How a robot's attention shapes the way people teach

Yukie Nagai[1] **Akiko Nakatani**[1] **Minoru Asada**[1,2]
[1] Graduate School of Engineering, Osaka University
[2]JST ERATO Asada Synergistic Intelligence Project
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
{yukie, akiko.nakatani, asada}@ams.eng.osaka-u.ac.jp

Abstract

We address the question of how a robot's attention shapes the way people teach. When demonstrating a task to a robot, human partners often emphasize important aspects of the task by modifying their body movement as caregivers do toward infants. This phenomenon has recently been investigated in developmental robotics; however, what causes such action modification is yet unrevealed. This paper presents an experiment examining influences that a robot's attention has on task demonstration of human partners. Our hypothesis is that a robot's bottom-up attention based on visual salience induces partners to exaggerate their body movement, to segment the movement frequently, to approach closely to the robot, and so on, which are homologous to modifications in infant-directed action. We present quantitative results supporting our hypothesis and discuss what properties of bottom-up attention contribute to eliciting such action modifications.

1. Introduction

Attention plays several roles in human-robot interaction. Robots, for example, show their interest by directing their attention to a favorite or a novel object and allocate their computational resources to attended signals. If robots are designed to learn action and/or objects from human partners, the robots' attention becomes even more important. Where to attend and hence what to learn significantly affect performance of the robots' learning.

Despite the importance of attention in task learning, there have not been so many studies focusing on it (see (Demiris and Khadhouri, 2008) and (Thomaz and Breazeal, 2008) for rare examples of studies focusing on a robot's attention). A big challenge here is that a top-down architecture to control attention cannot be adopted if robots are supposed not to know what to learn. Without semantic knowledge about the task, robots can only employ a bottom-up mechanism or simply fixate a certain location in the environment in order to obtain all relevant information from the fixed image.

Figure 1: A human teacher demonstrating a nesting-cup task to a humanoid robot

Inspired by human caregiver-infant interaction, we suggest that bottom-up attention shapes the way people teach so as to facilitate learning. More specifically, *motionese* is induced by bottom-up attention embedded in a robot. Motionese is modification in infant-directed action, which highlights important aspects of the action (Brand et al., 2002, Rohlfing et al., 2006, Nagai and Rohlfing, 2009). It is characterized by, for example, exaggeration of body movements, more pauses between movements, and high proximity to an infant. It is also known that motionese attracts stronger and/or longer attention of infants than adult-directed action (Brand and Shallcross, 2008, Koterba and Iverson, 2009) and facilitates more object exploration of infants (Koterba and Iverson, 2009). Regarding infants' attention, developmental studies have revealed that infants rely more on bottom-up signals than top-down preference or knowledge about the context when determining where to attend (Frank et al., 2009, Golinkoff and Hirsh-Pasek, 2006). Bottom-up attention based on visual salience better predicts young infants' attention than the locations of social signals like face-like stimuli do. We hypothesize that a robot endowed with infant-like bottom-up attention encourages human partners to modify their action so that the action is better structured as in motionese.

This paper presents an experiment of human-robot interaction, which examines influences of a robot's attention on task demonstration of human partners. A humanoid robot is equipped either with bottom-up attention based on visual salience or top-down attention controlled by an experimenter. We quantitatively analyze task relevant movements of the partners. Section 2 explains the experimental setup and two attention models used in our experiment. A quantitative analysis of the partners' task demonstration and a questionnaire are presented with their results in Sections 3 and 4, respectively. We then discuss in Section 5 what properties of a robot's attention contribute to eliciting action modifications in the partners and conclude the paper in Section 6.

