Motor Experience Alters Action Perception Through Predictive Learning of Sensorimotor Information

Jimmy Baraglia, Jorge L. Copete, Yukie Nagai, and Minoru Asada
Graduate School of Engineering, Osaka University
2-1 Yamada-oka, Suita, Osaka, 565-0871 Japan
Email: {jimmy.baraglia,jorge.copete,yukie,asada}@ams.eng.osaka-u.ac.jp

***Abstract*—Recent studies have revealed that infants' goal-directed action execution strongly alters their perception of similar actions performed by other individuals. Such an ability to recognize correspondences between self-experience and others' actions may be crucial for the development of higher cognitive social skills. However, there is not yet a computational model or constructive explanation accounting for the role of action generation in the perception of others' actions. We hypothesize that the sensory and motor information are integrated at a neural level through a predictive learning process. Thus, the experience of motor actions alters the representation of the sensorimotor integration, which causes changes in the perception of others' actions. To test this hypothesis, we built a computational model that integrates visual and motor (hereafter, visuomotor) information using a Recurrent Neural Network (RNN) which is capable of learning temporal sequences of data. We modeled the visual attention of the system based on a prediction error calculated as the difference between the predicted sensory values and the actual sensory values, which maximizes the attention toward not too predictable and not too unpredictable sensory information. We performed a series of experiments with a simulated humanoid robot. The experiments showed that the motor activation during self-generated actions biased the robot's perception of others' actions. These results highlight the important role of modalities integration in humans, which accounts for a biased perception of our environment based on a restricted repertoire of own experienced actions.**

I. Introduction

In early infancy, humans are not yet able to detect the goal of others' actions. Later on, infants undergo a developmental process that allows them to perceive others' actions as goal-directed. Several studies have been carried out to reveal when and how infants start understanding goal-directed actions. A remarkable work conducted by Woodward [1] shows that infants (5, 6 and 9 months old) with goal-directed action experience react differently to actors reaching for and grasping objects. For the experiments four actors were presented to the infants: a human arm, a rod, a flat occluder and a mechanical grasping tool. The experiment indicated that when infants were habituated to goal-directed actions (i.e., the human arm condition) they showed a stronger novelty response to test events that varied the goal of the action (e.g., the grasped object) than test events that varied the physical properties of the action (e.g., the motion path). On the other hand, if the actions were not goal-directed (i.e., the rod and the flat occluder conditions), or were goal-directed but difficult to infer the agency of the actor (i.e., the mechanical grasping tool condition), infants did not prefer one type of response to the other (i.e., the goal of the action versus the properties of the action).

Later, Sommerville et al. [2] studied the impact of infants' action production in perception of others' actions. The participants of the experiment were 3-month-old infants. They were divided in two groups. The first group of infants participated in an action task which consisted in interacting with two objects in order to acquire action experience. Due to the fact that 3-month-old infants are not yet able to perform coordinated gaze and manual contact with objects, they put a pair of mittens on the infants' hands so that they can easily make contact with them. The second group did not participate in the action task. Then, during the habituation phase, both groups of infants saw an actor reaching for and grasping one of two objects. Finally, during the test phase, the position of the objects was reversed and the infants saw new test events: a new goal event (i.e., the actor reached the same position as habituation but grasped a different goal), and a new path event (i.e., the actor reached a different position but grasped the same goal). Infants' looking time was measured during the experiment. The results showed that the looking time of the first group of infants was longer than the looking time of the second group in the first trial of the habituation phase. Also, the first group of infants looked significantly longer at the new goal event than the new path event, whereas the second group of infants looked equally to both events. According to Sommerville et al. [2], these findings reflect infants ability to detect the goal structure of action after experiencing object-directed behavior, and to apply this knowledge when perceiving others' actions. Therefore, in order to understand this phenomenon, we find important to clarify the underlying mechanism that accounts for the influence of the motor system on the perceptual system and enables infants to detect the goal in others' actions. Further, we need to explain the connection of this mechanism to the visual attention, in accordance to [2] which found that self-action production leads to an increase of visual attention to action goals. In regard to studies on visual attention, experiments conducted by Kidd et al. [3] showed that, when varying the complexity of a sequence of visual events, the probability of infants (7- and 8-month-old) to look away from those events became higher when the complexity of the stimulus was very low or very high. Their findings suggested that infants allocate their attention in order to maintain an intermediate level of

complexity.

