Physically and Emotionally Grounded Symbol Acquisition for Autonomous Robots

Masahiro Fujita, Rika Hasegawa, Gabriel Costa, Tsuyoshi Takagi, Jun Yokono, and Hideki Shimomura
6-7-35 Kitashinagawa, Shinagawa-ku, Tokyo 141-0001, Japan
mfujita@pdp.crl.sony.co.jp

Abstract

In this paper we present a novel concept of emotionally grounded symbols, which gives information about the importance of objects for the survivability of an autonomous robot. In addition emotionally grounded symbol provides information about emotions that were experienced during learning as well. Using this concept we implement an Emotionally Grounded (EGO) Architecture on a quadruped robot, which is able to acquire physically grounded symbols (objects' names) and emotionally grounded symbols, by interaction with humans and the environments

Introduction

One of the challenges in robotics and A.I. is to build an open-ended robot that learns new knowledge through interaction with a human and its environment. In the area of knowledge acquisition, acquiring names of objects has been well studied in many fields such as psychology and cognitive science. We have proposed the entertainment application of autonomous robots[12], and knowledge acquisition must be a key to engaging people to interact with entertainment robots such as pet-type robots.

Recently, some researchers [20][17][16] propose methods that can acquire names of objects using the physically grounded symbol concept. However, it is still a problem for autonomous robots to decide what kind of behavior should apply to the object, even if it has a name. How can we tell the meaning of an object? More basically, what is the meaning of an object?

According to the ethological model of behavior control[14], both external stimuli and internal motivations or emotions cause a robot system to select the appropriate behavior in a given situation. Thus, the meaning of the object is deeply related to the results of internal and emotional evaluations of the autonomous robot at least for this type of robots.

In this paper, we propose a novel concept for emotionally grounded symbols that gives a meaning to physically grounded symbol. The concept of emotionally grounded has two important aspects. First, an emotionally grounded symbol gives information about the importance of the object for the survivability of an autonomous robot. Based on this information it can be expected that if the robot applies a proper behavior to the object, it can get something in return and can keep its body in safe. Namely, the emotionally grounded symbol is a key to the homeostatic regulation [2] of the autonomous robot. Second, an emotionally grounded symbol gives information about emotions that were experienced during learning. Based on this information, the robot can remember the emotions if it applies a particular behavior to the object. This capability is essential for real animals and humans to survive in the real world. Damasio[8] proposes "the Somatic Maker Hypothesis", which explains the importance of remembering the experienced emotion, which enables essential and quick decision-making in the human brain.

In the remainder of this paper, first we describe the architecture for homeostasis regulation, where we present our ethological model for behavior control. Then, we explain physically grounded symbol acquisition, where we introduce an "information acquisition behavior". We also describe emotionally grounded symbol acquisition, where we again describe a concept of emotionally grounded symbols, using our emotional system.

We implement the EGO (Emotionally GrOunded) architecture in an autonomous quadruped robot that acquires information via interaction with a human and the environment. Some important features such as "shared attention" are handled in the architecture. Finally, we describe a preliminary experiment with the EGO architecture.

Architecture for Homeostatic Regulations

Terminologies

There are some terminologies, which are used differently by different researchers in the study of emotions[7][8][15][19][25]. In this paper, we use the following definitions:

- **Internal variables** are related to internal factors of the robot as a lifelike creature. For example, the amount of remaining battery is one of the internal variables. The temperature of CPU is also an example. Note that for entertainment purposes, we also define simulated internal variables. For example, we simulate an amount of "remaining water" so that the robot can exhibit a "drinking" behavior.
- **Drives** are signals from innate neural circuits in the brain, which are necessary for our body to keep alive.

Hunger, thirst and sleepiness are examples of drives. The hunger drive occurs when an amount of remaining battery or energy becomes less than a particular level. Based on the drives the robot will try to execute a proper behavior to maintain the internal variables in particular ranges. Thus, homeostasis regulation is a key to behavior selection.

- **Innate emotions** are emotions, which are derived from drives. For example, anger and fear are generated when the above drives are not satisfied. On the other hand, joy and pleasantness are generated when the above drives are satisfied.
- **Associated emotions** are emotions, which are linked to emotional experiences. For example, assume that you encounter a traffic accident and feel fear or threat (innate emotions). Then, the car or the place may be associated henceforward with that emotional experience. Afterwards, when you see the place, you may feel fear or threat. This is the associated emotion.