2. Experiment of human-robot interaction

2.1 Setting and robot

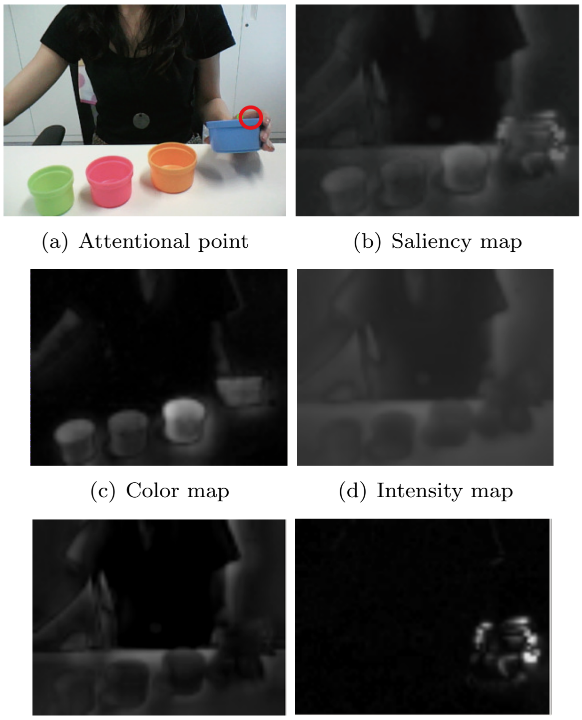
Figure 1 shows a scene from the experiment, where a human partner is demonstrating a nesting-cup task to a small humanoid robot. A nesting-cup task is often used to asses cognitive development in human children because seriated structures of cups seem formally homologous to grammatical constructions of language (Greenfield et al., 1972, Greenfield, 1991, Hayashi, 2007). For example, a strategy for paring cups can be represented by "cup A enters cup B," which corresponds to a simplest sentence structure of "subject-verb-object." We consider that examining teaching strategies for the task well illustrates how people support cognitive development in robots as well as in children.

The robot used in our experiment appears in the left side of Figure 1. It is about 45 cm tall and has 22 degrees of freedom, of which 2 are for the neck, 6 for the arms, and 14 for the legs. A camera is attached to the robot's head, and the direction of the camera is controlled by turning the head. Throughout the experiment, the robot moved only its head and arms while sitting on the table where the task was presented.

2.2 Two experimental conditions

In order to examine how the robot's attention shapes the way human partners teach a task, we designed two conditions in which a different type of visual attention was implemented into the robot.

Condition 1: The robot was endowed with a saliency model (Itti et al., 1998, Itti et al., 2003) to determine where to attend. The model computed bottom-up salience for image regions as contrast to the surrounding regions in terms of primitive features. Figure 2 shows a sample scene, where the salience was calculated from four features: color, intensity, orientation of edge features, and motion. Figure 2 (a) indicates the attentional location of the robot (i.e., the most salient location) with a red circle, (b) shows the corresponding saliency map, and (c)-(f) are the maps derived from the four features. Since the interaction partner was shaking the blue cup with her left hand, the model selected the region including the hand and the cup as an attentional point. Note that human features like face and hands as well as the cups could be the focus of the robot's attention due to their conspicuous color, edge, and/or motion even without using a priori knowledge about the features. Refer to (Itti et al., 1998, Itti et al., 2003, Nagai et al., 2008, Nagai and Rohlfing, 2009) for a more detailed explanation about the model.



(a) Attentional point (b) Saliency map

(c) Color map (d) Intensity map

(e) Orientation map (f) Motion map

Figure 2: A sample scene showing visual salience

A reason for adopting the bottom-up architecture is its similarity to infants' attention. As mentioned before, it has been demonstrated that bottom-up salience better predicts fixations of young infants than the locations of faces do (Frank et al., 2009). Young infants are known to rather ignore social cues and use perceptual salience to guide their attention when learning words from caregivers (Golinkoff and Hirsh-Pasek, 2006). They rely stronger on bottom-up salience than top-down preference or knowledge because of little semantic knowledge about the context. We hypothesized that the robot embedded with the saliency model was accepted as if it had immature abilities like an infant, which would induce motionese from human partners.

Condition 2: We made the robot behave like an older child or even an adult. In contrast to Condition 1, where the robot used only bottom-up signals, in Condition 2 the robot shifted its attention as if it understood the goal of the presented task.

