(19) Japan Patent Office (JP)

(12) Japanese Unexamined Patent Application Publication (A)

JP 2005-202076 A 2005.7.28
(11) Japanese Unexamined Patent Application Publication Number

2005-202076
(P2005-202076A)
(43) Publication date October 28, 2005 (7.28.2005)

(51) Int.C1. [7]		F1			Theme codes (reference)
G1OL	**13/08**	G1OL	3/00	H	2C150
A63H	**11/00**	A63H	11/00	Z	3C007
B25J	**13/00**	B25J	13/00	Z	5D045
G1OL	**13/00**	G1OL	3/00	Q	

Request for examination Not yet requested Number of Claims 24 OL (Total of 19 pages)

(21) Application number: Japanese Patent Application 2004-7306 (P2004-7306)
(22) Date of application: January 14, 2004 (1.14.2004)
(Japan Patent Office Note: the following is a registered trademark)
1. Bluetooth

(71) Applicant: 000002185
Sony Corporation
6-7-35 Kita Shinagawa, Shinagawa-ku, Tokyo-to
(74) Agent: 100082740
Patent attorney Keiki Tanabe
(72) Inventor: Hideki Shimomura
c/o Sony Corporation, 6-7-35 Kita Shinagawa, Shinagawa-ku, Tokyo-to

F terms (reference)
2C150 CA01 DA04 DF02 DF21 EB01 EE02 EF11 EF16
3C007 AS36 BS27 CS08 JS03 KS03 KS11 KS31 KS36 KS39 KX02 MT14 WA03 WA13 WB18 WC07 WC30
5D045 AA07 AA08 AB11

[54] [Title of the Invention] Speech control device and method and robot device

[57] [Abstract]
[Problem]
Depending on the distance from the device or robot device to the conversation partner, the conversation partner may have difficulty hearing the speech of the device or robot device.
[Means for Solving]
By changing the speech style of the device or robot device when interacting with the user as necessary depending on the distance between the device or robot device and the user, the device or robot device can always speak in a style that is easy for the user to hear, allowing for smooth interaction with the user, thus realizing a speech control device and method, and a robot device that can improve entertainment value.
[Representative Drawing] Fig. 13

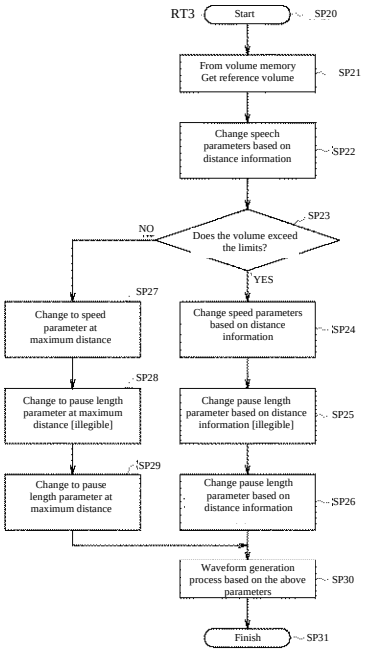

FIG. 13 Speech change processing procedure

Amazon v. SoundClear
US Patent 11,069,337
Amazon Ex. 1004

[Scope of Patent Claims]

[Claim 1]

A speech control device that controls the speech of a device having a dialogue function with a user during dialogue between the device and the user, characterized by comprising;

a distance detection means for detecting the distance between the device and the user who is the conversation partner; and

a speech form change means for changing the speech form of the device during conversation with the user as necessary, according to the distance between the device and the user detected by the distance detection means.

[Claim 2]

The speech control device according to Claim 1,

characterized in that the speech form change means

changes the speech volume as the speech form.

[Claim 3]

The speech control device according to Claim 1,

characterized in that the speech form change means

changes the speech speed as the speech form.

[Claim 4]

The speech control device according to Claim 1,

characterized in that the speech form change means

changes the speech intonation as the speech form.

[Claim 5]

The speech control device according to Claim 1,

characterized in that the speech form change means

changes the space between phrases as the speech form.

[Claim 6]

The speech control device according to Claim 5, characterized by comprising:

a character string output means for outputting a character string corresponding to a content to be spoken by the device;

a speech means for generating a speech signal of synthetic voice corresponding to the character string;

and a speaker for outputting audio based on the audio signal,

wherein the speech form changing means transforms the character string

output from the character string output means

as a method of changing the space between the phrases.

[Claim 7]

The speech control device according to Claim 2, further comprising:

a user identification means for identifying the user who is the conversation partner;

a memory means for storing a reference volume for each of the users;

and a reference volume change means for changing the reference volume of the user stored in the memory means in response to a request from the user during conversation,

wherein the speech style change means changes the speech volume

based on the reference volume of the user who is the conversation partner

stored in the reference volume change means.

[Claim 8]

The speech control device according to Claim 2,

wherein the speech style change means changes another speech style by a maximum amount when the changed speech volume exceeds a predetermined threshold value.

[Claim 9]

A speech control method for controlling speech of a device having a function for interacting with a user during a conversation between the device and the user, comprising:

a first step of detecting a distance between the device and the user who is a conversation partner;

and a second step of changing the speech form of the device during a conversation with the user as necessary, depending on the detected distance between the device and the user.

A speech control method comprising:

[Claim 10]
The speech control method according to Claim 9,
characterized in that the second step
changes the speech volume as the speech form.
[Claim 11]
The speech control method according to Claim 9,
characterized in that the second step
changes the speech speed as the speech form.
[Claim 12]
The speech control method according to Claim 9,
characterized in that the second step
changes the speech intonation as the speech form.
[Claim 13]
The speech control method according to Claim 9,
characterized in that the second step
changes the space between phrases as the speech form.
[Claim 14]
The speech control method according to Claim 13,
characterized in that the second step includes
a character string output step of outputting a character string corresponding to a content to be spoken by the device,
a speech generation step of generating a speech signal of synthetic voice corresponding to the character string,
and a speech output step of outputting speech based on the speech signal,
wherein in the speech generation step, the output character string is transformed
as a method of changing the intonation of the speech.
[Claim 15]
The speech control method according to Claim 10,
further comprising: a storing step of storing a reference volume for each user;
a user identifying step of identifying the user who is a conversation partner;
and a reference volume changing step of changing the stored reference volume of the user
in response to a request from the user during the conversation,
wherein in the second step, the speech volume is changed
based on the stored reference volume of the user who is a conversation partner.
[Claim 16]
The speech control method according to Claim 10,
wherein the second step changes another speech style by a maximum amount when the changed speech volume exceeds a predetermined threshold value.
[Claim 17]
A robot device having a function of interacting with a user, comprising:
a distance detection means for detecting a distance to the user who is a conversation partner;
and a speech form change means for changing an utterance form during a conversation with the user as necessary,
according to the distance to the user detected by the distance detection means.

