DECLARATION OF GWEN SNORTELAND

I, Gwen Snorteland, pursuant to 28 U.S.C. § 1746, hereby declare as follows:

1. I am a Translator at TransPerfect, Inc.

2. I submit this declaration to certify the accuracy of the English translation of Japanese Unexamined Patent Application Publication Number 2005-202076.

3. My statements are based on personal knowledge and my review of Japanese Unexamined Patent Application Publication Number 2005-202076 and its Japanese to English translation. If called as a witness about the facts contained in these statements, I could testify competently based on such personal knowledge and the investigation I have conducted.

4. Attached as Exhibit A is a true and accurate copy of Japanese Unexamined Patent Application Publication Number 2005-202076.

5. Attached as Exhibit B is a true and accurate copy of an English translation of Japanese Unexamined Patent Application Publication Number 2005-202076 ("the Japanese Unexamined Patent Application Publication Number 2005-202076 Translation").

6. The Japanese Unexamined Patent Application Publication Number 2005-202076 Translation is a true and accurate translation of Japanese Unexamined Patent Application Publication Number 2005-202076 from Japanese to English.

7. All statements made herein of my own knowledge are true, and all statements made on information and belief are believed to be true. Further, I am aware that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001. I declare under penalty of perjury that the foregoing is true and correct.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct. Executed on September 5, 2024, at Cleveland, Ohio.



Gwen Snorteland

Amazon v. SoundClear
US Patent 11,244,675
Amazon Ex. 1005

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号

特開2005-202076

(P2005-202076A)

(43)公開日　平成17年7月28日(2005.7.28)

(51)Int.Cl.[7]		F I			テーマコード（参考）
G10L	13/08	G10L	3/00	H	2C150
A63H	11/00	A63H	11/00	Z	3C007
B25J	13/00	B25J	13/00	Z	5D045
G10L	13/00	G10L	3/00	Q	

審査請求 未請求　請求項の数 24　ＯＬ　（全 19 頁）

(21)出願番号　特願2004-7306(P2004-7306)

(22)出願日　平成16年1月14日(2004.1.14)

（特許庁注：以下のものは登録商標）

１．Ｂｌｕｅｔｏｏｔｈ

(71)出願人　000002185

ソニー株式会社

東京都品川区北品川６丁目７番３５号

(74)代理人　100082740

弁理士　田辺　恵基

(72)発明者　下村　秀樹

東京都品川区北品川６丁目７番３５号ソニー株式会社内

Ｆターム(参考)　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)【発明の名称】発話制御装置及び方法並びにロボット装置

(57)【要約】

【課題】

機器又はロボット装置から対話相手までの距離によっては、機器又はロボット装置の発話が対話相手に聞き取り難い場合があった。

【解決手段】

機器又はロボット装置とユーザと間の距離に応じて、ユーザとの対話時における機器又はロボット装置の発話形態を必要に応じて変更させるようにしたことにより、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができ、かくしてエンターテインメント性を向上させ得る発話制御装置及び方法並びにロボット装置を実現できる。

【選択図】　図１３

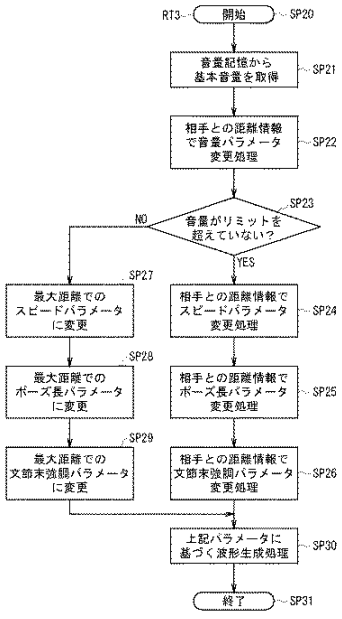


図１３　発話変更処理手順

【特許請求の範囲】
【請求項１】
ユーザとの対話機能を有する機器と上記ユーザとの対話時における当該機器の発話を制御する発話制御装置において、
上記機器及び対話相手の上記ユーザ間の距離を検出する距離検出手段と、
上記距離検出手段により検出された上記機器及び上記ユーザ間の上記距離に応じて、上記ユーザとの対話時における上記機器の発話形態を必要に応じて変更させる発話形態変更手段と
を具えることを特徴とする発話制御装置。
【請求項２】
上記発話形態変更手段は、
上記発話形態として、発話音量を変更する
ことを特徴とする請求項１に記載の発話制御装置。
【請求項３】
上記発話形態変更手段は、
上記発話形態として、発話スピードを変更する
ことを特徴とする請求項１に記載の発話制御装置。
【請求項４】
上記発話形態変更手段は、
上記発話形態として、発話のイントネーションを変更する
ことを特徴とする請求項１に記載の発話制御装置。
【請求項５】
上記発話形態変更手段は、
上記発話形態として、文節間の間を変更する
ことを特徴とする請求項１に記載の発話制御装置。
【請求項６】
上記機器が発話すべき内容に応じた文字列を出力する文字列出力手段と、
上記文字列に応じた合成音声の音声信号を生成する音声手段と、
上記音声信号に基づく音声を出力するスピーカと
を具え、
上記発話形態変更手段は、
上記文節間の間を変更する手法として、上記文字列出力手段から出力された上記文字列を変形する
ことを特徴とする請求項５に記載の発話制御装置。
【請求項７】
上記対話相手の上記ユーザを特定するユーザ特定手段と、
上記ユーザごとの基準音量を記憶する記憶手段と、
対話時における上記ユーザからの要求に応じて上記記憶手段に記憶された当該ユーザの上記基準音量を変更する基準音量変更手段と
を具え、
上記発話形態変更手段は、
上記基準音量変更手段に記憶された上記対話相手の上記ユーザの上記基準音量を基準として、上記発話音量を変更する
ことを特徴とする請求項２に記載の発話制御装置。
【請求項８】
上記発話形態変更手段は、
変更後の上記発話音量が所定の閾値を超える場合には、他の発話形態を最大量変更する
ことを特徴とする請求項２に記載の発話制御装置。
【請求項９】
ユーザとの対話機能を有する機器と上記ユーザとの対話時における当該機器の発話を制

御する発話制御方法において、
上記機器及び対話相手の上記ユーザ間の距離を検出する第１のステップと、
検出した上記機器及び上記ユーザ間の上記距離に応じて、上記ユーザとの対話時における上記機器の発話形態を必要に応じて変更する第２のステップと
を具えることを特徴とする発話制御方法。
【請求項１０】
上記第２のステップでは、
上記発話形態として、発話音量を変更する
ことを特徴とする請求項９に記載の発話制御方法。
【請求項１１】
上記第２のステップでは、
上記発話形態として、発話スピードを変更する
ことを特徴とする請求項９に記載の発話制御方法。
【請求項１２】
上記第２のステップでは、
上記発話形態として、発話のイントネーションを変更する
ことを特徴とする請求項９に記載の発話制御方法。
【請求項１３】
上記第２のステップでは、
上記発話形態として、文節間の間を変更する
ことを特徴とする請求項９に記載の発話制御方法。
【請求項１４】
上記第２のステップは、
上記機器が発話すべき内容に応じた文字列を出力する文字列出力ステップと、
上記文字列に応じた合成音声の音声信号を生成する音声生成ステップと、
上記音声信号に基づく音声を出力する音声出力ステップと
を具え、
上記音声生成ステップでは、
上記発話のイントネーションを変更する手法として、出力した上記文字列を変形する
ことを特徴とする請求項１３に記載の発話制御方法。
【請求項１５】
上記ユーザごとの基準音量を記憶する記憶ステップと、
対話相手の上記ユーザを特定するユーザ特定ステップと、
対話時における上記ユーザからの要求に応じて記憶した当該ユーザの上記基準音量を変更する基準音量変更ステップと
を具え、
上記第２のステップでは、
記憶した上記対話相手の上記ユーザの上記基準音量を基準として、上記発話音量を変更する
ことを特徴とする請求項１０に記載の発話制御方法。
【請求項１６】
上記第２のステップでは、
変更後の上記発話音量が所定の閾値を超える場合には、他の発話形態を最大量変更する
ことを特徴とする請求項１０に記載の発話制御方法。
【請求項１７】
ユーザとの対話機能を有するロボット装置において、
対話相手の上記ユーザまでの距離を検出する距離検出手段と、
上記距離検出手段により検出された上記ユーザまでの上記距離に応じて、上記ユーザとの対話時における発話形態を必要に応じて変更する発話形態変更手段と
を具えることを特徴とするロボット装置。

