An Infomax Controller for Real Time Detection of Social Contingency

Javier R. Movellan
University of California San Diego, & ATR, Kyoto
movellan@mplab.ucsd.edu

***Abstract*— We present a model of behavior according to which organisms react to the environment in a manner that maximizes the information gained about events of interest. We call the approach "Infomax control" for it combines the theory of optimal control with information maximization models of perception. The approach is reactive, not cognitive, in that it is better described as a continuous "dance" of actions and reactions with the world, rather than a turn-taking inferential process like chess-playing. The approach however is intelligent in that it produces behaviors that optimize long-term information gain. We illustrate how Infomax control can be used to understand the detection of social contingency in 10 month old infants. The results suggest that, while lacking language, by this age infants actively "ask questions" to the environment, i.e., schedule their actions in a manner that maximizes the expected information return. A real time Infomax controller was implemented on a humanoid robot to detect people using contingency information. The system worked robustly requiring little bandwidth and computational cost.**

***Index Terms*— Information Maximization, Contingency Detection, Control Theory.**

I. Introduction

John Watson proposed that infants use contingency information to define and recognize human beings [7, 8]. In 1986 [3, 6] Watson and the author of this document conducted an experiment to test how 10 month old infants use contingency information to detect novel social agents. Infants were seated in front of a robot that did not look particularly human. In the experimental group the robot was programmed to respond to the environment in a manner that simulated the contingency properties of human beings. Each infant in the control group was matched to an infant in the experimental group and was presented the same temporal distribution of lights, sounds and turns of the central robot as was experienced by his/her matched participant. However, in the control group the robot was not responsive to the infant's behavior or to any other events in the room.

A. Forty Three Seconds of an Infant's Day

In that study we found evidence that the infants in the experimental group treated the robot as if it were a social agent: For example, they exhibited 5 times more vocalizations than infants in the control group. Moreover they followed the "line of regard" of the robot when it rotated, showing some evidence for shared attention [3]. Most interesting was the fact that some infants appeared to actively "decide" in a few trials, and a matter of seconds, whether or not the robot was responsive.

Fig. 1. *Left: Schematic of the robot head used in [6]. Right: Baby-9. The image of the robot is seen reflected on a mirror positioned behind the baby.*

Particularly telling were the first 43 seconds of the experiment with one of the infants in the experimental group. We will refer to him as Baby-9 (see right side of Figure 1). He was 10 months old on 7/14/1986, when the study was run at UC Berkeley's Institute for Human Development. Most people that see the video of the experiment agree that by the third or fourth vocalization (25 seconds into the experiment) baby-9 has detected the fact that the robot was responsive to him. Most importantly, most people inescapably feel that the infant is actively querying the robot as if asking whether or not it is responding to him. This brings some interesting questions:

1) What does it mean to "ask questions" for an organism that does not have a language?
2) Why did Baby-9 schedule his vocalizations in the way he did? Why did he not vocalize, for example, at a much rapid or a much slower rate?
3) Was it reasonable for Baby-9 to decide within 3 to four responses and 20-30 seconds into the experiment that the robot was responsive? Why not more or less time and responses?

We will approach these questions by developing a formal model of the problem faced by Baby-9 and showing that his behavior was indeed optimal in terms of the expected information return about the responsiveness or lack of responsiveness of the robot.

II. Detecting Social Contingency: Causal Model

Our goal is to gain a better understanding of the problem of active contingency detection in simple social interactions between infants and caregivers. These are characterized by the existence of self-feedback (e.g., infants can hear themselves), significant delays and uncertainty in the caregiver's responses, and significant levels of background activity. We will investigate the problem from the point of view of a bare-bones *"social robot"* endowed with a single binary sensor (e.g., a sound detector) and a single binary actuator. There will be two players: (1) A *social agent*, which plays the role of the caregiver, and (2) A *robot*, which plays the role of the infant. Agent and robot are in an environment which may have random background activity. The role of the robot is to discover as soon as possible and as accurately as possible the presence of *responsive social agents*.

