IEEE CIS

# AMD Newsletter

Robotics
Psychology
Neuroscience
Machine Intelligence

The Newsletter of the Autonomous Mental Development Technical Committee

VOLUME 1, NUMBER 2 ISSN 1550-1914 OCTOBER 2004

## Message From The Editor

Welcome to the second issue of the AMD TC Newsletter. We are glad to announce that the AMD TC Newsletter has received the International Standard Serial Number! In addition to the existing news, you will find an interesting dialog column: the discussions on “Object Detection and Object Variance in Autonomous Mental Development.” Of course, you will have the call for papers, the call for participations, and the glossary, as usual. Enjoy the issue!

- Yilu Zhang, Editor of the AMD Newsletter

## Committee News

- June, 2004: IEEE Neural Networks Society changes its name to IEEE Computational Intelligence Society. The URL of the new Society website is http://ieee-cis.org.
- August, 2004: The AMD TC Newsletter has received the International Standard Serial Number (ISSN): 1550-1914. A single ISSN uniquely identifies a serial (such as magazines, newspapers, annuals, journals, proceedings) regardless of language or country in which published. ISSN are assigned by a network of over 60 centers worldwide coordinated by the ISSN International Centre located in Paris. ISSN are assigned to serials published in the United States by the National Serials Data Program (NSDP) of the Library of Congress.
- September, 2004: The AMD TC website has been modified and moved to http://www.cse.msu.edu/amdtc/ to be integrated with the new website of the Society: http://www.ieee-cis.org/amd
- September, 2004 – AMD TC meeting announcement: The next ICDL Advisory Board/AMD TC meeting will be held at 7pm, October 21, 2004, during ICDL '04, which will be held at The Salk Institute for Biological Studies, La Jolla, California. Everybody is invited to discuss AMD TC organization matters.

## Dialog Column

### Dialog: Object Detection and Object Variance in Autonomous Mental Development

Juyang Weng, Department of Computer Science and Engineering, Michigan State University East Lansing, MI 48824; Yilu Zhang, GM Research and Development Center, Warren, MI 48090

Suppose that we need to discover or design a set of mechanisms in a developmental program that will develop an agent who exhibits a capability of object detection under a wide variety of object variance, following the agent's rich experience of interacting with the real world. For example, a human adult is able to recognize an object in a cluttered background to be a cup (i.e., object detection), largely independent of the cup's position, size, orientation, and the lighting directions, etc. (i.e., object variance). Of course, the capability of handling object variance is not perfect.

In pattern recognition, early invariant features have been proposed for recognition (e.g., color, texture, etc.), but those invariant features are applicable only to a particular known setting for a particular task. What set of developmental mechanisms (which are innate) enables an AMD developmental being, biological or artificial, to develop cognitive capabilities of object detection with the associated object variance?

One is the “innate physical knowledge” school, which states that the capability of detecting objects and understanding

# Dialog Column

object permanence (although they vary in position, size, orientation and may be occluded) is largely a hereditary capability. E.g., newborn babies like to see a face-like pattern (which is helpful to detect a human face), and 3- to 4.5-month-old infants are surprised when they see some physical rules about objects are violated (e.g., [1] [2]).

The other is the "experience-shaping" school, which considers that the above capability of a 3- to 4.5-month-old is largely acquired from experience guided by a set of task-nonspecific developmental mechanisms (including some innate behaviors). E.g., human adults do not do well in recognizing upside-down faces.

On one hand, the "innate physical knowledge" school cannot disregard the developmental experience of a 3-month-old and, thus, it is hard to attribute the behaviors of 3- to 4.5-month-old to "innate physical knowledge." On the other hand, the "experience-shaping" school is not very convincing without showing what actual developmental mechanisms (presumably in the genes) are sufficient to develop cognitive capabilities of object detection under a wide variety of object variance.

References:
[1] E. S. Spelke. Perceptual knowledge of objects in infancy. In J. Mehler, E. C. F. Walker & M. Garrett (eds.) Perspectives on mental representation. Hillsdale, NJ: Erlbuam, 1982.
[2] R. Baillargeon. Object permanence in 3.5- and 4.5-month-old infants. Developmental Psychology, 23:655-664, 1987.

## Reply to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Stephen E. Levinson, Department of Electrical and Computer Engineering and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

The very same questions that arise in vision also arise for speech recognition. The speech signal is described by its spectrum which appears to yield features of high variability leading to high classification error rates for machines whereas human performance is near perfect with little apparent effort on the part of the listener. One approach to this problem is to build a system that develops its own method of speech recognition.

