Enhancing User Involvement via Joint Attention with a Listener Robot

Masashi Okamoto[*], Yoshiyasu Ogasawara[*], Yukiko I. Nakano[†], Toyoaki Nishida[‡]

[*]Graduate School of Information Science and Technology, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
{okamoto, yoshiyasu}@kc.t.u-tokyo.ac.jp

[†]Research Institute of Science and Technology for Society, Japan Science and Technology Agency
Atago Green Hills MORI Tower 18F, 2-5-1 Atago, Minato-ku, Tokyo, 105-6218, Japan
nakano@kc.t.u-tokyo.ac.jp

[‡]Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
nishida@i.kyoto-u.ac.jp

Abstract

The recent progress of technology makes computers and robots more intelligent than before. Notwithstanding, it is still difficult for people to establish natural communication with robots. People are not so willing to communicate with computers or robots as with humans. We believe that it is effective to apply our natural communication manner between a listener and a speaker to human-robot communication. Then the user of the robot will be smoothly involved in the human-robot communication. The purposes of this paper are: (1) to model a natural communication environment with computers from the cognitive perspective, and (2) to propose a method of establishing communication environment between a human and a listener robot from the cognitive model. In our method, their mutual intention is formed by establishing joint attention using the redundancy of behaviors. As a result of this, the user is expected to be involved more in the communication with the listener robot.

Keywords

Human-robot communication, join attention, User Involvement, empathy, redundancy of behaviors

1 Introduction

Nowadays, the rapid progress of technology makes computers more intelligent than before. Notwithstanding, it is difficult yet for people to establish natural communication with computers. People are not so willing to communicate with computers as with humans.

In fact, there are many examples where the natural communication is not established between humans and computers. One of the typical examples is making a video letter with a video cam. In this case the speakers in video letters often present unnatural way of speech and behavior. However, they will not do such a way of speaking when they talk to their close persons. We assume that this difference is caused by whether natural communication between a speaker and a listener is established or not.

The human listener responds to the speaker's behavior with various ways. The listener implicitly conveys his listening attitude to the speaker through his responses or gestures. These behaviors and responses establish the communication and let the speaker feel relaxed. On the contrary, in making a video letter, the video camera is placed at the listener's position instead of a human listener. Then the communication is not established because the camera does not response to the speaker at all. Therefore, the speaker feels stressed and cannot behave as usual.

The purpose of this research is twofold. One purpose is to propose a cognitive model for natural human-computer interaction where the user of a computer or a robot can naturally involved in the interaction or communication with it as with a human. If it were not for an adequate model for human-computer interaction environment, the designing and constructing of a robot would be ad hoc. We thus introduce the ‘User Involvement’ and ‘Empathy Channel’ as the key concepts for designing natural human-computer interaction environment.

The other is to build a listener robot whose natural behaviors as a listener allows its user to speak and behave in unrestrained ways. In particular, we focus on ‘joint attention’ to be realized between the user and the listener robot. As a result, it will be shown how the users get to be involved in the communication with robots effortlessly.

2 User Involvement in human-computer interaction

In this section we focus on the basis for designing and evaluating a natural human-computer interaction, in particular, human-robot communication. We believe that the design basis of natural human-computer interaction settings should be considered from the point of the user’s view. In other words, the user should be as naturally involved in the interaction or communication with a computer or a robot as he is in the real world. The ‘User Involvement’ theory is introduced here to verify the idea.

2.1 Requirements of User Involvement

The definition of ‘User Involvement’ and the main requirements to establish it are as follows (Okamoto et al. 2004):

User Involvement. The cognitive way humans willingly engage in, or forced to be involved in a virtual world which computers display, in a human-to-robot communication, or in a computer-mediated community.

Requirements.

1. Cognitive/Communicative/Social reality should be achieved.
2. Two (or more) cognitive spaces should be linked, and the user should cognitively move in and out those spaces.

