Conversation System of an Everyday Robot Robovie-IV

Noriaki Mitsunaga[1], Zenta Miyashita[2,1], Takahiro Miyashita[1],
Hiroshi Ishiguro[2,1] and Norihiro Hagita[1]
[1]ATR Intelligent Robotics and Communication Laboratories
[2]Graduate School of Engineering, Osaka University
Japan

1. Introduction

Robots are expected to play a much more active role in our daily lives in the near future. Since there are large differences in social aspects between short-term interaction in a laboratory and long-term interaction in real life, it is imperative to conduct experiments with a robot that interacts with people every day in their real daily lives to develop such robots in the future.

Jijo-2 (Matsui et al., 1997) was developed as an office-conversant robot that informs people where a person is, guides them to a room, and so on. Also, a museum guiding robot (Burgard et al. 1998) has been developed. They both interacted with people on a daily basis, and their interactions were designed to achieve tasks such as providing information about a person, guiding a visitor to a room, or explaining displays. However, communication is not just for achieving tasks but also for building social relationships. For example, when people chat with others, there may be no explicit intention to exchange information; it is just conversation. Although such kinds of interaction without explicit intention might seem unnecessary, we think they increase familiarity which serves as the basis of smooth communication to achieve tasks. This means we will prefer the robot most familiar to us when encountering two robots that are functionally identical.

We call such an interaction that increases familiarity and facilitates communication "social interaction." Our interest is in how to realize such interactions in a robot and what makes an interaction a social one. To investigate these issues it is necessary to perform experiments in daily life environments. For the first step of our research, we have developed a human-sized version of Robovie, Robovie-IV. Robovie-IV features interaction abilities including voice chats, which are intended to be "social interactions." Using Robovie-IV we conducted an experiment in which it interacted with people in their daily life in our office environment for six months.

This chapter is arranged as follows. In Section 2, we discuss how to implement daily interactions between a human and a robot, and then in Section 3 we discuss the requirements for a robot to carry out everyday communication. In Sections 4 and 5 respectively we briefly introduce the hardware and software architecture of Robovie-IV. Section 6 outlines the implementation of the base software for the daily interaction, and

Section 7 describes our experiment and observations. Finally, we discuss the results and offer conclusions in Section 8.

2. Daily interaction between humans and robots

Isbell et al. (Isbell et al., 2001) have realized an agent that can participate in a chat session over the Internet. The agent learns how to interact in a chat with rewards given by other human participants. Kubota et al. (Kubota et al., 2000), meanwhile, have created an agent that chats with people using a type of corpus based on exchanged E-mails on a mailing list that discusses drinks. These agents and the method to build the corpus are free from speech recognition, speech synthesis (text-to-speech), and so on since electrical texts are given.

It is still difficult, however, to dictate our daily conversations and to build a usable corpus by current voice recognition methods. Also, there are many differences between chats over the Internet and our daily face-to-face chats. To the best of our knowledge, there is no corpus presently usable for daily human-robot conversation. Consequently, we have decided to implement rule-based interactions between a human and a robot including chats. There are four modes in human-robot interaction:

* Mode A) the human makes the first move and does not wait for the robot's reaction;
* Mode B) the human makes the first move and expects a reaction from the robot;
* Mode C) the robot makes the first move and does not wait for the human's reaction; and
* Mode D) the robot makes the first move and expects a reaction from the human.

Patting a robot's shoulder or head is an example of interaction Mode A. In the case where a human passes by without expecting a reaction from the robot, it is not necessary to explicitly implement interactions. Approaching, touching, talking to, and making a gesture to the other are examples of Modes B and D, while saying "hello" is an example of interaction Mode C. We regarded Mode D as most important and mainly implemented the interactions where the robot actively tries to talk with a human when the distances between them are within a certain range. For Mode B, we mainly implemented reactions in which a human touched the robot's skin.

We classify the interactions in Mode D into four types, according to sensor information:

* Type1) interactions in which a robot can continue to play regardless of the sensor values, without making people feel that there is something wrong with them;
* Type2) interactions that may comprise reactions that rely only on simple sensor information;
* Type3) interactions that need voice recognition, but mis-recognitions do not lead to severe discomfort; and
* Type4) interactions that need voice recognition, and mis-recognitions lead to severe discomfort.

Greetings and performances like a dance are examples of Type1 interactions, and giving responses is an example of Type2 interaction. When a person is drinking something, the robot might ask "What are you drinking?" Without recognizing the answer, the robot can say "I would like to drink it too," A Type2 interaction does not need voice recognition even if the communication type is a chat. Both Type3 and 4 interactions use the result of the voice recognition system; the differences are their required recognition accuracy. Suppose the robot says "Let's play," and fails to recognize the person's vocal response of "no," and it starts to play. The human may feel it is selfish but this is not serious problem. However, if

the voice recognition fails for a word used in the robot's response from the robot, the human will easily detect the failure.

We are mainly implementing Type2 and 3 interactions because too many Type1 interactions reduces humans' interest in the robot, and Type4 interactions disappoint the human if voice recognition fails. For Type4 interactions, we needed to carefully construct a voice recognition dictionary in order to reduce the number of failures.

To talk about current events, we prepared fixed-form rule-based chats that included sensing results, chat histories, dates and times, and information from the Internet. For example, how many people the robot met today, who it met, current weather, and so on. We incrementally added interaction rules to Robovie-IV since we incrementally learned what were expected interactions. Robovie-IV had 12 Type1, 11 Type2, 49 Type3, and 2 Type4 interactions at the end of the experiment.

