Introduction: The Third International Workshop on Epigenetic Robotics

Luc Berthouze
Neuroscience Research Institute
Tsukuba AIST Central 2, Umezono 1-1-1
Tsukuba 305-8568 Japan
Luc.Berthouze@aist.go.jp

Christopher G. Prince
Department of Computer Science
University of Minnesota Duluth
Duluth, MN 55812 USA
chris@cprince.com

1. Introduction

This paper summarizes the paper and poster contributions to the Third International Workshop on Epigenetic Robotics. The focus of this workshop is on the cross-disciplinary interaction of developmental psychology and robotics. Namely, the general goal in this area is to create robotic models of the psychological development of various behaviors. The term "epigenetic" is used in much the same sense as the term "developmental" and while we could call our topic "developmental robotics", developmental robotics can be seen as having a broader interdisciplinary emphasis. Our focus in this workshop is on the interaction of developmental psychology and robotics and we use the phrase "epigenetic robotics" to capture this focus.

2. Invited papers

In a stimulating contribution, **György (George) Gergely** addresses his empirical research with infants and also ties the work into epigenetic robotics. Gergely and his colleagues have found that infants interpret their social world in terms of a *principle of rational action*. In particular, at about one year of age, infants interpret the actions of others (e.g., caregivers) in terms of the action's function to bring about goal states. Infants expect actions to be realized in efficient ways and also understand that actions may be constrained by the situation. In one fascinating study (Gergely et al., 2002), this research group has found that 14-month-old infants will use their rational interpretations to decide how to imitate an actor. When an actor presses a light switch with her head, the infants tend to use their head *less* to imitate the action when the actor had her hands occupied (wrapped in a blanket). When the actors hands were unencumbered, infants more frequently imitated by using the same head-press behavior. Gergely concludes that infants reason about an actor's behavior. The infants' reasoning is apparently analogous to: "If the actor used her head to press the light switch and her hands were unencumbered, then there must have been a good reason for her to press the light switch with her head, so I should do the same." Gergely coherently argues that these findings may be of use to epigenetic robotic models of imitation. In general, imitation will only serve a valuable function when it is selectively focused. "Automatically imitating [i.e., echoing] every human action that one is perceptually exposed to is a ... pathological condition." More generally, the principle of rational action in humans is important as it is a "central inferential component of [a] mature theory of mind." Theory-of-mind refers to reasoning about the mental states of self and other (e.g., Wellman, 1990). Hence, not only is this work important to robotic models of imitation, but also to models of theory-of-mind development (e.g., Scassellati, 2002). This research may also supply us with theories for the mechanisms on which rational action interpretation is based, and help us on the developmental path of building robots that model higher-level human cognitive functions.

Rod Grupen deals with the structure of learning and development in synthetic systems. Taking inspiration from Piaget's early stages, he forwards two hypotheses: (a) rich classes of behavior can be expressed in terms of relatively few, schematic structures that are grounded in physical activity called "physical schemata"; (b) physical schemata are the sensorimotor foundation for re-usable control knowledge. To defeat the curse of dimensionality, he suggests application of appropriately parameterized schema instead of learning from scratch. This parameterization or adaptation of a schema to specific run-time contexts can be done by exploiting *symmetries* in behavioral descriptions. Of interest is Grupen's integrated view of the role of reflexes in the developmental process. He contends that reflexes serve as an epigenetic computational basis and that some are short-lived and serve an ontogenetic

knowledge formation role [1]. He proposes a generative basis for reflexes that permits a huge variety of closed-loop stimulus-response options that can serve both epigenetic and ontogenetic developmental goals. This kind of native reflexive basis can participate in a knowledge structure whose growth is directed by temporally dependent systems of constraints and resulting control knowledge representations can later be generalized to novel circumstances. The author describes a computational framework that addresses both learning and development, and incorporates the above principles. The power of the approach is shown with the acquisition of a tracking schema using two primitives – saccade, foveate. Used in triangulation, this scheme can be used to build up a Localize schema for finding known targets. Finally, in conjunction with Reach and Grasp schema, a high-level Acquire schema can be developed. Grupen concludes that this framework is very suitable to support the interaction of a teleoperator with a learning robot by trying to generate internal representations of the teleoperated task expressed in terms of value functions native to the device.