In this study we build a computational model to explain how action production alters perception of others actions as reported by Sommerville et al. [2]. Specifically, we focus on the relation between the visual and motor systems and the effect of action experience on the perception of others' actions. Our hypothesis is that motor activity biases visual perception through the joint representation of visuomotor experiences. As an extension to the aforementioned hypothesis, we also propose that the allocation of visual attention is modulated by the prediction error that results from making correspondence between own experience and others' actions. We carried out experiments using iCub Simulator to validate our hypothesis.

II. Hypothesis

A. Findings in Infant Study

Sommerville et al. [2] reported that infants' action experience alters their perception when observing others' actions. Here we focused on two main findings:

1) Experience apprehending objects initially increases infants' attention to similar reaching events performed by another person (Figure 2 in [2]);
2) Infants with apprehension experience look significantly longer at new goal events than new path events, whereas infants without that experience looked equally to both events (Figure 3 in [2]).

Sommerville et al. [2] suggested that the first finding "may reflect infants ability to recognize correspondences across executed and observed actions and/or an increased motivation to attend to action after acting". Regarding the second finding, they indicated that "experience apprehending objects directly prior to viewing the visual habituation event enabled infants to detect the goal-directed structure of the event in another person's actions".

B. Our Hypothesis

We argue that action experience is acquired through the process of integrating motor and sensory information, and that the joint representation derived from this integration is used to learn to predict other's actions. Further, the motor information contains a representation of the action goal that alters the sensorimotor representation in terms of the goal [4], and that is how the motor experience allows infants to detect the goal in others' actions. Here, the motor information accounts for the set of signals that control the motion of the body, and the signals that encode the final target of the motion. For the case of reaching for an object with the hand, the motor signals controls the arms to move from their current position to a next position, and indicates the final target of the motion action. Based on this argumentation, our hypothesis is that when perceiving others' actions infants make predictions based on the sensory information (Fig. 1-a), whereas when generating actions infants learn to predict motor and sensory information, and consequently to encode the sensorimotor representation of experiences in terms of goals (Fig. 1-b). Therefore, since both action perception and action production share the same predictor, then action production has influence on action observation.

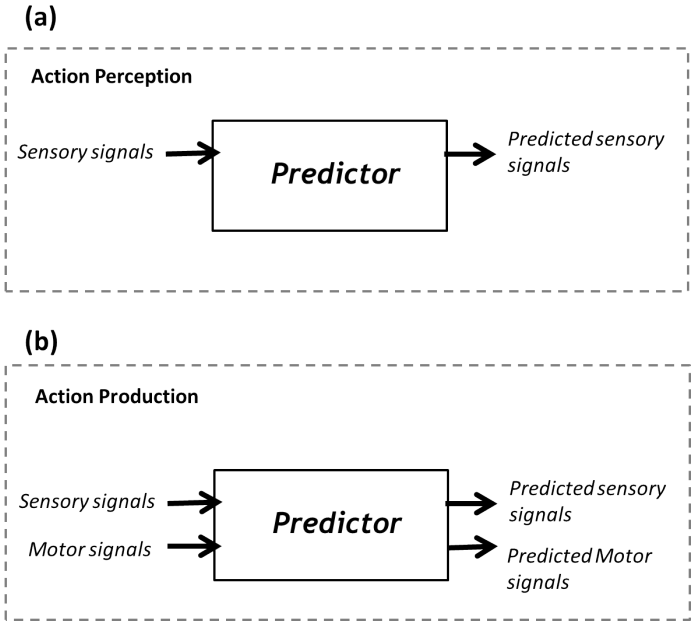

Fig. 1. Our hypothesis. (a) During the action observation infants receive sensory information and predict sensory information. (b) During the action observation infants receive sensory and motor information and predict sensory and motor information.

During this learning process a prediction error arises between the predicted sensory information and the actual one. The magnitude of the prediction error depends on the action experience. We argue that the prediction error plays a fundamental role in the development of several cognitive abilities [5]. Several robotic studies propose that predictive learning could account for the development of imitation [6], helping behavior [7] and goal-directed prediction in early infancy [8]. Kilner et al. [9] showed that prediction error could account for the development of the mirror neuron systems. Their hypothesis is that minimizing the prediction error could help inferring the cause of an observed action. Here, we hypothesize that the prediction error modulates the level of attention to external stimulus through an interest function [5], as shown in Fig. 2. We support our idea based on evidence in [3] according to which visual allocation depends on the complexity of the events, where scenes of middle level complexity draw more attention than those of low and high levels. Thus, the prediction error accounts for the complexity of an event.

