Proceedings of the 4th International Symposium on
Imitation in Animals and Artifacts
pp. 299-306, April 2007

Can Motionese Tell Infants and Robots "What to Imitate"?

Yukie Nagai [1] and **Katharina J. Rohlfing** [2]

Abstract. An open question in imitating actions by infants and robots is how they know "what to imitate." We suggest that parental modifications in their actions, called *motionese*, can help infants and robots to detect the meaningful structure of the actions. Parents tend to modify their infant-directed actions, e.g., put longer pauses between actions and exaggerate actions, which are assumed to help infants to understand the meaning and the structure of the actions. To investigate how such modifications contribute to the infants' understanding of the actions, we analyzed parental actions from an infant-like viewpoint by applying a model of saliency-based visual attention. Our model of an infant-like viewpoint does not suppose any a priori knowledge about actions or objects used in the actions, or any specific capability to detect a parent's face or his/her hands. Instead, it is able to detect and gaze at salient locations, which are standing out from the surroundings because of the primitive visual features, in a scene. The model thus demonstrates what low-level aspects of parental actions are highlighted in their action sequences and could attract the attention of young infants and robots. Our quantitative analysis revealed that motionese can help them (1) to receive immediate social feedback on the actions, (2) to detect the initial and goal states of the actions, and (3) to look at the static features of the objects used in the actions. We discuss these results addressing the issue of "what to imitate."

1 INTRODUCTION

Imitation learning is a promising approach for robotics researchers to enable their robots to autonomously acquire new skills from humans [21, 31]. It allows robots to learn new behaviors by first observing human movements and then reproducing them by mapping into their motor commands. It consequently reduces the efforts of designers in developing robots' behaviors. In addition to these engineering benefits, the research on imitation learning leads us to the deeper understanding of human intelligence [2]. Human infants, even neonate [25, 26], are able to imitate actions. In the course of their development, infants can reproduce actions and the goal of actions shown by another person. The ability to imitate is moreover discussed as a route to their further cognitive development, e.g., the differentiation of the self and other, the understanding of other's intention, and the use of language [9]. Thus, to investigate the mechanism for imitation learning from a constructivist viewpoint allows us to uncover human intelligence [2].

There are some advantages in robot imitation, however, we still have an open question of how robots know "what to imitate" and "how to imitate." Nehaniv and Dautenhahn [28, 29] discussed these two fundamental issues in robot imitation. Breazeal and Scassellati [7, 8] also pointed out the issues and reported the current techniques used in robot systems. When a robot attempts to imitate a human action or a sequence of his/her actions to achieve a goal-oriented task, it has to first detect the movements of the person and then determine which movements are relevant to the task. A robot without any a priori knowledge about the task does not know which actions of the person are important and necessary for the task, while he/she sometimes produces not only actions directly related to the task but also unrelated ones. It is also required to detect the initial and goal states of the actions and the objects involved in the actions so that a robot can imitate the sequence of the actions not only at a trajectory level but also at a goal level. These problems are stated as the issue of "what to imitate," and several approaches have been proposed from different perspectives (e.g., [4, 6, 10, 11, 34]).

Another issue to be solved in robot imitation is how a robot knows "how to imitate." A robot that tries to imitate human actions has to be able to transform the observed actions of a person into its motor commands so as to reproduce the same actions or to achieve the same goal of the actions. A difficulty in transforming the actions is that a robot cannot access to the somatosensory information of the person and is thereby unable to directly map the actions into the motor commands. Moreover, the body structure of a robot is usually different from the person's, which makes the problem more difficult. These issues are called "how to imitate" and have been investigated from various approaches (e.g., [1, 3, 4, 10]).

In addressing these issues from a standpoint of cognitive developmental robotics [2], we suggest that parental modifications in their infant-directed actions can help robots as well as infants to imitate the actions [12, 30]. When infants attempt to imitate actions presented by their parents, they also face the same problems: "what to imitate" and "how to imitate." Although infants are supposed to have little semantic knowledge about actions as robots do, they are surprisingly able to imitate the actions. They are skillful in processing a stream of ongoing activity into meaningful actions and organizing the individual actions around ultimate goals [33]. We thus consider that parental actions aid infants solving "what to imitate" and "how to imitate." It is known that parents tend to modify their actions when interacting with their infants (e.g., [5, 30]). They, for example, put longer and more pauses between their movements, repeat the same movements, and exaggerate their movements when interacting with infants compared to when interacting with adults. Such modifications, called *motionese*, are suggested to aid infants structuring the actions and understanding the meaning of the actions. However, we do not know yet how it actually affects and contributes to the infants' understanding of the actions. Because the current researches have analyzed motionese only from an adult's viewpoint, i.e., they focused

[1] Bielefeld University, Germany, email: yukie@techfak.uni-bielefeld.de
[2] rohlfing@techfak.uni-bielefeld.de

only on the actions relevant to a task, it is still unclear what aspects of parental actions would be attended to by infants and how they help infants to understand and imitate the actions.

