Goal-Directed Learning of Hand-Eye Coordination in a Humanoid Robot

Mathias Schmerling[1], Guido Schillaci[2] and Verena V. Hafner[2]

***Abstract*— Visuo-motor coordination is known to be highly important for the development of a broader range of cognitive and motor skills in human infants and can thus be considered one of the key skills for robots to master.**

In this paper, we investigate how a recent concept in developmental robotics, referred to as *goal babbling*, relates to a visuo-motor coordination task in the humanoid robot Aldebaran Nao that requires coordinated control of two sub-systems of motors, namely head and arm motors. The idea of goal babbling builds on findings in developmental psychology showing that human infants attempt goal-directed movements early on in their development enabling them to rapidly and efficiently bootstrap their motor system. Goal babbling has been shown to be superior to the classical idea of random motor babbling for the learning of body kinematics in robotic systems, in particular for systems with many degrees of freedom.

Our results not only support the utility of goal babbling for the acquisition of visuo-motor coordination skills but also suggest that goal babbling is particularly effective in the case where two separate motor sub-systems, head and arm, need to be coordinated.

I. INTRODUCTION

Control of movements and coordination of different parts of the body are a fundamental prerequisite for the development of complex motor and cognitive skills in humans, as reported in several studies in developmental psychology. Rehearsal of motor control through exploration behaviours has been observed already during pre-natal stages of development in foetuses [1]. Early coupling between visual and motor systems in infants has been reported and investigated also by Tükel [2]. Yu et al. suggested that there is a correlation between the development of hand-eye coordination, learning capabilities and social skills in humans [3].

Nonetheless, sensory-motor coordination, the ability to achieve desired sensory outcomes by means of an agent's own actions, is a key ability also for modern robots. In control theory, a prototypical instance of sensory-motor coordination is the ability of a robot to control its end-effector in a 2- or 3-dimensional task space by actions in an m-dimensional motor space. To effectively do so, the robot firstly needs to be able to predict the sensory outcome x of a motor action q, known as the forward model $f : Q \rightarrow X$ where X and Q refer to the task space and motor space, respectively. Secondly, the robot needs to be able to infer a motor command $\hat{q}$ to achieve a desired sensory outcome x^*, known as the inverse model $g : X \rightarrow Q$. Developmental robotics asserts that, for real-world, high-dimensional, redundant and possibly non-stationary robots, motor coordination skills need to be acquired through autonomous exploration, i.e. by generating samples (q, x) of the unknown forward and inverse function through an incremental learning process. This holds in particular for soft robotics with highly non-linear control. One class of exploration strategies, referred to as *motor babbling*, focuses on exploring the motor space Q: Motor actions are chosen and successively executed and their sensory outcome is observed.

In recent publications [4][5], we showed how a humanoid robot can acquire hand-eye coordination and reaching skills by exploring its movement capabilities through random body babbling and by using a biologically inspired internal model of the robot body. The random babbling approach, however, is limited by the fact that high-dimensional motor spaces cannot be exhaustively explored and, additionally, exploration of these spaces is often redundant as many different actions map to the same sensory outcome. Therefore, a recent concept, referred to as *goal babbling*, proposes to explore the sensory space X directly by goal-directed actions: The agent chooses goals x^* from the sensory space and infers actions $\hat{q}$ to reach them based on the knowledge acquired up to that point, updating the inverse and forward model on the fly. The idea of goal babbling is inspired by findings from developmental psychology showing that human infants attempt goal-directed movements early on in their development [6].

Many publications have since demonstrated the utility of goal babbling. To mention a few, Rolf & Steil [7] demonstrated how goal-directed exploration of the task space enables successful learning of the inverse function on a challenging robot platform, the Bionic Handling Assistant. Baranes & Oudeyer [8] did an extensive study comparing different variations of goal and motor babbling strategies on several simulated and robot platforms. Notably, they also introduced a curiosity-driven learning scheme [9], [10], [8] in which agents hold an interest measure for different subregions of the explored space. Curiosity-driven goal babbling was shown to be superior to other learning schemes in most scenarios.

It should be noted that in these publications on goal babbling as well as in the present study, the notion of goals always refers to a specific typology of goals, namely to the selection of coordinates in a defined, often cartesian, workspace by a single expert system. We emphasize that the

[1]Mathias Schmerling is with the Bernstein Center for Computational Neuroscience, Berlin, Germany mathias.schmerling@bccn-berlin.de

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

formulation of goals as adopted here is not meant to capture the entirety of the concept. Indeed, it is easy to identify a wide range of topics not accounted for by this low-level definition of goals, such as for example goals relating to the interaction with external objects (e.g. [11]) and with other agents and, furthermore, the existence of goals on different behavioral scales for the planning of complex actions and action sequences. However, we argue that the notion of goals in motor control problems is nonetheless informative and can potentially be extended to other contexts as well.