In this approach the 'reality' is classified into the following three aspects:

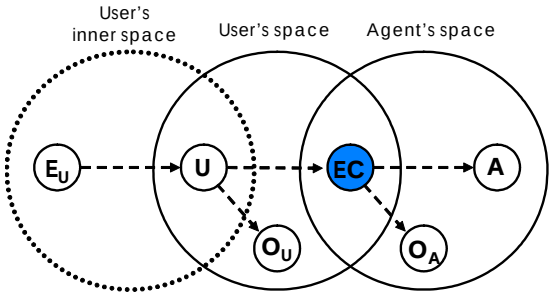
- **Cognitive reality.** The way of seeing objects, events and their relations in the real/virtual world as real.
- **Communicative reality.** The sense of reality that is achieved through communication with others.
- **Social reality.** The collective and intersubjective sense of reality based on sharing thoughts or opinions with one another.

In this research we deal with cognitive reality and communicative reality because social reality is considered to concern more broad interactive settings such a computer-mediated community as online communities on the Internet.

2.2 Astigmatic model of User Involvement

Since the User Involvement is strongly related to our sense of reality and is common to both verbal and nonverbal communications, linguistic researches help to comprehend how it works. In particular, recent cognitive lingusitc studies suggest many important characteristics of human cognition in conceptualizing the world.

Langacker (1993) points out the reference-point ability, which enables us to conceptualize an entity at a distance using a mental path from a more accessible entity as a reference point. Applying this concept to the User Involvement, it can be said that people in real world conceptualize an unfamiliar entity in another world using an ac-



E_U: ego of user. U: user's self or body. O_U: object in user's space.
A: virtual agent. O_A: object in agent's space. EC: Empathy Channel.

Figure 1: Astigmatic model for User Involvement

cessible entity as a reference point that lies in both worlds. At the time both of the worlds (i.e. cognitive spaces) need to be linked or overlapped with the reference point.

Such overlapping is not limited to the relation between the real space and a virtual space. At the very start of our life, we are living in two spaces, that is, thinking in the inner space and acting in the outer world using our body as a reference point.

Figure 1 illustrates an example of how the computer user conceptualizes respective objects in different cognitive spaces using each reference point, especially in virtual agent systems. We call this model as the *astigmatic model* for User Involvement in that multi-spaces are overlapped and linked there. For example, E_U (ego of user) locates O_U (object in user's space) using U (i.e. his self/body) as a reference point. It is one of our everyday activities in which we conceptualize something in the real world.

On the other hand, U can conceptualize O_A (object in agent's space) via a certain reference point which is cognitively accessible for the user and, at the same time, consistently functions as a constituent of the virtual world. We call that reference point 'Empathy Channel' in that the user will acquire the viewpoint of the virtual agent through empathizing with its image, gaze, behaviors, utterances and so on. In order to feel or experience the virtual world with a sense of reality, the user should be able to utilize such an Empathy Channel.

2.3 Empathy Channel

Empathy is thus another key concept for designing a natural human-computer interaction environment. According to Wispe (1991), the difference between empathy and sympathy is explained as follows:

> "…To know what something would be like for the other person is empathy. To know what it would be like to be that person is sympathy. In empathy one acts "as if" one were the other person...The object of empathy is understanding. The object of sympathy is the other person's well-being." (Wispe 1991: 80)

To put it differently, empathy is the ability to attain the other person's viewpoint toward the world. Considering the User Involvement, the user should feel and experi-

(a) USA. New York City. 1955

(b) USA. NYC. Felix, Gladys and Rover. 1974

Figure 2: The photographs of Elliott Erwitt

ence another world mediated by a computer as if he/she were there. Taking a virtual agent system for instance, the user should attain the agent's viewpoint to enter its world via a certain channel. We thus called the channel to connect the user's cognitive space and the agent's cognitive space as Empathy Channel.

