Grounding Cyber Information in the Physical World with Attachable Social Cues

Hirotaka Osawa
Keio University and PRESTO JST
Kanagawa, Japan
osawa@ayu.ics.keio.ac.jp

Kentaro Ishii
The University of Tokyo
Tokyo, Japan
kenta@sakamura-lab.org

Seiji Yamada
National Institute of Informatics
Tokyo, Japan
seiji@nii.ac.jp

Michita Imai
Keio University
Kanagawa, Japan
michita@ics.keio.ac.jp

Abstract

Unpredictable user behaviors in a physical process represent one of the fundamental obstacles to the realization of a Cyber-Physical System. In this paper, we propose the use of social cues such as body shape, expressions, and verbal timing to control user behaviors in the physical world. Social cues can control user behaviors in both their spatial and temporal aspects. As a result, user actions become more predictable in a CPS. We consider how social cues restrict user behaviors by referring to a number of psychological, cognitive, and human-robot interaction studies, and we propose a model of restriction based on social cues. Using this model, we created hardware and software in order to realize attachable social cues, and we seek to demonstrate the effect of social cues using the example of home appliances.

1. Introduction

Today's rapid development of information technologies and sensors such as parallel computing, distributed computing, and ubiquitous computing allows us to handle a large amount of information from the physical world, such as smart grids and intelligent transportation systems. The strong connection between physical information and cyber information requires us to create an appropriate computing framework, in a paradigm that is referred to as Cyber-Physical Systems (CPS) [1].

On a CPS, we need to handle the connection between the cyber world and physical processes more carefully. A CPS attempts to manage a combined computing model that includes both cyber and physical processes. Cyber information has several useful features, such as controllability, predictability, and reproductivity. The physical world, on the other hand, is filled with a great deal of unpredictable information. This means that a CPS needs to be able to confront problems that are rooted in the unpredictability of the physical world, such as unrestricted locations and disordered information.

One of the great difficulties in physical processes is undoubtedly that of human behavior. Computers can handle information in the cyber world almost perfectly, and it is possible to manage certain physical processes using simulation models. However, it is difficult to estimate how a human will behave in the system because human thoughts are the most unknown elements in computing, as has been shown in artificial intelligence studies. Users behave differently according to their gender, age, culture, inner contexts, and any number of other unpredictable hidden states. To include user behaviors in a predictable form is one of the most important challenges in a CPS.

We propose the use of social cues such as bodily expressions, social attitudes, and response times in order to resolve the uncertainties involved in the physical process of a CPS. Social cues could make it possible to regulate user behaviors by placing restrictions in space and time. Let us consider the example of an audio interface. An audio interface is driven by its user anytime and anywhere, even if there are no social cues. However, if the system has an interface that operates using social cues, then the user will be obliged to give a voice response to the these social cues. The user will then speak to and listen to the system, which would be based upon human timing.

1533-2306/11 $26.00 © 2011 IEEE
DOI 10.1109/RTCSA.2011.66

These social cues would restrict the user's actions to a limited sphere of activity and duration.

In this study, we create regulatory tools that ground cyber information in the physical world by using attachable social cues, and we demonstrate the effect of our approach in a training method for a home appliance. We also explain the hardware and software implementation of this tool. This tool regulates the user's behaviors and contributes a better explanation of cyber information to users in the physical world.

The rest of the paper is organized as follows. Section 2 explains the difference between the approach of this study and that of a previous study, and it offers examples of spatial and temporal constraints that are given by social cues. Section 3 explains the model of constraints. Sections 4 and 5 explain the implementation of hardware and software to realize attachable social cues. Section 6 presents the results of our demonstration and discusses how our system works. Section 7 concludes our work.

2. Related Studies for the effect of social cues

Several studies in the fields of engineering and psychology have tried to regulate human behavior through the use social cues. Picard proposed the use of emotion as an interface with computers in the concept of affective computing that he introduced [2]. Norman also showed the importance of emotional cues in the design of an interface [3].

Our proposed approach is based on their works, and succeeds in realizing their ideas more thoroughly. In this study, we focus especially on the possibilities of manipulating user's spatial and temporal behaviors.

2.1. Related studies of spatial and temporal constraints achieved by social cues

Human behaviors are controllers that range from a higher level of prediction to a lower level of prediction. It is possible to estimate lower levels of prediction with the use of several algorithms. For example, the Markov process model and particle filters make good approximations of locations for moving objects, including human movements [4]. However, it is still impossible for a computer to estimate a higher level of user behaviors that are caused by inferences in the human brain.

