Developing Social Action Capabilities in a Humanoid Robot using an Interaction History Architecture

Naeem Assif Mirza [1,2], Chrystopher L. Nehaniv [1], Kerstin Dautenhahn [1], René te Boekhorst [1]

[1]*Adaptive Systems Research Group, School of Computer Science*
University of Hertfordshire, College Lane, Hatfield, AL10 9AB. United Kingdom
$\{C.L.Nehaniv, K.Dautenhahn, R.teBoekhorst\}@herts.ac.uk$

[2]*Istituto Italiano di Tecnologia, Via Morego 30, Genoa, Italy*
$assif.mirza@iit.it$

***Abstract*— We present experimental results for the humanoid robot Kaspar2 engaging in a simple "peekaboo" interaction game with a human partner. The robot develops the capability to engage in the game by using its history of interactions coupled with audio and visual feedback from the interaction partner to continually generate increasingly appropriate behaviour. The robot also uses facial expressions to feedback its level of reward to the partner. The results support the hypothesis that reinforcement of time-extended experiences through interaction allows a robot to act appropriately in an interaction.**

I. Introduction

This paper reports the results of an experiment showing a humanoid robot (Kaspar2 - Fig 1) using its history of interaction to acquire the ability to engage in the early interaction game "peekaboo" with a human interaction partner. The robot is a simple upper-body humanoid that can display a range of facial and bodily expressions. The peekaboo engagement is developed by the robot using the Interaction History Architecture, a developmental control architecture based on the grounded history of sensorimotor interactions.

In earlier experiments (see [1]), this architecture was shown to be capable of supporting development of a turn-taking interaction in a non-humanoid robot which took appropriate sequences of actions or gestures based on its own grounded sensorimotor experience. This new experiment uses interaction history-based control architecture, relying on temporally extended grounded sensorimotor experiences, deployed on an expressive an expressive humanoid for the first time. The humanoid embodiment enhances the richness of the possible interaction for instance by adding the ability to feedback reward through facial gestures. An audio modality is also added to the visual and other sensorimotor data, and is employed in perception of reward along with face recognition. Furthermore, for the first time in a robotic platform, we show how continual modification of the space of experiences through merging and forgetting builds a more adaptive and focused interaction history.

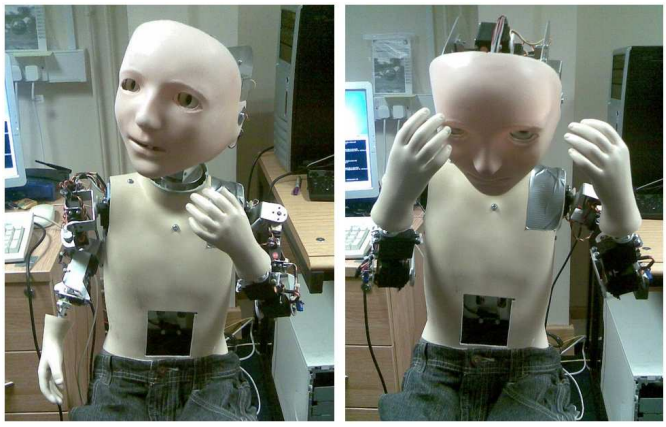

Fig. 1. The Kaspar2 robot (University of Hertfordshire) has two 5 DoF arms, a 3 DoF neck, two coupled 2 DoF eyes containing colour cameras and a flexible face actuated by two further motors at the mouth.

A. Interaction Histories

We define an interaction history for an embodied agent as *the temporally extended, dynamically constructed, individual sensorimotor history of an agent situated and acting in its environment, including the social environment, that shapes current and future action* [1]. The history is grounded in the sensorimotor coupling of the agent with its environment and therefore the development of the action capabilities of an agent based on such a history are also grounded and meaningful from the agent's perspective.

This aligns with the "embodied cognition" hypothesis, that *"cognition is a highly embodied or situated activity and suggests that thinking beings ought therefore be considered first and foremost as acting beings."* [2]. Lakoff & Johnsson [3] also argue that all cognition, including representations and memory of categories, eventually grounds out in embodiment and Glenberg [4] also argues that the purpose of perception and memory for the natural environment is to guide action, and that even abstract concepts can be interpreted in terms of physical actions and properties. In general we can say that

memory *manifests* itself as embodied action of some kind. That is, it is in actions resulting from recall that one witnesses memory and that recall itself is dependent on embodiment.

Autonomous embodied artificial agents that make use of interaction histories in guiding their actions can be thought of as extending their temporal horizon beyond that of a simple *reactive agent* and become *post-reactive* systems when acting with respect to a broad temporal horizon by making use of temporally extended episodes in interaction dynamics [5].