The famous photographer Elliott Erwitt[1] is conscious of the effects and functions of empathy. See Figure 2. The picture (a) features a view from the head window of a train, and a boy watching the view. The picture (b) captures a small dog taken by a woman which is watching us.

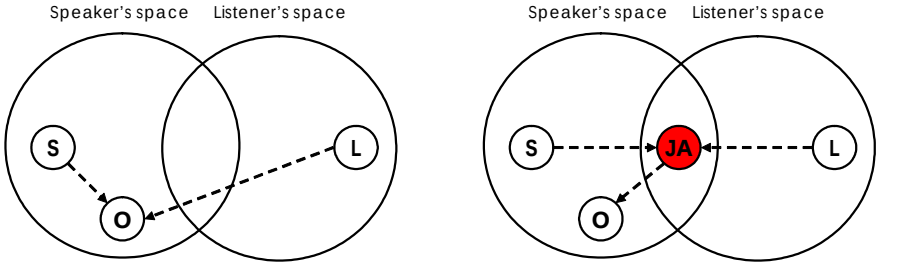
If the boy were not captured in the picture (a), the cognitive effect for viewers would be different. It is because, in watching (a), the viewers attain the boy's viewpoint via empathizing with the boy ('s back image) and feel the scenery from a train window as if they were there. We consider that it is the example for achieving cognitive reality through an empathized image as Empathy Channel.

Similarly, the viewer of the picture (b) feels an odd feeling as if he became a dog. It is assumed that the low camera angle and the gaze of a dog provoke that feeling, because both of them should originally spring from the viewpoint of a small animal that is about to communicate with the dog in the picture. We thus consider the picture (b) is a good example for establishing communicative reality via gazing at the viewer as a communicative partner.

In our approach it is also our task to find out what can be a candidate for Empathy Channel in human-computer interaction environment.

2.4 Joint attention as Empathy Channel

In this paper we focus on how to achieve the communicative reality in the main so as to enhance the User Involvement by implementing natural responses and reactions as a good listener into the listener robot. For it is difficult to fully establish the cognitive reality using a robot, in that current humanoid robots do not have so sufficient appearances or facial expressions as to make humans feel as if they the robots were living. Moreover, as Reeves and Nass (1996) point out, humans are likely to behave toward artefacts as if they were humans. Therefore, if a robot reacts to humans in unnatural ways, then they will assume it is churlish and non-cooperative and will not



(a) Communication without joint attention (b) Communication with joint attention

Figure 3: Joint attention as Empathy Channel

[1] http://www.elliotterwitt.com/

keep communicating with it.

From the view of the User Involvement, it can be said that respective participants in communication originally lie in a different cognitive space before the communication begins. But, once the communication starts, there has to be something to connect those cognitive spaces, which functions as a reference point (i.e. Empathy Channel) for one participant to access another. Then the communicative reality for the participant is achieved.

In human communication, Empathy Channel is that which is cognitively shared in the participants, such as exchanging verbal information, establishing eye contact, matching action and reactions, joint attention, and so on. We believe that those factors that enable human communication will be applied to human-robot communication as well.

Figure 3 illustrates how the joint attention functions as Empathy Channel in communication. A speaker can attend to an object without consideration of a listener. If a listener then accidentally attends to the same object, no joint attention is established (Figure 3a). On the contrary, if the speaker intentionally and explicitly pays attention to an object, the listener can attend to it via the speaker's attention behavior. Then the joint attention is achieved, and the listener can attain the speaker's viewpoint through the joint attention (Figure 3b). In other words, the listener sees the object in the way the speaker does and recognizes what the speaker is feeling toward it.

We thus focus on the 'joint attention' in particular in order to establish natural human-robot communication environment. Specifically, the speaker-listener communication using a listener robot is described in the following sections.

3 Requirements for natural speaker-listener communication

3.1 Mutual intention in speaker-listener communication

A speaker and a listener keep exhibiting their behaviors, such as gestures or utterances, toward each other in the communication. In view of the User Involvement, it might be interpreted to overlap each cognitive space for establishing an Empathy Channel to access the other's cognitive space.

