Exploration strategies in developmental robotics: a unified probabilistic framework

Clément Moulin-Frier, Pierre-yves Oudeyer

Exploration strategies in developmental robotics: a unified probabilistic framework

Clément Moulin-Frier
Flowers team, Inria / ENSTA-Paristech, France
Email:clement.moulin-frier@inria.fr

Pierre-Yves Oudeyer
Flowers team, Inria / ENSTA-Paristech, France
Email:pierre-yves.oudeyer@inria.fr

***Abstract*—We present a probabilistic framework unifying two important families of exploration mechanisms recently shown to be efficient to learn complex non-linear redundant sensorimotor mappings. These two explorations mechanisms are: 1) goal babbling, 2) active learning driven by the maximization of empirically measured learning progress. We show how this generic framework allows to model several recent algorithmic architectures for exploration. Then, we propose a particular implementation using Gaussian Mixture Models, which at the same time provides an original empirical measure of the competence progress. Finally, we perform computer simulations on two simulated setups: the control of the end effector of a 7-DoF arm and the control of the formants produced by an articulatory synthesizer.**

I. Introduction

The learning of sensorimotor tasks, for example reaching objects with the hand or controlling the shape of a vocal tract to produce particular sounds, involves the learning of complex sensorimotor mappings. This latter generally requires to build a model of the relationships between parts of the sensorimotor space. For example, one might need to predict the positions of the hand knowing the joint configurations, or to control the shape the vocal tract to produce the sound of particular words.

Let us introduce the problem more formally. A learning agent interacts with a surrounding environment through motor commands M and sensory perceptions S. We call $f : M \to S$ the unknown function defining the physical properties of the environment, such that when the agent produces a motor command $m \in M$, it then perceives $s \in S$. Classical robotic problems are e.g. the prediction of the sensory effect of an intended motor command through a forward model $\tilde{f} : M \to S$, or the control of the motor system to reach sensory goals through an inverse model $\tilde{f^{-1}} : S \to M$. The agent has to learn such models by collecting (m, s) pairs through its interaction with the environment, i.e. by producing $m \in M$ and observing $s = f(m)$. These learning processes are often difficult for several reasons:

- the agent has to deal with uncertainties both in the environment and in its own sensorimotor loop;
- M and S can be highly dimensional, such that random sampling in M to collect (m, s) pairs can be a long and fastidious process;
- f can be strongly non-linear, such that the learning of $\tilde{f}$ from experience is not trivial;
- f can be redundant (many M to one S), such that the learning of $\tilde{f^{-1}}$ is an ill-posed problem ($f^{-}1$ does not exist, or cannot be directly recovered from f).

When a learning process faces these issues, random motor exploration (or motor babbling) in M is not a realist exploration strategy to collect (m, s) pairs. Due to high dimensionality, data are precious whereas, due to non-linearity and/or redundancy, data are not equally useful to learn an adequate forward or inverse model.

Two important families of exploration mechanisms recently shown to be efficient to learn complex non-linear redundant sensorimotor mappings. The first one concerns the space in which the learning agent chooses points to explore, what we will call the *choice space*. Previous models [1], [2] have shown that learning redundant inverse models could be achieved more efficiently if exploration was driven by goal babbling (choice space: S), triggering reaching, rather than direct motor babbling (choice space: M). Goal babbling is especially efficient to learn highly redundant mappings (e.g. the inverse kinematics of a high dimensional arm). At each time step, the agent chooses a goal s_g in the sensory space S (e.g. uniformly), uses the current knowledge of an inverse model to infer a motor command $m \in M$ to reach s_g, observes the corresponding consequence $s = f(m)$ and update its inverse model according to the newly collected (m, s) pair. This exploration strategy allows the agent to cover the goal space more efficiently, avoiding to waist time in redundant parts of the sensorimotor space (e.g. executing many motor commands that actually reach the same goal). The second principle comes from the field of active learning, where exploration strategies are conceived as an optimization process. Samples in the input space (M in our sensorimotor framework) are collected in order to minimize a given property of the learning process, e.g. the uncertainty [3] or the prediction error [4] of the model. This allows the agent to focus on parts of the sensorimotor space in which exploration is supposed to improve the quality of the model.