【請求項１８】
上記発話形態変更手段は、
上記発話形態として、発話音量を変更する
ことを特徴とする請求項１７に記載のロボット装置。
【請求項１９】
上記発話形態変更手段は、
上記発話形態として、発話スピードを変更する
ことを特徴とする請求項１７に記載のロボット装置。
【請求項２０】
上記発話形態変更手段は、
上記発話形態として、発話のイントネーションを変更する
ことを特徴とする請求項１７に記載のロボット装置。
【請求項２１】
上記ロボット装置が発話すべき内容に応じた文字列を出力する文字列出力手段と、
上記文字列に応じた合成音声の音声信号を生成する音声手段と、
上記音声信号に基づく音声を出力するスピーカと
を具え、
上記発話形態変更手段は、
上記発話のイントネーションを変更する手法として、上記文字列出力手段から出力された上記文字列を変形する
ことを特徴とする請求項２０に記載のロボット装置。
【請求項２２】
上記発話形態変更手段は、
上記発話形態として、文節間の間を変更する
ことを特徴とする請求項１７に記載のロボット装置。
【請求項２３】
上記対話相手の上記ユーザを特定するユーザ特定手段と、
上記ユーザごとの基準音量を記憶する記憶手段と、
対話時における上記ユーザからの要求に応じて上記記憶手段に記憶された当該ユーザの上記基準音量を変更する基準音量変更手段と
を具え、
上記発話形態変更手段は、
上記基準音量変更手段に記憶された上記対話相手の上記ユーザの上記基準音量を基準として、上記発話音量を変更する
ことを特徴とする請求項１８に記載のロボット装置。
【請求項２４】
上記発話形態変更手段は、
変更後の上記発話音量が所定の閾値を超える場合には、他の発話形態を最大量変更する
ことを特徴とする請求項１８に記載のロボット装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、発話制御装置及び方法並びにロボット装置に関し、例えばインターテインメイントロボットに適用して好適なものである。
【背景技術】
【０００２】
近年、多くの企業や大学等の研究機関においてヒューマノイド型ロボットの開発が進められている。かかるヒューマノイド型ロボットは、ＣＣＤ（Charge Coupled Device）カメラ、マイクロホン及びタッチセンサ等の外部センサと、バッテリセンサ及び加速度センサ等の内部センサとが搭載され、これら外部センサ及び内部センサの出力に基づいて外部

及び内部の状態を認識し、認識結果に基づいて自律的に行動し得るようになされたものである（例えば非特許文献１参照）。

【０００３】

また近年では、音声認識機能及び対話制御機能が搭載され、ユーザとの間で簡単な日常会話を行い得るようになされたエンターテインメントロボットも多く見受けられるようになってきた。

【非特許文献１】特願２００３－２７０８３５

【発明の開示】

【発明が解決しようとする課題】

【０００４】

ところで、かかる音声認識及び対話制御機能が搭載された従来のロボットでは、ロボットとユーザとの物理的な距離にかかわりなく、常に予め設定された一定の発話音量でユーザと対話するように構築されている。

【０００５】

このため発話音量の設定によっては、ロボットの近くにいるユーザにとっては適切であったとしても、ロボットから少し離れた場所にいるユーザにとっては音が小さ過ぎて発話内容が聞き取り難かったり、逆にロボットから少し離れた場所にいるユーザにとっては適切であったとしても、ロボットの近くにいるユーザにとっては音が大き過ぎて発話内容が聞き取り難い場合がある問題があった。

【０００６】

かかる問題点を解決するための１つの方法として、スイッチ操作によりエンターテインメントロボットの発話音量を自在に変更し得るようにすることも考えられる。しかしながらこの方法によると、ユーザとロボットとのインタラクションの自然性が損なわれる問題があり、さらに発話音量の設定をその都度行わなければならないとすることは、ユーザにとって不便極まりない問題がある。

【０００７】

また残響が多い部屋など、周囲の環境を考慮した場合、単にロボットの発話音量を上げるだけでは、ロボットから離れた場所にいるユーザに対して常にロボットの発話を聞き取り易くさせ得ることになるとは限らない。

【０００８】

そしてこのようにロボットの発話内容がユーザに聞き取り難いという状況は、ユーザとロボットとのスムーズかつ自然な対話を阻害する要因となって、対話制御機能を有するロボットのエンターテインメント性を損なうものであり、何らかの解決策が望まれる。

【０００９】

本発明は以上の点を考慮してなされたもので、対話制御機能を有するロボットのエンターテインメント性を向上させ得る発話制御装置及び方法並びにロボット装置を提案しようとするものである。

【課題を解決するための手段】

【００１０】

かかる課題を解決するため本発明においては、機器とユーザとの対話時における当該機器の発話を制御する発話制御装置において、機器及びユーザ間の距離に応じて、ユーザとの対話時における機器の発話形態を必要に応じて変更させる発話形態変更手段を設けるようにした。

【００１１】

この結果この発話制御装置は、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができる。

【００１２】

また本発明においては、機器とユーザとの対話時における当該機器の発話を制御する発話制御方法において、機器及びユーザ間の距離に応じて、ユーザとの対話時における機器の発話形態を必要に応じて変更するようにした。

【００１３】
この結果この発話制御方法によれば、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができる。
【００１４】
さらに本発明においては、ロボット装置において、ユーザまでの距離に応じて、ユーザとの対話時における発話形態を必要に応じて変更する発話形態変更手段を設けるようにした。
【００１５】
この結果このロボット装置は、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができる。
【発明の効果】
【００１６】
本発明によれば、機器とユーザとの対話時における当該機器の発話を制御する発話制御装置及び方法において、機器及びユーザ間の距離に応じて、ユーザとの対話時における機器の発話形態を必要に応じて変更させるようにしたことにより、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができ、かくしてエンターテインメント性を向上させ得る発話制御装置及び方法を実現できる。
【００１７】
また本発明によれば、ロボット装置において、ユーザまでの距離に応じて、ユーザとの対話時における発話形態を必要に応じて変更する発話形態変更手段を設けるようにしたことにより、常にユーザに聞き取り易い発話形態で発話することができるため、ユーザとスムーズな対話を行うことができ、かくしてエンターテインメント性を向上させ得るロボット装置を実現できる。
【発明を実施するための最良の形態】
【００１８】
以下図面について、本発明の一実施の形態を詳述する。
【００１９】
（１）本実施の形態によるロボットの構成
（１－１）ロボット１のハードウェア構成
図１において、１は全体として本実施の形態によるロボットを示し、胴体部ユニット２の上部に首関節３を介して頭部ユニット４が取り付けられると共に、当該胴体部ユニット２の左右上部にそれぞれ肩関節４Ａ、４Ｂを介して腕部ユニット５Ａ、５Ｂが取り付けられ、かつ胴体部ユニット２の下部にそれぞれ股関節６Ａ、６Ｂを介して一対の脚部ユニット７Ａ、７Ｂが取り付けられることにより構成されている。
【００２０】
図２は、このロボット１の機能構成を模式的に示したものである。この図２に示すように、ロボット１は、全体の動作の統括的制御やその他のデータ処理を行う制御ユニット１０と、入出力部１１と、駆動部１２と、電源部１３とで構成される。
【００２１】
入出力部１１は、入力部としてロボット１の目に相当する一対のＣＣＤ（Charge Coupled Device）カメラ２０や、耳に相当する一対のマイクロホン２１、頭部や手及び足裏などの部位に配設されてユーザからの物理的な働きかけや、手と外部物体との接触、足裏面の接地等を感知するタッチセンサ２２、あるいは五感に相当するその他の各種のセンサを含む。
【００２２】
また入出力部１１は、出力部としてロボット１の口に相当するスピーカ２３、あるいは点滅の組み合わせや点灯のタイミングにより顔の表情を形成するＬＥＤ（目ランプ）２４などを装備している。これら出力部は、音声やランプの点滅など、脚などによる機械運動パターン以外の形式でもロボット１からのユーザ・フィードバックを表現することができる。

【００２３】
駆動部１２は、制御ユニット１０が指令する所定の運動パターンに従ってロボット１の機体動作を実現する機能ブロックであり、行動制御による制御対象物である。駆動部１２は、ロボット１の各関節における自由度を実現するための機能モジュールであり、それぞれの関節におけるロール、ピッチ、ヨーなど各軸毎に設けられた複数の駆動ユニット２５$_1$～２５$_n$で構成される。各駆動ユニット２５$_1$～２５$_n$は、所定軸回りの回転動作を行うモータ２６$_1$～２６$_n$と、モータ２６$_1$～２６$_n$の回転位置を検出するエンコーダ２７$_1$～２７$_n$と、エンコーダ２７$_1$～２７$_n$の出力に基づいてモータ２６$_1$～２６$_n$の回転位置や回転速度を適応的に制御するドライバ２８$_1$～２８$_n$の組み合わせで構成される。

【００２４】
電源部１３は、その字句通り、ロボット１内に各電気回路などに対して給電を行う機能モジュールである。本実施の形態に係るロボット１は、バッテリを用いた自律駆動式であり、電源部１３は、充電バッテリ２９と、充電バッテリ３０の充放電状態を管理する充放電制御部３１とで構成される。