We analyze motionese from an infant-like viewpoint and discuss how it can help infants and robots to detect "what to imitate." Our model of an infant-like viewpoint does not suppose any a priori knowledge about actions or objects used in the actions. It does not know which parental actions are relevant to a task, what the goal of the task is, or what objects are involved in the task. Furthermore, it is not equipped with any specific ability to detect a parent's face or his/her hands. Instead, it is able to detect and gaze at outstanding locations in a scene. To simulate such a capability of visual attention, we adopt a model of saliency-based visual attention [16, 17] inspired by the behaviors and the neural mechanism of primates. A salient location in this model is defined as a location which locally stands out from the surroundings because of its color, intensity, orientation, flicker, and motion [16]. It thus can demonstrate what low-level aspects of parental actions are highlighted in their action sequences and could attract the attention of young infants and robots. We analyze motionese with the model and discuss the results toward solving the issue of "what to imitate."

The rest of this paper is organized as follows. In Section 2, we summarize the current evidences of motionese from psychological and computational studies. In Section 3, we introduce the model of saliency-based visual attention and describe the benefits of using it for the analysis of motionese. Next, we show analytical experiments of motionese in Section 4, and discuss the experimental results in Section 5. Finally, we conclude with future directions in Section 6.

2 PARENTAL MODIFICATIONS IN INFANT-DIRECTED INTERACTIONS

It is well known that parents significantly alter the acoustic characteristics of their speech when talking to infants (e.g., [19]). They, for example, raise the overall pitch of their voice, use wider pitch, slow the tempo, and increase the stress. These phenomena, called *motherese*, are suggested to have the effects of attracting the attention of infants and providing easily structured sentences to infants, which consequently facilitates their language learning.

In contrast to motherese, motionese is phenomena of parental modifications in their actions. Parents tend to modify their actions when interacting with infants so that they maintain the attention of infants and highlight the structure and the meaning of the actions as in motherese. Brand et al. [5] revealed that mothers altered their actions when demonstrating the usage of novel objects to their infants. They videotaped mothers' interactions first with an infant and then with an adult, and manually coded them on eight dimensions: the proximity to the partner, the interactiveness, the enthusiasm, the range of the motion, the repetitiveness, the simplification, the punctuation, and the rate. Their results comparing the infant-directed interactions (IDI) and adult-directed interactions (ADI) revealed significant differences in the first six dimensions out of the eight (higher rates in IDI than in ADI). Masataka [22] focused on a signed language and found that deaf mothers also altered their signed language. He observed deaf mothers when interacting with their deaf infants and when interacting with their deaf adult friends, and analyzed the characteristics of their signs. His comparison indicated that, when interacting with infants, deaf mothers significantly slowed the tempo of signs, frequently repeated the same signs, and exaggerated each sign. His further experiments showed that such modifications in a signed language attracted greater attention of both deaf and hearing infants [23, 24]. Gogate et al. [14] investigated the relationship between maternal gestures and speech in a object-naming task. They asked mothers to teach their infants novel words by using distinct objects and observed how the mothers used their gestures along with their speech. Their results showed that mothers used the target words more often than non-target words in temporal synchrony with the motion of the objects. They thus suggested that maternal gestures likely highlighted the relationship between target words and objects, of which effects were demonstrated in their further experiment [13]. Iverson et al. [18] also revealed that maternal gestures tended to co-occur with speech, to refer to the immediate context, and to reinforce the message conveyed in speech in daily mother-infant interactions. Their analysis moreover showed positive relationships between the production of maternal gestures and the verbal and gestural productions and the vocabulary size of infants.

In contrast to the former studies, in which motionese was manually coded, Rohlfing and her colleagues [12, 30] applied a computational technique to evaluate motionese. They adopted a 3D body tracking system [32], which was originally developed for human-robot interactions, to detect the trajectory of a parent's hand when he/she was demonstrating a stacking-cups task to his/her infant first and then to an adult. Their quantitative analysis revealed that parents put longer and more pauses between actions and decomposed a rounded movement into several linear movements in IDI compared with in ADI. They suggested with these results that motionese can help infants and robots to detect the meaning of actions. This approach is very attractive for robotics researchers because their model can be immediately implemented into robots and enables them to leverage the advantages of motionese in imitation learning. However, it is still an open question how robots know "what to imitate." Although their study as well as the former studies showed that parents modify their task-relevant actions so as to be easily understood, robots as well as young infants do not know which parents' actions are relevant to a task. To address this problem, we apply a model of saliency-based visual attention to the analysis of motionese.