Combining both principles, recent works grounded in developmental psychology have concentrated on defining *empirical measures* of interest, either in the motor M or sensory S spaces. Computational studies have shown the importance of developmental mechanisms guiding exploration and learning in high-dimensional M and S spaces and with highly redundant and non-linear f [5], [2]. Among these guiding mechanisms, intrinsic motivations, generating spontaneous exploration in humans [6], [7], have been transposed in curiosity-driven learning machines [8], [9], [10] and robots [5], [2] and shown to yield highly efficient learning of inverse models in high-dimensional redundant sensorimotor spaces [2], [11]. Efficient

versions of such mechanisms are based on the active choice of learning experiments that maximize learning *progress*, for e.g. improvement of predictions or of competences to reach goals [8], [5]. This automatically drives the system to explore and learn first easy skills, and then explore skills of progressively increasing complexity. Such intrinsically motivated exploration was also shown to generate automatically behavioural and cognitive developmental structures sharing interesting similarities with infant development [5], [12], [13], [14]. This approach is grounded in psychological theories of intrinsic motivations [6], [15], explores several fundamental questions about curiosity-driven open-ended learning in robots [5], and allows to generate some novel hypotheses for the explanation of infant development, regarding behavioural [13], cognitive [12] and brain circuitry [16] issues.

This article extends previous modeling and results of the authors [17]. In the next section we describe several exploration strategies proposed in the literature to efficiently learn complex sensorimotor mappings. Then Section III offers an integration of these strategies into a unified probabilistic framework. In Section IV we implement this general formal framework using Gaussian mixture models. In particular, we suggest an original implementation of the empirical measure of the competence progress. Section V validates our approach on two developmental robotics experiments. The first one qualitatively shows how coherent developmental trajectories can emerge from the model in a setup involving a 7-DoF simulated arm. The second one uses an articulatory synthesizer (a model of the human vocal tract able to compute auditory features from articulatory commands), and shows quantitative performance comparisons of various exploration strategies on a control task.

II. Exploration strategies

Systematic comparisons of various exploration strategies have been performed [2], [11]. These strategies differ in the way the agent iteratively collects (m, s) pairs to learn forward and/or inverse models (comparing random vs. competence progress based exploration, in either motor M or sensory S choice spaces). These strategies are summarized below (the original name of the corresponding algorithm appears in parenthesis).

- **Random motor exploration (ACTUATOR-RANDOM):** at each time step, the agent randomly chooses an articulatory command $m \in M$ (choice space: M), produces it, observes $s = f(m)$ and updates its sensorimotor model according to this new experience (m, s).
- **Random goal exploration (SAGG-RANDOM):** at each time step, the agent randomly chooses a goal $s_g \in S$ (choice space: S) and tries to reach it by producing $m \in M$ using an inverse model $\tilde{f^{-1}}$ learned from previous experience. It observes the corresponding sensory consequence $s = f(m)$ and updates its sensorimotor model according to this new experience (m, s).
- **Active motor exploration (ACTUATOR-RIAC):** at each time step, the agent chooses a motor command m by maximizing an interest value in M based on an empirical measure of the competence progress in prediction in its recent experience. The agent uses a forward model $\tilde{f}$ learned from its past experience to make a prediction $s_p \in S$ for the motor command m. It produces m and observe $s = f(m)$. The agent updates its sensorimotor model according to the new experience (m, s). A measure of competence is computed from the distance between s_p and s, which is used to update the interest model in the neighborhood of m.
- **Active goal exploration (SAGG-RIAC):** at each time step, the agent chooses a goal s_g by maximizing an interest value in S based on an empirical measure of the competence progress to reach goals in its recent experience. It tries to reach s_g by producing $m \in M$ using a learned inverse model $\tilde{f^{-1}}$. It observes the corresponding sensory consequence $s \in S$ and updates its sensorimotor model according to this new experience (m, s). A measure of competence is computed from the distance between s_g and s, which is used to update the interest model in the neighborhood of s_g.