3. Requirements for a communication robot

What capabilities are necessary for a communication robot? At least the following requirements are needed to achieve natural and effective human-robot communication. First of all, the robot should be self-contained. Although its computers, database systems, and even sensors can be outside its body, it should have mechanisms to perform both verbal and nonverbal communication by itself. Furthermore, it should be able to move around with no wires for smooth communication because the distance between a person and a robot cannot be ignored when attempting to achieve effective communication. Also, it should not have an outward appearance that frightens or discomforts people.

Second is the capability of haptic communication because haptic communication is as important as vision and voice. People who are familiar with each other often touch each other's hair or hug each other; such haptic interaction reinforces their familiarity. If a communication robot equipped with tactile sensors over its entire body could have the same haptic interaction capability as human do, the robot would become more familiar, thus shortening its communicative distance from people. To study haptic communication, we have previously developed two types of humanoid robots, Robovie-IIS and Robovie-IIF, that have tactile sensors embedded in a soft skin that covers the robot's entire body (Miyashita et al., 2005). These robots were developed based on Robovie-II (Ishiguro et al. 2001). Fig. 1 shows overall views of Robovie-IIS, Robovie-IIF, and a scene of communication with a human. Robovie-IV has tactile sensors based on the technique we used for these robots.

Figure 1. From left to right, Robovie-IIS, Robovie-IIF, and haptic interaction between children and Robovie-IIS.

Third is the locomotion mechanism that can generate involuntary motions. Human motion can be classified into two types: voluntary and involuntary motion (Bruno, 2002, Zheng, 1990, Wu 1991). Voluntary motions are a set of motions made to achieve given tasks or intentions. Going to a certain place, moving an arm forward for a handshake, vocalization to say hello, and reacting to a pat in order to know who did it are such examples. Involuntary motions, on the other hand, are a set of incidental motions such as feedback control arising in response to physical stimuli from the environment without prior planning or motivation. These motions do not correspond to tasks directly but instead consist of motions that make the robot appear to behave more naturally. Robovie-III was developed to enable involuntary motion for a robot (Miyashita & Ishigruro, 2004). It uses a wheeled inverted pendulum mechanism for locomotion. Since the inverted pendulum is controlled by feedback on its posture, involuntary motion of the whole body occurs. In addition to effecting involuntary motion, the wheeled inverted pendulum has a unique feature: When some external force is applied to the body from its front, it moves backwards. Since its wheels are controlled to maintain its posture, it moves backwards to keep the posture as it tilts backwards due to the applied force. We also adopt the wheeled inverted pendulum mechanism for Robovie-IV.

Fourth is human recognition. People expect a robot to be able to find and identify them. Many methods for human detection and identification using images have been proposed; however, with current computational power and video camera resolution and viewing angles, conditions under which those methods work are still limited. Consequently, we decided to use a laser range sensor to find human leg candidates and an optical/RF tag system for human identification. Robovie-IV finds leg candidates using a laser range sensor, verifies the results with its camera, and identifies the detected human with a tag.

Based on the above discussion, Robovie-IV is designed as a robot 1) whose height is the same as a child's; 2) whose whole body is covered with soft, light-colored skin for a soft look and touch; 3) with many tactile sensors are embedded in the soft skin; 4) which can move in two modes, one is a normal wheeled robot mode with passive casters, and the other is the inverted pendulum mode; and 5) which has optical and RF tag readers and laser range sensors for human identification.

4. The hardware Implementation

Fig. 2 shows front and side views of Robovie-IV. The height of the robot is about 1.1 m, which is smaller than the Robovie-II/II-S/II-F models (which are 1.2 m high). As the figure shows, it has two arms each with four degrees of freedom, one head with pan, tilt, and roll joints, four support wheels (passive casters), two drive wheels, and one jack to enable the inverted pendulum state. Robovie-IV is equipped with two video cameras with pan and tilt joints, one camera with an omni-directional mirror, a microphone (attached on the omni-directional mirror as in Fig. 3) and speaker, an optical tag reader (easily detachable, not shown in the figure), and an RF tag reader (the system is based on the active tag system called Spider by RF Code Inc.). Robovie-IV also features two laser range sensors in the front and back, two gyros that sense the same axes (we use two to take the average of their outputs in order to reduce random noise, and to detect failure so that the robot does not fall over), and 56 tactile sensors.

camera with omni-
directional mirror
and a microphone

2 cameras with
pan and tilt

*neck
(3dof)*

1 optical
tag reader

*2 arms
(4dof x2)*

1 speaker

56 tactile
sensors

1 RF tag
reader

gyro sensor
(1 axis)

56 tactile
sensors

2 laser range
sensors

(diameter of wheels)

4 support wheels
(passive casters)

200

*1 jack
(1dof)*

2 wheels

Front view
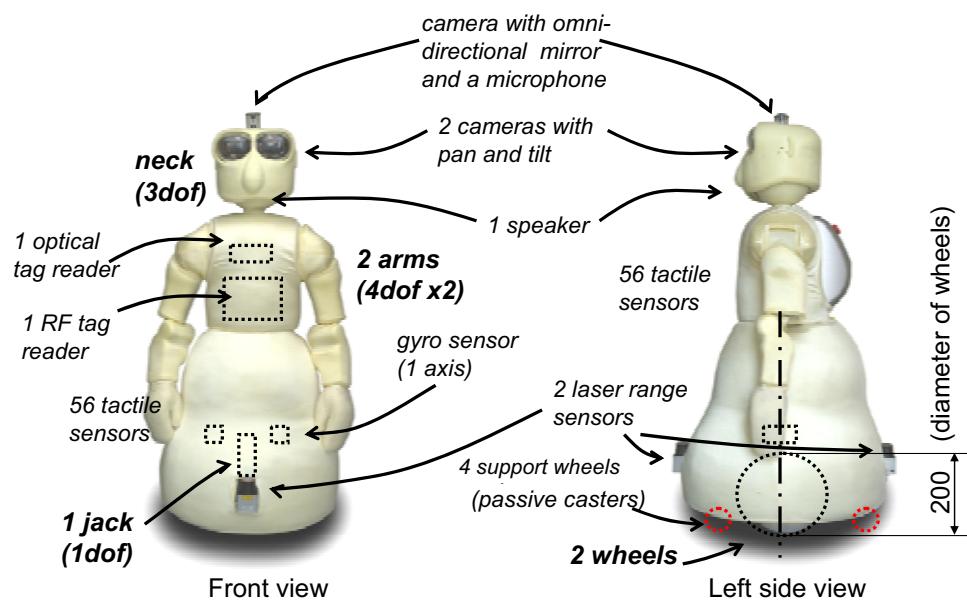
Left side view

Figure 2. Front and left-side views of Robovie-IV. Fitted actuators and sensors are shown.

