NAIST-IS-DD0461042

Doctoral Dissertation

Home Robot Interface Innovated by Familiar Behaviors

Daisuke YAMAMOTO

February 1, 2007

Department of Information Processing
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to the Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Thesis Committee:

Masatsugu KIDODE, Professor (Supervisor)
Tsukasa OGASAWARA, Professor
Shigeki SUGANO, Professor (Waseda University)
Yasuyuki KONO, Associate Professor

親和行動を導入したホームロボットインタフェース*

山本 大介

内容梗概

近年，ブロードバンドの普及に伴い家庭内でも情報家電をはじめとするホームネットワークの構築が進んでいる．さらに，これらの情報家電と各種センサが融合したユビキタスネットワークに関する研究も盛んである．しかし，これら情報家電のユーザインタフェースは，必ずしも使いやすいものとは言えない．それは各家電が，例えば，冷蔵庫の在庫を示したり，電子番組表(EPG)と連携し番組表を見ながら番組を選んだり，と多くの便利な機能を持ち，様々な使い方ができるようになる一方，その多様さがゆえに，ユーザの欲する機能を実行するために複雑な入力が強いられるためである．この問題を解決するために，情報家電とユーザとの仲立ちとなり，簡単に操作できるインタフェースとして，音声で応答し親しみやすい身体性を持つロボット用いたインタフェースが注目されている．

しかし，これには2つの課題がある．一つはロボット側の音声認識技術の課題である．家庭環境の多様な騒音に加え，ロボットの音声認識ではマイクと発話者の位置関係が定まらず音響特性は変動する．さらに子供から老人まで幅広くかつ不慣れなユーザに対応する必要があり，その誤認識は避けられない．もう一つは情報家電側の課題である．従来からあるリモコン等の手で操作する入力に代わり，音声入力を用いた音声リモコンの研究がされている．しかし個々の操作は容易になるが，多くの機能を持った情報家電を音声で扱うには，予め決められた音声指示語を覚えなければ使えず，ユーザには大きな負担となる．

*奈良先端科学技術大学院大学情報科学研究科情報処理学専攻学位論文，NAIST-IS-DD0461042，2007 年　月　日．

本研究では，間違っても許され次第に言葉を覚える幼児を参考にした親和行動を、ロボットに導入することでこの解決を図っている．”間違っても許され”については，発達心理学に基づく親和行動（発話・注視・移動）を実装した．そして一般の被験者28人による評価実験を行い，幼児を参考とした親和行動が誤認識を許容させる効果があることを確認した．

”次第に言葉を覚える”については，ユーザが家電を操作した際に，幼児のようにその意味を聞くことで，音声指示語を獲得する方法を実装した．そして，一般被験者12人での実験により，次第に音声指示語を誤認識も含めて獲得し，操作率が従来方法を上回り，指示語のゆらぎと認識結果のゆらぎを含めて獲得することで，操作率も向上することを確認している．また，印象評価の結果から親和性も向上していることが分かった．

これらの効果は高齢者（６０代）に対して特に高く，情報家電など複雑な機械を敬遠しがちな高齢者向けインタフェースとして，本インタフェースが有効であることが示唆された．

キーワード

ロボットインタフェース，幼児の親和性，誤認識の許容，音声指示習得，
高齢者向けインタフェース

Home Robot Interface Innovated by Familiar Behaviors*

Daisuke YAMAMOTO

Abstract

Recently, robotic interfaces have been attracting attention in view of their potential to serve as intermediaries between complicated information equipment and users. For the robotic interface, the user wants to operate the robotic interface by voice instruction that is easy to input. However there are two problems. One is that the voice recognition is difficult especially at home and by inexperienced users. Another is that the voice instructions are decided beforehand. I propose to make the robot act like an infant. The user will accept recognition error by the robot. And the user will be able to teach the user-defined instructions, naturally.

At first, I innovated by infant's familiar behaviors to the robotic interface. The effects of the familiar behaviors installed in a robotic information home appliance, "ApriAlpha", are studied. The experimental results show that the task execution with speech repetition, object gazing and moving to object has the highest score for the familiar factor. The familiar factor the case in which the user orders to the robot with the familiar behaviors and it makes a mistake, and in case the user orders to the television and it operates correctly was almost equal. Therefore, it checked that the introduced the familiar behaviors make a tolerant of the error recognition. Moreover elderly people tend to prefer familiar behaviors more than young people do.

For the 2nd problem, I developed the learning method like an infant. This method is

.*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0461042, ???, 2007.

when the unknown operation for the robot is carried out, the robot ask the meaning of the operation to a user. Then, the robot memorizes the user's word and the operation related with the word. Next, if the user orders the word, the robot will perform the memorized operation related with the word. I installed this method to the ApriAlpha, and evaluated the effect by the subject experiment. As a result, the robot memorizes the user-defined instructions automatically by this method. It enables the user to operate it at a high operation rate as the robot memorizes the instructions.

In conclusion, that introducing the familiar behaviors in character with an infant makes the tolerance of the robot's mistakes, and it is useful for the interface that is possible with natural word.

Keywords:

Robotic Interface, Infant's Familiarity, Tolerance of Mistakes, Learning of Word, Interface for Elderly People

Contents

List of Figures

List of Tables

1. Introduction

In recent years, home network systems are being introduced in line with the progress of broadband technology. A home network system connects home appliances and information equipment in the home(Fig.1.1). The home appliance having many convenient functions is composed for a user to variously utilize.

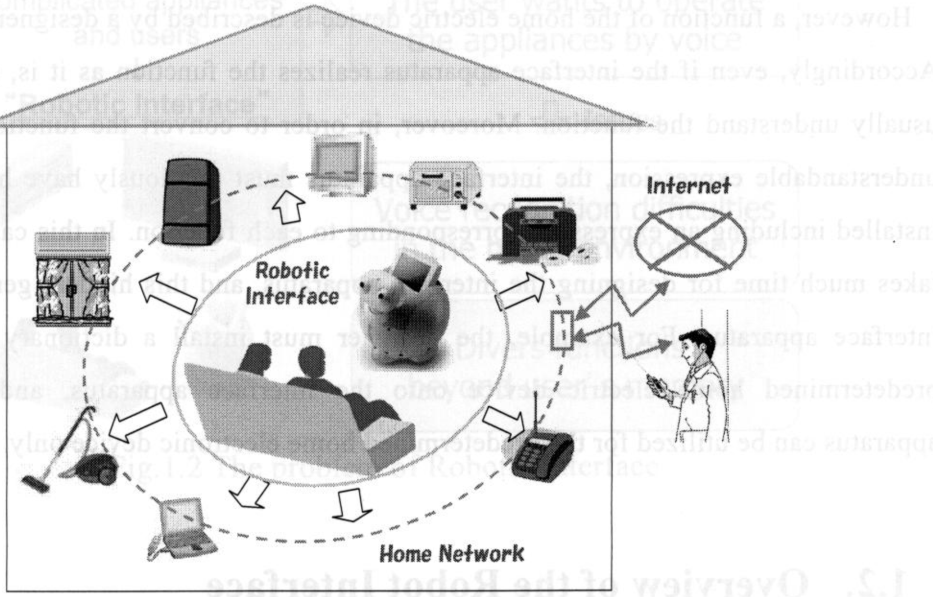


Fig.1.1 Home Environment in the Near Future

1.1. Overview of the Information Home Appliance

Although the expectation is that people will be able to enjoy a convenient and comfortable quality of life by using these networked and advanced home appliances, at present the user interfaces of these apparatus are complicated and not human friendly. The user must select his/her desired function from many functions. In case of not finding the desired function, it is inconvenient for the user. Generally, the difficulties are particularly acute for the elderly.

Accordingly, up to the present, an interface apparatus between the information home appliance and the user is taken into consideration. For example, an interface apparatus preparing a dictionary in which the user's operation by speech or gesture and the meaning contents of the operation are edited by each user. Furthermore, by using a directory service such as UPnP (Universal Plug and Play) or UDDI (Universal Description Discovery and Integration), an apparatus for presenting functions of the home electric device to the user (client device) is known. In case that a plurality of client devices cooperatively has a plurality of functions, the client device for presenting contents of this cooperative function to the user is researched. In this case, the client device functions as the interface apparatus.

However, a function of the home electric device is described by a designer's original rule. Accordingly, even if the interface apparatus realizes the function as it is, the user cannot usually understand the function. Moreover, in order to convert the function to the user's understandable expression, the interface apparatus must previously have had a dictionary installed including an expression corresponding to each function. In this case, the designer takes much time for designing the interface apparatus, and this hinders general-use of the interface apparatus. For example, the designer must install a dictionary related with a predetermined home electric device onto the interface apparatus, and this interface apparatus can be utilized for the predetermined home electronic device only.

1.2. Overview of the Robot Interface

Recently, robotic interfaces have been attracting attention in view of their potential to serve as intermediaries between complicated information equipment and users. Until now, industrial robots perform tasks precisely according to the instructions, pet robots move and act like dogs, cats, or cute bears. Secretary robots and delivery robots in offices can deal with limited tasks interactively, and carry letters and documents to specified places in offices. All of these robots, however, can move around autonomously and communicate with human beings by limited media interactions. Since current robots convey a certain feeling of familiarity and also demonstrate a set of friendly behaviors, these robots are expected to be used in the home.

For the robotic interface, the user wants to operate the robotic interface by voice instruction that is easy to input. However there are two problems (Fig.1.2). One is that the voice recognition is difficult especially at home and by inexperienced users. Another is that the voice instructions are decided beforehand. I propose to make the robot act like an infant. The user will accept recognition error by the robot. And the user will be able to teach the user-defined instructions, naturally.

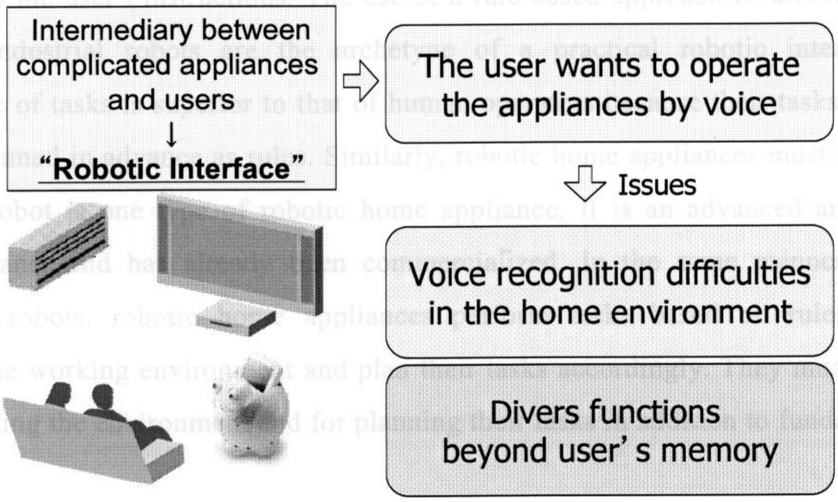


Fig.1.2 The problem of Robotic Interface

1.3. Outline of the Dissertation

In chapter 2, the concept of the familiarity for a home robot interface is discussed. As familiarity, I investigated familiarity of an infant (1-2 years old). This familiarity has the capacity for mistake and the curiosity for getting new words. In chapter 3, the tolerant interface for error recognition approach based on infant's familiarity is described. I implemented a home robot interface and studied how to install it in a robotic information home appliance, "ApriAlpha^{TM}". In chapter 4, the learning user-defined method is explained. This method is how a robot gains a user's spoken instruction word so that an infant may memorize language. In chapter 5, the experiment to evaluate the effect of familiar behaviors described in chapter 3 and the operation rate of user spoken words described in chapter 4 is evaluated. As a result, introducing the familiar behaviors similar to those of an infant increases tolerance of the robot's errors and it is useful for a natural language interface.

2. Familiarity of Home Robot Interface

2.1. Conventional Robot Interface

The practicality of a robotic interface is an embedded function to execute tasks exactly according to the user's instructions. The use of a rule-based approach results in a practical interface. Industrial robots are the archetype of a practical robotic interface. Their performance of tasks is superior to that of human operators because their tasks are defined and programmed in advance as rules. Similarly, robotic home appliances must be practical. A cleaner robot is one type of robotic home appliance. It is an advanced and robotized vacuum cleaner and has already been commercialized. In the same manner as factory automation robots, robotic home appliances perform tasks based on rules and must recognize the working environment and plan their tasks accordingly. They must have rules for recognizing the environment and for planning their tasks in addition to fundamental task rules.

In this research, the fundamental knowledge for practicality is discussed in advance in the form of rules that are as abstract as possible. The insufficiency is compensated by the familiarity as described below.

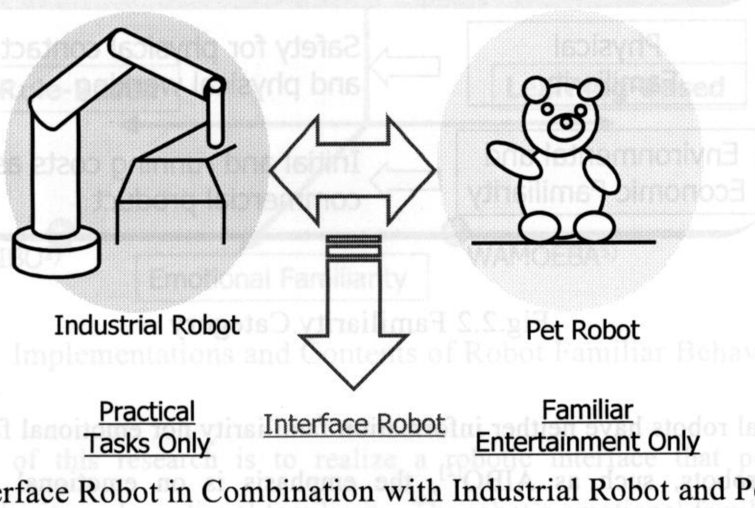


Fig.2.1 Interface Robot in Combination with Industrial Robot and Pet Robot

2.2. Related Researches

According to the report of the Committee on Human Friendly Robots, a human friendly robot requires four aspects of familiarity: information familiarity, emotional familiarity, physical familiarity, environmental and economic familiarity (Fig.2.2).[1] The last two aspects of familiarity are concerned with the robots as a product. In this research, since the robot is considered as an interface without physical work, it requires two types of familiarity, namely, information familiarity and emotional familiarity. Information familiarity is the mutual understanding between users and robots. Emotional familiarity is the situation representation. The robot recognizes the user's mental state and expresses its own state in order to enhance the user's understanding of the robot's condition.

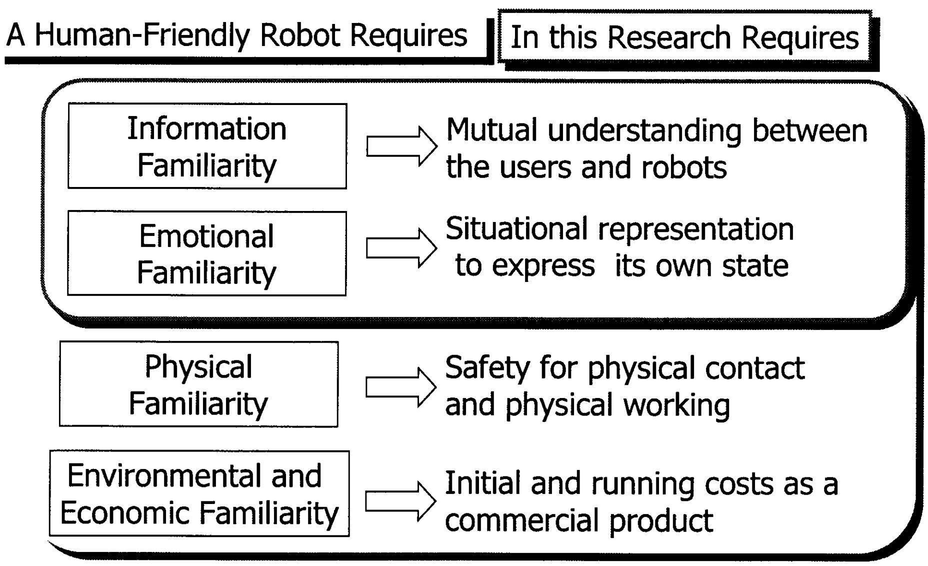


Fig.2.2 Familiarity Category

Conventional robots have neither information familiarity nor emotional familiarity. In the case of pet robots, such as AIBO[2], the emphasis is on emotional familiarity. The tail-wagging motion expresses a contact state detected by a touch sensor etc. as the robot's own mental condition. On the other hand, information familiarity is important in the case of Infanoid[3] or Robovie[4]. For mutual understanding of the user and the robot, information sharing is required. Then, the robot shares the object that both it and the user are looking at and plans to carry out information sharing.