We hypothesize that a dynamically constructed history that is used to generate and select actions in an embodied agent can also serve as the basis for *ontogenetic development* of the agent. Self-organization (merging and deletion of) experiences in the history can provide abstraction as well as anticipation [6]. Development in this case can be seen as *the increasing richness of the connections of experience with action*, mediated by suitable mechanisms. Such a history can facilitate incremental development at the borders of experience (*cf.* Vygotsky's "zone of proximal development" [7])

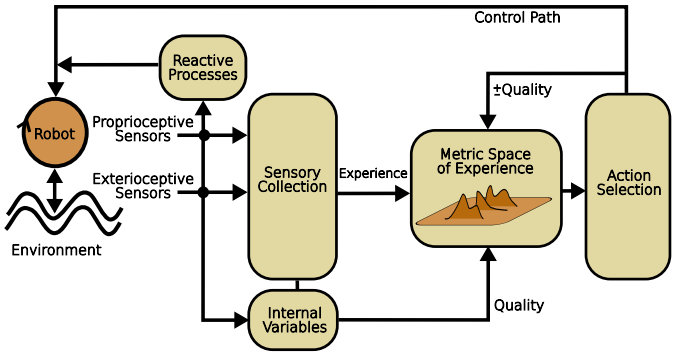

Fig. 2. Schematic of the Interaction History Architecture

II. Interaction History Architecture

The Interaction History Architecture is shown schematically in Figure 2. The approach is as follows:

1) to continually gather sensorimotor data and find "suitable" episodes of sensorimotor experience in the history *near* (in terms of the experience metric) to the current episode;
2) depending on the course of subsequent experience, to choose from among actions that were executed when these episodes were previously encountered;
3) where no suitable experiences are found, to choose random actions.

There are two key aspects of this architecture. The first is the *metric space of experience* whereby new experiences appear as points in a growing and changing high-dimensional metric space. The metric space is enhanced with *quality* information, potentially received from the environment, from internal drives or from other sources such as affective state. Each experience is also associated with actions executed during the experience. The second is the *action selection* system. This "closes the perception-action loop" and also closes an internal loop feeding back and modifying the experience space. The quality associated with each experience combined with proximity in the metric space is used to select experiences from the history and select actions associated with those experiences.

A. Interaction History Space

Briefly[1], the *Interaction History Space* consists of:

Sensorimotor Experiences: Time-series of sensor readings from all available sensors of a robot, from time t to another time $t+h$ where h is the *horizon length* of the experience.

The Experience Metric: A metric measure of distance between sensorimotor experiences. Based on an information-throetic measure of distance between sensor time-series viewed as values of random variables. (Crutchfield-Rényi Information Metric [8]).

Next Action information: The next action executed after an experience is associated with that experience.

Quality information: A value representing environmental reward received after the experience (for a particular time span).

Thus the metric space of experience in the Interaction History Architecture, the *interaction history space*, can be described by the tuple $(\boldsymbol{\epsilon}, \boldsymbol{D}, \boldsymbol{q}, \boldsymbol{a})$, where $\boldsymbol{\epsilon}$ is a collection of quantized "experiences", $\boldsymbol{D}$ is the a matrix of distances between elements of $\boldsymbol{\epsilon}$, $\boldsymbol{q}$ is a vector of quality values and $\boldsymbol{a}$ a vector of actions.

The metric space is constructed continuously as the robot experiences its environment. A new experience is created every *Granularity* G timesteps, and consists of Horizon h timesteps counting back from the current timestep. Where $h > G$ the experiences will overlap. Each sensor reading is quantized into Q evenly-sized bins. Each new quantized experience is compared to other experiences in order to determine its neighbours. This process, if all experiences are compared, results in a distance matrix between experiences which defines the structure of the metric space as it is experienced by an individual robot.

B. Action Selection

A simple mechanism is adopted for action selection whereby the robot can execute one of a number of "atomic" actions (or no action) at any timestep. The actual action selected will either be a random selection of one of the atomic actions, or will be an action that was previously executed *after* an experience in the history. Both "quality" and proximity to the current episode in the space affect the chance of an historical experience (and therefore action) being selected.

This process ensures the robot may still choose a random action as this may potentially help to discover new, more salient experiences This has the advantage of emulating body-babbling, i.e. apparently random body movements that have the (hypothesized) purpose of learning the capabilities of the

[1]For further details see [1].

body in an environment [9]. Early in development, there are fewer, more widely spread experiences in the space, so random actions would be chosen more often. Later in development, it is more likely that an the action selected will come from past experience.

An advantage of this approach is that behaviour can be bootstrapped from early random activity, and later behaviour built on previous experience.

1) Roulette-Wheel Action Selection: An experience is selected from K *candidate experiences* near to the current experience $E_{current}$. The chance of random action selection is also represented in that list. The probabilities are calculated using a "gravitational model" where each experience is represented as a point mass a particular distance from $E_{current}$. The probability of selecting an experience E_i from $E_1, \ldots, E_K$ is:

$$p_i = \frac{m_i q_i}{D(E_{current}, E_i)^2} \tag{1}$$

where q_i is the *quality value* of E_i, m_i is the mass (*i.e.* how many experiences have been merged into this experience) and $D(E_{current}, E_i)$ is the experience distance[2].