【００２５】
充電バッテリ２９は、例えば、複数本のリチウムイオン２次電池セルをカートリッジ式にパッケージ化した「バッテリ・パック」の形態で構成される。

【００２６】
また、充放電制御部３０は、バッテリ２９の端子電圧や充電／放電電流量、バッテリ２９の周囲温度などを測定することでバッテリ２９の残存容量を把握し、充電の開始時期や終了時期などを決定する。充放電制御部３０が決定する充電の開始及び終了時期は制御ユニット１０に通知され、ロボット１が充電オペレーションを開始及び終了するためのトリガとなる。

【００２７】
制御ユニット１０は、ロボット１の「頭脳」に相当し、例えば頭部ユニット４又は胴体部ユニット２内に搭載されている。この制御ユニット１０は、図３に示すように、メイン・コントローラとしてのＣＰＵ（Central Processing Unit）３１が、メモリやその他の各回路コンポーネントや周辺機器とバス接続された構成となっている。バス３７は、データ・バス、アドレス・バス、コントロール・バスなどを含む共通信号伝送路である。バス３７上の各装置にはそれぞれに固有のアドレス（メモリ・アドレス又はＩ／Ｏアドレス）が割り当てられている。ＣＰＵ３１は、アドレスを指定することによってバス３７上の特定の装置と通信することができる。

【００２８】
ＲＡＭ（Read Access Memory）３２は、ＤＲＡＭ（Dynamic RAM）などの揮発性メモリで構成された書き込み可能メモリであり、ＣＰＵ３１が実行するプログラム・コードをロードしたり、実行プログラムによる作業データの一時的な保存のために使用される。

【００２９】
ＲＯＭ（Read Only Memory）３３は、プログラムやデータを恒久的に格納する読み出し専用メモリである。ＲＯＭ３３に格納されるプログラム・コードには、ロボット１の電源投入時に実行する自己診断テスト・プログラムや、ロボット１の動作を規定する制御プログラムなどが挙げられる。

【００３０】
ロボット１の制御プログラムには、ＣＣＤカメラ２０やマイクロホン２１などの各種センサからの入力を処理してシンボルとして認識する「センサ入力・認識処理プログラム」、短期記憶などの記憶動作を司りながらセンサ入力と所定の行動制御モデルとに基づいてロボット１の行動を制御する「行動制御プログラム」、行動制御モデルに従って各関節モータの駆動やスピーカ２２の音声出力などを制御する「駆動制御プログラム」などが含まれる。

【００３１】
不揮発性メモリ３４は、例えばＥＥＰＲＯＭ（Electrically Erasable and Programmab

le ROM）のように電気的に消去再書き込みが可能なメモリ素子で構成され、逐次更新すべきデータを不揮発的に保持するために使用される。逐次更新すべきデータには、暗記鍵やその他のセキュリティ情報、出荷後にインストールすべき装置制御プログラムなどが挙げられる。

【００３２】

インターフェース３５は、制御ユニット１０外の機器と相互接続し、データ交換を可能にするための装置である。インターフェース３５は、例えば、入出力部１１内のＣＣＤカメラ２０やマイクロホン２１及びスピーカ２２との間でデータ入出力を行う。また、インターフェース３５は、駆動部１２内の各ドライバ28_1～28_nとの間でデータやコマンドの入出力を行う。

【００３３】

また、インターフェース３５は、ＲＳ（Recommended Standard）－２３２Ｃなどのシリアル・インターフェース、ＩＥＥＥ（Institute of Electrical and Electronics Engineers ）１２８４などのパラレル・インターフェース、ＵＳＢ（Universal Serial Bus）インターフェース、ｉ－Ｌｉｎｋ（ＩＥＥＥ１３９４）インターフェース、ＳＣＳＩ（Small Computer System Interface ）インターフェース、ＰＣカードやメモリ・スティックを受容するメモリ・カードインターフェース（カードスロット）などのような、コンピュータの周辺機器接続用の汎用インターフェースを備え、ローカル接続された外部機器との間でプログラムやデータの移動を行い得るようにしてもよい。また、インターフェース３５の他の例として、赤外線通信（ＩｒＤＡ）インターフェースを備え、外部機器と無線通信を行うようにしてもよい。

【００３４】

さらに、制御ユニット１０は、無線通信インターフェース３６やネットワーク・インターフェース・カード（ＮＩＣ）３８などを含み、Ｂｌｕｅｔｏｏｔｈのような近接無線データ通信や、ＩＥＥＥ８０２．１１ｂのような無線ネットワーク、あるいはインターネットなどの高域ネットワークを経由して、外部のさまざまなホスト・コンピュータとデータ通信を行うことができる。

【００３５】

このようなロボット１とホスト・コンピュータ間におけるデータ通信により、遠隔のコンピュータ資源を用いて、ロボット１の複雑な動作制御を演算したり、リモート・コントロールすることができる。

【００３６】

（１－２）ロボット１のソフトウェア構成

図４は、ＲＯＭに格納された制御プログラム群により構成されるロボット１の行動制御システム４０の機能構成を模式的に示したものである。

【００３７】

この行動制御システム４０は、オブジェクト指向プログラミングを採り入れて実装されている。この場合、各ソフトウェアは、データとそのデータに対する処理手続きとを一体化させた「オブジェクト」というモジュール単位で扱われる。また各オブジェクトは、メッセージ通信と共有メモリを使ったオブジェクト間通信方法によりデータの受け渡しとＩｎｖｏｋｅを行うことができる。

【００３８】

ここで行動制御システム４０は、ＣＣＤカメラ２０（図２）、マイクロホン２１（図２）及びタッチセンサ２２（図２）などの各種センサからのセンサ出力に基づいて外部環境を認識するための画像認識部４１、音声認識部４２及び接触認識部４３を有している。

【００３９】

画像認識部４１は、ＣＣＤカメラ２０から与えられる画像信号Ｓ１に基づいて顔認識や色認識などの画像認識処理や特徴抽出を実行する。そして画像認識部４１は、かかる顔認識結果であるその人物に固有の顔ＩＤ（識別子）、顔画像領域の位置及び大きさといった顔認識情報や、色認識結果である色領域の位置や大きさ及び特徴量といった色認識情報な

どの各種画像認識結果と、画像信号Ｓ１とを短期記憶部４４に送出する。また画像認識部４１は、ＣＣＤカメラカメラ２０からの画像信号Ｓ１に基づいて、いわゆるステレオビジョン法によって撮像対象までの距離を検出し、検出結果を短期記憶部４４に送出する。

【００４０】

音声認識部４２は、マイクロホン２１から与えられる音声信号Ｓ２に基づいて音声認識や話者認識、音源方向認識などの各種音に関する認識処理を実行する。そして音声認識部４２は、かかる音声認識結果である認識した単語の文字列情報と、音響的特徴等に基づく話者認識処理結果であるその話者に固有の話者ＩＤ情報と、音源方向認識結果である音源方向情報となどの各種音声認識結果を短期記憶部４４に送出する。また音声認識部４２は、これら各種音声認識結果と併せて、これら音声信号Ｓ２を短期記憶部４４に送出する。

【００４１】

さらに接触認識部４３は、頭部ユニット４（図１）の上部や腕部ユニット５Ａ、５Ｂ（図１）の先端部である手、脚部ユニット７Ａ、７Ｂ（図１）の底部である足底等に配設されたタッチセンサ２２からそれぞれ与えられる圧力検出信号Ｓ３に基づいて「撫でられた」、「叩かれた」、「物を把持した」、「足裏面が接地した」という外部との物理的な接触を認識し、得られたこれら接触認識結果を短期記憶部４４に送出する。また接触認識部４４は、これら接触認識結果と併せて、各タッチセンサ２２からの圧力検出信号Ｓ３を短期記憶部４４に送出する。

【００４２】

短期記憶部４４は、ロボット１の外部環境に関する情報を比較的短い時間だけ保持するオブジェクトであり、画像認識部４１から与えられる各種画像認識結果及び画像信号Ｓ１と、音声認識部４２から与えられる各種音声認識結果及び音声信号Ｓ２と、接触認識部４３から与えられる各種接触認識結果及び圧力検出信号Ｓ３とを受け取り、これらを短期間だけ記憶する。

【００４３】

また短期記憶部４４は、これら受け取った画像認識結果、音声認識結果及び接触認識結果と、画像信号Ｓ１、音声信号Ｓ２及び各圧力検出信号Ｓ３とを統合的に用いて顔画像領域、人物ＩＤ、話者ＩＤ及び文字列情報等の対応付けを行うことにより、現在どこにどの人物がいて、発した言葉がどの人物のものであり、その人物とはこれまでにどのような対話を行ったかというターゲット情報及びイベント情報を生成し、これを行動選択制御部４５に送出する。