Human-robot interaction studies have revealed that a human-like virtual agent or robot can regulate user actions and decrease the uncertainty involved in human behaviors. For example, Kuzuoka et al. found that a robot's head and body can regulate the standing position of a user. Imai et al. succeeded in shifting a user’s attention by making use of the theory of joint attention in the field of cognitive science [5]. Shiomi et al. showed that the success rate in guiding a robot is influenced by the robot's directions [6].

What about the time constraints that change in response to social cues? Generally, users prefer a quick response of a system. Several studies have shown that the system response needs to be under two seconds [7][8]. However, an appropriate social cue extends the user's preference about response time. Shiwa et al. showed that a human-like robot makes it possible for users to engage their imagination in a thinking time of longer duration. This in turn creates a preference for an extended response time among users [9]. Verbal fillers can also serve to maintain response times.

2.2. Availability of social cues in computing

The framework for the handling of social cues in computing has not been well examined, in spite of the possibilities mentioned in the previous subsections, because social cues are used in closed agents, such as virtual agents and robots. In this approach, every form of advantage in regard to social cues is centralized in the anthropomorphic agent.

However, this is a lost opportunity, in that social cues are only used within the anthropomorphic agent. Reeves and Nass discovered an innate human tendency to assume an anthropomorphic attitude [10]. For example, if a user performs a task on a computer and then evaluates its performance, the user's evaluation becomes more favorable if the working computer and the asking computer are the same. On the other hand, the user’s expression becomes more objective if the working computer and asking computer are different. This tendency is found not just in beginners but also among computer experts. This result shows that humans have an intrinsic anthropomorphic attitude that accepts environmental information. This work shows that human behaviors are restricted even by slight social cues such as an interactive interface. We think that a perfect anthropomorphic appearance and social cues can have an overwhelming influence on setting restrictions for users.

3. Model for user's spatial and temporal attitudes

Imagine that a computing system is implemented in a room and it helps the user's achieve a task in response to the user's verbal input. If there are no monitoring sensors in the room, the computing system cannot determine where the user is. This means that the distribution of the user's location is constant, as shown on the left side of Fig. 1.

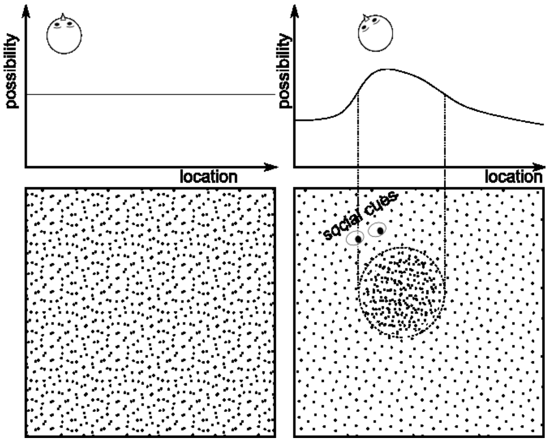


Fig. 1. Spatial restrictions given by social cues

However, if the room has social cues such as human-like eyes and gives responses to the user, it is reasonable to think that the user will be directed to be in front of the social cues even in the absence of sensor monitoring. The right side of Fig. 1 shows how spatial restriction operates upon the user. We can use several psychological theories to estimate how social cues restrict user location, such as Hall's social distance theory [11] and interaction with a baby-like agent. For example, personal distance is about 45cm circle area. If the object pronounce to the user with several social cues like "let's watch my head," the system can estimate that the possibility of user's position after reply is bigger in less than 45cm distance from the object.

Temporal restrictions can also be at work in this situation. If there are no social cues in the system, the user accepts the system as a tool. Dennett calls this human attitude toward the tool a design stance [12]. If the user attends to the system in the design stance, then the user expects a quick response. As a result, the user's allowance for a response delay becomes like that shown in Fig. 2. However, if the system expresses a number of social cues, the user considers that the system has its own intentions. Dennett calls this attitude an intentional stance. If the user interacts with the system with an intentional stance, the user considers that it is more natural to have sufficient thinking time to understand the commands given. A previous study has shown that this allows for a one-second delay of the system, as shown in Fig. 2 [9]. This result directly helps to increase computing time in a system.