Two different aspects of this approach are evident. First, some components of the developmental process appear to be instinctual, present at birth, having been formed in evolutionary time. Second, there is an adaptive component that develops during the life of the individual. Although both components exist in humans, as a practical matter we treat the first one by trying to design feature invariance directly into the machine while the latter component can be made truly adaptive.

In my research [1], the adaptive or developmental processes are based on two fundamental principles. First, cognitive functions are not isolated. They work synergetically. That is, cognitive function is the result of adapting to input signals from the entire sensorimotor periphery including audio, video, tactile, haptic, and proprioceptive data. Information from the sensorimotor periphery is organized in associative memory so that mutual information from all channels is used simultaneously for all perceptual/cognitive functions. The mathematical idea here is that sensorimotor features are points in a high dimensional space in which nearly invariant features are located. Whereas, an isolated sensory channel is a projection of the composite signal down onto a space of much lower dimension in which it appears to be of high variability.

Second, the method of adaptation is reinforcement learning resulting from real time interactions with the physical world. Reinforcement simply rewards successful behavior and punishes unsuccessful responses. The rewards and punishments are not internal but must also come from the sensorimotor system. Rewards and punishments are used to adjust the contents of the associative memory so that useful memories are retained and those leading to poor behavior are erased.

Reference
[1] Stephen E. Levinson, Weiyu Zhu, Danfeng Li, Kevin Squire, Ruei-Sung Lin, Matthew Kleffner, Matthew McClain, and Johnny Lee, "Automatic Language Acquisition by Autonomous Robot", Proc. Int. Joint Conf. on Neural Networks, Portland, OR, July, 2003, (invited).

# Dialog Column

## Reply to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Olaf Sporns, Department of Psychology, Indiana University, Bloomington, IN 47405

Weng and Zhang have raised a very important issue in their dialog statement. The problem of object detection and recognition under "natural" and unconstrained conditions (noisy, variable, cluttered, obstructed etc.) remains a major challenge for the design of robust visual robotic platforms. Is this sort of ability innate ("hardwired" into the architecture) or is it the result of experience-dependent learning and plasticity? While newborn humans exhibit some degree of seemingly innate preferences (for example, as Weng and Zhang correctly point out, towards faces or face-like stimuli), there is also increasing evidence that neural populations involved in object identification and recognition undergo plastic changes even in the adult. Neurophysiological studies in primates have shown that the development of visual expertise is accompanied by lasting changes in the response properties of cells in the inferior temporal cortex (Kobatake et al., 1998) [3]. More recent work by Jagadeesh et al. [2] has demonstrated that the behavioral salience of a stimulus (its relevance to the organism) can be a critical factor in determining and controlling changes in inferior temporal cortex object representations. These studies and others attest to a significant level of neural plasticity within visual cortical regions involved in object recognition in the adult primate. While systematic develop-mental studies of how this neural architecture grows and matures are still lacking, the evidence strongly suggests that plasticity, learning, and saliency play a crucial role.

Another issue in this discussion relates to the role of sensorimotor activity (or to use a more modern term, of embodiment) in the development of the ability to recognize objects. A possible hypothesis is that at least some invariant visual properties of objects (such as translation, rotation or size) may actually depend on such embodied interactions. In a robot model, self-generated movement and the temporal dynamics of some visual cortical circuits were found to interact to produce invariant neural activity [1]. That motor behavior can support and generate "useful" correlations in visual input was also elegantly demonstrated by Metta and Fitzpatrick [4]. In their experiments, the touching and pushing of an object creates coherent motion that is an important clue for the segregation of the object from the background.

In my opinion, plasticity and embodiment are related. The coupling of sensorimotor behavior and neural dynamics supports the development of object identification and recognition by generating additional statistical structure in the input stream. This coupling, in addition to some level of innate "likes and dislikes", may point to an important and general developmental mechanism.

References:
[1] N. Almassy, G.M. Edelman, and O.Sporns. Behavioral constraints in the development of neuronal properties: A cortical model embedded in a real world device. Cerebral Cortex 8, 346-361, 1998.
[2] B. Jagadeesh, L. Chelazzi, M. Mishkin, and R. Desimone. Learning increases stimulus salience in anterior inferior temporal cortex of the macaque. J. Neurophysiol. 86, 290-303, 2001
[3] E. Kobatake, G. Wang, and K. Tanaka. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J. Neurophysiol. 80, 324-330, 1998
[4] G. Metta, and P. Fitzpatrick. Early integration of vision and manipulation. Adaptive Behavior 11, 109-128, 2003.