【００４４】

行動選択制御部４５は、短期記憶部４４から与えられるターゲット情報及びイベント情報と、短期記憶部４４の記憶内容とに基づいて、予め複数用意されている行動の中からロボット１が現在置かれている状況及び以前の行動に依存して選択した行動（状況依存行動）や、外部刺激に応じた反射的な行動（反射行動）、又は与えられた状況若しくはユーザからの命令に応じた比較的長期に渡る行動計画に基づく行動（熟考行動）などを次のロボット１の行動として決定する。そして行動選択制御部４５は、このようにして決定した行動を出力管理部４６に通知する。

【００４５】

出力管理部４６は、行動選択制御部４５からの通知に応じて、状況依存行動及び反射行動などの複数の行動が競合した場合の調停処理や、動作、音声及びＬＥＤ２４の点滅の同期をとる処理を行いながら、対応する駆動ユニット25_1～25_nのモータ26_1～26_nを駆動したり、ＬＥＤ２４を所定パターンで点滅駆動させる。

【００４６】

また行動選択制御部４５は、次の行動としてユーザとの対話を決定した場合には、この後音声認識部４２により短期記憶部４４に順次格納されるユーザの発話の音声認識結果を常時監視し、この音声認識結果に基づいて、ロボット１に発話させるべき内容を順次決定する。そして行動選択制御部４５は、この決定結果に基づいて必要な文字列を予めＲＯＭ３３（図３）に格納された発話文字列データベース４７から読み出し、これを出力管理部

４６に送出する。

【００４７】

このとき出力管理部４６は、行動選択制御部４５から与えられる文字列を音声合成部４８に送出する一方、音声合成部４８は、供給される文字列に基づいて合成音声の音声信号Ｓ４を生成し、これをスピーカ２３（図２）に送出する。この結果この音声信号Ｓ４に基づく音声がスピーカ２３から出力される。

【００４８】

このようにしてこのロボット１においては、ＣＣＤカメラ２０、マイクロホン２１及びタッチセンサ２３などの各種センサのセンサ出力に基づき認識される外部状況等に基づいて自律的に行動し得るようになされている。

【００４９】

（２）ロボット１における発話制御機能

次に、このロボット１に搭載された発話制御機能について説明する。

【００５０】

このロボット１には、対話相手までの物理的な距離に応じて、発話音量や、発話スピード、イントネーション及び文節間の間などの発話形態を制御する発話制御機能が搭載されている。またこのロボット１には、このような発話形態のうちの特に発話音量に関する感じ方の個人差が大きいことを考慮して、ユーザからの要求に応じて基準とすべき発話音量（以下、これを基準音量と呼ぶ）をユーザごとに変更する発話制御機能も搭載されている。

【００５１】

実際上、このロボット１の場合、基準音量をユーザごとに変更する発話制御を行うための手段として、行動制御システム４０には、ユーザごとの基準音量を記憶保持するための基準音量記憶部４９が設けられている。因みに、この基準音量記憶部４９は、不揮発性メモリ３４（図３）により構成されるものである。

【００５２】

そして行動選択制御部４５は、画像認識処理部４１の画像認識結果に基づいて短期記憶部４４が新たなユーザを検出するごとに、図５に示すように、そのユーザの人物ＩＤと当該ユーザに対する基準音量（初期設定値は「３」）とを対応付けて、これら人物ＩＤ及び基準音量を基準音量記憶部４９に格納する一方、この後そのユーザとの対話時に図６に示す基準音量変更処理手順ＲＴ１に従って、基準音量記憶部４９に記憶保持されたそのユーザの基準音量を必要に応じて変更するようになされている。

【００５３】

すなわち行動選択制御部４５は、ユーザとの対話を開始するとこの基準音量変更処理手順ＲＴ１をステップＳＰ０において開始し、続くステップＳＰ１において、「もう少し大きな声で話して」又は「聞こえない」などといった、ロボットの発話音量を上げることを要求する意味合いの言葉や、「もう少し小さな声で話して」又は「うるさい」などといった、ロボットの発話音量を下げることを要求する意味合いの言葉が与えられるのを待ち受ける。

【００５４】

そして行動選択制御部４５は、この後上述のようなロボット１の発話音量を上げ又は下げることを要求する意味合いの言葉がユーザから発せられたことを短期記憶部４４に記憶保持された音声認識部４２の音声認識結果に基づいて認識すると、ステップＳＰ２に進んで、短期記憶部４４から与えられるターゲット情報及びイベント情報に基づいて、その言葉が現在の対話相手から発せられたものであるとの特定ができているか否かを判断する。

【００５５】

行動選択制御部４５は、このステップＳＰ２において否定結果を得るとステップＳＰ１に戻り、これに対して肯定結果を得るとステップＳＰ３に進んで、基準音量記憶部４９に記憶保持されたそのユーザの基準音量を、そのユーザが発した言葉の意味内容に応じて、発話音量を上げることを要求する意味合いの言葉であった場合には規定値（例えば「１」

）だけ上げ、これに対して発話音量を下げることを要求する意味合いの言葉であった場合には規定値だけ下げる。なお、音量の数値はロボット１の発話音量の音量レベルの大小を表すものであり、数値が大きくなるほど大きな所定の音量レベルがそれぞれ対応付けられている。

【００５６】

そして行動選択制御部４５は、その後ステップＳＰ１に戻り、この後ステップＳＰ１～ステップＳＰ３について同様の処理を繰り返す。このようにして行動選択制御部４５は、ユーザからの発話音量の変更要求に応じて、ユーザごとに基準音量を変更する。

【００５７】

一方、このロボット１では、対話相手までの物理的な距離に応じて、意味内容の区切りがはっきりと分かるように文節間の間を変更するための手法として、必要に応じて各文節の後ろに「ね」を入れるように発話制御を行うようになされ、そのための手段として行動制御システム４５に発話文字列変形部５１が設けられている。

【００５８】

この場合、発話文字列変形部５１は、図７に示すように、対話相手までの距離に対して各文節の後ろに「ね」を入れるように文字列を変形するか否かを規定した文字列変形判断テーブル５０をＲＯＭ３３（図３）に記憶保持している。

【００５９】

そして行動選択制御部４５は、ユーザとの対話時、ロボット１が発話すべき内容（言葉）に対応した文字列を発話文字列データベース４７から読み出し、これを文字列情報として発話文字列変形部５１に順次送出する一方、これと併せて短期記憶部４４に記憶保持された画像認識部４１により検出された対話相手までの距離を表す距離情報を発話文字列変形部５１に送出する。

【００６０】

発話文字列変形部５１は、行動選択制御部４５から文字列及び距離情報が与えられると、当該距離情報及び文字列変形判断テーブル５０に基づき、図８に示す文字変形処理手順ＲＴ２に従って、その文字列情報に基づく文字列を変形すべきか否かを判断し、必要時にはこの文字列を変形する。

【００６１】

すなわち発話文字列変形部４９は、行動選択制御部４５から文字列情報及び距離情報が与えられると、この文字列変形処理手順ＲＴ２をステップＳＰ１０において開始し、続くステップＳＰ１１において行動選択制御部４５からの距離情報に基づき認識される対話相手までの距離と、文字列変形判断テーブル５０（図７）とに基づいて、この文字列情報に基づく文字列を変形すべきか否かを判断する。

【００６２】

具体的に発話文字列変形部４９は、文字列変形判断テーブル５０に従って、画像認識部４１が認識した対話相手までの距離が350〔cm〕未満である場合には文字列を変形すべきでないと判断し、これに対して対話相手までの距離が350〔cm〕以上である場合には文字列を変形すべきと判断する。

【００６３】

発話文字列変形部４９は、この判断結果としてこのステップＳＰ１１において否定結果を得ると、その文字列に対して変形処理を施すことなくステップＳＰ１４に進んでこの文字列変形処理手順ＲＴ２を終了し、この後この文字列の文字列情報をそのまま行動選択制御部４５、出力管理部４６を順次介して音声合成部４８に送出する。

【００６４】

これに対して発話文字列変形部４９は、ステップＳＰ１１において肯定結果を得ると、ステップＳＰ１２に進んでその文字列に対して形態素解析処理を行ってその文字列に含まれる各文節末助動詞をそれぞれ検出する。

【００６５】

さらに発話文字列変形部４９は、この後ステップＳＰ１３に進んで、ステップＳＰ１２

において検出したこれら文節末助動詞の後ろに順次「ね」の文字を付加する。従って、発話文字列変形部４９は、例えば行動選択制御部４５から「今日の東京の天気は晴れだよ」という文字列を与えられた場合、「今日の」、「東京の」及び「天気は」という各文節末助動詞の後ろにそれぞれ「ね」の文字を付加した「今日のね、東京のね、天気はね、晴れだよ」といった文字列を生成することとなる。

【００６６】

そして発話文字列変形部４９は、この後ステップＳＰ１４に進んでこの文字列変形処理手順ＲＴ２を終了し、この後このようにして得られた変形した新たな文字列を表す文字列情報を行動選択制御部４５、出力管理部４６を順次介して音声合成部４８に送出する。