In this work, we implemented and compared different exploration methods on the humanoid robot Aldebaran Nao for the development of reaching skills using visual feedback from one of its cameras. Experiments took place in the Cyberbotics Webots simulator [12]. The difference to the previously mentioned studies on goal babbling is the use of an on-board instead of an external sensory system. Previous studies used feedback from externally mounted cameras or position information of a simulator to generate samples for learning. Arguably, externally installed cameras are not as compact and as mobile as the on-board visual systems that today's humanoid robots are equipped with. Using Nao's built-in camera for sensory feedback, however, poses an additional challenge for learning. The visual range of Nao is comparatively narrow for a fixed camera position. Therefore, covering a substantial volume of end-effector positions requires coordinated control not only of Nao's hand but also of its head. In turn, the learning problem in this case breaks down into two concurrent sub-problems: *1) learning of the inverse kinematics of the arm to control the position of the end-effector* and *2) learning of the inverse kinematics of the head to be able to generate sensory feedback for 1)*.

We propose a two-level model architecture to tackle this learning problem, introducing two separate inverse models: one for the arm and one for the head kinematics. We investigate the roles of motor babbling and curiosity-driven goal babbling in this learning setup. To do so, a recently published toolbox for robot exploration called Explauto [13] was employed as part of the proposed architecture, demonstrating its utility and its compatibility with the humanoid robot Nao. Our results indicate an important role of goal babbling for visuo-motor coordination tasks. Furthermore, we illustrate how the two levels of the model relate to each other and how a synergic relationship emerges between them when the Nao robot explores in a curiosity-driven manner.

II. Methods

A. Robotic Setup

Nao possesses 5 joints to control each one of its arms and two HD cameras positioned on its chin and forehead. For the following experiments, only the bottom camera, its 2 neck joints (Head-Pitch and Head-Yaw), and 4 of the 5 joints in its left arm were used (Shoulder-Roll, Shoulder-Pitch, Elbow-Roll and Elbow-Yaw). Low-level built-in functions (NAOqi v1.14.5 [14]) were used to control the joints in absolute angles.

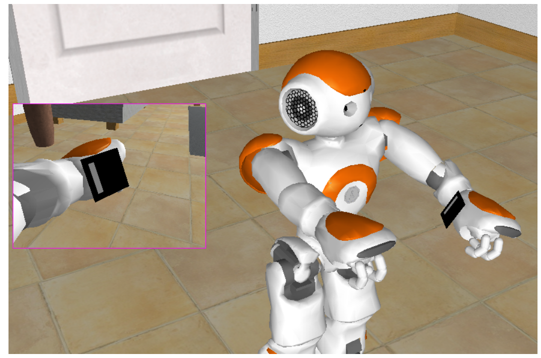


Fig. 1. Illustration of the robotic setup. A marker is placed on the hand of the robot providing visual feedback of its end-effector position.

Sensory information was provided to Nao by placing a marker tag of size $4cm \times 4cm$ on the visible side of its hand. The image and the spatial position of the marker relative to the camera, serving as a proxy to the robot's hand position, was calculated using the computer vision library ArUco [15]. To obtain absolute spatial coordinates from the original relative coordinates, the data was transformed from the camera frame to the torso frame via built-in functions. In each iteration, head and arm movements were executed as timed angle interpolations with gaussian velocity profile. Each interpolation to a new motor position was fixed to take $5s$, taking 20 samples of image and joint angle data at $4Hz$ during execution. Learning models were updated batch-wise on the 20 collected samples at the end of each movement. All experiments took place in the Cyberbotics Webots [12] simulator v7.4.3 and ran until $10,000$ movements were executed ($200,000$ samples collected).

B. Model architecture

First and foremost Nao's task is to learn through exploration the forward and inverse relationship between the joints of its arm $q_A \in Q_A$ and the position of its end-effector $x_S \in X_S$. To this end, the model comprises forward and inverse functions $g_1 : X_S \rightarrow Q_A$ and $f : Q_A \rightarrow X_S$ together representing an internal model for the arm kinematics. It is important to point out that we assume that the sensory space X_S explored by the robot's arm is the body-centered cartesian space as derived from identifying the position (including depth) of the marker in the camera image. The analysis of vision data was thus decoupled from the motor learning and the existence of an extrinsic coordinate system was taken as a given. While there is evidence that extrinsic coordinate systems exist in the brain [16] (although critically [17]), the question of how they could emerge during learning was not addressed in this work.