The chance of random is added to the list as:

$$p_0 = \frac{\sum_{i=1}^{K} p_i}{(r_{max}/\tau)^2} \tag{2}$$

where r_{max} is the radius of the ball that includes the ranked experiences and τ is a *temperature* factor, that controls the chance of random action selection.

Then the weighting on the "roulette wheel" is given by:

$$w_i = \frac{p_i}{\sum_{i=0}^{K} p_i} \tag{3}$$

C. Update of Environmental Reward

The quality value q has bearing on the selection of the experience, and in turn on the action-selection process. The quality value is intended to reflect how useful the experience is in terms of positive or negative environmental feedback, and is derived directly from the internal reward function or an external reward measured by the robot's sensors.

In the simplest case, the immediate (instantaneous) reward received from the environment is associated with the current experience. An alternative scheme is for the quality associated with an experience to be dependent not only on the current reward, but also on the future reward. In the present implementation the *future reward* for an experience $E_{t,h}$ for some given horizon h_{future} is the maximum reward over the next h_{future} following the experience.

D. Merging and Deletion of Experiences in the Interaction History Space

It is necessary to employ strategies such as *merging* and *forgetting* if storage and computation requirements are to be controlled. However, employing such a strategy also provides a powerful mechanism for continually changing and adapting the experience space and is therefore of fundamental importance.

The merging strategy is to merge any two experiences closer than a threshold T_{merge}. T_{merge} was fixed for the most part, however alternative strategies were trialled during development of the algorithm, including adapting the threshold such that the maximum number of experiences in the space remained constant.

The meta-information associated with experiences that are merged are also assimilated. Actions from both merged experiences are accumulated, resulting in an action probability distribution; the quality values are averaged; and, a weight value, indicating the number of experiences that have been merged together, is set to the sum of the weights of the merged experiences.

Experiences may also be deleted, that is, forgotten. There are a number of different strategies to decide which experiences should be forgotten, and the one used here is to forget those experiences which have lower quality values and thus will have little or no impact on future action selection. Specifically, experiences older than h_{future} with a quality less than or equal to T_{purge} will be deleted.

III. Development using Interaction Histories Through Playful Interaction

We describe an experiment that illustrates how a robot can develop action capabilities based on its history of interaction with the environment through the use of the architecture presented. The scenario is a simple communicative interaction game, "peekaboo", that uses simple non-verbal gestures. The peekaboo game as a research tool is discussed, followed by a description of an experiment using an upper-body humanoid robot that uses its interaction history to develop the capability to engage in a peekaboo interaction with a human partner.

A. Peekaboo as a Research Tool

The development of gestural communicative interaction skills is grounded in the early interaction games that infants play. In the study of the ontogeny of social interaction, gestural communication and turn-taking in artificial agents, it is instructive to look at the kinds of interactions that children are capable of in early development and how they learn to interact appropriately with adults and other children. A well known interaction game is "peekaboo" where classically, the caregiver having established mutual engagement through eye-contact, hides their face momentarily. On revealing their face again the care-giver cries "peek-a-boo!', "peep-bo!", or something similar, resulting in pleasure for the infant before the cycle repeats.

In relation to the development of social cognition in infants, cyclic social interaction games are important as they are considered to contribute developmentally to infant understanding and practise of social interaction. Peekaboo provides the caregiver with the scaffolding upon which infants can co-regulate their emotional expressions with others, build social expectations and establish primary intersubjectivity [11].

[2]The "Experience Metric" -see [10].

It is as as a simple example of a socially-based interaction, that peekaboo is used in these experiments, but we expect our architecture to operate in many other situations.

B. Peekaboo with the Humanoid Robot Kaspar2

We describe an experiment that demonstrates how a robot can use its history of interactions with a human partner to engage in the peekaboo game. This implementation uses audio both as an extra sensory modality and as reward feedback.

1) Method: The robot and human partner[3] were positioned facing each other at a distance of a few feet at the same eye-level. The robot control software was started with the interaction history containing no previous experiences. Interaction then commenced with the robot executing various actions and the human offering vocal encouragement when it was thought appropriate. The interaction then continued for approximately two to three minutes.

Three different conditions were tried differing in the vocal reward feedback during the interaction. Either "peekaboo" was encouraged, an alternative action sequence was encouraged, or no vocal encouragement was offered at all.

The experimental hypothesis was that encouraging the hiding action would result in a higher rate of peekaboo sequences than would be expected from random action selection. Furthermore, this should also be the case when other actions are encouraged instead. Finally, this hypothesis was also tested by the no-encouragement condition with the expectation that no action would be selected in preference to any other.