We will develop a discrete-time model of the problem. The parameter $\Delta t \in \Re$ will represent the sampling period, i.e., the time between time steps, in seconds. The activity of the robot's actuator is represented by the binary random process $\{U_t\}$. The variable U_t takes value 1 when the robot's actuator is active at time t, and zero otherwise. The presence or absence of responsive social agents is indicated by the random variable H. We refer to $\{H = 0\}$, the absence of a responsive agent, as the *"null hypothesis"*, and $\{H = 1\}$, the presence of a responsive agent, as the *"alternative hypothesis"*. The parameter π represents the prior probability of the alternative hypothesis, i.e., the robot's initial belief about the presence of a social agent, prior to the gathering of sensory information.

A. Modeling the Social Agent

We let the behavior of the social agent depend on two auxiliary processes: A timer $\{Z_t\}$ and an indicator $\{I_t\}$. The timer takes values in $\{0, \cdots, \tau_2^a\}$ where $\tau_2^a \in \mathcal{N}$ is a parameter of the model, whose meaning will be explained below. The timer keeps track, up to τ_2^a, of the number of time steps since the last robot action, i.e.,

$$Z_1 = \tau_2^a + 1 \tag{1}$$

$$Z_t = h(Z_{t-1}, U_t) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } U_t = 1 \\ Z_{t-1} & \text{if } U_t = 0 \text{ and } Z_{t-1} = \tau_2^a + 1 \\ 1 + Z_{t-1} & \text{else} \end{cases} \tag{2}$$

for $t = 2, 3, \cdots$ The indicator vector $I_t = (I_{1,t}, I_{2,t}, I_{3,t})^T$ consists of three binary variables that indicate whether or not time t belongs to the following categories: (1) *"Self Period"*, indicated by $I_{1,t}$; (2)*"Agent Period"*, indicated by $I_{2,t}$, and (3) *"Background Period"*, indicated by $I_{3,t}$. The meaning of these three periods is explained below.

The reaction times of social agents is bounded by the parameters $0 \leq \tau_1^a \leq \tau_2^a$, i.e., it takes agents anything from τ_1^a to τ_2^a time steps to respond to an action from the robot. "Agent periods", which are designated by the indicator process $\{I_{2,t}\}$ are periods of time for which responses of agents to previous robot actions are possible if an agent were to be present. Thus,

$$I_{2,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^a, \tau_2^a] \\ 0 & \text{else} \end{cases} \tag{3}$$

During agent periods, the robot's sensor is driven by the Poisson process $\{D_{2,t}\}$ which has rate R_2. The distribution of R_2 depends on whether or not a responsive agent is present in a manner that will be specified below.

B. Modeling Self-Feedback and Background Processes

We allow for the robot sensor to respond to the robot actuators, e.g., the robot can hear its own vocalizations, and allow for delays and uncertainty in this self-feedback loop. In particular we bound self-feedback delays to occur within the interval $[\tau_1^s, \tau_2^s]$, where $\tau_1^a > \tau_2^s$. The indicator variable for self-feedback period is thus defined as follows:

$$I_{1,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^s, \tau_2^s] \\ 0 & \text{else} \end{cases} \tag{4}$$

During Self periods, the activation of the sensor is driven by the Poisson process $\{D_{1,t}\}$ with rate R_1.

With regard to the background process, we model it as a Poisson process $\{D_{3,t}\}$ with rate R_3. The background process is responsible for driving the sensor's activity that is not due to self-feedback and is not due to social agent responses to the robot's behaviors. Note background activity can include, among other things, the actions from external social agents who are not responding to the robot (e.g., two social agents may be talking to each other thus activating the robot's sound sensor). We endow the background rate R_3 with an uninformative prior Beta distribution to reflect the fact that the background activity may change dramatically from situation to situation in ways that are not known to the robot. The background indicator keeps track of periods for which self-feedback or responsive actions from a social agent may not happen, i.e.,

$$I_{3,t} = (1 - I_{1,t})(1 - I_{2,t}) \tag{5}$$

C. Modeling the Robot's Sensor

The activity of the sensor is a switched Poisson process: during self-feedback periods it is driven by the Poisson process $\{D_{1,t}\}$, during agent periods it is driven by $\{D_{2,t}\}$ and during background periods it is driven by $\{D_{3,t}\}$, i.e.,

$$Y_t = I_t \cdot D_t = \sum_{i=1}^{3} I_{i,t}\, D_{i,t} \tag{6}$$

We still need to specify the distribution of the response rate R_2 during agent periods. If an agent is present, i.e., $H = 1$, we let R_2 be independent of R_1 and R_3 and endow it with an uninformative Beta prior distribution. This reflects the fact that different agents respond at different rates in ways that the robot does not know apriori. If an agent is not present, i.e., $H = 0$, then the response rate during agent periods is not different from the response rate during background periods, i.e., $R_2 = R_3$.