The task of learning the functions g_1 and f is equivalent to the ones already discussed in the literature [8], [18], [7]. However, the learning task contains an additional challenge as sensory feedback on the end-effector position is only available to Nao if it can infer an appropriate head motor command $q_H \in Q_H$ to center gaze on its hand. The proposed model therefore introduces an additional mapping function

g_2. This function maps image positions $x_I \in X_I$ in retinal coordinates to head motor commands $q_H \in Q_H$. Additionally, the model needs to be provided with some information on what to focus on, which in our case is the dynamically changing arm position. In this work, we decided to augment the model with efferent copies of arm movements: $g_2 : (Q_A, X_I) \to Q_H$. Formalized in this way, g_2 infers, provided with an efferent copy of the next arm movement, a motor command for the head that will result in the marker ending up at a desired image position. The validity of this approach is supported by studies on gaze control in humans showing that subjects incorporate proprioceptive and especially efferent signals when anchoring gaze to pointing movements [19] and are able to do so even in the absence of visual feedback [20]. An obvious disadvantage of this approach, however, is the fact that the learned model does not readily generalize to gaze control tasks not involving the arm, such as tracking objects. Practically, this can still be done by translating a given object location to a reaching movement via g_1. This, on the other hand, would mean to say that any gaze control is implicitly a reaching movement, a stance we do not want to take up in this work. We chose our formalization for its simplicity and practicality but we also admit to the fact that other formalizations would have been possible, such as invoking the forward model f to translate the efferent copy to a 3-dimensional location and training a gaze controller on those locations. Summing up, the overall model thus comprises a total of three mapping functions to be learned:

$$g_1 : X_S \to Q_A \tag{1}$$

$$f : Q_A \to X_S \tag{2}$$

$$g_2 : (Q_A, X_I) \to Q_H \tag{3}$$

These mapping functions need to be learned by Nao through exploration. During the experiment, Nao will execute head and arm movements and observe the sensory consequences of these actions, thus generating observations $(q_A^t, q_H^t, x_I^t, x_S^t)$. These observations are then fed to a regression model to obtain estimates of increasing precision for the unknown functions f, g_1, g_2. Thereby the regression models for g_1 and f are updated on (q_A^t, x_S^t) while g_2 is updated on $((q_A^t, x_I^t), q_H^t)$. Note that g_2 is different from g_1 in the sense that there is no causal relation between arm configuration and head configuration. That is, while g_2 infers actions q_A that *achieve* outcomes x_S, g_2 infers actions q_H that achieve outcomes x_I *given* q_A. Nonetheless, g_2 describes a valid, learnable and, also different from g_1, non-redundant mapping: for a given arm configuration q_A, there is at most one head configuration q_H that will achieve x_I. Thereby the arm configuration is always provided by the arm controller and represents an efferent copy of the next movement to be executed, thus binding head movements to arm movements.

In principle, the proposed model architecture is independent of the specific choice of the regression model. For simplicity and because the primary goal of the present work was to investigate the role of goal babbling and motor babbling in the proposed learning task, a variation of the k-nearest-neighbor algorithm, namely a weighted-nearest-neighbor scheme with adaptive choice of k, was employed. The regression model was provided by the Explauto framework [13].

C. Goal Babbling and Motor Babbling

The terms goal babbling and motor babbling refer to the type of exploration employed by the agent to generate observations (q^t, x^t). In motor babbling schemes, exploration focuses on the action space: Actions q are chosen from the space of all available actions Q and then executed to observe their sensory outcome x. As opposed to this, goal babbling, a more recent concept, refers to the idea to explore the sensory space directly by goal-directed actions. The agent chooses goals x^* from the sensory space X and tries to reach them with an action it infers from its current inverse estimate $\hat{q} = g(x^*)$. Exploratory noise drives the execution of previously unexplored actions in the vicinity of $\hat{q}$. Note that while the term *action* can in principle denote both low-level motor commands as well as motor primitives (as e.g. in some of the experiments in [8]) or even higher-level actions, throughout this paper it is exclusively used to describe joint angle commands.

This general setup can be transferred to the presented variation of the learning task with a few modifications. The idea is to understand exploration of head and arm movements as two concurrent learning tasks which can both be explored either via goal babbling or via motor babbling. In the simplest case, both the head and arm kinematics are explored via motor babbling. This means that in each iteration Nao chooses actions for its head as well as for its arm and then executes them, observing the sensory outcome, i.e. image and spatial coordinates of the marker, if visible. In this case, Nao is in pure exploration, learning passively and not exploiting g_1 or g_2. In the second condition, head movements are instead chosen in a goal-directed manner. In this case, Nao chooses an arm movement q_A in each iteration and then tries to infer a head movement appropriate for that arm movement, invoking g_2 with the desired image position x_I^* acting as a goal: $\hat{q}_H = g_2(q_A, x_I^*)$. We refer to this condition where g_2 is exploited from the start as the goal/motor babbling hybrid condition. Finally, exploration can take place in a goal-directed manner on both levels. In this case an end-effector position x_S^* is chosen as a goal and an arm movement $\hat{q}_A = g_1(x_S^*)$ is inferred to reach it. Exploratory noise η is added to $\hat{q}$. Afterwards, a head movement $\hat{q}_H = g_2(\hat{q}_A, x_I^*)$ is inferred to achieve a desired retinal position of the marker, again adding exploratory noise. Exploratory noise on joint angles was Gaussian distributed with zero mean and $\sigma = 0.1rad$. The chosen head and arm movements in each iteration were executed simultaneously as timed angle interpolations. All these approaches, i.e. the goal babbling, the motor babbling and the hybrid approach are again outlined as pseudocode (Algorithms 1, 2, 3).