2) Interaction History Architecture Components and Settings: **Metric Space of Experiences:** The sensor rate during these experiments resulted in an average timestep length of approximately 300*ms*. Experiences were created every $G = 2$ timesteps - permitting real-time creation of the metric space, quantizing the sensor data into $Q = 5$ bins. The horizon h for experiences was either 16 or 20 depending on the run. Quality was assigned to experiences as the maximum environmental reward received in the subsequent $h_{future} = 32$ or $h_{future} = 40$ timesteps (again, depending on the run). These values were chosen as reasonable values, the horizon approximately matching the duration of a single behavioural sequence.

The thresholds for merging and deletion were set at $T_{merge} = 0.6bits$ and $T_{purge} = 0.9bits$ respectively. With these values, a combination of the merging and forgetting processes resulted in a manageable sized metric space for real-time operation.

Action Selection: The closest $K = 4$ neighbours of the current experience within a radius of $r_{max} = 2.0$*bits* of $E_{current}$ were considered in the action-selection process.

3) Motivational Dynamics: In this experiment, motivation feedback (reward) is provided through two mechanisms: observation of a face, and audio feedback.

[3]Note that for all these experiments the lead author took the role of the human partner and so was fully aware of the capabilities of the robot and of the software.

TABLE I

KASPAR2 PEEKABOO: ACTIONS

Group	Number	Action	Description
Movement Actions	3	HL	Head Left
	4	HR	Head Right
	6	HID	Hide Head with Hands
	8	RAU	Right Arm Up
	9	LAU	Left Arm Up
	12	RAW	Wave Right Arm
	13	LAW	Wave Left Arm
	14	TR	"Think" Right - raise right arm to chin and look right
	15	TL	"Think" Left - raise left arm to chin
Facial Expressions	1	Smi	Smile
	2	Neu	Neutral
	16	Frn	Frown
Resetting Actions	0	Rst	All motors to resting position
	7	NA	No Action
	5	HF	Head to forward position
	10	RAD	Right Arm Down
	11	LAD	Left Arm Down

Face: Human-like faces were detected in the robot's camera image[4] and this provided direct positive reward R_f, constrained to be in the range $[0, 1]$. Habituation causes this reward to drop-off over time.

Sound: Sound was captured from a microphone, and used both as an additional sensory signal as well as providing further environmental reward. The sum of the amplitudes of the sound signal samples over the period of a timestep, ε_{sound}, provides a new sensory input to the robot and is normalized to the range [0,1].

Resulting Reward Signal: The final reward signal R generated by the robot in response to it's environmental interaction is a combination of the sound and face reward signals. $R = \max(1, \alpha(R_f + R_s))$ where α, in the range [0,1] attenuates the reward signal and is set at 0.75 for this experiment meaning that neither reward signal on its own can result in a maximum R, but requires support from the other reward signal.

4) Experimental Materials and Methods: **Robot:** The robot used was the upper-body humanoid Kaspar2 robot created at the University of Hertfordshire, see Figure 1. The robot has 17 individually controlled DC servo motors: three in the neck controlling head orientation, two controlling the coupled eyes, two controlling the mouth for facial expression, and five controlling each arm. The interaction history architecture and control software was written in C++ as multiple interacting modules, with the communication layer and abstraction of hardware control provided by the YARP framework [13].

Actions: A total of 17 actions were available to the robot, and these can be considered in 3 groups: movement actions, facial expressions and resetting actions. These are listed in Table I. The types of action that the robot can execute at any time depends on which action was last executed. This is so that the

[4]Using the OpenCV library implementation [12] of Viola-Jones HAAR cascades.

robot does not attempt to execute actions that could possibly damage it. The configuration therefore defines the set of next actions possible after any given action and the action selection process is responsible for ensuring that these conditions are met.

5) Defining a Peekaboo Sequence: A "peekaboo" sequence is defined to be a sequence of actions beginning with the robot hiding its face (action 6 - HID), followed by any number of "no-action" actions (action 7 - NA) and ending with the robot back in the resting position (action 0 - Rst). Furthermore, for the purposes of evaluating the results of this experiment the actions should be selected from previous experience rather than executed randomly.

To measure the relative amounts of peekaboo in any given period of behaviour, $p_{sel}(A^{HID})$, the percentage of times the hiding action was *selected* as compared to other "movement" actions, was used as a measure and is calculated as follows. Given N possible actions $\{A^1, A^2, \ldots A^N\}$ and a period of behaviour consisting of K actions executed (selected or random), action A^n will be executed $F(A^n) = F_{rand}(A^n) + F_{sel}(A^n)$ times, where F_{rand} indicates the frequency of random executions and F_{sel} the frequency of the action being deliberately selected. Then the percentage of times the Hiding action A^{HID} was selected is given by $P_{sel}(A^{HID}) = 100F_{sel}(A^{HID})/K$ Note that for the purpose of evaluating "peekaboo", only actions in the "movement actions" group were considered (see Table I).