【００６７】

このとき音声合成部４８は、図９に示すように、対話相手までの距離に対するロボット１の発話音量の変更量を規定した音量変更量規定テーブル５２と、図１０に示すように、当該距離に対するロボット１の発話スピードの変更量を規定したスピード変更量規定テーブル５３と、図１１に示すように、当該距離に対する文節間の間（ポーズ長）の変更量を規定したポーズ長変更量規定テーブル５４と、図１２に示すように、当該距離に対して発話のイントネーションを変えて文節末を強調するか否かを規定した文節末強調判断テーブル５５とをＲＯＭ３３（図３）に記憶保持している。また音声合成部４８には、短期記憶部４４からそのときの対話相手の人物ＩＤが人物ＩＤ情報として与えられる一方、当該対話相手までの距離が距離情報として順次与えられる。

【００６８】

そして音声合成部４８は、出力管理部４６を介してロボット１が発話すべき内容の文字列情報が与えられると、短期記憶部４４から与えられる人物ＩＤ情報及び距離情報と、ＲＯＭ３３に格納された各テーブル５２～５５と、基準音量記憶部４９に記憶保持されたそのユーザの基準音量とに基づき、図１３に示す発話変更処理手順ＲＴ３に従って、ロボット１の発話音量や、発話スピード等を必要に応じて変更する。

【００６９】

すなわち音声合成部４８は、かかる文字列情報が与えられるとこの文字列変形処理手順ＲＴ３をステップＳＰ２０において開始し、続くステップＳＰ２１において、短期記憶部４４から与えられた人物ＩＤ情報に基づいて、基準音量記憶部４９から対応する対話相手の基準音量を読み出し、この後ステップＳＰ２２に進んで距離情報から認識される対話相手までの距離に基づき、音量変更量規定テーブル５２（図９）を参照して、ロボット１の発話音量を規定する音量パラメータのパラメータ値を必要に応じて変更する。

【００７０】

具体的に音声合成部４８は、距離情報として与えられた対話相手までの距離が例えば50〔cm〕未満である場合には、基準音量記憶部４９から読み出した基準音量をそのときの発話音量として決定し、対話相手までの距離が50～80〔cm〕の範囲である場合には、基準音量よりも所定の規定値（例えば「１」）だけ大きい音量をそのときの発話音量として決定し、以降同様にして、当該ユーザまでの距離が80～150〔cm〕、150～250〔cm〕又は250〔cm〕以上の場合には、それぞれ基準音量よりも「２」、「３」又は「５」だけ大きい音量をそのときの発話音量として決定する。そして音声合成部４８は、この後この決定結果に応じて音量パラメータのパラメータ値を必要に応じて変更する。

【００７１】

続いて音声合成部４８は、ステップＳＰ２３に進んで、ステップＳＰ２２において上述のように決定した音量が予め設定された閾値を超えたか否かを判断する。

【００７２】

すなわちロボット１の発話音量には一定の限界があるため、例えば基準音量が既に限界値やこれに近い値に設定されている場合には、ステップＳＰ２２において音量を決定する際に、そのユーザとの距離に応じて本来的には発話音量をもっと大きな音量に決定すべき場合においてもできないことがある。そしてこのような場合、かかるユーザにとって、ロボット１の発話の内容が聞き取り難く感じられるおそれがある。

【００７３】
そこで音声合成部４８は、このステップＳＰ２３において肯定結果を得た場合には、ステップＳＰ２４～ステップＳＰ２６において発話音量以外の発話スピード等の他の発話形態について通常の変更処理を行い、これに対してステップＳＰ２３において否定結果を得た場合には、ステップＳＰ２７～ステップＳＰ２９においてかかる他の発話形態について特殊な変更処理を行うようになされている。
【００７４】
実際上、音声合成部４８は、ステップＳＰ２３において肯定結果を得ると、ステップＳＰ２４に進んで、短期記憶部４４から与えられる距離情報により認識される対話相手までの距離に基づき、ＲＯＭ３３に格納されたスピード変更量規定テーブル５３（図１０）を参照して、ロボット１の発話スピードを規定するスピードパラメータのパラメータ値を必要に応じて変更する。
【００７５】
具体的に音声合成部４８は、距離情報として与えられた対話相手までの距離が例えば200〔cm〕未満である場合には、予め定められた所定の初期設定値を発話時のスピードとして決定し、対話相手までの距離が200～350〔cm〕の範囲である場合には、初期設定値よりも20〔％〕だけ遅いスピードをそのときの発話スピードとして決定し、対話相手までの距離が350〔cm〕以上の場合には、初期設定値よりも50〔％〕だけ遅いスピードをそのときの発話スピードとして決定する。そして音声合成部４８は、この後この決定結果に応じてスピードパラメータのパラメータ値を必要に応じて変更する。
【００７６】
続いて音声合成部４８は、ステップＳＰ２５に進んで、短期記憶部４４から与えられる距離情報により認識される対話相手までの距離に基づき、ＲＯＭ３３に格納されたポーズ長変更量規定テーブル５４（図１１）を参照して、その文字列を発話する際の文節間の間の長さ（ポーズ長）を規定するポーズ長パラメータを必要に応じて変更する。
【００７７】
具体的に音声合成部４８は、距離情報として与えられた対話相手までの距離が例えば100〔cm〕未満である場合には、予めポーズ長の初期設定値として定められた値をそのときのポーズ長として決定し、対話相手までの距離が100～350〔cm〕の範囲である場合には、初期設定値よりも30〔％〕だけ長い時間をそのときのポーズ長として決定し、対話相手までの距離が350〔cm〕以上の場合には、初期設定値よりも60〔％〕だけ長い時間をそのときのポーズ長として決定する。そして音声合成部４８は、この後この決定結果に応じてポーズ長パラメータのパラメータ値を必要に応じて変更する。
【００７８】
次いで音声合成部４８は、ステップＳＰ２６に進んで、短期記憶部４４から与えられる距離情報により認識される対話相手までの距離に基づき、ＲＯＭ３３に格納された文節末強調判断テーブル５５（図１２）を参照して、その文字列の内容を発話する際に各文節末を強調するイントネーションとなるように、対応するパラメータである文節末強調パラメータのパラメータ値を必要に応じて変更する。
【００７９】
具体的に音声合成部４８は、距離情報として与えられた対話相手までの距離が例えば200〔cm〕未満である場合には、その文字列を音声として発話する際に各文節末を変化させないことを決定し、対話相手までの距離が200〔cm〕以上の場合には各文節末をそれぞれ上げることを決定する。そして音声合成部４８は、この後この決定結果に応じて文節末強調パラメータのパラメータ値を必要に応じて変更する。
【００８０】
これに対して音声合成部４８は、ステップＳＰ２３において否定結果を得ると、ステップＳＰ２７に進んで、短期記憶部４４から与えられる距離情報により認識される対話相手までの距離にかかわりなく、ロボット１の発話スピードを、常に、スピード変更量規定テーブル５３（図１０）において最大距離に対して規定されている発話スピードに決定する

。従って、この実施の形態においては、発話スピードとして初期設定値よりも50〔％〕だけ遅いスピードがそのときの発話スピードとして決定される。そして音声合成部４８は、この後この決定結果に応じてスピードパラメータのパラメータ値を必要に応じて変更する。

【００８１】

続いて音声合成部４８は、ステップＳＰ２８に進んで、ロボット１の発話時のポーズ長を、常に、ポーズ長変更量規定テーブル５４（図１１）において最大距離に対して規定されている発話スピードに決定する。従って、この実施の形態においては、ポーズ長として初期設定値よりも60〔％〕だけ長い時間がそのときの発話スピードとして決定される。そして音声合成部４８は、この後この決定結果に応じてポーズ長パラメータのパラメータ値を必要に応じて変更する。

【００８２】

次いで音声合成部４８は、ステップＳＰ２９に進んで、その文字列の内容を発話する際における文節末強調判断パラメータを、常に、文節末強調判断テーブル５５（図１２）において最大距離に対して規定された状態に決定する。従って、この実施の形態においては、常に各文節末を強調すべき旨が決定される。そして音声合成部４８は、この後この決定結果に応じて文節末強調パラメータのパラメータ値を文節末を強調するように必要に応じて変更する。

【００８３】

そして音声合成部４８は、このようにして音量パラメータやスピードパラメータ等の各種パラメータを設定し終えると、ステップＳＰ３０に進んで、与えられた文字列情報に基づく文字列に応じた音声波形を生成すると共に、この音声波形を上述のようにして設定した音量パラメータ、スピードパラメータ、ポーズ長パラメータ及び文節末強調パラメータに応じて変形し、かくして得られた音声波形の音声信号Ｓ４をスピーカ２３に送出する。