An important part of the goal babbling architecture is the way goals are chosen from the sensory space. In the present scenario, this concerns both the goals x_I^* and x_S^*. For simplicity, the goals in the image space were always chosen to be the image center, i.e. $x_I^* = x_I^{center}$. Note that the other part of the input to g_2, an efferent copy of the arm configuration, is always provided by the arm kinematic learner, never chosen independently. If the gaze kinematic learner would operate independently, i.e. if the input to g_2 was different from the actual arm movement to be executed, then Nao would look for its hand based on meaningless input and thus fail to find the marker even if g_2 was learned perfectly.

The goals x_S^* were chosen from the space described as $X_S = [0.0m; 0.25m] \times [0.0m; 0.3m] \times [-0.1m; 0.3m]$ which was known a priori to be bigger than the reachable space. Instead of choosing goals randomly, they were chosen based on a curiosity-driven learning schedule that was proposed in and is provided by the Explauto framework [13]. Briefly outlined, in this learning schedule, goals are chosen such that empirical learning progress is maximized by partitoning the sensory space and by sampling preferentially from subregions that exhibit a decrease in learning error. To facilitate the calculation of this curiosity measure and to speed up the learning process in general, the number of sensorimotor experiments was increased by retrieving sensory and motor information at a constant sampling rate during the interpolation to a new motor position. Execution of each motor command was fixed to take $5s$ and 20 observations were retrieved at $4Hz$ during execution. In the case of goal babbling, each of the 19 in-between observations was considered as a goal reached with maximal precision.

Algorithm 1 Goal Babbling

$f, g_1, g_2, t \leftarrow 0$
loop
    choose x_S^*, x_I^*
    $\hat{q}_A \leftarrow g_1(x_s^*) + \eta$
    $\hat{q}_H \leftarrow g_2(\hat{q}_A, x_I^*) + \eta$
    execute $\hat{q}_A, \hat{q}_H$
    observe *(including samples)* x_S, x_I
    if x_S, x_I *is valid* **then**
        update models f, g_1, g_2
    end if
    $t \leftarrow t + 1$
end loop

Algorithm 2 Motor Babbling

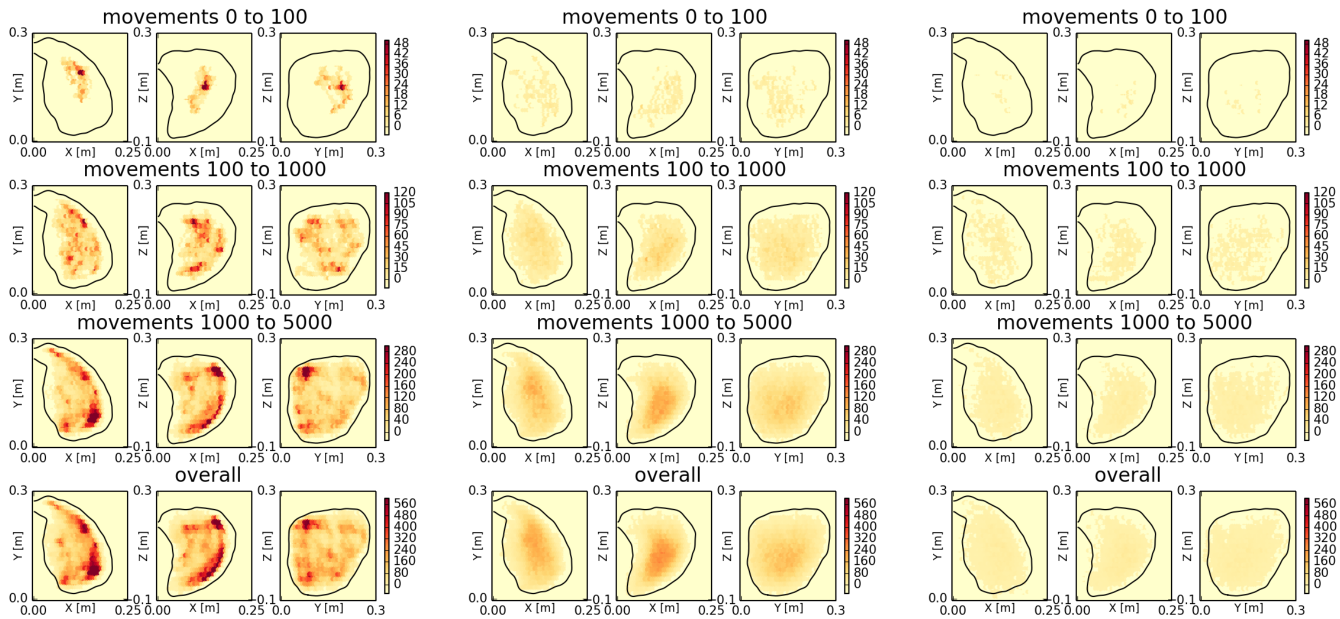
$f, g_1, g_2, t \leftarrow 0$
loop
    choose q_A, q_H
    execute q_A, q_H
    observe *(including samples)* x_S, x_I
    if x_S, x_I *is valid* **then**
        update models f, g_1, g_2
    end if
    $t \leftarrow t + 1$
end loop

