathcal{X}_c^i, X_b \in \mathcal{X}_c^j]$$

In this paper we used the Needleman-Wunsch global alignment algorithm [21] to calculate $sim(X_a, X_b)$. The algorithm calculates the cost of aligning two discrete sequences

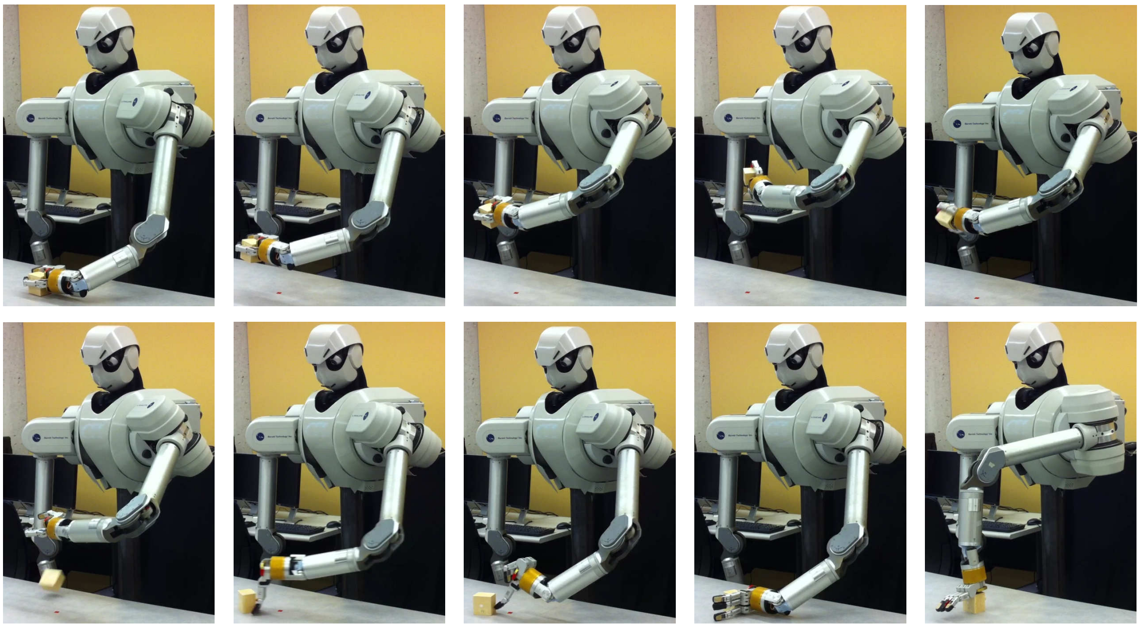

Fig. 3. The ten exploratory behaviors that the robot performed on all objects. From left to right and top to bottom: *grasp*, *lift*, *hold*, *shake*, *rattle*, *drop*, *tap*, *poke*, *push*, and *press*. The object in this figure is one of the sound boxes. The red marker on the table indicates the initial position of the objects at the beginning of each trial. The object was placed back in that position by the experimenter after some of the behaviors (e.g., drop).

(strings), which in our case correspond to sequences of most highly-activated SOM states (see the previous section). The expected similarity s^c_{ij} is estimated as

$$\frac{1}{|\mathcal{X}^i_c| \times |\mathcal{X}^j_c|} \sum_{X_a \in \mathcal{X}^i_c} \sum_{X_b \in \mathcal{X}^j_c} sim(X_a, X_b)$$

Next, the robot estimates the $|\mathcal{O}| \times |\mathcal{O}|$ pairwise object similarity matrix $\mathbf{W}^c$ for a specific sensorimotor context $c \in \mathcal{C}$. Each entry W^c_{ij} in $\mathbf{W}^c$ is defined as the similarity s^c_{ij} between two objects i and j in the specific context c. Figure 4 shows the similarity matrices for the *sound cylinders* for each of the 20 contexts.

B. Combining Sensorimotor Contexts

It has been shown that combining information from different sensorimotor contexts has a boosting effect for tasks such as object recognition [22]. Since object matching is a similar task, it is likely that combining contexts will be useful in this case as well. Thus, in this paper, we propose three methods to combine sensorimotor contexts: uniform combination, recognition accuracy based combination, and pairing accuracy based combination. The result of combining different contexts is a consensus matrix $\mathbf{W}$ that represents the similarity between object pairs for the specific set of contexts that was used to create it.