【００８４】

このようにしてこのロボット１においては、対話相手までの物理的な距離に応じて発話音量や、発話スピード、イントネーション及び発話時の間などの発話形態を制御する一方、さらに発話音量についてはユーザごとに制御することにより、常に対話相手に聞き取り易い発話を行い得るようになされている。

【００８５】

（３）本実施の形態の動作及び効果

以上の構成において、このロボット１は、対話相手のユーザが離れた場所にいるほど、より大きな音量で、よりゆっくりと、文節間により大きな間を開けて発話する一方、そのユーザまでの距離によっては、さらに文節末を強調するようなイントネーションで発話したり、各文節の後ろに「ね」という言葉を挿入して発話する。

【００８６】

従って、このロボット１は、対話相手のユーザまでの距離にかかわりなく常にユーザに聞き取り易い発話音量で発話することができることは無論のこと、例えば対話場所が残響が多い部屋などである場合においても、常にユーザに聞き取り易い発話を行うことができ、この結果としてユーザとスムーズかつ自然な音量での対話を行うことができる。

【００８７】

この場合において、このロボット１では、かかる発話音量の基準音量をユーザからの要求に応じてユーザごとに変更するため、対話相手のユーザの聴覚特性に応じた適切な音量で各ユーザと対話を行うことができ、その分ユーザに不快感を与えることなく、より一層とユーザとスムーズな対話を行うことができる。

【００８８】

以上の構成によれば、対話相手までの物理的な距離に応じて発話音量や、発話スピード、イントネーション及び発話時の間などの発話形態を制御するようにしたことにより、常にユーザに聞き取り易い発話形態で発話を行うことができる。かくするにつきユーザとスムーズ対話を行うことができ、かくしてエンターテインメント性を向上させ得るロボット

を実現できる。

【００８９】

（４）他の実施の形態

なお上述の実施の形態においては、本発明をヒューマノイド型のロボット１に適用するようにした場合について述べたが、本発明はこれに限らず、この種々の形態のロボット装置及びロボット装置以外の対話機能を有する種々の形態の機器に広く適用することができる。

【００９０】

また上述の実施の形態においては、対話相手までの距離に応じて発話音量、発話スピード、イントネーション及び発話時の間を変更するようにした場合について述べたが、本発明はこれに限らず、これら発話形態のうちの一部のみを変更し又はこれら発話形態に加えて他の発話形態をも変更するようにしても良い。

【００９１】

この場合において、上述の実施の形態においては、ロボット１がイントネーションを変更する際に各文節末を強めるようなイントネーションで発話するようにした場合について述べたが、本発明はこれに限らず、文節末以外の例えば助詞を強調するようにイントネーションを変更するようにしても良い。

【００９２】

さらに上述の実施の形態においては、対話相手までの距離に応じて発話音量、発話スピード、イントネーション及び発話時の間を段階的に変更するようにした場合について述べたが、本発明はこれに限らず、これらを対話相手までの距離に応じて関数式を用いた演算等により連続的に変更するようにしても良い。この場合において例えばイントネーション（各文節末の強調）の変更については、対話相手までの距離に応じて各文節末の強調の程度を連続的に変化させるようにすれば良い。また各文節末の後ろに「ね」を挿入することによる対話時の間については、全ての文節末の後ろに挿入するのではなく、対話相手までの距離に応じて段階的に「ね」を挿入する文節末の数を変化させるようにしても良い。なお、文節間の間を稼ぐために各文節末の後ろに挿入する言葉としては「ね」以外の言葉を適用できることは言うまでもない。

【００９３】

さらに上述の実施の形態においては、対話相手のユーザまでの距離を検出する距離検出手段を、一対のＣＣＤカメラ２０と、その出力に基づきステレオビジョン法により当該距離を検出する短期記憶部４４とにより構成するようにした場合について述べたが、本発明はこれに限らず、例えば距離センサなどの他の手段を広く適用することができる。

【００９４】

さらに上述の実施の形態においては、対話相手のユーザまでの距離に応じて、当該ユーザとの対話時におけるロボット１の発話形態を必要に応じて変更させる発話形態変更手段としての機能を、発話文字列変形部５１及び音声合成部４８に分散させるようにした場合について述べたが、本発明はこれに限らず、かかる全ての機能を音声合成部４８に搭載するようにしても良い。

【００９５】

さらに上述の実施の形態においては、対話相手のユーザを特定するユーザ特定手段としての機能を短期記憶部４４にもたせるようにした場合について述べたが、本発明はこれに限らず、かかる機能を行動選択制御部４５にもたせるようにしても良い。

【産業上の利用可能性】

【００９６】

本発明は、エンターテインメントロボットのほか、対話機能を有する他の用途のロボット装置やロボット装置以外の他の機器に広く適用することができる。

【図面の簡単な説明】

【００９７】

【図１】本実施の形態によるロボットの外観構成を示す斜視図である。

【図２】本実施の形態によるロボットの内部構成を示すブロック図である。
【図３】制御ユニットの構成を示すブロック図である。
【図４】本実施の形態によるロボットの行動制御システムの具体的構成を示すブロック図である。
【図５】基準音量記憶部に記憶保持されたユーザごとの基準音量の説明に供する概念図である。
【図６】基準音量変更処理手順を示すフローチャートである。
【図７】文字列変形判断テーブルを示す概念図である。
【図８】文字列変形処理手順を示すフローチャートである。
【図９】音量変更量テーブルを示す概念図である。
【図１０】スピード変更量テーブルを示す概念図である。
【図１１】ポーズ長変更量テーブルを示す概念図である。
【図１２】文節末強調判断テーブルを示す概念図である。
【図１３】発話変更処理手順を示すフローチャートである。
【符号の説明】
【００９８】
１……ロボット、４０……行動制御システム、４１……画像認識部、４２……音声認識部、４４……短期記憶部、４５……行動選択制御部、４７……発話文字列データベース、４８……音声合成部、４９……基準音量記憶部、５０……文字列変形判断テーブル、５１……発話文字列変形部、５２……音量変更量テーブル、５３……スピード変更量テーブル、５４……ポーズ長変更量テーブル、５５……文節末強調判断テーブル、Ｓ１……画像信号、Ｓ２、Ｓ４……音声信号、ＲＴ１……基準音量変更処理手段、ＲＴ２……文字列変形処理手順、ＲＴ３……発話変更処理手順。