Deb Roy presents research "aimed towards developing conversational robots that ground language in action and perception." One project from Roy's laboratory involves the robot *Ripley*. Ripley is designed as a platform to support exploration of questions of grounded language and language acquisition. Ripley is a sophisticated robot arm and gripper, equipped with cameras, proprioception, and microphones. Questions being addressed include the effects of changes of visual perspective, and grounding of verbs such as *touch* and *lift*. The sensory-motor system is used to maintain an online simulation or "mental model." For example, part of this mental model is *Objecter*, an "object permance" module which maintains the state of objects in the world. Using the camera inputs, a color-based technique is used to segment image foreground from background, generating a set of connected regions. These regions are fed at 15Hz to Objecter which uses this information to update its object representations. The mental model also simulates the dynamics of 3-D rigid objects based on Newtonian physics. The system provides a memory by using run length encoding to represent mental model history. An earlier project (Roy and Pentland, 2002), used a sensory-only robot to learn associations between shapes and speech audio. This model detected phoneme sequences that were highly predictive of shapes (and vice versa), and formed speech-object prototypes on this basis.

[1] Another way to describe these short-lived ontogenetic processes are as *ontogenetic adaptations*. Ontogenetic adaptations are not immature versions of adult morphology or behavior, but are rather specific developments that are solely useful for the young organism (e.g., see Oppenheim, 1981; Thelen and Smith, 1994).

3. Survey

In a survey of the field, **Max Lungarella and Giorgio Metta** define epigenetic robotics as a two-pronged approach to robotics, located at the intersection of developmental psychology and robotics. On the one hand, it employs robots to instantiate and investigate models originating from developmental psychology, the so-called synthetic neural modeling or synthetic methodology. On the other hand, it seeks to design better robotic systems – more autonomy, more adaptability, more sociability – by applying insights gained from studies in ontogenetic development. Since the field has been rapidly growing in recent years, the authors restrict their survey to studies conducting robotic experiments, i.e., studies situated in the real world because only the world crystallizes the really hard problems, and studies with clear intent to address hypotheses put forward in either developmental psychology or developmental neuroscience. They identify 6 major categories of contribution: social interaction, sensorimotor control, categorization, value system, developmental psychology, motor skill acquisition and morphological changes. Interestingly, a somewhat different distribution of focus has been observed in this year's contributions.

4. Infant models

Matthew Schlesinger and Patrick Casey further our understanding of how infants understand objects. The question in this research is regarding when infants understanding of object motion can legitimately be said to involve expectations, anticipations, and prediction. As adults, we use these terms (expectations, anticipations, prediction) freely in regards to objects to describe the behavior of adults. However, presumably these understandings undergo development, and there is some controversy as to whether six-month-old infants possess understandings that can be interpreted in these terms. The authors approach this topic with a "strictly bottom-up model of perceptual processing in infants" implemented as a connectionist model. The position embodied in the model is that infants tend to respond in terms of the perceptual salience of objects and events. They used this model to generate predictions of infant behavior. The authors then carried out a replication of Baillargeon (1986)'s "car study" with six-month-old infants. In this study, infants are familiarized with a car rolling down a ramp, and then receive test trials in which an 'impossible' event occurs – the car appears to roll through a solid object (in control trials, the car rolls in front of the solid object). The results are found to provide support for both a perceptually-oriented explanation and a more 'adult' knowledge-oriented explanation.

In their poster, **Ryuta Fukuda, Michael**

Spratling, Denis Mareschal, and Mark Johnson study the development of sequential information encoding in human infants. They propose a connectionist model consisting of 30 cortico-basal ganglionic loops (prefrontal cortex) and characterized by a (non-modifiable) random matrix of synaptic weights and competition between output units. Temporal information is transformed into a spatial pattern of neuronal activations using neurophysiologically plausible activation functions and circuitry. The authors tested this architecture against data from two experiments in infant psychology: (a) visual expectation (at 1 year of age) – infants show better anticipation in response to alternating patterns of presentations of toys; (b) sound sequence discrimination (in newborns) – infants show better discrimination of rising and falling tones at low-time resolution. Since the model successfully differentiated patterns without employing any learning mechanism, the authors conclude that this structure has an intrinsic ability to differentiate sequential information.