During the process, one participant's intention is approaching that of the other, and then the mutual intention is being formed. The mutual intention here means rather the 'intentionality', which Dennett (1987) suggests, than the 'intention' as used in a usual context. For the mutual intention concerns what is aimed both by the speaker and the listener.

Take the explanation task for communication example. The state of the communication environment changes depending on what is being explained, or what is being attended to. When applying the model to the human-robot communication with a listener robot, the robot needs to change their behaviors according to the state of the communication environment. Therefore, in order to decide the robot's behavior at each moment of the communication process, it will be necessary to fully analyze what comprises the given domain of explanation. In other words, building a listener robot concerns the verbal/nonverbal behavior model of a speaker and a listener in their communication environment.

3.2 Joint Attention

As shown in the previous section, joint attention is the interaction where a speaker and a listener cognitively share an object that they attend to. For instance, the listener looks at the thing the speaker points at.

Tomasello (1999) suggests that joint attention is divided into the following three types:

- **Check Attention:** Attention to the partner, or the object he shows.
- **Follow Attention:** Attention to the object that the partner points at or his eye gaze turns to.
- **Direct Attention:** Making the partner pay attention to the object to which a listener attends, with voice, eye gaze, and so on.

In human communication, these interactions are executed unconsciously and make the participants feel the communication natural. Thus, joint attention is also one of the important requirements to establish communicative reality between the speaker and the listener.

In view of the User Involvement, the joint attention leads humans to be mutually involved in the same cognitive space in the communication, which is the overlapped part in both participants' cognitive spaces. As a result, the joint attention enables the listener to access the speaker's attention object through the joint attention as Empathy Channel, and vice versa. Therefore, joint attention achieves the smooth transition among the cognitive spaces of the participants, which helps to establish the communicative reality for them.

3.3 Social skills for communication

We also propose the idea to use the social skills, as a policy for establishing more advanced communicative reality. Social skill is the theory to establish human relations effectively in view of the communication techniques. Aikawa (2000) describes the

Capable Pose	No interruption, No rush
Open Question	Prompting, Explaining in details
Reflection	Verbal response, Repeat, Paraphrase, Summary
Using Non-verbal Channel	Posture, Gaze, Nodding, Distance, Hand's movement
Decoding Speaker's Non-verbal Channel	Voice (pause, speed, pitch), Emotion, Gaze, Hand's movement

Table 1: Skills for listening to speech

social skills for listening to the partner's speech. Table 1 summarizes them. Among these skills, response and nodding are important for building the listener robot. Many people have experienced that a listener's nodding frequently improves the rhythm of a speaker's speech. While it is difficult for robots to understand the contents of the speech, it is possible for them to behave as if they were listening to the speaker through using these skills.

4 Establishing joint attention between a speaker and a listener

4.1 Establishing joint attention

The correspondence of the listener's reaction with the precedent actions of the speaker is necessary for the communication between a speaker and a listener to be established. It is through this repetitive process that natural speaker-listener communication is achieved. Among the correspondent action and reaction couplings, the most fundamental and effective one is seemingly the joint attention.

In order to achieve the joint attention, the listener has to recognize the speaker's attention and its target from the observation of the speaker's behaviors, and should react to the attention appropriately and instantly. We thus classified the speaker's behaviors according to each modality, and observed what can be attention behaviors.

Moreover, it is also essential for the joint attention to be established whether the attention behaviors are put out by the speaker intentionally or not. For reacting against non-intentional behaviors would become unnatural in communication. Therefore, we proposed the method of detecting the intensity of the speaker's intention based on the redundancy of the behavior.

4.2 Modality of attention behavior

Table 2 summarizes attention behaviors in view of each modality. A listener and a speaker use these types of behaviors to express their own attention or to interpret the other's.