In the two active strategies, the measure of interest was obtained by recursively splitting the choice space (M in ACTUATOR-RIAC, S in SAGG-RIAC) into sub-regions during the agent life. Each region maintains its own empirical measure of competence progress from its competence history in a relative time window. The competence is defined as the opposite of the distance between s_p and s (i.e. $-\|s_p - s\|_2$) in the active motor strategy, between s_g and s (i.e. $-\|s_g - s\|_2$) in the active goal one.

With the goal of unification, we can extract the following general principles from these strategies.

- Whatever the strategy used, the agent has to sample points in a given space. This space is M for the first and the third strategy, S for the second and the fourth. We call it the *choice space* X.
- In all but the first strategy, the agent has to make an inference from the choice space X to its "complement" in $M \times S$ (which is S if $X = M$ and M if $X = S$). We call this latter the *inference space* Y.
- In the active exploration strategies, the agent has to maintain an empirical measure of interest in the choice space X. In the other strategies, the agent makes a random sampling in X.

Table I thus suggests to classify these four strategies along two dimensions. The first one corresponds to the choice space X, which is here either M (motor strategies) or S (goal strategies). The second dimension is the kind of interest measure used by this agent at each time step to choose a point in its choice space, either uniform leading to a random sampling in X (random strategies), or based on empirical measurements, here the competence progress in prediction or control (active strategies).

III. Probabilistic modeling

The general principles we extracted in the previous section make the probabilistic framework appear as a good candidate

TABLE I. *Exploration strategies classification.*

choice space X	Interest measure	
	Uniform sampling	Competence-progress
M	Random motor exploration (ACTUATOR-RANDOM)	Active motor exploration (ACTUATOR-RIAC)
S	Random goal exploration (SAGG-RANDOM)	Active goal exploration (SAGG-RIAC)

to provide a general model which encompasses all the suggested exploration strategies.

The notations and principles of this formalization are inspired by [18], [19]. Upper case A denotes a probabilistic variable, defined by its continuous, possibly multidimensional and bounded domain $\mathcal{D}(A)$. The conjunction of two variables $A \wedge B$ can be defined as a new variable C with domain $\mathcal{D}(A) \times \mathcal{D}(B)$. Lower case a will denote a particular value of the domain $\mathcal{D}(A)$. $p(A \mid \omega)$ is the probability distribution over A knowing some preliminary knowledge ω (e.g. the parametric form of the distribution, a learning set ...). Practically, ω will serve as a model identifier, allowing to define different distributions of the same variable, and we will often omit it in the text although it will be useful in the equations. $p(A\ B \mid \omega)$ is the probability distribution over $A \wedge B$. $p(A \mid [B = b]\ \omega)$ is the conditional distribution over A knowing a particular value b of another variable B (also noted $p(A \mid b\ \omega)$ when there is no ambiguity on the variable B). For simplicity, we will often confound a variable and its domain, saying for example *"the probability distribution over the space A"*.

Considering that we know the joint probability distribution over the whole sensorimotor space, $p(M\ S \mid \omega_{SM})$, Bayesian inference provides the way to compute every conditional distribution over $M \wedge S$. In particular, we can compute the conditional distribution over Y knowing a particular value x of X, as long as X and Y correspond to two complementary subdomains of $M \wedge S$ (i.e. they are disjoint and $X \wedge Y = M \wedge S$). Thus, the prediction of $s_p \in S$ from $m \in M$ in the active motor exploration strategy, or the control of $m \in M$ to reach $s_g \in S$ in the active or random goal exploration strategies, correspond to the probability distributions $p(S \mid M\ \omega_{SM})$ and $P(M \mid S\ \omega_{SM})$, respectively. More generally, whatever the choice and inference spaces X and Y, as long as they are disjoint and that $X \wedge Y = M \wedge S$, Bayesian inference allows to compute $p(Y \mid X\ \omega_{SM})$.

Such a probabilistic modeling is also able to express the interest model, that we will call ω_I, such that the agent draws points in the choice space X according to the distribution $p(X \mid \omega_I)$. In the random motor and goal exploration strategies, this distribution is uniform, whereas it is a monotonically increasing function of the empirical interest measure in the case of the active exploration strategies. We will provide more details about the way to iteratively compute $p(M\ S \mid \omega_{SM})$ and $p(X \mid \omega_I)$ from the experience of the agent in the next section.