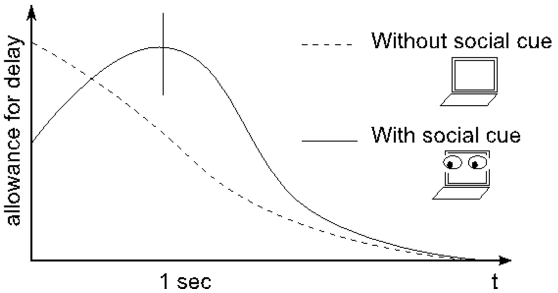


Fig. 2. Temporal restrictions given by social cues

We might ask then, what types of elements are required to realize social cues in the physical world? Studies of the human anthropomorphic tendency have reported that several elements, such as body shape, emotional clues, and verbal clues are important to the realization of social cues [13]. In the next section, we initiate the realization of these elements through the implementation of hardware and software.

4. Hardware implementation

We have created several robotics parts and sensors in order to realize spatial and temporal restrictions that are given as social cues, as was discussed in the previous section. The total components that create social cues toward the physical world is similar to the physical robot like communication robots and humanoids. However, these parts must be variant according to attached target space.

To focus on the more general availability of social cues, we implemented all parts separately. For example, separate eyes and arms make it possible to attain different scales of social bodies.

4.1. Total components

Our hardware components are a laptop with a camera, body parts, sensor modules, and visual markers. Content author uses the computer where authoring software is running. The author inputs interactive contents by using the software. Body parts are main components for giving social cues. We implement eye- and arm-like parts as shown in Fig. 3 and Fig. 5. Sensor modules enable to recognize listener’s response such as opening a door and pushing a button.

On performing phase, body parts are controlled based on interactive contents created by the author and listener’s response acquired by sensor modules. Visual markers are used to specify pointing targets (Fig. 9). Each component is wirelessly connected using ZigBee communication.

We use an off-the-shelf laptop and webcam, while we develop body parts, sensor modules, and visual markers. Body parts, sensor modules, and visual markers can be attached on any flat surface with adhesive gum. Body parts come with visual markers to calculate relative position to pointing targets. The visual markers on body parts are used only on authoring phase, thus they can be removed on performing phase.

4.2. Bodily parts

Bodily biological cues gives strong presence to the agent's body as shown in our previous study [14]. The human eye (1) enables vision and (2) indicates what a person is looking at [15].

Eye-like body parts contain a small OLED that displays eye patterns. Each eye pattern is used to express emotions and point some location. Fig. 3 shows their appearance, and Fig. 4 shows implemented eye patterns. These patterns are based on Ekman's categorization [16]. Arm-like body parts have four servo motors with skeletal structure formed by modeling machine. They are also used to express emotions and point some location. Fig. 5 shows their appearance. Eye- and arm-like parts can be controlled via ZigBee connection.

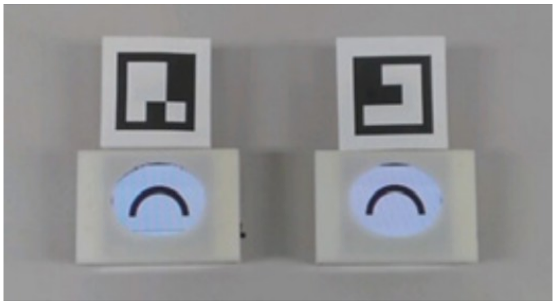


Fig. 3. Appearance of eye-like body parts

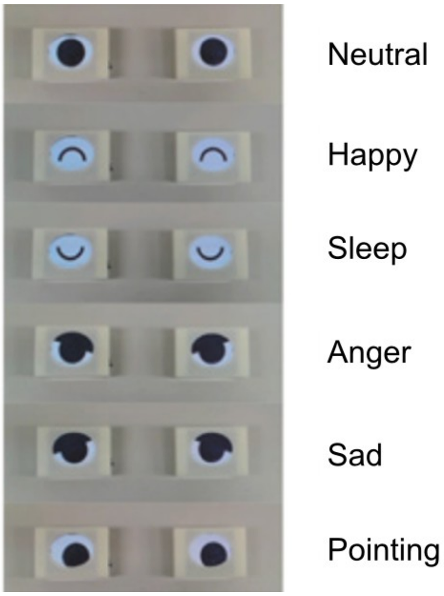


Fig. 4. Eye pattern by eye-like bodily parts

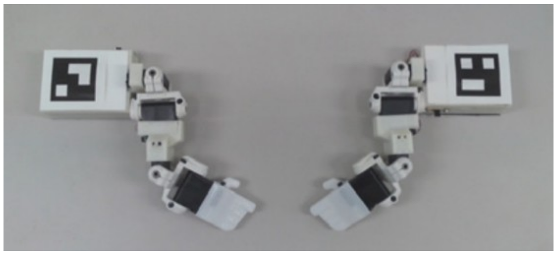


Fig. 5. Appearance of arm-like bodily parts

4.1. Sensor parts

Inputs from sensors are also essential to achieve interaction between the system and the user. Appropriate input and reactions strengthen social cues.