Algorithm 3 Goal/Motor Babbling

$f, g_1, g_2, t \leftarrow 0$
loop
    choose q_A
    choose $x_I^* = x_I^{center}$
    $\hat{q}_H \leftarrow g_2(q_A, x_I^*) + \eta$
    execute $q_A, \hat{q}_H$
    observe *(including samples)* x_S, x_I
    if x_S, x_I *is valid* **then**
        update models f, g_1, g_2
    end if
    $t \leftarrow t + 1$
end loop

D. Evaluation

All experiments were replicated 5 times for each of the three conditions. Performance of the model was evaluated under several aspects. Firstly, the development of Nao's visuo-motor coordination skills was assessed in terms of how often the marker was detected in the camera image over the course of the experiments. Secondly, it was evaluated how well Nao covers the volume of reachable space with observations. Lastly, the development of its reaching skills was assessed over time. Unbiased estimation of how well Nao can move its end-effector to desired positions in the task space requires a set of testcases uniformly sampled from the reachable space. However, in the given task, the space of reachable positions is, on the one hand, constrained by the properties of the arm and, on the other hand, depends on coordinated gaze control. It is thus difficult to model a priori. Instead, we obtained an approximation from data collected in independent, long-running experiments of all babbling types. This strongly non-uniformly distributed data set was then resampled to an approximate uniform distribution. In detail, a uniform grid of target positions spanning a volume larger than the reachable space (regular spacing $5mm$) was constructed and each grid point was tested for reachability, i.e. it was retained as long as there was at least one observation from the indepedent data set not further away than $2mm$. As a result, 33539 grid points were retained and are displayed in Fig. 2, illustrating the limits of reachability. Black contour lines represent a qualitative estimation of the perimeter. Of note, these contour lines do not represent a systematic estimation of the reachable space but were calculated only for the purpose of visualization. Each time the reaching performance was evaluated, Nao was asked to reach 200 testcases randomly chosen from the 33539 grid points by its current inverse estimate, and the Euclidean distance between desired and actual marker position was considered as reaching error.



(a) goal babbling (b) hybrid (c) motor babbling

Fig. 4. Distribution of observations in a prototypical experiment for the goal babbling, motor babbling, and hybrid approach. Rows represent different time periods of the experiment. Columns represent the 2D-projections of the 3D task space on the respective axes.

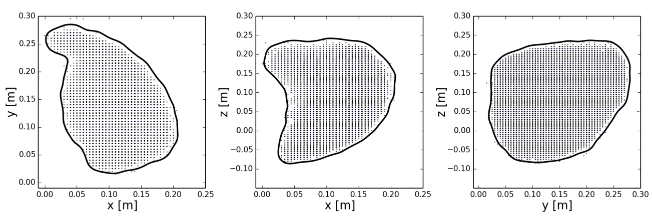


Fig. 2. Illustration of the reachable space. Plots correspond to the 2D projections of the 3D sensory space. The x-axis is positive towards Nao's front, the y-axis is positive towards Nao's left side and the z-axis is vertical. Points represent reachable hand positions. Black contour lines illustrate the reachable space qualitatively.

III. Results

A. Marker Detection

One of Nao's learning tasks is to learn its head kinematics to generate the sensory feedback necessary for the learning of its arm kinematics. The results (Fig. 3) show strong qualitative and quantitative differences between the three previously outlined settings. A prototypical replication is shown. As expected, the pure motor babbling setting is clearly inferior to the two other configurations in which there is an active mechanism to center gaze on the marker. This is simply due to the fact that the number of possible arm configurations is large while the visual range of the bottom camera is comparatively narrow. For random head movements the robot's hand will only occasionally cross the visual field and thus the number of valid samples, i.e. samples where the marker is inside the camera image, on average only amounts to 3.1% of the total number (200,000) of samples. Of note, this does not by itself imply an inferior performance of the learned function g_2 but is merely due to the fact that g_2 is not exploited in this purely explorative condition. However, a critical point in the present setup is the generation of observations necessary to learn the arm kinematics. It will be shown later that learning of the arm model, relying only on the few samples generated by purely explorative head movements, is inferior to the other two approaches. The present findings therefore highlight the need for exploitative head movements.