3 SALIENCY-BASED VISUAL ATTENTION

3.1 Architecture of model

To analyze motionese from an infant-like viewpoint, i.e., without any a priori knowledge about actions or objects used in the actions, we adopt a model of saliency-based visual attention [16, 17]. The model, inspired by the behavior and the neuronal mechanism of primates, can simulate the attention shift of humans when they see natural scenes. Humans are able to rapidly detect and gaze at salient locations in their views. A salient location here is defined as a location which locally stands out from the surroundings because of its color, intensity, orientation, flicker, and motion [16]. For example, when we see a white ball in a green field, we can rapidly detect and look at the ball because of its outstanding color, intensity, and orientation. When a dot is moving left while a number of dots moving right, the former dot will be tracked visually because of its distinguished motion. The model of saliency-based visual attention imitates such a primal but adaptable attention mechanism of humans.

Figure 1 shows the overview of the model used in our experiment. This is the same as the model proposed in [16] excepting the absence of the mechanism of "inhibition of return," which inhibits the saliency of locations that have been gazed at. It means that our model determines attended locations frame by frame independently. The model works as follows:

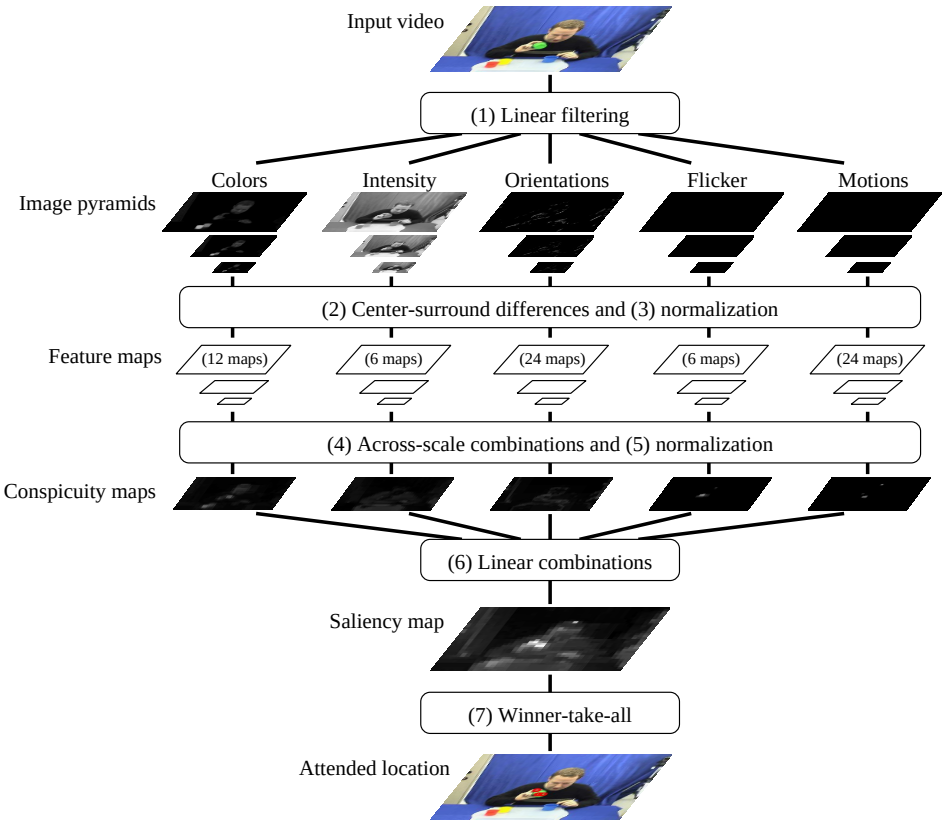


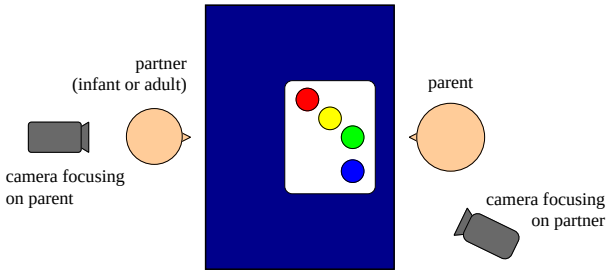
Figure 1. A model of saliency-based visual attention, which was revised from original one proposed in [17]

1. Five visual features (colors, intensity, orientations, flicker, and motions) are first extracted by linearly filtering a frame of an input video, and then image pyramids with different scales are created.
2. The differences between a center-fine scale and a surround-coarser scale image are calculated to detect how much each location stands out from the surroundings.
3. The center-surround differences are normalized to first eliminate modality-dependent differences and then globally promote maps containing a few conspicuous locations while globally suppressing maps containing numerous conspicuous peaks. The results are called feature maps.
4. The feature maps are combined through the across-scale addition to get together the different scales into one map.
5. The combined maps are normalized again to obtain conspicuity maps.
6. The conspicuity maps of the five features are linearly summed into a saliency map.
7. Finally, the most salient locations in the saliency map are selected as the attended locations in the frame.

In our analysis, image locations of which saliency were higher than the maximum $\times$ 0.9 in each frame were selected as the attended locations. That is, not only one location but several locations could be attended to in a frame. Refer to [16, 17] for more detail explanations of the processing.