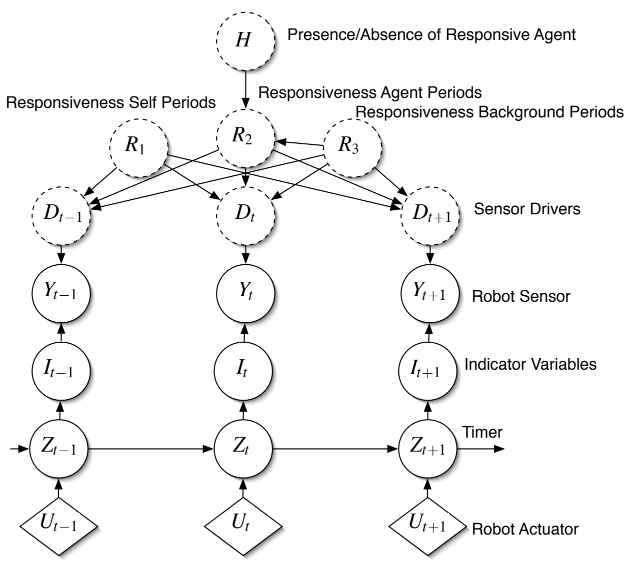

Fig. 2. *Graphical Representation of the Causal Model. Arrows represent dependency relationships between variables. Dotted figures indicate unobservable variables, continuous figures indicate observable variables. Diamonds indicate control variables.*

D. Auxiliary Processes and Constraints

The processes $\{O_t, Q_t\}$ register the sensor activity and lack-thereof up to time t during self, agent, and background periods. In particular for $t = 1, 2, \cdots$

$$O_{i,t} = \sum_{s=1}^{t} I_{i,t} Y_t, \text{ for } i = 1, 2, 3 \tag{7}$$

$$Q_{i,t} = \sum_{s=1}^{t} I_{i,t}(1 - Y_t), \text{ for } i = 1, 2, 3 \tag{8}$$

Figure 2 display Markovian constraints in the joint distribution of the different variables involved in the model. An arrow from variable X to variable Y indicates that X is a “parent” of Y. The probability of a random variable is conditionally independent of all the other variables given the parent variables.

III. Development and Learning. Inference, and Control

We refer to *“development”* as the problem of discovering the causal structures underlying social interaction, i.e., discovering a model of the kind displayed in Figure 2. This is a difficult problem that may require large amounts of data gathered over months or years. We refer to *”learning”* as the problem of discovering contingencies, i.e., making inferences about unobservable variables of a given model. This is a process that in general requires less data than model development and may occur within seconds, minutes or hours.

Development and learning rely on two basic processes: inference and control. Inference refers to the problem of combining prior information with sensor data in a principled manner. Control refers to the problem of scheduling the behavior in real time to achieve the goals of the organism.

In this document we will assume the robot has already developed a causal model and focus on the learning problem, i.e, how to decide about the presence or absence of a social agent based and how to use the actuator to make such decisions as fast and as accurately as possible.

A. Learning: Inference

Let $(y_{1:t}, u_{1:t}, o_t, q_t, z_t)$ be an arbitrary sample from $(Y_{1:t}, U_{1:t}, O_t, Q_t, Z_t)$. It can be shown that the log-likelihood ratio between the two hypotheses is as follows:

$$\log \frac{p(y_{1:t} \mid u_{1:t}, H = 1)}{p(y_{1:t} \mid u_{1:t}, H = 0)} \tag{9}$$

$$= \log \frac{\Gamma(2 + o_{2,t} + o_{3,t} + q_{2,t} + q_{3,t})}{\Gamma(1 + o_{2,t} + o_{3,t})\,\Gamma(1 + q_{2,t} + o_{3,t})} \tag{10}$$

$$+ \sum_{i=2}^{3} \log \frac{\Gamma(1 + o_{i,t})\,\Gamma(1 + q_{i,t})}{\Gamma(2 + o_{i,t} + q_{i,t})} \stackrel{\text{def}}{=} f(o_{2,t}, o_{3,t}, q_{2,t}, q_{3,t}) \tag{11}$$