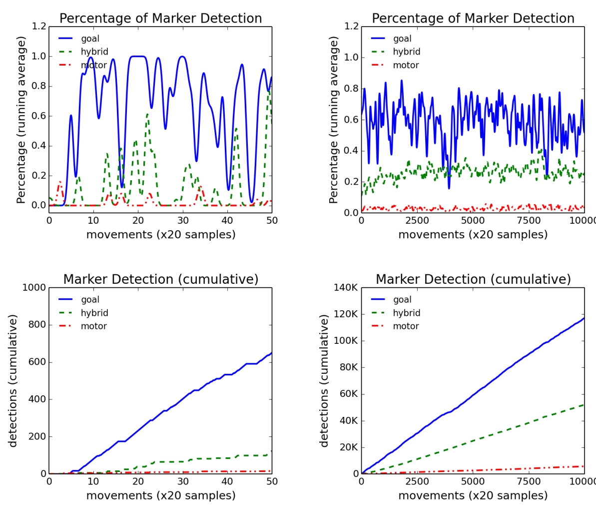


Fig. 3. Plots show the frequency (upper plots) and cumulative number (lower plots) of marker detections for the **goal babbling**, **motor babbling** and **hybrid** approach. A prototypical replication of the experiment is shown. Left plots show the initial phase of the experiment, right plots the whole experiment. **Goal babbling** outperforms the other two approaches.

The other two settings with goal-directed head movements perform better with 24.2% and 60.3% marker detection on average for goal/motor and pure goal babbling, respectively. The difference between these two approaches is striking as both of them use the same mechanism for gaze control. The explanation lies in the fact that the two learning tasks, i.e. arm and gaze control, are not independent but instead mutually depend on each other since they are bound together by g_2 and since both head and arm movements determine which observations are made during learning. That is, if Nao chooses arm movements in a goal-directed fashion, this will also benefit the gaze control learning for the following reason: Goal-directed exploration will focus on arm positions where the learning progress is maximal. Learning progress can only be made where there is sensory feedback and thus where the head inverse model g_2 has already acquired some knowledge. This will automatically constrain Nao's arm movements to regions of the observation space that have been uncovered also by the head model. Consequently, Nao makes efficient use of the knowledge *both* inverse models acquire over time. Or, in other words, goal-directed arm movements allow the learning of gaze control to keep up with the learning of arm control.

This, on the other hand, is not the case for the hybrid approach. In the hybrid condition, arm movements are unconstrained and thus extremely challenging for the head model. The detection frequency is lower for two reasons. The first reason is that the arm moving to random positions will frequently end up in regions of the observation space still unexplored by the head model. Secondly, many arm positions are simply not mappable to an appropriate head configuration at all since the workspace of the hand is larger than the visual workspace. This concerns hand positions that are too far up or too far down, causing Nao to run into the limits of its neck pitch joint, and also positions that are concealed by Nao's own body, i.e. by its shoulder and its chest. While curiosity-driven arm movements will refrain from these positions lacking visual feedback, random arm movements will not.

Fig. 3 illustrates the discussed effects quantitatively for one prototypical replication. Initially, all the different approaches start out with random head movements - the inverse models did not acquire enough knowledge to make any predictions yet. Consequently, there are only occasional detections. Once the number of detections surpasses a certain threshold that is determined by the specific regression model (in the present case 13), the frequency of successful detection increases for the two settings in which there is an active mechanism for head movements. The increase is particularly striking in the pure goal babbling setup. This is due to the fact that, as mentioned earlier, goal-directed arm movements will focus on subregions where the marker has been detected before. This subregion is initially very narrow and thus detecting the marker is very simple. Exploratory noise combined with the curiosity-driven learning then continually drives both head and arm movements to leave already explored regions of the observation space and to uncover new ones. This is also what explains the high variability of detection frequency during the goal babbling experiment. In fact, Nao alternates between episodes of almost perfect detection and episodes of very low detection frequency throughout the whole experiment, owing to the heterogeneity of the observation space in terms of learning difficulty. Episodes of low detection frequency were found to occur whenever Nao focuses on certain challenging regions of the observation space such as hand positions far out to the left that are easily obscured by its shoulder as well as hand positions close to its body. It is important to note that while the marker is often missed in these cases due to partial concealment, this does not necessarily mean that the gaze learner makes altogether wrong inferences. However this is not captured in the present setup as the visual sensors can only detect the marker when it is fully visible. This is one of the drawbacks of the simple visual sensors employed in the experiment. It is also the reason why the detection frequency does not increase beyond the initial incline for the goal babbling experiment and stays flat at around 60%. Since observations where the marker is not fully visible lack sensory feedback entirely and thus are simply ignored, the learner does not register the extent to which it is failing in challenging subregions of the observation space, thus staying focused on those regions overly long. This behavior continues to occur even towards the end of the experiment.