In this study, we used three kinds of sensor modules: light sensor, bend sensor, and touch sensor. Each sensor module periodically obtains and sends data from the sensor. We intend to use light sensors to detect form change of the physical object such as opening a door (Fig. 6). We intend to use bend sensors to detect physical contact between parts of the physical object such as a telephone and a receiver (Fig. 7). We intend to use touch sensors to detect listener’s touch action to the physical object such as pushing a button (Fig. 8).

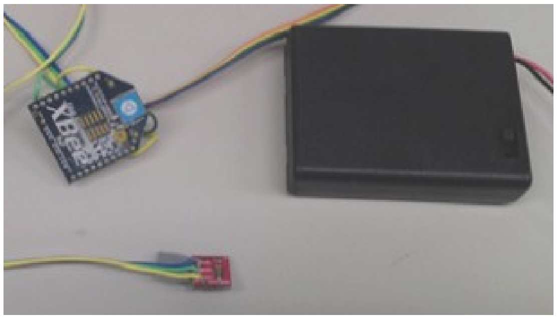


Fig. 6. Light sensor module

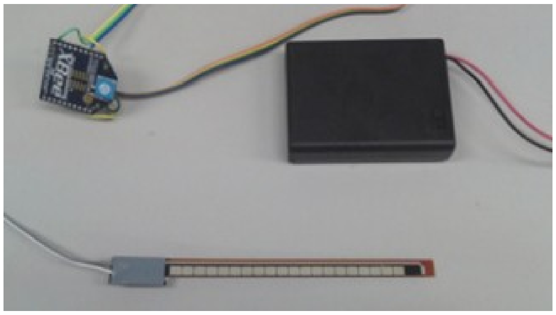


Fig. 7. Bend sensor module

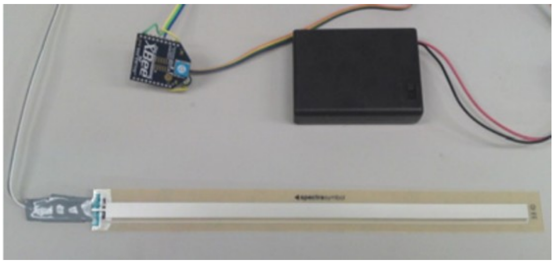


Fig. 8. Touch sensor module

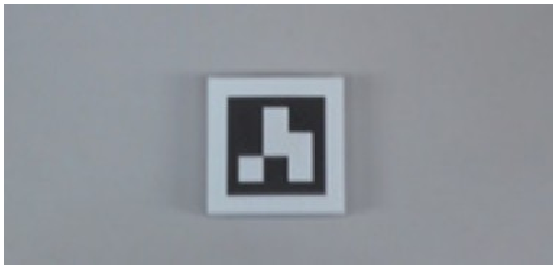


Fig. 9. Visual marker

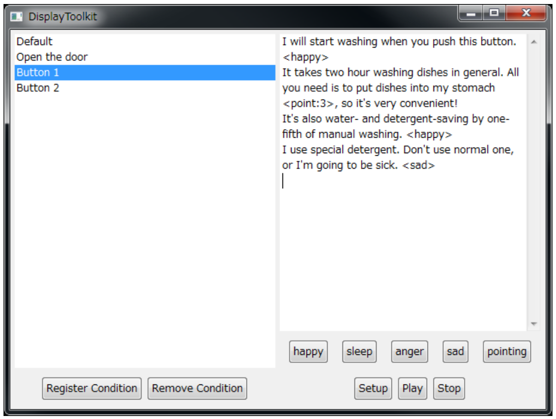
5. Software implementation

We created authoring software to regulate above social cues. This software makes it easier to create interactive responses according to the determined scenario toward the target.

The authoring software is comprised of three modules: content input module, content performing module, and visual marker recognition module. The content input module is main component for the author to create interactive contents. The content performing module controls speech and body parts based on interactive contents created by the author and sensor values from sensor modules. Visual marker recognition module processes image acquired by the camera and recognizes visual markers.