We explain findings in [2] in correspondence to our hypothesis by saying that the motor information when infants produce actions changes the sensorimotor representation used to predict sensory information from new events. Then, for the first finding (subsection A-1), we argue that the action experience makes others' actions become more predictable although not completely predictable, which generates an error in the prediction that changes infants' attention. For the second finding (subsection A-2), we say that the action experience

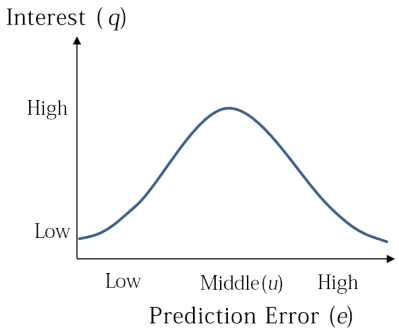

Fig. 2. Visual attention. Curve of interest value in function of the prediction error.

enables infants to make predictions based more on the goal than on the trajectory due to the influence of the motor information, which produces a change in the visual attention as a function of the prediction error. This underline the important role of the motor information representing action goals.

III. Computational Model

We propose a computational model based on our hypothesis which consists of four modules: the visual module, the motor module, the sensorimotor integration, and the visual attention module. The details of each module are explained in following subsections.

A. Motor Module

The motor module generates motor activation signals when the system performs actions. Here, we represent the motor activity as a distinctive set of signals which encode the target of the action and the motor primitive currently generated. This module will output:

- the motor primitives $\mathrm{P} = [p_1, ..., p_m]$ represented as a vector of m binary signals whose components take values 0 or 1, where m is the number of action primitives,
- and the target of the ongoing action $\mathrm{G} = [g_1, ..., g_n]$ as a vector of n binary and mutually exclusive signals whose components take values 0 or 1, where n is the number of target objects in the scene.

For the case of two objects (n=2) and two motor primitives (m=2): arm reaching primitive and arm retracting primitive, the motor module will output a vector $\mathbf{M}$ composed of four activation signals,

$$\mathbf{M}(t) = [g_1(t), g_2(t), p_1(t), p_2(t)], \tag{1}$$

where t represents the time. The choice of variables is based on the idea that infants' actions are goal directed (see goal babbling theory [10]). Thus, it is important to represent both the motion primitives used to perform an action and the targets (goals).

B. Vision Module

The vision module extracts visual information when the system observes actions and then provides the position of the moving effector and the relations between objects in the scene. To do so, the module first extracts and tracks the objects in the scene, then measures the dynamic of the moving effector relative to the objects and finally employs the resulting motion vectors to calculate their relations. Here relations refer to the relative dynamic between objects and the moving effector. For example the moving effector getting closer to (or getting away from) an object is considered a relation. This module will output:

- the position [x, y, z] of the moving effector,
- the vector $\mathbf{R} = [r_{11}, ..., r_{1s}, ..., \mathrm{r}_{n1}, ..., r_{ns}]$ of $s \times n$ possible combinations between the moving effector and n objects for s relations, whose components take values 0 or 1,
- and the vector $\mathbf{RG} = [rg_1, ..., rg_s]$ of s possible relations between the moving effector and any object, whose components take values 0 or 1.

This choice is justified by the fact that infants can be expected to distinguish between objects and actors (see [1]), and therefore to be potentially able to recognize dynamic relations between them. Note that visual relations contained in the vector $\mathbf{R}$ are dependent on the identity of the objects regardless of their positions, while the relations contained in the vector $\mathbf{RG}$ takes accounts all the objects for each relation (e.g., the relation getting closer takes value 1 if the moving effector is getting closer to any object), which guarantees a differentiated representation of the dynamic of the moving effector regardless of the identity of the targeted object.

Thus, for the case of two objects (n=2) and two relations (s=2): getting closer and getting away, the vision module will output a vector $\mathbf{V}$ made of nine signals,

$$\begin{aligned}\mathbf{V}(t) = [&x(t), y(t), z(t), r_{11}(t), r_{12}(t),\\ &r_{21}(t), r_{22}(t), rg_1(t), rg_2(t)],\end{aligned} \tag{2}$$

In our model implementation, the transformation of the visual input is simplified by the use of iCub simulator. The simulator provides the position of the different objects in the scene, and the robot's and others' hand positions. The relation values are calculated based on the velocities of the hand and the objects, namely the derivative of their positions in respect to time. Then, the relations indicate, for example, when the hand's velocity vector points toward or away from an object.

