On the Sense of Agency and of Object Permanence in Robots

Sarah Bechtle[1], Guido Schillaci[2] and Verena V. Hafner[2]

***Abstract*—This work investigates the development of the sense of object permanence in humanoid robots. Based on findings from developmental psychology and from neuroscience, we link the mechanisms behind the development of the sense of object permanence to those behind the development of sense of agency and to processes of internal simulation of sensory activity. In this paper, we present two experiments. First, a humanoid robot has to learn the forward relationship between its movements and their sensory consequences perceived from the visual input. In particular, we implement a self-monitoring mechanism that allows the robot to distinguish between self-generated movements and those generated by external events. In a second experiment, once having learned this mapping, we exploit the self-monitoring mechanism to suppress the predicted visual consequences of intended movements. We speculate that this process can allow for the development of the sense of object permanence. We will show that, using these predictions, the robot maintains an enhanced simulated image where an object occluded by the movement of the robot arm is still visible, due to sensory attenuation processes.**

I. INTRODUCTION

Adults usually do not have any difficulty in recognising their own body and their own movements. This apparently simple skill is, however, very important to efficiently and naturally interact with the environment and with people. Interacting with objects, for example, would not be that easy, if we were not aware of our body and of the consequences of our actions. However, the mechanisms behind the capability to recognise ourselves are not fully understood. Researchers suggested that self-recognition requires at least awareness of one's body and one's actions [1]. In particular, self-awareness would have two important aspects: a sense of ownership - that is the sense that it is *my body* that is moving - and a sense of agency - that is the sense that *I* am the initiator of the movements and of their consequences [2]. One of the proposals that explain the sense of agency states that our brain implements a self-monitoring mechanism that constantly anticipates the sensory consequences of our actions [3]. Sense of agency would be therefore dependent on the congruence between predicted and observed sensory outcomes of bodily actions.

In this paper, we present an implementation of a self-monitoring mechanism that allows a humanoid robot to anticipate the sensory consequences of self-generated movements. We show that the predictive processes implemented by such a mechanism can lead to the distinction between self-generated actions and those generated by an external subject. In addition, we speculate that the same self-monitoring mechanism adopted for distinguishing between self-generated movements and those generated by other individuals - which could produce markers for a sense of agency - may be behind the development of a sense of object permanence. We present a robotic experiment in support for this assumption.

The rest of the paper is structured as follows. Firstly, we provide a brief introduction to self-recognition (section I-A), to sense of object permanence (section I-B) and to related robotics works (section I-C). Then we present the experiments in section II and the results in section III. Finally, we discuss the outcomes of this study in section IV.

A. *Self Recognition*

How are we able to recognise ourselves? Researchers suggested that our brain implements an active process that refers the body part we observe to a representation of the whole body [1]. This representation of the body that would allow the experiencing of the self is known as body image [1], [2], [7]. Developmental psychologists suggest that self awareness develops over time from very early stages of development. As the development of self-awareness unfolds (already around the age of two months, as suggested by Rochat [6]), infants start having a sense of how their own body is situated in relation to other entities in the environment. Infants at 5 months of age, for example, are already able to distinguish between their own leg movements from those of another infant, when they are displayed in a mirror [6]. By the second year, when linguistic competences start to come into play, self-awareness remains implicit. The sense of agency is thought to be an important aspect of self-awareness. This experience has been proposed to be dependent on the degree of congruence between predicted and actual sensory consequences of our bodily actions [3]. Our brain is thought to implement predictive processes that anticipate the sensory outcomes of our motor actions. Experiencing an action as self-generated would be therefore caused by the anticipation and, thus, the *attenuation* of the sensory consequence of such a motor action, which would be allowed by "the privileged access [that our brain has] to internally generated efferent information during one's own action" [8]. Researchers proposed that sensory attenuation would be the result of a self-monitoring mechanism implemented by our brain, whose existence would explain, for example, why tickling sensations cannot be self-produced [9] or why people perceive the loudness of sounds as less intensive when they are

[1]Sarah Bechtle is affiliated with the Bernstein Center for Computational Neuroscience, Berlin, Germany sarahb@bccn-berlin.de

[2]Guido Schillaci and Verena V. Hafner are affiliated with the Adaptive Systems Group, Humboldt-Universität zu Berlin, Germany guido.schillaci@informatik.hu-berlin.de, hafner@informatik.hu-berlin.de

self-generated than when they are generated by other people [8].