5.1. Content Input Module

With the content input module, the author inputs conditions of content switching, utterance contents, and action of body parts. Figure 10 shows a screen shot of the content input module. The left pane lists conditions of content switching. Utterance contents and action of body parts of the selected condition are displayed in the right pane.



I will start washing when you push this button.
<happy>
It takes two hour washing dishes in general. All you need is to put dishes into my stomach <point:3>, so it's very convenient!
It's also water- and detergent-saving by one-fifth of manual washing. <happy>
I use special detergent. Don't use normal one, or I'm going to be sick. <sad>

Fig. 10. Screen shot of content input module

Each condition of content switching is identified by sensor value received from sensor modules. The author registers by keeping sensor state as content switching and pressing the "Register Condition" button. For example, the author can register open and switch by leaving the door opened and pressing the "Resister Condition" button. When the author presses the "Remove Condition" button, the selected condition is removed.

An action of body parts is incorporated in utterance contents as a tagged word in the right pane. We prepare five tagged words corresponding to eye expression shown in Fig. 4: <happy>, <sleep>, <anger>, <sad>, and <point:#>. The tagged word <point:#> takes one parameter that specifies ID of the pointing target. The ID of the target is presented in the visual marker recognition module. The author needs to check the vision marker recognition module to see IDs of the visual markers.

The content input module has buttons named "happy", "sleep", "angry", "sad", and "pointing." If the author presses these buttons, the corresponding tagged word is inserted at the cursor position. The "Setup" button is used to memorize marker positions. When the button is pressed, the relative positions of the markers are automatically calculated. The "Play" button is used to start the current interactive contents, and the "Stop" button is used to stop the interactive contents. The author can iteratively edit interactive contents after checking to perform the contents.

5.2. Content Performing Module

The content performing module manages speech, body movement, and content switching. This module plays speech using a text-to-speech engine based on the author's input. When a tagged word appears in the sentence, this module sends a control signal to make a body expression. In case sensor data matches pre-registered switching condition, this module stops speech and starts another interactive content corresponding to the condition.

5.3. Visual Marker Recognition Module

The visual marker recognition module recognizes visual marker pattern in the camera image (Fig. 9). Visual markers consist of 3x3 binary patterns in a square. This module binarizes the camera image and finds square blobs in the image. After that, it decodes ID of each square blob based on the 3x3 binary patterns.

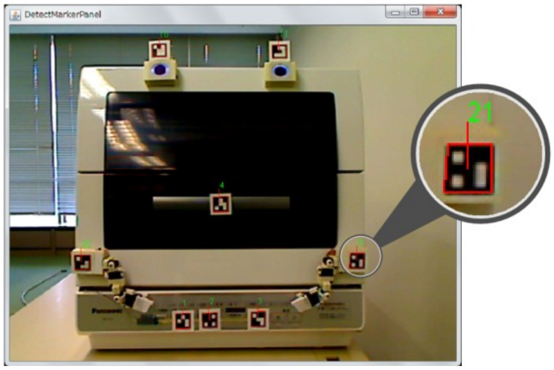


Fig. 11. Visual marker recognition module

6. Demonstration and Discussion

We demonstrated our implementation in two conferences (shown in Fig. 12). In the demonstration, we attached these parts to a refrigerator, a shredder, a printer, and a dishwasher. We attached touch sensors on front of the buttons and detected user's touching motions. We also inserted light sensors and bend sensor toward the appliances' doors to detect users' opening motion. We also created interactive contents to explain the features of appliances.

During these demonstrations, more than hundred participants interacted to our system. The result of demonstration gives us several findings.

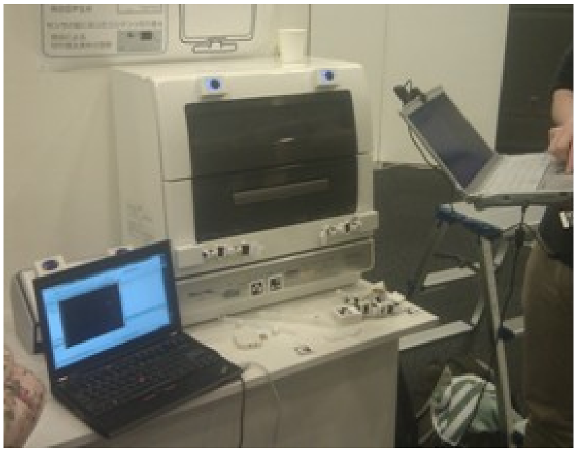


Fig. 12. Demonstration

Although, replies from the users are very different according to their features, we can find the general tendency of the users' responses. First, the participants interacted in front of the attached home-appliances. This means that our model in Section 3 is partially worked in direction. We could not measure how close they approached to the target.