1) Uniform Combination: Given some set of contexts $\mathcal{C}'$, where $\mathcal{C}' \subseteq \mathcal{C}$, the similarity matrices $\mathbf{W}^c$ for each of these contexts can be used to construct the consensus matrix $\mathbf{W}$ by simply averaging their individual values, i.e.,

$$W_{ij} = \frac{1}{|\mathcal{C}'|} \sum_{c \in \mathcal{C}'} W^c_{ij}$$

for all pairs of objects i and j.

2) Recognition Accuracy Based Combination: This method assumes that contexts that are useful for object recognition will also be useful for object pairing. The object recognition accuracy r_c for context c is estimated by performing 10-fold cross validation on all the data from context c using a classifier that attempts to recognize object identities from sensory feedback sequences. To create the consensus matrix for a given set of contexts $\mathcal{C}'$ ($\mathcal{C}' \subseteq \mathcal{C}$), a weighted combination was used:

$$W_{ij} = \sum_{c \in \mathcal{C}'} \alpha_c \times W^c_{ij}$$

where α_c is the normalized recognition accuracy r_c for context c such that $\sum_{c \in \mathcal{C}'} \alpha_c = 1.0$. The classifier used in this paper was the k-nearest neighbor classifier with k set to 3 and using the global alignment similarity function as a similarity metric.

3) Pairing Accuracy Based Combination: The third combination method allowed the robot to get feedback on its attempts to pair some of the objects to refine its ability to pair the remaining objects. In order to determine the usefulness of each context, the robot split the set of objects such that either 2, 3, or 4 of the six pairs were in the training set and the rest remained in the testing set. Then, for each context c, using the objects in the training set, the robot would attempt to pair them (using the pairing method described below) and evaluate the pairing accuracy p_c for that context. To construct the consensus matrix $\mathbf{W}$, a weighted combination was used similar to the previous method:

$$W_{ij} = \sum_{c \in \mathcal{C}'} \alpha_c \times W^c_{ij}$$

where α_c is the normalized pairing accuracy p_c for context c such that $\sum_{c \in \mathcal{C}'} \alpha_c = 1.0$. After generating the consensus matrix $\mathbf{W}$, the robot would then attempt to pair only the

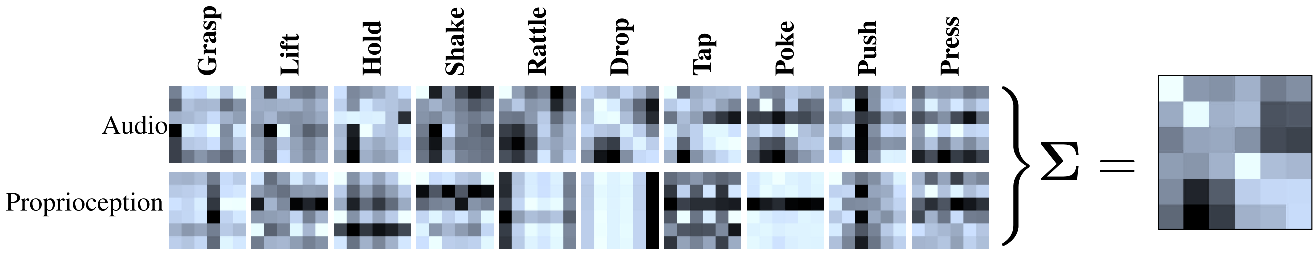

Fig. 4. The similarity matrices used to perform matching given two sets of six objects each for the *sound cylinders*. The matrices for each individual context are shown as well as the consensus matrix for all 20 contexts. The pairing accuracy combination method using four pairs for training was used to combine the individual matrices.