Inspired by the studies mentioned above, we implemented a self-monitoring mechanism in a humanoid robot for the prediction and attenuation of the sensory consequences of self-generated actions. In section III-A, we will show how these processes can enable the distinction between self-generated movements and those generated by external entities.

B. Object Permanence

Many cultures share a popular game, where adults play with infants by hiding their face with the hands and, all of a sudden, showing it back to the kids by saying an utterance[1]. This game seems to demonstrate the difficulties that infants have in understanding object permanence at early stages of development. Some researchers claim that young infants do not understand that an object's existence continues even if the object itself is covered: "out of sight is literally out of mind" [10], [11]. Even though in the literature there is a widely divergent conclusion about how and when object permanence develops [5], [11], [12], [13], Piaget [14] states that by the age of 18-24 months children fully understand object permanence and therefore children first need to be able to recognise their own body as a separate entity in the world in order to be able to understand object permanence. Investigations on the correlation between self-agency and object permanence can be found in the literature in developmental psychology [4], [6], [2]. In this paper, we assume that the self-monitoring mechanism behind the development of sense of agency can allow as well for the development of a sense of object permanence. In particular, we show a robotic experiment in support for this assumption.

How do we test if children understand object permanence or not? What happens in our brain when an object is occluded? In cognitive science, most of the studies correlate the understanding of object permanence to the looking times at possible or impossible occlusion events [11], [12], [13]. In the neurosciences, researchers looked at the EEG signals during occlusion of an object. Consistently over multiple studies [10], [15], a burst of gamma band EEG activity over the temporal lobe was measured during an occlusion event and whenever an object was expected to appear from behind the occluder. Gamma oscillations are thought to be related to active mental representations of objects [10]. Therefore, researchers concluded that the aforementioned burst is related to the infant's mental representation of the occluded object [10]. This brings Kaufman et al. [10] to conclude that increased looking time at an impossible event may not be related to object permanence itself but more to a conflict between the actual visual input and the current mental representation of an object.

In this paper, we show how a robot equipped with a self-monitoring mechanism - implemented for distinguishing between self-generated movements and those generated by other individuals or events - can maintain a mental representation of an occluded stationary object. By executing predictions of the sensory consequences of the motor actions, the robot brain is able to attenuate the movements from the visual input and to generate an enhanced simulated visual input, where the originally occluded object is actually still visible. We speculate, therefore, that mechanisms for self-monitoring and for sensory attenuation, which would be requirements for having a sense of agency, are also behind the development of a sense of object permanence.

C. Related Works

In this section, we introduce robotics studies related to the topics mentioned before.

In a visuo-motor coordination study, Saegusa et al. [16] show how a robot can learn to recognise its own body, implementing a function that correlates the visual input to the arm/head proprioception of the robot. In particular, the authors monitor the correlation between the speed of a moving blob in the camera image and the proprioceptive input. Every time this correlation exceeded a certain threshold, the visuo-motor information is stored in the memory of the robot. The proposed system can allow the robot to query the visuo-motor memory, for example for finding the arm configuration that resulted in high visuo-motor correlation to the current head position, leading the robot to move the arm into its visual field. The same memory also provides a cue for predicting the appearance and location of body parts by predicting the expected visuo-motor correlation of the current body posture.

Michel et al. [17] presented also a motion-based approach for robotic self-recognition. They use a characteristic time window between the initiation of a motor movement and the perception of arm motions to learn self-generated actions. Their implementation of self-recognition conceptually consists of two components. One module incrementally learns the characteristic time delay in the action-perception loop from a sequence of random arm motions within the visual field. A separate classification module uses the learned delay model to identify newly occurring moving objects that satisfy the delay window (thus conceptually belonging to the self). Pitti et al. [18] presented a biological plausible model of spike-timing dependent synaptic plasticity for the emergence of self-agency. In the model, different maps encoded different modalities, including the visual one. Connections between maps are strengthened whenever there is a simultaneous spiking activity. The model is used for executing sensory predictions of intended motor commands. The authors measured the level of self-agency analysing the error between the predicted and observed sensory information.