5. Perception and action

Artur Arsenio, Paul Fitzpatrick, Charles Kemp, and Giorgio Metta further the work presented last year by Metta and Fitzpatrick at EpiRob and focus on object segmentation. They argue that object segmentation is a key resource for development since it is a first step of a *developmental trajectory* towards a robust, well-adapted vision system. Achieving high-quality object segmentation, for example, is useful to train up high-level visual modules (e.g., object recognition) and enhance low-level vision. The authors describe an array of methods for segmenting objects, using a selection of active or interactive cues. In the first scenario, the robot arm probes an area, seeking to trigger object motion and then identifies the boundaries of an object through its motion. In the second scenario, the authors take the original point of view of a wearable system viewing its wearer's actions. The system takes an active role, by monitoring human actions, issuing requests, and using active sensing to detect grasped objects held up to view. Finally, the authors attempt to exploit *natural human showing behavior* such as finger tapping, arm waving, or object shaking. Preliminary results are presented showing that combining those various approaches is useful in learning to recognize objects. The authors intend to investigate the developmental mechanism allowing the combination of these hypothetical building blocks into complex behaviors.

Frédéric Kaplan and Pierre-Yves Oudeyer address motivational principles in visual development. The authors are interested "to find general principles to design robots capable to extend their sensori-motor competences during their lifetime." The authors experiment with a Sony AIBO robot, and have the goal to develop visual competencies from scratch, driven only by internal motivations. They specify three task-independent motivational variables: predictability, familiarity, and stability. Predictability is defined by the robot being able to predict the current sensory state based on the previous sensory-motor state. Familiarity is defined in terms of the frequency of specific sensory-motor transitions for a time interval. Stability is defined as the deviation of the sensory inputs from their average values in a time interval. Each motivational variable is associated with a reward function. For the stability variable, the reward received by the robot is proportional to the value of the variable, and for predictability and familiarity, the robot is rewarded when it experiences an increase in the particular variable relative to the last time step. "As [the robot] learns, sensory-motor trajectories that used to give rewards tend to be less" motivating for the system. When motivational variables have peaked (e.g., increases in the predictability of the situation no longer occur), then the system should attempt to find other learning situations.

Frank Pollick, Joshua Hale, and Phil McAleer are interested in the features that make the movements of a system – be it a humanoid robot, or a CG (computer graphic) character – seem natural to a human observer. They suggest that studying the perception of humanoid movement might reveal unique aspects of human visual perception and visual cognition. Specialized mechanisms seem to be devoted to the perception of movement as evidenced by the perception of two-frame apparent motion sequences. With a short inter-frame interval, the sequence appears meaningless. With a longer duration interval, however, a solid constraint is imposed and coherent movement is observed. The authors examined adult human similarity judgements of arm movements generated by 14 different control strategies (e.g., minimum velocity, minimum torque) in both a 30-DOF humanoid robot and a CG character (with data extracted from a motion capture system). Two types of movement were considered – fast motion and slow motion – and pairwise similarity comparisons of all elements of the set of movements was performed. MDS (multi-dimensional analysis) of the data show that while fast movements were primarily evaluated along a single dimension (velocity of the movement), slow movements were evaluated equally along two dimensions (path of the hand, and body posture). This would partially support a recent hypothesis that perception of human movement is not mediated through motion per se, but rather through the recognition of discrete postures.

In their poster, **Tian Lan, Michael Arnold,**

Terrence Sejnowski, and Marwan Jabri propose a biologically-inspired architecture to implement sequence learning. Learning sequences of primitive skills is an important step to realize more complex tasks in a skill-based machine design. To model the role of basal ganglia in sequential execution and learning, they use an actor-critic architecture with a temporal difference (TD) learning algorithm. They tested the model on a simplified – one-dimensional – version of the block copying task (BCT) and show that a 9-action sequence was successfully learned after 2000 trials.

In another poster, **Williams Paquier, Nicolas Do Huu, and Raja Chatila** present an architecture for developmental robotics. The core of this architecture is a new type of pulsed neural network. These networks are comprised of columns, which act as pipes for data processing. Columns receive input bursts in one layer, and release output bursts in another layer. Hypercolumns are sets of columns that share the same input. Columns compete and tend to adapt to particular patterns of input. The simulator *NeuSter* implements this new type of neural network. Experiments with robots are underway.