## Reply to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Chris Brown, Department of Computer Science, University of Rochester, Rochester, NY 14627, USA

As stated, the supposition is that we have an agent with "rich experience of interacting with the real world". For me the interesting issue is not how rich experience with the world leads to getting acquainted with it, but how agents can be acquainted with the world through minimal, seemingly irrelevant, experience [1]. Just what experiences (visual or not) are needed to support development of general vision? How do diferent cultures (and even different species) develop a parsing of a confusing world (with familiar and novel objects) upon which they all (arguably) agree? How can it all be so independent of individual experience when no current learning models are?

# Dialog Column

Vision and language are a couple of "faculties", or "modules" [2], which means (among other things) they are functions of our brains that are not introspectable and are very difficult to "control", especially to turn off. E.g. it is hard to view a normal scene and only see color patches. Language is more like something learned (from individual-specific human experience) - animals don't have it and different languages do not agree on the "objects". However there do seem to be universal constraints on language structure, the mature linguistic faculty acts like a module, and language development is still a mystery.

How individualizable is the visual faculty? Perhaps one could test the "customization" of visual development with methods something like the well-known raising of kittens in restricted visual environments. The difference is that only the high level behavior of the environment would be manipulated so that laws governing objects and their appearance are not "natural". Unfortunately, tactile, proprioceptive, and other unknown types of stimuli could well be relevant. Then perhaps one could measure differences in the development of visual recognition, generalization, invariance, etc. and explore the relevant dimensions of the causes and effects of such differences.

I am much more an innatist than an empiricist on these matters: I predict a surprisingly small exposure to a surprisingly small subset of a surprisingly wide range of stimuli will do the trick, and that no existing learning algorithms have anything to do with what goes on in natural development.

Last, there is a further issue of what good such biological experiments will do us. My colleague Randal Nelson says: "Biological structures are pretty complicated things, and it seems entirely possible that the system has pretty strong biases towards what sorts of things it is good at acquiring without having 'explicit' innate information one could dig out. One might do better asking how it might be possible to describe the sort of biases that are present, and if there is any simpler explanation than the system itself" [3].

References
[1] C. M. Brown. Vision, learning, and development. Technical Report 492, Department of Computer Science, University of Rochester, May 1994. In DSpace: http://hdl.handle.net/1802/752.
[2] J. A. Fodor. The Modularity of Mind. Bradford, MIT Press, Cambridge, MA, 1983.
[3] R. Nelson. Personal communication.

## Reply to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Ming Xie, Department of Mechanical and Production Engineering, Nanyang Technological University, Singapore 639798

I express my opinion from an engineering point of view. A human being is a reality of an organized complexity. A human being has an innate body at the time of birth, which constantly undergoes physical development during the life span. A human being also has an innate mind at the time of birth, which also undergoes constant psychic (or mental) development during the life span. Therefore, it is a fact that a human being has, at the time of birth, a) an organized body for physical development, and b) an organized mind for psychic (or mental) development. This organized mind enables a human being to incrementally build up his or her mental abilities in many areas, such as the ability to learn and understand a natural language, a natural scene, etc.

Next, consider the crucial question of what constitutes an organized mind. It is obvious that an organized mind must be embodied in a physical entity, called the brain, which is responsible for information processing. However, it is less obvious that an organized mind should also have an organized memory, which is indispensable to the cognitive state of "knowing" (e.g. know-how). But, the least obvious entity in an organized mind is the innate mental processes, or innate principles, solutions or algorithms, which are behind the mental abilities of learning, analysis, synthesis and understanding. Let's admit that an organized mind consists of three constituent entities, namely: a) the brain for information processing, b) the memory for knowledge modeling and representation, and c) the innate principles, solutions or algorithms for learning, analysis, synthesis and understanding.