The posterior distribution about the hypothesis of interest is as follows:

$$p(H = 1 \mid y_{1:t}, u_{1:t}) \tag{12}$$

$$= \text{logistic}\left(\log \frac{\pi}{1 - \pi} + f(o_{2,t}, o_{3,t}, q_{2,t}, q_{3,t})\right) \tag{13}$$

This posterior distribution, contains all the information available to the robot about the presence of a responsive agent. It has two important properties: (1) It does not depend on $o_{1,t}, q_{1,t}$, i.e., the self-periods are uninformative about the hypothesis, and (2) If $o_{1,t} + q_{1,t} = 0$ or $o_{2,t} + q_{2,t} = 0$ the log-likelihood ratio is 0. In other words, if no data has been gathered in either the agent or the background condition then we have gained no information about H. Thus in order to gain information about H the robot must use its actuator at least once and not use it at least once.

B. Learning: Infomax Control

In this section we focus on how to schedule the behavior of the robot’s sensor in real time in order to maximize the information received about the presence or absence of social agents. Let t represent the current time and suppose by time

t we have observed $y_{1:t}, u_{1:t}$. For a future time $s > t$ let us consider the mutual information between the observable variables and the hypothesis of interest H

$$\mathrm{I}(H, Y_{t+1,s}, U_{t+1:s} \mid y_{1:t}, u_{1:t}) \quad (14)$$
$$= \mathcal{H}(H \mid y_{1:t}, u_{1:t}) - \mathcal{H}(H \mid Y_{t+1:s}, U_{t+1:s}, y_{1:t}, u_{1:t}) \quad (15)$$

where $\mathcal{H}$ stands for entropy. The equation tells us that the information about H provided by the observable processes $Y_{t+1:s}, U_{t+1:s}$ equals the reduction of uncertainty about H provided by those observables. Since the term $\mathcal{H}(H \mid y_{1:t}, u_{1:t})$ does not depend on future actions then maximizing the information return provided by future actions is equivalent to minimizing the future entropy of H. We will let the controller's objective W_s be the negative entropy of H, i.e., the controller will choose actions that are expected to minimize the uncertainty about H

$$W_s \stackrel{\text{def}}{=} -\mathcal{H}(H \mid Y_{t+1:s}, U_{t+1:s}, y_{1:t}, u_{1:t}) \quad (16)$$

An Infomax controller is an open-loop controller that maximizes the expected value of W at future times $t+1, \cdots, T$.

The causal model we are working with belongs to the family of partially observable Markov processes. Finding optimal controllers for these processes is in general difficult. In this case however the problem simplifies because it is possible to find a recursive statistic that summarizes the observable sequences without any loss of information about H. In particular it is possible to show that the optimal controller satisfies the following form of Bellman's optimality equation [2]:

$$C(y_{1:t}, u_{1:t}) = C'(s_t) \stackrel{\text{def}}{=} \operatorname*{argmax}_{u_{t+1}} N_t(s_t, u_{t+1}) + F_t(s_t, u_{t+1}) \quad (17)$$

$$V_t(y_{1:t}, u_{1:t}) = V'(s_t) \stackrel{\text{def}}{=} \max_{u_{t+1}} N_t(s_t, u_{t+1}) + F_t(s_t, u_{t+1}) \quad (18)$$

where $S_t \stackrel{\text{def}}{=} (Y_t, O_t, Q_t, Z_t)$ carries all the information about W_t in the observed sequences $Y_{1:t}, u_{1:t}$. The term C' is a controller that makes decisions based on s_t, V' is the value of s_t, and

$$N_t(s_t, u_{t+1}) \stackrel{\text{def}}{=} \underbrace{\mathrm{E}\left[W_{t+1} \mid s_t, u_{t+1}\right]}_{\text{Next Step Expected Return}} \quad (19)$$

$$F_t(s_t, u_{t+1}) \stackrel{\text{def}}{=} \underbrace{\mathrm{E}[V_{t+1}(s_t, Y_{t+1}, u_{t+1}) \mid s_t, , u_{t+1}]}_{\text{Future Expected Return}} \quad (20)$$

This equation can be solved using dynamic programming techniques [1, 2].