[Claim 18]

The robot device according to Claim 17,

characterized in that the speech form change means

changes the speech volume as the speech form.

[Claim 19]

The robot device according to Claim 17,

characterized in that the speech form change means

changes the speech speed as the speech form.

[Claim 20]

The robot device according to Claim 17,

characterized in that the speech form change means

changes the speech intonation as the speech form.

[Claim 21]

The robot device according to Claim 20, characterized by comprising:

a character string output means for outputting a character string corresponding to a content to be spoken by the device;

a speech means for generating a speech signal of synthetic voice corresponding to the character string;

and a speaker for outputting audio based on the audio signal,

wherein the speech form changing means transforms the character string

output from the character string output means

as a method of changing the speech intonation.

[Claim 22]

The robot device according to Claim 17,

characterized in that the speech form change means

changes the space between phrases as the speech form.

[Claim 23]

The robot device according to Claim 18, further comprising:

a user identification means for identifying the user who is the conversation partner;

a memory means for storing a reference volume for each of the users;

and a reference volume change means for changing the reference volume of the user stored in the memory means in response to a request from the user during conversation,

wherein the speech style change means changes the speech volume

based on the reference volume of the user who is the conversation partner

stored in the reference volume change means.

[Claim 24]

The robot device according to Claim 18,

wherein the speech style change means changes another speech style by a maximum amount when the changed speech volume exceeds a predetermined threshold value.

[Detailed Description of the Present Invention]

[Technical Field]

[0001]

The present invention relates to a speech control device and method, and a robot device, which is suitable for application to, for example, an entertainment robot.

[Background art]

[0002]

In recent years, many companies and research institutions such as universities have been developing humanoid robots. Such humanoid robots are equipped with external sensors such as Charge Coupled Device (CCD) cameras, microphones, and touch sensors, and internal sensors such as battery sensors and acceleration sensors, and are designed to recognize the external and internal conditions based on the output of these external sensors and internal sensors, and to act autonomously based on the recognition results (see, for example, Non-Patent Literature 1).

[0003]

In recent years, many entertainment robots have also been installed with speech recognition and interaction control functions to enable simple daily conversations with users.

[Non-Patent Literature 1] Japanese Patent Application 2003-270835

[Disclosure of the Invention]

[Problem to be Solved by the Invention]

[0004]

However, conventional robots equipped with such voice recognition and dialogue control functions are constructed to always converse with the user at a constant speech volume set in advance, regardless of the physical distance between the robot and the user.

[0005]

For this reason, depending on the setting of the speech volume, even if it is appropriate for a user close to the robot, the volume may be too low for a user a little distance away from the robot to hear the speech content, or conversely, even if it is appropriate for a user a little distance away from the robot, the volume may be too high for a user close to the robot to hear the speech content.

[0006]

One method for solving such problems is to make it possible to freely change the speech volume of the entertainment robot by operating a switch. However, this method has the problem of impairing the naturalness of the interaction between the user and the robot, and furthermore, having to set the speech volume each time is a problem of extreme inconvenience for the user.

[0007]

Also, when the surrounding environment, such as a room with a lot of reverberation, is taken into consideration, simply increasing the robot's speech volume does not necessarily make it easier for users who are far from the robot to hear the robot's speech.

[0008]

Such a situation in which the content of a robot's speech is difficult for the user to hear is a factor that hinders smooth and natural dialogue between the user and the robot, and impairs the entertainment value of the robot having a dialogue control function, and some kind of solution is desired.

[0009]

The present invention has been made in consideration of the above points, and aims to propose a speech control device and method, and a robot device, that can improve the entertainment value of a robot having a dialogue control function.

[Means for Solving the Problem]

[0010]

In order to solve such a problem, in the speech control device that controls the speech of the device during a dialogue between the device and a user, a speech form changing means is provided that changes the speech form of the device during a dialogue with the user as necessary according to the distance between the device and the user.

[0011]

As a result, the speech control device can always speak in a speech form that is easy for the user to hear, and can have a smooth dialogue with the user.

[0012]

Also, in the present invention, in the speech control method that controls the speech of the device during a dialogue between the device and a user, the speech form of the device during a dialogue with the user is changed as necessary according to the distance between the device and the user.

[0013]

As a result, according to this speech control method, the robot device can always speak in a speech form that is easy for the user to hear, and therefore a smooth dialogue with the user can be carried out.

[0014]

Furthermore, in the present invention, the robot device is provided with a speech form changing means for changing the speech form when interacting with the user as necessary, depending on the distance to the user.

[0015]

As a result, the robot device can always speak in a speech form that is easy for the user to hear, and can thus have a smooth dialogue with the user.

[Effect of the Invention]

[0016]

According to the present invention, in a speech control device and method for controlling the speech of a device during a dialogue between the device and a user, the speech form of the device during dialogue with the user is changed as necessary according to the distance between the device and the user, so that the device can always speak in a speech form that is easy for the user to hear, allowing for smooth dialogue with the user, and thus improving entertainment value.

[0017]

In addition, according to the present invention, a robot device is provided with a speech style change means for changing the speech style when interacting with a user as necessary depending on the distance to the user, so that the robot device can always speak in a style that is easy for the user to hear, allowing for smooth interaction with the user and thus realizing a robot device that can improve entertainment value.

[Best Embodiments of the Invention]

[0018]

One embodiment of the present invention will be described in detail below with reference to the drawings.

[0019]

(1) Constitution of the robot according to this embodiment

(1-1) Hardware constitution of robot 1

In FIG. 1, 1 indicates the robot according to this embodiment as a whole, which is configured by attaching a head unit 4 to the upper part of a body unit 2 via a neck joint 3, attaching arm units 5A and 5B to the upper left and right parts of the body unit 2 via shoulder joints 4A and 4B, respectively, and attaching a pair of leg units 7A and 7B to the lower part of the body unit 2 via hip joints 6A and 6B, respectively.

[0020]

FIG. 2 schematically illustrates a functional constitution of this robot 1. As shown in FIG. 2, the robot 1 is configured with a control unit 10 for controlling overall operation and other data processing, an input/output unit 11, a drive unit 12, and a power supply unit 13.

[0021]

The input/output unit 11 includes, as input units, a pair of CCD (Charge Coupled Device) cameras 20 corresponding to the robot 1's eyes, a pair of microphones 21 corresponding to the ears, touch sensors 22 arranged on the head, hands, soles of the feet, etc., to detect physical actions from the user, contact between the hands and external objects, and the soles of the feet touching the ground, as well as various other sensors corresponding to the five senses.