C. Sensorimotor Integration Module

1) Sensorimotor Integration: Biological systems use multiple sensorimotor modalities to interact with their environment. In line with this, studies in robotics have adopted multi-modal integration approaches to model multiple behavior patterns using visuomotor information [11]. We argue that the process of integrating multi-modal information, such as visual and motor information, encodes our experience and is vital to understand others' action goals. Another important aspect to

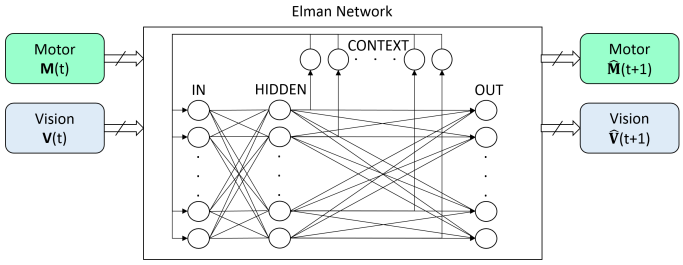

Fig. 3. Internal structure of the sensorimotor integration module based on a Recurrent Neural Network

consider is that humans tend to anticipate near future events when interacting with their environment based on the sensory data they perceive [12]. Recent studies have shown that predictive learning is important to shape infants' perception of the world and direct their attention model [13]. Here, we take advantage of the structure and functionality of the Elman Recurrent Neural Network (RNN) [14] which can perform integration of multiple data and learn the characteristics of temporal sequence data. The model of the neural network is shown in Fig. 3. The inputs of the neural network $\mathbf{I}(t)$ are the outputs from the visual module and the motor module,

$$\mathbf{I}(t) = [\mathbf{V}(t), \mathbf{M}(t)], \tag{3}$$

and the outputs $\mathbf{O}(t)$ are the predicted visual and motor data,

$$\mathbf{O}(t+1) = [\hat{\mathbf{V}}(t+1), \hat{\mathbf{M}}(t+1)], \tag{4}$$

where $\hat{\mathbf{V}}(t+1)$ is the predicted visual information, and $\hat{\mathbf{M}}(t+1)$ is the predicted motor information. The internal composition of $\hat{\mathbf{V}}(t+1)$ and $\hat{\mathbf{M}}(t+1)$ is equivalent to $\mathbf{V}(t)$ and $\mathbf{M}(t)$, respectively. The neural network is trained using the back propagation through time method to minimize the learning error which includes visual and motor prediction errors. We used 13 neurons in the input and output units, and 50 neurons in the hidden and context units, which was empirically determined as the minimum number of neurons for the network to converge.

2) Prediction Error: Humans' capacity to make predictions is strongly influenced by their experience and the lack of experience causes the prediction outcome to differ largely from the actual one. We define this difference as the prediction error. Here, the prediction error $e(t+1)$ when observing others performing an action is calculated as,

$$e(t+1) = |\hat{\mathbf{V}}(t+1) - \mathbf{V}(t+1)|, \tag{5}$$

where $\hat{\mathbf{V}}(t+1)$ is the predicted sensory data, and $\mathbf{V}(t+1)$ is the actual sensory data.

D. Visual Attention Module

For implementing the visual attention module we adopted the findings of Kidd et al. [3] who suggested that infants allocate their attention in order to maintain an intermediate level of complexity. Hereafter we will refer to this as the principle of predictability, where the complexity is represented by the prediction error. Accordingly, the visual attention is assumed to be proportional to an interest value q (Fig. 2). The interest value q is defined as follows,

$$q(t) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{\frac{-(e-u)^2}{2 \cdot \sigma^2}} \tag{6}$$

where α is a scaling factor, σ is the variance and u is the intermediate value of the prediction error, respectively. The interest function is maximized when the prediction error is moderate, that is, when the observed action is not too predictable (i.e., prediction error is low) or not too unpredictable (i.e., prediction error is high).