Note that primary emotions and secondary emotions used in Damasio's book[8] are related to "Innate emotions" and "Associated emotions" in this paper. However, in his context, the primary emotions are pre-organized emotions by which an animal can react to avoid dangerous situations very quickly. The secondary emotions are the same as associated emotions. On the other hand, some researchers[15][25] use the terms primary emotions and the secondary emotions differently, in the same way as the drives (primary) and the innate emotions (secondary) defined above. Therefore, we clarify these different terminologies to avoid confusion.

Behaviors for Homeostatic Regulation

Basically, behaviors are selected to keep the internal variables in proper ranges. However, the study of ethology tells us that behavior selection does not depend only upon internal variables, but external stimuli as well[14]. For example, assume that an animal is hungry and thirsty based on its internal variables. In addition, assume that it is more hungry than thirsty, but there is a cup of water in front of the animal. Then, ethological study shows the animal tends to select a drinking behavior, not searching a food. Thus, both external stimuli and the states of internal variables are important for behavior selection.

We developed an architecture[1] to realize the facts described above. Fig. 1 shows the architecture of ethological behavior modeling. This is basically a behavior-based architecture, containing many behavior modules. Each behavior evaluates both external stimuli and the internal variables. The external stimuli come from a perception module, whose importance to the corresponding behavior is evaluated by the releasing mechanism module, which outputs a release signal to the corresponding behavior module. The internal variables are evaluated in the drive module, which outputs a motivation value to the corresponding behavior module. These two factors, the release signal and the motivation value are evaluated in the behavior module, which then outputs a behavior value. The behavior selector basically selects the behavior with the maximum behavior value.

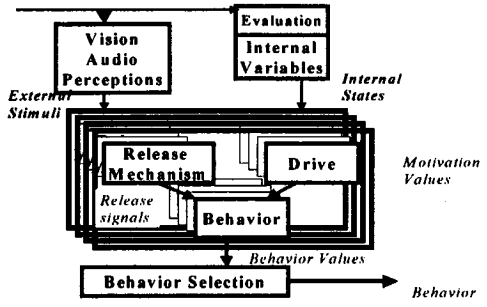

Fig. 1: Ethological Modeling of Behavior Selection for Homeostatic Regulation

This architecture and mechanism generate proper behaviors to regulate the internal variables in particular ranges, namely, behaviors in support of homeostasis.

Physically Grounded Symbol Acquisition

Kaplan[17] and Roy[20] studied the word acquisition capability of a robot in the real world.

Fig. 2 shows a simple diagram showing the essence of the word acquisition. Assume that there are two perceptual channels, the visual perception channel and the auditory perception channel. The visual perceptual channel outputs visual events (VEs), which are category IDs of the visual perception module. The auditory perceptual channel outputs auditory events (AEs), which are also category IDs of the auditory perception module. These VEs and AEs can be considered as grounding to the physical world through the perceptual channels. For example, a particular VE (VE-1) is a color segmentation event, which indicates a "red" object in the visual input of the robot. An AE (AE-1) is a phoneme sequence [red]. If these two events occur simultaneously, these two are first stored in a Short-Term-Memory (STM), and then memorized in an associative memory or a Long-Term-Memory (LTM)(Fig. 2 (a)). As we describe in a later section, the real implementation includes dialogue with the human to learn the object name.

After the learning episode is over, if only one event, e.g. the AE-1 (phoneme sequence [red]), is input to associative memory, then the memorized VE-1 is output from the associative memory, which is the category indication of "red object. Thus, the symbol is grounding to both visual and audio perceptual channel.

Of course if only the VE-1("red" object) is presented, then the associative memory can recall AE_1 (phoneme sequence [red]) .

In order to incorporate physically grounded symbol acquisition into the behaviors of an autonomous robot, specifically a pet-type robot, we introduce an internal variable, "remaining information" and a corresponding drive "curiosity". Then, information acquisition becomes a

homeostatic behavior striving to keep the remaining information within a particular range.

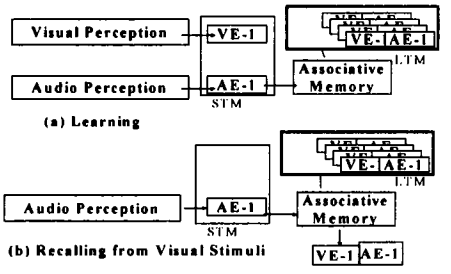

Fig. 2: Physically Grounded Symbol Acquisition: (a) Associative learning of visual event and audio event (name) (b) Recalling its name from Visual event

The internal variable, "remaining information" increases when the robot memorizes a name by the word acquisition, and decreases exponentially by time.