[0022]

The input/output unit 11 is also equipped with a speaker 23 equivalent to the mouth of the robot 1, and LEDs (eye lamps) 24 that form facial expressions by combining flashing lights or timing the lights on and off, as output units. These output units can express user feedback from the robot 1 in forms other than mechanical motion patterns using legs, such as voice and flashing lights.

[0023]

The drive unit 12 is a functional block that realizes the body movement of the robot 1 according to a predetermined motion pattern commanded by the control unit 10, and is the object to be controlled by behavior control. The drive unit 12 is a functional module for realizing the degree of freedom at each joint of the robot 1, and is composed of multiple drive units 25_1 to 25_n provided for each axis, such as roll, pitch, and yaw, at each joint. Each drive unit 25_1 to 25_n consists of a combination of motors 26_1 to 26_n performing rotational motion around a given axis, encoders 27_1 to 27_n detecting rotational positions of motors 26_1 to 26 n, and drivers 28 1 to 28 n that adaptively control rotational positions and rotational speeds of motors 26_1 to 26_n based on the output of encoders 27_1 to 27_n.

[0024]

The power supply unit 13 is literally a functional module that supplies power to each electric circuit in the robot 1. The robot 1 according to this embodiment is an autonomous type that uses a battery, and the power supply unit 13 is composed of a charging battery 29 and a charge/discharge control unit 31 that manages the charge/discharge state of the rechargeable battery 30.

[0025]

The charging battery 29 is composed, for example, in the form of a "battery pack" in which multiple lithium-ion secondary battery cells are packaged in a cartridge format.

[0026]

In addition, the charge/discharge control unit 30 also measures the terminal voltage, charge/discharge current, and ambient temperature of the battery 29 to grasp the remaining capacity of the battery 29 and determine the start and end times of charging. The start and end time of the charge determined by the charge/discharge control unit 30 is notified to the control unit 10 and becomes a trigger for the robot 1 to start and end the charge operation.

[0027]

The control unit 10 corresponds to the “brain” of the robot 1 and is mounted, for example, in the head unit 4 or the body unit 2. As shown in FIG. 3, the control unit 10 is configured with a CPU (Central Processing Unit) 31 as the main controller, which is connected to memory, other circuit components, and peripheral devices via a bus. Bus 37 is a common signal transmission path including a data bus, an address bus, a control bus, and the like. Bus 37 is a common signal transmission path including a data bus, an address bus, a control bus, and the like. CPU 31 is assigned a unique address (memory address or I/O address) to each device on the bus 37.

[0028]

RAM (Read Access Memory) 32 is a writable memory composed of volatile memory such as DRAM (Dynamic RAM), and is used to load program code executed by the CPU 31 and to temporarily store work data by the executed program.

[0029]

Read Only Memory (ROM) 33 is a read-only memory that permanently stores programs and data. The program code stored in the ROM 33 includes a self-diagnostic test program that is executed when the robot 1 is powered on, and a control program that specifies the operation of the robot 1.

[0030]

The control program for the robot 1 includes a "sensor input/recognition processing program" that processes input from various sensors such as the CCD camera 20 and microphone 21 and recognizes them as symbols, a "behavior control program" that controls the behavior of the robot 1 based on the sensor input and a predetermined behavior control model while managing memory operations such as short-term memory, and a "drive control program" that controls the drive of each joint motor and the voice output of the speaker 22 according to the behavior control model.

[0031]

The non-volatile memory 34 is composed of a memory element that can be electrically erased and rewritten, such as an EEPROM (Electrically Erasable and Programmable ROM), and is used to store data that needs to be updated in a non-volatile manner. Examples of data that needs to be updated include secret keys, other security information, and device control programs that need to be installed after shipment.

[0032]

Interface 35 is an apparatus for interconnecting with equipment outside the control unit 10 and enabling data exchange. The interface 35, for example, inputs and outputs data between the CCD camera 20, microphone 21, and speaker 22 in the input/output unit 11. The interface 35 also inputs and outputs data and commands between the drivers 28_1 to 28_n in the drive unit 12.

[0033]

Also, the interface 35 may be provided with a general-purpose interface for connecting peripheral devices of a computer, such as a serial interface such as RS (Recommended Standard)-232C, a parallel interface such as IEEE (Institute of Electrical and Electronics Engineers) 1284, a USB (Universal Serial Bus) interface, an i-Link (IEEE 1394) interface, a SCSI (Small Computer System Interface) interface, a memory card interface (card slot) that accepts a PC card or a memory stick, etc., so that programs and data can be transferred between the computer and a locally connected external device. In addition, as another example of the interface 35, an infrared communication (IrDA) interface may be provided to perform wireless communication with external devices.

[0034]

Furthermore, the control unit 10 includes a wireless communication interface 36 and a network interface card (NIC) 38, and can communicate data with various external host computers via short-distance wireless data communication such as Bluetooth, a wireless network such as IEEE802. 11b, or a high-bandwidth network such as the Internet.

[0035]

By using such data communication between the robot 1 and the host computer, it is possible to use remote computer resources to calculate the complex motion control of the robot 1 and to remotely control it.

[0036]

(1-2) Software constitution of robot 1

FIG. 4 schematically illustrates a functional constitution of a behavior control system 40 of a robot 1 configured by a group of control programs stored in a ROM.

[0037]

This behavior control system 40 is implemented incorporating object-oriented programming. In this case, each software is handled in a module called an "object" that integrates the data with the processing procedures for the data. In addition, each object can pass and invoke data by means of an object-to-object communication method using message communication and shared memory.

[0038]

Here, the behavior control system 40 includes an image recognition unit 41, a speech recognition unit 42, and a contact recognition unit 43 for recognizing an external environment based on sensor output from various sensors such as a CCD camera 20 (FIG. 2), a microphone 21 (FIG. 2), and a touch sensor 22 (FIG. 2).

[0039]

The image recognition unit 41 performs image recognition processing and feature extraction, such as facial recognition and color recognition, based on the image signal S1 provided by the CCD camera 20. The image recognition unit 41 then sends out to the short-term memory unit 44 the various image recognition results, such as facial recognition information such as the person’s unique facial ID (identifier), the position and size of the facial image region, and the color recognition information such as the position, size, and feature of the color region such as the color recognition result, and the image signal S1. The image recognition unit 41 also detects the distance to the imaging object by means of a so-called stereo visual method based on the image signal S1 from the CCD camera camera 20, and sends the detection results to the short-term memory unit 44.

[0040]

The speech recognition unit 42 performs various sound-related recognition processes such as speech recognition, speaker recognition, and sound source direction recognition based on the speech signal S2 provided by the microphone 21. The speech recognition unit 42 then sends various speech recognition results, such as character string information of the recognized words, which is the speech recognition result, speaker ID information unique to the speaker, which is the speaker recognition processing result based on acoustic features, etc., and sound source direction information, which is the sound source direction recognition result, to the short-term memory unit 44. In addition, the speech recognition unit 42, along with these various speech recognition results, sends these speech signals S2 to the short-term memory unit 44.