Two methods for generating action are applied in conventional robots. One is a rule-based method which involves designing action generation rules and implementing them in advance. The other is a learning-based method which involves acquisition of the action generation method for human evolution or a development process by studying a reference. Although these have various purposes and techniques, from the authors' viewpoint, they are classified based on both information familiarity and emotional familiarity as two techniques for conventional robots. AIBO uses a rule-based method to achieve emotional familiarity. On the other hand, WAMOEBA[5] uses a learning-based method based on the concept of self-preservation. To achieve information familiarity, Infanoid uses a learning-based method, whereas Robovie uses a rule-based method.

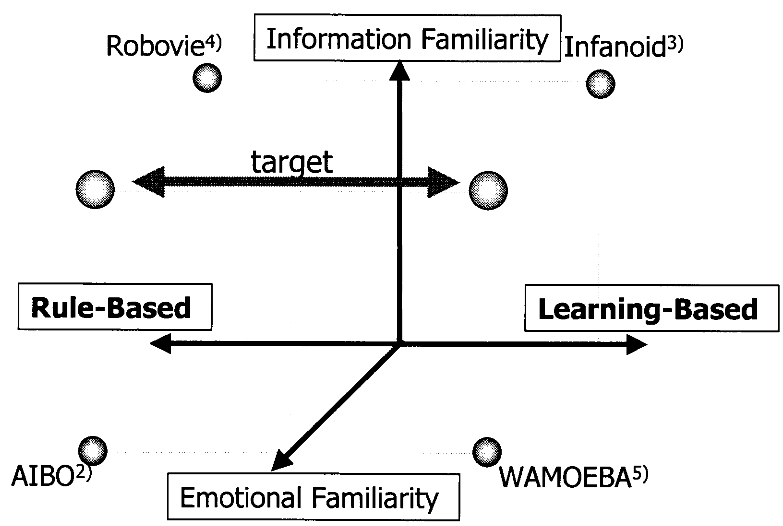


Fig.2.3 Implementations and Contents of Robot Familiar Behaviors

The objective of this research is to realize a robotic interface that possesses both information familiarity and emotional familiarity. The robot's emotional familiarity helps to clarify the user's ambiguous instructions in a voice dialog. If a robot acts like an infant, the user will give intelligible instructions to the robot in a manner similar to that of parents giving instructions to an infant. The user will accept incorrect recognition by the robot and will be willing to repeat an instruction in the same way that parents are tolerant of their infant's mistakes. Furthermore, the robot can increase information acquisition opportunities by means of emotional familiarity and can thus increase the information shared with the user. With regard to the increase in information familiarity, as an infant gradually comes to understand what his or her parents say and takes appropriate action, the robot gradually offers a suitable service in response to the user's ambiguous directions. As for the method of realizing these two aspects of familiarity, we refer to the developmental psychology of infant.

2.3. Demonstrations of Behaviors Fusion

As an example of a robotic interface that has both practicality and familiarity (information and emotion), we describe a demonstration conducted using the ubiquitous home constructed as part of the UKARI project[8],[9] (Fig.2.4).

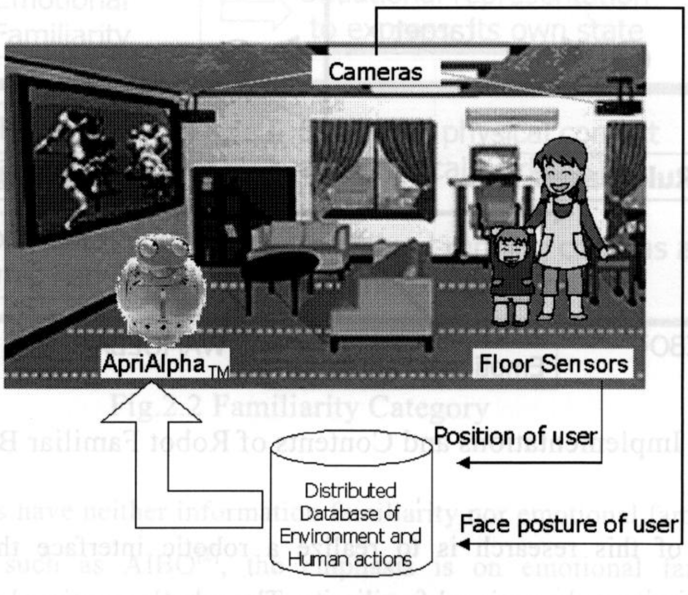
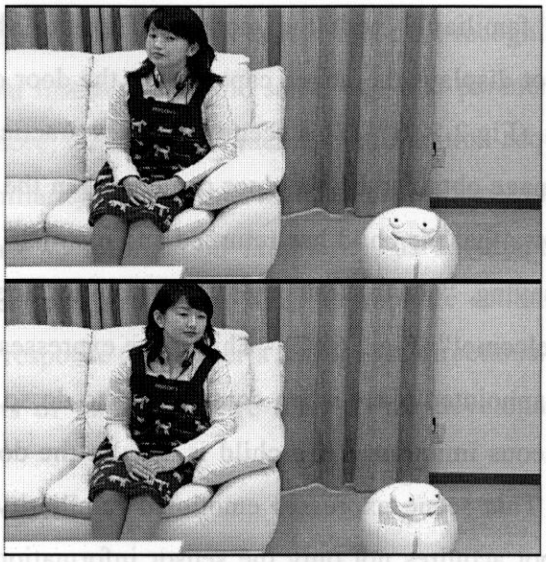


Fig.2.4 Experiment in the UKARI project

While the mother is watching TV in the sitting room, the child returns home. First, the robot (ApriAlpha^{TM}) looks in the same direction as the mother. (Fig.2.5 (A) - This scene expresses information familiarity with the mother.) Then, as the child returns and the doorbell rings, the robot displays the image captured by the door camera on its display and shows it to the mother. (Fig.2.5 (B) - This scene expresses practicality based on the rule of showing the user an image obtained by the door camera when the doorbell rings) From the image, the mother knows that the child has returned and instructs the robot to welcome the child at the door by saying, "Go to Welcome!" Accordingly, the robot goes to the door, saying, "Welcome! Welcome!" (Fig.2.6 (C) - This scene expresses practicality based on the rule of moving to the appointed place when commanded to do so. Emotional familiarity is expressed by simultaneous imitation.) If a child is seen at the door, the robot follows the child. (Fig. 2.6 (D) - This scene expresses emotional familiarity with the child.) In this demonstration, the robot acquires not only the sensor information on its own but also the sensor information from the ubiquitous home via the distributed database of the environment and human actions that have been developed for the UKARI project and use this information. For example, the robot turns its face according to the direction in which the mother looks from the environmental camera and projects the image information from the door camera. Following the child, the robot acquires the positional information of the child from floor sensors via the distributed database. In this demonstration, the effects of practicality and familiarity proposed in this research can be roughly estimated.

The above demonstration includes three fusion methods; the switch method, the parallel method, and the suppress method. (Fig.2.7) The switch method changes the practical behavior and the familiar behavior. The Fig.2.5 (A) and (B) show the switch method. While the mother is watching TV in the living room, the robot executes face tracking of the familiar behavior. The doorbell ringing is a trigger to switch the familiar behavior to the practical behavior. In the parallel method, the practical behavior and the familiar behavior are executed at the same time. The familiar behavior embellishes practical behavior and makes various. The Fig.2.6 (C) shows the parallel method. The practical moving motion to the door and the familiar speaking motion "Welcome! Welcome!" are paralleled. The suppress method is that the familiar behavior suppresses the practical behavior. The Fig.2.6

(D) shows the suppress method. ApriAlpha™ likes the child so that he forgets the practical behavior of going to the door and achieves only the familiar behavior of following the child.



(A) Looking to the same direction as mother looks

(B) Showing the door camera's view

on the attached display

Fig.2.5 Example Scene of Fusion Experiment (1)

(C) Going to the door and Saying

"Welcome! Welcome!"

(D) Following the child

Fig.2.6 Example Scene of Fusion Experiment (2)

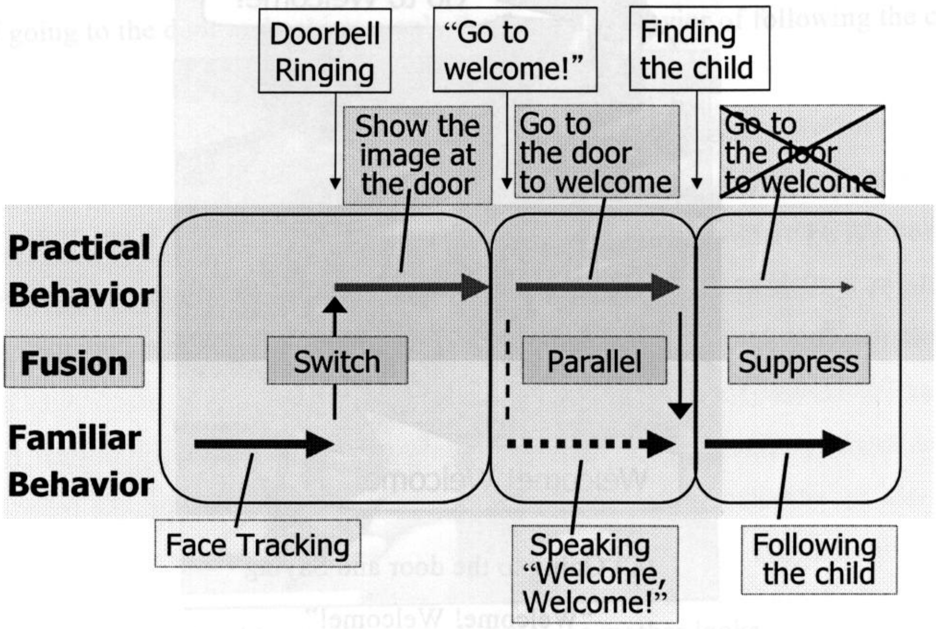


Fig.2.7 Example of Fusion Control Sequence

2.4. Familiar Behavior Innovation

2.4.1. Capacity for Mistake Recognition

The voice instruction is important for the home robot interface because it allows the user to operate the robot naturally. Using current technology, the voice recognition accuracy is insufficient for practical use on a routine basis. In the foreseeable future, voice recognition technology is unlikely to advance to the level such that complete voice recognition is realized. Thus, the home robot interface must have built-in tolerance against misunderstanding.

I refer to an infant's familiar behaviors[6] for the interface design solution. An infant's familiar behaviors elicit the parent's intelligible instructions when the infant misunderstands the parent's voice instructions (Fig2.8). An infant's familiar behaviors are as follows;

Speaking: babble, a word speaking, repeated speaking,

Gazing: face tracking, gazing to face, etc.

Moving: person following, object following, etc.

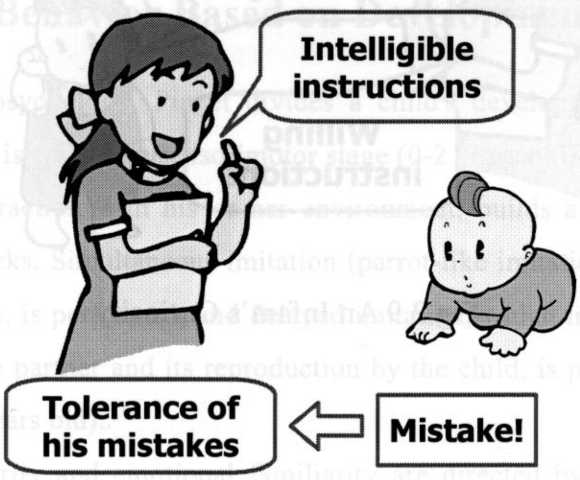


Fig.2.8 An infant's Familiarity

2.4.2. Operation by User-defined Instructions

For practical use on a routine basis, the need to achieve highly accurate recognition of voice instructions constitutes the other major issue. The verbal problem is the vocabulary. If all vocabulary is registered, highly accurate voice recognition is realized. The user's vocabulary in everyday use includes diverse abbreviations, buzzwords and so on. So it is unrealistic to register all vocabulary in everyday use.

Our solution is increasing the vocabulary to learn the user's words using home appliances. In such an approach, the curiosity of an infant is effective (Fig.2.9). An infant is full of curiosity regarding new things and new experiences. And an infant is a keen questioner. Parents want to come to let an infant know spontaneously to the question. I utilized the characteristics of an infant in the present work.

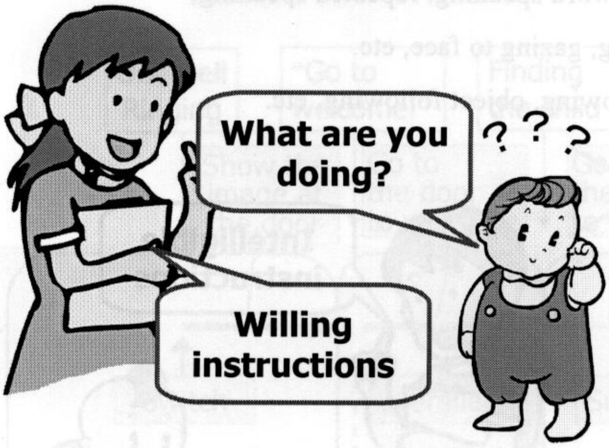


Fig.2.9 An Infant's Curiosity

2.5. Summary

In this chapter, the difference from the conventional robot was indicated to be the home robot interface described in the preceding chapter. And the hypothesis that familiarity and practicality were important was built up for the realization. The following chapter, by utilizing this familiarity describes solution of two subjects.

3. Tolerant Interface Error Recognition by Familiar Behaviors

3.1. Familiar Behavior Based on Developmental Psychology

In developmental psychology, Piaget divides a child's development process into four stages. The first stage is called the sensorimotor stage (0-2 years old). In this stage, the child, through physical interaction with his or her environment, builds a set of concepts about reality and how it works. Simultaneous imitation (parrot-like imitation), which imitates the previous partner's acts, is performed, and delayed imitation, which involves a lapse of time between the act of the partner and its reproduction by the child, is performed at the end of this stage (around 2 years old). [6]

Information familiarity and emotional familiarity are directed by combining these two imitative actions. Simultaneous imitation shows how the robot has recognized the user's instructions. Delayed imitation demonstrates the adaptability to the user because the robot imitates the user's past actions. Simultaneous imitation and delayed imitation can lead to new actions in addition to the rule-based actions mentioned in chapter 2. A practical interface is realized by the rule-based actions and a familiar interface is realized by the simultaneous and delayed imitations.

The proposed robot acts in accordance with a rule-based algorithm for the practical interface and in accordance with a learning-based algorithm for the familiar interface. Since it acts like a child, the control mechanism based on the two algorithms looks like the actions of a child. While a child obeys its parents' instructions, he or she is distracted by interesting things and forgets the instructions. Children do not wish to lose an opportunity to attract their parents' attention. The robot alternates these interfaces like a child according to the recognized conditions. While the robot performs tasks by means of practicality behavior control, the robot finds opportunities to act by means of familiarity behavior control. While the robot does not perform any tasks, the robot is controlled by the familiarity behavior control in order to attract the user's attention.

3.2. Implementation of the Robot Interface

The above-mentioned framework was installed in a robotic information home appliance, ApriAlpha™, which is currently under development (Table 3.1 and Fig.3.1) [7] at Toshiba Corporation. Furthermore, ApriAlpha™ cooperates with a distributed database of environments and human actions and the interaction field model proposed in the UKARI (Universal Knowledgeable Architecture for Real-Life appliances) project.[8] The distributed database stores information concerning person-to-person, person-to-appliance and person-to-service interfaces, which is obtained from sensors and other devices, by structuring it. The interaction field model is designed to express the interactions between human beings and appliances as potential energy functions and to propose a framework for producing services through the interactions of potential energy.[9]

A block diagram of the ApriAlpha™ modules and the UKARI system is shown in Fig.3.2 Moreover, we consider that the imitation action of a robot has an embodiment different to that of a human. The ApriAlpha™ system consists of two interface modules and two control modules. The interface modules are an actuator interface and a sensor interface. The actuator interface module mainly controls the three functions; gazing, speaking, and moving. These three functions are used for expressing familiarity to the user. The sensor interface module controls cameras that serve as "eyes", microphones, wheel encoders, and ultrasonic sensors for obstacle avoidance and sends the sensor data to the distributed database. The control modules are separated into two parts; a basic control module part and a fusion control module part. The basic control module part serves as a bridge between the actuator interface and the sensor interface by a feedback loop. The fusion control module part combines practical behaviors and familiar behaviors based on the following method.

ApriAlpha™ has the three motions; gazing, speaking and moving. Using these three motions, ApriAlpha™ achieves practical behaviors. Gazing motion is to announce to the user that ApriAlpha™ awaits the user's instructions. Speaking motion is to attract the user's attention and notify some information. Moving motion is to go to the target position.

In addition to the three motions, ApriAlpha™ does useful tasks via the network. The attached display of ApriAlpha™ shows the door camera's view or any other information from Web sites. ApriAlpha™ recognizes the user's instruction by speech recognition function and controls the various home information appliances.