Figure 3. A microphone, attached on a omni-directional mirror, is located on top of the head
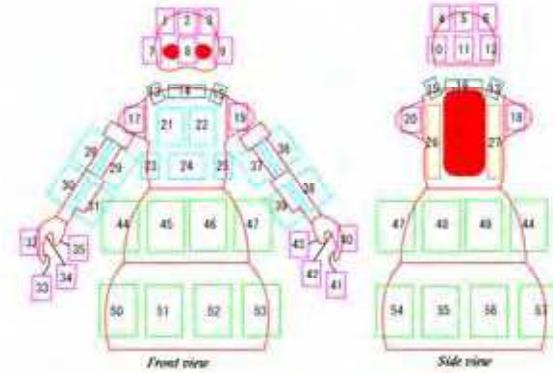
Figure 4. Arrangement of the skin sensor elements. There are 56 sensors in its soft skin (#16 is not implemented).

Fig. 4 displays the arrangement of the sensor elements that are embedded in the soft skin. The figure clearly shows that there are tactile sensors under the nose, ears, and in the thumbs for Robovie-IV to sense contact to those areas.

Fig. 5 shows Robovie-IV's hardware architecture. There are four PID motor controllers connected to a PC via RS-232 dedicated for wheels, left and right arms, and neck joints. The jack and the inverted pendulum are controlled by the motor controller. Two laser range sensors and optical and RF tag readers, and cameras with pan/tilt joints are also connected to the PC via RS-232. Signals from the tactile sensors are fed to five skin processors. The processed signals are then sent to the PC via the RS-422 bus. Images from three cameras are captured by a frame grabber installed on the PC. A speaker and a microphone are connected to the sound output and input of the PC. Each controller has a SH2 micro-processor (Renesas Technology Corp.) and the main computer has a Pentium-M processor that runs at 2 GHz. The PC can be connected to the sensors via a wireless LAN.

The flow of sound signals is shown in Fig. 6. The sound signal recorded by a microphone (monaural) is fed to two high-pass filters (f_c = 100Hz) followed by a low-pass filter (f_c=2kHz). The LPF's output is fed to the input of the PC's sound card. The two HPFs and the LPF are 2nd order active filters which consist of OP-amps. The HPFs are used to reduce wind noise. The sound signal from the microphone is sampled at a 16kHz sampling rate with 16 bits of accuracy by a speech recognition program. The output from the PC is amplified and fed to a speaker which is located in the head of Robovie-IV.

Fig. 7 shows the structure of a tactile sensor element embedded in the soft skin. As the figure illustrates, the soft skin consists of four layers. The outside layer is made of thin silicone rubber, and the middle layer is made of thick silicone rubber. We use these silicone rubber layers to realize humanlike softness. The inner layer is made of urethane foam, which has a density lower than that of the silicone rubber; the densities of the urethane foam and the silicone rubber are 0.03 g/cm³ and 1.1 g/cm³, respectively. The total density of the soft skin, including all layers, is 0.6 g/cm³. Robovie-IV's tactile sensor elements are film-type piezoelectric sensors inserted between the thin and thick silicone rubber layers. These film-type sensors, consisting of polyvinylidene fluoride (PVDF) and sputtered silver, output a high voltage proportionate to changes in applied pressure. Since the middle and the inner