We adopted a wizard of Oz technique to avoid difficulties in developing such sophisticated attention. An experimenter controlled the robot's attention by selecting the next attentional point in the robot's camera image. The following three rules were applied to selecting the attentional point:

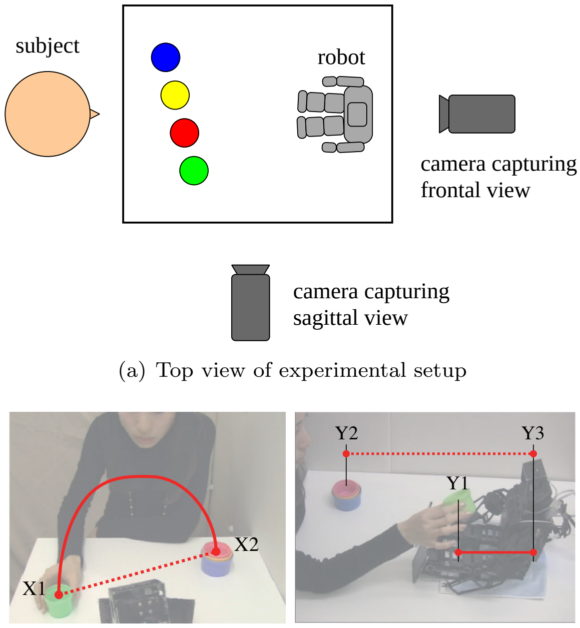
- To select a cup held by a partner when he/she is moving it toward the goal position, which can be another cup or an empty space on the table
- To select the goal position as the attentional point when a partner is doing anything else with the holding cup except moving it to the goal (e.g., showing the cup to the robot)
- To direct the robot's attention to a partner's face when he/she does not hold any cup in his/her hand

These rules were developed based on heuristics. If older children know the goal of the task or can predict the demonstrated action, they would smoothly track the movement and even shift attention to the goal position before the actual movement is presented. They would examine where the goal is and what is placed there. For older children, a partner's face is also attractive and important to receive social cues. They would look at the partner's face to establish eye contact, to achieve joint attention, and/or to read emotional expressions especially when the partner's body movement is not so prominent. A wizard of Oz technique enabled the robot to reproduce such matured adult-like attention.

2.3 Subjects and task

Each condition had 8 subjects (7 male and 1 female university students) between the ages of 22 to 30 (i.e., 16 independent subjects in total). They studied engineering and had some experiences of interacting with robots though they met our robot for the first time and knew *nothing* about the nature of the experiment.

An experimenter instructed the subjects to demonstrate a nesting-cup task to the humanoid robot so that the robot could learn to perform the task. They were allowed to use speech as well as action to teach the task although speech was not



(a) Top view of experimental setup

(b) Frontal view (c) Sagittal view

Figure 3: Analysis of demonstrators' movement

included in our analysis. The instruction they were given was only about the position of the robot's camera and its ability to shift the attention by reacting to the subjects' movement, but not about the mechanism of the attention. They were instead allowed to get familiar with the robot and the task by performing it once before the experiment. The experiment took about 5 minutes followed by a questionnaire.

3. Analysis of cup manipulation

We quantitatively analyzed cup manipulation of the subjects. Their body movement was recorded with two cameras as illustrated in Figure 3 (a): one captured the subjects' movement in a frontal plane and the other captured it in a sagittal plane. Figures 3 (b) and (c) show sample images, in which positions of $X1$, $X2$, $Y1$, $Y2$, and $Y3$ used in the analysis are denoted.

3.1 Six characteristics of cup manipulation

We measured six characteristics of cup manipulation by tracking the movement of the cups:

(a) Roundness of cup movement:

$$R = \frac{\text{travel distance } (X1, X2)}{\text{linear distance } (X1, X2)}$$

The travel and the linear distances are denoted by the solid and the dashed lines in Figure 3 (b), where $X1$ and $X2$ are the initial and final position of a cup. That is, roundness represents how large arc subjects formed to move a cup.