[0041]

Furthermore, the contact recognition unit 43 recognizes physical contact with the outside, such as "being stroked," "being hit," "grasping an object," and "the sole of the foot touching the ground," based on the pressure detection signal S3 provided by the touch sensors 22 arranged on the upper part of the head unit 4 (FIG. 1), the hands that are the tips of the arm units 5A and 5B (FIG. 1), and the soles that are the bottoms of the leg units 7A and 7B (FIG. 1), and sends these contact recognition results to the short-term memory unit 44. The contact recognition unit 44 also sends a pressure detection signal S3 from each touch sensor 22 to the short-term memory unit 44 in conjunction with these contact recognition results.

[0042]

The short-term memory unit 44 is an object that holds information about the external environment of the robot 1 for a relatively short period of time, and receives and stores for a short period of time various image recognition results and image signals S1 provided by the image recognition unit 41, various speech recognition results and speech signals S2 provided by the speech recognition unit 42, and various contact recognition results and pressure detection signals S3 provided by the contact recognition unit 43.

[0043]

In addition, the short-term memory unit 44 uses the image recognition results received, the speech recognition results and the contact recognition result, and the image signal S1, the speech signal S2, and each pressure detection signal S3 to coordinate the facial image area, person ID, speaker ID, and character string information, etc., to generate target information and event information about where the person is currently, what person the words are, and what kind of interaction has been made with that person, and sends this information to the behavior selection control unit 45.

[0044]

Based on the target information and event information provided by the short-term memory unit 44 and the contents stored in the short-term memory unit 44, the behavior selection control unit 45 determines the next behavior of the robot 1 from among a plurality of previously prepared behaviors, such as a behavior selected depending on the current situation and previous behavior of the robot 1 (situation-dependent behavior), a reflexive behavior in response to an external stimulus (reflexive behavior), or an action based on a relatively long-term action plan in response to a given situation or a command from the user (deliberative behavior). The behavior selection control unit 45 then notifies the output management unit 46 of the behavior determined in this way.

[0045]

In response to the notification from the behavior selection control unit 45, the output management unit 46 performs arbitration processing when multiple behaviors such as situation-dependent behaviors and reflex behaviors compete, and processing to synchronize the movement, speech, and blinking of the LED 24, while driving the motors 26_1 to 26_n of the corresponding drive units 25_1 to 25_n and driving the LED 24 to blink in a predetermined pattern.

[0046]

In addition, if the behavior selection control unit 45 decides to interact with the user as the next action, the speech recognition results of the speech of the user stored sequentially in the short-term memory unit 44 are continuously monitored by the speech recognition unit 42 after this, and the content to be spoken by the robot 1 is sequentially determined based on the speech recognition results. Then, the behavior selection control unit 45 reads the necessary character string from the speech string database 47 stored in advance in the ROM 33 (Fig. 3) based on this determination result, and sends it to the output management unit 46.

[0047]

At this time, the output management unit 46 sends the character string provided by the behavior selection control unit 45 to the speech synthesis unit 48, while the speech synthesis unit 48 generates a speech signal S4 of synthetic speech based on the supplied character string and sends it to the speaker 23 (FIG. 2). As a result, a speech based on this speech signal S4 is output from the speaker 23.

[0048]

In this way, the robot 1 can act autonomously based on external conditions, etc. recognized based on sensor output of various sensors such as CCD camera 20, microphone 21, and touch sensor 23.

[0049]

(2) Speech control function in robot 1

Next, the speech control function installed in this robot 1 will be described.

[0050]

The robot 1 is equipped with a speech control function that controls speech forms such as speech volume, speech speed, intonation, and intervals between phrases according to the physical distance to the conversation partner. In addition, the robot 1 is equipped with a speech control function that changes the speech volume to be used as a standard (hereinafter referred to as the reference volume) for each user in response to a request from the user, taking into consideration the large individual differences in the sense of speech forms, especially speech volume.

[0051]

In practice, in the case of the robot 1, as a means for performing speech control that changes the reference volume for each user, the behavior control system 40 is provided with a reference volume memory unit 49 for storing and holding the reference volume for each user. Incidentally, this reference volume memory unit 49 is configured from the non-volatile memory 34 (FIG. 3).

[0052]

Then, each time the short-term memory unit 44 detects a new user based on the image recognition result of the image recognition processing unit 41, the behavior selection control unit 45 associates the person ID of the user with the reference volume for that user (the initial setting value is "3") as shown in FIG. 5, and stores the person ID and the reference volume in the reference volume memory unit 49. Meanwhile, during subsequent conversations with the user, the reference volume for the user stored and held in the reference volume memory unit 49 is changed as necessary according to the reference volume change processing procedure RT1 shown in FIG. 6.

[0053]

That is, when the behavior selection control unit 45 starts a dialogue with the user, it starts this reference volume change processing procedure RT1 in step SP0, and in the following step SP1, it waits for a word that requests to increase the volume of the robot's speech, such as "speak a little louder" or "I can't hear you," or a word that requests to decrease the volume of the robot's speech, such as "speak a little quieter" or "it's too noisy."

[0054]

Then, when the behavior selection control unit 45 recognizes that a word that requests to increase or decrease the volume of the robot's speech has been uttered by the user based on the speech recognition result of the speech recognition unit 42 stored and held in the short-term memory unit 44, it proceeds to step SP2, and judges whether or not the word has been specified as coming from the current conversation partner, based on the target information and event information given from the short-term memory unit 44.

[0055]

If the behavior selection control unit 45 obtains a negative result in step SP2, it returns to step SP1. If the behavior selection control unit 45 obtains a positive result, it proceeds to step SP3, where it increases the reference volume of the user stored in the reference volume memory unit 49 by a specified value (e.g., "1") if the words uttered by the user are words that request an increase in the speech volume, and decreases the reference volume by the specified value if the words uttered by the user are words that request a decrease in the speech volume. It should be noted that the volume value represents the volume level of the speech volume of the robot 1, and the larger the value is, the higher the predetermined volume level is associated with each other.

[0056]

The behavior selection control unit 45 then returns to step SP1 and repeats the same process for step SP1 through step SP3. In this way, the behavior selection control unit 45 changes the reference volume for each user in response to a request from the user to change the speech volume.

[0057]

On the other hand, in this robot 1, as a method for changing the interval between phrases so that the division of meaning can be clearly understood according to the physical distance to the conversation partner, speech control is performed to insert "ne" after each phrase as necessary, and as a means for this, the speech character string transformation unit 51 is provided in the behavior control system 45.