IV. Experiments

A. Experimental settings

We reproduced similar experimental settings to those described in [2]. This experiment procedure is summarized in Fig. 4. We conducted experiments with the simulated version of the humanoid robot iCub. The experiments considered two scenarios: the watch-first and the reach-first condition. For each experiment, the robot was placed 40 centimeters away of two objects, separated from each other by another 40 centimeters (see Fig. 5). In the watch-first scenario, the system first observed another individual reaching for one of the objects from the same location as the robot. We took advantage of the functionality of iCub simulator to generate the robot's and the others' motions from the same origin and following similar trajectories. We referred to experimental settings in psychological studies that used infants' own visual perspective so they could easily find similarities between self-motion and others' motion [1] [2]. This phase is called the visual habituation. Then, the position of the objects was swapped and the system observed two more actions: reaching for the other object (new goal event) and reaching for the same object (new path event). In the reach-first scenario, the same process was repeated, but this time the system previously experienced reaching for both objects in the action task, before the visual habituation. The three experiments (action task, habituation and new event) were repeated 20 times with random initialization of the weights of the neural network.

B. Action Task

During the action task, the robot's arm moved toward and touched one of two objects, then came back to the initial position and repeated the same action for the same or the other object, randomly (see Fig. 5). We assume that the robot is already able to perform goal directed actions. The robot actions are pre-programmed and executed through the inverse kinematic library of iCub which find the trajectory that minimizes the jerk [15]. This task corresponds to the first stage of training for the reach-first condition, and accounts for the experience of coordinated visuomotor interactions. During the action task, the neural network was trained with both vision and motor data for 500 reaching actions, where each of them

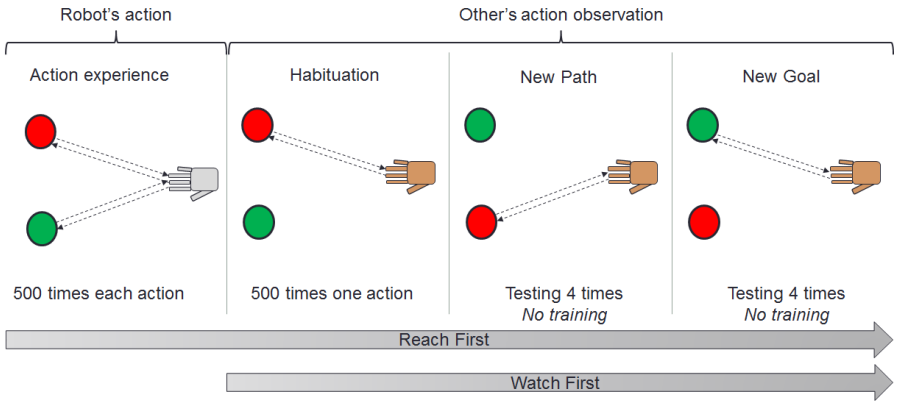

Fig. 4. Procedure of the experiment

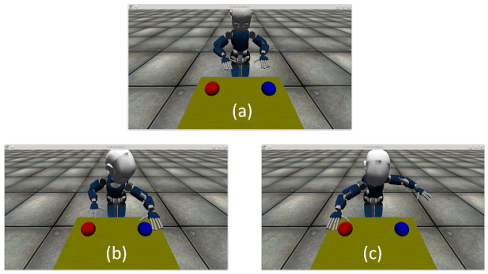

Fig. 5. Experiment setting. (a): initial position before reaching; (b): reaching for object to the left of the robot; (c): reaching for object to the right of the robot.

was composed of 175 steps (i.e., 87500 action steps). Fig. 6 (a) depicts the mean error of the action task over all training trials. The mean error e_m was calculated as the average of the prediction errors in a defined time window of size w = 50 (chosen empirically) in order to attenuate the noise due to the dynamic of the reaching actions, which is not the main target of this study. The mean error is defined as follows:

$$e_{mean}(t+1) = \frac{1}{w}\sum_{i=1}^{w} e(t-i). \quad (7)$$

C. Visual Habituation

The visual habituation procedure consisted of training the neural networks with the visual data when observing others' reaching actions. Here, the neural network was the same as the one trained during the action task for the reach-first condition. Because the motor module was not used, the motor inputs were fixed to 0 and the back-propagation was disabled for both the motor inputs and outputs so that the network does not unlearn the previously acquired motor prediction abilities (in the reach-first condition). During the habituation, the neural network was trained with only the vision for 500 reaching actions, where each of them was composed of 175 steps. Fig. 6 (b) and 6 (c) show the mean error (blue lines) for reach-first condition and watch-first condition, respectively (the initial action step between both habituation graphs differs as the action steps in reach-first condition continues from the last action step in action task). The grey curve in Fig. 6 represents the interest function (see Eq. 6). In our approach, we applied straightforward the evidence in [3] which showed the visual attention shows a bell-shaped distribution in function of the error, and thus we selected a Gaussian-shaped curve to represent the interest function. Here, the intermediate error u (Eq. 6) used for the interest function N was defined as half of the maximum prediction error, where the maximum prediction error value was considered to be 0.365, which corresponded to the maximum mean error value of the condition without experience (e.g., watch-first condition). Hereafter in our discussion, the intermediate error stands as a reference to define whether an error is considered high or low. The variance σ (Eq. 6) was arbitrarily defined to be 0.7 for illustration purposes since it does not alter the relation between high and low errors e and high and low interest q. Further aspects that should be considered for future works will be discussed in Section V.

We can observe from Fig. 6 (b) and (c) that in the reach-first condition the mean error values increased in the first steps (one habituation action), and in the watch-first condition the error was high. However, the error for the reach-first condition was significantly lower than the error in the watch-first condition. This difference is due to the fact that the visual information learned during the action task in the reach-first condition contributed to keep the error relatively low. Fig. 6 also shows that the interest value (grey line) for the reach-first condition was higher than the interest value for the watch-first condition.

In relation to the findings of Sommerville et al. [2], these results may explain why infants in the reach-first condition looked longer to a first visual habituation compared to infants in the watch-first condition. We suggest that the coordinated visuomotor experience contributes to form a representation of own visual experience, which makes others' actions become more predictable but not fully predictable generating a prediction error that increases the visual attention.

D. New Path and New Goal

Finally, we measured the mean error when the goal or the trajectory were changed respect to the original action

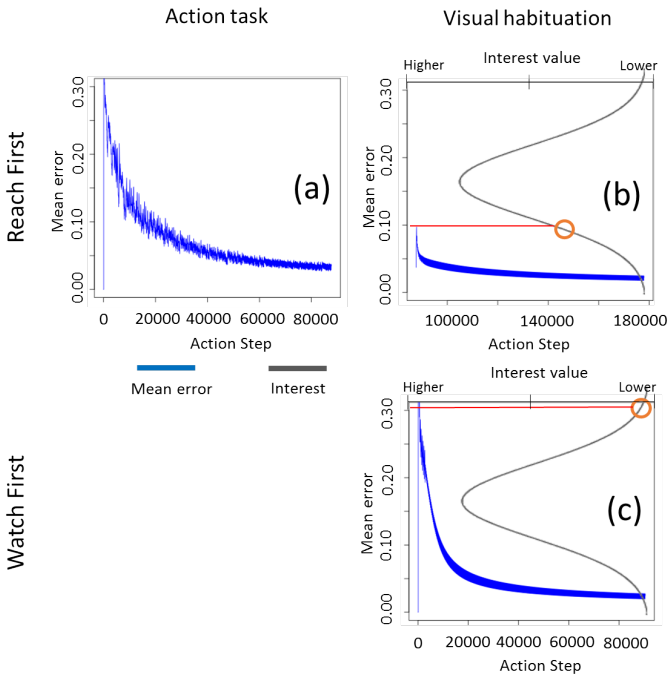

Fig. 6. Training of the neural network for watch-first condition and reach-first condition. The bottom horizontal axis represents the action step, the vertical axis represents the mean error and the top horizontal axis represents the interest value. The blue line and the gray line are the mean error in function of the action step and the interest value in function of the mean error, respectively. The red line and the red point represent the intersection of the mean error with the curve of interest. (a) Mean error during action task for reach-first condition over all training trials; (b) Mean error and interest value during habituation task for reach-first condition; (c) Mean error and interest value during habituation task for watch-first condition