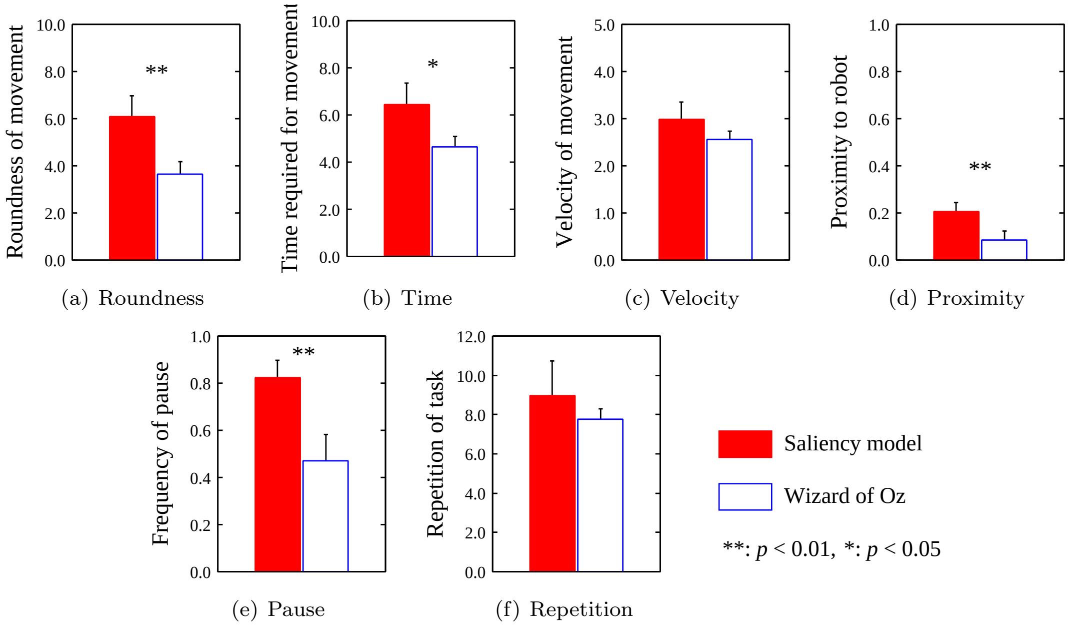


Figure 4: Comparative results for demonstrators' movement

(b) Time required for an action, where an action is defined as relocation of a cup from $X1$ to $X2$: T [sec]

(c) Velocity of moving a cup:

$$V = \frac{\text{travel distance } (X1, X2)}{T} \text{ [pixel/sec]}$$

(d) Proximity to the robot:

$$P = 1 - \frac{\text{horizontal distance } (Y1, Y3)}{\text{horizontal distance } (Y2, Y3)}$$

$Y1$, $Y2$, and $Y3$ are the horizontal position of a moving cup, that of the blue cup (i.e., the goal position), and that of the robot's head, respectively, as indicated in Figure 3 (c). The more closely a subject approached the robot, the larger the proximity became.

(e) Frequency of pauses between movement:

$$F = \frac{\text{num. of pauses}}{\text{num. of actions}}$$

A pause was counted when subjects stopped their hand movement while transporting a cup. The frequency became 1 if a subject took one pause while moving a cup.

(f) Number of repetition of presenting the task, where the task is defined as the relocation and seriation of the four cups: N

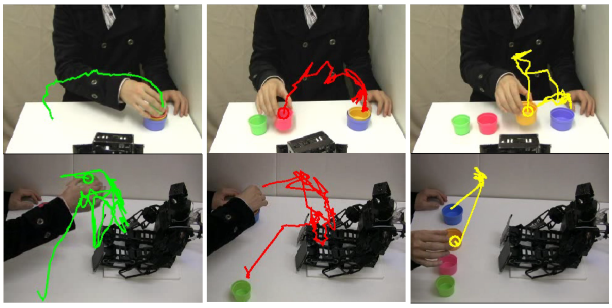
These parameters were adopted from (Brand et al., 2002, Rohlfing et al., 2006, Vollmer et al., 2009), where motionese was characterized by higher roundness, longer time for an action, higher proximity to an infant/a robot, more pauses, and higher repetitiveness. Compared to these studies, our experiment aimed not only at showing differences and/or similarities between infant-, adult-, and robot-directed action, but also at revealing properties of a learner's attention which influences a teacher's demonstration of a task.