[0058]

In this case, the speech character string transformation unit 51 stores and holds in the ROM 33 (Fig. 3) a character string transformation determination table 50 that specifies whether or not to transform the character string to insert "ne" after each phrase according to the distance to the conversation partner, as shown in FIG. 7.

[0059]

The behavior selection control unit 45 reads out character strings corresponding to the contents (words) to be spoken by the robot 1 when conversing with the user from the speech string database 47, and sequentially sends these as character string information to the speech character string transformation unit 51. At the same time, the behavior selection control unit 45 sends to the speech character string transformation unit 51 distance information indicating the distance to the conversation partner detected by the image recognition unit 41 and stored and held in the short-term memory unit 44.

[0060]

When the speech character string transformation unit transformation unit 51 is given character string and distance information from the behavior selection control unit 45, it judges whether or not to transform the character string based on the character string information according to the character string transformation processing procedure RT2 shown in FIG. 8 based on the distance information and the character string transformation determination table 50, and transforms the character string if necessary.

[0061]

That is, when the speech character string transformation unit 49 is given character string information and distance information from the behavior selection control unit 45, it starts this character string transformation processing procedure RT2 in step SP10, and in the following step SP11, it judges whether or not to transform the character string based on this character string information based on the distance to the conversation partner recognized based on the distance information from the action selection control unit 45 and the character string transformation determination table 50 (FIG. 7).

[0062]

Specifically, the speech character string transformation unit 49 judges that the character string should not be transformed if the distance to the conversation partner recognized by the image recognition unit 41 is less than 350 [cm] according to the character string transformation determination table 50, and judges that the character string should be transformed if the distance to the conversation partner is 350 [cm] or more.

[0063]

If the speech character string transformation unit 49 obtains a negative result in step SP11 as the judgment result, it proceeds to step SP14 without performing transformation processing on the character string, and ends the character string transformation processing procedure RT2. After that, it sends the character string information of the character string as it is to the speech synthesis unit 48 via the behavior selection control unit 45 and the output management unit 46 in sequence.

[0064]

On the other hand, if the speech character string transformation unit 49 obtains a positive result in step SP11, it proceeds to step SP12, where it performs a morphological analysis process on the character string to detect each clause-final auxiliary verb contained in the character string.

[0065]

The speech character string transformation unit 49 then proceeds to step SP13, where it sequentially adds the character "ne" to the end of each clause-final auxiliary verb detected in step SP12. Therefore, when the speech character

string transformation unit 49 is given the character string "Today's weather in Tokyo is sunny" from the behavior selection control unit 45, it will generate a character string such as "Today's, Tokyo's, the weather is, sunny" by adding the character "ne" to the end of each clause-final auxiliary verb, "Today's," "Tokyo's," and "The weather is."

[0066]

The speech character string transformation unit 49 then proceeds to step SP14 to end the character string transformation processing procedure RT2, and sends character string information representing the thus obtained new transformed character string to the speech synthesis unit 48 via the behavior selection control unit 45 and the output management unit 46 in sequence.

[0067]

At this time, the speech synthesis unit 48 stores in the ROM 33 (Fig. 3) a volume change amount specification table 52 that specifies the amount of change in the speech volume of the robot 1 with respect to the distance to the conversation partner, as shown in Fig. 9, a speed change amount specification table 53 that specifies the amount of change in the speech speed of the robot 1 with respect to the distance, as shown in Fig. 10, a pause length change amount specification table 54 that specifies the amount of change in the interval between phrases (pause length) with respect to the distance, as shown in Fig. 11, and a phrase-end emphasis determination table 55 that specifies whether or not to change the intonation of the speech to emphasize the end of a phrase with respect to the distance, as shown in Fig. 12. The speech synthesis unit 48 is also provided with the person ID of the conversation partner at that time as person ID information from the short-term memory unit 44, and the distance to the conversation partner is sequentially provided as distance information.

[0068]

Then, when the Speech synthesis unit synthesis unit 48 is provided with character string information of the content to be spoken by the robot 1 via the output management unit 46, the voice synthesis unit 48 changes the voice volume and speech speed of the robot 1 as necessary according to the speech change processing procedure RT3 shown in FIG. 13 based on the person ID information and distance information provided by the short-term memory unit 44, each table 52 to 55 stored in the ROM 33, and the reference voice volume of the user stored and held in the reference volume memory unit 49.

[0069]

That is, when the speech synthesis unit 48 is given such character string information, it starts this character string transformation processing procedure RT3 in step SP20. In the following step SP21, based on the person ID information given from the short-term memory unit 44, it reads out the reference volume of the corresponding conversation partner from the reference volume memory unit 49. Then, it proceeds to step SP22, and based on the distance to the conversation partner recognized from the distance information, it refers to the volume change amount specification table 52 (FIG. 9) and changes the parameter value of the volume parameter defining the speech volume of the robot 1 as necessary.

[0070]

Specifically, when the distance to the conversation partner given as distance information is, for example, less than 50 cm, the speech synthesis unit 48 determines the reference volume read from the reference volume memory unit 49 as the current speech volume, and when the distance to the conversation partner is in the range of 50 to 80 cm, determines a volume that is a predetermined specified value (for example, "1") higher than the reference volume as the current speech volume, and similarly thereafter, when the distance to the user is 80 to 150 cm, 150 to 250 cm, or 250 cm or more, determines a volume that is "2", "3", or "5" higher than the reference volume as the current speech volume. The speech synthesis unit 48 then changes the parameter values of the volume parameters as necessary according to the result of this determination.

[0071]

Next, the speech synthesis unit 48 proceeds to step SP23 to determine whether the volume determined as described above in step SP22 exceeds a predetermined threshold.

[0072]

That is, there is a certain limit to the speech volume of the robot 1, so for example, if the reference volume is already set to a limit value or a value close to this, it may not be possible even if the speech volume should be determined to be a larger volume in accordance with the distance with the user when determining the volume in step SP22. In such cases, the content of the utterance of robot 1 may be difficult to hear for such users.

[0073]

The speech synthesis unit 48, therefore, performs normal change processing for other speech forms, such as speech speed other than speech volume, in step SP24 to step SP26, if a positive result is obtained in step SP23, and in step SP27 to step SP29, if a negative result is obtained in step SP23, special change processing is performed for such other speech forms.

[0074]

In effect, upon obtaining a positive result in step SP23, the speech synthesis unit 48 proceeds to step SP24 and changes the parameter value of the speed parameter defining the speech speed of the robot 1 as necessary, referring to the speed change amount specification table 53 (FIG. 10) stored in ROM 33, based on the distance to the conversation partner recognized by the distance information provided by the short-term memory unit 44.