Given this probabilistic framework, Algorithm 1 describes our generic exploration algorithm.

Algorithm 1 Generic exploration algorithm

1: set choice space X
2: **while** true **do**
3: $\quad x \sim p(X \mid \omega_I)$
4: $\quad y \sim p(Y \mid x\ \omega_{SM})$
5: $\quad m = M\left((x, y)\right)$
6: $\quad s = exec(m)$
7: $\quad update(\omega_{SM}, (m, s))$
8: $\quad update(\omega_I, (x, y, m, s))$
9: **end while**

Line 1 defines the choice space of the exploration strategy. For example X is set to M for the motor strategies and to S for the goal strategies described in Section II, but the formalism can also deal with any part of $M \wedge S$ as the choice space. Line 3, the agent draws a point x in the choice space X according to the current state of its interest model ω_I, through the probability distribution $p(X \mid \omega_I)$ encoding the current interest over X. This distribution is uniform in the case of the random strategies and related to the competence progress in prediction or control in the active strategies of Section II. Line 4, the agent draws a point y in the inference space Y (remember that Y is such that $X \wedge Y = M \wedge S$) according to the distribution $p(Y \mid x\ \omega_{SM})$, using Bayesian inference on the joint distribution $p(M\ S \mid \omega_{MS})$. If $X = M$, and therefore $Y = S$, this corresponds to a prediction tasks $p(S \mid [M = x])$; if $X = S$, and therefore $Y = M$, this corresponds to a control task $p(M \mid [S = x])$. Line 5, the agent extracts the motor part m of (x, y), noted $M\left((x, y)\right)$, i.e. x if $X = M$, y if $X = S$. Line 6, the agent produces m and observe $s = exec(m)$, i.e. $s = f(m)$ with possible sensorimotor constraints and noises. Line 7 the agent updates its sensorimotor model according to its new experience (m, s). Line 8 the agent updates its interest model according to the choice and inference (x, y) it made and its new experience (m, s).

In this framework, we are able to more formally express each algorithm presented in Section II. The random motor strategy (ACTUATOR-RANDOM) is the simpler case where the choice space is $X = M$ and the interest model of line 3 is set to a uniform distribution over X. Inference in line 4 is here useless because motor extraction (line 5) will return the actual choice x and that there is no need to update the interest model in line 8. The active motor strategy (ACTUATOR-RIAC) differs from the previous one by the interest model of line 3 which favors regions of X $(= M)$ maximizing the competence progress in prediction. This latter is computed at the update step of line 8 using the history of previous competences, defined as the opposite differences between the prediction $y \in Y$ computed on line 4 (with $Y = S$) and the actual realization $s \in S$ of line 6. The random goal strategy (SAGG-RANDOM) is the case where the interest model is uniform and the choice space is S, implying that the inference corresponds to a control task to reach $x \in X$ by producing $y \in Y$ (with $X = S$ and therefore $Y = M$). Finally, the active goal strategy (SAGG-RIAC) differs from the previous one by the interest model which favors regions of X $(= S)$ maximizing the competence progress in control. This latter is computed in the same way that for ACTUATOR-RIAC, except that the opposite difference is here between the chosen goal $x \in X$ and the actual realization $s \in S$ (with $X = S$).

IV. Implementation with Gaussian mixture models

In the present paper, we only provide the principles of our implementation of the sensorimotor $p(M\ S \mid \omega_{SM})$ and interest $p(X \mid \omega_I)$ distributions, and leave its detailed description to a further paper. Both the sensorimotor and the interest distributions involve learning of Gaussian mixture models (GMM) using the Expectation-Maximization (EM) algorithm [20]. The values of the parameters we will use in the experiments of the next section appear in parenthesis.

$p(M\ S \mid \omega_{SM})$ involves K_{SM} (=28) components (i.e. it corresponds to a weighted sum of K_{SM} Gaussian distributions). It is learnt using an online version of EM proposed by [21] where incoming data are considered in lots in an incremental manner. Each update corresponds to line 7 of Algorithm 1 but is executed once each sm_step (=400) iterations of Algorithm 1. The ω_{SM} model is thus refined incrementally during the agent life, updating it each time sm_step new (m, s) pairs are collected. Moreover, we adapted this online version of EM to introduce a *learning rate* parameter α (from 0.1 to 0.01 in a logarithmic decreasing manner over time), allowing to set the relative weight of the new learning data with respect to the old ones.