Weng et al. [19] and Chen et al. [20] studied the development of sense of object permanence in robots. Both papers used an approach based on a novelty parameter leading the robot to look longer at new situations. As mentioned in section I-B, increased looking times can be a possible evidence for the understanding of object permanence in psychology experiments. In typical experiments, young children

[1]"Peek-a-boo" is the English name of this game. The utterance may change from language to language.

are presented with an object that moves along a track part of which is occluded. If infants understand object permanence they should expect that the object continues to pursue its trajectory behind the occluder and to reappear on the other side of the occluder [11]. This paradigm is tested with impossible and possible events. In the case of impossible events, the trajectory of the moving object is obstructed by an obstacle behind the occluder, but the moving object still reappears on the other side of the occluder [11], [12], [13]. Looking time is used as a measurement for the development of sense of object permanence (prolonged looking time corresponds to impossible events). This experimental approach to test whether object permanence is developed or not, was also used by Chen et al. [20] in a robotic experiment where they calculated a novelty parameter depending on the novelty of a situation. In their experiment, the robot would turn its head whenever the novelty parameter would fall below a certain value. They measured the time the robot looked at a specific situation before turning the head. This looking time was an indicator for development of sense of object permanence. Although Chen et al. [20] could measure increased looking time at impossible events whenever impossible events were presented first, they could not show increased looking time at impossible events when they were shown after the possible event.

II. Methodology

In this work, we want the robot to learn what the sensory consequences of its own motor actions are. In particular, the task for the robot is to look at its own movements and to learn the mapping between the motor commands and the movements detected from the visual input. The initial phase of the experiment consisted in training a forward model. A forward model is a type of internal model that can predict the sensory consequence of an action [21]. As shown in figure 1, a forward model takes as input the current sensory state $S(t)$ and the motor command $M(t)$ and outputs the predicted sensory state $S^{*}(t+1)$.

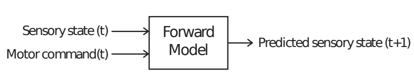

Fig. 1: Schematic representation of a forward model.

A forward model can encode the motor characteristics of the system. Here, we trained the model with sensorimotor data generated executing a self-exploration behaviour (see section II-B). Once having learned the mapping between motor commands and perceived movements, the forward model can be used for implementing a self-monitoring mechanism. Sensory states, such as movements in the visual input, can be predicted - or anticipated - if we feed the forward model with the current sensory state of the system and the intended motor command. Therefore, the predicted sensory state can be compared to the observed one, producing a prediction error that can be monitored, for example, for detecting the occurrence of an unexpected event. The capability of the forward model to anticipate sensory states, such as movements in the visual input, can be also exploited to suppress the predicted visual consequences of intended movements from the actual visual input. We speculate that a similar mechanism can allow for the development of the sense of object permanence. We will show that using these predictions, the robot maintains an enhanced simulated image where an object occluded by the movement of the robot arm is still visible, due to sensory attenuation processes.

A. Robotic Platform

An Aldebaran Nao humanoid robot with 25 degrees of freedom was used. Nao has 5 joints on each of its arms and two HD cameras positioned on the forehead and on the chin of the robot head. In the experiments presented in this paper, only 2 of the 5 joints of the left arm were used, namely the elbow joints (roll and yaw). Built-in functions (NAOqi SDK v1.14.5) were used to control the joints. From these two joints, information about the sensory state $S(t)$ and the motor commands $M(t)$ was extracted, as explained in the next section. As an additional sensory modality, and for encoding $S(t+1)$, we recorded the visual input from the upper camera of the robot.

B. Experimental Setup

The experiments presented in this paper have been carried out using the Cyberbotics Webots robot simulator. Before starting each experiment, the head of the Nao was slightly turned to the left (by 0.3 and 0.1 radians), in order to have a better view of the movements of its left arm. A learning session was performed, consisting in the robot executing a self-exploration behaviour for gathering sensorimotor data. We adopted a simple random walk exploration behaviour in the joint space. Observations consisted in mappings between motor commands and sensory data. The exploration behaviour has been implemented as follows. At a certain point in time t, an action $M(t)$ consisting in a rotation of each of the two elbow joints (sampled from a Gaussian distribution with mean 0.0 and standard deviation 1.0), was applied from the current elbow joint configuration $S(t)$. $S(t)$ and $M(t)$ constitute the sensorimotor context at time t. After the execution of $M(t)$ from the current joint position $S(t)$, new sensory information $S(t+1)$ was gathered from the visual input, containing the information about perceived movements within the Nao's visual field of the upper camera. The images from the camera were collected with a frame-rate of 5 frames per second. Images were collected in the RGB colorspace and then converted into gray scale. The size of the images was 320×240 pixels. The sensorimotor context $(S(t), M(t))$, as well as the observed sensory outcome $S(t+1)$, was collected and saved. In section II-C, we will present in detail how we performed the movement detection and the feature extraction process for estimating $S(t+1)$. In section II-D, we will present the model architecture that we used for learning the sensorimotor mappings.