[0075]

Specifically, the speech synthesis unit 48 determines a predetermined default value as the speed at the time of speech when the distance to the conversation partner is less than 200 [cm] as distance information, and determines a speed 20 [%] slower than the default value when the distance to the conversation partner is in the range of 200 to 350 [cm], and a speed 50 [%] slower than the default value when the distance to the conversation partner is 350 [cm] or more. The speech synthesis unit 48 then changes the parameter value of the speed parameter as necessary according to the result of this determination.

[0076]

Next, the speech synthesis unit 48 proceeds to step SP25 and, based on the distance from the short-term memory unit 44 to the conversation partner recognized by the distance information, refers to the pause length change amount specification table 54 (FIG. 11) stored in ROM 33 to change the pause length parameter, which defines the length (pause length) between sentences when uttering the character string, as necessary.

[0077]

Specifically, the speech synthesis unit 48 determines a value set as the initial set value of the pause length in advance as the pause length when the distance to the conversation partner provided as distance information is less than 100 [cm], and determines a pause length when the distance to the conversation partner is in the range of 100 to 350 [cm], 30 [%] longer than the default value, and 60 [%] longer than the default value when the distance to the conversation partner is 350 [cm] or more. The speech synthesis unit 48 then changes the parameter values of the pause length parameters as necessary according to the result of this determination.

[0078]

The speech synthesis unit 48 then proceeds to step SP26 and, based on the distance from the short-term memory unit 44 to the conversation partner recognized by the distance information, refers to the phrase-end emphasis determination table 55 (FIG. 12) stored in ROM 33 to change the parameter value of the corresponding parameter, the phrase-end emphasis parameter, as necessary, so that it becomes an intonation that emphasizes each end of sentence when uttering the content of the character string.

[0079]

Specifically, the speech synthesis unit 48 decides not to change the end of each sentence when speaking the character string as audio if the distance to the conversation partner provided as distance information is less than 200 [cm], and decides not to change the end of each sentence if the distance to the conversation partner is 200 [cm] or more. The speech synthesis unit 48 then changes the parameter values of the phrase-end emphasis parameters as necessary according to the result of this determination.

[0080]

In contrast, the speech synthesis unit 48, upon obtaining a negative result in step SP23, proceeds to step SP27, and determines the speech speed of the robot 1 at all times, regardless of the distance to the conversation partner recognized by the distance information provided by the short-term memory unit 44, to the speech speed specified for the maximum distance in the speed change amount specification module 53 (FIG. 10). Thus, in this embodiment, the speed of speech is determined as the speed of speech at which it is 50 [%] slower than the default value. The speech synthesis unit 48 then changes the parameter value of the speed parameter as necessary according to the result of this determination.

[0081]

Next, the speech synthesis unit 48 proceeds to step SP28, where it determines the pause length of the robot 1 when it speaks to be the speech speed that is always specified for the maximum distance in the pause length change amount specification table 54 (FIG. 11). Thus, in this embodiment, a pause length that is 60% longer than the initial setting is determined as the speech speed at that time. The speech synthesis unit 48 then changes the parameter values of the pause length parameters as necessary according to the result of this determination.

[0082]

Next, the speech synthesis unit 48 proceeds to step SP29, where it determines the phrase-end emphasis judgment parameter when speaking the contents of the character string to be always the state specified for the maximum distance in the phrase-end emphasis determination table 55 (FIG. 12). Thus, in this embodiment, it is always determined that the end of each phrase should be emphasized. The speech synthesis unit 48 then changes the parameter value of the phrase-end emphasis parameter according to the result of this determination, as necessary, so as to emphasize the end of the phrase.

[0083]

After the speech synthesis unit 48 has finished setting various parameters such as the volume parameter and the speed parameter in this way, it proceeds to step SP30, generates a speech waveform according to the character string based on the given character string information, and transforms this speech waveform according to the volume parameter, speed parameter, pause length parameter, and phrase-end emphasis parameter set as described above, and sends the speech signal S4 of the speech waveform thus obtained to the speaker 23.

[0084]

In this way, in this robot 1, while controlling the speech type such as speech volume, speech speed, intonation, and the time between speech, depending on the physical distance to the conversation partner, it is possible to always make easy speech to the conversation partner by controlling the speech volume for each user.

[0085]

(3) Operation and effect of the present embodiment

In the above constitution, the robot 1 speaks louder, more slowly, and with larger gaps between phrases the farther away the user/conversation partner is, while depending on the distance to the user, it speaks with an intonation that emphasizes the end of a phrase and inserts the word "ne" after each phrase.

[0086]

Therefore, the robot 1 can of course always speak at a volume that is easy for the user/conversation partner to hear, regardless of the distance to the user, and even if the conversation takes place in a room with a lot of reverberation, the robot can always speak at a volume that is easy for the user to hear. As a result, the robot can converse with the user smoothly and at a natural volume.

[0087]

In this case, the robot 1 changes the reference volume of the speech sound for each user in response to the user's request, so that the robot can converse with each user at an appropriate volume according to the hearing characteristics of the user/conversation partner, thereby enabling smoother conversation with the user without causing discomfort to the user.

[0088]

According to the above constitution, by controlling the speech volume, speech speed, intonation, and speech form such as between speeches according to the physical distance to the conversation partner, the speech form can always be spoken in a speech form that is easy for the user to hear. This makes it possible to realize a robot that can smoothly communicate with the user, thus improving entertainment value.

[0089]

(4) Other embodiments

It should be noted that, in the above embodiment, the present invention has been described as being applied to a humanoid robot 1; however, the present invention is not limited to this, and can be widely applied to various types of robot devices and various types of devices other than robot devices that have interactive functions.

[0090]

In the above-described embodiment, we also discussed cases where the speech volume, speech speaker, intonation, and speech time are changed depending on the distance to the conversation partner, but the present invention is not limited to this, but only some of these speech forms may be changed, or other speech forms may be changed in addition to these speech forms.

[0091]

In this case, in the above embodiment, the robot 1 changes the intonation so as to speak with an intonation that emphasizes the end of each phrase. However, the present invention is not limited to this, and the intonation may be changed so as to emphasize, for example, a particle other than the end of a phrase.

[0092]

Furthermore, in the above-described embodiment, it is stated that the speech volume, speech speed, intonation, and speech time are gradually changed according to the distance to the conversation partner, but the present invention is not limited to this, and it is also acceptable to continuously change these by operation using a function formula, etc., according to the distance to the conversation partner. In this case, for example, for changes in intonation (emphasis at the end of each phrase), it is necessary to continuously change the degree of emphasis at the end of each sentence according to the distance to the conversation partner. In addition, regarding the pause during conversation caused by inserting "ne" after each phrase end, instead of inserting it after every phrase end, the number of phrase ends into which "ne" is inserted may be changed in stages according to the distance to the conversation partner. It goes without saying that words other than "ne" can be used as words to be inserted after each phrase end to create a gap between phrases.