3.2 Benefit of applying model to analysis of motionese

Applying the model to the analysis of motionese enables us to reveal what visual features of parental actions are highlighted in their action streams and could attract the attention of young infants and robots. Over the first year of life, infants semantic knowledge of actions, such as environmental, social, and psychological constraints on their organization and structure, is quite limited in comparison to adults. Thus, infants do not clearly understand the meaning or the structure of the actions when they see the actions for the first time. They also have limited information about objects, e.g., what objects are involved in the actions and what the initial and goal states of the objects are. Instead, they are certainly able to detect and gaze at salient locations in their views. For example, when colorful toys are shown to infants (usually, infants' toys have bright colors like yellow, red, and blue), they will look at the toys because of their salient colors. When a parent moves his/her hand to grasp and manipulate the toys, the hand as well as the toys will attract the attention of infants. Assuming only perceptual saliency, a parent's face can also attract the infants' attention because both of its static visual features and of its movement caused by his/her smiling and talking. Note that a parent's face and his/her hands can be attended to as salient locations without supposing any specific capability to detect their features or even skin color. We aim at evaluating how much meaningful structures of parental actions are detected without any knowledge about actions, objects, or humans, and how they can contribute to solving the problem of "what to imitate."



(a) top-view of experimental setup

(b) image frame of video focusing on parent, which was used as input to saliency-based attention model

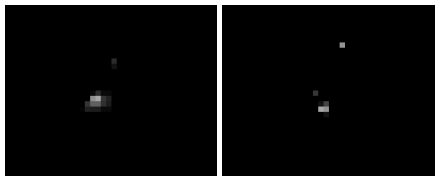
(c) image frame of video focusing on infant

Figure 2. Experimental setup and sample image frames of videos

(a) input image, in which attended locations denoted by circles

(b) saliency map (sum of (c)-(g))

(c) color map (d) intensity map (e) orientation map



(f) flicker map (g) motion map

Figure 3. Example of saliency map equally summing up five conspicuity maps and attended locations

4 ANALYSIS OF MOTIONESE WITH SALIENCY-BASED ATTENTION MODEL

4.1 Method

We analyzed the videotaped data used in [30]. In contrast to [30], in which only the task-related parental actions were analyzed, we dealt with all visual features in the scenes.

4.1.1 Subjects

Subjects were 15 parents (5 fathers and 10 mothers) of preverbal infants at the age of 8 to 11 months ($M = 10.56$, $SD = 0.89$). We chose this age because infants start to imitate simple means-end actions such as acting on one object to obtain another [33] and to show the understanding of goal-directed actions at 6 months of age [20].

4.1.2 Procedure

Parents were instructed to demonstrate a stacking-cups task to an interaction partner while explaining him/her how to do it. The interaction partner was first their infants and then an adult. **Figure 2** (a) illustrates the top-view of the experimental setup, and (b) and (c) show sample image frames of cameras which were set behind a parent and a partner and focused on each of them. The stacking-cups task was to sequentially pick up the green, the yellow, and the red cups and put them into the blue one on the white tray.

4.1.3 Analysis

We analyzed videos recording the parents' actions as shown in Figure 2 (b). The videos were input to the model of saliency-based visual attention, and image locations with high saliency were detected as the attended locations frame by frame. **Figure 3** shows how the attended locations were determined in a frame: (a) shows an input image (320 $\times$ 256 [pixels]), in which three attended locations are denoted by red circles, and (b) shows the saliency map of the scene (40 $\times$ 32 [pixels]), which sums up the five conspicuity maps: (c) the color, (d) the intensity, (e) the orientation, (f) the flicker, and (g) the motion maps. The view of the maps corresponds to the input image, and the brightness of the pixels represents the degree of saliency, i.e., white means high saliency while black means low. In the example, the father was showing the green cup to his infant by shaking it, and therefore the cup and his right hand were attended to by the model. The color map extracted the green, the yellow, and the red cups as well as the father's face and hands as salient locations, while the intensity map detected the white tray and the father's black cloth. The orientation map detected the father's face, his hands, and the contour of the tray because of their rich edges. The flicker and the motion maps extracted the father's right hand with the green cup because of their movement. As a result, the saliency map, which equally summed up

the five conspicuity maps, detected the three highly salient locations in the scene (see Figure 3 (a)). Note that our model selected the locations of which saliency was higher than the maximum $\times$ 0.9 in each frame, which allows us to evaluate the general tendency of parental actions. Through our experiment, the blue cup was not salient due to the blue background.

4.2 Results

4.2.1 Proportion of attended locations

We first compared how often a parent's face, his/her hands, and the cups were attended to by the model in IDI and in ADI. The attended locations were automatically classified using the predefined colors and positions of the targets. The results were compared separately in three time periods: before, during, and after the task. The start and the end of the task were defined when a parent picked up the first cup and when he/she put down the final cup into the blue one, respectively. The length of the periods before and after the task was 2 [sec].