# Dialog Column

Then, come to this relevant question of how a human recognizes objects in an effortless manner, in spite of object variation. One hypothesis, which could explain the human mystical mental ability in recognizing objects from real-time images or off-line photos, is as follows:

Object recognition has three prioritized sub-goals. They are: (a) object detection (e.g. is there a real object?), (b) object categorization (e.g. is it a car?), and (c) object identification (e.g. is it my car?). In order to achieve the second and third sub-goals of object recognition, the observer must have knowledge in his or her memory, which serves as reference or criterion for categorization and identification. Thus, an autonomous agent must have "internalized knowledge" [1]. In addition, this implies that object categorization or identification should be knowledge-based, or modell-based. Most importantly, it is certain that model-based recognition could easily cope with objection variation in size, pose and appearance. In literature, there are many proposed solutions to tackle the issues of object variation in size and appearance. But, the most critical issue is how to cope with object variation in pose. In literature, ad hoc solutions such as the aspect-graph approach and its variants have been proposed. Recently, it is discovered that the deterministic three-view approach for robotic head-eye coordination could be extended to elegantly solve the issue of object variation in pose, under the context of model-based object recognition [2].

On the other hand, we may raise the question of where the observer gains the knowledge (i.e. models) necessary for successful object recognition. The answer is through the process of autonomous learning. Hence, an autonomous agent should be "experience-shaped".

References

[1] M. Xie, J. Kandhasamy and H. F. Chia. Meaning-centric Framework for Natural Text/Scene Understanding by Robots. International Journal of Humanoid Robotics. Vol 1, No. 2, pp 375—407, 2004

[2] M. Xie. Fundamentals of Robotics: Linking Perception to Action. World Scientific, 2003

## Reply to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Zhengyou Zhang, Microsoft Research, Redmond, WA 98052, USA

Although the examples listed in John and Yilu's Call for Dialogue only concern the visual world, the same problem arises in other areas such as auditory scene analysis. Take as an example our recent work on office presence determination [1]. One of the tasks is to classify auditory singles as human conversations, phone rings, or others. All phone rings can be considered as a class of auditory objects. Solving this problem seems to be trivial to a human adult, but it turned out to be extremely difficult to develop a robust technique to cover all the variations. There exist a large number of different phones (mechanical and digital) and ring tones used in cell phones have even larger variations. Autonomous Mental Development (AMD) came out to be a very attractive approach. E.g., it does not require a hand-built object model and, thus, it enables the user to train new ring types conveniently at the user end, if his phone rings are not recognize well by the default trained "brain."

With that said, AMD faces several challenges. The first is the long period of time necessary for the developmental process. Even if we knew how our brain works and a computer were as powerful as our brain, it would still take several years to develop an adult system with someone living intimately with the system. It will be a tough job to sell the idea to a government agency and much harder to a company. We have to limit our scope and show measurable progress, stage by stage, within a reasonable period. The office presence project is an example of such stage progress.

The second is how to define the measure of success. Take again the phone ring as an example. Nowadays, you can record anything on your cell phone as your ring tone. Let's say you recorded "You've got a call". When a human hears "You've got a call", it is easy for him to distinguish whether it is your personalized phone ring or it is part of your conversation (e.g., when you tell your friend: "I just recorded 'You've got a call' as my ring tone"). The key point here is that the domain is open. Can an AMD machine achieve this level of intelligence? If GOFAI (Good Old Fashion AI) has failed to reach this goal, can AMD reach it eventually?

# Dialog Column

The third challenge is the paradigm transition. AMD requires a radical change in architecture and working process from traditional approaches. Governments and companies have invested a huge amount of resources in, say, face recognition and speech recognition. How can we convince them to invest in a completely new technique ? Of course, it would be fantastic if we could demonstrate the AMD is indeed superior to the traditional approach for the same task. Considering the huge investment already spent and the fine tunings already integrated in the traditional approaches, it might take many years before AMD is able to catch up to the same task. In addition to working on problems that AMD technique has clear advantages (e.g., the office presence project), another possible direction is for AMD approach to leverage the rich knowledge already acquired in developing the traditional systems.

Reference
[1] X. Huang, J. Weng, and Z. Zhang, "Office Presence Detection Using Multimodal Context Information", in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), May 17-21, 2004, Montreal, Quebec, Canada.

## Reply to replies to "Dialog: Object Detection and Object Variance in Autonomous Mental Development"

Juyang Weng, Computer Science and Engineering, Michigan State University, East Lansing, MI 48824
Yilu Zhang, GM Research and Development Center, Warren, MI 48090

First, the most basic mechanism should be the AMD mental architecture, whose study is currently lacking. The SAIL and Dav robots in our EI laboratory have been using a general-purpose mental architecture outlined in Weng 2004 [2] to deal with object variance for not only visual sensorimotor behaviors, but also auditory sensorimotor behaviors (as S. Levinson correctly pointed out), and hopefully invariance of any sensory modality.