In the Table 2, Hand movements and Eye gaze are considered to strongly represent attention. If a listener robot is able to recognize such attention behaviors of the speaker, then the establishment of joint attention will be easier. The other types of behaviors are also used for representing attention. However, they are often used with

Modality	***Behaviors***
Hand movements	*Pointing, Grasping, Showing*
Eye gaze	*Direction (head, eyes)*
Posture	*Approaching, Direction (body), Standing-up, Sitting-down*
Speech	*Direction words, Verbal response, Mentioning*

Table 2: Modality of attention behavior

a certain degree of redundancy. The redundancy of attention behavior is described in the next subsection.

4.3 Redundancy of attention behavior

Attention behaviors are frequently represented with more than one modality or in repeating fashion. Such redundant manners strongly suggest that those behaviors are intentional. In other words, the redundancy of behavior manner helps a listener to recognize the speaker's intention.

It is assumed that there are the following two types of redundancy in the speaker-listener communication:

- **Redundancy of modality:** Using multiple modalities of behaviors simultaneously
- **Redundancy of time:** Using the repetitive or persistent behaviors

Since the redundant behaviors are often intentional, communicative robots should react rapidly and appropriately against those intentional behaviors of humans. Conversely, the robots also can convey strong intention to humans using redundant actions or behaviors. This effect of redundancy is discussed in Tajima (2004) through the human-robot communication experiment.

A human can adjust himself to a robot with ease if the robot can interpret those redundancies. It is often observed that a human behaves in a redundant way when he cannot communicate with others smoothly in such a situation as noisy surroundings and teaching restless students. Thus, even the robots with insufficient recognition abilities can communicate with a human as long as he is willing to use the redundant ways of communication.

5 Constructing a Listener Robot

In this section, we describe the implementing method of the attributes of establishing natural speaker-listener communication environment, suggested in Section 4, into the listener robot.

5.1 Target

In this research, we construct a listener robot, which is to participate in such human-robot communication environment as the situation where a human explains the procedure of assembling a piece of furniture or an appliance to the listener robot.

Since assembling the parts of furniture is usually done in sequence with several tools, its procedure, each placement of the parts, and which tool is needed when, become the major topics on which the speaker should put his emphasis respectively. The speaker thus explains about such topics while drawing the listener's attention to them with pointing gesture, eye gazes, or voices.

This explaining task regarding the ways of assembling furniture contains the basic behaviors that are usually used by a speaker and a listener. Therefore, the following

discussions on that task will be applied to the problems of many other explanation tasks.

5.2 Hardware

In constructing a listener robot, we use Robovie[2], a humanoid robot. Robovie is nearly as tall as a human and can move its hands, head, and eyes. Additionally, it can make a move with its wheels.

The motions of the speaker's body or the tools he uses are recognized by the motion capture. The markers of the motion capture are attached to the speaker's head, arms, body, and the objects to use or to point at in his explanation.

5.3 The communication modes

As described before, the behaviors of a listener robot should be precisely changed corresponding to what is being explained or what is being attended to.

We thus design a listener robot so that it can change its behaviors according to what the speaker is attending to, which we call the communication mode. The communication mode is classified into the following four types (also see Figure 4):

- Talking-*to* mode
- Talking-*about* mode
- Confirmation mode
- Busy mode

In general, the speaker engaged in the explanation task attends to either the listener or the target to be explained. In the '**talking-*to***' mode, the speaker is mainly watching the listener and is involved in the cognitive space based on the relation between the speaker and the listener. As the speaker in the talking-*to* mode expects the listener to be involved in the same conversation, the listener should pay attention to the speaker