Table 3.1 Specifications of "ApriAlpha™"

Dimensions	350 x 350 x 380mm (w x d x h) (Weight 9.5kg)
Sensors Interfaces	CCD Cameras (2), Ultrasonic Sensors (6), Microphones (3), Speakers (2), TFT Liquid Crystal Display with Touch-Screen
Movement	2 Drive Motors and Wheels; Speed: 0.5m/s 3 Motors of Camera's Pan/Tilt Motion 1 Motors of Display Open-Close Motion
Battery	High Output Lithium-Ion Battery (Integrated) 2 Hours of Uninterrupted Operation with Battery
Transmission/ Communication Functions	Communication with External PCs, Home Servers, PDAs and Mobile Phones through Wireless LAN Communication with Bluetooth™ Equipment through Home Server

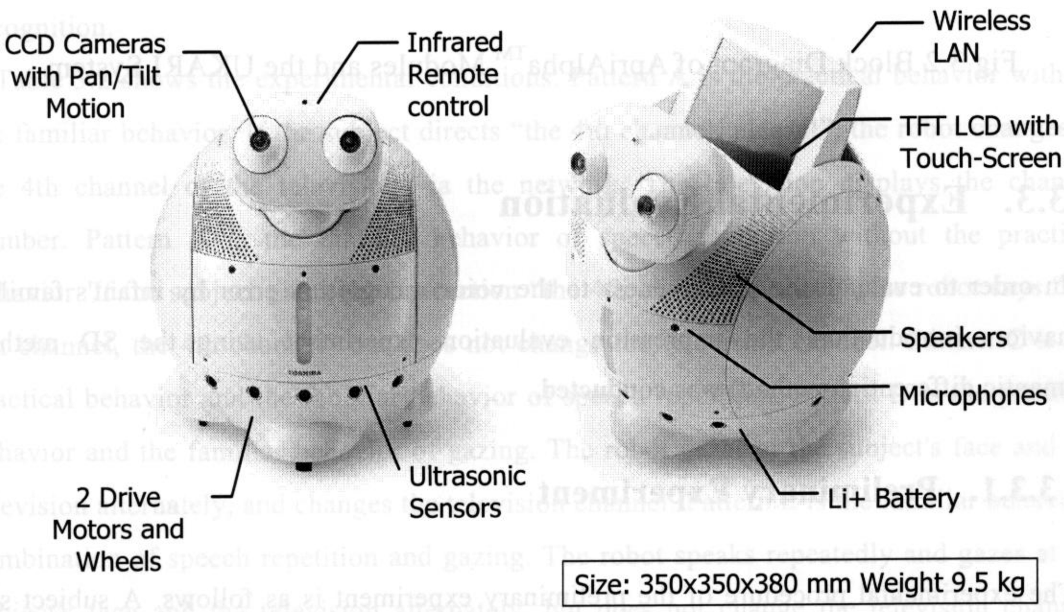


Fig.3.1 Overview of Robotic Information Home Appliance, "ApriAlpha™"

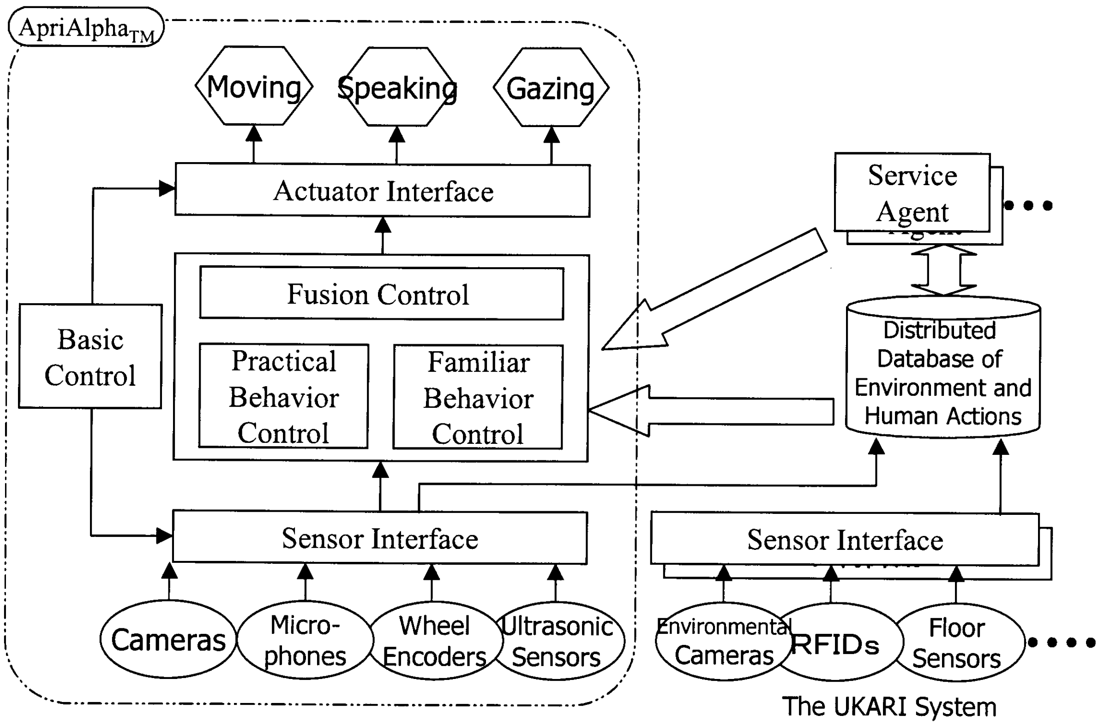


Fig.3.2 Block Diagram of ApriAlpha™ Modules and the UKARI System

3.3. Experimental Evaluation

In order to evaluate the effectiveness to the voice recognition error by infant's familiar behaviors introduction, the impression evaluation experiment using the SD method (semantic differential method) was conducted.

3.3.1. Preliminary Experiment

The experimental procedure of the preliminary experiment is as follows. A subject sits down in front of a television and the robot. The subject gives instructions to the robot by voice and changes the television channel. (Fig.3.3) The responses of the robot are controlled by 6 conditions.

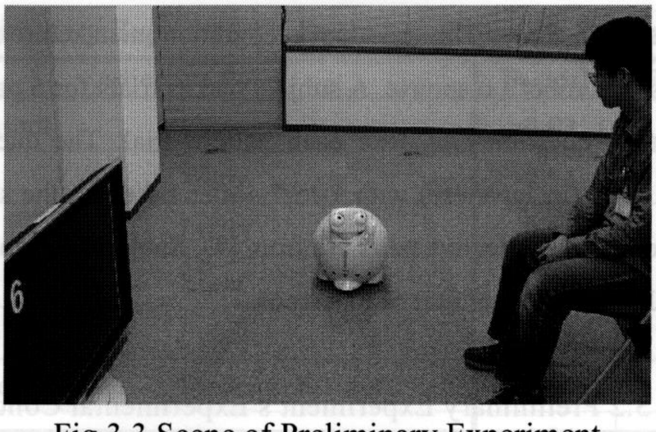


Fig.3.3 Scene of Preliminary Experiment

The behavior executions are as follows: the practical behavior without the familiar behavior, the familiar behavior without the practical behavior, and the practical and familiar behaviors. The practical behavior is to change the television channel. For the familiar behaviors, two motions are prepared. One is speech repetition. The other is gazing at the subject's face and the television. The other two conditions are for the execution of speech recognition.

Table 3.2 shows the experimental conditions. Pattern A is the practical behavior without the familiar behavior. If the subject directs "the 4th channel, please!", the robot changes to the 4th channel of the television via the network. The television displays the channel number. Pattern B is the familiar behavior of speech repetition without the practical behavior. If the subject gives the instruction "the 4th channel, please!", the robot says "the 4th channel, the 4th channel" but does not change the television channel. Pattern C is the practical behavior and the familiar behavior of speech repetition. Pattern D is the practical behavior and the familiar behavior of gazing. The robot gazes at the subject's face and the television alternately, and changes the television channel. Pattern E is the familiar behaviors combination of speech repetition and gazing. The robot speaks repeatedly and gazes at the subject's face and the television alternately, but does not change the television channel. Pattern F is a combination of all the behaviors. The familiar behavior of gazing without the practical behavior is neglected because the subject cannot understand the meaning of this behavior. Each subject tries 6 times for each pattern. The robot does not recognize subject instructions 3 times per 6 times. The orders of recognition mistake are randomized and the subject cannot guess the robot's response. A subject makes trials for 6 patterns in a different order and replies to a questionnaire after each pattern trial. The questionnaire involves rating 11 adjective pairs (in Japanese) with 1-to-7 scales based on the semantic differential method (SD method). These adjective pairs are from [4]. Subjects are non-researchers. They are three men and three women, a total of 6 persons.

Table 3.2 Preliminary Experiment's Experimental Conditions

	Practicality (Channel Changing)	Familiarity (Speaking)	Familiarity (Gazing)
A	○	–	–
B	–	○	–
–	–	–	○
C	○	○	–
D	○	–	○
E	–	○	○
F	○	○	○

Table 3.3 shows the results of the ratings. The adjective pairs in the table are translated from Japanese words. Principal component analysis was performed according to the SD method rating for the 11 adjective pairs. Based on the difference in eigenvalues, we adopt a solution that consists of one factor only. The second factor's eigenvalue is lower than 1.0. The first factor is named the “familiarity factor" because several pairs of familiarity adjectives, such as “Unpleasant-Pleasant” and “Cold-Warm”, have high loadings for the first factor. The second factor, the eigenvalue of which is low, is named the “motion factor” because several pairs of adjectives of motion, such as “Slow-Quick” and ”Slow-Rapid”, have high loadings for the second factor.

Table 3.3 Preliminary Experiment's Evaluated Adjective Pairs and the Results

Adjective Pairs		FactorI	FactorII
Unpleasant	Pleasant	**-0.93**	-0.06
Cold	Warm	**-0.91**	**-0.29**
Ugly	Pretty	**-0.91**	-0.01
Dark	Light	**-0.90**	-0.07
Rigid	Frank	**-0.90**	0.05
Comprehensible	Incomprehensible	-0.86	0.00
Quiet	Showy	-0.84	**-0.20**
Mechanical	Humanlike	-0.83	**-0.37**
Slow	Quick	-0.82	**0.54**
Slow	Rapid	-0.82	**0.54**
Boring	Interesting	-0.81	-0.09
Proportion of variance		75.0%	7.83%

The experimental conditions for the familiarity factor are shown in Fig.3.4. As expected, the behavior fusion of task execution with speech repetition and object gazing achieves the highest score for the familiarity factor. This result does not depended on whether voice recognition is correct or mistaken. The order of familiarity factor scores is F>>B>C>E>>D>>A (">>" means $p<0.05$, where p means statistical significance of the comparison).

For all conditions, except pattern A, correct recognition is higher than mistaken recognition and is significantly different ($p<0.05$). No significant difference is observed in the case of Pattern A. It shows the subject does not notice change of the television channel because Pattern A is only the channel change without any familiar behavior. This result shows the familiar behavior of gazing is effective for the speech interaction.

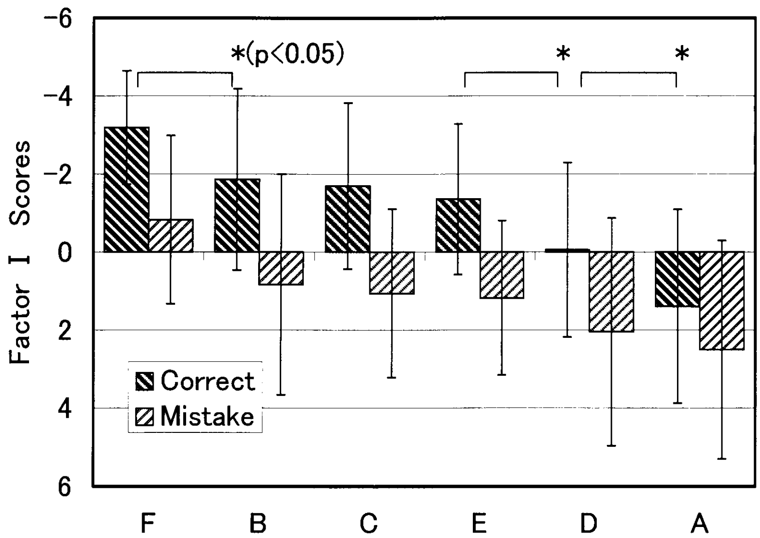


Fig.3.4 Preliminary Experiment's Comparison of Familiar Behaviors

(Correct and Mistake of Voice Recognition)

3.3.2. Experiment

Although the experiment method is almost the same as that of the preliminary experiment, there are partly different points. As experimental conditions, the added familiar behavior is that of moving. When changing the television channel, the robot moves to the television, when hearing a command, it moves to the subject. It turns out that the familiarity factor goes up as familiar behavior is added like an utterance < utterance + gazing from the result of the preliminary experiment. So each familiar behavior of utterance, gazing, and moving decided to add with not a total combination but pattern B, C, and D (Table 3.4). In order to measure the influence of existence of a robot as an experimental condition, the case that a subject directs to TV directly, pattern E was added. Pattern E constitutes a valuation basis.

As a valuation method, the impression evaluation experiment uses the SD method as same as the preliminary experiment. In order to evaluate the influence of a voice recognition error, each four trial pattern performed 50% of voice recognition error at a time at random. In the preliminary experiment, 11 adjective pairs performed the questionnaire that subjects answer after each end of a pattern. However, the contribution of factors except the familiarity factor was low, and was not able to evaluate. Then, the adjective pairs considered to be related with practicality were added, and the number of adjective pairs in the questionnaire is increased to 22 (Table 3.5).

Table 3.4 shows the experimental conditions, and Fig. 3.5 shows the familiar behaviors of the experiment. The responses of the robot are controlled by 5 conditions. Subjects are non-researchers. They are 15 men and 13 women, a total of 28 persons. In order to reveal differences that may be related to age, the subjects consisted of 10 people in their 20's, 9 in their 30's, and 9 in their 60's.

Table 3.4 Experiment's Experimental Conditions

	Practicality (Channel Changing)	Familiarity (Speaking)	Familiarity (Gazing)	Familiarity (Moving)
A	○	—	—	—
B	○	○	—	—
C	○	○	○	—
D	○	○	—	○
E	○	TV. directly (without robots)		

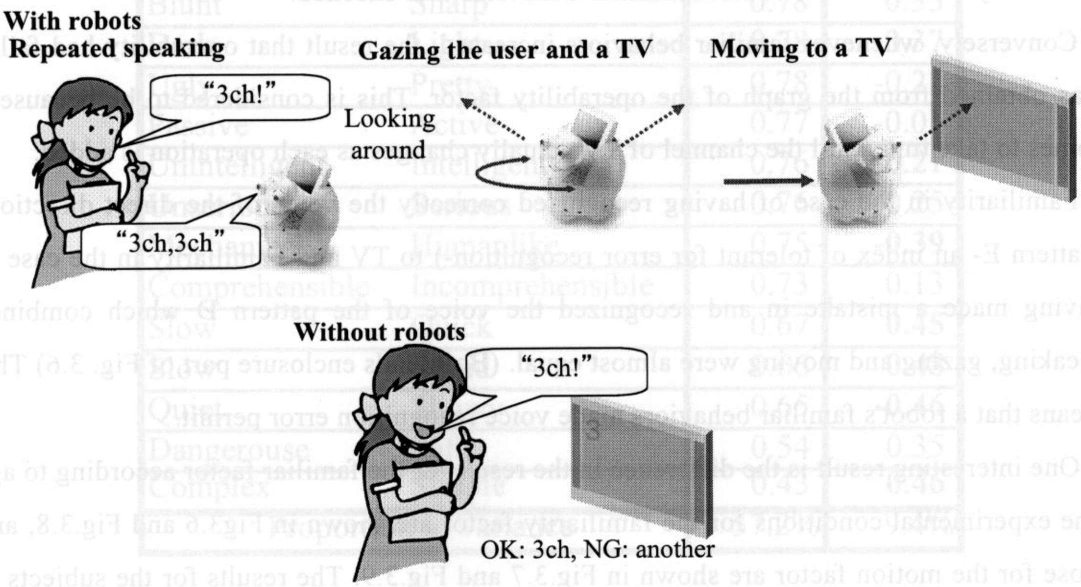


Fig.3.5 Experiment's Conditions of Familiar Behaviors

Table 3.5 shows the results of the ratings. The 22 adjective pairs in the table are translated from Japanese words. In the preliminary experiment, in order to check existence of factors, principal component analysis was used. Then, since existence of the familiarity factor and the motion factor was checked, factor analysis was used in this experiment. As a result, the proportion of variance of the 1st factor got about 57%, and the 2nd factor got about 9.4% (Table 3.5).

As in the preliminary experiment, the first factor is named the “familiarity factor" because several pairs of familiarity adjectives, such as “Unpleasant-Pleasant” and “Unfriendly-Friendly”, have high loadings for the first factor. The second factor is named the “operability factor” because several pairs of adjectives of operation, such as “Slow-Rapid” and "Complex-Simple", have high loadings for the second factor.