6) Success Criteria: To consider a run successful the encouraged behaviour should be executed repeatedly for some extended period of the run. Remembering that the system starts by executing random actions and building-up experience before potentially using its history to execute the appropriate action repeatedly, then we might reasonably consider the run to be successful if the behaviour made up at least a third to half of overall behaviours executed. Furthermore, a full peekaboo cycle would be comprised of more than one (usually 2 or 3) selected actions that together make up the selected behaviour. So from an action perspective if the encouraged action was selected more than around $10 - 15\%$ of the time, then the run could be considered successful. However, the percentage of selection alone was not the sole criteria for judging success. Instead, each trace was examined to see when, if, and how often repeated behaviour was executed. Ultimately however, some runs were still considered borderline - that is they may have failed to satisfy some aspect of the criteria. The comments in Table II offer explanations for the decisions in these and other cases.

C. Results

A total of 22 runs were completed. 16 of these for the first condition (encouraging the Hiding action), 3 for the second condition and 3 for the no-encouragement condition. The results are summarized in Table II.In most of the experimental runs it was fairly straightforward to estimate whether the experiment successfully supported, or clearly failed, the hypothesis that the interaction history would result in increases in frequency of the encouraged action. However, in 2 of the runs, this was not possible ("?" in Table II). In run d0039, the hiding action was the only one to be selected (rather than chosen randomly) however the run was too short for successful evaluation. In run d0052, the figures for the whole run do not indicate success, however, the results are borderline as the peekaboo behaviour was clearly beginning to occur towards the end of the run.

TABLE II
IHA ON KASPARII: EXPERIMENTAL RUNS SUMMARY

Run	Type	h	Comment	HID Chosen	Result
d0032	Pkb	16	HID executed early and repeated	55.17%	Success
d0033	Pkb	16	HID executed early and repeated	41.18%	Success
d0034	None	16	HID only twice randomly	0.00%	Success
d0035	Alt HL	16	HL action chosen often. HID also chosen. HL=36.59%	14.63%	Success
d0036	Pkb	16	HID chosen often.	42.11%	Success
d0037	Pkb	16	3 HID actions selected, but RAW selected more often	13.64%	Fail
d0038	Pkb	16	No random HID to encourage.	0.0%	Fail
d0039	Pkb	16	Run too short	12.50%	?
d0041	Pkb	16	Mixed actions - some HID	5.49%	Fail
d0042	Pkb	16	Mixed actions	9.68%	Fail
d0043	Pkb	16	HID only twice	1.09%	Fail
d0044	Pkb	16	HID throughout	18.87%	Success
d0045	None	16	Few random HID actions	0.00%	Success
d0046	Alt HL	16	HL chosen many times HL=11.84%	2.63%	Success
d0049	Pkb	20	Few HID actions	3.26%	Fail
d0050	Pkb	20	HID chosen often	26.32%	Success
d0051	Pkb	20	HID chosen often	19.32%	Success
d0052	Pkb	20	HID not chosen enough for success over run. However, regular peekaboo was begining to occur at the end.	4.96%	?
d0053	Pkb	20	HID chosen often	17.46%	Success
d0054	Pkb	20	HID chosen often	61.76%	Success
d0055	Alt TR	20	TR (Think-Right) encouraged. TR=26.00%	0.00%	Success
d0056	None	20	Some HID chosen	2.53%	Success

Where a result could be determined, 14 out of 20 runs (70%) were successful. In the following sections representative results from each condition are discussed.

1) Peekaboo Encouragement Condition: Figure 3 shows for the first run (d0032), how the motivational variables (face, sound and resultant reward) vary with time, along with the