3.2 Results: motionese induced by salience-based attention

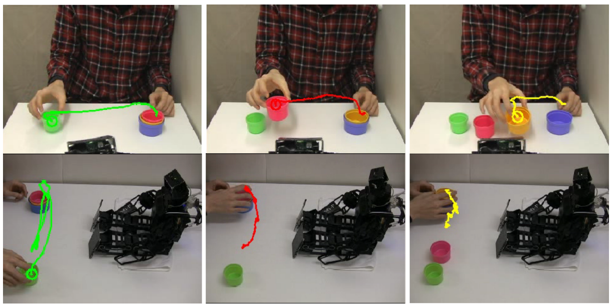
Figure 4 shows the result of the analysis: (a) to (f) are the mean and standard deviation for roundness, time, velocity, proximity, pause, and repetition, respectively. A filled bar and an open bar show the results for Condition 1 (the saliency model) and Condition 2 (the wizard of Oz technique), respectively. A t-test on the two conditions revealed significant differences (indicated by "$**$" if $p < 0.01$ and "$*$" if $p < 0.05$) in four out of the six characteristics: (a) roundness, (b) time, (d) proximity, and (e) pauses.

3.2.1 Roundness of movement

The roundness of cup movement was significantly higher in Condition 1 ($M = 6.08$, $SD = 0.896$) than in Condition 2 ($M = 3.65$, $SD = 0.529$), $p < 0.01$ (see Figure 4 (a)). Subjects in Condition 1 moved a cup in a larger arc than those in Condition 2, suggesting that the robot's attention based on bottom-up salience induced exaggeration of cup movement. Figures 5 (a) and (b) show an example of the tra-



(a) Motionese induced by salience-based attention (Condition 1)



(b) Smooth movement facilitated by adult-like attention (Condition 2)

Figure 5: Trajectory of cup movement

jectories of cup movement observed in Conditions 1 and 2, respectively. The colored lines are the traveling path of the cups, which correspond to the solid line in Figure 3 (b). The paths qualitatively demonstrate how cup movement was exaggerated by being elicited by bottom-up attention.

A reason is considered as follows: The saliency model made the robot's attention sensitive to cup movement. As seen in Figure 2 (f), motion was a strong cue to attract the robot's attention. When a subject started handling a cup, the robot fixated the cup and tracked the movement of the cup with high salience produced by the movement. However, the shift of the robot's attention might be too small to recognize because of the small body of the robot and of spatial continuity in salient motion. Subjects therefore exaggerated cup movement so as to examine the robot's attention. In contrast, the robot's attention in Condition 2 might easily be examined. The robot largely shifted attention between a cup and a subject's face depending on the situation, which eased the identification of the attentional location. Moreover, the robot's attention directed to the goal position ahead of an actual movement facilitated smooth and linear transition of a cup as seen in Figure 5 (b). This comparative result suggests that the smallness of the attentional shift of the robot was a key to elicit the exaggeration of partners' body movement.

3.2.2 Time required for an action and velocity of movement

Exaggeration of cup movement produced a secondary effect: Subjects in Condition 1 spent significantly longer time for relocating a cup ($M = 6.45$, $SD = 0.905$) than those in Condition 2 ($M = 4.66$, $SD = 0.417$), $p < 0.05$, due to the longer travel distance (see Figure 4 (b)). Salience-based attention influenced partners' task demonstration in terms not only of space (i.e., high roundness of movement) but also of time.

Note that the velocity of cup movement did not differ between the two conditions. The movement in Conditions 1 was as fast ($M = 2.99$, $SD = 0.362$) as in Condition 2 ($M = 2.56$, $SD = 0.169$), $p = 0.199$ (see Figure 4 (c)), suggesting that longer time required for an action was caused only by longer distance for traveling a cup.