Finally, **Anil Seth, Jeffrey McKinstry, Gerald Edelman, and Jeffrey Krichmar** forward an approach to the binding problem with the Darwin VIII robot. The binding problem involves the question of how the diverse and functionally segregated visual regions of the brains of biological organisms "coordinate their activities in order to associate features belonging to individual objects and distinguish among different objects." The authors focus on a pre-attentive mechanism involving "the linkage of neuronal groups by selective synchronization." In Darwin VIII, reentrant connectivity is used to achieve synchronously active neuronal circuits. Darwin VIII simulates a nervous system comprising 28 neuronal areas, 53450 neuronal units, and approximately 1.7 million synaptic connections. Resulting behavior in the robot includes conditioning to prefer one visual target over multiple distractors via an innately preferred auditory cue.

6. Acquisition of representations

Georgi Stojanov and Andrea Kulakov address the issue of representations by casting it in the frame of *interactivism* (Bickhard, 1999). To quote Bickhard, *interactivism is a complex philosophical and theoretical system ... [and] involves a commitment to a strict naturalism – a regulative assumption that reality is integrated: that there are no isolatable and independent grounds of reality, such as would be the case if the world were made of Cartesian substances.* Representations emerge out of agent-environment interaction. Within interactivism, an agent is regarded as an action system: an autonomous, self-organizing and self-maintaining entity, which can exercise actions and sense their effects in the environment it inhabits. The authors argue that this view is especially suited for treatment of the problem of representation in epigenetic agents since (a) representations cannot be treated in isolation, (b) they should be grounded and (c) we should account for their emergence: they are either learned in the development of the agent, or inherited (as a result of evolution). The authors attempt to draw the lines of a generic interactivist architecture. At first, an innate structure determines actions that can be initiated. An action-sensory flow monitor transmits its output to a pattern-detector which is triggered by certain patterns in the flow and can have top-down influence on action selection as well as generate anticipations about the incoming sensors. With these mechanisms, we obtain a *recursively self-maintained action system* that generates "trajectories in the good subspaces." As the agent grows, structural changes introduce meta-pattern detectors, and with this mechanism more complex value systems appear which provides a basis for more complex behavior.

Li Yang and Marwan Jabri discuss the issue of efficient representations for sensorimotor control. Efficiency is defined in terms of Barlow's principle of redundancy reduction (Barlow, 1961). The idea is to reduce redundancies due to complex statistical dependencies among elements of input streams. The authors focus on sparse representations – a distributed population coding in which very few neurons are excited in a represented pattern – because they have *instrinsic advantages in terms of fault-tolerance and low-power consumption potential, and are therefore attractive for robot sensorimotor control with powerful dispositions for decision-making.* Inspired by the mammalian brain and its visual ventral pathway, they propose a hierarchical sparse coding network architecture that extracts visual features for use in sensorimotor control. V1 complex cells (sensitive to direction of motion of stimulus) are modelled by combining multi-dimensional independent component analysis (ICA) with invariant-feature subspaces. V2 end-stopped cells (sensitive to contour length in addition to being tuned to position and orientation) are modelled as the minimization of an energy function. Experiments on images show that the model performs well in detecting corners, curvature or sudden breaks in lines. Since these contours are very important for shape representation, this is an important result for developing object representation and recognition.

In his poster, **Eric Berkowitz** approaches the problem of developing concepts in an artificial system. The system, named 'ASPARC', has primitive mechanisms including spatial perception and "instinctive knowledge that an act will have some re-

sult" (i.e., a kind of cause-effect knowledge). ASPARC focuses on spatial knowledge, representing a path as "ordered locations in space and the acts that cause motion between them", and the understanding that "'path' is actually an abstract concept built upon perceptions of walking." This system takes the concept of a 'path' to be "the basic prototype for all action."

Rémi Driancourt's poster shares a similar concern to that of **Yang and Jabri** and deals with the acquisition of efficient visual representations. The author proposes a developmental schedule for a robot to learn to build a representation of its physical environment (in which it is the only agent of change). The stages go from calibration and self-organization (1), reflexes (2), assimilation (repeat and perturb basic control policies) to a 6th stage in which "object permanency"-like ability develops. Statistical methods (multidimensional scaling, PCA, etc.) are used to construct methods which can self-organize perceptual input and perform hierarchical perceptual abstraction.