The graph summarized in order of the familiarity factor for correct and incorrect recognitions of the speech of each pattern is shown in Fig. 3.6. The graph summarized in order of the operability factor is shown in Fig. 3.7. The graph of this familiarity factor shows that familiarity goes up it as both of correct and incorrect recognitions add each familiar behaviors of speaking, gazing and moving. As in the preliminary experiment, the familiarity factor tends to increase as familiar behaviors are added. The validity of the familiarity factor for the mounted familiar behaviors was checked.

Conversely, whenever familiar behaviors increased, the result that operability had fallen was obtained from the graph of the operability factor. This is considered to be because it comes to take time until the channel of TV actually changes as each operation is added.

Familiarity in the case of having recognized correctly the voice of the direct directions (pattern E- an index of tolerant for error recognition-) to TV and Familiarity in the case of having made a mistake in and recognized the voice of the pattern D which combined speaking, gazing and moving were almost equal. (Both-ends enclosure part of Fig. 3.6) This means that a robot's familiar behaviors made voice recognition error permit.

One interesting result is the difference in the result for the familiar factor according to age. The experimental conditions for the familiarity factor are shown in Fig3.6 and Fig.3.8, and those for the motion factor are shown in Fig.3.7 and Fig.3.9. The results for the subjects in their 60’s are shown on the left-hand side, and those for the subjects in their 20’s are shown on the right-hand side. In Fig.3.8, the result for the familiar factor is almost the same as that for speech recognition. On the other hand, for the subjects in their 20's a clear difference appeared for the correct or mistaken speech recognition. In Fig.3.9, these results are attributable to the motion factor. For the subjects in their 60's, motion factors differ little, but for those in their 20's motion factors do differ. This is considered to reflect the difference in values between elderly people and young people.

Table 3.5 Experiment's Evaluated Adjective Pairs and the Results

Adjective Pairs		Factor I	Factor II
Unpleasant	Pleasant	**0.88**	-0.05
Unfriendly	Friendly	**0.85**	-0.32
Boring	Interesting	**0.83**	-0.12
Dull	Exciting	**0.83**	-0.25
Rigid	Frank	**0.83**	-0.15
Difficult	Easy	0.82	0.34
Disobedient	Obedient	0.81	0.18
Cold	Warm	0.80	-0.38
Weary	Unwearied	0.79	-0.05
Blunt	Sharp	0.78	0.35
Dark	Light	0.78	-0.37
Ugly	Pretty	0.78	-0.21
Passive	Active	0.77	-0.01
Unintelligent	Intelligent	0.76	0.21
Unserious	Serious	0.76	0.25
Mechanical	Humanlike	0.75	**-0.39**
Comprehensible	Incomprehensible	0.73	0.13
Slow	Quick	0.67	**0.45**
Slow	Rapid	0.66	**0.48**
Quiet	Showy	0.65	**-0.46**
Dangerouse	Safe	0.54	0.35
Complex	Simple	0.43	**0.46**
Proportion of variance		57.2%	9.4%

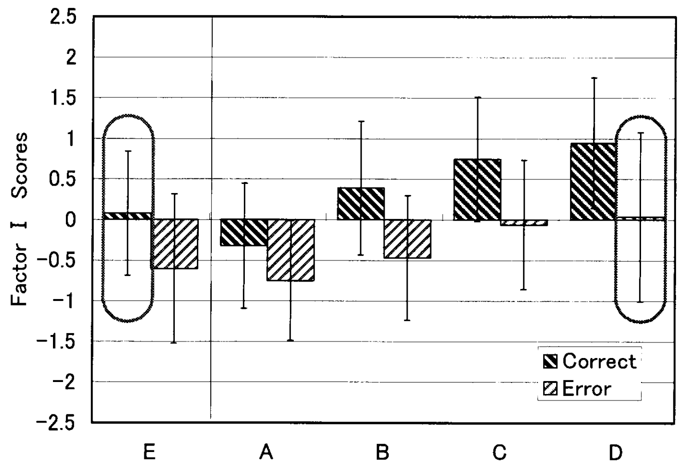


Fig.3.6 Experiment's Comparison for Factor I (All)

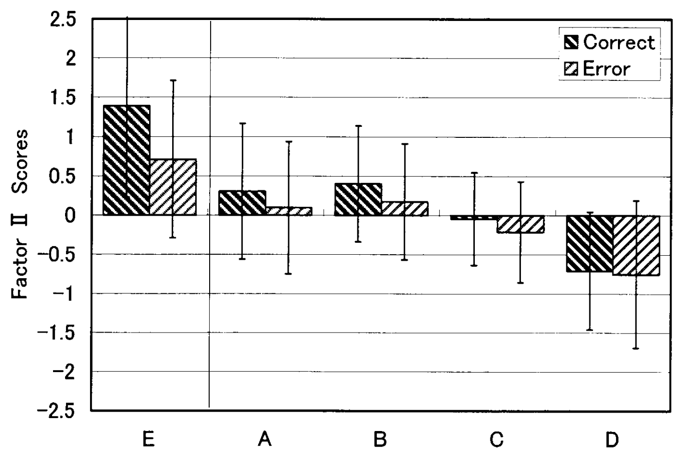


Fig.3.7 Experiment's Comparison for Factor II(All)

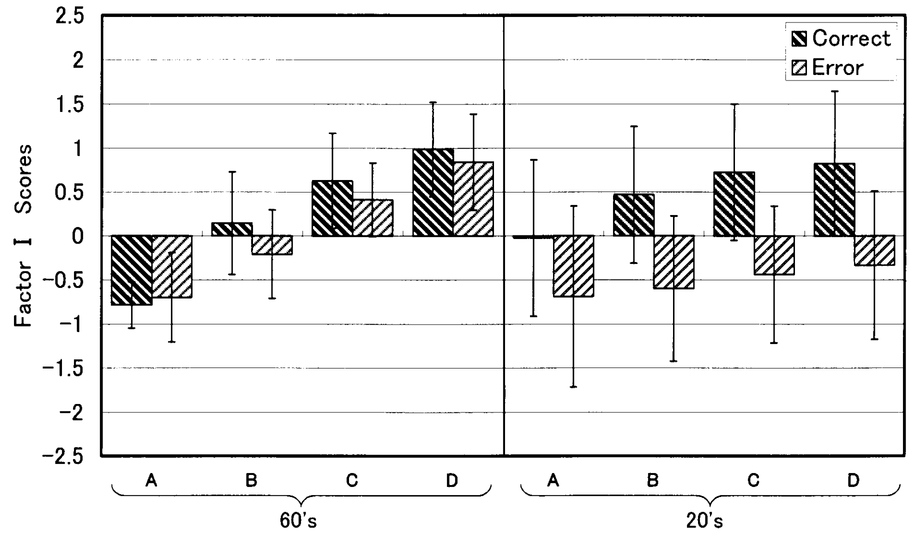


Fig.3.8 Experiment's Comparison for Factor I

(Differences between the subjects in their 60's and those in their 20's)

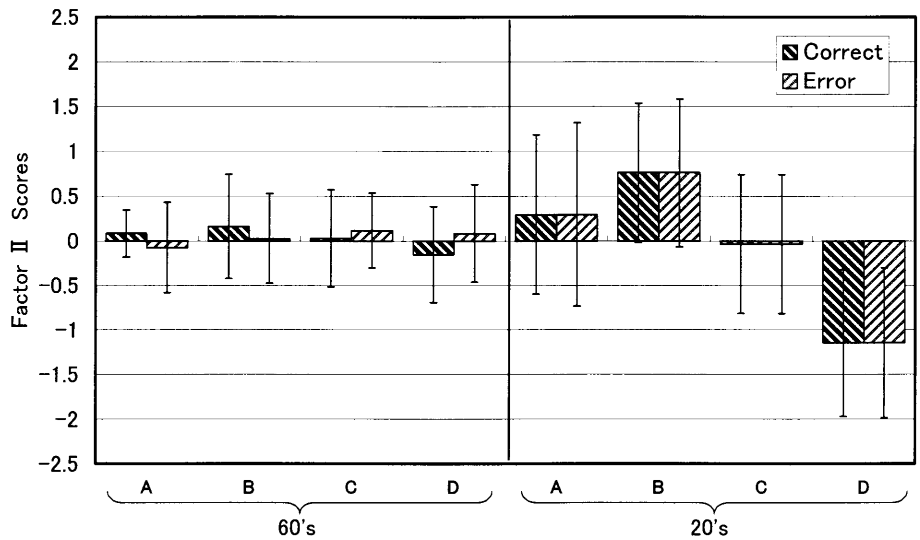


Fig.3.9 Experiment's Comparison for Factor II

(Differences between the subjects in their 60's and those in their 20's)

3.3.3. Evaluation of the Robots in the Different Shapes

This section reports the effects of the familiar behaviors installed in two kinds of robots in the different shapes. As a candidate for comparison, I used the dialog interface robot Phyno [10] (Fig.3.10) manufactured in the National Institute of Information and Communications Technology open lab. Phyno is the height of about 200mm and has 1 DOF (Degree of Freedom) at waist, 2 DOF at wings and 3 DOF at head. The impression evaluation experiment by the SD method in the general subject of a total of 21 men and women aged from 19 to 69 was conducted using ApriAlpha and Phyno. Experiment scenery is shown in Fig. 3.13 and Fig.3.14. The contents of an experiment are the same as the former. A different point is to add pointing with wing of the direction to TV for Phyno's familiar behavior (Fig.3.11 and Fig.3.12). Moreover, ApriAlpha's moving is performed by two sorts, normal speed (speed about 0.3 m/s) and high-speed which doubled speed. There are seven experiment patterns in total used as A and G of the valuation basis of the existence of B-F which added movement, high-speed movement, and pointing with wing to utterance and a gaze, and familiar behavior, and a robot's existence (Table 3.6), and experiment turn presupposed that it is random. The operator operated the experiment and it added 50% of incorrect recognition at random so that a subject might not understand for the impact evaluation of each pattern 6 trial deed and incorrect recognition. When it had been incorrect-recognized on the question paper which consists of a seven-step 24 adjective pair (Table 3.7) after each end of a pattern as the case where it has recognized correctly, it divided into it, and the reply was obtained.

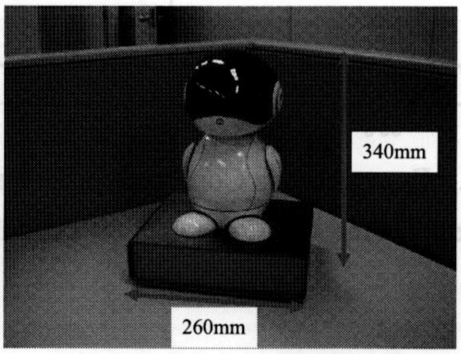


Fig.3.10 Overview of Phyno

As a result of conducting factor analysis of the impression evaluation (Table 3.7), the 1st factor (66.5% of contribution) had the high influence on the adjective pair "Dark - Light" and "Cold - Warm", "Rigid - Frank", so the first factor is named the "familiarity factor". The graph which compared each pattern by the familiarity factor is shown in Fig. 3.15. This graph to B-F that the subject orders to the robot with the familiar behavior of the familiarity factor is higher than G that the subject orders to TV directly. Therefore, it was checked that the familiar behavior introduced also in which robot is effective to voice error recognition.

Table 3.6 Experimental Conditions of Experiments

	Robots	Familiar behaviors				
		Speaking	Gazing	Moving	Moving (high	Pointing with wing
A	ApriAlpha™					
B	ApriAlpha™	O	O			—
C	ApriAlpha™	O	O	O		—
D	ApriAlpha™	O	O		O	—
E	Phyno	O	O	—	—	
F	Phyno	O	O	—	—	O
G	TV. directly	—	—	—	—	

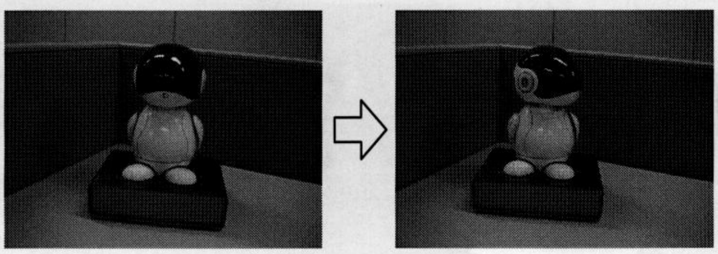


Fig.3.11 Motion of Pointing

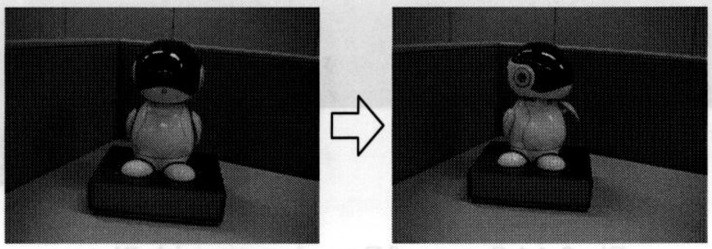


Fig.3.12 Motion of Pointing with Wing

Fig.3.13 Scene of Experiments with ApriAlpha™

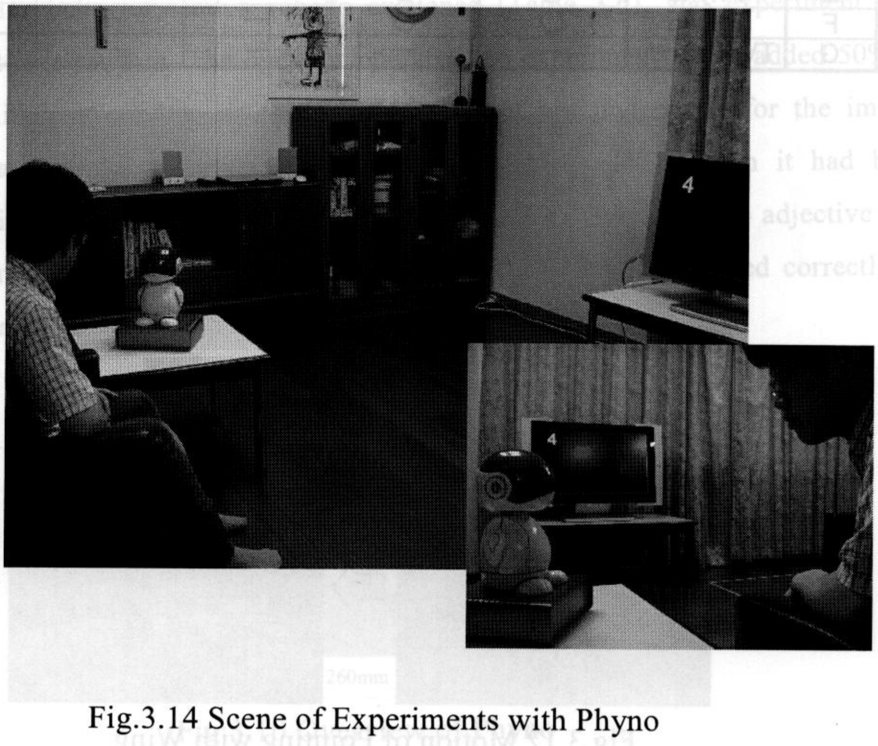


Fig.3.14 Scene of Experiments with Phyno

Table 3.7 Evaluated Adjective Pairs and the Results of Experiments

Adjective Pairs		Factor I	Factor II	Factor III
Dark	Light	**0.83**	0.13	0.05
Cold	Warm	**0.82**	0.30	0.04
Rigid	Frank	**0.79**	0.19	0.07
Unfriendly	Friendly	**0.77**	0.38	0.09
Quiet	Showy	**0.76**	0.11	0.11
Mechanical	Humanlike	0.73	−0.07	0.04
Ugly	Pretty	0.69	0.37	0.17
Dull	Exciting	0.63	0.38	0.17
Unpleasant	Pleasant	0.63	0.53	0.29
Boring	Interesting	0.47	0.49	0.26
Passive	Active	0.46	0.51	0.33
Unkind	Kind	0.39	**0.70**	0.26
Unintelligent	Intelligent	0.32	0.47	**0.47**
Weary	Unwearied	0.31	0.18	0.29
Unserious	Serious	0.31	0.47	0.35
Disobedient	Obedient	0.31	**0.69**	0.30
Comprehensible	Incomprehensible	0.30	**0.67**	0.38
Difficult	Easy	0.25	**0.59**	**0.45**
Unnecessary	Necessary	0.21	0.29	0.29
Dangerouse	Safe	0.16	**0.69**	0.15
Blunt	Sharp	0.12	0.32	**0.77**
Slow	Quick	0.08	0.25	**0.90**
Slow	Rapid	0.05	0.23	**0.91**
Complex	Simple	−0.04	0.38	0.28
Proportion of variance		25.7%	18.9%	15.4%