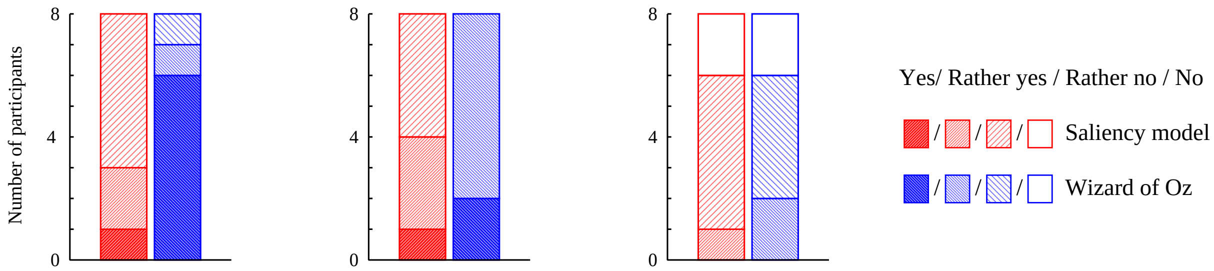
3.2.3 Proximity to the robot

The proximity to the robot was significantly higher in Condition 1 ($M = 0.207$, $SD = 0.0361$) than that in Condition 2 ($M = 0.0839$, $SD = 0.0395$), $p < 0.01$ (see Figure 4 (d)). Subjects in Condition 1 more closely approached the robot than those in Condition 2, that is, the robot's attention based on salience encouraged subjects to intensify their movement. Figure 5 shows qualitative difference. The paths drawn in the sagittal view (the lower pictures in Figures 5 (a) and (b)) show how closely subjects presented a cup to the robot. In Condition 1 they brought a cup to the front of the robot's head whereas subjects in Condition 2 did not. They rather linearly moved a cup to the target position in Condition 2.

The robot's attention based on salience was sensitive to signals. It could easily be distracted by irrelevant stimuli while rapidly responding to new relevant stimuli. It seemed subjects in Condition 1 intuitively understood that intensive movement such as shaking a cup and closely approaching the robot was effective to attract and strengthen the robot's attention. Therefore they tried to draw the robot's attention, when it was distracted, by closely presenting a cup to the robot. In Condition 2, by contrast, the proximity was low over the experiment. The robot's attention controlled by the wizard of Oz technique rather elicited distant movement from the robot. Reliability and predictability of the robot's attention encouraged subjects to efficiently demonstrate the task.

3.2.4 Frequency of pauses between movements

We also found significantly higher frequency of pauses in Condition 1 ($M = 0.823$, $SD = 0.0754$) than in Condition 2 ($M = 0.471$, $SD = 0.111$),



(a) Do you think the robot was looking at you?

(b) Do you think the robot could understand and learn the task?

(c) Do you think the robot can imitate the task?

Figure 6: Questionnaire about the robot's attention, learning, and imitation

$p < 0.01$ (see Figure 4 (e)). Subjects took more pauses between movements in Condition 1, suggesting that salience-based attention induced more action segmentation in the task. Figure 5 (a), especially the trajectories in the sagittal view, shows how actions were segmented. The cups were first linearly lifted to the front of the robot's head, stayed there for a while except small movement, and then put down on the table. Subjects' action of presenting a cup closely to the robot resulted in segmenting the cup movement into two sub-actions: lift-up and put-down.

As explained in the former sections, salience-based attention was difficult to examine. The attentional shift of the robot was rather small and unpredictable. Thus subjects would try to examine the robot's attention by creating rhythm in their action like move-stop-move. In Condition 2, by contrast, subjects easily understood the strategy for the robot's attention. Reliability and predictability of the robot's attention promoted smooth and continuous movement of subjects as seen in Figure 5 (b).

3.2.5 Repetition of task demonstration

Repetition, the last characteristic we analyzed, did not show difference between the two conditions: Subjects in Condition 1 repeated demonstrating the task as many ($M = 8.95$, $SD = 1.78$) as those in Condition 2 ($M = 7.76$, $SD = 0.524$), $p = 0.129$ (see Figure 4 (f)). This is an artifact of the fixed duration of the experiment. An experimenter asked the subjects to stop demonstrating the task after about 5 minutes so as to focus on the beginning of the interaction, in which the subjects were more enthusiastic about the interaction. However, other studies analyzing infant-/robot-directed action found higher repetitiveness of task demonstration (Brand et al., 2002, Rohlfing et al., 2006, Vollmer et al., 2009). We will thus conduct another experiment without limitation in the interaction duration and analyze temporal changes in motionese over long interaction.

4. Questionnaire about the robot's attention and capability

4.1 Three questions

We conducted a questionnaire to gain better insights into why subjects modified their actions. After the interaction experiment, all the subjects were asked to answer the following three questions by "yes," "rather yes," "rather no," or "no":

(a) Do you think the robot was looking at you?