layers deform easily upon human contact with the skin, the sensor layer can easily detect the contact.
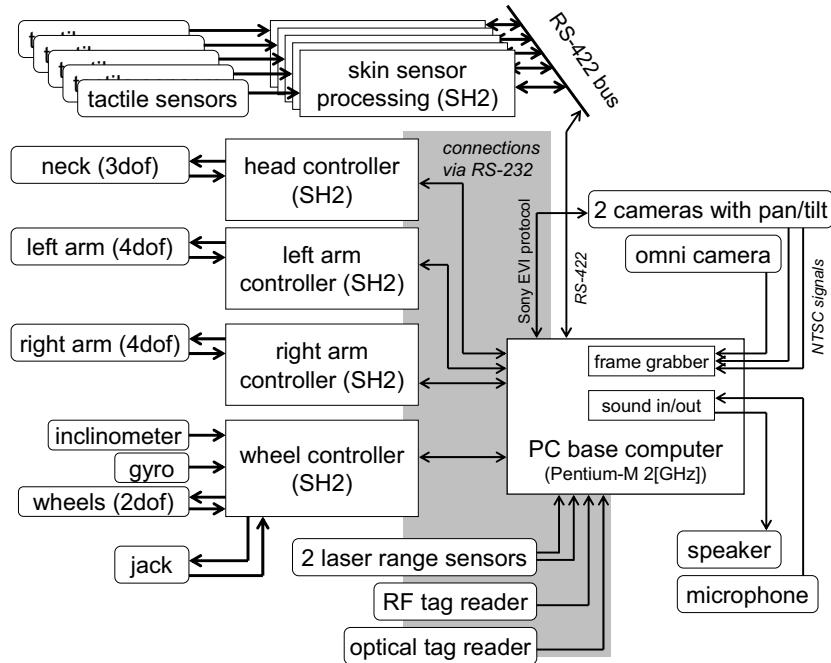
Figure 5. Robovie-IV's hardware architecture. There are four motor controllers, five skin processors, and one main computer.
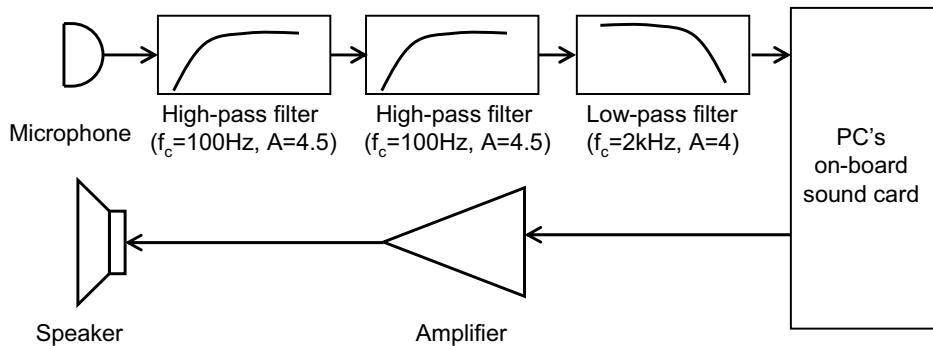
Figure 6. The flow of sound signals. The sound signal from a microphone is fed to the PC's on-board sound card via two HPFs and a LPF. The sound from the PC is amplified and fed to a speaker.
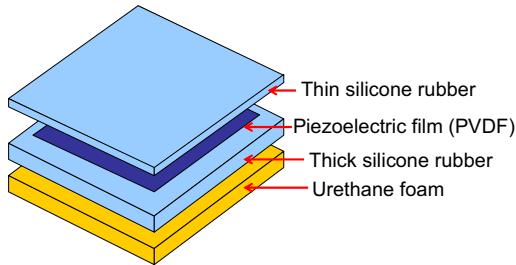
Figure 7. The skin's structure, which consists of four layers. The piezoelectric sensor is inserted between the outer and middle layers.

Figure 8. The drive wheels are connected by a jack to the main body and support wheels. The figure shows the inverted pendulum mode.

Fig. 8 illustrates how the drive wheels, the support wheels, and the jack are connected. The support wheels are connected to the main body directly, while the drive wheels are connected via the jack. Robovie-IV can select two locomotion modes, with or without support wheels, by controlling the jack. With the support wheels, Robovie-IV moves as a normal robot with two drive wheels, but without them, it moves in the inverted pendulum mode. We use the wheeled inverted pendulum controller proposed by Ha and Yuta (Ha & Yuta, 1994).

5. The Software Implementation

Fig. 9 presents an overview of Robovie-IV's software architecture. The PC's operating system is Linux, and a box with bold line in the figure indicates a process running on Linux. There are five processes, *robovie4, robobase4, robomap4, PostgreSQL*, and *Julian*. The processes are connected through named pipes, sockets, and shared memory. The process *robovie4* makes decisions for the robot from the internal state, information in the database, and the

sensor information gathered and processed by the other processes. Processors on the motor controllers control motors as directed by a program called *robobase4* running on the PC. Process *robobase4* handles the communication between controllers and sensors while hiding the differences of the protocols from *robovie4*. Process *robomap4* is the one for self-localization. It receives information from the two laser range sensors and odometry from *robobase4* and compares this information with the map of the environment. The estimated position is returned to *robobase4* and sent to *robovie4*. The *PostgreSQL* (see http://www.PostgreSQL.org/) is a popular database engine. We use the database to store and recall co-experiences with humans.
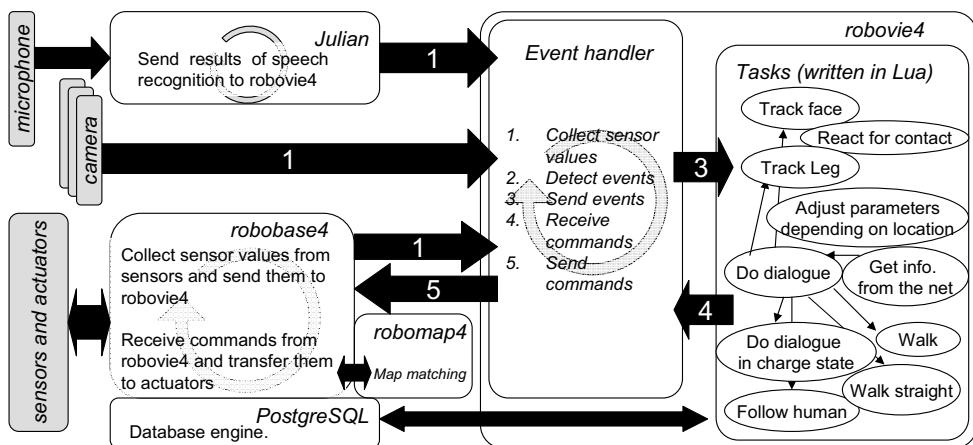

Figure 9. Robovie-IV's software architecture. The PC's OS is Linux. There are five processes *robovie4, robobase4, robomap4, PostgreSQL,* and *Julian.*

The *Julian* process is that of a grammar-based recognition parser called "*Julian.*" *Julian* is a modified version of *Julius*, which is a high-performance, two-pass large-vocabulary continuous speech recognition (LVCSR) decoder software for speech-related researchers and developers. We have prepared manually designed deterministic finite automata (DFA) grammar as the language model and use it for the speech recognition by Julian. The *robovie4* process connects to the *Julian* process and commands dictionary selection and so on. The recognized words are sent back to *robovie4*.