Figures 4, **5**, and **6** show the results for the periods before, during, and after the task. In each graph, the horizontal axis denotes the label of the subjects, and the vertical axis denotes the proportion at which (a) a parent's face, (b) his/her hands, and (c) the cups were attended to over the period. When an attended location was at none of them, e.g., at a parent's cloth and at the tray, it was counted as (d) the others. The means and the standard deviations are listed in **Table 1**.

Before task: The non-parametric test (the Wilcoxon test) revealed significant differences in the proportion of attention on the cups (Figure 4 (c); $Z = -2.045,\ p < 0.05$) and in that on the others ((d); $Z = -1.988,\ p < 0.05$). It indicates that the cups attracted more attention in IDI than in ADI, and that the others were less attended to in IDI than in ADI.

During task: The non-parametric test revealed a significant difference in the proportion of attention on a parent's face (Figure 5 (a); $Z = -2.556,\ p < 0.05$). It also showed a statistical trend in the proportion of attention on parent's hands ((b); $Z = -1.817,\ p = 0.069$). A parent's face attracted much more attention in IDI than in ADI while his/her hands attracted less attention in IDI than in ADI.

After task: The non-parametric test revealed a statistical trend in the proportion of attention on a parent's face (Figure 6 (a); $Z = -1.874,\ p = 0.061$). The parametric t-test showed a trend in the proportion of attention on the cups ((c); $t(14) = 1.846,\ p = 0.086$). These results suggest that a parent's face was attended to in ADI more than in IDI, and that the cups were attended to in IDI more than in ADI.

4.2.2 Contribution of static features to saliency of objects

We next analyzed how much the static visual features of the cups contributed to their saliency in IDI and in ADI. Here the static features include the color, the intensity, and the orientation while the motion features include the flicker and the motion. The sum of the degrees of saliency derived from the static features was compared between IDI and ADI.

Figure 7 shows the contribution rate of the static features to the saliency of the cups (a) before, (b) during, and (c) after the task. **Table 2** lists the means and the standard deviations. The non-parametric test (the Wilcoxon test) revealed significant differences in the contribution rates before the task (Figure 7 (a); $Z = -2.040,\ p < 0.05$) and during the task ((b); $Z = -3.045,\ p < 0.05$). It indicates that in the two time periods the static features much more contributed to the saliency of the cups in IDI than in ADI.

Table 1. Proportions of attended locations

		IDI		ADI	
		M	SD	M	SD
before task	parent's face	0.070	0.104	0.049	0.047
	parent's hands	0.583	0.171	0.521	0.192
	cups	0.289	0.145	0.196	0.185
	others	0.216	0.184	0.356	0.220
during task	parent's face	0.040	0.038	0.019	0.017
	parent's hands	0.680	0.150	0.715	0.127
	cups	0.448	0.117	0.433	0.112
	others	0.089	0.088	0.089	0.083
after task	parent's face	0.085	0.103	0.154	0.117
	parent's hands	0.484	0.311	0.475	0.239
	cups	0.306	0.198	0.180	0.123
	others	0.230	0.232	0.270	0.176

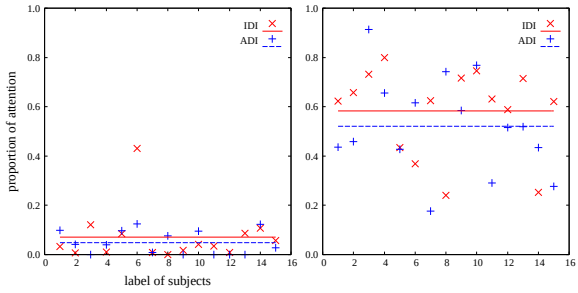
Table 2. Contribution of static features to saliency of cups

	IDI		ADI	
	M	SD	M	SD
before task	0.461	0.331	0.240	0.267
during task	0.256	0.203	0.090	0.100
after task	0.650	0.349	0.421	0.405

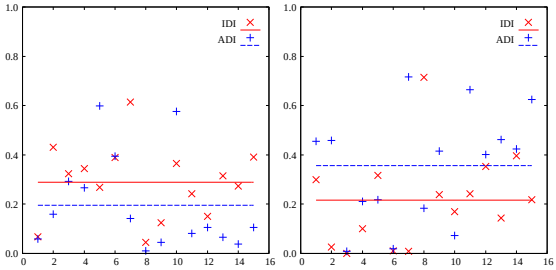
5 DISCUSSIONS

Our first focus of analysis revealed that a parent's face attracted much more attention in IDI than in ADI during the task while it attracted less attention in IDI than in ADI after the task. A reason is that the parents in IDI often talked to and smiled at their infants when demonstrating the task. They commented on each action while executing it, tried to maintain the infants' attention by addressing them verbally, and tried to get the infants interested in the task by showing emotional expressions. These behaviors caused movements on the parents' faces and made them more salient than others (see **Figure 8** (a)). By contrast, in ADI the parents rarely talked to or smiled at the adult partner during the task but explained the task after finishing it. Thus, their faces attracted more attention after the task. The result that the parents' hands were more attended to in ADI than in IDI during the task also indicates that their faces did not often move compared to their hands. We suggest from these results that parents give their infants immediate feedback on their actions, which helps infants to detect what actions are important and relevant to the task.