IV. Analysis of the the optimal controller

The dynamic programming problem was solved on a cluster of 24 2.5Ghz PowerPC G5 CPUs. The computation time was in the order of 1 hour. The parameters of the model were set as follows: $T = 40, \tau_1^s = 0; \tau_2^s = 0; \tau_1^a = 1; \tau_2^a = 3; \pi = 0.5$. We then used logistic regression to model the behavior of the controller for times $15 < t < 25$, since these are times which are not too close to the beginning and end of the controller's window of interest, i.e., $t \in [1, 40]$. We did not expect for logistic regression to provide a perfect prediction since in some cases the value function is equal for both actions and in such occasions the optimal controller may arbitrarily chooses one action over the other. Surprisingly logistic regression approximated the optimal controller with 96.46 % accuracy over all possible conditions. This model prescribed to respond if and only if $I_{3,t} = 1$, i.e., we are in a background period, and in addition

$$\frac{\mathrm{Var}(R_3 \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{3,t} + q_{3,t} + 3} > 9\, \frac{\mathrm{Var}(R_2 \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{2,t} + q_{2,t} + 3} \quad (21)$$

Interpretation: Greedy one-step controllers [4, 5] that ignore the future expected return would fail on this task. The reason is that when making a response the next time steps are occupied by self-feedback, that happens to be uninformative, thus a greedy controller ends up deciding to never act. Including future expected return allows the controller to implicitly look ahead and see that in the long run making an action can provide a better information return than being inactive. The statistic

$$\frac{\mathrm{Var}(R_i \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{i,t} + q_{i,t} + 3} \quad (22)$$

is used by the controller to decide when to act. This statistic captures the expected reduction in the uncertainty about R_i provided by a new observation from the period under which R_i actively drives the sensor: a self-feedback period for R_1, an agent period for R_2 and a background period for R_3. The optimal controller thus appears to want to keep the uncertainty about R_3 and R_2 within a fixed ratio. Actions are more costly, in terms of information return, than lack of action. If the robot acts at time t it gains no information during the times $[t+\tau_1^s, t+\tau_2^s]$ since self-feedback observations are not informative about H. Moreover during times $[t+\tau_1^a, t+\tau_2^a]$ the robot cannot act and thus can only get information about R_2, not R_3. This may help explain why uncertainty about the agent activity rate R_2 needs to be 9 times larger than the uncertainty about the background activity rate, R_3, before an action occurs.

V. Understanding 43 Seconds of an Infant's Day

Here we examine whether the Infomax controller can provide a qualitative understanding of the first 43 seconds of the experimental session with Baby-9, as described in Section I-A. During this time Baby-9 produced 7 vocalizations, which occurred at the following times in seconds from the start of the experiment: $\{5.58, 9.44, 20.12, 25.56, 32.1, 37.9, 41.7\}$.

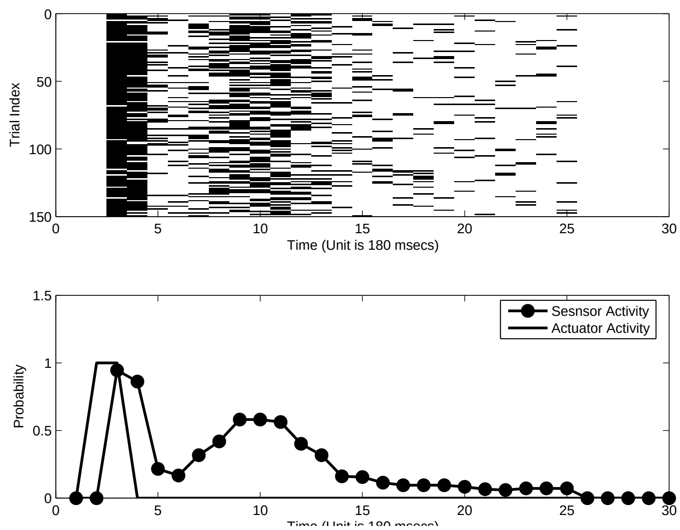

Fig. 3. *Above: Raster plot of 150 trials. On each trial an animated character made a sound and subjects were asked to talk back to the character and let it know that they were listening. Dark indicates that the audio sensor was active. Below: Probability of the audio sensor being active as a function of time. The probabilities is estimated by averaging across the 150 trials in the figure above.*

Each of these vocalizations were followed by a combination of sounds and lights from the robot. The intervals, in milliseconds, between the beginning of two consecutive infant vocalizations were as follows: $\{4.22, 10.32, 5.32, 6.14, 5.44, 3.56\}$. Most people agree that by the third or 4th vocalization the infant knows that there is a responsive agent in the room.