We adopted a text-to-speech program called *CHATR* (Campbell & Black, 1996) to produce Robovie-IV's voice. The *robobase4* process calls *CHATR* when it is commanded to utter text by the *robovie4* process. The synthesized voice is fed to the speaker through the on-board sound card in the PC.

The *robovie4* process mainly consists of an event handler written in C++ and tasks written in Lua (see http://www.Lua.org/), which is a scripting language that can be embedded in a C or C++ program. Tasks written in Lua run in parallel by the collaborative multithreading supported by Lua. There are task threads to do dialogue, react to tactile events, track a human's face, track legs, and so on. The events from sensor values are detected in the event

handler, negating the need to copy an event detection for each task thread. The event handler in *robovie4* continuously,

1. collects sensor values from *julian* and *robobase4*;
2. detects events;
3. sends events to task threads;
4. receives commands from task threads; and
5. sends commands to *robobase4* as in Fig. 9.

Each of the task thread continuously,

1. waits for events from the event handler;
2. reads sensor values sent from *julian* or *robobase4*; reads from and writes to the database; and
3. decides the next action.

The dialogue task is responsible for human-robot interactions and controls other task threads. In terms of the four modes of human-robot interaction defined earlier, the dialogue task does not react to Mode A (human-initiated non-waiting) interactions since no reaction is expected. It just records the event, such as someone patting the robot's head. When a Mode B (human-initiated waiting) interaction, such as a human approaching the robot, is detected, the task tries to take the initiative and move to the interaction Mode D (robot-initiated waiting) by initiating a dialogue. Although people may feel that the robot interacted in Mode C (robot-initiated non-waiting), no Mode C interaction is explicitly implemented.

An interaction in Mode D repeats the following until its end.

1. The robot speaks and makes a gesture,
2. the robot waits for an answer or reaction from the human, and
3. the robot reacts according to the answer.

The Type1 interactions do not have steps 2) and 3). The Type2 interactions do not use the result of *Julian* but just use the fact that the human utterance ends. The Type3 and Type4 interactions react according to the result of voice recognition. The Type3s do not repeat a word which the human uttered, while the Type4s directly include it in the robot's answer.

6. The daily interactions

Basically, Robovie-IV repeats (a) moving around, seeking a human to interact with; and (b) interaction with the found human until he or she quits the interaction as Fig. 10 shows. Robovie-IV detects humans by first finding leg candidates from front laser range-sensor values by background subtraction, large sound, or reading of RF-ID tag. The laser range sensor measures distances at the height of an ankle. Fig. 11 illustrates the readings of the range sensor plotted on the floor map. The circle on the right and the line indicates the position and the posture of Robovie-IV from self-localization data based on map matching. Clearly, the walls and cartoon boxes are detected by the sensor, whereas distant walls are not detected since the sensor's maximum range is 4 m. The two small circles in front of the circle representing the robot indicate detected legs. The circle on the left of the robot, however, is a falsely detected leg candidate. When a leg candidate is found, Robovie-IV rotates itself and gazes in the direction of the leg candidate. If a large skin-colored blob or a large blob in a differential image is detected by Robovie-IV's camera, or the tag reader detects a valid ID, it assumes the candidate to be a human leg. During interaction, it tracks the skin-colored blob at the shortest distance in the direction of the leg candidate and

maintains that distance and posture with respect to the leg candidate. The tag reader's attenuator is adjusted to detect tags within 1.5m of Robovie-IV.

Fig. 12 shows the flow of one of Robovie-IV's typical interaction. Once it has detected a human (Fig. 12 (a)), it will try to interact with him or her. It will say "hi" in Japanese and wait for the response (Fig. 10 (a) and 12 (b)). If it does not detect a response it will begin to seek another human; otherwise, it will recognize the person as an interaction partner. Then it starts a dialogue as in Fig . 10 (b) and Fig. 12(c). It continues by selcting a topic and trying to chat with the person. In the chat, it repeats the following sequence of actions: 1) stop speech recognition, 2) do a gesture and/or say something, 3) change DFA grammar of *Julian* and start speech recognition (not for Type2 interactions), 4) wait for a reply, a reaction, or some time to elapse. The end condition of the chat is met when the person says "bye-bye" or goes away. Then, it says "bye-bye" and searches for the next person to talk with (Fig. 12 (d).) Interaction rules are derived from the Lua program (functions), gestures, and DFA grammars for Julian. There are more than 70 interaction rules and Lua codes implemented, with more than 5,000 lines of code. There are three branches on average on the conditional branches of Type2, 3, and 4 interactions.