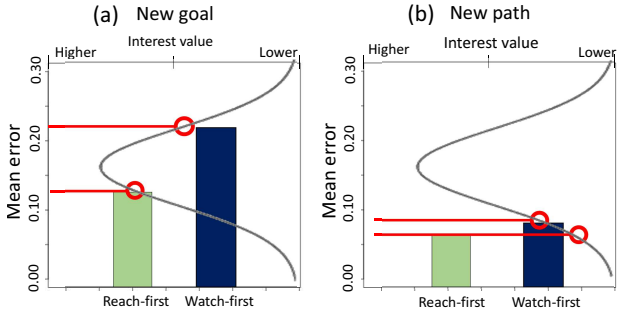

Fig. 7. Error and interest during the test event. In both graphs the bottom horizontal axis represents the condition, the vertical axis represents the mean error and the top horizontal axis represents the interest function. The green and blue bars represent the mean error for the reach-first condition and the watch-first condition, respectively. The gray line, whose independent axis is the top horizontal axis, represents the interest value in function of the mean error. The red line and point represent the intersection of the mean error with the curve of interest. (a) New goal event, (b) New path event.

employed during the visual habituation, namely new goal and new path event, respectively. The graphs of the mean error and the interest value for watch-first and reach-first conditions are shown in Fig. 7. We can see that the change of the path or the goal produced an increase of the prediction error for all conditions, in comparison with their values in the last phase of habituation (see Fig. 6). The results showed that the prediction error of the system without motor experience (i.e., the watch-first condition) took distant values from the middle value of the prediction error u. However, the prediction error of the system with motor experience (i.e., the reach-first condition) took closer values to the middle value u, especially for the new goal event. Therefore, the increase of interest was higher for new goal and reach-first condition than for other conditions.

An explanation for this result is that, in the case of watch-first condition the system learned visual trajectories (x, y, z) and visual relations $(r_{11}(t), r_{12}(t), r_{21}(t)$ and $r_{22}(t))$ that are specific to the targeted object in the habituation, and therefore the system produced a high prediction error for the new goal event (respect to the middle prediction error) since the system could not associate the new targeted goal due to the lack of familiarity, but produced a low prediction error for the new path event since the targeted object was the same. This indicates that in our experiment the visual relations had stronger effect on the prediction error than the trajectory had. Subsequently, if we compare these results to the case of the system with action experience, which had visual experience with both objects, we can say that the experience of motor target signals ($g_1(t)$ and $g_2(t)$) allowed the system to encode action experience in terms of goals, and therefore for the new goal event the visuomotor representation analized the event not in terms of visual appearance but also in terms of the action goal. Consequently we can see that the prediction error in the reach-first condition increased in comparison with the prediction error for the new path event..

V. Discussion

Our computational approach brings significant results that may shed a light on the influence that motor experience has on the perception of others' actions. Our experimental results showed that the system acquired a joint representation of visual and motor experience of coordinated interactions in the action task (i.e., reach-first condition), and was able to apply that visuomotor experience to make predictions of the visual and motor components of others' actions in the visual habituation. The prediction error increased with respect to the new path (same goal) and was closer to the intermediate value when the system had interaction experience (i.e., reach-first condition) than when the system did not have interaction experience (i.e., watch-first condition). Our interpretation of these results is that the system with coordinated visuomotor experience associates visual experience of own actions and visual information of others' actions, which results in a moderate prediction error that stimulates an increase of attention.

The experiment for new path event (same goal, different path of the hand) and new goal event (different goal, same path of the hand) tested both systems (watch-first system and reach-first system) after the habituation stage when they were already familiarized with the scene. The prediction error increased for both watch-first and reach-first conditions, but the prediction

error for new goal (same path) under reach-first condition was closer to the middle error, which resulted in a higher attention, similarly to the experimental results reported by Sommerville et al. [2]. We attribute this result to the contribution of the motor component. During habituation in reach-first condition the system formed an association between its motor experience, including motion target, and the visual information of other's actions by using the joint visuomotor representation. Therefore, since the motor information, which encodes the goal, got encoded in the joint visuomotor representation, when the goal was changed the system perceived the action in terms of the goal and produced a prediction error that led to similar attention increases as those reported by [2]. For the watch-first condition, the results showed an increase in the prediction error for the new goal with respect to the new path. However, we attribute that result to the lack of visual experience with the new object. Our experiments suggest that for future works it is necessary to introduce modifications to the experimental setting proposed in [2]. We consider that the experimental settings proposed in [16] could be adopted, where additional visual experience was provided in the watch-first condition in order to cancel the effect of the lack of visual information. Thus, the link between our results and the psychological findings is straightforward in terms of the strong connection between motor and sensory components, and the influence that own motor experience has on visual attention.