【図２】

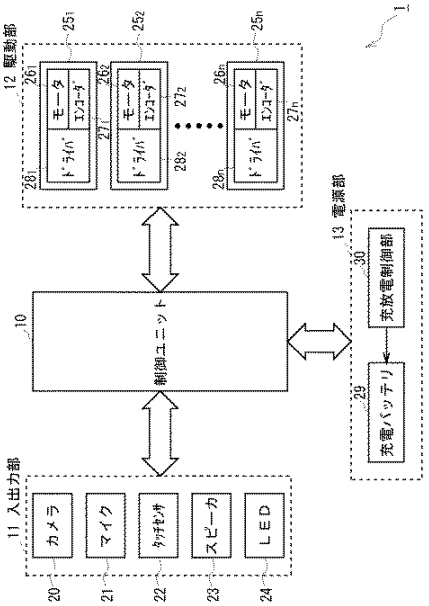


図２　本実施の形態によるロボットのハードウェア構成（２）

【図３】

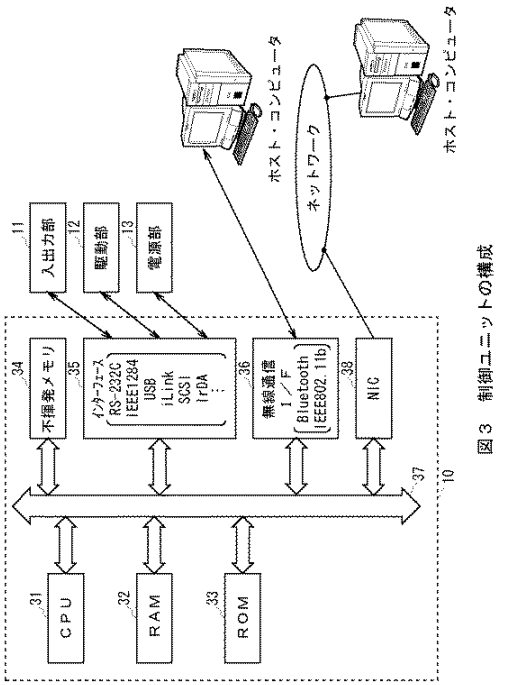


図３　制御ユニットの構成

【図４】

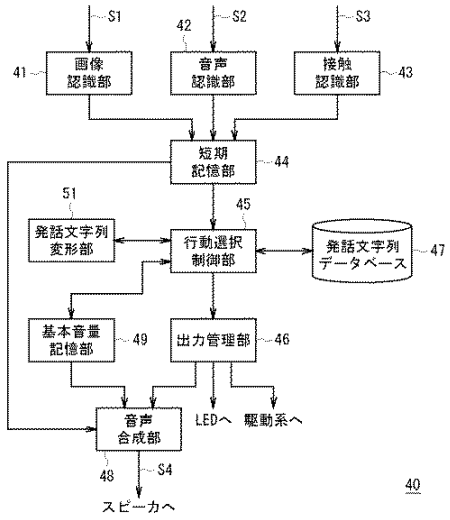


図４　本実施の形態によるロボットにおける行動制御システムの構成

【図５】

人物ID	基本音量
1	3
2	3
3	5

図５　ユーザごとの基準音量

【図６】

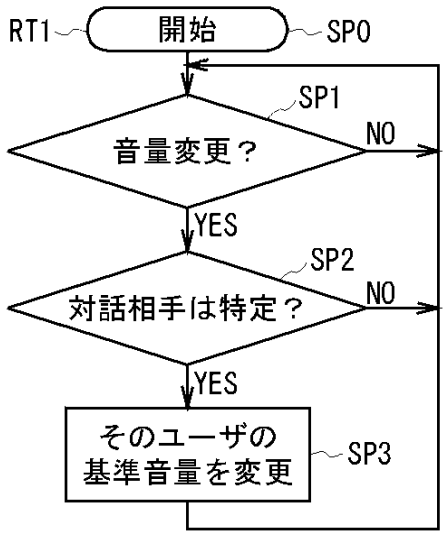


図６　基準音量変更処理手順

【図７】

350cm未満	350cm以上
なし	実施

50

図７　文字列変形判断テーブル

【図８】

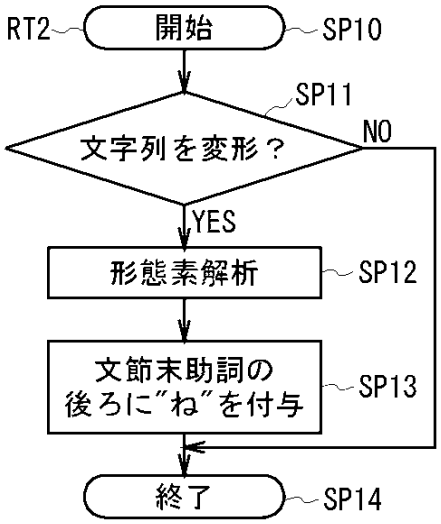


図８　文字列変形処理手順

【図９】

50cm未満	50～80cm	80～150cm	150～250cm	250以上
0	+1	+2	+3	+5

52

図９　音量変更量テーブル

【図１０】

200cm未満	200～350cm	350cm以上
0%	+20%	+50%

53

図１０　スピード変更量テーブル

【図１１】

100cm未満	100～350cm	350cm以上
0%	+30%	+60%

54

図１１　ポーズ長変更量テーブル

【図１２】

200cm未満	200以上
なし	上げる

55

図１２　文節末強調判断テーブル

【図１３】

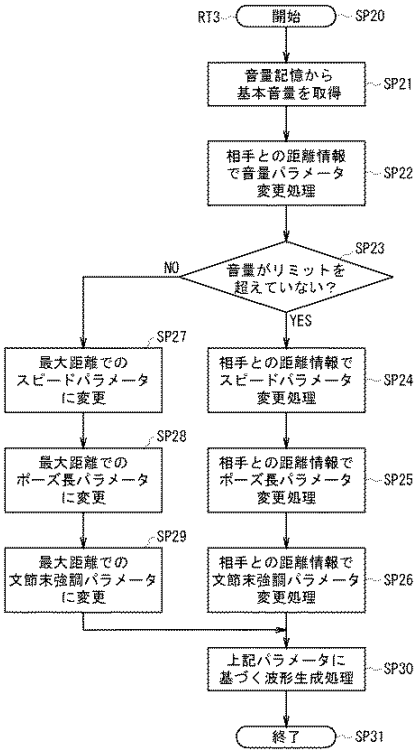


図１３　発話変更処理手順

【図１】

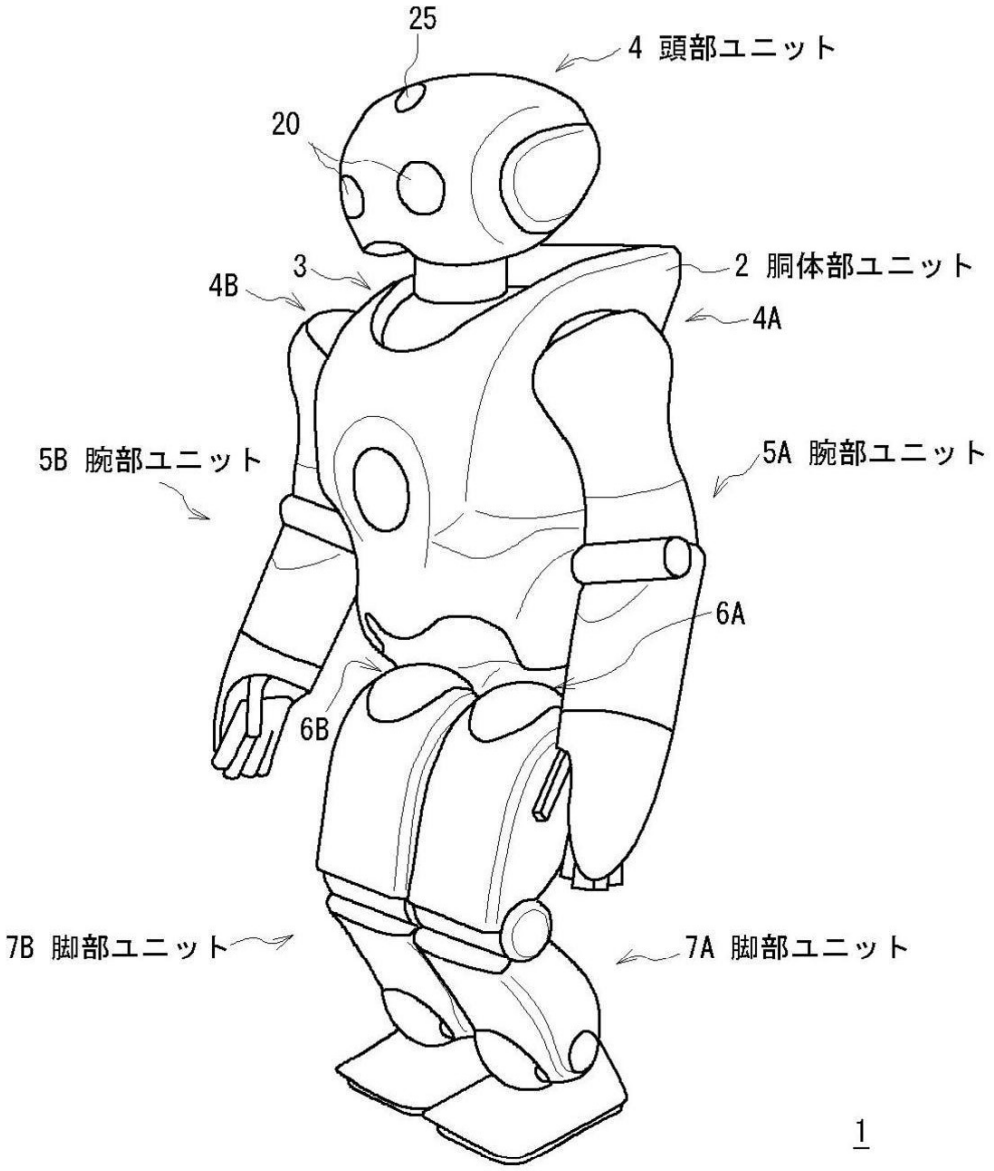


図１　本実施の形態によるロボットのハードウェア構成（１）

EXHIBIT B

(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

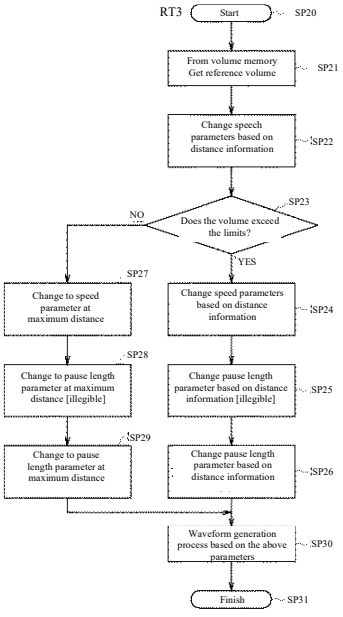


FIG. 13 Speech change processing procedure

[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]