(a) flow of moving around and
seeking a human to interact with

(b) flow of interaction with a human
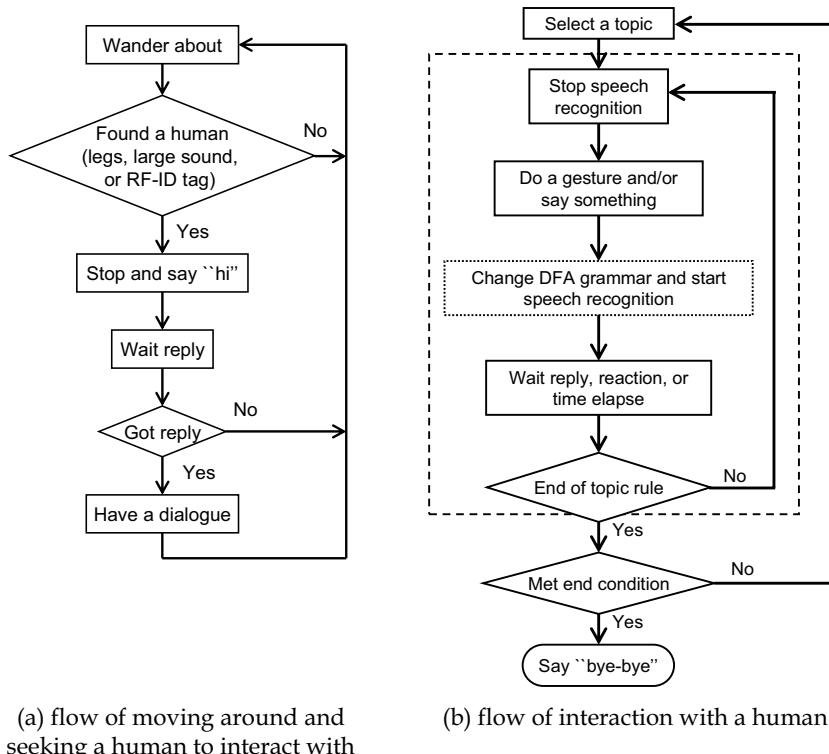
Figure 10. Flow of Robovie-IV's daily interaction. Robovie-IV wander around to search for a human to talk with (a) and have a dialogue (b).
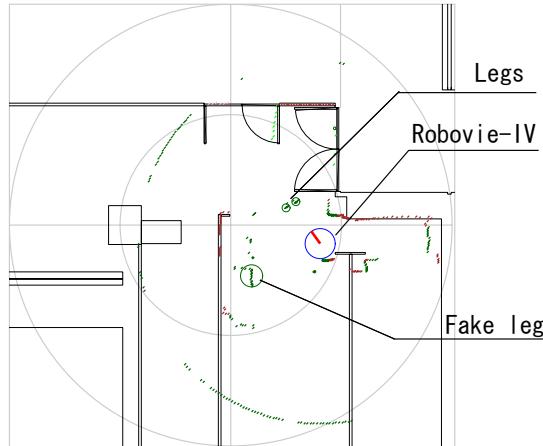
Figure 11. A depth image plotted on the floor map by a laser range sensor. A circle and a line on the right indicate Robovie-IV's position and posture. Two circles in front of the robot are detected legs. The larger circle to the left of the robot is a false detection.

1) search for a human to talk with

2) find a human and say "hi"

3) have a chat
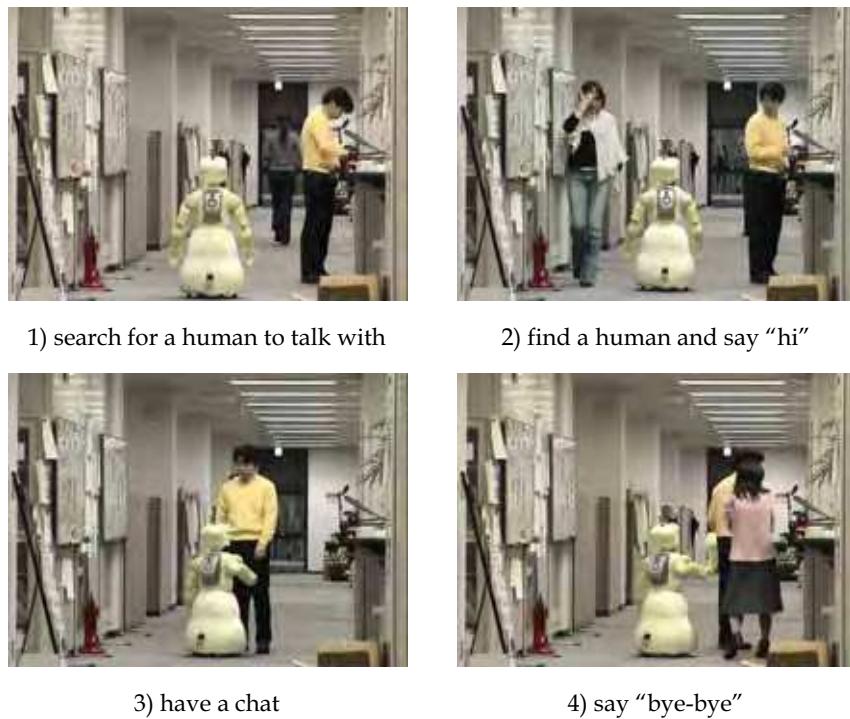
4) say "bye-bye"

Figure 12. Flow of a typical interaction that Robovie-IV tries to do. 1) it searches for a human to talk with, 2) say "hi", 3) have a chat, and 4) say "bye-bye".

Robovie-IV records the results of interaction to the database. Examples of results include with whom it interacted, which rule it used, when the interaction rule started and finished being used, the response from the human, how long the human moved, how many times it met with a person, and so on. It also records the number of leg candidates, skin-colored blobs, and Ids detected by the tag reader in a certain interval. Robovie-IV uses the data in the database in order not to repeat same interaction, to say it has met many people today, and so on. It also talks about current weather based on the information from the Internet.

7. Observation of Daily Human Robot Interaction in an Office Environment

We have conducted a long-term experiment in our office at ATR's IRC laboratory for six months. Robovie-IV roamed around the office and interacted in our daily life. In the office there were about 100 people including researchers, engineers, and secretaries. The total length of the corridor in which Robovie-IV could move was about 150 meters. We implemented interactions resembling those of a child in an office who comes along with its parent, since the height of the Robovie-IV is about 1.1 m, similar to that of a child. We incrementally added interactions and updated other items of software during the experiment. Fig. 13 (a) and (b) show a scene in the office and an example of interaction.

Fig. 14 shows some scenes from the daily experiment, (a) Robovie-IV trying to communicate, (b) a person patting the robot's head while passing by, (c) two people talking to each other, one toying with the robot's nose, (d) a person covering the robot's eyes, and (f) Robovie-IV interacting with a person. In scene (e), following its usual human-seeking behavior, Robovie-IV moved into the middle of a conversation in progress. At this stage, of course, Robovie-IV is unable to meaningfully join such a conversation. Robovie-IV did not always interact with detected humans as in Fig. 14 (a) because the experimental environment was an office and people were busy with their jobs. Though the novelty of having Robovie-IV in the office had worn off, people still noticed it out of the corner of their eye while working, indicating that they were still aware of its presence, and that it was gradually changing its behavior. Robovie-IV is still under development, but it has become a robot that makes its presence felt by people.