(b) Do you think the robot could understand and learn the task?

(c) Do you think the robot can imitate the task?

4.2 Results

The results are shown in Figure 6. In each graph, the left and right bars present the results for Condition 1 (the saliency model) and Condition 2 (the wizard of Oz technique), respectively. A darkest bar denotes the number of subjects who answered "yes" while an open bar "no."

4.2.1 Focus of the robot's attention

The first insight is about the focus of the robot's attention. The result shown in Figure 6 (a) indicates that the robot equipped with salience-based attention focused less on the demonstrated task than the robot with top-down attention did. More than half of the subjects in Condition 1 answered "rather no" whereas majority answered "yes" in Condition 2.

This result is consistent with our interpretation described in Section 3. The saliency model made the robot's attention sensitive and even distracted, which actually induced subjects to exaggerate their movement. Subjects amplified their body movement and closely approached the robot in order to draw and maintain the robot's attention. Note that there were three subjects answering "yes / rather yes" and *no* participant answering "no" in Condition 1, indicating that the saliency model nonetheless enabled the robot to look at relevant locations despite no knowledge about the task.

4.2.2 Robot's learnability

The result concerning the robot's learnability shows that the saliency model enabled the robot to learn the task to some extent. The half of the subjects in Condition 1 answered the question (b) by "yes / rather yes" (see Figure 6 (b)), indicating that the robot was detecting task relevant targets.

In our experiment, the robot did not learn the task or even improve the strategy for attention. The same attentional model with the same parameters was used over the experiment. An interesting finding is that the difference between the two conditions became less for the question (b) than for (a). The answers in Condition 1 were more positive for the question (b) than for (a) whereas the contrary in Condition 2. It may suggest that although the robot's attention based on salience was not always directed to the subjects, it captured the important aspects of the demonstrated task, which was emphasized by the action modifications of the subjects.

4.2.3 Robot's capacity to imitate

The result concerning the robot's capacity to imitate did not really reflect the difference in the robot's attention but rather reflected the fact of no hands installed in the robot. The number of subjects answering "no / rather no" were almost the same between the two conditions, and no one answered "yes" unlike the other questions. We consider this result an artifact caused by limited capacity of the robot's action, and there must be some dependencies between attention, learning, and imitation. We will further examine the relation using a robot equipped with sophisticated hands.

5. Discussion

The experiment verified our hypothesis that bottom-up attention of a robot learner induces motionese of human teachers. Figure 7 summarizes what properties of a learner's attention elicit what aspects of modifications in teachers' actions.

We found mainly two types of links between attention and action modification. First, teachers exaggerate their actions responding to a learner's attention with respect to *space*. Teachers' movements show high roundness and high proximity, which are induced by small shift and high distraction of a learner's attention. Teachers may try to amplify the attentional shift of a learner by spatially exaggerating their body movement or to concentrate a learner's attention by narrowing their movement in order to examine where the learner looks. Secondly, teachers synchronize their body movement with a learner's attention in terms of *time*. They spend long time to demonstrate a task and create rhythm in the movement responding to slow and unpredictable attention of a learner. Unlike exaggeration in space, teachers do not try to accelerate the attentional shift of a learner but adjust their movement to the learner. Although our findings might not cover all aspects of attention or action modification, we can see some main structure concerning how a learner's attention shapes teaching.

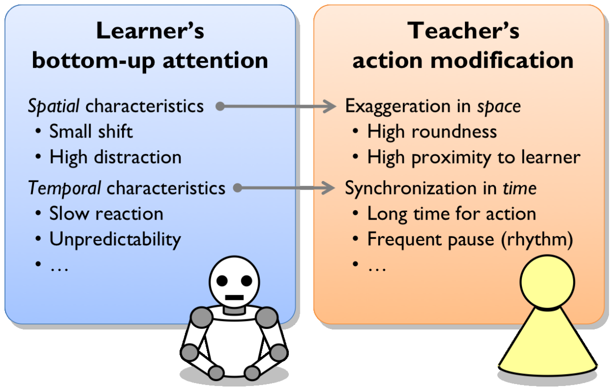


Figure 7: Why and how bottom-up attention of a learner induces motionese of a teacher

Similar to our findings, Shimojo (Shimojo, 2006) stated three types of modifications in parental actions: modifications in terms of space, time, and emotion. Teachers' emotion was not in our focus of the analysis, but it is surely important for a learner to perceive what is more important in demonstrated actions. Nagai and Rohlfing (Nagai and Rohlfing, 2009) revealed that social cues from teachers can be used to detect sub-goals of an action. We will investigate how emotional exaggeration of teachers influences task learning.