The poster of **Patrick Poelz and Erich Prem** focuses on concept acquisition in mobile robotics using multi-dimensional scaling. The authors apply a method called "isomap" (Tenenbaum et al., 2000) to realize symbol anchoring – the mapping of objects in the robot's environment on structures internal to the robot. Isomap is a nonlinear dimensionality reduction method capable of finding the intrinsic dimensionality of the data it processes, and preserving its nonlinear structure as captured in the geodesic manifold distances between all pairs of data points. The authors use this method to construct prototypes that correspond to object-like features of the environment. An interesting property of those isomaps is that analogical reasoning can be carried out by linear operations in the feature space created.

Finally, **Alexander Stoytchev**'s poster addresses the issue of the acquisition of a body-schema – a postural model of the body and a model of the surface of the body – by a robot system. As shown by recent studies on Japanese macaque monkeys, these schema are not static, but can be modified dynamically to include *extensions* such as tools. The author proposes a model built around body icons, pairs of vectors representing motor and sensory components of a specific joint configuration of the robot. These body icons are acquired through a motor babbling phase. A gradient ascent strategy (in a potential field determined by the position of the target) is used to move the robot from one configuration to the other. Body extensions are implemented as offset vectors. The model was sucessfully tested with a simulated two-dimensional arm manipulator with a gripper.

7. Emotions

In his poster, **Antonio Chella** applies the theory of conceptual spaces (Gärdenfors, 2000) to models of emotion. "Conceptual space[s] ... are metric space[s] whose dimensions are related with the quantities processed by the robot[s] sensors" and these "dimensions do not depend on any specific linguistic description." The hypothesis here is that an emotional system may enhance the flexibility of complex decision making skills in robotic systems. "Emotions are modeled by enhancing the robot[s] [conceptual space] by adding new dimensions" of rage-relief, and pleasure-fear. In the implemented robotic system, a person that runs towards the robot generates the fear emotion in the robot. An ART network is used to associate emotional responses with sensory input in the case of the primary emotions (i.e., rage-relief, pleasure-fear).

In another poster, **Andrea Kleinsmith and Nadia Bianchi-Berthouze** focus on robotics for "active and affective communication with humans." This research assessed human responses to computer animated body motions, finding that in some cases, the simplicity of the computer animated motions interfered with human subjects recognizing affective states. The authors also "implemented a system that incrementally learns the mapping function between [real human] body postures and [the] emotional labels" of happy, angry, and sad. They modeled "the mapping of posture features into emotional labels as a categorization problem." In a test, this system correctly categorized 107 of 108 images of the three affective postures.

8. Language and related skills

Brendan Burns, Charles Sutton, Clayton Morrison, and Paul Cohen focus on an associative approach to word learning. They define associative word learning as "the acquisition of meanings through the observation of the co-occurrence of the words and an example of their meaning." The authors collected their language learning data by first taking nine video movies of a robot performing five primitive actions: moving backward or forward, turning left or right, and sitting still. Each movie had a corresponding sequence of perception vectors – discrete perceptual information recorded by the robot while it was performing the actions. The movies were shown to human raters, who wrote textual descriptions of the movies. A Multi-Stream Dependency Detection (MSDD) algorithm, an algorithm based on information-theory, was used to find associations between the textual descriptions generated by the raters and the perception vectors corresponding to the particular movie. For example, MSDD found the perception vector "(zero positive)" to be associated with natural language phrases from

the raters including "* moves forward." The perception vector "(zero positive)" indicates that the robot is not turning and is moving forward.

Yukie Nagai, Koh Hosoda, and Minoru Asada approach the task of modeling the joint attention skills of an infant. Joint attention comprises triadic social interactions between child, caregiver, and objects or events. Being strongly implicated in cognitive skills such as language (Baldwin, 1995), it is of keen interest for epigenetic robotics to model joint attention. The authors instantiate the idea of development by transitioning between two phases of operation of the model. In the initial phase, the model attends to salient objects. If it finds a salient object in its view, it turns its cameras to more centrally focus on that object. When its cameras detect a salient object, the system may also detect a human face. In this case, the face stimuli plus the current camera angle are input to a learning module. In this way, the head turn angle required to turn the cameras and centrally focus on a salient object, which is in view along with a face, is used as input to a supervised learning algorithm. Thus, in the second phase, the system locates a face in its input, and turns its cameras to centrally focus on a salient object, deriving the head turn angle from previous associations between faces and turn angles. A sigmoid function is used to transition between the source of control for the head turn – from that based on salient objects (the first phase) to that based on using a face (the second phase).