If an external stimulus corresponds to an "unknown object", which has not been learned by the word acquisition yet, the release mechanism generates a release signal to the information acquisition behavior. This will invoke the information acquisition behavior when the robot learns about the unknown object.

Emotionally Grounded Symbol Acquisition

Assume that we implement the information acquisition behavior. Then, we can give a name to an object. For example, we give a name to an object <apple>. Assume that when we finger-point towards the object and ask the robot, "what is it?" Then the robot can answer, "I know it. It is an <apple>". The robot knows the <apple> is "red", which is a color segmentation category. However, the robot does not know the function of <apple>, namely that <apple> is used to increase the internal variable, remaining energy, if the robot eats the <apple>. This is the basic idea of our concept of emotionally grounded symbols. Physically grounded symbols are associated with its internal meaning in terms of a change of the robot's internal variables.

In our implementation, a symbol has three different yet associated channels, the visual perceptual channel (color segmentation), the audio perceptual channel (word recognition), and changes of internal variables (ΔI).

In addition to generate a release signal to the corresponding behaviors, ΔI plays another important roll in our EGO architecture, which is to generate associated emotions.

Fig. 3 shows how emotions are generated in our system. In the figure, the internal variables generate drives and evaluate external stimuli so that the selected behaviors maintain the internal variables within particular ranges (Homeostatic regulation).

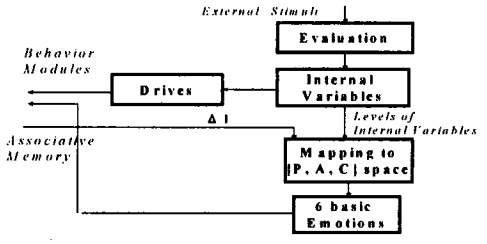

Fig. 3: Emotion System

In parallel, the internal variables are mapped into 3 dimensional emotional space, (Pleasantness, Activation(Arousal), Certainty), which is a modification of Takanishi's approach[24]. The current state of the internal variables can be categorized as one of the 6 basic emotional states proposed by Ekmann[11]. Note that it is not necessary in our EGO architecture to map the internal variables into the 3 dimensional space. Kismet project[7] uses another 3 dimensional space, and other researchers use other emotional spaces[15].

Pleasantness is a measure if a change of internal variables is satisfied, or not. Activation is a measure if a robot is awake, or not. Certainty is a measure if a stimulus is from known object or not. Takanishi introduces this certainty in his facial expression study using a robot[24]. His measure, however, is a kind of a confidence measure of recognition of the stimuli. Our certainty is a measure of memorized information of a corresponding object.

A change in emotions is applied in two ways. One is by directly monitoring the internal variables. Another is from the associative memory, which memorizes symbols with ΔI. Monitoring the internal variables is to measure the actual state of the current internal variables. This is the subjective "feeling" of the robot.

On the other hand, obtaining ΔI from the associative memory is not an actual status of the internal variables, but an estimated or predicted state of the internal variables. For example, if the robot sees an apple, and ΔI is already learned, which tells increase of "energy", then, by only seeing the apple and not eating it, the robot can feel increase of pleasantness, which results in emotion joy. Thus, the emotionally grounded symbol "apple" enables the robot to feel the experienced emotions, and to express its emotions by motions or voices.

EGO Architecture

Now, we can summarize the overall view of the EGO architecture (Fig. 4).

Basically the architecture combines the functions we described in the previous sections. The framework is the same as shown in the ethological model in Fig. 1. Now the internal variables are used to generate the emotions described in the previous section, and the associative function for 3 channels, which are visual perception, audio perception, and change of internal variables, is realized with Short-Term-Memory (STM) and Long-Term-Memory

(LTM). The association of the change of internal variables by simple object presentation can generate the emotions, which influence behavior selection and motion generation.