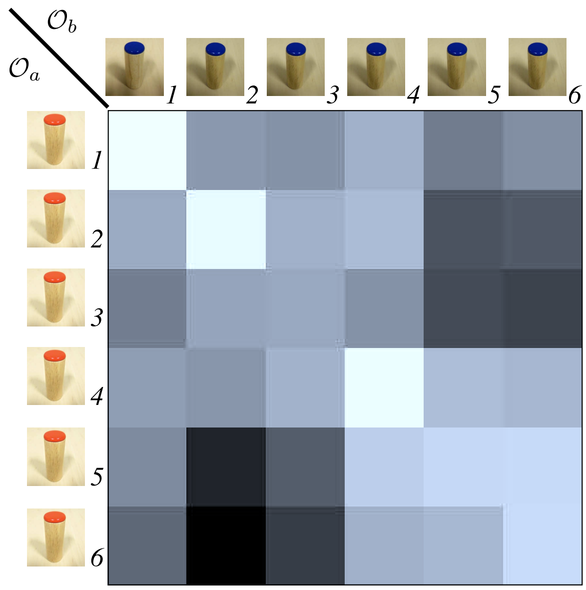

Fig. 5. The consensus weight matrix for the *sound cylinders* using all 20 sensorimotor contexts for matching two groups of six objects. The pairing accuracy combination method using four pairs to train was used to combine the individual similarity matrices for each context. The subscripts indicate correct matches.

objects from the testing set. Figures 4 and 5 show a consensus matrix generated by combining the similarity matrices from all 20 contexts when training using 4 pairs of objects.

C. Generating Matchings

The robot was tasked with generating matchings among the objects in the four Montessori toys. The objects were split into two groups of six and the robot was tasked with selecting one object from each group to generate a match. This split into two groups of six is naturally suggested by the Montessori toys. For example, the sound cylinders have either red or blue caps; the pressure cylinders have either black or white buttons (see Fig. 2).

More formally, given a 6x6 non-symmetric similarity matrix $\mathbf{W}^c$ or a consensus matrix $\mathbf{W}$ and objects $\mathcal{O}$ partitioned into two sets of equal size $\mathcal{O}_a$ and $\mathcal{O}_b$, matches were generated by picking pairs that maximized similarity between the objects in the pair and minimized similarity between those objects and the remaining objects. One such matrix is shown in Fig. 4. Formally, the objects $i \in \mathcal{O}_a$ and $j \in \mathcal{O}_b$ that maximize

$$q(i,j,\mathbf{W}) = W_{ij} - \gamma \left[\sum_{k \in \mathcal{O}_b/j} W_{ik} + \sum_{k \in \mathcal{O}_a/i} W_{kj} \right]$$

were selected and then removed from $\mathcal{O}_a$ and $\mathcal{O}_b$. The first term captures the pairwise similarity between objects i and j; the last term captures the pairwise similarity between objects i and j and the rest of the objects. The constant γ is a normalizing weight, which ensures that this function is not biased toward any of the terms. In our case, it was set to

$$\gamma = \frac{1}{2(|\mathcal{O}| - 1)}.$$

This process was repeated until no more objects remained to be paired.

D. Evaluation

Given a set of objects (e.g., the weight cylinders), the robot's model was queried in order to group the objects into pairs. Five interactions were randomly picked for each object from the set of ten interactions that were performed on each object and used to create the weight matrix $\mathbf{W}^c$ for each sensorimotor context $c \in \mathcal{C}$. Consensus matrices $\mathbf{W}$ were generated using the three methods described above for a given set of contexts. Matchings were then generated using the method described above. This process was repeated 100 times for every group of contexts. For each size from 1 to $|\mathcal{C}|$, 100 sets of contexts were randomly generated and tested ($1,721$ in total)[1]. Results are reported as the average accuracy or as Cohen's kappa statistic [23] over all 100 iterations. Accuracy is computed as

$$\%Accuracy = \frac{\#correct\ matchings}{\#total\ matchings} \times 100.$$

The kappa statistic is computed as

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

In our experiments, $P(a)$ is the pairing accuracy of the robot and $P(e)$ is the accuracy a random matching would be

[1]For sets of size 1, $|\mathcal{C}| - 1$, and $|\mathcal{C}|$ all sets of that size were tested since there were fewer than 100 sets of those sizes.