C. Movement Detection and Feature Extraction

The sensory consequences $S(t+1)$ of motor movements $M(t)$ have been represented by the movements detected from the visual input. In particular, we implemented a motion detection algorithm using the computer vision library OpenCV v2.4.11, in order to allow the robot to visually detect the movements of its hand. Movement detection was performed on gray scale images grabbed from the Nao upper camera. In particular, the absolute difference between two consecutive blurred images is calculated and a threshold is applied to segment out the moving parts of the image. The image size was $320 \times 240 = 76800$ pixels. The outcome of the initial step of feature extraction resulted therefore in a feature vector of size 76800. In order to decrease the dimensionality of the feature vector encoding $S(t+1)$, we applied a 20×20 grid to the difference image. Every dot of the grid stores the sum of all the neighbour pixels that are moving in the surrounding square. By applying this, we were able to reduce the dimensionality of $S(t+1)$ to 400 elements.

D. Model Architecture

We aim at learning the forward relationship between the robot actions and the resulting sensory consequences. As mentioned before, a forward model can be used to predict the sensory consequences of movements and to generate prediction errors, when predictions are compared to actual observations. In particular, the forward model we want to train implements the following mapping:

$$(S(t), M(t)) \quad \longrightarrow \quad S(t+1)$$

In this work, training consisted simply in gathering all the sensorimotor data into a knowledge base. Predictive processes were instead implemented using a simple k-nearest-neighbor (k-NN) algorithm. In particular, a prediction is executed as follows. An input consisting of $S(t)$ and $M(t)$, that is the current elbow joint configuration and the rotations that the system wants to apply to the joints, is fed into a search algorithm. The k training examples closest to the input in the $S(t)$ and $M(t)$ space are extracted. The corresponding values of $S(t+1)$ in the k nearest neighbors are thus averaged and presented as the output of the prediction. In the experiments presented here, we used $k = 3$.

In the training phase, the forward model learns the mapping between motor commands and the movements detected from the visual input. Once the model is trained, we run processes of sensorimotor predictions feeding the forward model with input data gathered during the robot movements. Moreover, prediction errors can be computed by comparing the predicted sensory state $S^*(t+1)$, that is the predicted vector representing the movements in the visual input, to the actual one, that is $S(t+1)$. According to the studies mentioned in the introduction of this paper, prediction errors, as a result of sensory attenuation processes, could represent a cue for sense of agency.

In the following, we will present two experiments. In the first experiment (section III-A), we show an implementation of a self-monitoring mechanism, which computes prediction errors and adopts them as a cue for the sense of agency. The sense of agency, as mentioned in the introduction of this paper, is thought to be dependent on the level of coherence vs. incoherence of sensory predictions and sensory observations. In the experiment, we execute sensory predictions using the forward model trained as explained before and we perform sensory attenuation by subtracting the predictions from the actual sensory observations. We foresee that whenever an unexpected event is detected from the visual input, such as an external agent moving in the environment, we will observe an increase in the prediction error.

In a second experiment (section III-B), we implement sensory attenuation processes, where sensory predictions are subtracted from actual sensory observations. Sensory attenuation, which in the current experiment is performed in the visual domain, generates *enhanced* images where the movements of the robot are attenuated. This leads to situations where stationary objects positioned in front of the robot, which in normal cases would be just hidden by the body of the robot itself, are still visible in the enhanced image. We speculate, therefore, that sensory attenuation processes produced by self-monitoring mechanisms could be at the basis of the development of the sense of object permanence.