Second, we find several clues that the participant's verbal reply timing toward the target is also influenced by social meaning in the content. When the target respond with greetings, the participants replied same verbal content toward the target object (like hello, good morning, good evening). It is very simple and ordinary interaction. However, it means that the social cues forces the user to repeat same pronunciation. For example, we can calibrate the auditory interface's parameters like direction and pitch control by this returning voice.

7. Conclusion

In this paper, we propose the use of social cues such as body shape, expressions, and verbal timing to control user behaviors in the physical world. Social cues can control user behaviors in both their spatial and temporal aspects. As a result, user actions become more predictable in a CPS.

We consider how social cues restrict user behaviors by referring to a number of psychological, cognitive, and human-robot interaction studies, and we propose a model of restriction based on social cues. Using this model, we created hardware and software in order to realize attachable social cues, and we seek to demonstrate the effect of social cues using the example of home appliances. The demonstration suggests that our approach will extends to decrease unpredictability of a user's behaviors. Our model gives the researchers arm to conquer users' capricious behaviors in computing and contributes to find more applications in CPS.

In future, we will make more precise models based on the clues found by this study. We are planning to use these attachable social cues to several CPS applications in home electronics like health care. We also planning to make better tools to achieve attachable social cues.

8. Acknowledgements

This work was supported by the JST PRESTO program.

9. References

[1] Lee, E. A., "Cyber-Physical Systems - Are Computing Foundations Adequate?", Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap, 2006

[2] Picard, R.W., *Affective Computing*, The MIT press, MA, 1997.

[3] Norman, D.A., *Emotional Design: Why We Love (or Hate) Everyday Things*, Basic Books, 2003.

[4] Thrun, S., Burgard, W., Fox, D., *Probabilistic Robotics*, The MIT press, MA, 2005.

[5] Kuzuoka, H., Suzuki, Y., Yamashita, J., Yamazaki, K., "Reconfiguring spatial formation arrangement by robot body orientation", Proceeding of the 5th ACM/IEEE international conference on Human-Robot Interaction, pp. 285-292., 2010.

[6] Shiomi, M., Kanda, T., Ishiguro, H., Hagita, N., "A Larger Audience, Please! Encouraging People to Listen to a Guide Robot", Proceeding of the 5th ACM/IEEE international conference on Human-Robot Interaction, pp. 31-38., 2010.

[7] Miller, R. B., "Response time in man-computer conversational transactions," Proc. Spring Joint Computer Conference, AFIPS Press, Montvale, NJ, pp. 267-277., 1968.

[8] Starner, T., "The Challenges of Wearable computing: Part 2," IEEE Micro, Vol. 21, Issue 4, pp. 54-67., 2001.

[9] Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., Hagita, N., "How Quickly Should Communication Robots Respond?", Proceeding of the 3rd ACM/IEEE international conference on Human-Robot Interaction, pp. 153-160., 2008.

[10] Reeves, B., Nass, C., *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge Univ. Press, 1996.

[11] Hall, E. T., *The Hidden Dimension*, Doubleday & Company., 1966.

[12] Dennett, D., *Three kinds of intentional psychology* in Heil, J. - Philosophy of Mind: A guide and anthology, Clarendon Press, Oxford, 2004.

[13] DiSalvo, C., Gemperle, F., "From seduction to fulfillment: the use of anthropomorphic form in design." In DPPI ’03: Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp. 67–72, 2003.

[14] Osawa, H., Mukai, J., Imai, M., "Towards Anthropomorphized Spaces: Human Responses to Anthropomorphization of a Space Using Attached Body Parts", Proceedings of 17th International Symposium on Robot and Human Interactive Communication, pp.148–153., 2008.

[15] Kobayashi, H., Kohshima, S., "Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye." J Hum E, Vol. 40, No. 5, pp. 419–435, 2001.

[16] Ekman, P., *Basic Emotions*, in Dalgleish, T; Power, M, Handbook of Cognition and Emotion, John Wiley & Sons., 1999.