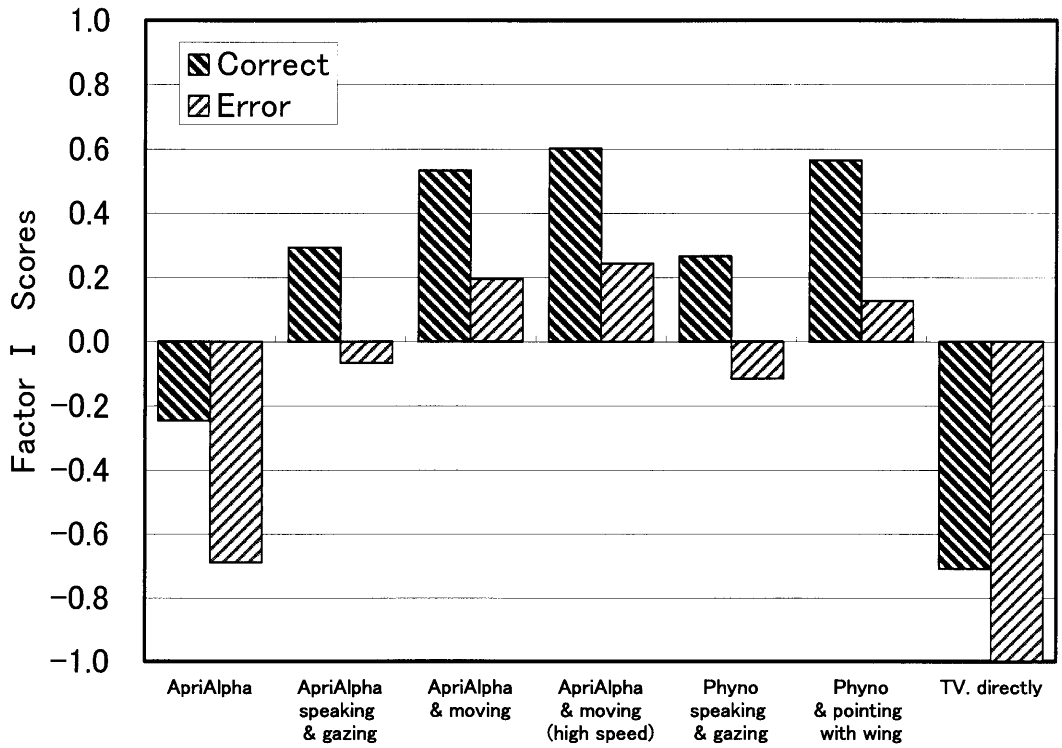


Fig.3.15 Experiment's Comparison for Factor I

3.4. Summary

In this chapter, I have proposed a new robotic interface which is tolerant voice recognition error by infant's familiar behaviors. I implemented a home robot interface and studied how to install it in a robotic information home appliance, "ApriAlpha^{TM}". The effect of infant's familiar behaviors is experimentally evaluated using the semantic differential method and principal component analysis or factor analysis. The familiar behaviors combined with speech repetition, object gazing and moving achieves the highest score for the familiarity factor. It is especially effective for the subjects in their 60's.

4. Interface Learning User-Defined Voice Instructions by Familiar Behaviors

4.1. Learning Method of User's Voice Instruction

4.1.1. Learning of User's Voice Instruction

In order to operate a home appliance without recourse to voice instruction words registered beforehand, but instead by the user's words, a robot needs to acquire the spoken instruction words for home appliance operation from the user. However, it is hard to register a spoken instruction word using a keyboard, for example. It is unrealistic to register many operation words. In this research, we devised an approach whereby a robot learns a user's words in a natural voice dialog. Just as an infant asks its parents what it does not understand, similarly a robot learns a user's words and associates them with home appliance operation because a robot asks a user questions in order to understand aspects of operation that it does not understand.

Concretely, I consider the case in which the robot changes to a news program on a multi-channel television (Fig.4.1). In order to teach the robot, a user operates the television via a remote control as well as in the traditional manner. The robot will detect that the television was operated via a network, if the television is connected to a network. If the television is not connected to a network, the robot receives a remote control signal directly and detects operation of the television. If the operation command of the television is a command that the robot has not encountered previously and the robot lacks information about it, the robot questions the user: "What are you doing now?". An operation command is <SetNewsCh> etc. if the television is connected to a network. If the television is a conventional home appliance, it is the code of a remote control signal itself. If the user answers "I changed the channel to the news. (ニュースにした)" to the question from the robot, the robot will repeat the recognized language and will learn the correspondence to the command <SetNewsCh> "changed channel to the news." When operating the television, the user commands "Change the channel to the news. (ニュースにして)". Calling it "the news, the news", the robot sends a network signal, or sends a remote control signal, and changes the television channel.

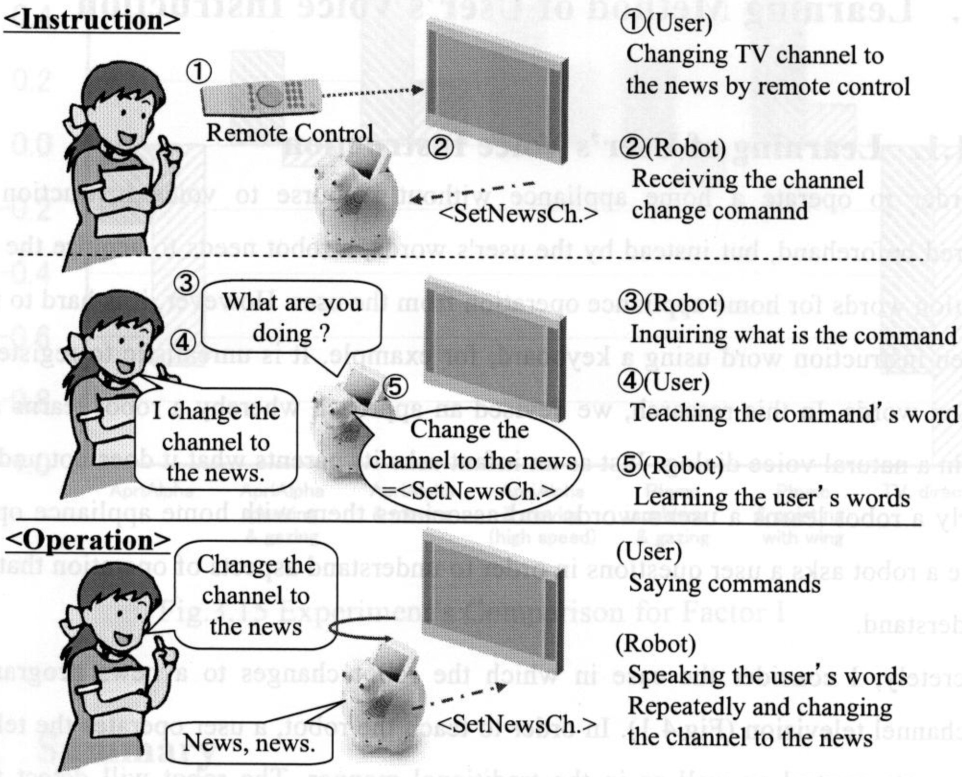


Fig.4.1 Explanation of Learning Method of user's voice instruction

If study has not progressed yet even if taught once, the robot asks a question each time, "What are you doing?" If study is progressing to some extent, the already learned words will be uttered. If this utterance is wrong, the user can teach again and the robot can learn the right user's words gradually. When teaching, if the voice recognition mistaken (for example , the voice recognition error of the "the news" recognized as "Muse"), the robot repeats the words recognized by mistake. Hearing the repetition, the user can teach again or can operate it using remote control again.(Fig.4.2)

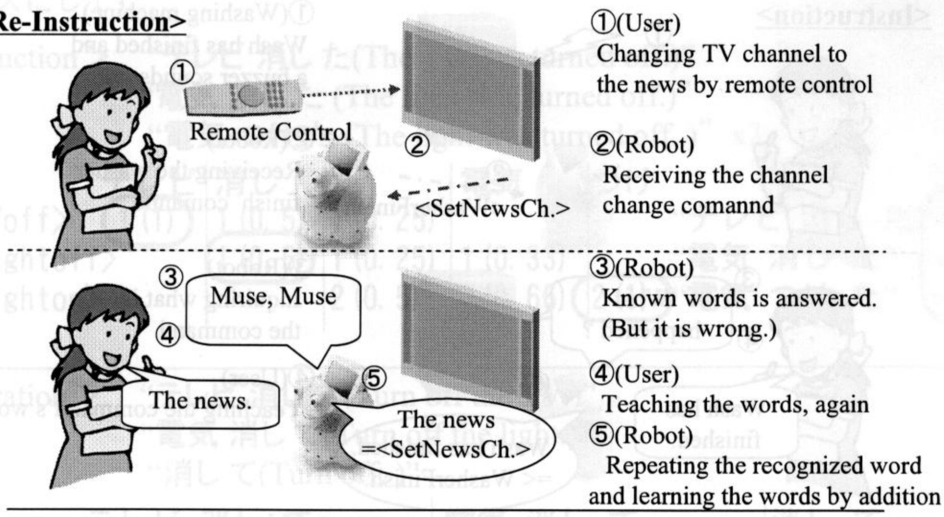


Fig.4.2 Explanation of Learning Method of Re-Instruction

4.1.2. Notice of User's Voice Instruction

Repetition ("the news, the news") of the user's words at the time of a robot operating a home appliance is also notifying a user by means of words. It is almost meaningless when a user operates this. However, in a home appliance with the function of which a user is notified, it is meaningful. For example, the end of wash is told at a buzzer like a washing machine. That is, when a washing machine is completed, the command is sent via a network. A robot receives the command and asks the user, "What happened". And a user teaches "Wash has finished." Then, it can be reported when wash was completed, "Wash has finished. Wash has finished." as the same framework as the time of operating it. (Fig.4.3)

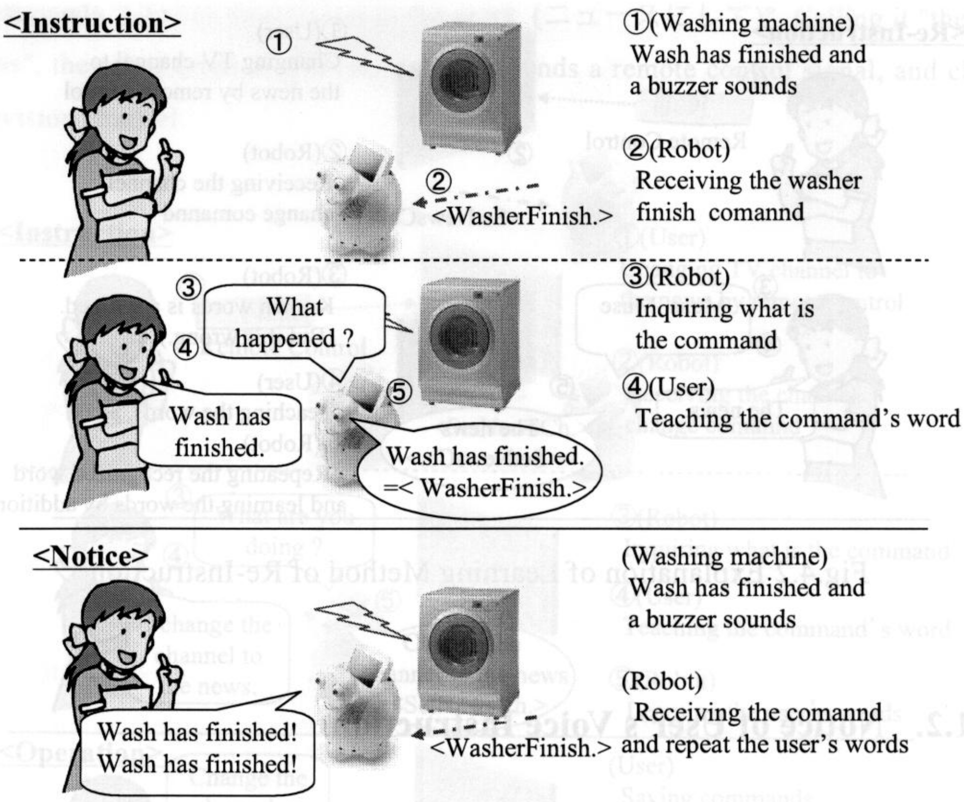


Fig.4.3 Explanation of Learning Method of Notice

4.1.3. Voice Instruction matching using a morphological analysis

In the case of instruction and in the case of operation, the word ending differs in Japanese. In the above example, it is "I changed the channel to the news (ニュースにした)" in the case of instruction. In the case of operation, it becomes "Change the channel to the news (ニュースにして)"and the word ending differs from "て" and "た." In order to cope with the problem posed by the change of the word ending, by means of a morphological analysis, in the case of instruction, it separates by a clause and matches with "ニュース","にし","て" separately at the command <SetNewsCh> and is memorized. In the case of operation as well as in the case of instruction, a user's utterance is separated by a clause. The adaptive value of the clause and each command is computed from the frequency of the clause in each command. And the command of the highest adaptive value is output. Thus, operation becomes possible even if the word ending differs such as "change the channel to the news

(ニュースにして)."

<Instruction> "テレビ 消し た(The TV was turned off.)"
"電気 消し た (The light was turned off.)"
"電気 つけ た (The light was turned off.)" x2

	テレビ	消し	た	電気	つけ		
<SetTVoff>	1 (1)	1 (0. 5)	1 (0. 25)			"テレビ 消し た"	=テレビ
<SetLightoff>		1 (0. 5)	1 (0. 25)	1 (0. 33)		"電気 消し た"	=消し
<SetLighton>			2 (0. 5)	2 (0. 66)	2 (1)	"電気 つけ た"	=つけ

<Operation> "テレビ 消し て(Turn off the TV.)"
"電気 消し て(Turn off the light.)"
"消し て(Turn off.)"

	テレビ 消し て	電気 消し て	消し て
<SetTVoff>	1 +0. 5 +0 =1. 5	0 +0. 5 +0 =0. 5	0. 5 +0 =0. 5
<SetLightoff>	0 +0. 5 +0 =0. 5	0. 33 +0. 5 +0 =0. 83	0. 5 +0 =0. 5
<SetLighton>	0 +0 +0 =0	0. 66 +0 +0 =0. 66	0 +0 =0
	↓ <SetTVoff>	↓ <SetLightoff>	↓ <SetLightoff> or<SetTVoff>

Fig.4.4 Explanation of Learning Method of user's voice instruction

Next, the details of the calculation method are explained. For example, as for instruction words, A and B were taught every 1 time, C was taught twice (Fig. 4.4). Voice recognition is right and presupposes that it was assigned to the command <SetTVoff>, <SetLightoff>, and <SetLighton>, respectively.<SetTVoff>: "テレビ消した"(The TV was turned off.), <SetLightOff>: "電気消した"(The light was turned off), <SetLightOn>: "電気つけた"(The light was turned on) "テレビ消した" , "電気消した" and "電気つけた" are decomposed by morphological analysis with "テレビ／消し／た", "電気／消し／た" and "電気／つけ／た". Each command is connected for every morpheme of voice instruction words, and computes a goodness of fit according to the frequency. That is, "テレビ" gets frequency 1 (goodness of fit 1) at <SetTVoff>. "消し" gets frequency 1 at <SetTVoff> by "テレビ消した", and gets frequency 1 at <SetLightoff> by "電気消した". "消し" goodness of fit turns into frequency / sum total frequency, and gets every 0.5 (the inside of Fig. 4.4 parenthesis is a goodness of fit). Thus, whenever it gains voice instruction words, the goodness of fit of each morpheme is computed.

When operating it, for example “テレビ消して” (Turn off the TV.) is inputted, The sum of a goodness of fit with each command will be calculated for every morpheme of “テレビ/消し/て”, and a command with the highest goodness of fit <SetTVoff> (goodness of fit 1.33) will be executed. Thus, even if “テレビ消した” teaches at the time of instruction, at the time of operation, it can be operated by “テレビ消して”.

Moreover, when voice instruction words are ambiguous like “消して” (Turn off) at the time of operation and a goodness of fit is that it is the same (or slim margin), A robot throws back a question like “消して って何？” (Turn off, what?). This will cancel the ambiguity of a user's voice instruction words.

It is comparing the size of the goodness of fit within each command, the morpheme which shows the greatest value can be obtained as representative words of a command. This can also be used as words at the time of notifying at which it was shown with Section 4.2.2. Moreover, it can register as voice command in the discrete speech recognition whose recognition precision is higher than the continuous speech recognition.

4.2. Advantage of Learning Voice Instruction

The method of learning the user's voice instructions has two advantages over the conventional method of registering the voice instructions beforehand (Fig.4.5).