[0093]

Furthermore, in the above embodiment, the distance detection means for detecting the distance to the conversation partner user is configured by a pair of CCD cameras 20 and a short-term memory unit 44 that detects the distance by a stereo vision method based on the output of the cameras, but the present invention is not limited to this, and other means such as a distance sensor can be widely applied.

[0094]

Furthermore, in the above embodiment, the function of the speech form changing means for changing the speech form of the robot 1 when interacting with a user as necessary according to the distance to the user/conversation partner is distributed to the speech character string transformation unit 51 and the speech synthesis unit 48, but the present invention is not limited to this, and all such functions may be installed in the speech synthesis unit 48.

[0095]

Furthermore, in the above embodiment, the function of the user identification means for identifying the conversation partner user is provided in the short-term memory unit 44, but the present invention is not limited to this, and such function may be provided in the behavior selection control unit 45.

[Industrial Applicability]

[0096]

The present invention can be broadly applied to entertainment robots as well as other robotic devices and other devices other than robotic devices for other applications having interactive functions.

[Brief Description of the Drawings]

[0097]

[Figure 1] A perspective view showing the external appearance constitution of the robot according to the present embodiment.

[Figure 2] A block diagram showing the internal constitution of the robot according to the present embodiment.

[Figure 3] A block diagram showing the constitution of the control unit.

[Figure 4] A block diagram showing a specific constitution of a robot's behavior control system in accordance with the present embodiment.

[Figure 5] A conceptual diagram that provides an explanation of the reference volume for each user stored in the reference volume memory section.

[Figure 6] A flowchart showing the procedure for changing the reference volume.

[Figure 7] A conceptual diagram showing a string deformation judgment table.

[Figure 8] A flowchart showing the string deformation processing procedure.

[Figure 9] A conceptual diagram showing the volume change amount table.

[Figure 10] A conceptual diagram showing the speed change amount table.

[Figure 11] A conceptual diagram showing a pause length change amount table.

[Figure 12] A conceptual diagram showing the phrase-end emphasis determination table.

[Figure 13] A flowchart showing the utterance change processing procedure.

[Explanation of References]

[0098]

1...Robot, 40...Behavior control system, 41...Image recognition unit, 42...Speech recognition unit, 44...Short-term memory unit, 45...Behavior selection control unit, 47...Speech string database, 48...Speech synthesis unit, 49...Reference volume memory unit, 50...Character string transformation determination table, 51...Speech character string transformation unit, 52...Volume change table, 53...Speed change table, 54...Pause length change table, 55...Phrase-end emphasis determination table, S1...Image signal, S2, S4...Speech signal, RT1...Reference volume change processing method, RT2...Character string transformation processing procedure, RT3...Speech change processing procedure.

[FIG. 2]

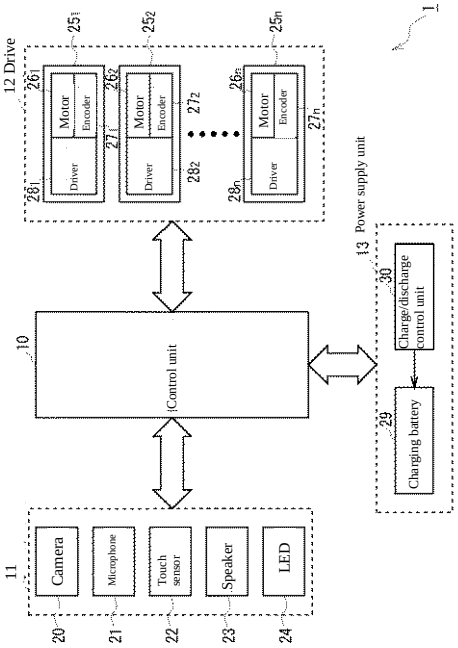

FIG. 2 Hardware constitution of the robot according to this embodiment (2)

[FIG. 3]

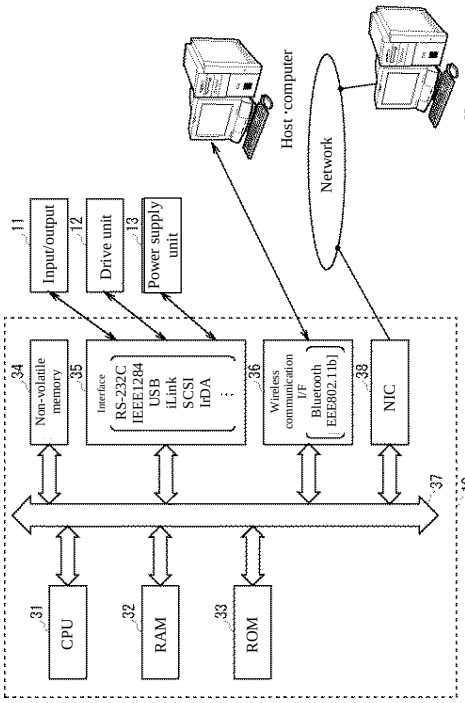

FIG. 3 Constitution of the Control Unit

[FIG. 4]

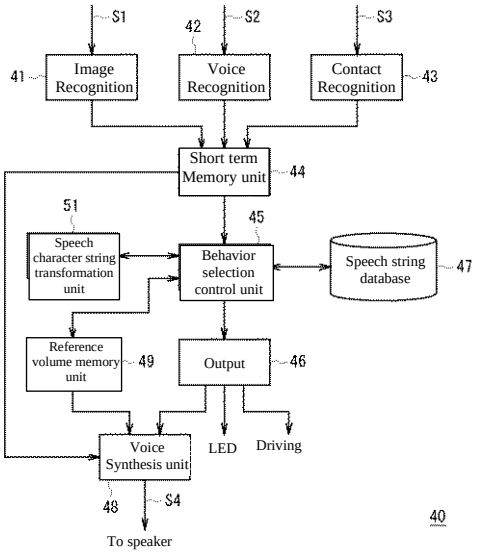

FIG. 4 Constitution of a behavior control system in a robot in accordance with the present embodiment

[FIG. 5]

Person ID	Reference volume
1	3
2	3
3	5

FIG. 5 Reference volume per user

[FIG. 6]

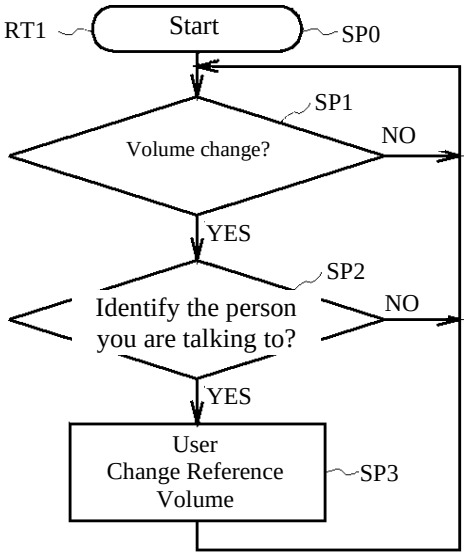

FIG. 6 Reference volume change processing procedure

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

[FIG. 8]