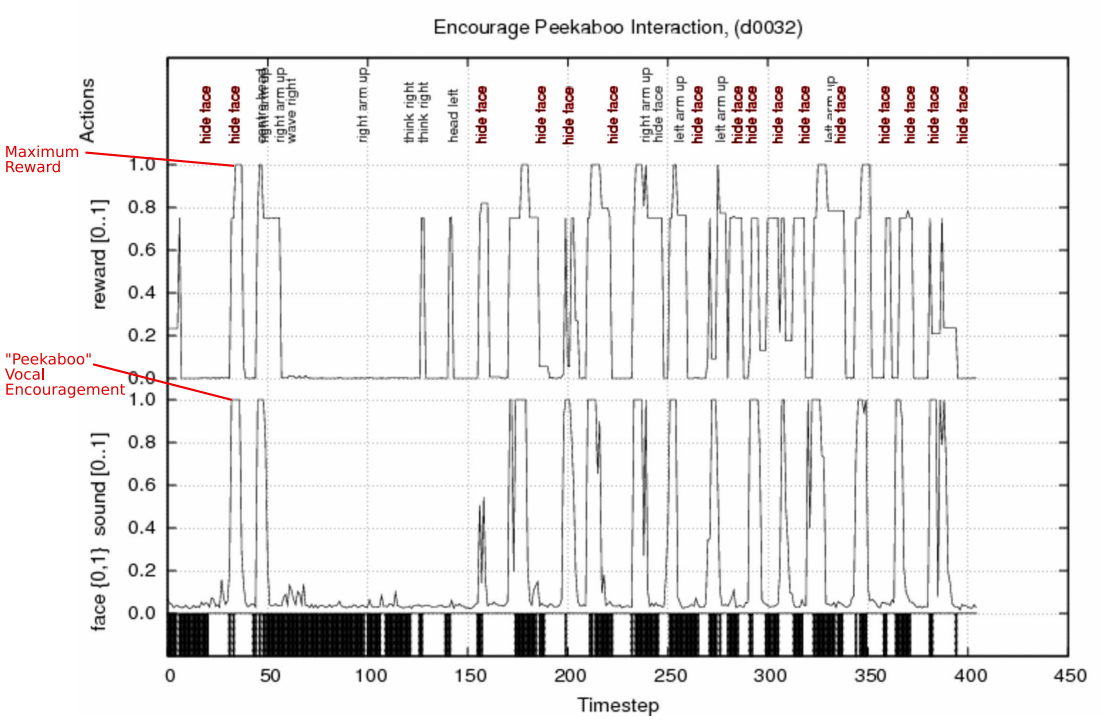

Fig. 3. *Kaspar2 Results d0032. Example of Peekaboo Encouragement Condition.* The trace shows, against time, the detection of the face and audio encouragement as well as the resulting reward. Along the top are shown the actions executed.

actions being executed . The interaction partner encourages the first “peekaboo” sequence (“hide-face” on the diagram). Note that a “peekaboo” action is actually a combination of the action to hide the face (action 6), any number of “no-action” actions (action 7) and an action to return to the forward resting position (action 0) (for clarity only the primary action is shown on the trace). This results in a maximal reward shortly after the hide-face action, and as the interaction partner continues to reinforce the peekaboo behaviour with vocal reward, this pattern can be seen repeated throughout the trace.

As the chance of choosing a random action rather than selecting one using the history gradually declines the early part of the run will be more exploratory (have more randomly selected actions) whereas towards the end of the run, actions will be more likely to be deliberately selected using past experience. It can be seen that during the first half of the run various different actions are tried, but during the second half of the run, the “hide-face” action is chosen regularly.

The timing of the motivational feedback given by the interaction partner to the robot is important in determining what actions are executed. In Figure 4 from run d0050, the encouragement for the hiding action (and subsequent actions to return the robot to the resting position) is only received *after* the robot additionally turns its head to the side. The result is that when the robot decides to repeat the hiding action, it generates experiences which are likely to generate the actions that were executed following the original hiding action, *i.e.* the robot hides its face, returns to face the front and immediately turns its head to the side.

This behaviour (of the architecture) is an important part of how not just single actions are repeated, but instead how sequences of actions and robot behaviour are replayed, and it is this that encourages a fuller development of capabilities of the robot. It is important to note also that a specific sequence of actions are not learnt, instead it is the continuing generation of experience through the structural coupling of the embodied agent and its environment that drives this observed repeated behaviour. This can be clearly seen from Figure 4 in that the timing of the subsequent head-turn following a hiding action is not always the same, and indeed does not always occur.

2) Alternative Action Encouragement Condition: To illustrate that the operation of the interaction history is not limited to the peekaboo behaviour, the interaction partner also encouraged certain alternative actions rather than hiding. In two cases the “head left” (HL) action was encouraged (once also with a different call of “hello!” instead of “peekaboo!”) and in one case the “think right” (TR) action was encouraged instead. In each of these cases the predominant action after some time was the encouraged one.

3) No Encouragement Condition: The final condition where the interaction partner offered no or very little encouragement resulted in various kinds of behaviour, none of which reinforced any particular action over any other, other than “doing nothing”.

Run d0045 was completed without an interaction partner present and so offered no reward feedback at all. The result showed some random actions being chosen at first but as time goes on, “movement actions” are not chosen and the robot