One is the robustness with respect to the divergence of the user's voice instructions. For example, when changing the channel to a news program, the user will give various instructions, such as "The news" and "The news program". In the conventional method, in order to deal with this divergence, it is necessary to register all reference instructions supposed to be spoken by users. However, since spoken instructions change depending on the user, it is difficult to register beforehand by supposing all the users' spoken instructions. Moreover, even if they are registered, the registered number of voice instructions will increase, and the possibility of making a mistake in speech recognition will become high. On the other hand, in the case of the proposed method, the robot learns the words of the user who actually says the spoken instruction. Therefore, the actual user can operate home appliances by means of the instructions he speaks. Since the registered instructions are limited, the rate of speech recognition improves.

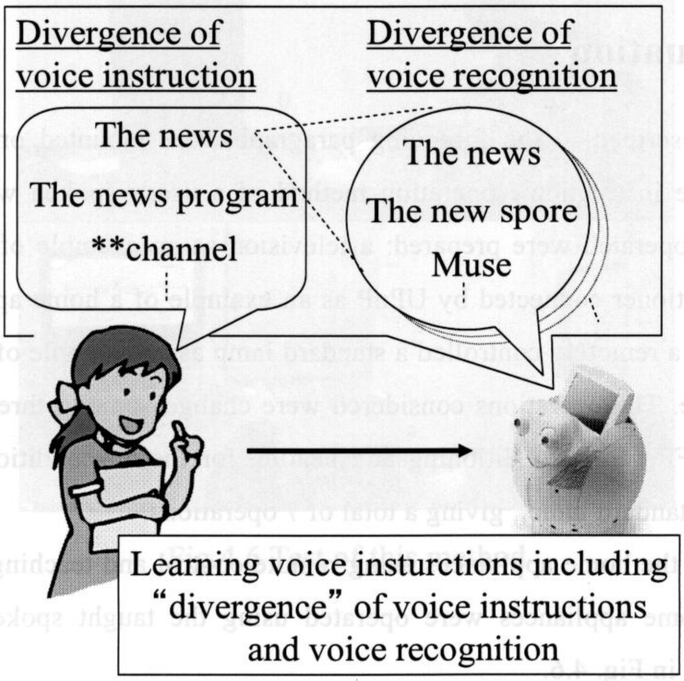


Fig.4.5 Advantage of Learning Voice instruction

Another advantage is the robustness with respect to the divergence of the recognition result due to the voice recognition error of a robot. For example, when changing the TV channel to the news program, the user said "news" but the robot recognized it as "muse." In the case of the conventional method of registering the voice instructions, "muse" is not registered as the voice instruction to change the TV channel to a news program. Therefore, the robot does not change the TV to a news program. On the other hand, in the case of the proposed method, even if the robot erroneously recognizes it as "muse", it may be able to change the TV channel correctly. Because, in the case of the proposed method, the voice instructions are learned using the actual speech recognition result of instruction. Thus, a robot may erroneously recognize "news" to be "muse". However, if this is what the robot has learned, even if it erroneously recognizes the user's instruction as "muse", the robot can change the channel to the news program correctly. In fact, the possibility that the same erroneous speech recognition will occur for the same words in the same environment and for the same user is high.

4.3. Evaluation

The method described in the foregoing paragraphs was mounted on AapriAlpha^{TM} for verification of the instruction / operation method of a user's spoken words. Three home appliances to be operated were prepared: a television as an example of AV equipment, a network air-conditioner connected by UPnP as an example of a home appliance connected to a network, and a remotely controlled a standard lamp as an example of a home appliance in widespread use. The operations considered were changed among three channels for the television, ON/OFF of air conditioning and heating for the air-conditioner, and ON/OFF switching of the standard lamp,, giving a total of 7 operations.

After operating the home appliances using remote control and teaching a robot a spoken command, the home appliances were operated using the taught spoken command. The situation is shown in Fig. 4.6.

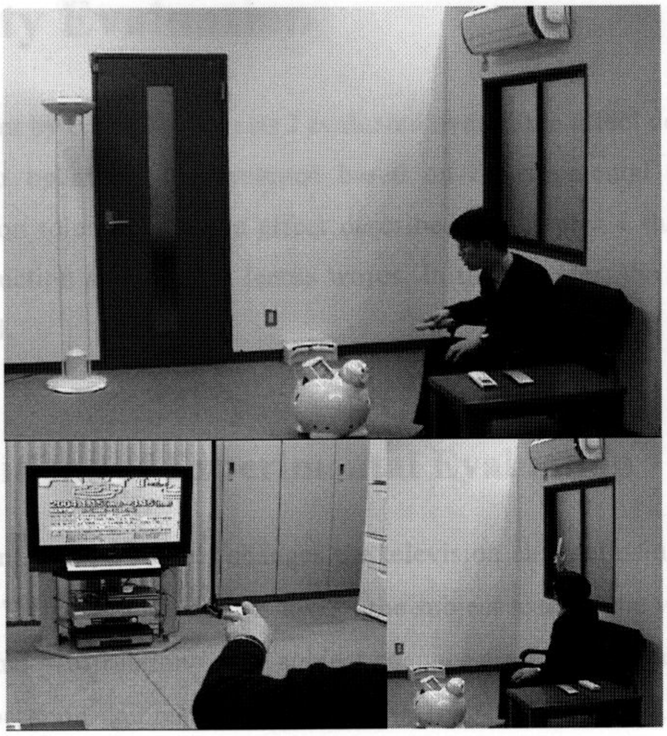


Fig.4.6 Test of this method

4.4. Summary

In this chapter, while incorrect recognition is allowed, how a robot gains a user's spoken instruction word so that an infant may memorize language is proposed. And the validity of the proposal method was evaluated through the experiment which gains the spoken command word by 12 general subjects, and changes the channel of television with the reference term. After operating the home appliances using remote control and teaching a robot a spoken command, the home appliances in the taught spoken command was operated.

5. Usability Evaluation

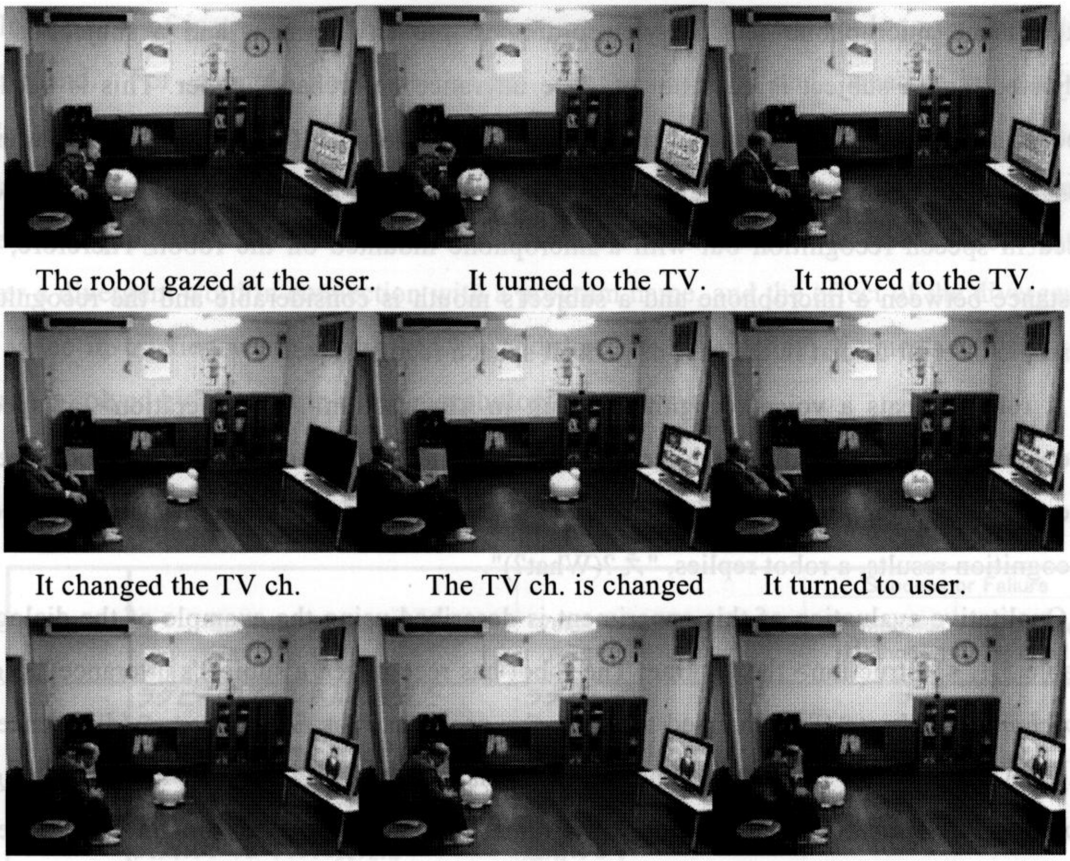
By the experiment by common subjects I evaluated overall the effect described in Chapter 3 that the familiar behaviors implemented based on developmental psychology makes incorrect recognition tolerate, and the effect described in Chapter 4 that a robot learns a user's spoken instruction as an infant learns words. In this chapter, the experiment and its result are described.

5.1. Conditions of Experimental Evaluation

In this experiment, first, a subject changes the television channel using a remote control. Since a robot asks "What are you doing, now?", the subject teaches the voice instruction for changing the channel <Instruction>. The subject changes are from the 1st channel to the 5th channel. (The operation commands are from <Set1ch> to <Set5ch>). A program of the genre typical of each channel was projected on the TV so that a subject could teach by free utterance. (The 1st channel is for news programs. The 2nd is for weather reports. The 3rd is for information programs. The 4th is for sports programs (rugby). The 5th is for movies. <Instruction> was repeatedly performed for about 5 minutes. Next, using the taught instructions, the subject commanded the robot and the robot changed the channel <Operation>. < Operation > was repeatedly performed for about 2 minutes. One trial was set as <instruction> and <operation> (for about 7 minutes). In order to compare the stage at which the spoken instructions words are not learned (at the start of an experiment) with the stage at which the spoken instruction words are learned (when instruction progresses), three trials were performed per subject. The subject completed a questionnaire at the end of each trial. The questionnaire involves rating 24 adjective pairs (in Japanese) (Table 5.2) with 1-to-7 scales based on the semantic differential method (SD method) , as well as the questionnaire described in Chapter 3. The subject distinguished his/her impressions between the case of erroneous recognition and the case of correct recognition, and completed each questionnaire. Moreover, when changing the television channel, a robot performs familiar behaviors modeled on those of an infant, whose effect of permitting voice recognition errors had already been verified by experiment. Specifically, the familiar behaviors are the following three actions: repeating a user's voice instruction, gazing at the television and the user in turn and approaching the user when receiving directions from the user and approaching the television when changing the channel. The subjects were a total of 12 men and women of a young man (his 20s-30s) and elderly people (his 60s-70s). The scene of the experiment is shown in Fig.5.1 and Fig.5.2. The robot familiar behavior motions are shown in Fig.5.3.

Fig.5.1 Scene of Usability Experiment (Younger)

Fig.5.2 Scene of Usability Experiment (Senior)



The robot gazed at the user. It turned to the TV. It moved to the TV.

It changed the TV ch. The TV ch. is changed It turned to user.

It returned to the user. It turned to the user It gazed at the user.

Fig.5.3 Robot motions of Usability Experiment

5.2. Qualitative evaluation of the proposed method

There were 110 - 209 dialogs per subject (an average of 148 dialogs per subject) in three experiments (about 7 minutes per experiment, giving a total of about 21 minutes). The instruction dialog of was 44 - 157 times (an average of 95 times), and the operation dialog was 36 - 81 times (an average of 53 times).

As an example, a part of the dialog of a certain subject is shown in Table 5.1. The subject's spoken words in Table 5.1 are copied from the video of the experiment. The voice recognition result in Table 5.1 is the result of the robot's correct recognition. (In this experiment, microphone sensitivity according to the environment and a subject is not adjusted, and a subject is does not practice utterance beforehand, either. This is because speech recognition in the case of an actual home environment and an unfamiliar user is assumed. Furthermore, experiments were usually conducted not with a headset microphone used in speech recognition but with a microphone mounted on the robot. Therefore, the distance between a microphone and a subject's mouth is considerable and the recognition result is inferior to that in the case of general speech recognition.)

A robot repeats a voice recognition result in <Instruction>. In <operation>, if speech recognition results already match the words of a learned voice instruction , the robot repeats the words of an applicable voice instruction. If there is no reference term matching speech recognition results, a robot replies, "え?(What?)".

Qualitative evaluation of this experiment is described using the example of the dialog in Table 5.1. <Instruction> the 1st line: The robot has recognized a subject's utterance "*Rugby no channel ni kaeta* (I've changed to the rugby program.)" as *"Rugby no s*a (difference of rugby.)" by mistake. <Operation> the 1st line: It has recognized a subject's utterance "*Rugby no channel ni kaete* (Change the rugby program!)" as "*Rugby no san nennkan de* (in 3 years of rugby)" by mistake. The robot divides the subject's utterance by a morphological analysis, and is learning the relation between "Rugby" and an operation command <Set4ch>. Therefore, if "Rugby" is recognized, it can perform the operation correctly.

<Operation> the 2nd line: It has recognized a subject's utterance "*Eiga no channel ni kaete* (Change to the movie program!)" as "*Reagan no san nenkan de* (In three years of Mr. Reagan)" by mistake. In the case of the conventional method, since there is no relation with a movie program when the robot recognizes "In three years of Mr. Reagan", it cannot perform the operation. However, in the case of the proposed method, the operation can be performed correctly. Because, <Instruction> the 2nd line: The robot has recognized a subject's utterance "*Eiga no channel ni kaeta* (I've changed to the movie program.)" as "*Reagan san* (Mr. Reagan)" by mistake, too. Thus, there are many examples of the correct operation being performed even if the voice recognition was incorrect, because the same incorrect recognition occurred in the case of both instruction and operation. This shows the effect of the divergence of voice recognition.

In the case of instruction, the robot gained different instruction words to one operation command. For example, the robot gained "weather report" and "weather bulletin" as <Set2ch>, and gained "the 5th channel" and "movie" as <Set5ch>. In the case of operation, a subject might utter a divergent verb for example "Show the movie program", "Change to the movie program", etc. However, in the short experiment time, it was rare for a subject to utter a divergent word in connection with a program name, and the effect of the divergence of voice instruction was not checked. In addition, the divergence of voice instruction of verbs is solved by processing using morphological analysis.

Table 5.1 Judgments of Success or Failure

	Subject's Spoken Word	Voice Recognition Result	Success or Faliure	
			Conventional Method	Proposed Method
Instruction	*ラグビーのチャンネルに変えた* *Rugby no channel ni kaeta* (I've changed the Rugby Program.)	***ラグビーの差*** ***Rugby** no sa* (difference of **Rugby**)	Success	—
	映画のチャンネルに変えた *Eiga no channel ni kaeta* (I've changed the Movie Program.)	*レーガンさん* *Reagan san* (Mr. Reagan)	Failure	—
	天気のチャンネルに変えた *Tenki no channel ni kaeta* (I've changed the Weather Program.)	*延期の3、* *Enki no san* (the 3rd of postponement)	Failure	—
Operation	*ラグビーのチャンネルに変えて* *Rugby no channel ni kaete* (Change the Rugby Program!)	***ラグビー**の3年間で、* ***Rugby** no san nenkan de* (in 3 years of **Rugby**)	Success	Success
	映画のチャンネルに変えて *Eiga no channel ni kaete* (Change the Movie Program!)	*レーガンの3年間、* *Reagan no san nenkan de* (in 3 years of Mr. Reagan)	Failure	Success
	天気 *Tenki !* (Weather!)	***天気*** ***Tenki !*** (**Weather** !)	Success	Failure

Table 5.2 The Example of Instruction Dialogs (the 1st Traial)