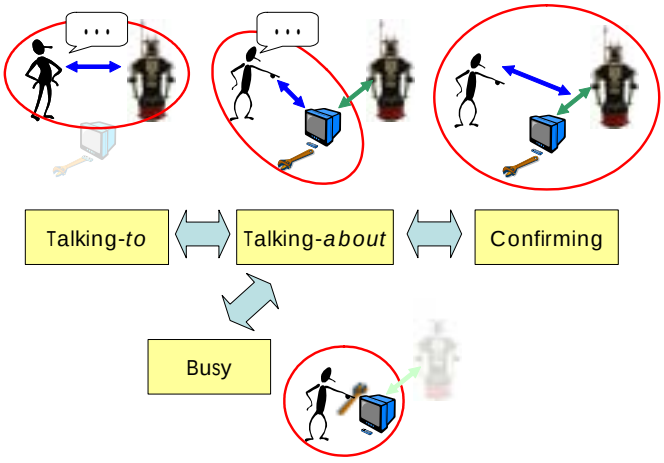


Figure 4: The modes of communication

[2] http://www.mic.atr.co.jp/~michita/everyday-e/

himself.

On the other hand, when the speaker is mainly watching the target to be explained, he is in the '**talking-*about***' mode. In this mode, the speaker expects the listener to cognitively share the target. Therefore, the listener should attend to the target in turn.

Additionally, in the cases where the speaker switches his gaze between toward the listener and toward the target, he is interested in the relation of the listener and the target, that is, whether or not the listener is paying attention to the target. We thus call the mode as the '**confirming**' mode in that he tries to confirm the listener's attention. In the confirming mode, the listener should sensitively react to the behaviors of the speaker. It is because the speaker's confirmation might be motivated by the significance of the current information flow.

The '**busy**' mode is the mode when the speaker is devoting himself to his work without talking to his listener. In this mode the speaker is attending more to the target than in the talking-*about* mode, and tends to ignore the listener. This situation is not favorable for any explanation task, but frequently occurs when the speaker is not skillful. A listener robot can keep attending to the target during the busy mode, which is one of the advantages for using robots.

These modes of the speaker's attention differ in which cognitive space the speaker attends to. The speaker creatively uses each mode to establish communicative reality. If the listener robot does not appropriately react according to the current communication mode, the speaker will not be able to smoothly switch these modes, and then he will not feel the communication with the robot as real.

5.4 Architecture of listener robot system

The architecture of the listener robot is shown in Figure 5.

The robot receives motion capture data of the speaker's behaviors and the speech sound as input. As the result of processing the data, the robot outputs bodily expressions as feedback to the communication environment.

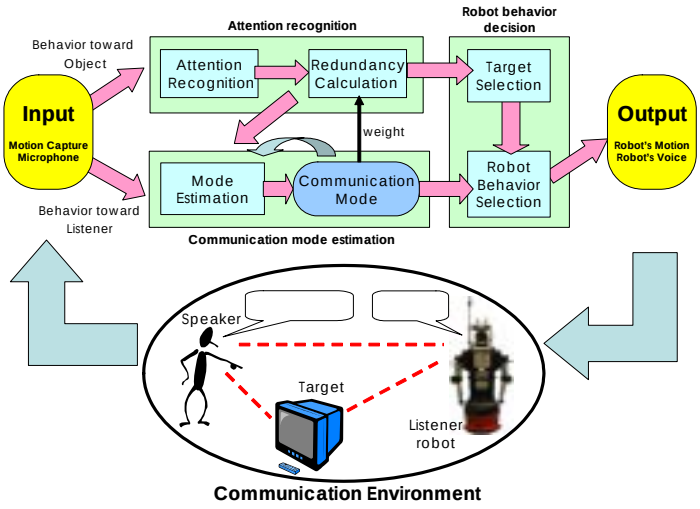


Figure 5: Architecture of listener robot

- **Input:** Motion capture data, and speech sound.
- **Output:** Robot behaviors (head, eyes, arms, head nod, voice, and body direction)

The following subsections describe the methods used in each module in the robot system architecture.

5.4.1 Attention recognition

In order to recognize the attention behaviors of a speaker, the listener robot processes each data from respective markers attached to all the possible targets. The details of the process are described below.