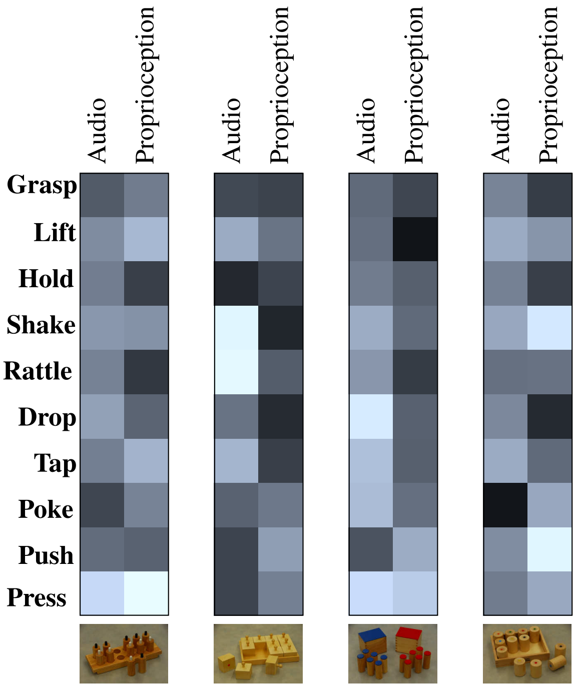

Fig. 6. The accuracy of each context when matching between two sets of six objects. Lighter values indicate higher accuracy with completely white being 100%. Darker values indicate lower accuracy with completely black being 0%. The images from left to right are: *pressure cylinders*, *sound boxes*, *sound cylinders*, and *weight cylinders*.

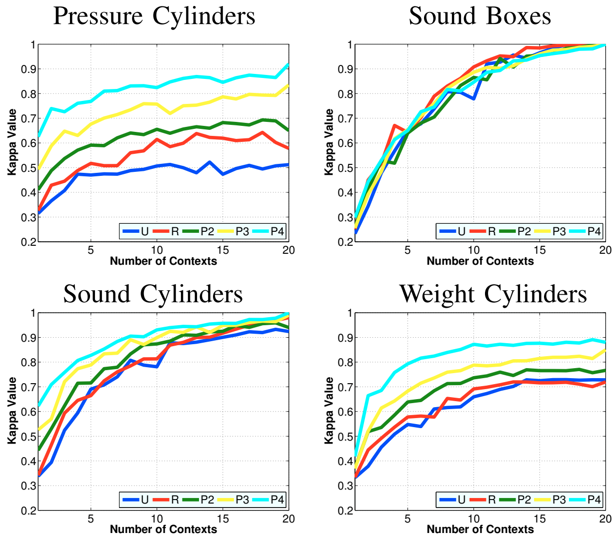

Fig. 7. The kappa statistic for each set of objects. Each line represents a different method for combing the sensorimotor contexts. The line labels are as follows: U-uniform combination; R-recognition accuracy based combination; P2-pairing accuracy using two pairs for training; P3-pairing accuracy using three pairs for training; P4-pairing accuracy using four pairs for training.

expected to get. Kappa is used to allow for direct comparisons between the different sensorimotor context combination methods, since for the pairing accuracy based method, chance accuracy is different than it is for the other methods. The kappa statistic controls for chance accuracy.

The evaluation was performed off-line after the robot interacted with all 48 objects (4 Montessori tasks $\times$ 12 objects in each).

VI. Results

A. Object Matching with a Single Context

Figure 6 shows the matching accuracy for each context for all four Montessori tasks. For the *pressure cylinders*, the best sensorimotor context was *proprioception-press* (97.5% pairing accuracy), which was expected. Surprisingly, *audio-press* also did well (80.7%), which was not expected since (at least to the authors' ears) all the cylinders sound the same when pressed. Also interesting is the *audio-drop* context for the *sound cylinders* (89.3% accuracy), which outperformed both *shake* (60.3%) and *rattle* (51.3%) behaviors for *audio*. *Audio-press* (82.3%) for the *sound cylinders* also did well, which is likely due to the fact that they would fall over while being pressed. It is also worth noting that for the *weight cylinders*, the best contexts were *proprioception-shake* (87.7%) and *proprioception-push* (94.3%) rather than contexts that more directly measure the weight such as *proprioception-lift* (50.7%) and *proprioception-hold* (18.8%).

In summary, the robot was able to identify the relevant behaviors and sensory modalities and use them to pair the objects in each of the four Montessori tasks with a high degree of accuracy.