Yilu Zhang and Juyang (John) Weng focus on the problem of "imperfect alignment" or "non-strict coupling" in learning relations between auditory and visual sensory information. For example, the visual appearance of an object may change over time while the auditory information is present. That is, the object may be rotated by a person holding the object, and the viewing angle may change because of a change in viewer position. Additionally, auditory energy is, of course, distributed over time. In their approach, they use an architecture based on Incremental Hierarchical Discriminant Regression (IHDR). Their system is based on Level Building Elements (LBE's), each of which is comprised of two IHDR trees. Each LBE also has a *prototype updating queue*, which is central to the authors' method of addressing the imperfect alignment problem. This queue contains representations obtained from one of the IHDR trees and serves to propagate information from more recent contexts to older contexts. Three LBE's were used in the overall system. Two LBE's serve to receive auditory and visual sensory inputs respectively, and one serves to learn associations between auditory and visual information. In the training task, the system receives audio inputs ("name?" or "size?") and visual inputs (a view of one of 12 stimulus objects) and supervised training output was provided in the form of a label (button press) categorizing the name or size of the object. Testing evaluated the system's ability to generalize on new audio and visual inputs, outputting the name or size information.

In his poster, **Robert Clowes** pursues Vygotskyian inspired robotics using simulation techniques. The question addressed in this work is: Can techniques for simulating language evolution "model the involvement of language-like systems in the development of higher practical activity?" This work uses a multi-agent system in which agents talk about the objects in their world, and also act on these objects. The agents have maps of the world, and are reinforced for making the world conform to their maps. A fitness function is used to reward social behavior. He argues that this research enables "a route to understanding ... how signs [may] mediate activity."

9. Imitation and skill acquisition

Mamiko Abe, Tomoyuki Yamamoto, and Tsutomu Fujinami look at how learning affects the kinematic and dynamic profiles of physical movements during a task performance. They compared the physical movements of beginners and experts in a kneading task (the preparation of clay for shaping). This task was selected because all body parts need to work together (joint synergy), the evaluation criterion is simple, and the learning process is long (about 3 years). The authors argue that such tasks cannot be imitated easily because of issues in both the learning and the teaching of the task. If visual instruction is assumed, mimicry may be regarded as similarity in the spatial domain. To master a skill, however, similarity in acceleration and velocity should be achieved. While hand-in-hand teaching could be considered, humans' sense of velocity is poor and the mapping between acceleration and velocity is unknown. Finally, the authors argue that learning through verbal descriptions is unlikely because it is difficult to linguistically describe parallel events. The authors performed a dynamical analysis on data from both beginners and experts based on results from a motion capture system. It was found that skilled subjects showed high-organization. Body parts could be fit into two groups – a torso group, and an arm group – with a constant phase difference. Cluster analysis revealed a good phase differentiation. The novices' motions instead were highly variable and did not share any common structure. In addition, the wave forms were not as regular, and phase differentiation could not be established. As a nice illustration of embodiment, it was shown that while novices tended to push down the clay, experts exploited their whole-body rocking motion. The authors suggest that appropriate auditory signals could

help mastering phase relationships.