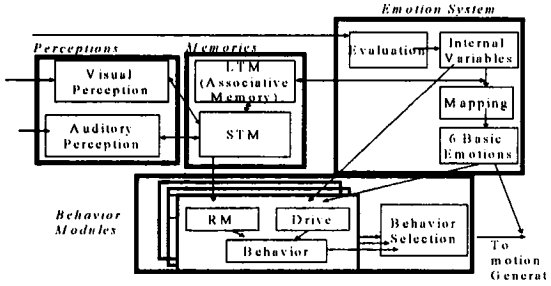

Fig. 4: EGO architecture

IMPLEMENTATION

Enhanced Four-legged Robot Platform

We implemented the architecture as described in the previous sections on the four-legged robot system shown in Fig. 5. The robot is an enhanced version of Sony's autonomous four-legged robot platform, which has a wireless LAN card to communicate with workstations on the Ethernet through the TCP/IP. The embedded CPU board uses a MIPS R4000 series RISC CPU with 100MIPS performance and 16Mbytes main memory.

We implemented a set of primitive behaviors including basic posture transition, search for an object, track an object, approach an object, kick an object, eat an object, and so on. We also implemented a speech synthesizer by which the robot can speak natural language (Japanese), and an LED-flashing pattern for facial expressions.

All of the input signals such as image, joint angles, servo-pulses, values of acceleration meter, etc, are transmitted from the robot to the workstation via the wireless LAN except for the audio signal, for which we use the microphone signal of the workstation.

Fig. 5: The enhanced four-legged robot system

The Perception, Emotion system, Behavior Generator and Association Memory are all implemented on the workstation.

Implemented Functions and Experimental Results

Visual Perception

We use single colored objects as the targets to learn. An unsupervised color segmentation algorithm is used to extract color patches from a YUV-image. We use normalized 2 dimensional color space, (Nu, Nv) = (atan(U/Y) ,atan(V/Y), to compare an input pixel with prototype color vectors. In our implementation the comparison doesn't need to evaluate the function atan(), but computes the projection to the prototype color vectors.

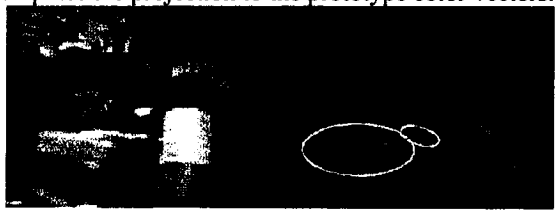

Fig. 6: Color Segmentation and Shared Attention

Shared Attention

Pixel regions with skin color are further processed to detect a hand with a finger, so that we can implement shared attention as shown in Fig. 6 to signal the target of the learning process to the robot. Shared attention not only allows the robot to recognize the finger pointed object, but also allows the human to recognize the robot's visual tracking of the object. It is not implemented, but if the robot could also finger point an object, it would help to construct shared attention more easily.

Note that there is no special function of shared attention in the architecture. Interaction of human and robot behaviors (finger-pointing) is a key to construct shared attention.

Audio Perception

Regarding auditory signal perception, we use a Hidden Markov Model (HMM) for continuous speech recognition developed at our laboratory, which has a Japanese phoneme dictionary. Since our first goal is to aim at acquiring the names of objects, we register some verbs such as "kick" and "stop". If it acquires a "ball", then we can apply the command of "kick a ball"

For unknown word acquisition, we use another HMM, where all phonemes are connected with each other so that all possible phoneme sequences should be captured in the HMM. If a phoneme sequence is input to both HMMs, the ordinary HMM evaluates the likelihood functions for possible HMM candidates. The added HMM generates the applied phoneme sequence. If the maximum value of likelihood of the registered HMM is too small, then the system determines that the input is a new phoneme sequence. Then, the generated HMM by the added HMM is registered with the ordinary HMMs.

Emotional Model

The implemented internal variables are (1) enrgy-1 for battery charging behavior, (2) fatigue for taking a rest behavior, (3) enrgy-2 for simulating-eating behavior, (4)

excrement for simulating-defecate behavior, (5) energy-3 for simulating-drinking behavior, (6) urine for simulating-urinate behavior, and (7) information for information acquisition behavior.

Physically and Emotionally Grounded Symbol Acquisition

In this subsection, we add an explanation of the Information Acquisition Behavior subsystem. The drive for this subsystem generates a motivation value when the corresponding internal variable moves out of proper range.

$$DriveVal = 1 - \tanh(IntVal - \theta) + \varepsilon 1$$

where DriveVal is a value applied to a corresponding behavior module, *IntVal* is a value of an internal variable, and θ is a threshold parameter for the proper range, and ε2 is a small positive value for the behavior. A release signal is generated by checking object information. If there is no information in the name or the ΔI field, a release signal is generated. The release signal value is in proportion to the number of the fields of no information(#nilfield).

$$RMsig\,1 = \alpha \times (\#\ nilfields),$$

where α is a proper positive number.