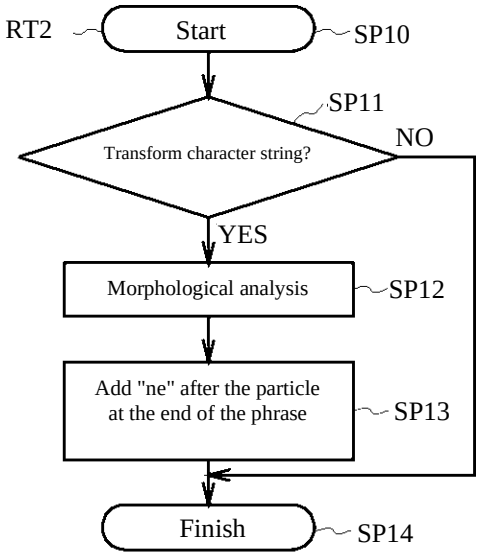

FIG. 8 Character string transformation processing procedure

[FIG. 9]

Less than 50 cm	50~80cm	80~150cm	150~250cm	More than 250
0	+1	+2	+3	+5

52

FIG. 9 Volume Change Table

[FIG. 10]

Less than 200 cm	200~350cm	350 cm or more
0%	+20%	+50%

53

FIG. 10 Speed change table

[FIG. 11]

Less than 100 cm	100~350cm	350 cm or more
0%	+30%	+60%

54

FIG. 11 Pause length change table

[FIG. 12]

Less than 200 cm	Less than 200 cm
None	Raise

55

FIG. 12 Phrase-end emphasis determination table

[FIG. 13]

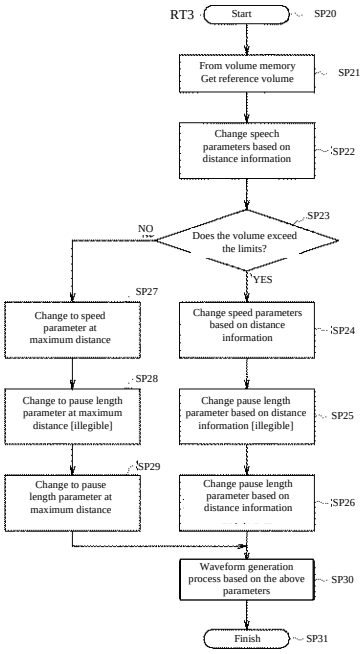

FIG. 13 Speech change processing procedure

[FIG. 1]

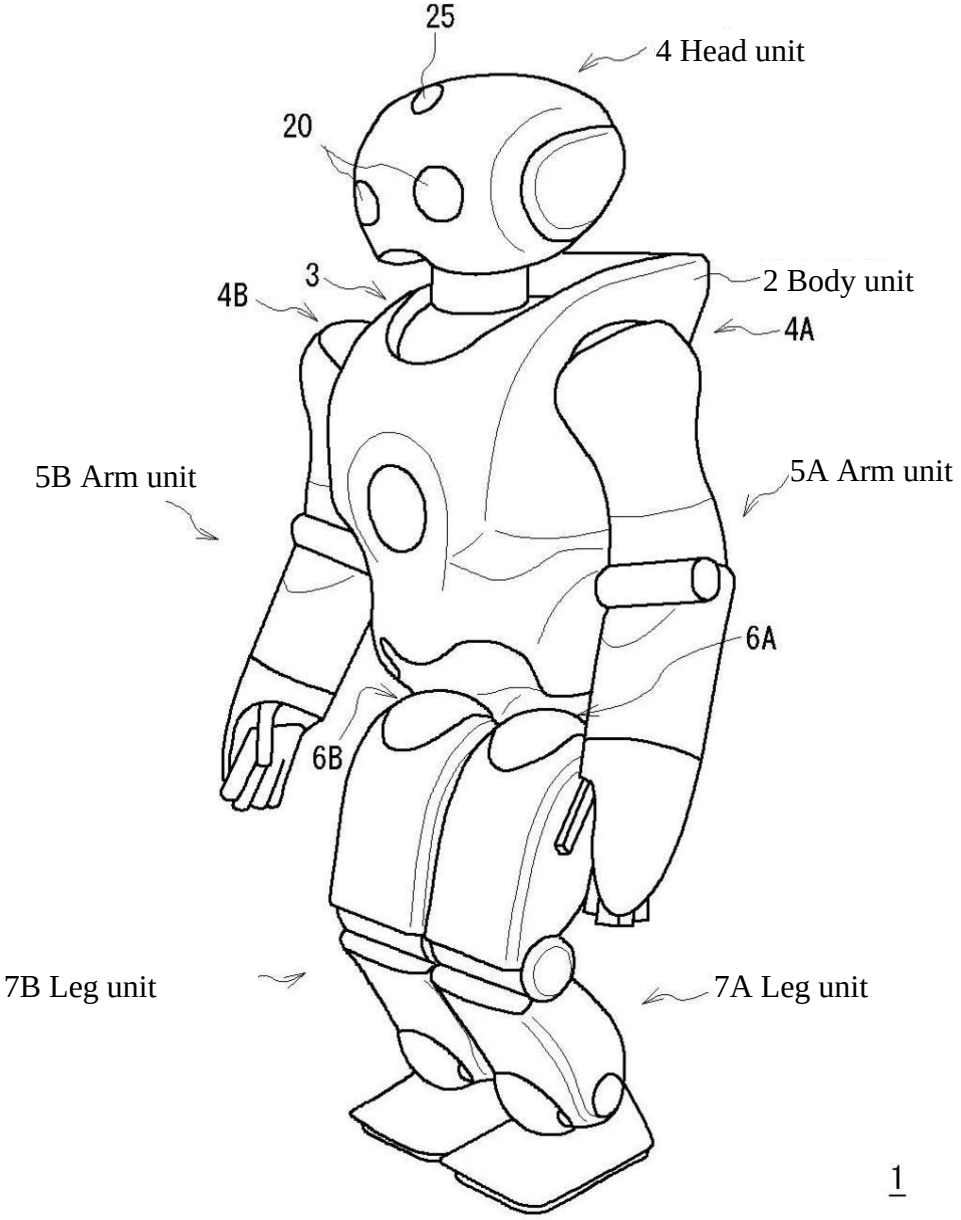

FIG. 1 Hardware constitution of the robot according to this embodiment (1)