First, the confidence for attention toward each target is calculated from the positions of the motion capture markers to decide the following behaviours:

- Eye gaze (head direction)
- Pointing
- Grasping
- Repetitive hand gestures (e.g. tapping)
- Physical relationship with objects (distance, body direction)

Respective behaviors are recognized based on simple calculation of such as distance and angles between markers. The precise direction of eyes is difficult to recognize from motion capture information. Therefore, we are not concerned it in this listener robot.

Repetitive gesture of hands is calculated by employing methods proposed by Tajima (2003) and Anuchitkittikul (2004). Their method detects human's repetitive gestures by using an autocorrelation function and the Fast Fourier Transform (FFT).

The calculation of the confidence for respective gestures is based on the relational network of sensor data and behaviors. Outputs from this process are calculated by using the method proposed by Hatakeyama (2004). This theory applies Bayesian Network to the processing of sensor information and the decision of robot's behaviour. In this method, inputs and outputs are represented as random variables. The causal relationships between inputs and outputs are expressed as a conditional probability table.

This method has an advantage over other methods that describe the relationship between input and output as rules, because it is more robust against the noisy input and the changes of communication environment.

Furthermore, the confidence values of the respective behaviors are regarded as scores, and then the scores are weighed and added up into the redundancy of attention behaviors toward the target (i.e. Redundancy of modality). Then the weighing scores depend on the current communication mode and the duration of behaviors (i.e. Redundancy of time). For example, in the Talking-*to* mode, weighed values are less than in any other mode because gazing attention to the target does not frequently occur in the Talking-*to* mode.

5.4.2 Communication mode estimation

The communication modes suggested in 5.3 are classified according to what relation

is established between the speaker, the listener and the target. Our listener robot decides each communication mode based on the speaker's behaviours.

Which communication mode the speaker and the listener are involved in is estimated from the following inputs using the similar algorithm to that of attention recognition:

- Target attention (with the largest redundancies)
- Gaze to the listener robot
- Body direction toward the listener robot
- Attention behavior (e.g. posture change)
- Speech (volume, duration)
- Preceding communication mode

The general conditions for recognizing each communication mode are as follows:

- **Talking-*to***: with frequent speech, speaker's body and face toward the listener
- **Talking-*about*:** with attention behaviours and gaze to the target
- **Confirming:** with momentary gaze toward the listener during target attention
- **Busy:** with no speech, with continuous attention to the target

5.4.3 Robot behavior decision

The listener robot's behaviors are decided by if-then rules based on the redundancy of attention behavior and the communication mode. The object with the largest redundancy is selected as the target for the listener robot's attention.

The manners of the listener robot's behaviors are summarized as follows:

- **Talking-*to* mode:** turns the head to the speaker and randomly nods when the speech is aborted.
  - **Low redundancy:** *no attention*
  - **Medium redundancy:** *gaze attention in a short period*
  - **High redundancy:** *slight head move and gaze attention in a short period*

- **Talking-*about* mode:** uses various methods.
  - **Very low redundancy:** *no attention*
  - **Low redundancy:** *gaze attention*
  - **Medium redundancy:** *head attention after gaze attention*
  - **High redundancy:** *prompt head attention*
  - **Very high redundancy:** *head move and verbal responses*

- **Confirming mode:** takes a glance at the speaker, nods with verbal responses, and then returns to the original state. Nodding and verbal responses are randomly produced.

- **Busy mode:** in the same fashion as talking-*about* mode, but with no utterances.

6 Conclusions

6.1 Discussion

We built a listener robot system that enables joint attention with a user as speaker, which is based on the recognition of the redundancy of behaviors and the communication mode transition.

In the present state, the recognition accuracy is not sufficient especially in the case of low redundancy of behaviors observed. We are thus trying to apply a learning algorithm to the recognition module for acquiring adequate parameters. Additionally, the manners of the robot behavior still remain ad-hoc though they are based on some preliminary analysis of actual human-communication data. We believe more data analysis will be helpful to decide respective adequate responses for the listener robot.