$p(X \mid \omega_I)$ is a uniform distribution in the random strategies, whereas it has to reflect an interest measure in the active strategies, which is here related to an empirical measure of the *competence progress*. For this aim, we compute a measure of competence for each 4-tuple (x, y, m, s) collected at each iteration of Algorithm 1. We define the competence c of each iteration as $c = e^{-\|(x,y)-(m,s)\|_2}$, i.e. the exponential of the opposite of the Euclidean distance between the concatenation of the choice and inference points (x, y) and the actual realization (m, s) (remember that $X \wedge Y = M \wedge S$). As $m = M\left((x, y)\right)$ (line 5 of Algorithm 1), we actually have $c = e^{-\|S((x,y))-s\|_2}$, where $S((x, y))$ is the sensory part of (x, y). Thus, each episode is associated with a tuple (t, x, c), where t is the (normalized) time index of the iteration. We then consider the competence progress as a correlation between time and competence (the higher the correlation, the higher the competence progress). For this aim, we learn the joint distribution of this data $p(T\ X\ C \mid \omega_{TXC})$ (where T and C are the mono-dimensional variables defining the time and competence domains, with values in $\mathbb{R}^+$) using a classical version of EM on a GMM of K_I (=12) components using the last $sm\_step * im\_step$ tuples (t, x, c), on the time window corresponding to the last im_step (=12) updates of the sensorimotor model. After convergence of the EM algorithm, we bias the result by setting the a priori distributions of the model ω_{TXC} (i.e. the weight of each Gaussian) to the resulting value of the covariance between t and c (normalized to sum up to 1 and considering only the positive correlations). Finally, the interest model $p(X \mid \omega_I)$ corresponds to the Bayesian inference $p(X \mid [T = t^+]\ \omega_{TXC})$, where t^+ is the time index of the future update of the sensorimotor model (e.g. if the t values of the learning set are $\{1, \ldots, n\}$, then $t^+ = n + 1$). This allows line 3 of Algorithm 1 to sample values in regions of X which maximize the expected competence progress at the next update of the sensorimotor model.

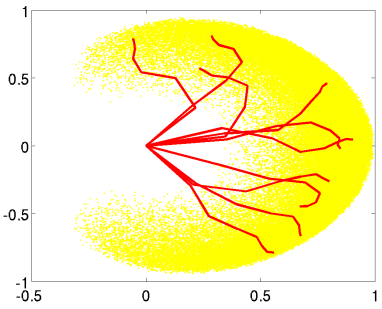

Fig. 1. The 7-DoF simulated arm. Yellow (or light gray) area shows the reachable space of the end effector in the 2-D plan. It corresponds to the position of the end effector for 100,000 motor configurations uniformly sampled in the 7-D motor space. Red (or dark gray) lines show 10 particular arm configurations randomly chosen in this set.

V. Results

This section presents two developmental robotics experiment to show (A) how the implementation of the competence progress measure as a correlation between time and competence can lead to coherent developmental trajectories and (B) how our probabilistic framework allows to compare various exploration strategies in a unified way. Note however that these results are preliminary and mainly illustrative, and that more thorough analysis are still to be done.

A. Experiment 1: developmental trajectories on a 7-DoF arm

This experiment involves a motor space M corresponding to the 7 joint angles of a simulated arm constrained on a 2-D plan. Each segment is $2/3$ shorter than the previous one and the total length is 1. The sensory space S corresponds to the 2-D Cartesian coordinates on the plan of the end effector. Joint angles are constrained in the range $[-\pi/3, \ldots, \pi/3]$. Figure 1 illustrates this sensorimotor space. We observe that random motor configurations favors positions of the end effector "in front of the agent" (high values on the abscissa) due to the redundancy of the sensorimotor mapping in this area.