Our further analysis focusing on the objects involved in the task revealed that the objects were more salient in IDI than in ADI before and after the task. The saliency emerged because the parents interacting with their infants tended to put longer pauses before and after the task. While many of the parents in ADI started the task without checking whether the adult partner looked at the task-relevant locations, in IDI, they looked at the infants first and then started the task after confirming the infants' attention on the cups (see Figure 8 (b)). They also tried to attract the infants' attention on the cups by shaking them before the task. The result that the other locations attracted less attention in IDI than in ADI before the task also indicates that the parents made much effort to attract the attention of infants on the

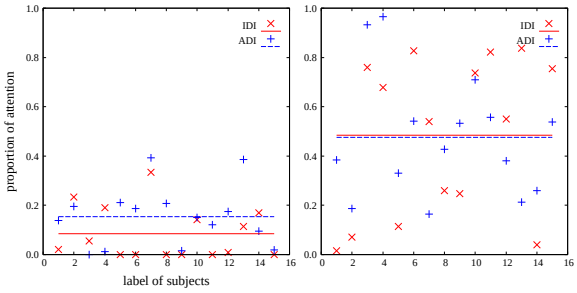


(a) parent's face (b) parent's hands

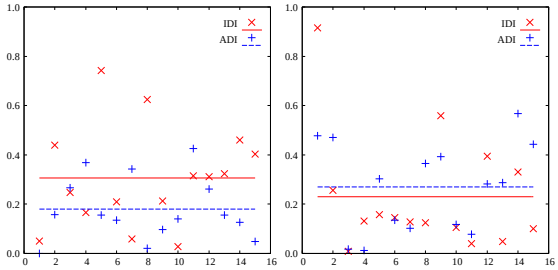


(c) cups (d) others

Figure 4. Proportions of attended locations before task (2 [sec])

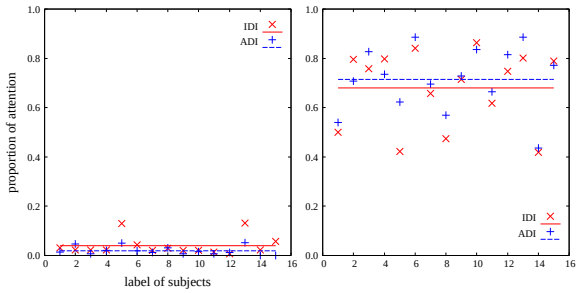


(a) parent's face (b) parent's hands

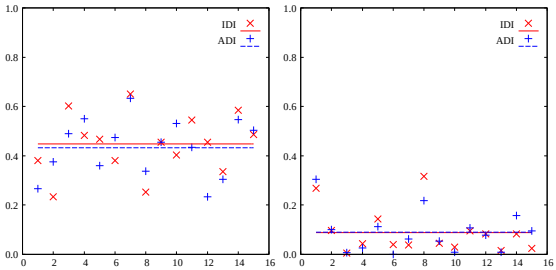


(c) cups (d) others

Figure 6. Proportions of attended locations after task (2 [sec])

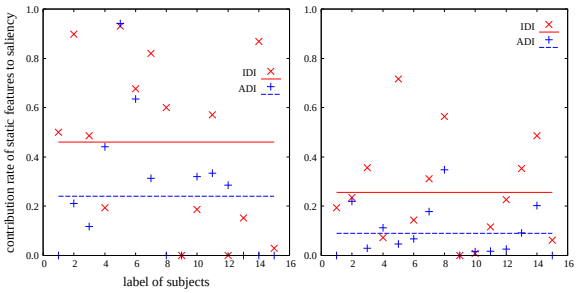


(a) parent's face (b) parent's hands

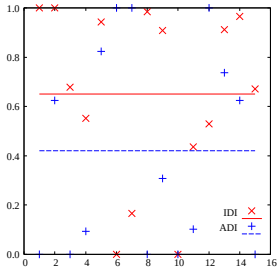


(c) cups (d) others

Figure 5. Proportions of attended locations during task



(a) before task (b) during task



(c) after task

Figure 7. Contribution of static features to saliency of cups

(a) parent's face attended to during task in IDI

(b) cups attended to before task in IDI

(c) cups attended to after task in IDI

Figure 8. Examples of attended locations, which are indicated by a red, a green, or a blue box if they are on a parent's face, on his/her hands, or on the cups, respectively

task-related locations. In addition, the parents in IDI tended to stop their movement and look at the infants for a while after the task (see Figure 8 (c)) while the parents in ADI continued to move and commented a lot on the task. They likely showed the goal state of the task to the infants. We therefore suggest that parents aid their infants detecting the initial and goal states of the actions by inserting longer pauses before and after the task.