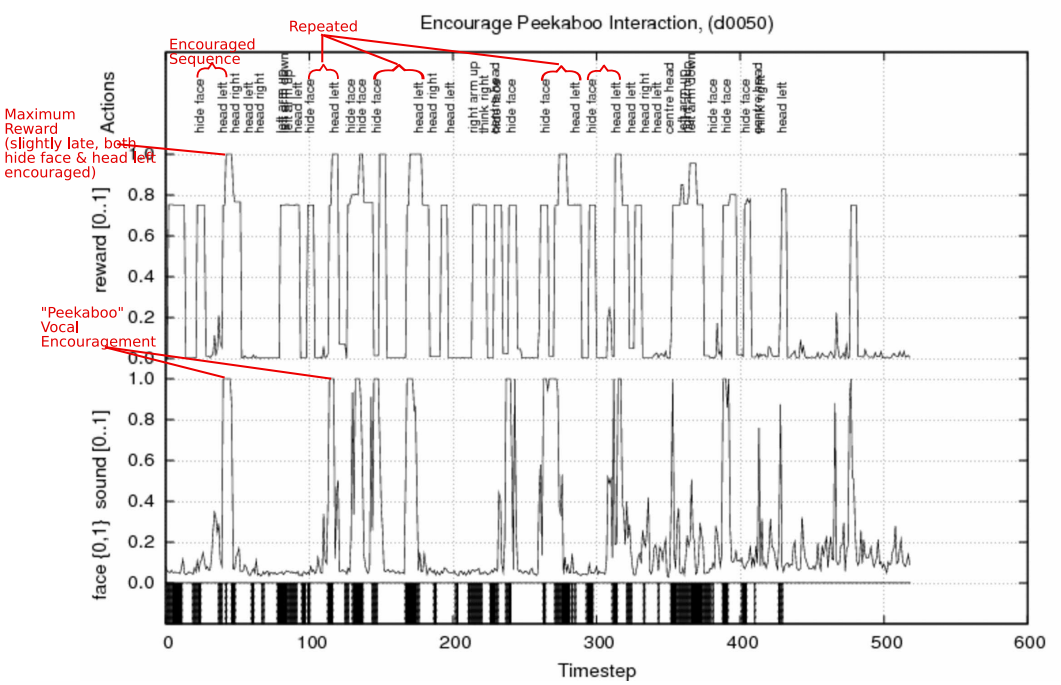

Fig. 4. *Kaspar2 Results d0050. Showing a repeated action sequence.* A multiple action sequence is encouraged and repeated here.

executed actions that keep it stationary.

In the other cases where no encouragement was offered (runs d0034 and d0056) the robot did receive some reward albeit not a maximum reward. In these cases the robot did have actions from recent behaviour to choose from, however, the behaviour did not become repeated over the long term as continual merging and purging of experiences that do not result in near maximal reward resulted in only transitory behaviour. Thus the modification of the space through merging and deletion plays an important role.

D. Emergent Classes of Experience

Analysis of the results shows that there was an extensive reduction in the number of experiences in the metric space through forgetting and merging, usually reducing the number of experiences by between 40% and 90%. Between 5 and 20% of experiences were merged, the others were deleted ("forgotten").

Examining a typical example; run d0033, a successful peekaboo run, merged 15 experiences out of a total of 181 experiences and deleted 63. One experience that was merged with many later ones was experience number 1 (the second experience). That experience was merged with 8 other experiences and was associated with action 6 (HID - the "hiding" action). Often when the HID action was chosen, it was experience number 1 which was found to be similar to the current experience. Thus it is possible to say that a class of experiences was emerging during this run that "represented" to the robot that it should next execute the peekaboo "hiding" action.

This results in a developed history that has become adapted to the interaction and focused around rewarded experience.

IV. Related Work

The concept of an agent learning from its past experience is one also used by the Case-Based Reasoning (CBR) approach [14]. Extension to the continuous domain [15] and combination with a Reinforcement Learning approach, however, brings the approach much closer to our IHA. However, in our approach, the use of an information theoretic metric measure to compare past experience with present experience can potentially uncover different and more interesting relationships in the history of experience as well as offering an ordered list of near experiences to choose from. Furthermore, the application to the social domain is unique and challenging.

Our approach is also related to reinforcement learning [16], particularly those examples that use intrinsic motivation *e.g.* [17] [18] and memory-based approaches *e.g.* [19] [20] [21]. In contrast to traditional reinforcement learning, the Interaction History Architecture approach uses temporally extended experience rather than the instantaneous values of the sensorimotor and internal variables (*state*). This distinction is important as, particularly where there is an interaction partner or other agents, the environment cannot be modelled as a simple Markov Decision Process.

[22] also studies the acquisition of a peekaboo-style communicative ability although in a virtual agent. The human caregiver hides the face instead of the robot while also saying "peek-a-boo" as reassurance and surprise. The model matches simplified state (internal emotion state, face sensor and reward) to predict when to expect a reward. Our work thus differs from

this in many important ways, the most significant being the generality of our approach, using complex sensor stream and episodes of experience, and the potential to develop and adapt action capabilities over ontogeny.

V. Future Work

While short term behaviour acquisition is illustrated here, future research work should look at how behaviour can be altered over the long term in response to changing encouragement and reward by the interaction partner. Furthermore, showing how different behavioural responses can be developed for different experiences would be important next step.

Further experiments should also utilize interaction partners that do not have prior knowledge regarding the operation of the robot and software.

VI. Conclusion

The Interaction History Architecture was implemented for the upper-body humanoid robot Kaspar2. The peekaboo interaction game was used to evaluate the architecture in terms of how the robot could use its own personal interaction history to develop the capability to engage in the game. Results, while limited, indicate that giving appropriate encouragement to the robot as it executes certain series and groups of behaviours can result in those behaviours being selected in preference to others in equivalent conditions. This result supports the hypothesis that encouraging the hiding action would result in a higher rate of peekaboo sequences than would be expected from random selection. Furthermore, encouraging alternative action sequences resulted in those actions being repeated, inviting the conclusion that this behaviour of the architecture is general and not limited to the peekaboo game. Additional support for the hypothesis was found in the conditions that offered no encouragement. In these cases no single action or sequence was selected in preference to any other, emphasizing the importance of the interaction of the environment with the robot in producing a history of interaction that can be used to develop action capabilities.