The Infomax controller presented in Section III-B requires setting five parameters: The sampling period for the time discretization, the self-delay parameters, and the agent delay parameters. To get rough estimates for the agent latency parameters τ_1^a, τ_2^a, we asked 4 people, unaware of the purpose of the study, to talk to a computer animated character. The ages of the 4 participants were $4, 6, 24$ and 35 years. We used an optimal encoder to binarize the activity of an auditory sensor and plotted the probability of activation of this binary sensor as a function of time over 150 trials. Each trial started with a vocalization of the animated character and ended 4 seconds later. The results are displayed on Figure 3. The top graph in the figure shows the activity of the acoustic sensor as a function of time from the beginning of the character's vocalization over 150 trials. Each horizontal line is a different trial. The first vertical bar is due to self-feedback from the character. By about 1200 to 1440 msec after the end of the vocalization from the animated character there is another peak of activity in the sensor, which is now caused by the vocalizations of the human participants. The lower graph of the Figure shows the probability of sensor activity as a function of time collapsed across trials. Note the first peak in activity due to self-feedback, and the gradual raise and fall in sensor activity due to the human response.

Based on this graph we run a simulation of the optimal controller with the following parameters: $\Delta t = 800$ msec, $\tau_1^s = \tau_2^s = 0$, $\tau_1^a = 1; \tau_2^a = 3$. In other words, we let self-delay to be negligible with respect to the expected delays in human responses, and we bracket the human activity to occur within 800 to 2400 milliseconds. Note these parameter values were chosen to reflect the time delays and the levels of uncertainty of social interactions, not to fit Baby-9's data. We let $\pi = 0.01$ to simulate a worst case scenario, thus requiring more data to decide that there is a responsive system. Figure 4 shows the results of the simulation. The horizontal axis in all the graphs is time, measured in seconds. The top graph shows the vocalizations of the optimal controller, which now plays the role of Baby-9. The controller produced 6 vocalizations over a period of 43 seconds. The average interval between vocalizations was 5.92 seconds, compared to 5.833 secs for Baby-9. The difference is not significant using a standard T-test $(T(9) = 0.08,\ p = 0.94)$.

The second graph from the top of Figure 4 shows the system's belief's about the presence of a responsive agent. By the fourth response, thirty seconds into the experiment, this probability passes the 0.5 level. The third graph shows the posterior probability distributions about the the agent and background response rates by the end of the 43 second period. The last graph shows the ratio between the uncertainty about the sensor rate during agent periods and the rate during background periods. Note when this ratio reaches the value of 9, the simulated baby makes a response.

The model thus shows that Baby-9 scheduled his responses and made decisions about the responsiveness of social agents in an optimal manner, given the statistics of times delays and levels of uncertainty typically found in social interactions. The model also is consistent with the idea that Baby-9 was "asking questions" to the robot, in the sense that his vocalizations were scheduled in a manner that maximized the information returned about the responsiveness of the robot. Another point of interest is that the optimal controller exhibits turn-taking, i.e., after an action is produced the controller waits for a period of time, an average of 5.92 seconds, before vocalizing again. The period between vocalizations is not fixed and depends on the relative uncertainty about the levels of responsiveness of the agent and the background.

VI. Real Time Robot Implementation

We implemented the optimal Infomax controller developed above on RobovieM, a humanoid robot developed at ATR's Intelligent Robotics laboratory. While the robot was not strictly necessary to test the real time controller, it greatly helped improve the quality of the interactions developed between humans and machine thus providing a more realistic way for testing the controller. The current version of the Infomax controller requires a 1 bit sensor and a 1 bit actuator.