Björn Breidegard and Christian Balkenius deal with the use of imitation to both produce and recognize speech sounds. The authors discriminate between "internally-controlled imitation" and "externally-controlled imitation." An instance of the former is *babbling*, whereby the system produces sounds at random, and subsequently attempts to produce that sound again (Piaget's circular reactions). Externally-controlled imitation involves the recognition of a human caregivers' speech sounds and its subsequent imitation. To get the system to develop its speech representations based on its own motor output, the authors use the Double Cone Model (DCM), a functional representation of a number of hierarchically organized cortical regions. In effect, those regions act as sequential self-organizing maps (SSOM) which are an extension of the traditional self-organizing maps (SOM). Experiments were carried out using a computer program that can perceive real-time sound input, and produce real-time sound output. It was found that the system used whatever training it had received in an attempt to reproduce speech. For example, when trained with a piano piece, the machine was somehow capable of *imitating* incoming speech sounds. The authors compare this ability with the use of sounds from one's native language to acquire sounds of a second language. To produce intelligible speech output however, training on speech signals was necessary. The authors discuss two interesting future directions to this work: (a) the simulation of impairements in language acquisition by introducing lesions and other disturbances in the model (e.g., aphasia); (b) implementation of the model on a 40-parameter virtual tongue for a good exploration of the problem of motor babbling in the acquisition of sound.

Tetsuya Ogata, Noritaka Masago, Shigeki Sugano, and Jun Tani investigate the issue of collaborative learning between a human operator and a robot. Here, it is not just the system's development which is considered, but the development of the coupling between the human operator or partner, and the system. According to the authors, the difficult issue in such coupled tasks is one of stability of the system for long periods of time because learning tends to generate complex dynamics, especially in terms of the relationship between operator and system. This study deals with a navigation task in which a humanoid robot and a human subject navigate together in a given workspace. A clever experimental setup has the machine and the operator become closely interdependent: the human operator needs the robot for local navigation, and the robot needs the human operator for global navigation. Recurrent neural networks (RNN's) are used by the system to construct a model for anticipating future sensory information and for generating appropriate motor commands. The authors' results show that consolidation learning is beneficial to learning performance by enabling the iterative rehearsing of past experiences. Interestingly, the learning development features phase transitions which are characteristic of the development of human collaboration. Coherent phases alternate with incoherent phases, possibly as a result of the human generating and/or modifying hypothetical strategies by using context information.

Yuichiro Yoshikawa, Junpei Koga, Minoru Asada, and Koh Hosoda share a similar goal to that of **Breidegard and Balkenius** and are interested in the developmental acquisition of phonemes in a robotic system. This study is inspired by the observation that infants acquire phonemes common to adults without having the capability to articulate, and without having explicit knowledge about the relationship between the sensorimotor system and phonemes. The authors believe that insights can be gained on this process by building a robot that reproduces a similar developmental process. Two (very general) issues must then be addressed: what are the interactive mechanisms involved and what should be the behavior of the caregiver/teacher? The authors use evidence in developmental pyschology to conjecture that (a) the caregiver's vocalization in response to infants' cooing reinforces the infant's articulation of cooing – this cooing is then interpreted as phonemes –, and (b) the caregiver's parroting helps specify the correspondence between cooing and the caregiver's phonemes as well as determines the acoustic properties of the phonemes. The authors experiment with an artificial articulatory system consisting of a 5-DOF mechanical system deforming a silicon-made vocal track, connected to an artificial larynx. The processing unit of the system consists in an extractor of formants and a learning mechanism with self-organizing auditory and articulatory layers. Starting off with random vocalization, the system uses the caregiver's parroting to boostrap its learning. As an illustration of the power of imitation, the system's utterance doesn't necessarily have to be the same as the caregiver's. However, this abstraction also brings what the authors call *arbitrariness* in determining the proper articulations. The authors address this issue by selecting articulations that minimize the necessary torque and intensity of deformation changes.

Finally, in a poster, **Pierre Andry, Philippe Gaussier, Jacqueline Nadel, and Michele Courant** argue that to imitate, a system doesn't need to build a model of the demonstrator's geometry. Instead, a simple homeostatic perception-action control loop can suffice. Inspired by self-organizing Kohonen networks (SOM's), the neural architecture consists of clusters learning many-to-one

proprioceptive-to-visual associations. The imitative task is simplified by controlling the system (a mobile robot equipped with an arm and camera) in the visual space instead of the motor space and exploiting a perceptual ambiguity – the perception of the experimenter is assumed to be the perception of the effector's position. By acting as an homeostat, the system attempts to reduce the error, which results in basic imitation. The authors claim that deferred imitation of a trajectory is possible (e.g., observing a head movement, and reproducing it later with the arm).