B. Object Matching with Multiple Contexts

Figure 7 shows the kappa statistic for each set of objects as the number of contexts is varied from 1 to 20. The graphs show that as the number of sensorimotor contexts used to perform matching increases, so does the kappa statistic. In all cases, the pairing accuracy based combination using four pairs for training (the cyan line) outperforms all the other combination methods. The only exception to this is for the *sound boxes*, since accuracy reaches 100%, all methods reach a kappa value of 1.0. In most cases, the pairing accuracy based combination using three pairs for training (the yellow line) also outperforms the other methods (except for the method that uses four pairs for training). The pairing accuracy based combination using two pairs for training performs about the same as the recognition accuracy combination method, which usually performs slightly better than the uniform combination method. All the combination methods perform better than chance for all object sets, which is indicated by a 0.0 kappa value.

C. Repeating the Same Behavior

In all results reported up to this point, five interactions were randomly chosen from the ten for each object during each iteration. Figure 8 shows the average kappa statistic as the number of trials vary, averaged over all the sets of objects and number of contexts. The accuracies quickly converge after only a few trials, implying that repeating the same behavioral repertoire multiple times on an object has quickly diminishing returns. In most cases and for all combination methods, after four repetitions there is very little gain. Diminishing returns is most quickly realized for the pairing accuracy combination method using four pairs for training. The largest gain when increasing interactions was realized by the uniform combination method. This suggest that the uniform combination method benefited the most from a decrease in noise due to its lack of weighted preferences between the contexts, whereas the pairing accuracy combination methods didn't benefit as much because the weights assigned to each context already decreased the noise.

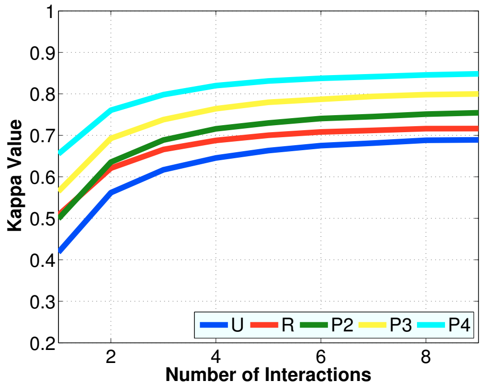

Fig. 8. The kappa statistic averaged across all four sets of objects while varying the number of interactions used to generate the similarity matrices $\mathbf{W}^c$ for each context $c \in \mathcal{C}$. The number of randomly sampled interactions was varied from 1 to 9. The line labels are the same as in Fig. 7.

VII. Conclusion and Future Work

This paper demonstrated a framework that allows a robot to solve object matching tasks by estimating the pairwise similarity of objects in specific sensorimotor contexts. The performance of this framework was evaluated with four standard Montessori tasks that require pairing a set of objects based on their perceived similarities across multiple sensory modalities. The results showed that for a given set of objects, certain contexts are best suited to extract the information necessary to perform object pairing (e.g., *audio-shake* for the *sound boxes*), while others are not useful for that set of objects (e.g., *proprioception-lift* for the *sound cylinders*).

The robot was also able to combine similarity measures from different contexts using three different methods: uniform combination, recognition accuracy based combination, and pairing accuracy based combination. The robot was able to achieve the best performance in almost every case when it was allowed to train on four of the six object pairs before being tested on the remaining two. These results show that embodied sensorimotor similarity measures between objects can be extremely useful for performing matching tasks.

This paper introduced the domain of Montessori tasks to the field of robotics and showed how embodied learning could be used to solve object pairing tasks. For each set of objects the robot learned which set of contexts are most useful for pairing the objects and which are not. The objects in each Montessori task implicitly capture an important concept that the robot can discover on its own through sensorimotor exploration. In the future similar tasks could be used to teach robots not only matching skills, but also important concepts such as ordering, sorting, and relating.

Future work can also expand upon this research by improving the feature extraction methods, the similarity measure, the combination methods, or by using a better matching algorithm. It would also be useful to develop methods that can discover novel exploratory behaviors. This framework can also be applied to other tasks such as object categorization and object recognition. For example, a robot could match previous experiences with objects with new experiences in order to label the objects.

References

[1] M. Montessori, *The Montessori method.* Frederick A. Stokes Co., New York City, USA, 1912.