It was found that classes of experiences emerged through the process of merging of experiences as the interaction progressed. These classes of experience and their associated next-action can be said to be emergent, grounded "representations" that have "meaning" from the robot's own perspective in the actions they generate.

Acknowledgement

This work was conducted within the EU Integrated Project RobotCub ("Robotic Open-architecture Technology for Cognition, Understanding, and Behaviours"), funded by the EC through the E5 Unit (Cognition) of FP6-IST under Contract FP6-004370.

References

[1] N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, and R. te Boekhorst, "Grounded sensorimotor interaction histories in an information theoretic metric space for robot ontogeny," *Adaptive Behaviour*, vol. 15, no. 2, pp. 167–187, 2007.

[2] M. L. Anderson, "Embodied cognition: A field guide," *Artificial Intelligence*, vol. 149, pp. 91–130, 2003.

[3] G. Lakoff and M. Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York, 1999.

[4] A. M. Glenberg, "What is memory for?" *Behavioral and Brain Sciences*, vol. 20, no. 1, March 1997.

[5] C. L. Nehaniv, D. Polani, K. Dautenhahn, R. te Boekhorst, and L. Cañamero, "Meaningful information, sensor evolution, and the temporal horizon of embodied organisms," in *Artificial Life VIII*. MIT Press, 2002, pp. 345–349.

[6] N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, and R. te Boekhorst, "Anticipating future experience using grounded sensorimotor informational relationships," in *Artificial Life XI 11th International Conference on the Simulation and Synthesis of Living Systems*. Winchester, UK: University of Southampton, August 2008, in press.

[7] L. Vygotsky, *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press., 1978.

[8] J. Crutchfield, "Information and its metric," in *Nonlinear Structures in Physical Systems - Pattern Formation, Chaos and Waves*, L. Lam and H. Morris, Eds. New York: Springer-Verlag, 1990, pp. 119–130.

[9] A. Meltzoff and M. Moore, "Explaining facial imitation: a theoretical model," *Early Development and Parenting*, vol. 6, pp. 179–192, 1997.

[10] C. L. Nehaniv, "Sensorimotor experience and its metrics," in *Proc. 2005 IEEE Congress on Evolutionary Computation*, vol. 1. Edinburgh, Scotland: IEEE Press, 2-5 Sept. 2005, pp. 142–149.

[11] P. Rochat, J. G. Querido, and T. Striano, "Emerging sensitivity to the timing and structure of protoconversation in early infancy," *Developmental Psychology*, vol. 35, no. 4, pp. 950–957, 1999.

[12] OpenCV, "Open computer vision library (gpl licence)," http://sourceforge.net/projects/opencvlibrary/, 2000.

[13] G. Metta, P. Fitzpatrick, and L. Natale, "YARP: Yet Another Robot Platform," *International Journal of Advanced Robotic Systems*, vol. 3, no. 1, pp. 43–48, 2006.

[14] J. Kolodner, *Case-based Reasoning*. Morgan Kaufman, 1993.

[15] A. Ram and J. C. Santamaria, "Continuous case-based reasoning," *Artificial Intelligence*, vol. 90, no. 1-2, pp. 25–77, 1997.

[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, March 1998.

[17] A. G. Barto and Ö. Şimşek, "Intrinsic motivation for reinforcement learning systems," in *Proceedings of the Thirteenth Yale Workshop on Adaptive and Learning Systems*, 2005, pp. 113–118.

[18] A. Bonarini, A. Lazaric, M. Restelli, and P. Vitali, "Self-development framework for reinforcement learning agents," in *Proceedings the of 5th International Conference on Development and Learning (ICDL 2006)*, 2006.

[19] L.-J. Lin and T. Mitchell, "Reinforcement learning with hidden states," in *From animals to animats 2: Proceedings of the second international conference on simulation of adaptive behavior*, J.-A. Meyer, H. L. Roitblat, and S. W. Wilson, Eds. MIT Press, Cambridge, MA, 1992, pp. 271–280.

[20] B. Bakker, "Reinforcement learning with long short-term memory," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, Cambridge, MA, 2002.

[21] R. McCallum, "Hidden state and reinforcement learning with instance-based state identification," *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, vol. 26, no. 3, pp. 464–473, 1996.

[22] M. Ogino, T. Ooide, A. Watanabe, and M. Asada, "Acquiring peekaboo communication: Early communication model based on reward prediction," in *Proceedings of the 6th IEEE International Conference on Developoment and Learning*. IEEE CDROM, 2007.