6.2 Evaluation as Empathy Channel

We also have to evaluate how effectively the joint attention established by a listener robot functions as Empathy Channel for communicative reality in human-robot communication environment and enhances the User Involvement of the user.

The astigmatic model of the User Involvement suggests what will be required in the Empathy Channel:

(1) The high accessibility in one's space
(2) The consistent functioning in the other's space
(3) The continuous correspondence of behaviors in both spaces

The reason for (1) is that the Empathy Channel should be highly accessible to a participant in one space to function as the reference point. Moreover, since the Empathy Channel also lies in the other space, it should consistently function as a constituent of the space to maintain the reality of that world, so it is the reason for (2). At the same time, the Empathy Channel connects the two cognitive spaces simultaneously. If its behaviors in each space are not corresponding to each other, the connection will be broken, so (3) will be required, too.

6.3 Toward evaluating the joint attention for the User Involvement

Therefore, in evaluating how much the joint attention established by our listener robot enhances the User Involvement, these three aspects of Empathy Channel should be taken into consideration.

First, the accessibility of the joint attention should be evaluated. It will be estimated by asking the speaker whether or not the joint attention with the listener robot is easily recognized. On the other hand, it should be also estimated how quickly the listener robot can recognize the speaker's attention behaviors.

Secondly, it is attested that the joint attention of the listener robot seems consistent with the other listening behaviors for the speaker. Conversely, it should be clarified how precisely the listener robot can detect the speaker's attention behaviors from other unintentional behaviors.

Lastly, it should be observed that the correspondences between the speaker's action and the listener robot's reactions are naturally established in a sustainable fashion.

6.4 Summary and future plans

In this paper, we introduced the User Involvement theory to realize and evaluate the natural communication environment with computers, and then proposed the method to establish the natural human-robot communication environment with a listener robot. In particular we developed the listener robot that can establish joint attention with a speaker.

In the near future, we will conduct an experiment to measure psychological burden of the user in the explanation task toward the listener robot. Then, we will develop the system supporting creation of the video contents using this listener robot. If this robot's behavior makes a speaker relaxed, the speaker is expected to be able to create good video contents.

References

Mitsuru Aikawa. *Techniques of Human Relation –Psychology of Social Skills (in Japanese)*. Number 20 in Selection of Social Psychology. Science-sha, 2000.

Burin Anuchitkittikul, Masashi Okamoto, Hidekazu Kubota, and Toyoaki Nishida. Gestural interface for the creation of personalized video-based content. In *Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004)*, China, 2004.

Cynthia Breazeal and Brian Scassellati. A context dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, 1146–1151, Stockholm, Sweden, 1999.

Makoto Hatakeyama. Human-Robot Interaction based on Interaction Schema (*in Japanese*). Master's thesis, Graduate School of Information Science and Technology, The University of Tokyo, 2004.

Ronald W. Langacker. Reference-Point Constructions. *Cognitive Linguistics* 4: 1-38, 1993.

Masashi Okamoto, Yukiko I. Nakano, and Toyoaki Nishida. Toward enhancing User Involvement via Empathy Channel in human-computer interface design. In *Proceedings of IMTCI*, 2004.

Byron Reeves and Clifford Nass. *The Media Equation: How People treat computers, television, and new media like real people and places*. CSLI Publications, 1996.

Takashi Tajima and Toyoaki Nishida. Manual-less interaction based on synchronization and modulation (*in Japanese*). In *Proceedings of IEICE Human Information Processing*. The Institute of Electronics, Information and Communication Engineers, 2003.

Michael Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, 1999.

Tomio Watanabe and Hiroki Ogawa. InterRobot for human interaction and communication support. In *Proceedings of world Multi-conference on Systems, Cybernetics and Informatics (SCI2001)*, 466–471, 2001.

Lauren Wispe. *The Psychology of Sympathy*. New York: Plenum Press, 1991.