[2] A. Lillard, *Montessori: The science behind the genius.* Oxford University Press, New York City, USA, 2008.

[3] A. Lillard and N. Else-Quest, "The early years: Evaluating Montessori," *Science*, vol. 313, no. 5795, pp. 1893–1894, 2006.

[4] M. Pitamic, *Teach Me to Do It Myself.* Elwin Street Productions, London, UK, 2004.

[5] J. Sinapov, T. Bergquist, C. Schenck, U. Ohiri, S. Griffith, and A. Stoytchev, "Interactive object recognition using proprioceptive and auditory feedback," *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1250–1262, 2011.

[6] K. Saenko and T. Darrell, "Object category recognition using probabilistic fusion of speech and image classifiers," in *Machine Learning for Multimodal Interaction: 4th International Workshop, MLMI 2007, Brno, Czech Republic, June 28-30, 2007, Revised Selected Papers*, vol. 4892. Springer-Verlag, New York, 2008, pp. 36–47.

[7] F. Endres, C. Plagemann, C. Stachniss, and W. Burgard, "Unsupervised discovery of object classes from range data using latent Dirichlet allocation," in *Proc. of RSS*, 2009, pp. 113–120.

[8] J. Sinapov and A. Stoytchev, "The odd one out task: Toward an intelligence test for robots," in *Proceedings of the 9th IEEE International Conference on Development and Learning (ICDL)*, 2010, pp. 126–131.

[9] K. Rathunde and M. Csikszentmihalyi, "Middle school students motivation and quality of experience: A comparison of Montessori and traditional school environments," *American Journal of Education*, vol. 111, no. 3, pp. 341–371, 2005.

[10] M. Daehler, R. Lonardo, and D. Bukatko, "Matching and equivalence judgments in very young children," *Child Dev.*, pp. 170–179, 1979.

[11] A. Leslie and M. Chen, "Individuation of pairs of objects in infancy," *Developmental Science*, vol. 10, no. 4, p. 423, 2007.

[12] B. Younger, "The segregation of items into categories by ten-month-old infants," *Child Development*, pp. 1574–1583, 1985.

[13] W. McPherson and P. Holcomb, "An electrophysiological investigation of semantic priming with pictures of real objects," *Psychophysiology*, vol. 36, no. 01, pp. 53–65, 1999.

[14] S. Nolfi and D. Marocco, "Active perception: A sensorimotor account of object categorization," in *From Animals to Animats: The Seventh International Conference on the Simulation of Adaptive Behavior*, 2002, pp. 266–271.

[15] L. Natale, G. Metta, and G. Sandini, "Learning haptic representation of objects," in *Proceedings of the International Conference on Intelligent Manipulation and Grasping*, 2004.

[16] T. Nakamura, T. Nagai, and N. Iwahashi, "Multimodal object categorization by a robot," in *Proceedings of the IEEE/RSJ IROS*, 2007, pp. 2415–2420.

[17] S. Takamuku, K. Hosoda, and M. Asada, "Object category acquisition by dynamic touch," *Advanced Robotics*, vol. 22, no. 10, pp. 1143–1154, 2008.

[18] J. Sun, J. Moore, A. Bobick, and J. Rehg, "Learning visual object categories for robot affordance prediction," *The International Journal of Robotics Research*, vol. 29, no. 2-3, p. 174, 2010.

[19] J. Sinapov, C. Schenck, K. Staley, V. Sukhoy, and A. Stoytchev, "Grounding semantic categories in behavioral interactions: Experiments with 100 objects," *Robotics and Autonomous Systems (to appear)*.

[20] A. Chan and E. Pampalk, "Growing hierarchical self organizing map (GHSOM) toolbox: visualizations and enhancements," in *Proc. of the 9th NIPS Conference*, vol. 5, 2002, pp. 2537–2541.

[21] G. Navarro, "A guided tour to approximate string matching," *ACM computing surveys (CSUR)*, vol. 33, no. 1, pp. 31–88, 2001.

[22] J. Sinapov and A. Stoytchev, "The boosting effect of exploratory behaviors," in *Proc. of the 24th National Conference on Artificial Intelligence (AAAI)*, 2010, pp. 1613–1618.

[23] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.