B. Distribution of Observations

Fig. 4 shows the distribution of observations for the goal babbling, hybrid and motor babbling approach over the course of the experiment in one prototypical replication. Both the hybrid and motor babbling approach are characterized by a distribution of samples that has been shown to be typical for motor babbling [8]. Samples largely concentrate on the center of the reachable space and thin out at its edges because of the way random motor configurations non-uniformly map to the sensory space. As opposed to this, goal babbling more strongly explores the limits of the reachable space but also explores central positions. Note again that the goal babbling approach does not only differ in the distribution of samples but also in their absolute number. The results suggest that goal babbling more efficiently explores the sensory space, most notably positions at the limits of the reachable space.

C. Learning Curve

Fig. 5 shows the learning curves for the goal babbling, motor babbling and hybrid approach. The results suggest that an inverse model for the arm kinematics can be learned following the goal babbling and hybrid approach with a mean error converging to $0.8(\pm 0.05)cm$ on average for goal babbling and $1.0(\pm 0.06)cm$ on average for the hybrid condition. Performance was assessed for the first time at $t = 100$ and regularly from then onwards on 200 randomly chosen representative sensory goals (see section Methods). Pure motor babbling is inferior for obvious reasons, as only very few valid observations can be accumulated over the experiment. Despite the fact that also the hybrid and goal babbling approach differ considerably in the amount of valid observations, their performance is similar at later stages

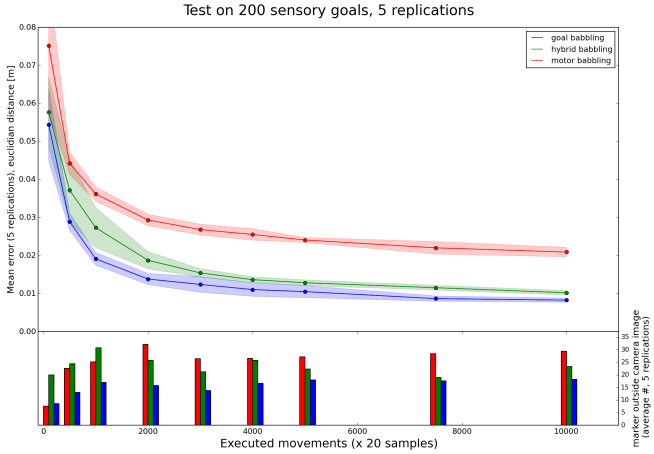


Fig. 5. Learning curves for the goal babbling, motor babbling and hybrid approach. The upper plot shows the learning error averaged over 5 replications over time on 200 randomly chosen representative sensory goals. Shaded areas indicate the error level corresponding to one standard deviation. The lower plot indicates the number of attempts where the marker was not in the camera image.

of the experiment. However, a slightly faster decrease in performance error for the goal babbling approach can be observed at earlier stages of the experiment. Performance could not be assessed for some goals at each evaluation because of the head inverse model's failure to suggest an appropriate head movement. The number of these masked goals is shown in the lower plot of Fig. 5 and suggests that only a small portion of goals is visually missed at each evaluation time for all three settings.

IV. Discussion

In this paper, we showed that two recent concepts inspired by findings of developmental psychology, goal babbling [7], [18], [8] and maximization of learning progress [9], [10], [8], support efficient learning of inverse kinematics in a humanoid robot that uses visual sensory feedback from its on-board camera. We argued that using visual feedback poses an additional challenge for learning because it requires coordinated control of an additional set of motors, namely the head joints. We therefore proposed to introduce an additional inverse kinematic model, mapping arm configuration and image space to head configuration and showed that learning of both models concurrently from the start via an incremental, goal-directed exploration is indeed possible. Furthermore, it is important to highlight the strong interdependence of both models and their respective exploration schemes. The results show that learning of the head inverse model benefits from goal-directed arm movements. This suggests that the benefits of using curiosity-driven goal babbling can extend from one level of the model architecture to another. We therefore venture that curiosity-driven learning schemes are a key element for multimodality and scalability in robot systems of increasing complexity.

To test this hypothesis in future work, it would be interesting to compare the three conditions investigated in this paper with other settings. One possible setting would be a hybrid condition in which arm configurations are chosen randomly but restricted to the visual workspace of the head. This would make the hybrid and goal babbling condition more comparable. However, we did not follow this line of investigation for two reasons. Firstly, the exact shape of the constrained motor space is quite complex for reasons mentioned earlier (i.e. Nao concealing its hand with its own body) and modeling it a priori is thus challenging and questionable in regard to biological plausibility. Secondly, we were interested to see if Nao was able to uncover the limits of its workspace, both for reaching and gaze control, autonomously. In fact, it is one of the known virtues of curiosity-driven learning to be able to effectively uncover the limits of reachability [10]. Another relevant setting would be a goal babbling approach with random instead of curiosity-driven choice of goals so as to better disentangle the effects of goal babbling and curiosity-driven learning.