Robot's Question/Utteranc	User's Answer / Direction	(Recognition Result)	Robot Repetition/Utterance
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	る機関の *Ru kikan no* (No meaning)	る機関の *Ru kikan no* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	気象情報を選んだよ *Kisyou jyouhou wo erandayo* (I chose the weather information.)	真言 を *Shingen wo* (No meaning)	真言 を *Shingen wo* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	3chを選んだよ *Santyanneru wo erandayo* (I chose 3ch.)	た ！ *Ta !* (No meaning)	た ！ *Ta !* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	4chを選んだよ *Yontyanneru wo erandayo* (I chose 4ch.)	る文言があった *Ru mongon ga atta* (No meaning)	る文言があった *Ru mongon ga atta* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	資産反感 を *Shisan hankan wo* (No meaning)	資産反感 を *Shisan hankan wo* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	5chを選んだよ *Gotyanneru wo erandayo* (I chose 5ch.)	ターン *Turn* (Turn)	ターン *Turn* (Turn)
今何したの？ *Ima nani shitano?* (What are you doing?)	5chを選んだよ *Gotyanneru wo erandayo* (I chose 5ch.)	さんの絵が *San no e ga* (No meaning)	さんの絵が *San no e ga* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	さんら *San ra* (No meaning)	さんら *San ra* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	気象情報を選んだよ *Kisyou jyouhou wo erandayo* (I chose the weather information.)	塩味を放映が *Shioaji wo houei ga* (No meaning)	塩味を放映が *Shioaji wo houei ga* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	3chを選んだよ *Santyanneru wo erandayo* (I chose 3ch.)	ならない の なら *Naranai no nara* (No meaning)	ならない の なら *Naranai no nara* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	4chを選んだよ *Yontyanneru wo erandayo* (I chose 4ch.)	4ch を やら *Yontyanneru wo yara* (No meaning)	4ch を やら *Yontyanneru wo yara* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	5chを選んだよ *Gotyanneru wo erandayo* (I chose 5ch.)	残念を やら *Zannen wo yara* (No meaning)	残念を やら *Zannen wo yara* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	さんへ の選んだら *Sanhe no erandara* (No meaning)	さんへ の選んだら *Sanhe no erandara* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	気象情報を選んだよ *Kisyou jyouhou wo erandayo* (I chose the weather information.)	資本の本 が *Shihon no hon ga* (No meaning)	資本の本 が *Shihon no hon ga* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	3chを選んだよ *Santyanneru wo erandayo* (I chose 3ch.)	3段目の長屋は *Sandanme no nagayaha* (No meaning)	3段目の長屋は *Sandanme no nagayaha* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	4chを選んだよ *Yontyanneru wo erandayo* (I chose 4ch.)	の長年のエラー *No naganen no error* (No meaning)	の長年のエラー *No naganen no error* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	5chを選んだよ *Gotyanneru wo erandayo* (I chose 5ch.)	もチャンネル を *Mo tyannneru wo* (No meaning)	もチャンネル を *Mo tyannneru wo* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	し た姉のエラー *Shi ta ane no error* (No meaning)	し た姉のエラー *Shi ta ane no error* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	3chを選んだよ *Santyanneru wo erandayo* (I chose 3ch.)	な ん じゃない の なら *Nanjyanainonara* (No meaning)	な ん じゃない の なら *Nanjyanainonara* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	気象情報を選んだよ *Kisyou jyouhou wo erandayo* (I chose the weather information.)	商標法 や *Syouhyouhou ya* (and trademark law)	商標法 や *Syouhyouhou ya* (and trademark law)
今何したの？ *Ima nani shitano?* (What are you doing?)	5chを選んだよ *Gotyanneru wo erandayo* (I chose 5ch.)	残念 のエラー *Zannen no error* (No meaning)	残念 のエラー *Zannen no error* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	4chを選んだよ *Yontyanneru wo erandayo* (I chose 4ch.)	のチャンネル が *no channel ga* (No meaning)	のチャンネル が *no channel ga* (No meaning)
今何したの？ *Ima nani shitano?* (What are you doing?)	気象情報を選んだよ *Kisyou jyouhou wo erandayo* (I chose the weather information.)	首相の本音 が *syusyou no honnne ga* (The prime minister's real intention)	首相の本音 が *syusyou no honnne ga* (The prime minister's real
今何したの？ *Ima nani shitano?* (What are you doing?)	1chを選んだよ *Ittyanneru wo erandayo* (I chose 1ch.)	残念そうな *Zannen souna* (No meaning)	残念そうな *Zannen souna* (No meaning)

Table 5.3 The Example of Operation Dialogs (the 1st Traial)

User's Answer / Direction	(Recognition Result)	Robot Repetition/Utterance
1chを選んで *Ittyanneru wo erande* (Choose 1ch.)	1chを選んで *Ittyanneru wo erande* (Choose 1ch.)	え? Ehe (What?)
1chを選んで *Ittyanneru wo erande* (Choose 1ch.)	ますチャンネルを選んで *Masu channel wo erande* (Choose "Masu channel".)	え? Ehe (What?)
変わらないなぁ *Kawaranaina* (It does not change.)	た母は *ta haha ha* (No meaning)	<Silence>
3chを選んで *Santyanneru wo erande* (Choose 3ch.)	3chを選んで *Santyanneru wo erande* (Choose 3ch.)	<Silence>
3chを選んで *Santyanneru wo erande* (Choose 3ch.)	3chを選んで *Santyanneru wo erande* (Choose 3ch.)	<Silence>
気象情報を選んで Kisyou jyouhou wo erande (Choose the weather information.)	小情報を選んで *Syou Jyouhou wo erande* (Choose small information.)	<Silence>
1chを選んで *Ittyanneru wo erande* (Choose 1ch.)	1chを選んで *Ittyanneru wo erande* (Choose 1ch.)	え? Ehe (What?)

Fig.5.4 Scene of the subject shows "1ch." by gesture

Table 5.4 The Example of Operation Dialogs (the 3rd Traial)

User's Answer / Direction	(Recognition Result)	Robot Repetition/Utterance
1ch を選ん で *Ittyanneru wo erande* (Choose 1ch.)	1ch を選ん で *Ittyanneru wo erande* (Choose 1ch.)	た姉のエラー、た姉のエラー *Ta ane no error, ta ane no error,* (,,Sister's error, sister's error)
そうそうそう、いいよ Sou sou sou, iiyo (Yes, yes, It's good.)	論争の *Ronsou no* (of dispute)	え？ *Ehe* (What?)
映画を選んで *Eiga wo erande* (Choose the movie.)	やアンテナ を選んで *Ya anntena wo erande* (Choose the Antenna or,,.)	<Silence>
映画を選んで *Eiga wo erande* (Choose the movie.)	原案を選ん で *Genan wo erande* (Choose the draft proposal)	<Silence>
わからない？ *Wakaranai?* (Do you understand?)	案内 *Annnai* (Guidance)	<Silence>
映画を選んで *Eiga wo erande* (Choose the movie.)	映画を選んで *Eiga wo erande* (Choose the movie.)	だ矢、だ矢 *Daya, daya* (No meaning)
そうそうそう *Sou sou sou* (Yes, yes.)	はオゾン層 の *Ha Ozone sou no* (,,is of Ozone layer,,)	え？ *Ehe* (What?)
いいよそれで Iiyo sore de (Yes, that's OK.)	ねぇ *Nehe* (Hi)	え？ *Ehe* (What?)
じゃあ4chを選んで *Jya yontyanneru wo erande* (Next,choose 4ch.)	や亜鉛4ch を選ん で *Ya aen yontyannneru wo erande* (Choose Zinc,4ch or,,)	だ矢、だ矢 *Daya, daya* (No meaning)
4chを選んで *Yontyanneru wo erande* (Choose 4ch.)	今ちゃんネルを選ん で *Ima tyannneru wo erannde* (Choose ch. Now)	なら、なら *Nara, nara* (No meaning)
違う違う。 *Chigau chigau* (No,no.)	な きゃ *Nakya* (No meaning)	え？ *Ehe* (What?)
4chを選んで *Yontyanneru wo erande* (Choose 4ch.)	4ch *Yontyanneru* (4ch)	4ch、4ch *Yontyanneru, yontyanneru* (4ch, 4ch)
そうそうそういいよ4chね *Sou sou sou iiyo yontyannnerune* (Yes yes it's good. It's 4ch.)	仰ぎ人気 とわざわざハサミを,,, *Aogi ninnki to wazawaza hasami wo,,,*	うるさくてわかんないよー *Urusakute wakannaiyo* (it is noisy so I don't understand)
気象情報を選んで *Kisyou jyouhou wo erande* (Choose the weather information.)	気象情報を選んで *Kisyou jyouhou wo erande* (Choose the weather information.)	商標法、商標法 *Syouhyouhou, syouhyouhou* (Trademark law,trademark law)
そうそう *Sou sou* (Yes.)	意味 を *Imi wo* (Meaning,,)	え？ *Ehe* (What?)
3ch を選ん で *Santyanneru wo erande* (Choose 3ch.)	した3ch を得 Shita 3ch wo e (,,,get 3ch,,,)	え？ *Ehe* (What?)
3ch を選ん で *Santyanneru wo erande* (Choose 3ch.)	た ん じゃ *Tanjya* (No meaning)	なら、なら *Nara, nara* (No meaning)
1ch *Ittyanneru* (1ch.)	1ch *Ittyanneru* (1ch.)	残念そうな、残念そうな *Zannen souna, zannen souna* (No meaning)
あら映画にかわっちゃった *Ara eiga ni kawattyatta* (Oh, it has changed to the movie.)	払え ない ん か だっ た *Harae nainnka datta* (,,couln't pay,,),	<Silence>
1ch *Ittyanneru* (1ch.)	し 、年齢を1ch *Shi, nennrei wo 1ch* (age 1ch,,)	<Silence>
1chだよ *Ittyanneru dayo* (It's 1ch.)	10ch を選ん で *Jyuttyannneru wo erande* (Choose 10ch.)	原案、原案 *Genan, genan* (Draft proposal, draft proposal)
これは映画 *Kore ha Eiga* (This is the movie.)	これは映画で *Kore ha Eiga de* (This is the movie,,.)	<Silence>
映画はダメ *Eiga ha dame* (No,the movie!)	し て半年を選ん で *Shite hanntoshi wo erande* (,,,Choose half year,,,)	原案、原案 *Genan, genan* (Draft proposal, draft proposal)
映画じゃないの。 *Eiga jya naino.* (It isn't the movie.)	例外で ない の *Reigai de naino* (It isn't exception.)	<Silence>
1ch を選ん で *Ittyanneru wo erande* (Choose 1ch.)	し て判例 の上欄で *Shite hannrei no uerann de* (,,Upper column of precedent,,)	<Silence>

Fig 5.4

5.3. Operation Rate Evaluation

In order to evaluate the operation rate in the case of using the proposed method, it was judged whether the operation of the robot was correct based on the instructions of a subject for every dialog of the result of the experiment. Moreover, the operation rate for the conventional method was estimated for comparison. The conventional method was considered to be the ideal case in which all the spoken instruction words assumed for each operation are registered. In detail, it was assumed that the main words in connection with a program name were registered as spoken instruction words beforehand. And the operation rate was estimated based on whether the main words are contained in a speech recognition result. For example, in the voice recognition result of Table 5.1, to relate to the program names, "Rugby" and the "weather" presuppose that it is registered. Therefore, in the 1st line of <instruction>, since "Rugby" (italicized part of Table 5.1) is recognized to be "a difference of Rugby", it is considered a success. Moreover, in the 3rd line of <operation>, the reference term the "weather" is not gained at the time of instruction, and the actual experiment using the proposal method has gone wrong. However, since the word related to the weather report program "weather" is recognized, it is estimated to be a success by the conventional method.

The operation rate of the conventional method was computed as the number of presumed successes / the total number of dialogs. The operation rate of the proposal method was computed as the number of operation successes / the total number of operations. As a result, the average of all the subjects' operation rates was 19.5% for the conventional method and 19.3% for the proposed method. Thus, the difference between the result for the conventional method and that fro the proposed method was slight. However, the operation rate of the conventional method is an ideal value based on the assumption of registration of all reference terms. It has sufficient meaning that the ideal value and the actual experimental value in the proposal method based on an experiment are equivalent.

Next, transition of the operation rate according to the number of trials is shown in Fig 5.5. Using the conventional method, in the 1st to the 2nd transition, since the subject gained experience in speech recognition the operation rate has increased, but it does not change in the 2nd to the 3rd transition. On the other hand, using the proposal method, the operation rate is low in the 1st trial since the spoken command words are not gained. However, the operation rate continued to increase through the whole, and it exceeded the conventional method's rate in the 3rd trial. The proposed method can obtain an operation rate exceeding the rate of speech recognition by gaining a subject's spoken command word despite the divergence of the recognition result by mistake. On the other hand, the conventional method cannot obtain an operation rate exceeding the rate of speech recognition. Therefore, the operation rate of the proposal method is considered to have exceeded the operation rate of the conventional method.

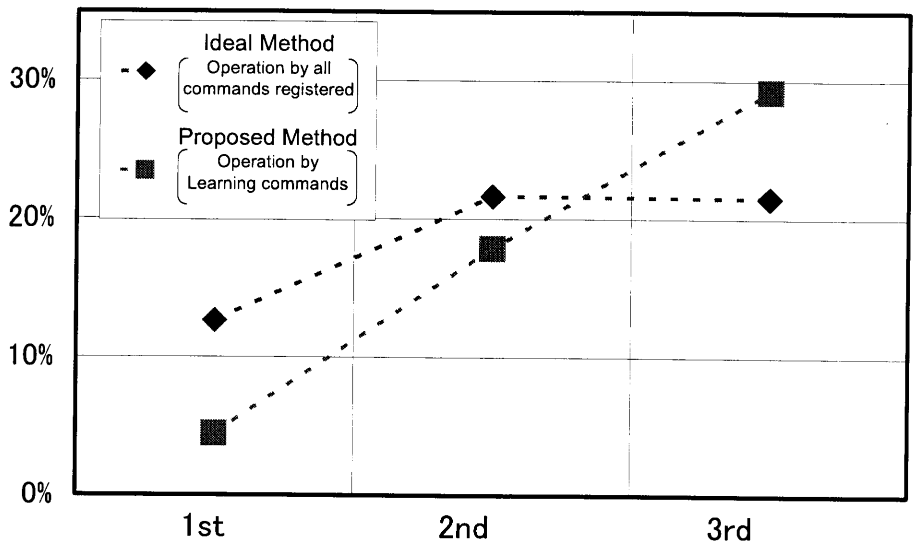


Fig.5.5 Rate Change of Succeeded Operations According to the Experiments

Table 5.5 Number of Dialogs and Operation Results (All trials)

All trials

No.	Gender	Generation	All dialogs	Instuructons	Opeations	Correct recognitions	Rate[%]	Correct operations	Rate[%]
1	Male	Younger	125	44	81	20	16.0%	19	23.5%
2	Female	Younger	197	156	41	23	11.7%	9	22.0%
3	Female	Younger	209	157	52	30	14.4%	17	32.7%
4	Male	Senior	127	89	38	26	20.5%	11	28.9%
5	Male	Senior	134	74	60	12	9.0%	12	20.0%
6	Male	Younger	110	55	55	26	23.6%	11	20.0%
7	Female	Senior	139	93	46	43	30.9%	8	17.4%
8	Female	Senior	132	88	44	34	25.8%	4	9.1%
9	Female	Senior	141	74	67	29	20.6%	5	7.5%
10	Male	Younger	157	83	74	46	29.3%	13	17.6%
11	Female	Senior	125	80	45	23	18.4%	11	24.4%
12	Male	Younger	177	141	36	24	13.6%	5	13.9%
Total			1773	1134	639	336	19.0%	125	19.6%
Average			147.8	94.5	53.3	28.0	19.5%	10.4	19.7%
Max			209	157	81	46	30.9%	19	32.7%
Min			110	44	36	12	9.0%	4	7.5%

Table 5.6 Number of Dialogs and Operation Results (the 1st trial)

1st trial

No.	Gender	Generation	All dialogs	Instuructons	Opeations	Correct recognitions	Rate[%]	Correct operations	Rate[%]
1	Male	Younger	42	28	14	3	7.1%	0	0.0%
2	Female	Younger	31	25	6	2	6.5%	0	0.0%
3	Female	Younger	43	31	12	5	11.6%	1	8.3%
4	Male	Senior	26	21	5	4	15.4%	0	0.0%
5	Male	Senior	48	37	11	0	0.0%	0	0.0%
6	Male	Younger	22	11	11	2	9.1%	1	9.1%
7	Female	Senior	55	42	13	9	16.4%	2	15.4%
8	Female	Senior	51	33	18	13	25.5%	0	0.0%
9	Female	Senior	51	27	24	12	23.5%	0	0.0%
10	Male	Younger	42	29	13	4	9.5%	0	0.0%
11	Female	Senior	28	20	8	3	10.7%	2	25.0%
12	Male	Younger	58	56	2	12	20.7%	0	0.0%
Total			497	360	137	69	13.9%	6	4.4%
Average			41.4	30.0	11.4	5.8	13.0%	0.5	4.8%
Max			58	56	24	13	25.5%	2	25.0%
Min			22	11	2	0	0.0%	0	0.0%

Table 5.7 Number of Dialogs and Operation Results (the 2nd trial)

2nd trial

No.	Gender	Generation	All dialogs	Instuructons	Opeations	Correct recognitions	Rate[%]	Correct operations	Rate[%]
1	Male	Younger	45	9	36	10	22.2%	9	25.0%
2	Female	Younger	63	46	17	5	7.9%	1	5.9%
3	Female	Younger	51	44	7	6	11.8%	4	57.1%
4	Male	Senior	55	35	20	14	25.5%	5	25.0%
5	Male	Senior	32	10	22	10	31.3%	9	40.9%
6	Male	Younger	45	26	19	11	24.4%	3	15.8%
7	Female	Senior	45	24	21	17	37.8%	1	4.8%
8	Female	Senior	35	27	8	6	17.1%	1	12.5%
9	Female	Senior	36	20	16	0	0.0%	0	0.0%
10	Male	Younger	68	35	33	25	36.8%	3	9.1%
11	Female	Senior	41	21	20	12	29.3%	3	15.0%
12	Male	Younger	59	45	14	7	11.9%	3	21.4%
Total			575	342	233	123	21.4%	42	18.0%
Average			47.9	28.5	19.4	10.3	21.3%	3.5	19.4%
Max			68	46	36	25	37.8%	9	57.1%
Min			32	9	7	0	0.0%	0	0.0%

Table 5.8 Number of Dialogs and Operation Results (the 3rd trial)

3rd trial

No.	Gender	Generation	All dialogs	Instuructons	Opeations	Correct recognitions	Rate[%]	Correct operations	Rate[%]
1	Male	Younger	38	7	31	7	18.4%	10	32.3%
2	Female	Younger	52	41	11	7	13.5%	4	36.4%
3	Female	Younger	63	40	23	10	15.9%	5	21.7%
4	Male	Senior	46	33	13	8	17.4%	6	46.2%
5	Male	Senior	53	27	26	2	3.8%	2	7.7%
6	Male	Younger	43	18	25	13	30.2%	7	28.0%
7	Female	Senior	39	27	12	17	43.6%	5	41.7%
8	Female	Senior	46	28	18	15	32.6%	3	16.7%
9	Female	Senior	54	27	27	17	31.5%	5	18.5%
10	Male	Younger	47	19	28	17	36.2%	10	35.7%
11	Female	Senior	56	39	17	8	14.3%	6	35.3%
12	Male	Younger	60	40	20	5	8.3%	2	10.0%
Total			597	346	251	126	21.1%	65	25.9%
Average			49.8	28.8	20.9	10.5	22.1%	5.4	27.5%
Max			63	41	31	17	43.6%	10	46.2%
Min			38	7	11	2	3.8%	2	7.7%

5.4. Impression Evaluation

The subjective evaluation carried out at the same time is described. As a result of conducting factor analysis of the impression evaluation (Table 5.8), the 1st factor (66.5% of contribution) had a high influence on the adjective pair "Unkind-Kind" and “Unpleasant - Pleasant", and so the first factor is named the "familiarity factor". Regarding transition according to the number of times the experiment was performed, when speech recognition was correct and incorrect, the familiarity factor became high as the experiment progressed. This result is connected also with the result of a previous operation rate and shows that the familiarity increased as the user's instruction words were gained.