III. Experiments and Results

In the following section we will present the experiments performed and the obtained results. Firstly, we will look at the question related to self-agency; therefore, we will discuss the experiments and results regarding object permanence. The prediction error at time $t+1$, that is $e(t+1)$, between the actual sensory outcome $S(t+1)$ and the predicted sensory outcome $S^*(t+1)$, was calculated in all experiments as follows:

$$e(t+1) = \sqrt{\sum_{i}^{n} \left(S(t+1) - S^*(t+1)\right)^2}$$

where i represents one of the $n = 400$ regions in the visual input, as described in section II-C.

A. Sense of agency

In this section, we show a series of experiments where we ran sensorimotor simulation processes under two different conditions. In the first condition, only the robot is present in the scene. While executing different robot behaviours, we ran sensorimotor predictions by feeding a trained forward model with the sensorimotor data gathered during the movements. In a second condition, we had the robot executing the same behaviours and the same prediction processes, although this time facing a moving object. As we will show at the end of this section, prediction errors are higher under the second condition than under the first one, due to the unexpected event of a moving object detected from the visual input. We argue that this phenomenon is resembling those argued by the behavioural studies introduced at the beginning of this paper, suggesting that self-monitoring mechanisms and

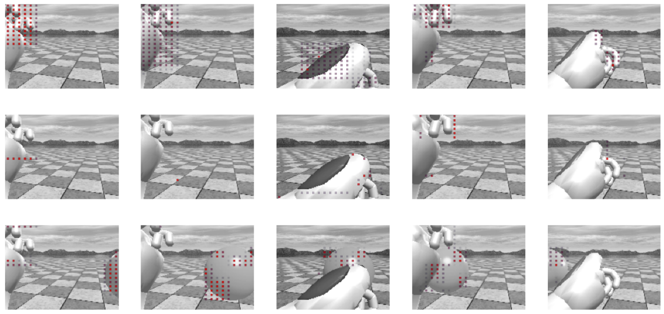

Fig. 2: A sample trajectory of the robot arm. The first row shows the actual sensory outcome that encodes the movements of the robot detected from the visual input; the second row shows the original sensory outcome attenuated using the predictions from the forward model; the third row shows how the sensory attenuation, as in the previous row, looks like when there is an object moving in the background. A coloured dot represents the centroid of a region where a movement has been detected. The transparency of a dot encodes the number of pixels in that region that are moving (fully opaque: all pixels are moving; fully transparent: no movements detected). As evident from the second and third sequences, sensory consequences of self-generated movements are better attenuated than those produced by other agents in the environment.

sensory attenuation processes may be behind the development of a sense of agency in robots.

The robot behaviours executed in this experiment consisted in arm movements where only two joints of the left arm were activated (the elbow joints, as described above). In particular, we implemented 7 different robot behaviours: three behaviours consisted in random movements of the left arm that were observable from the visual input of the robot; two behaviours consisted in random movements of the left arm *not* visible from the robot camera; one behaviour consisted in the robot holding its arm in an idle position visible from its camera; the last behaviours consisted in the robot holding its arm in an idle position *not* visible from its camera. The 7 different behaviours have been run under two conditions: (1) robot alone in the scene and (2) robot facing a moving object. During the execution of each of the behaviours, we fed the forward model with the sensorimotor data gathered from the robot and executed sensory predictions in the visual domain, producing vectors of prediction errors, $E = e_1, e_2, ..., e_T$. As depicted in Table I, we observed, for each behaviour, a statistically significant difference between the mean of the prediction errors under the first condition (robot alone in the scene) and the mean of the prediction errors under the second one (robot facing a moving object). All the p-values of the t-tests we have run resulted to be smaller than 0.05 showing that, in all cases, the null hypothesis has been confirmed and that the two conditions produced a statistically significant difference in the averages of the prediction errors. As expected, in each

TABLE I: The t-test results for the robot behavior in the two conditions (robot alone vs. robot not alone in the scene).