In addition, in order to incorporate finger-pointing into the Information Acquisition Behavior, it is checked to see if the object is marked by the finger-pointing.

$$RMsig\ \ 2 = \beta \times (\#\ fpmark\ )$$

where (#fpmark) is 1 if there is a marking of finger-pointing, otherwise 0, and β is a proper positive number. *RMsig2* is designed as larger than *RMsig*1, because finger-pointing is usually a request from a human, which should be treated as an important interaction for a robot. This is a requirement from the design side.

The final release signal for the object is

$$RMsig = RMsig\ 1 + RMsig\ 2 + \varepsilon 2$$

where ε2 is a small positive value. The *RMsig* is evaluated for each object and selects the object with the maximum *RMsig* value.

The behavior value for a subsystem is calculated by multiplying *RMsig* by Dri*veVal.*

$$Behval = RMsig \times DriveVal$$

For all subsystems, the *BehVals* are calculated and the subsystem with the maximum *Behval* is selected to execute. There are several mechanisms to sustain this selection for a while, however, we omit the explanation in this paper (See [1] for a more complete description).

The selected subsystem is now evaluated in the next level, the Mode layer. For the Information Acquisition Behavior, there are two modes, which occur when the subsystem is released by finger-pointing or an unknown stimulus. Then, in the lowest level, named the Module layer, more evaluations are performed.

Examples of Physically and Emotionally Grounded Symbol Acquisition Behaviors

Here is an example of dialogue of physically grounded symbol acquisition. Note that a real dialogue is in Japanese.

- H(uman): Finger-points an object that is unknown for a robot.
- R(obot): Looks at the object. (Shared attention is constructed)
- H: "Wha is this?"
- R: "I don't know"
- H: "Apple"
- R: "Is this apple?" (Is this an apple?)
- H: "Yes"
- R: "O.K. I got it"

After a while,

- H&R: construct shared attention to the object by finger-pointing behavior.
- H: "What is this?"
- R: "I know. This is apple"

There are some different versions depends on the situations. Here is an example of emotionally grounded symbol acquisition

- R: searches or just finds an unknown object, which has not been associated with ΔI.
- R: tries to apply some behaviors such as eating, drinking, and kicking, in order to complete the association with ΔI. Assume that "increase of simulating energy-2 (food)" is obtained by the learning.

After a while,

- R: finds the object. Assume that the robot is hungry.
- R: selects eating behavior to increase the simulating energy-2, which makes the emotion system joy.
- R: eats the object.

Thus, at the first time, the robot couldn't know that the object is a food (to increase the simulating energy-2). But after learning, the robot can select the eating behavior by only visual presentation.

RELATED WORKS

Regarding our emotion system and our concept of emotionally grounded symbols, there are important related works in Kismet project[7], Blumberg's project [27], and Velasquez's emotion model[26]. Kismet and Blumberg's projects use affective tags and associate them with objects, which is inspired Somatic Maker Hypothesis. Their association between objects (external stimuli) and emotions are done in 3 dimensional emotion space. On the other hand, our model makes association between the object and changes of internal variables, which are mapped to the reduced 3 dimensional emotional space.

Velasques also discussed emotional memories, in which the object is directly associated with the emotions though releasers. In our model, the variance of internal variable is associated with an object, and the variance generates emotions.

Regarding shared attention, we implement only finger-pointing behavior that is developed following the detection of gaze and face direction in cognitive science studies[21][22]. Details of the comparison and limitation of our architecture are described in [13].

CONCLUSION

In this paper we have presented an architecture capable of emotionally grounded symbol acquisition that aims to be the basis of an Open-ended System that can learn beyond pre-recorded behaviors. Some preliminary experiments using a 4-legged robot have also been reported.
We incorporate physically and emotionally grounded symbol acquisition into an autonomous behavior architecture based on an ethological model, by introducing a new internal variable. This makes it possible for the robot to generate *Information Eating Behaviors* such as searching unknown stimuli, closing to it, including a dialogue such as asking its name, as homeostatic activity. Furthermore, the important problem of symbol acquisition in the real world, namely Shared Attention, is naturally implemented by attentive autonomous behaviors in the architecture.