With this experiment, we want to qualitatively evaluate our original measure of competence progress as a correlation between time and competence. Figure 2 shows the evolution of the reached and choice points in a simulation implementing an active goal exploration strategy (bottom-right cell of Table I, i.e. $X = S$ and $p(X \mid \omega_I)$ reflects the competence progress as defined in Section IV). We observe that the system explores the positions that the end effector is able to reach in a developmental manner. In the first two plots (top-left and top-middle), the agent sets sensory goals corresponding to "easy configurations" in front of him (similar to those displayed in Figure 1). Then, it progressively sets goals in part of the space harder to reach, i.e. behind him (but still continues to explore easier parts, a possible reason being a tendency of our adapted version of the EM algorithm to forget previously learned data).

B. Experiment 2: performance comparison on a vocal control task

This experiment involves the articulatory synthesizer of the DIVA model described in [22]. This synthesizer is based on the

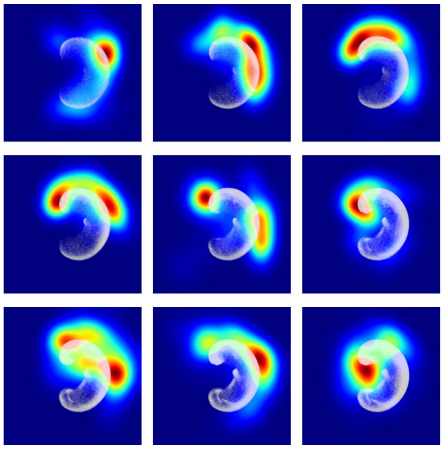

Fig. 2. Visualization of the reached points and interest distribution evolution over time in the active goal strategy on the 7-DoF arm experiment. Each plot represents the state of the reached points and the interest distribution every 100 updates of the interest model (from left to right, then top to bottom). White points are the points reached by the agent since the beginning of the simulation, corresponding to line 6 of Algorithm 1 (e.g. the top-right plot shows the reached points from the start to the 300^{th} update). The color map is a visualization of the interest distribution in the corresponding time window (e.g. the top-right plot shows the distribution in the time window from the 200^{th} to the 300^{th} update). It is obtained by computing an histogram of the goals drawn on line 3 of Algorithm 1 with 100 bins per dimension in S (hence 10,000 bins) and applying a 5-bins wide Gaussian filter.

Maeda's model [23], using 13 articulatory parameters: 10 from a principal component analysis (PCA) performed on sagittal contours of images of the vocal tract of a human speaker, plus glottal pressure, voicing and pitch. It is then able to compute the formants of the signal (among other auditory and somatosensory features). In the present study, we only use the 7 first parameters of the PCA (motor space M) and the two first formants (sensory space S), approximately normalized in the range $[-1, 1]$. We refer to [14] for more details on the Maeda model (also used in a developmental robotics setup) and to [22] for more details on the particular synthesizer of the DIVA model. Figure 3 explains the general principles of speech production.

We ran the implementation of the algorithm described in the previous sections with different choice spaces and interest distributions corresponding to the four strategies ACTUATOR-RANDOM, ACTUATOR-RIAC, SAGG-RANDOM and SAGG-RIAC described in Section II. We evaluate the efficiency of the obtained sensorimotor models to achieve a control task, i.e. to reach a test set of goals uniformly distributed in the reachable auditory space.

Figure 4 shows the performance results of the four exploration strategies on a control task during the life time of learning agents. We observe that the strategies with S as the choice space (random and active goal) are significantly more efficient that those with M (random and active motor), i.e. both convergence speed (say around 100 updates) and generalization at the end of the simulation (500 updates) are better. Moreover, both convergence speed and generalization are better for the active than for the random strategies. These results are similar (though less significant) to those obtained in previous experiments [2], [11] in other sensorimotor spaces (e.g. a arm reaching points on a plan as in Experiment 1), and

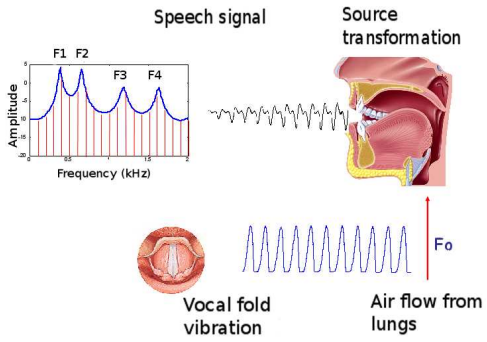

Fig. 3. Speech production general principles. The vocal fold vibration by the lung air flow provides a source signal: a complex sound wave with fundamental frequency F_0. According to the vocal tract shape, acting as a resonator, the harmonics of the source fundamental frequency are selectively amplified or faded. The local maxima of the resulting spectrum are called the formants, ordered from the lower to the higher frequencies. They belong to the major features of speech perception.