Our analysis on the contribution of the static features to the saliency of the objects showed that the features of the color, the intensity, and the orientation of the cups contributed much more to their saliency in IDI than in ADI. When the cups are attended to as salient locations, two reasons are considered: motion and static visual features. In IDI the saliency of the cups was derived not only from their movement but also from their intrinsic features, i.e., the color, the intensity, and the orientation, while in ADI the saliency was mostly came from their movement. The reason is that the parents in IDI often stopped their movement during the demonstration of the task and tried to attract the infants' attention not on their hands' motion but on the cups they were holding. Thus, the cups were attended to as salient locations because of their intrinsic features. We suggest with these results that parental actions help infants to detect the static features of the objects, which consequently enables them to better perceive the physical structure of the objects.

Although these findings are already very significant, some results are considered to be improved. Our analysis, for example, found a trend but did not reveal a statistically significant difference between the proportions of attention on the cups in IDI and in ADI after the task. Before the experiment, we hypothesized that the cups would attract much more attention in IDI than in ADI after the task as before the task. The reason why the cups were not so salient after the task is the blue background. In the goal state, all of the green, the yellow, and the red cups were put in the blue one, which means only the blue one was visible. Thus, the blue cup in the blue background was not detected as a salient location. We will therefore analyze other tasks using other colored objects to evaluate our hypothesis.

The position of the camera which recorded parents' actions also can be optimized. The camera was set higher than the head position of infants so that the view of the camera was not occluded by the infants. This position caused less saliency of the parents' faces because they always looked down to gaze at infants. We will thus change the position of the camera so that we can analyze motionese from a real infant viewpoint.

6 CONCLUSION

Our analysis on parental actions using a saliency-based attention model revealed that motionese can help infants (1) to receive immediate social feedback on the actions, (2) to detect the initial and goal states of the objects used in the actions, and (3) to look at the static features of the objects. In imitation learning, immediate feedback on the actions may allow infants to detect what actions are important and should be imitated. To look at the initial and goal states of the objects may be helpful in understanding the intention of the actions and in imitating the actions not only at the trajectory level but also at the goal level. To attend to the static features of the objects may also help infants to perceive the structure and the configuration of the objects. Therefore, all these results indicate that parental actions contribute to highlight the meaningful structures of the actions. We conclude that motionese can help infants to detect "what to imitate" and that the saliency-based attention model enables a robot to leverage these advantages in its imitation learning.

In contrast to current studies on robot imitation, in which a robot was given the knowledge about task-related actions and/or the goal of actions, our analysis showed that motionese enables a robot to detect these features autonomously. The model of saliency-based visual attention could highlight them in the sequences of parental actions. However, to solve the problem of "what to imitate," we still need to answer the following question. Which characteristics of actions, i.e., the trajectory or the goal of actions, should be imitated? We intend to further analyze motionese with respect to this problem.

We will also address the issue of "how to imitate." A robot that attempts to imitate human actions has to know how to transform the human movement into its own movement. To approach this problem, we propose a simple mapping from human movement detected in a robot's vision to the motion primitives of the robot represented in its somatic sense is enough to make the robot roughly imitate the actions [15, 27]. The motion primitives are designed with a set of neurons that are responsible to different motion directions while human movement is also detected and represented with neurons that are responsible to different motion directions [27]. We will develop such a mechanism and evaluate together with the attention model if they enable robots to imitate human actions by leveraging motionese.

REFERENCES

[1] Aris Alissandrakis, Chrystopher L. Nehaniv, and Kerstin Dautenhahn, 'Action, state and effect metrics for robot imitation', in *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, (2006).

[2] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi, 'Cognitive developmental robotics as a new paradigm for the design of humanoid robots', *Robotics and Autonomous Systems*, **37**, 185–193, (2001).

[3] Aude Billard, 'Learning motor skills by imitation: A biologically inspired robotic model', *Cybernetics and Systems: An International Jounal*, **32**, 155–193, (2001).

[4] Aude G. Billard, Sylvain Calinon, and Florent Guenter, 'Discriminative and adaptive imitation in uni-manual and bi-manual tasks', *Robotics and Autonomous Systems*, **54**(5), 370–384, (2006).

[5] Rebecca J. Brand, Dare A. Baldwin, and Leslie A. Ashburn, 'Evidence for 'motionese': modifications in mothers' infant-directed action', *Developmental Science*, **5**(1), 72–83, (2002).

[6] Cynthia Breazeal and Brian Scassellati, 'A context-dependent attention system for a social robot', in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pp. 1146–1151, (1999).

[7] Cynthia Breazeal and Brian Scassellati, 'Challenges in building robots that imitate people', in *Imitation in Animals and Artifacts*, eds., K. Dautenhahn and C. L. Nehaniv, 363–389, MIT Press, (2002).

[8] Cynthia Breazeal and Brian Scassellati, 'Robots that imitate humans', *Trends in Cognitive Sciences*, **6**(11), 481–487, (2002).