The basic idea is not to hand-design which features are object invariant (they are typically too weak to be applicable to practical settings). Instead, we proposed two types of mechanisms to autonomously select category relevant features: (1) the bottom-up, unconscious mechanisms, e.g., the most discriminating feature subspaces in the self-generated HDR tree [2] and (2) the top-down, "conscious" mechanisms, e.g., attention selection as internal behaviors [4] that select relevant parts of input. The mechanism, (1) which is crucial for early cortical development, requires relatively rich experiences to establish statistics. The mechanism (2) can effectively deal with a large amount of variation with somewhat minimal experience as C. Brown wished. In Zhang & Weng 2003 [4], after learning to respond correctly to petal-specific verbal commands, the SAIL robot was taught, with minimal training experience, to draw a 4-petal flower, without a need to issue any petal-specific command. The attention behavior is triggered by the context priming mechanism: consistent temporal pairing of last-context (e.g., end of drawing the i-th petal), the primed context (the (i+1)-th verbal command) and the further primed context (the action that executes the (i+1)-th command). Note that no symbolic logic programming is involved. The representation is in terms of numeric vectors in feature subspaces. This is an example of gaining "models" from embodied experience, as M. Xie alluded.

Interestingly, the SAIL developmental robot also displayed [1] human 3- to 4-month-old behaviors in Baillargeon's well-known draw-bridge experiment. Thus, at least the context priming mechanism can explain the "object permanence" raised by Jean Piaget, while no object model is built in.

The context priming mechanism has further demonstrated its power in learning object variance while each object is rotating while being learned [3]. The same mechanism can be extended to any type of object variance in principle, as long as the experience shows such a variation. More experimental studies are needed to show the power and the efficiency of this mechanism. However, it seems not necessary to impose that machine AMD must be as slow as biological AMD, since machine AMD does not have to follow the biological maturation clock.

References:
[1] Y. Chen, J. Weng and X. Huang, " `Object Permanence': Results From Developmental Robotics," in Proc. 2004 International Joint Conference on Neural Networks, Budapest, July 26 - 29, 2004.
[2] J. Weng, "Developmental Robotics: Theory and Experiments," International Journal of Humanoid Robotics, vol. 1, no. 2, pp. 199-235, 2004.

# Dialog Column

[3] Y. Zhang and J. Weng, “Conjunctive Visual and Auditory Development via Real-Time Dialogue,” in Proc. 3rd Int'l Workshop on Epigenetic Robotics, Boston, MA, pp. 974 - 980, August 4-5, 2003.
[4] Y. Zhang and J. Weng, “Chained Action Learning through Real-time Interactions,” in Proc. of IEEE/INNS International Joint Conference of Neural Networks (IJCNN 2002), Honolulu, HI, May 12-17, 2002.

# Call For Participation

**THIRD INTERNATIONAL CONFERENCE ON DEVELOPMENT AND LEARNING (ICDL ‘04)**

October 20-22, 2004, The Salk Institute for Biological Studies,
La Jolla, California
Register now online at http://www.icdl.cc/

# Call For Papers

SPECIAL ISSUE ON AMD IN THE IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION.
http://www.cse.msu.edu/~weng/TEC_AMD_CFP.htm
Deadline for camera-ready papers: December 31, 2004

The 4th International Conference on Development and Learning (From Interaction to Cognition), July 19-21, 2005, INTEX Osaka, Japan
http://www.icdl05.org
Deadline for camera-ready papers: February 18, 2005

# Glossary

**Object permanence**: The understanding that objects continue to exist even when they are out of sight. The term object permanence and its measure was first introduced by Jean Piaget. In his book "The Construction of Reality in in the Child", Piaget described that 7- to 12-month-old infants failed to retrieve a completely hidden object, suggesting "out of sight is literally out of mind." Recent studies have suggested that Piaget's manual search task was too conservative as a test for object permanence since it required sophisticated motor skills and efficient memory. Renee Baillargeon's "drawbridge" experiment used looking time as a measure, making it possible to test this concept for very young infants. However, the explanation of "innate knowledge" about object permanence has been challenged.

Editor: Yilu Zhang Editorial Assistant: Nicholas Wassil Logo: Joe Zimmerman