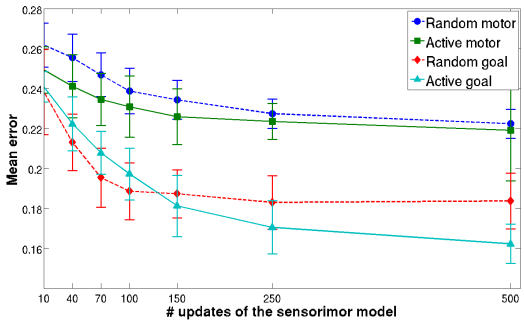

Fig. 4. Performance comparison of the four exploration strategies. X axis: number of update of the sensorimotor model. Y axis: Mean error distance on a control task where an agent has to reach 30 test points uniformly distributed in the reachable area of S. For each evaluation point $s_g \in S$, the agent infers 10 motor commands in M from the distribution $p(M \mid s_g\ \omega_{SM})$, where ω_{SM} is the state of the sensorimotor model at the corresponding time step (number of update on the X axis). The error of an agent at a time step is the mean distance between the sensory points actually reached by the 10 motor commands and the evaluation point s_g. Each curve plots the mean and standard deviation of the error for 10 independent simulations with different random seeds, for each of the four exploration strategies described in the previous sections.

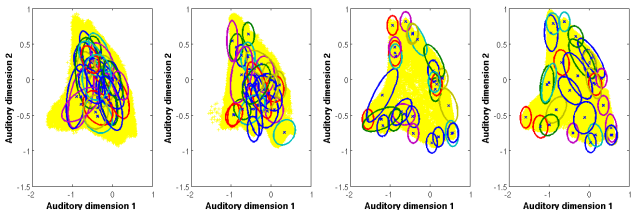

Fig. 5. State of the sensorimotor model at the end of the simulations for the four exploration strategies (from left to right: ACTUATOR-RANDOM, ACTUATOR-RIAC, SAGG-RANDOM and SAGG-RIAC). Auditory parameters are the two first formants computed by the articulatory synthesizer. Yellow (or grey) area is the auditory area reached by the agent at the end of the simulation. Ellipses represents the Gaussians of the sensorimotor GMM $p(M\ S \mid \omega_{SM})$ projected on S (1 standard deviation).

we refer to the corresponding paper for a thorough analysis of these results. This shows that our unified probabilistic framework is suitable to encompass all these exploration strategies.

Figure 5 shows the state of the sensorimotor model ω_{SM} projected in the sensory space S at the end of one simulation for each of the four exploration strategies. We observe that the position of the Gaussians are relatively disorganized when M (two first plots) is the choice space, whereas some structure appears when it is S (two last plots). Self-organization seems to spontaneously appear in the choice space where points are sampled from the interest model (either uniformly or actively). When this latter is S, this probably provides a sensorimotor model allowing to better control the vocal tract to reach auditory goals. Another observation is that the auditory space seems to be covered more uniformly in the active than in the random goal exploration strategy. The reason is that the random strategies more often choose goals outside the reachable space, thus favoring reaching at the borders of the sensory space S, whereas the active ones focus on the competence progress. To summarize these preliminary results, using S as the choice space is more efficient than using M because self-organization in S is adequate to achieve a control task, and the active goal strategy is more efficient than the random goal one because it allows to focus on the reachable part of S (and perhaps to set goals of increasing difficulties, as suggested in the arm experiment).