Next, the elderly people's result is divided and compared with the young people's result (Fig.5.6). When recognition is correct, for both elderly people and young people, the familiarity becomes higher as an experiment progresses. On the other hand, when the voice recognition is erroneous, the young people's familiarity falls remarkably, and it does not become higher as the experiment progresses. In elderly people, even if the voice recognition is erroneous, the elderly people's familiarity is high on the whole and it become higher as the experiment progresses. This shows that learning user-defined voice instructions is also particularly effective for elderly people as well as the tolerant error recognition described in chapter 3.

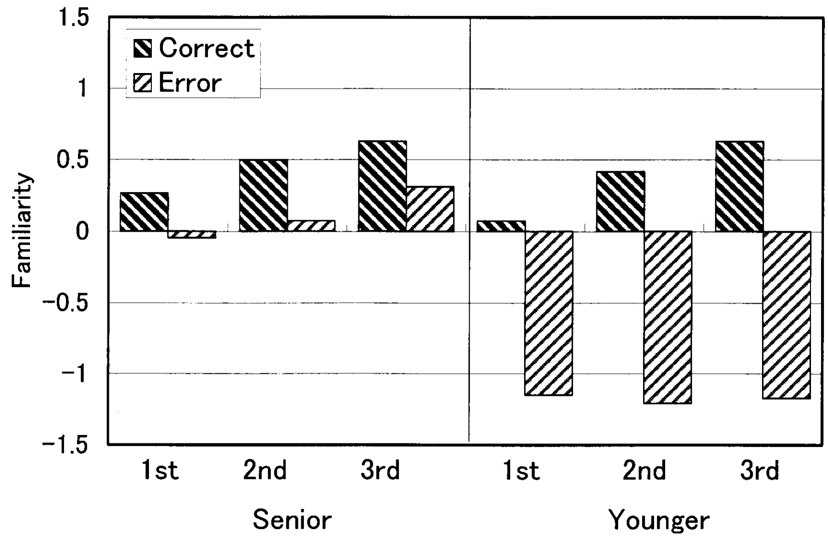


Fig.5.6 Experiment's Comparison of Familiarity

Table 5.9 Evaluated Adjective Pairs and Results

Adjective Pairs		Factor I
Unkind	Kind	**0.91**
Unpleasant	Pleasant	**0.88**
Unintelligent	Intelligent	**0.88**
Disobedient	Obedient	**0.87**
Unfriendly	Friendly	**0.87**
Boring	Interesting	0.86
Unserious	Serious	0.86
Unnecessary	Necessary	0.86
Ugly	Pretty	0.85
Dark	Light	0.85
Difficult	Easy	0.84
Rigid	Frank	0.83
Cold	Warm	0.83
Weary	Unwearied	0.83
Blunt	Sharp	0.81
Slow	Rapid	0.81
Quiet	Showy	0.81
Passive	Active	0.80
Dull	Exciting	0.80
Slow	Quick	0.78
Complex	Simple	0.74
Mechanical	Humanlike	0.73
Comprehensible	Incomprehensible	0.70
Dangerouse	Safe	0.46
Proportion of variance		66.5%

6. Conclusion

At first, I added an infant's familiar behaviors to the robotic interface. The effects of the familiar behaviors installed in a robotic information home appliance, ApriAlpha^{TM}, are studied. The experimental results show that the task execution with speech repetition, object gazing and moving to object has the highest score for the familiar factor. The familiar factors for the case in which the user commands the robot with the familiar behaviors and it makes a mistake and in case that the user commands the television and it operates correctly were almost equal. Therefore, it checked that the introduced the familiar behaviors make a tolerant of the error recognition. Moreover elderly people tend to prefer familiar behaviors more than young people do.

Next, I developed a learning method similar to that employed by an infant. According to this method, when the robot is commanded to perform an operation of which it has not prior knowledge, the robot asks the user to explain the meaning of the operation. Then, the robot memorizes the user's words and the operation related to the words. Next, if the user utters a command using the words, the robot will perform the memorized operation related to the words. We installed this method on the ApriAlpha^{TM} and evaluated the effect by means of a subject experiment. The result indicated that the robot memorizes the user-defined instructions automatically by this method. It enables the user to operate it at a high operation rate as the robot memorizes the instructions.

In conclusion, introducing the familiar behaviors similar to those of an infant increases tolerance of the robot's errors and it is useful for a natural language interface.

Future works; I will carry out prolonged evaluation experiments by a living environment involving various people and assess the validity of the results. Moreover, only directions with voice are used for operation in this paper. However, human uses gesture etc. in addition to voice. If these are also combined, I will think that an interface gets what it is easier to use. I would like to advance the research after this.

Acknowledgements

This research was completed by many people's support and advice.I would like to thank:

Prof. Masatsugu Kidode of NAIST for his research instruction, who pointed out the insufficient part to my research and taught the attitude as a researcher,

Dr. Miwako Doi at TOSHIBA for her exact advice, who show the solution,

Prof. Shigeki Sugano of Waseda University for his instruction as my first boss for a researcher and his advice and comments on improving this thesis,

Prof. Tsukasa Ogasawara and Associate Prof. Yasuyuki Kono of NAIST for their advice and comments on improving this thesis,

Dr. Hirotada Ueda (the present Visiting Prof. of Kyoto University) and the members of the UKARI project for their valuable advice,

Dr. Syun Egusa at TOSHIBA who gave me the chance to go to NAIST, and my bosses in my office,

Dr. Nobuto Matsuhira at TOSHIBA and all members of Robot Development Group in Corporate Research & Development Center for their supports to my research,

and all members in Artificial Intelligence Laboratory of NAIST.

References

[1] Research Committee on Human Friendly Robot: “Technical Targets of Human Friendly Robots”'(in Japanese), The Robotics Society of Japan, Vol.16 No.3, pp.288-294, 1998.

[2] AIBO Official Site: http://www.jp.aibo.com.

[3] Hideki Kozima and Jordan Zlatev: “An Epigenetic Approach to Human-Robot Communication”, IEEE International Workshop on Robot and Human Communication (ROMAN-00), 2000.

[4] Takayuki Kanda, et al.: “ Development and Evaluation of an Interactive Humanoid Robot "Robovie",” IEEE International Conference on Robotics and Automation (ICRA 2002), pp.1848-1855, 2002.

[5] Tetsuya Ogata, et al.: "Development of Emotional Communication Robot: WAMOEBA-2R -Experimental Evaluation of the Emotional Communication between Robots and Humans-", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2000), pp. 175-180, Nov. 2000.

[6] J. Piaget: “Play dreams and imitation in childhood”, New York W. W. Norton, 1962.

[7] Takashi Yoshimi, et al.: “Development of a Concept Model of a Robotic Information Home Appliance, ApriAlpha”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04), 2004.

[8] Michihiko Minoh: “The purpose and abstract of the UKARI project” (in Japanese), Information Processing Society of Japan, The 66th National Convention of IPSJ, Vol.5, No.3TTC-2, pp.5-8, March 10, 2004

[9] Doi Miwako: “Interface Design Based on a Distributed Database of Environment and Human Actions and an Information Flow Energy in a Ubiquitous Environment”(in Japanese), The 66th National Convention of IPSJ, 2004.

[10] Hirotada Ueda: Development of Context-aware Services for the Ubiquitous-Home and a Conversational Interface Robot, IEICE Mobile Multimedia Communications (MoMuC), Vol.105, No.264, pp.1-4, September 8, 2005

[11] Hirotada Ueda, et al.:”A Prototyping of a Dialog Interface Robot at the Ubiquitous-Home”(in Japanese), IPSJ SIG Technical Report, 2005-MBL-32(34),

2005-UBI-7(34), pp239-246(2005)

[12] Hiroshi Shinjo, et al.: "Intelligent User Interface Based on Multimodal Dialog Control for Audio-Visual Systems" (in Japanese), HITACHI-HYORON(日立評論), Vol.87, No.10, pp.47-50(2005)

[13] Yasushi Ishikawa, et al.: "The Current Status and Future Problems of Speech Interfaces for Practical Use" (in Japanese), The 66th National Convention of IPSJ, 1TTB-1 (2004).

[14] Ryo Yoshida, et al.: "Prototyping and Evaluation of a Voice Remote Control System for Home Appliances" (in Japanese), IPSJ Human Interface (HI), 2005-HI-116, pp.65-70 (2005).

[15] Masaki Ida, et al.: "Spoken Language Interface for Network Household Appliances of the Next Generation" (in Japanese), The 66th National Convention of IPSJ, 1TTB-2 (2004).

[16] Takuro Ikeda, et al: "An Experimental Consideration to Achive User Satisfaction for Speech Interface of In-Vehicle Computer" (in Japanese), Human Interface 2006, 2314 (2006).

[17] Naokazu Uchida, et al: "A dialogue Interface for Home Automation Systems using a Semantic Representation Enabling Semantic Structure Mapping" (in Japanese), The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 1B1-07 (2005).

[18] Naoto Iwahashi: "Active and unsupervised learning of spoken words through a multimodal interface", Proc. 13th IEEE Workshop Robot and Human Interactive Communication, pp.437-442 (2004).

[19] Naoto Iwahashi: "Language Acquisition by Robots Towards New Paradigm of Language Processing" (in Japanese), Transactions of the Japanese Society for Artificial Intelligence, Vol.18, No.1, pp.49-58 (2003).

[20] Takayuki Kanda, et al.:" An evaluation on interaction between humans and an autonomous robot Robovie" (in Japanese), The Robotics Society of Japan, Vol.20, No.3, pp.1-9 (2002).

[21] Terrence Fong, et al.: “A survey of socially interactive robots”, Robotics and Autonomous Systems, 42, 143-166, 2003.

[22] Cynthia Breazeal: “Toward sociable robots”, Robotics and Autonomous Systems, 42, 167-175, 2003.

[23] Masatomo Hashimoto, et al.: “Ubiquitous Computers in Living Room - An User Interface System for Network Home Appliances –”, IPSJ Magazine Vol.48, No.1, pp.53-58, 2007

[24] Belinda Buckley: “Children's Communication Skills: From Birth to Five Years”, Routledge ISBN-13: 978-0415259941 , 2003

[25] Tetsunori Kobayashi: “Trend of Spoken Dialogue Research”, Journal of the Japanese Society for Artificial Intelligence, Vol.17, No.3, pp.266-270, 2002

[26] Nielsen, J.: “Usability Engineering”. Morgan Kaufmann Publishers, San Francisco, USA. 1994

Publication list

Journal paper

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Practical Robotic Interface Innovated by Familiar Behaviors -Tolerant Behaviors of Voice Recognition Error-(in Japanese)", Journal of Human Interface Society, Vol.8 No.2 pp.247(45)-253(51), May 2006

International conference (Reviewed)

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Behavior Fusion in a Robotec Interface for Practicality and familiarity -Approach by Simultaneous Imitations-", 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2004) , No.8A1, September 22, 2004

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: Familiar Behaviors Evaluation for a Robotic Interface of Practicality and Familiarity, 4th IEEE International Conference on Development and Learning (ICDL-05), pp.149-154, July 20, 2005

Takashi Yoshimi, Nobuto Matsuhira, Kaoru Suzuki, **Daisuke Yamamoto**, Fumio Ozaki, Junko Hirokawa and Hideki Ogawa: "Development of a Concept Model of a Robotic Information Home Appliance, ApriAlpha", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04), 2004.

International conference

Miwako Doi, **Daisuke Yamamoto**, Hirotada Ueda and Masatsugu Kidode: "Familiar Styles of Robotic Interactions Using Home Appliances", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06) Workshop: Network Robot Systems, 2006

Domestic conference

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "A robotic interface mechanism for fusion of practical behaviors and familiar behaviors (in Japanese)", 28th Human Interface Society Correspondences on Human Interface, Vol.6, No.3, pp.65-68, June 8, 2004

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "A robotic interface that combines practical behaviors (in Japanese)", JSME Conference on Robotics and Mechatronics '04 (ROBOMEC2004), 2P2-H-36, June 20, 2004

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Behavior Fusion in a Robotic Interface for Practicality and Familiarity -Approach by Imitations for Familiar Behaviors- (in Japanese)", 3rd IEICE/IPSJ Forum on Information Technology (FIT2004), Vol.3, No.K-095, pp.607-608, September 9, 2004

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Behavior Fusion in a Robotic Interface for Practicality and Familiarity -Evaluation of the Usability by Familiar Behaviors- (in Japanese)", 67th National Convention of IPSJ, Vol.4, No.3H-5, pp.4-29 - 4-30, March 3, 2005

Daisuke Yamamoto, Tetsu Sato, Toshikazu Wada, Hideki Ogawa, Miwako Doi, Hirotada Ueda and Masatsugu Kidode: "Development of Locomotion Control Method by Tracking with Floor Pressure Sensors-Approach by Human Tracking for Practicality and Familiarity- (in Japanese)", JSME Conference on Robotics and Mechatronics '05 (ROBOMEC2005), 2P1-N-050, June 11, 2005

Daisuke Yamamoto, Miwako Doi, Hirotada Ueda and Masatsugu Kidode: "Practical Robotic Interface Innovated by Infant's Familiar Behaviors (in Japanese)", 1st Conference

of Transdisciplinary Federation of Science and Technology, No.B2-23, pp.421-422, November 26, 2005

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Practical Robotic Interface Innovated by Familiar Behaviors -Interface by User-defined Instructions- (in Japanese)", IPSJ Human Interface (HI), Vol.2006-HI-117, No.16, pp.111-116, January 13, 2006

Daisuke Yamamoto, Yosuke Tajika and Miwako Doi: "A Research for Middleware and Robotic Interface Using Ubiquitous Information (in Japanese)", 3rd International Workshop on Ubiquitous Home, pp.65-72, March 2, 2006

Daisuke Yamamoto, Miwako Doi, Hideki Ogawa, Hirotada Ueda and Masatsugu Kidode: "Practical Robotic Interface Innovated by Familiar Behaviors-Evaluation of the Robots in the Different Shapes- (in Japanese)", JSME Conference on Robotics and Mechatronics '06 (ROBOMEC2006), 1A1-E35, May 27 , 2006

Daisuke Yamamoto, Miwako Doi, Hirotada Ueda and Masatsugu Kidode: "Robotic Interface Learning Voice Instructions Defined by User (in Japanese)", 38th Human Interface Society Correspondences on Human Interface, pp.11-16, May 29, 2006

Daisuke Yamamoto, Miwako Doi, Nobuto Matsuhira, Hirotada Ueda and Masatsugu Kidode: "Home Robotic Interface Innovated by Familiar Behaviors (in Japanese)", 49th Japan Joint Automatic Control Conference, SU1-3-4, November 26, 2006