Behaviour	p-value	Condition 1		Condition 2	
		MEAN	STD	MEAN	STD
1	$1.86e^{-5}$ (t=-4.31)	4.60	3.7	5.07	4.19
2	$6e^{-10}$ (t=-6.27)	2.91	3.48	4.10	3.93
3	$2.82e^{-2}$ (t=-2.20)	5.95	4.64	6.70	4.60
4	$1.12e^{-27}$ (t=-11.36)	0.0	0.0	1.88	3.17
5	$1.09e^{-50}$ (t=-16.20)	0.0	0.0	2.0	2.35
6	$1.40e^{-25}$ (t=-10.85)	0.0	0.0	1.28	2.25
7	$2.10e^{-29}$ (t=-11.76)	0.0	0.0	1.87	3.05

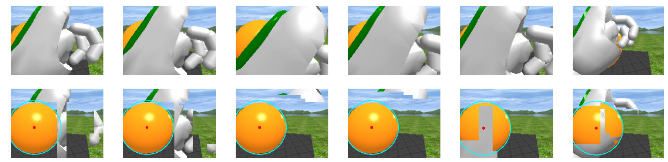

Fig. 3: Comparison between imaginary (bottom) and real frames (top) showing that, in the enhanced images, the red ball is still visible, although it is actually hidden by the Nao's hand.

of the seven different behaviours, the mean of the prediction errors resulted to be lower under the first condition (only the robot is moving in the scene), than under the second one (both the robot and an external object are moving in the scene). Figure 2 illustrates a sample trajectory of the robot arm. In particular, the second row shows the result of the sensory attenuation produced by subtracting the predictions of the forward model from the real sensory observation (first row). The third sequence shows how the attenuation looks like when there is an object moving in the background. As expected, the forward model is not able to predict these movements, and thus the system cannot attenuate them. This demonstrates that the predictive capabilities of forward models and the prediction errors they produce could be adopted for detecting movements produced by external agents. This is in line with the argumentations presented at the beginning of this paper related to sensory attenuation as a cue for the sense of agency.

B. *Object Permanence*

In a second experiment, we show how sensory predictions and sensory attenuation may lead as well to the development of a sense of object permanence. Inspired by the findings of [10], we created enhanced images where the movements of the robot are attenuated using the sensory predictions of a self-monitoring mechanism implemented through a forward model [9]. The experimental setup is shown in figure 3. The robot executes a random arm trajectory in front of an orange ball. Using the prediction of a forward model, the robot was able to attenuate its movements from the visual input, by replacing the pixels where the movement of the arm was expected with the corresponding pixels of a background image collected at the beginning of the experiment. We quantified the visibility of the ball, that is we measured how often the robot

TABLE II: The percentage of frames where the ball was visible in the imaginary compared to the real image.

Traject.	Ball Position 1		Ball Position 2		Ball Position 3	
	Real	Imaginary	Real	Imaginary	Real	Imaginary
1	71%	87%	74%	93%	83%	92%
2	87%	99%	85%	98%	96%	100%
3	79%	98%	77%	97%	82 %	96%
4	90%	94%	89%	99%	92%	99%
5	70%	88%	69%	89%	86%	75%

would perceive the ball from the enhanced visual input and from the real one. A video of a test sequence can be found at https://youtu.be/WKyZGcSCFak, which shows the actual visual input against the enhanced one. As evident from the video and from figure 3, the sight of the ball is occluded by the robot movements in the real visual input, but this is not always the case for the enhanced visual input. This is in line with what Kaufman et al. [10] and [15] suggest is happening in our brain during object occlusion. In Table II, the results for 5 different arm trajectories and 3 different locations of the ball are presented. For each trajectory and for each position of the ball, the percentage of frames (on average 150 frames per trajectory) where the ball was detected was always higher in the imaginary sequences than in the real ones.

IV. Conclusions

Processes of sensory prediction and sensory attenuation have been proposed by studies in psychology and neuroscience as having an important role in the development of the sense of agency and of the sense of object permanence. The experiments and the results presented in this work are in line with these findings. In particular, we investigated how self-monitoring mechanisms implemented through forward models can make a robot able to distinguish between self-generated movements and those generated by others, and to create mental representation of objects occluded by self-generated movements.

However, further research is needed in order to strengthen what has been argued in this work. Firstly, future experiments should address more complex sensory and motor spaces and more ecologic settings (e.g. more complex visual scenes, real robotic experiments, etc.). In the experiments presented here, in fact, only the movements of few joints of the robot's left arm have been taken into consideration, as well as only the visual modality. Although the scientific community recognises to the visual modality a prevalent role in self-recognition, several works assessed manifestations of self-recognition also in other modalities. Robots represent a perfect test bed for investigating these phenomena.