VI. Conclusion

We have integrated in this paper two important exploration principles of developmental robotics (exploration in the sensory space and active learning based on an empirical measure of the competence progress) into an integrated probabilistic framework able to express various exploration strategies in a compact and unified manner. We then suggest an original implementation of the underlying algorithm using GMMs where the competence progress is measured as a statistical correlation between time and competence. Finally, we showed that this modeling can be applied to various sensorimotor spaces (bracchio-visual and articulatory-auditory), that it is able to match performance comparison results obtained in previous works and seems to have interesting properties in terms of developmental trajectories and self-organization.

Further works should rely the approach to other tentatives of exploration strategy unification (e.g. [24], [25]). We also want to study the effect of an online adaptation of the choice space, taking advantage of the fact that our formalism does not restrict it to be either M or S. For example, we could study how the agent iteratively adapts which part of the sensorimotor space it is interested in at a given time of its development, favoring exploration in sensorimotor dimensions which display higher measures of competence progress. Finally, we are currently using the GMM implementation to model the emergence of articulated vocalizations in an intrinsically motivated agent.

Acknowledgment

This work was partially financed by ERC Starting Grant EXPLORERS 240 007. The authors would like to thank Louis-Jean Boë for the design of Figure 3 (vocal tract by Sophie Jacopin).

References

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

[2] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, 2012.

[3] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[4] S. Thrun, "Exploration in active learning," *Handbook of Brain Science and Neural Networks*, pp. 381–384, 1995.

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

[6] D. E. Berlyne, "A theory of human curiosity," *British Journal of Psychology*, vol. 45, pp. 180–191, 1954.

[7] E. Deci and R. M. Ryan, *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum Press, 1985.

[8] J. Schmidhuber, "A possibility for implementing curiosity and boredom in model-building neural controllers," in *Proc. SAB'91*, J. A. Meyer and S. W. Wilson, Eds., 1991, pp. 222–227.

[9] A. Barto, S. Singh, and N. Chenatez, "Intrinsically motivated learning of hierarchical collections of skills," in *Proc. 3rd Int. Conf. Dvp. Learn.*, San Diego, CA, 2004, pp. 112–119.

[10] J. Schmidhuber, "Formal theory of creativity, fun, and intrinsic motivation (1990-2010)," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 230–247, 2010.

[11] A. Baranes and P.-Y. Oudeyer, "Intrinsically motivated goal exploration for active motor learning in robots: a case study," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), Taipei, Taiwan*, 2010.

[12] F. Kaplan and P.-Y. Oudeyer, "The progress-drive hypothesis: an interpretation of early imitation," *Models and mechanims of imitation and social learning: Behavioural, social and communication dimensions*, pp. 361–377, 2007.

[13] P.-Y. Oudeyer and F. Kaplan, "Discovering communication," *Connection Science*, vol. 18, no. 2, pp. 189–206, 06 2006.

[14] C. Moulin-Frier and P.-Y. Oudeyer, "Curiosity-driven phonetic learning," in *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, 2012, pp. 1–8.

[15] M. Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention*. HarperCollins, 1997.

[16] F. Kaplan and P.-Y. Oudeyer, "In search of the neural circuits of intrinsic motivation," *Frontiers in neuroscience*, vol. 1, no. 1, p. 225, 2007.

[17] C. Moulin-Frier and P.-Y. Oudeyer, "The role of intrinsic motivations in learning sensorimotor vocal mappings: a developmental robotics study," in *Proceedings of Interspeech*, Lyon, France, 2013, p. In press.

[18] E. T. Jaynes, *Probability Theory: The Logic of Science*, G. L. Bretthorst, Ed. Cambridge University Press, June 2003.

[19] O. Lebeltel, P. Bessiere, J. Diard, and E. Mazer, "Bayesian robot programming," *Autonomous Robots*, vol. 16, p. 4979, 2004.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, Jan. 1977.

[21] S. Calinon, *Robot Programming by Demonstration*. CRC, 2009.

[22] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, "Neural modeling and imaging of the cortical interactions underlying syllable production," *Brain and language*, vol. 96, no. 3, pp. 280–301, 2006.

[23] S. Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," *Speech production and speech modelling*, pp. 131–149, 1989.

[24] M. Lopes and P.-Y. Oudeyer, "The strategic student approach for life-long exploration and learning," in *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, 2012, pp. 1–8.

[25] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? a typology of computational approaches," *Frontiers in Neurorobotics*, vol. 1, 2007.