6. Conclusion

This study has addressed the question of how a learner's attention shapes the way a teacher teaches. The learner's attention is not only guided by the teacher's movement but also influences the teacher's demonstration of actions. Our experimental results showed that a robot equipped with bottom-up attention induces partners to amplify their body movement, to closely approach the robot, to take longer time for demonstrating an action, and to segment frequently movements, which are consistent with findings about infant-directed action (Brand et al., 2002, Rohlfing et al., 2006, Vollmer et al., 2009). Exaggeration in space and synchronization in time seem to be strategies for teachers to modify their movement.

Based on the results for the questionnaire, we are going to synthetically investigate the relation between attention, learning, and imitation. Our experiment showed that attention shapes interaction. Open questions are what a robot can learn from motionese, how it can imitate, and how it influences further attention and thus interaction. Nagai (Nagai, 2009) demonstrated that examining continuity in the information detected by bottom-up attention enables a robot to extract key actions from motionese. We will extend this study by developing an architecture to link attention, learning, and imitation.

Acknowledgments

This research was supported in part by Global COE Program "Center of Human-Friendly Robotics Based on Cognitive Neuroscience" of the Ministry of Education, Culture, Sports, Science and Technology, Japan. We express our thanks to Dr. Takashi Minato and Mr. Fabio Dalla Libera at JST ERATO Asada Synergistic Intelligence Project for their kind help to maintain the robot.

References

Brand, R. J., Baldwin, D. A., and Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*, 5:72–83.

Brand, R. J. and Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, 11(6):853–861.

Demiris, Y. and Khadhouri, B. (2008). Content-based control of goal-directed attention during human action perception. *Interaction Studies*, 9(2):353–376.

Frank, M. C., Vul, E., and Johnson, S. P. (2009). Development of infants attention to faces during the first year. *Cognition*, 110:160–170.

Golinkoff, R. M. and Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, 15(1):30 – 33.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14:531–595.

Greenfield, P. M., Nelson, K., and Saltzman, E. (1972). The development of rulebound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology*, 3:291–310.

Hayashi, M. (2007). A new notation system of object manipulation in the nesting-cup task for chimpanzees and humans. *Cortex*, 43:308–318.

Itti, L., Dhavale, N., and Pighin, F. (2003). Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceediongs of the SPIE 48th Annual International Symposium on Optical Science and Technology*, volume 5200.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

Koterba, E. A. and Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants attention and object exploration. *Infant Behavior and Development*, 32(4):437–444.

Nagai, Y. (2009). From bottom-up visual attention to robot action learning. In *Proceedings of the 8th IEEE International Conference on Development and Learning*.

Nagai, Y., Muhl, C., and Rohlfing, K. J. (2008). Toward designing a robot that learns actions from parental demonstrations. In *Proceedings of the 2008 IEEE International Conference on Robotics and Automation*, pages 3545–3550.

Nagai, Y. and Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, 1(1):44–54.

Rohlfing, K. J., Fritsch, J., Wrede, B., and Jungmann, T. (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Advanced Robotics*, 20(10):1183–1199.

Shimojo, S. (2006). *Beginning to See: Genius of Mind, and the New Infant Science (in Japanese)*. Shinyosha.

Thomaz, A. L. and Breazeal, C. (2008). Experiments in socially guided exploration: lessons learned in building robots that learn with and without human teachers. *Connection Science*, 20(2-3):91–110.

Vollmer, A.-L., Lohan, K. S., Fischer, K., Nagai, Y., Pitsch, K., Fritsch, J., Rohlfing, K. J., and Wrede, B. (2009). People modify their tutoring behavior in robot-directed interaction for action learning. In *Proceedings of the 8th IEEE International Conference on Development and Learning*.