[9] J. Gavin Bremner, *Infancy*, Blackwell Publishers Limited, 1994.

[10] Sylvain Calinon, Florent Guenter, and Aude Billard, 'Goal-directed imitation in a humanoid robot', in *Proceedings of the International Conference on Robotics and Automation*, (2005).

[11] Yiannis Demiris and Gillian Hayes, 'Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model', in *Imitation in Animals and Artifacts*, eds., K. Dautenhahn and C. L. Nehaniv, 321–361, MIT Press, (2002).

[12] Jannik Fritsch, Nils Hofemann, and Katharina Rohlfing, 'Detecting 'when to imitate' in a social context with a human caregiver', in *Proceedings of the ICRA Workshop on Social Mechanisms of Robot Programming by Demonstration*, (2005).

[13] Lakshmi J. Gogate and Lorraine E. Bahrick, 'Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations', *Infancy*, **2**(2), 219–231, (2001).

[14] Lakshmi J. Gogate, Lorraine E. Bahrick, and Jilayne D. Watson, 'A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures', *Child Development*, **71**(4), 878–894, (2000).

[15] Verena V. Hafner and Yukie Nagai, 'Imitation behaviour evaluation in human robot interaction', in *Proceedings of the 6th International Workshop on Epigenetic Robotics*.

[16] L. Itti, N. Dhavale, and F. Pighin, 'Realistic avatar eye and head animation using a neurobiological model of visual attention', in *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology*, pp. 64–78, (2003).

[17] Laurent Itti, Christof Koch, and Ernst Niebur, 'A model of saliency-based visual attention for rapid scene analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(11), 1254–1259, (1998).

[18] Jana M. Iverson, Olga Capirci, Emiddia Longobardi, and M. Cristina Caselli, 'Gesturing in mother-child interactions', *Cognitive Development*, **14**, 57–75, (1999).

[19] Joseph L. Jacobson, David C. Boersma, Robert B. Fields, and Karen L. Olson, 'Paralinguistic features of adult speech to infants and small children', *Child Development*, **54**, 436–442, (1983).

[20] I. Kiraly, B. Jovanovic, W. Prinz, G. Aschersleben, and G Gergely, 'The early origins of goal attribution in infancy', *Consciousness and Cognition*, **12**, 752–769, (2003).

[21] Yasuo Kuniyoshi, Masayuki Inaba, and Hirochika Inoue, 'Learning by watching: Extracting reusable task knowledge from visual observation of human performance', *IEEE Transactions on Robotics and Automation*, **10**, 799–822, (1994).

[22] Nobuo Masataka, 'Motherese in a signed language', *Infant Behavior and Development*, **15**, 453–460, (1992).

[23] Nobuo Masataka, 'Perception of motherese in a signed language by 6-month-old deaf infants', *Developmental Psychology*, **32**(5), 874–879, (1996).

[24] Nobuo Masataka, 'Perception of motherese in japanese sign language by 6-month-old hearing infants', *Developmental Psychology*, **34**(2), 241–246, (1998).

[25] Andrew N. Meltzoff and M. Keith Moore, 'Imitation of facial and manual gestures by human neonates', *Science*, **198**, 75–78, (1977).

[26] Andrew N. Meltzoff and M. Keith Moore, 'Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms', *Developmental Psychology*, **25**(6), 954–962, (1989).

[27] Yukie Nagai, 'Joint attention development in infant-like robot based on head movement imitation', in *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts*, pp. 87–96, (2005).

[28] Chrystopher L. Nehaniv and Kerstin Dautenhahn, 'Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications', in *Interdisciplinary Approaches to Robot Learning, World Scientific Series in Robotics and Intelligent Systems*, eds., J. Demiris and A. Birk, volume 24, (2000).

[29] Chrystopher L. Nehaniv and Kerstin Dautenhahn, 'Like me? measures of correspondence and imitation', *Cybernetics and Systems: An International Journal*, **32**, 11–51, (2001).

[30] Katharina J. Rohlfing, Jannik Fritsch, Britta Wrede, and Tanja Jungmann, 'How can multimodal cues from child-directed interaction reduce learning complexity in robot?', *Advanced Robotics*, **20**(10), 1183–1199, (2006).

[31] Stefan Schaal, 'Is imitation learning the route to humanoid robots?', *Trends in Cognitive Science*, **3**, 233–242, (1999).

[32] Joachim Schmidt, Jannik Fritsch, and Bogdan Kwolek, 'Kernel particle filter for real-time 3d body tracking in monocular color images', in *Proceedings of the Automatic Face and Gesture Recognition*, pp. 567–572, (2006).

[33] J. A. Sommerville and A. L. Woodward, 'Pulling out the intentional structure of action: the relation between action processing and action production in infancy', *Cognition*, **95**, 1–30, (2005).

[34] Ales Ude, Curtis Man, Marcia Riley, and Christopher G. Atkeson, 'Automatic generation of kinematic models for the conversion of human motion capture data into humanoid robot motion', in *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots*, (2000).