The study of the development of the sense of object permanence should not be limited either to self-occluding events of stationary objects. Further investigations on this topic will be carried out. Nonetheless, the experiment we presented has a particular value in robotics and in psychology. The *internal* images generated by the proposed computational model cannot simply be observed with any neuroscientific or psychology technique. The ability to directly visualize enhanced images under the conditions of the simulation - or, more generally, to generate and to observe enhanced sensory information - can have potential applications and advantages as compared to many *indirect* techniques in the neuroscience and psychology.

Acknowledgment

The research leading to these results has partially received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 609465 (EARS (Embodied Audition for RobotS) Project). The authors would like to thank the reviewers for their valuable help in improving the quality of this paper.

References

[1] E. van den Bos and M. Jeannerod, "Sense of body and sense of action both contribute to self-recognition," *Cognition*, vol. 85, no. 2, pp. 177–187, Septmeber 2002.

[2] S. Gallagher, "Philosophical conceptions of the self: implications for cognitive science," *Trends in Cognitive Neuroscience*, vol. 4, no. 1, pp. 14–21, January 2000.

[3] S.-J. Blakemore and C. Frith, "Self-awareness and action," *Current Opinion in Neurobiology*, vol. 13, pp. 219–224, 2003.

[4] G. Dawson and F. C. McKissick, "Self-recognition if autistic children," *Journal of Autism and Developmental Disorders*, vol. 14, no. 4, pp. 383–394, 1984.

[5] B. I. Bertenthal and K. W. Fischer, "Development of self-recognition in the infant," *Developmental Psychology*, vol. 14, no. 1, pp. 44–50, 1978.

[6] P. Rochat, "Five levels of self-awareness as thay unfold early in life," *Consciousness and Cognition*, vol. 12, pp. 717–731, 2003.

[7] G. Knoblich, "Self-recognition: body and action," *Trends in Cognitive Neuroscience*, vol. 6, no. 11, pp. 447–449, 2002.

[8] C. Weiss, A. Herwig, and S. Schütz-Bosbach, "The self in action effects: Selective attenuation of self-generated sounds," *Cognition*, vol. 121, no. 2, pp. 207–218, 2011.

[9] S. J. Blakemore, D. Wolpert, and C. Frith, "Why can't you tickle yourself?" *NeuroReport*, vol. 11, no. 11, pp. 11–16, 2000.

[10] J. Kaufman, G. Csibra, and M. H. Johnson, "Representing occludet objects in the human infant brain," *Proceedings of the Royal Society of London. B. Biological Sciences*, vol. 270, pp. 140–143, 2003.

[11] R. Baillargeon, "Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants," *Cognition*, vol. 23, pp. 21–41, 1986.

[12] ——, "Object permanence in 3 1/2- and 4 1/2- month-old infants," *Developmental Psychology*, vol. 23, no. 5, pp. 655–664, 1987.

[13] R. Baillargeon, E. S. Spelke, and S. Wasserman, "Object permanence in five-month-old infatns," *Cognition*, vol. 20, pp. 191–208, 1985.

[14] J. Piaget, *The Construction of Reality in the Child*. Basic. SAS User's Guide: Statistics, 1954.

[15] J. Kaufman, G. Csibra, and M. H. Johnson, "Oscillatory activity in the infant brain reflects object maintenance," *Proceedings of the National Academy of Sciences*, vol. 102, no. 42, pp. 15 271–15 274, 2005.

[16] R. Saegusa, G. Metta, and G. Sandini, "Own body perception based on visuomotor correlation," *Intelligent Robots and Systems (IROS)*, pp. 1044 – 1051, 2010.

[17] P. Michel, K. Gold, and B. Scassellati, "Motion-based robotic self-recognition," *Intelligent Robots and Systems*, pp. 2763–2768, 2004.

[18] A. Pitti, H. Mori, S. Kouzuma, and Y. Kuniyoshi, "Contingency perception and agency measure in visuo-motor spiking neural networks," *Transactions on Autonomous Mental Development*, vol. 1, no. 1, pp. 86–97, 2009.

[19] J. Weng, Y. Zhang, and Y. Chen, "Developing early sense about the world: "object permanence" and visuoauditory real-time learning," *Neural Networks*, vol. 4, pp. 2710 – 2715, 2003.

[20] Y. Chen and J. Weng, "A case study of developmental robotics in understanding "object permanence"," 2004.

[21] D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspective and problems in motor learning," *Trends in Cognitive Sciences*, vol. 5, no. 11, pp. 487–494, 2001.