10. Conclusions

This year again, two issues noted by C. Prince in his conclusion of last year's workshop are evident in almost all contributions: What behaviors should be provided innately in an epigenetic robot? And, how can these behaviors be combined? While **Grupen** proposes a progressive build-up from parameterized reflexes along the lines of a somewhat classical behaviorist view, other papers have focused on the information theoretic aspects of the issue, mentioning the importance of sparse coding, redundancy reduction, statistical inference, etc. (e.g., **Yang and Jabri, Poelz and Prem**). With respect to learning, imitation appears prominently. Interestingly, two contributions (**Yoshikawa et al., Breidegard and Balkenius**) exploit imitation learning in similar fashion, combining imitation of self-movements (leading to Piagetian circular reactions) and imitation of a caregiver's movements. While **Andry et al.** aim to show that imitation can be simplified by exploiting ambiguous perception, **Abe et al.** focus on a task which shows the possible limits of imitation: humans may be good at imitating trajectories, but not velocity or acceleration profiles. And as **Pollick et al.** demonstrate, humans do not always look at trajectories to evaluate other's movements. The invited contribution of **Gergely** should be particularly noted in regards to imitation. From infancy, humans are selective about the acts that they imitate. Indeed, perhaps it would be stereotypically robotic (and not epigenetically robotic) to imitate *every* action observed!

Thinking about epigenetic robotics more broadly, **Gergely** noted that epigenetic robotics presently has a bias towards sensorimotor issues. While this is perhaps not surprising in a new field tackling a broadly integrative subject matter (see also Kitano, 2002), serious thought should be given to this observation. Just skimming the contributions of our present workshop shows research including language grounding (e.g., the invited talk of **Roy**, and the papers of **Burns et al.**, and **Zhang and Weng**), the social skill of joint attention (**Nagai et al.**), and visual object segmentation (**Arsenio et al.**). Can we consider these projects to be breaking ground in epigenetic robotic models of higher-level cognition? On the one hand, these projects tackle skills that in humans rely on psychological development. On the other hand, the extent to which developmental processes are involved in these projects can be furthered. This issue has broad application: It seems rare for epigenetic robotics projects to move past emergent behaviors, to *successively emergent* behaviors. Skills such as language should emerge in generally ordered, perhaps domain specific, phases. For example, the first words comprehended by a 10-month-old infant are learned in a predominantly associative manner and do not yet integrate the use of joint attention (K. Hirsh-Pasek, personal communication). A few months later, the first words produced by a one-year-old are further signs of language learning, but it is not until still months later that the child combines her words into longer utterances (e.g., Golinkoff and Hirsh-Pasek, 2000). While various robotic systems have been shown to demonstrate emergent behaviors, demonstrations of successive series of emergent behaviors are not well established. Future research needs to address the question: What epigenetic robotic architectures can provide for such ongoing emergence of skills?

In closing this introduction, we thank the Communications Research Laboratory (CRL) of Japan for their generous support of this workshop. Boston University has been a patient local sponsor through the process of arranging this workshop. We also thank the program committee members for their efforts in reviewing submissions.

References

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, (23):21–41.

Baldwin, D. (1995). Understanding the link between joint attention and language. In Moore, C. and Dunham, P., (Eds.), *Joint attention: its origins and role in development*, pages 131–158. Hillsdale, NJ: Lawrence Erlbaum.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W., (Ed.), *Sensory communication*, pages 217–234. Cambridge, MA: MIT Press.

Bickhard, M. (1999). Interaction and representation. *Theory and psychology*, 9(4):435–458.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gergely, G., Bekkering, H., and Király, I. (2002).

Rational imitation in preverbal infants. *Nature*, (415):755.

Golinkoff, R. and Hirsh-Pasek, K. (2000). *How babies talk*. NY: Plume.

Kitano, H. (2002). Computational systems biology. *Nature*, (420):206–210.

Oppenheim, R. (1981). Ontogenetic adaptations and retrogressive processes in the development of the nervous system and behaviour: A neuroembryological perspective. In Connolly, K. and Prechtl, H., (Eds.), *Maturation and development: Biological and psychological perspectives*, pages 73–109. Suffolk, England: The Lavenham Press Ltd.

Roy, D. and Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive science*, (26):113–146.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous robots*, (12):13–24.

Tenenbaum, J., Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, (290):2319–2323.

Thelen, E. and Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Wellman, H. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.