ACKNOWLEDGMENTS

We thank Dr. Arkin at the Georgia Institute of Technology for his discussion of the Ethological Model, and Dr. Luc Steels and Dr. Kaplan at CSL Paris for their discussion of Language Acquisition.

REFERENCES

[1] Arkin, R.C., Fujita, M., Takagi, T., and Hasegawa, R. Ethological Modeling and Architecture for an Entertainment Robot, in proc. of ICRA-2001

[2] Arkin, R.C., Homeostatic Control for a Mobile Robot: Dynamic Replanning in Hazardous Environments, Journal of Robotic Systems, Vol. 9(2), March 1992, pp. 197-214.

[3] Bates, J. The nature of character in interactive worlds and the oz project. Technical Report CMU-CS-92-200, Carnegie Mellon University, Oct. 1992

[4] Blumberg, B. Old Tricks, New Dogs: Ethology and Interactive Creatures, Ph.D. Thesis, MIT Media Laboratory, 1996

[5] Bonasso, R. P., Huber, E. and Kortenkamp, E, Recognizing and Interpreting Gestures within the Context of an Intelligent Robot Control Architecture, Embodied Language and Action, AAAI Fall Symposium, 1012, 1995

[6] Bruner, J. Learning how to do things with words, in J. Bruner and A. Garton (Eds.) Human growth and development, Wolfson College Lectures, Claredon Press, 1978

[7] Breazeal, C. : Robot in Society: Friend or Appliance? Agents99 workshop on emotion-based agent architectures, Seatle WA, pp.18—26 ,1999

[8] Damasio, A. Descartes' Error: Emotion, Reason, and the Human Brain, Putman Publishing Group, 1994

[9] Devroye, L., Gyorfi, L. and Luqosi, G., A Probabilistic Theory of Pattern Recognition, Springer, 1996

[10] Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, John, Wiley, and Sons, 1973

[11] Ekman, P. and Friesen, W.V. Unmasking the Face, Prentice-Hall, 1975

[12] Fujita, M. and Kitano, H., Development of an Autonomous Quadruped Robot for Robot Entertainment. Autonomous Robots, 5, 7-18, Kluwer Academic Publisher, 1998

[13] Fujita, M. et. al., An Autonomous Robot That Eats Information via Interaction with Humans and Environment, IEEE RoMAN, 2001

[14] Halliday, T. R. and Slater, P. J. B(eds). Animal Behavior, Blackwell Scientific Publications, 1983

[15] Hori, T., Brain and Emotions (in Japanease), Kyouritsu Shuppann, 1991.

[16] Iwahashi, N. and Tamura, M., "Spoken language acquisition based on the conceptual pattern analysis in perceptual information", 1999

[17] Kaplan, F. Talking AIBO: First experimentation of verbal interactions with an autonomous four-legged robot. In proceedings of the CELE-Twente workshop on interacting agents, October, 2000

[18] Pook, P and Ballard, D, Deictic Telesaaistance, in proceedings of IROS-94, 245—252, 1994

[19] Rolls, E.T., Emotions, pp. 325—344, Japan Sci. Soc. Press/Karger, 1986

[20] Roy, D. and Pentland A. Learning words from natural audio-visual input, in proceedings of International Conference on Spoken Language Processing, 1998

[21] Scaiffe, M. and Bruner, J., The capacity for joint visual attention in the infant, Nature, 253, pp.265--266

[22] Scassellati, B., Imitation and Mechanisms of Joint Attention, in Computation for Metaphors, Analogy, and Agents LNCS 1562, pp. 176—195, Spronger-Verlag, 1999

[23] Steels, L. Perceptually Grounded Meaning Creation, In proceedings of the International Conference on Multi-Agent Systems, 1996,

[24] Takanishi, A. An Anthropomorphic Robot Head having Autonomous Facial Expression Function for Natural Communication with Human, The 9th International Symposium Robotics Research (ISRR), 297—304, 1999

[25] Turner, J. H. On the origins of human emotions, Stanford University Press, Chapter 3

[26] Velasquez, J. Modeling Emotion-Based Decision-Making, in Emotional and Intelligent: The Tangled Knot of Cognition AAAI FS Technical Report FS-98-03, pp164—169, 1998

[27] Yoon, S-Y. Blumberg, B., and Schneider, G., Motivation Driven Learning for Interactive Synthetic Characters, in proceedings of International Conference on Autonomous Agents,