In Fig. 6 and 7 we presented the results including the correspondence with the visual attention based on our visual attention model. In our work we employed a Gaussian-shaped curve and the middle value of the prediction error to establish a relation between prediction error and attention allocation, and the experimental results demonstrated to be in favor of our selection. Nonetheless, we consider that tuning those parameters, including σ (Eq. 6), still requires additional information that should be obtained by considering additional experimental conditions to those in [2] (e.g., measuring the looking time of infants with interaction experience without gloves in the habituation phase). Here, we must highlight the importance of the results in terms of the patterns of the prediction error obtained under the different conditions. Our results demonstrated a clear influence of the motor experience on the perception reflected in the prediction error which we hypothesize ultimately alters the visual attention. Our results proved to be consistent with the perceptual changes in [2], and we consider they may constitute a significant foundation for the understanding and design of cognitive development.

VI. Conclusion

We proposed a computational model to explain psychological findings showing that action production alters the perception of other's actions [2]. The experimental results showed that the influence of the motor experience on the joint representation of own visuomotor experience lead to changes in perception of others' actions. Our hypothesis proved to be valid to explain main findings relating the visuomotor experience and the allocation of visual attention in infants.

As a future work, we propose analyzing the hidden layer of the neural network in order to measure the effect of the motor information in the sensorimotor representation. We believe that the prediction error play a vital role in several domains at a cognitive level. Recent psychological studies have shown that infants make distinctions between path and goal for prediction purposes [12]. Thus, we find relevant to investigate the possible developmental connections between visual attention and prediction of others' actions.

Acknowledgment

This work is partially supported by MEXT/JSPS KAKENHI (Research Project Numbers: 24119003, 24000012, 25700027) and JSPS Core-to-Core Program, A. Advanced Research Networks.

References

[1] A. L. Woodward, "Infants selectively encode the goal object of an actor's reach," *Cognition*, vol. 69, no. 1, pp. 1–34, 1998.

[2] J. A. Sommerville, A. L. Woodward, and A. Needham, "Action experience alters 3-month-old infants' perception of others' actions," *Cognition*, vol. 96, no. 1, pp. B1–B11, 2005.

[3] C. Kidd, S. T. Piantadosi, and R. N. Aslin, "The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex," *PloS one*, vol. 7, no. 5, p. e36399, 2012.

[4] G. Rizzolatti, L. Cattaneo, M. Fabbri-Destro, and S. Rozzi, "Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding," *Physiological reviews*, vol. 94, no. 2, pp. 655–706, 2014.

[5] Y. Nagai, "A model of infant preference based on prediction error: How does motor development influence perception?," in *the Biennial Meeting of the Society for Research in Child Development*, March 2015.

[6] T. Minato, D. Thomas, Y. Yoshikawa, and H. Ishiguro, "A model of the emergence of early imitation development based on predictability preference," in *Development and Learning (ICDL), 2010 IEEE 9th International Conference on*, pp. 19–25, IEEE, 2010.

[7] J. Baraglia, Y. Nagai, and M. Asada, "Prediction error minimization for emergence of altruistic behavior," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, pp. 281–286, IEEE, 2014.

[8] J. L. Copete, Y. Nagai, and M. Asada, "Development of goal-directed gaze shift based on predictive learning," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, pp. 351–356, IEEE, 2014.

[9] J. M. Kilner, K. J. Friston, and C. D. Frith, "Predictive coding: an account of the mirror neuron system," *Cognitive processing*, vol. 8, no. 3, pp. 159–166, 2007.

[10] M. Rolf and J. J. Steil, "Goal babbling: a new concept for early sensorimotor exploration," *Osaka*, vol. 11, p. 2012, 2012.

[11] K. Noda, H. Arie, Y. Suga, and T. Ogata, "Multimodal integration learning of robot behavior using deep neural networks," *Robotics and Autonomous Systems*, vol. 62, no. 6, pp. 721–736, 2014.

[12] E. N. Cannon and A. L. Woodward, "Infants generate goal-based action predictions," *Developmental science*, vol. 15, no. 2, pp. 292–298, 2012.

[13] H. E. Den Ouden, P. Kok, and F. P. De Lange, "How prediction errors shape perception, attention, and motivation," *Frontiers in psychology*, vol. 3, 2012.

[14] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[15] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 1668–1674, IEEE, 2010.

[16] S. A. Gerson and A. L. Woodward, "Learning from their own actions: The unique effect of producing actions on infants' action understanding," *Child development*, vol. 85, no. 1, pp. 264–277, 2014.