(a) Robovie-IV is roaming in the office at ATR's IRC laboratory

(b) A person is communicating with Robovie-IV. The robot is asking if the person likes watching baseball

Figure 13. Scenes from the experiment in our office. (a) shows Robovie-IV and the office that it roamed around. (b) shows an example of interaction

(a)Robovie-IV trying to communicate

(b)A person patting the robot's head while passing by

(c) Two people talking to each other, one toying with the robot's nose

(d) A person covering the robot's eyes

(e) Robovie-IV moving into a conversation in progress

(f) Robovie-IV interacting with a person
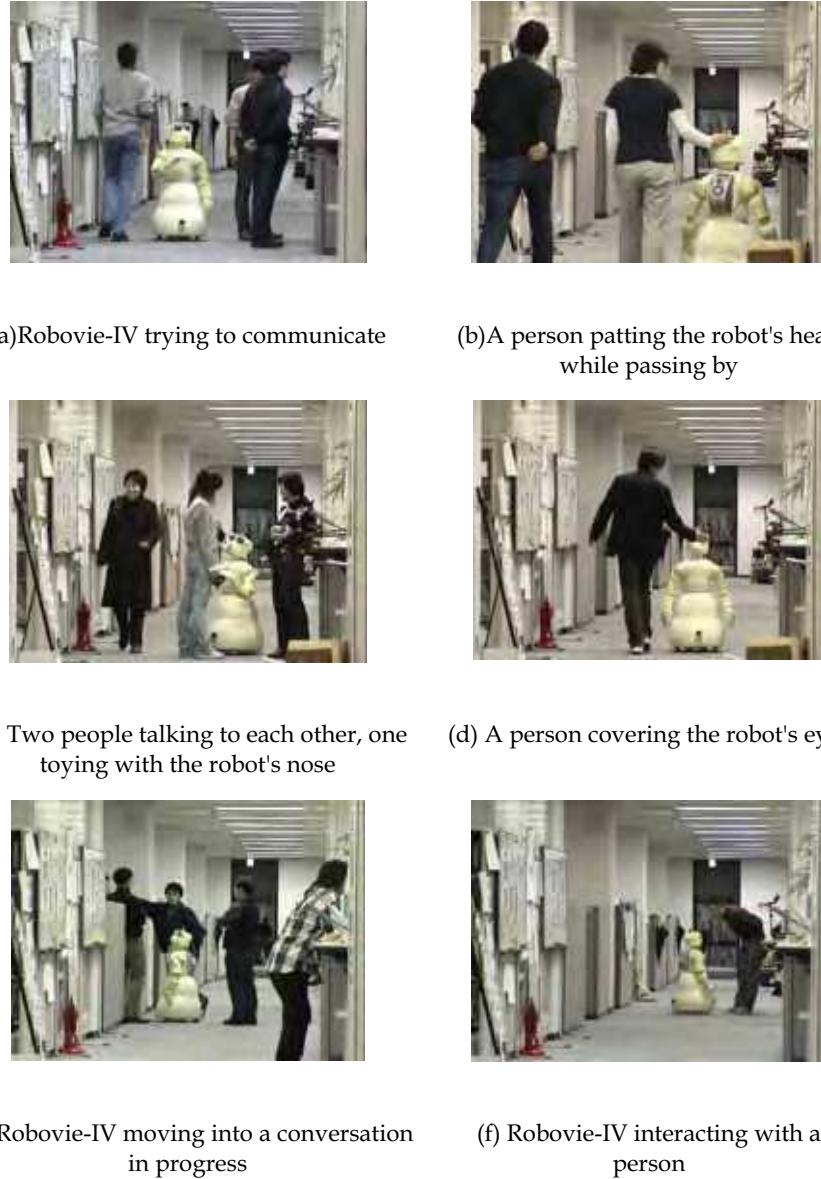
Figure 14. Some scenes from the daily experiment.

There still remain some technical issues to be solved. One is the timing to say hello to a person passing by. There were cases that he or she would walk away before Robovie-IV started its greeting. The second issue is that there are people who are not getting used to conversing with Robovie-IV. *Julian* does not allow overlapping of human and robot voices

without external sound processing. Also there is a sweet spot in the microphone where the success rate of recognition increases. We do not have answers to deal with these issues yet.

Furthermore, there are different feelings toward what a robot actually does. When Robovie-IV asks, "What are you drinking?", even if a person is not drinking, one would feel it is very strange that the robot does not understand the situation. The person might feel that is interesting, however, that the robot is making small talk, so he or she answers the question. What kinds of interaction increase familiarity and make communication smoother is still open question. We think this question will be answered by having more experiment and adding more types of interaction.

8. Fundamental Issues and Conclusions

An interaction robot communicating with people in everyday life is the key to revealing the social relationships of a robot among people. Long-term everyday interactions are necessary to observe and analyze the following;

- Influence of a robot on people's sense of time, e.g. whether a robot's behavior reminds us of the current time; for example, if the robot establishes a "habitual" routine of behaviors based on the time of day, could it serve to make people aware of the time?
- The relationship between a robot's task and its preferred social interactions, e.g. what kinds of interactions are necessary for a cleaning robot, a secretary robot, a guiding robot, or a robot that carries heavy objects in tandem with a human;
- Preferred types of human-robot dialogues according to the social relationship and the robot's task, e.g. whether simple statements are preferred, colloquial language is preferred, or new type of language emerges;
- The social role of a robot among people, e.g. whether a robot will be perceived to be "alive" in a way that is different from humans, plants, pets, inanimate objects, and so on, and what factors contribute to this perception;
- What kind of social behaviors would be acceptable to different communities.