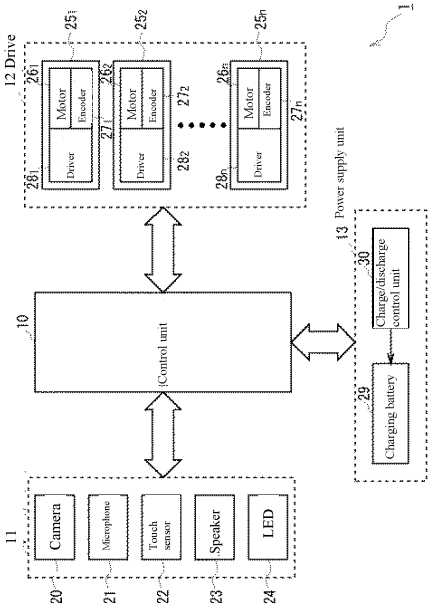


FIG. 2 Hardware constitution of the robot according to this embodiment (2)

[FIG. 3]

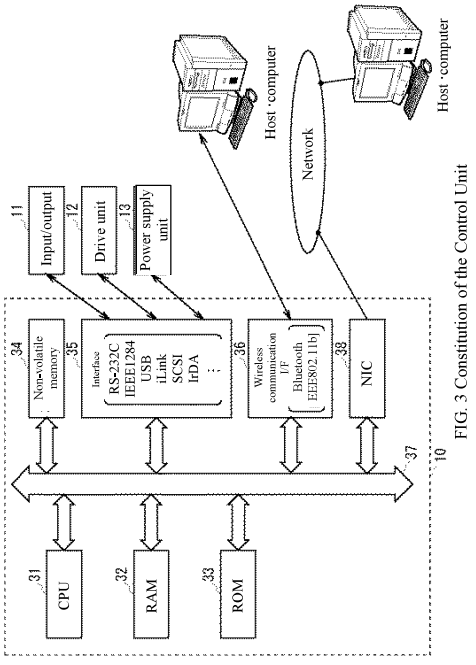


FIG. 3 Constitution of the Control Unit

[FIG. 4]

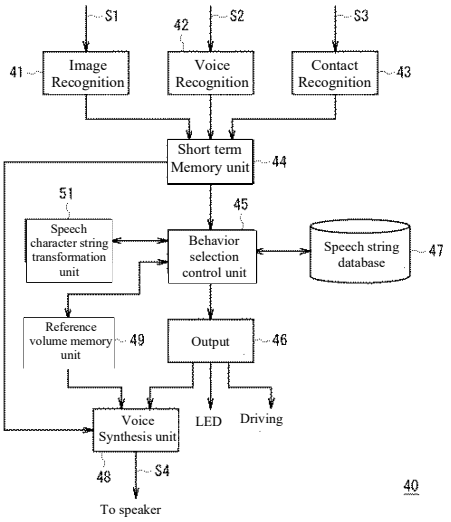


FIG. 4 Constitution of a behavior control system in a robot in accordance with the present embodiment

[FIG. 5]

Person ID	Reference volume
1	3
2	3
3	5

FIG. 5 Reference volume per user

[FIG. 6]

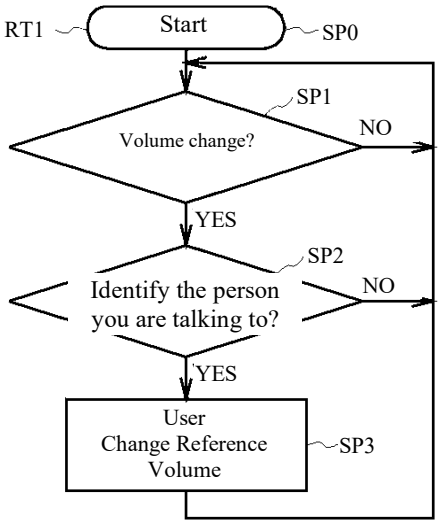


FIG. 6 Reference volume change processing procedure

[FIG. 7]

Less than 350 cm	350 cm or more
None	Implementation

FIG. 7 Character string transformation determination table

[FIG. 8]

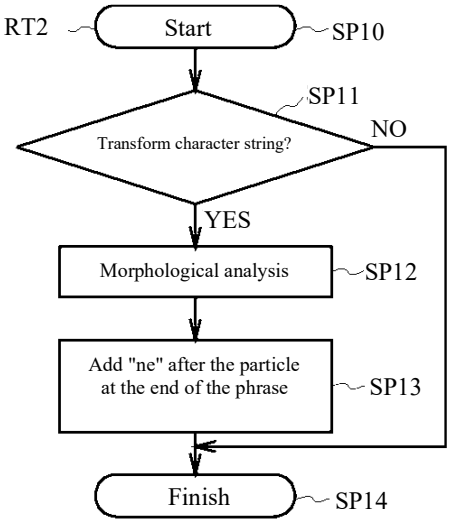


FIG. 8 Character string transformation processing procedure

[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]

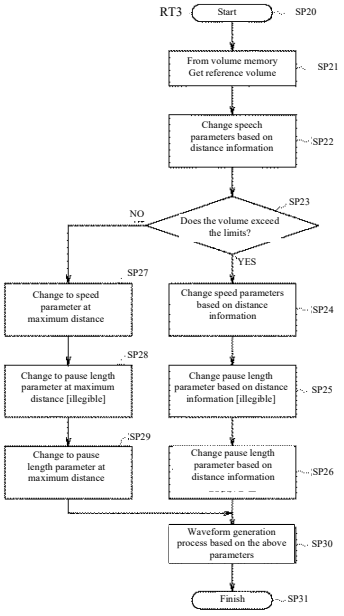


FIG. 13 Speech change processing procedure

[FIG. 1]

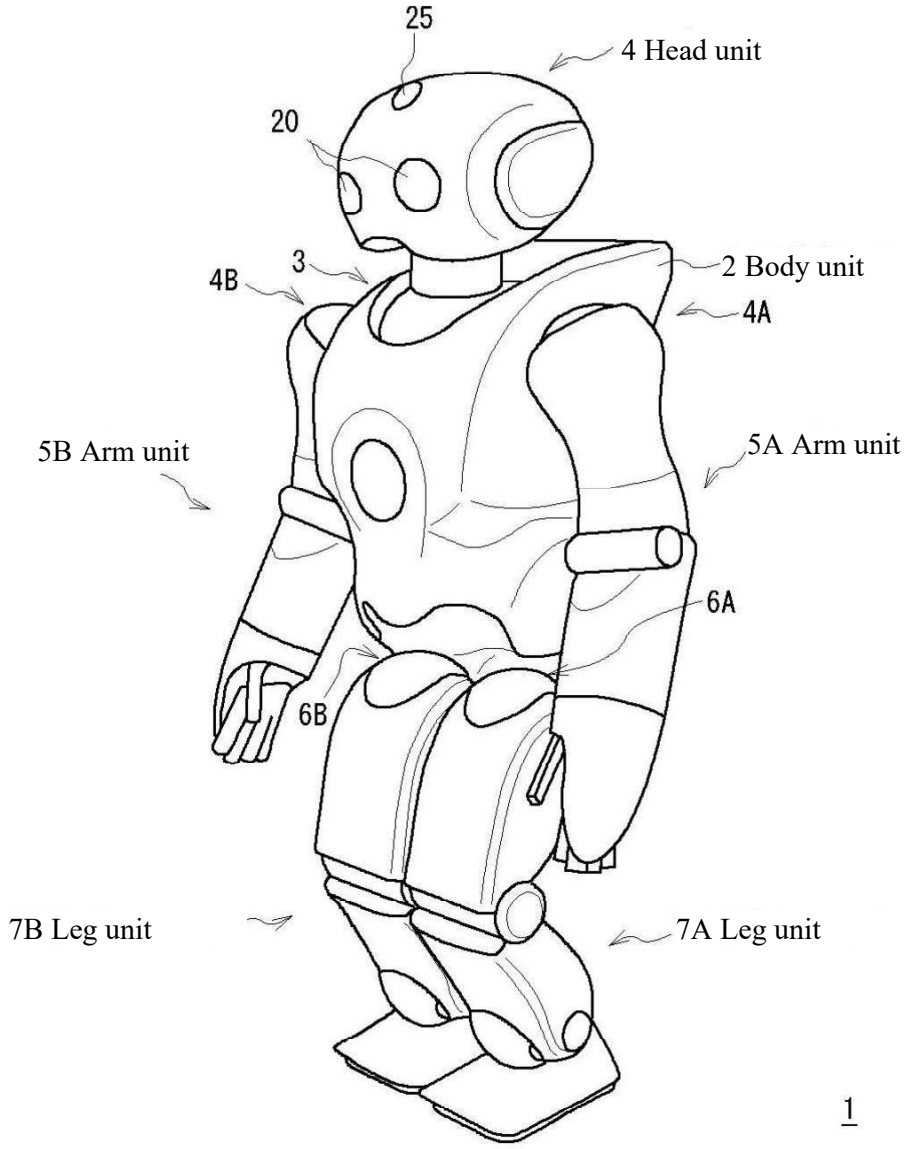


FIG. 1 Hardware constitution of the robot according to this embodiment (1)