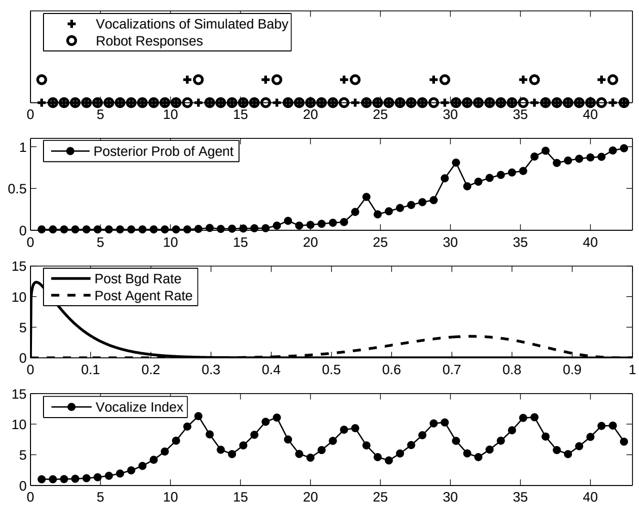

Fig. 4. *The horizontal axis represents time in seconds. From top to bottom: (1) Responses of the Infomax controller (which simulates a baby). Note the environment was silent and the controller was responded to every time it vocalized. (2) Posterior probability for the presence of a responsive agent as a function of time. (3) Posterior distribution for the agent and background rates after 43 seconds. (4) Ratio of the uncertainty about the agent's response rate vs the uncertainty about the background's response rate.*

For sensor we chose to average acoustic energy over 500 msec windows and discretized it using a 1 bit optimal coder. The actuator was a small loudspeaker producing a 200 msec robotic sound. Lacking quantitative evaluations we will simply state that the controller works remarkably well. In standard office environments, with relatively high levels of noise, the controller decides in a few trials whether or not a responsive agent is present. Particularly effective are transition points in which agents switch from talking to the robot to talking to somebody else. We have demonstrated the system at several scientific talks, and conferences with very good results even in relatively noisy conditions like poster rooms.

VII. Conclusions

We introduced the idea of Infomax control as a self-supervised form of motor control. No external reinforcer is required. Instead, Infomax controllers modify their internal states to better explain the available data and produce actions that are expected to provide highly informative data. Infomax control does not fit well the mold of standard reinforcement learning approaches. Classical and instrumental learning models emphasize the role of external stimulus as reinforcers of behaviors. In Infomax control however, stimuli and responses do not have intrinsic reinforcement value. Instead what determines behavior is the expected information return about hypotheses of interest.

In this paper we used the ideas of Infomax control to understand the detection of social contingency in 10 month old infants. Interestingly when the controller makes a response it follows it by a period of silence, as if waiting for the outcome of a question. This “turn-taking” behavior was not built onto the system. Instead it emerged from the requirement to maximize information gain given the time delays and levels of uncertainty typical of social interactions. The results suggest that, in spite of lacking language, some infants actively ask questions to humans and to other aspects of the environment, scheduling their actions in a manner that maximizes the expected information return. The approach proposed here works well in practice when applied in robots that need to operate in real time in everyday life situations.

Acknowledgment

This research was sponsored by a UC Discovery Grant, and a grant from National Association for Autism Research. The optimal controller was computed using a Cluster funded by NSF grant IIS0223052. The real time robot implementation of the model was conducted by the author at ATR's Intelligent Robotics and Communication Laboratory, as part of an “Innovational Interaction Media,” project with funding from the National Institute of Information and Communications Technology (NICT) of Japan.

References

[1] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, New Jersey, 1957.

[2] J. R. Movellan. Optimal control of dynamic Bayes nets. Technical report, Kolmogorov Tutorials, UCSD, 2005.

[3] J. R. Movellan and J. S. Watson. The development of gaze following as a Bayesian systems identification problem. In *Proceedings of the International Conference on Development and Learning (ICDL02)*. IEEE, 2002.

[4] J. D. Nelson and J. R. Movellan. Active inference in concept induction. In T. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, number 13, pages 45–51. MIT Press, Cambridge, Massachusetts, 2001.

[5] J. D. Nelson, J. B. Tenenbaum, and J. R. Movellan. Active inference in concept learning. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 692–697. LEA, Edinburgh, Scotland, 2001.

[6] Movellan J. R. and J. S. Watson. Perception of directional attention. In *Infant Behavior and Development: Abstracts of the 6th International Conference on Infant Studies*, NJ, 1987. Ablex.

[7] J. S. Watson. Smiling, cooing and ”the game”. *Merrill-Palmer Quarterly*, 18:323–339, 1972.

[8] J. S. Watson. The perception of contingency as a determinant of social responsiveness. In E. B. Thoman, editor, *Origins of the Infant's Social Responsiveness*, pages 33–64. LEA, New York, 1979.