Furthermore, several aspects of the proposed experiments are still to be refined in future work. Firstly, we trained the internal models entirely on positive feedback, considering only trials where the marker was visible and discarding the rest. This, however, led Nao to become over-interested in challenging subregions of the observation space. Given a learning mechanism that includes also the binary information on the visibility of the marker, the detection frequency might increase for the goal babbling setup. The second major aspect concerns the way we modeled the sensory space presented to the robot. We introduced a marker tag as a proxy to the robot end-effector and used computer vision techniques to determine its position. Furthermore, the sensory space from which to sample goals was handcrafted. This suggests that the robot holds a preconceived notion of its hand position and of the task space in which to function. Consequently, there is no need for self-detection which greatly simplifies the experiments but is less biologically plausible and, arguably, less scalable. However, the question of how an agent can learn to perform self-detection and to abstract the task space from a richer sensory context is an open and challenging question. A recent publication [21] has addressed this issue in a simple simulated experiment where the agent had to learn to abstract goals from a richer sensory context while simultaneously performing goal-directed exploration. We think that this is an interesting line of investigation because it highlights the important question of how goals can appear in artificial agents without the designer specifying them beforehand.

Acknowledgment

The research leading to these results has been partially funded from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 609465, related to the EARS (Embodied Audition for RobotS) project.

References

[1] S. Zoia, L. Blason, G. D'Ottavio, M. Bulgheroni, E. Pezzetta, A. Scabar, and U. Castiello, "Evidence of early development of action planning in the human foetus: a kinematic study," *Experimental Brain*

Research, vol. 176, no. 2, pp. 217–226, 2007. [Online]. Available: http://dx.doi.org/10.1007/s00221-006-0607-3

[2] S. Tükel, "Development of visual-motor coordination in children with neurological dysfunctions," 2013.

[3] C. Yu and L. B. Smith, "Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination," *PLoS ONE*, vol. 8, no. 11, p. e79659, 11 2013.

[4] I. Kajic, G. Schillaci, S. Bodiroza, and V. V. Hafner, "A biologically inspired model for coding sensorimotor experience leading to the development of pointing behaviour in a humanoid robot," in *Proceedings of the Workshop "HRI: a bridge between Robotics and Neuroscience". 9th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2014)*, 2014.

[5] G. Schillaci, V. V. Hafner, and B. Lara, "Online Learning of Visuo-Motor Coordination in a Humanoid Robot. A Biologically Inspired Model." in *IEEE Int. Joint Conf. Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2014, pp. 145–151.

[6] C. Von Hofsten, "Eye–hand coordination in the newborn." *Developmental psychology*, vol. 18, no. 3, p. 450, 1982.

[7] M. Rolf and J. J. Steil, "Efficient exploratory learning of inverse kinematics on a bionic elephant trunk," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 6, pp. 1147–1160, 2014.

[8] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, 2013.

[9] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[10] C. Moulin-Frier and P.-Y. Oudeyer, "Exploration strategies in developmental robotics: a unified probabilistic framework," in *IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, 2013, pp. 1–6.

[11] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Autonomous selection of the what and the how of learning: an intrinsically motivated system tested with a two armed robot," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*. IEEE, 2014, pp. 434–439.

[12] Webots, "http://www.cyberbotics.com," commercial Mobile Robot Simulation Software. [Online]. Available: http://www.cyberbotics.com

[13] C. Moulin-Frier, P. Rouanet, and P.-Y. Oudeyer, "Explauto: an open-source python library for active and online sensorimotor learning," in *IEEE Int. Joint Conf. Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2014.

[14] Aldebaran Robotics. [Online]. Available: http://doc.aldebaran.com/1-14/index.html

[15] ArUco, "http://www.uco.es/investiga/grupos/ava/node/26," arUco: a minimal library for Augmented Reality applications based on OpenCv.

[16] S. Kakei, D. S. Hoffman, and P. L. Strick, "Muscle and movement representations in the primary motor cortex," *Science*, vol. 285, no. 5436, pp. 2136–2139, 1999.

[17] W. Wu and N. Hatsopoulos, "Evidence against a single coordinate system representation in the motor cortex," *Experimental brain research*, vol. 175, no. 2, pp. 197–210, 2006.

[18] M. Rolf, J. J. Steil, and M. Gienger, "Goal babbling permits direct learning of inverse kinematics," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 216–229, 2010.

[19] J.-L. Vercher, G. M. Gauthier, O. Guedon, J. Blouin, J. Cole, and Y. Lamarre, "Self-moved target eye tracking in control and deafferented subjects: roles of arm motor command and proprioception in arm-eye coordination," *Journal of Neurophysiology*, vol. 76, no. 2, pp. 1133–1144, 1996.

[20] S. F. Neggers and H. Bekkering, "Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal," *Journal of Neurophysiology*, vol. 86, no. 2, pp. 961–970, 2001.

[21] M. Rolf and M. Asada, "Autonomous development of goals: From generic rewards to goal and self detection," in *IEEE Int. Joint Conf. Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2014.