We believe our daily experiment is the first step to reveal these fundamental issues.

In this chapter, we proposed a basic framework for analyzing the daily interactions between a human and a robot. We first discussed the daily interactions between a human and a robot. Human-robot interaction falls into four modes according to who initiates the interaction and whether he or she expects a reaction from the other. Then, we subdivided one mode of that interactions, which a robot initiates and expects a reaction from the human, into four types according to the sensor information the robot uses. We focused on the interactions which require voice recognition but for which misrecognitions do not lead to severe discomfort.

Next, we explained what is necessary for a communication robot and introduced the hardware and software architectures of Robovie-IV. Robovie-IV has been developed as a self-contained robot that fulfills the requirements of a communication robot. That is, it is as tall as a human child, its entire body is covered by a soft-skin containing 56 embedded tactile sensors, it moves by an inverted pendulum mechanism that causes involuntary motions, and it has sensors for human recognition. Finally, we briefly described the current status of our ongoing daily experiment in our office and discussed behaviors of the robot and how humans relate to it.

9. Acknowledgement

10. References

Bruno, H. (2002). Repp.: phase correction in sensorimotor synchronization: nonlinearities in voluntary and involuntary responses to perturbations. *Human movement science*, Vol. 211, No. 1, (2002) (1-37)

Burgard, W.; Cremers, A. B., Fox, D. Haehnel, D. Lakemeyer, G. Schulz, D. Steiner, W., & Thrun, S. (1998). The interactive museum tour-guide robot, In *Proceedigns of the Fifteenth National Conference on Artificial Intelligence*, 1998

Campbell, W.N. & Black, A. (1996). Prosody and the selection of source units for concatenative synthesis, In: Progress in Speech Synthesis, van Santen, J. et al. (Ed.), Springer

Ha, Y. & Yuta, S. (1994). Trajectory tracking control for navigation of self-contained mobile inverse pendulum, In *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems 1994*, pp. 1875-1882, 1994

Isbell, C.; Shelton, C. R. Kearns, M. Singh, S. & Stone, P. (2001) A social reinforcement learning agent, In *Proceedings of the Fifth International Conference on Autonomous Agents*, pp. 377−384, Montreal, Canada, 2001, ACM Press

Ishiguro, H.; Ono, T. Imai, M. Maeda, T. Kanda, T. & Nakatsu R. (2001). Robovie: A robot generates episode chains in our daily life. In *Proceedings of the 32nd International Symposium on Robotics*, pp. 1356−1361, 2001

Kubota, H.; Nishida, T. & Koda, T. (2000). Exchanging tacit community knowledge by talking-virtualized-egos. In *Fourth International Conference on AUTONOMOUS AGENTS*, pp. 285−292, 2000

Lee, A.; Kawahara, T. & Shikano, K. (2001). Julius --- an open source real-time large vocabulary recognition engine, In *Proceedings of the European Conference on Speech Communication and Technology*, pp. 1691-1694, 2001

Matsui, T.; Asoh, H. Hara, I. & Otsu, N. (1997). An event-driven architecture for controlling behaviours of the office conversant mobile robot, Jijo-2, In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, pp. 3367−3372, April 1997

Miyashita, T. & Ishiguro, H. (2004). Human-like natural behaviour generation based on involuntary motions for humanoid robots, *International Journal of Robotics and Autonomous Systems*, Vol. 48, (2004) (203-212)

Miyashita, T.; Tajika, T. Ishiguro, H. Kogure, K. & Hagita, N. (2005). Haptic communication between humans and robots, In *Proceedings of the 12th International Symposium of Robotics Research*, 2005

Wu, C. (1991). Analysis of voluntary movements for robotic control, In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 326−331, 1991

Zheng, Y. (1990). A neural gait synthesizer for autonomous biped robots, In *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems*, pp. 601−608, 1990
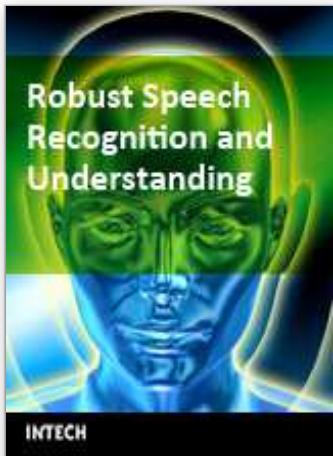
Robust Speech Recognition and Understanding

Edited by Michael Grimm and Kristian Kroschel

This book on Robust Speech Recognition and Understanding brings together many different aspects of the current research on automatic speech recognition and language understanding. The first four chapters address the task of voice activity detection which is considered an important issue for all speech recognition systems. The next chapters give several extensions to state-of-the-art HMM methods. Furthermore, a number of chapters particularly address the task of robust ASR under noisy conditions. Two chapters on the automatic recognition of a speaker's emotional state highlight the importance of natural speech understanding and interpretation in voice-driven systems. The last chapters of the book address the application of conversational systems on robots, as well as the autonomous acquisition of vocalization skills.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Noriaki Mitsunaga, Zenta Miyashita, Takahiro Miyashita, Hiroshi Ishiguro and Norihiro Hagita (2007). Conversation System of an Everyday Robot Robovie-IV, Robust Speech Recognition and Understanding, Michael Grimm and Kristian Kroschel (Ed.), ISBN: 978-3-902613-08-0, InTech, Available from: http://www.intechopen.com/books/robust_speech_recognition_and_understanding/conversation_system_of_an_everyday_robot_robovie-iv

INTECH
open science | open minds