

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6705410号
(P6705410)

(45) 発行日 令和2年6月3日(2020.6.3)

(24) 登録日 令和2年5月18日(2020.5.18)

(51) Int.Cl.		F I			
G 1 0 L	15/22	(2006.01)	G 1 0 L	15/22	2 0 0 V
G 1 0 L	15/24	(2013.01)	G 1 0 L	15/24	Z
G 1 0 L	15/10	(2006.01)	G 1 0 L	15/10	3 0 0 G
G 1 0 L	15/00	(2013.01)	G 1 0 L	15/00	2 0 0 H

請求項の数 16 (全 22 頁)

(21) 出願番号	特願2017-61357 (P2017-61357)	(73) 特許権者	000001443
(22) 出願日	平成29年3月27日 (2017.3.27)		カシオ計算機株式会社
(65) 公開番号	特開2018-163313 (P2018-163313A)		東京都渋谷区本町1丁目6番2号
(43) 公開日	平成30年10月18日 (2018.10.18)	(74) 代理人	100095407
審査請求日	平成30年11月2日 (2018.11.2)		弁理士 木村 満
		(72) 発明者	中込 浩一
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内
		(72) 発明者	山谷 崇史
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内
		(72) 発明者	島田 敬輔
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内

最終頁に続く

(54) 【発明の名称】 音声認識装置、音声認識方法、プログラム及びロボット

(57) 【特許請求の範囲】

【請求項1】

音源までの距離を判別する距離判別部と、
 音声信号を認識する音声認識部と、
 前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、を備え、
 前記認識登録語設定部は、前記距離判別部が判別した距離が第2の長さの場合における認識可能な登録語の集合が、前記判別した距離が第2の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する、
 音声認識装置。

10

【請求項2】

音源までの距離を判別する距離判別部と、
 音声信号を認識する音声認識部と、
 前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、を備え、
 前記認識登録語設定部は、前記判別した距離に基づいて、第1の登録語を認識可能に維持しつつ、該第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更する、
 音声認識装置。

【請求項3】

20

音源までの距離を判別する距離判別部と、
音声信号を認識する音声認識部と、
前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離
に応じて変更する認識登録語設定部と、を備え、

前記認識登録語設定部は、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、

前記音声認識部は、音声信号を、各登録語に付されている重みに従って認識する、
音声認識装置。

【請求項 4】

前記認識登録語設定部は、認識可能な登録語と認識不可な登録語とを特定し、前記距離判別部で判別された距離に応じて認識可能な登録語の数を変更し、

前記音声認識部は、音声信号を、前記認識可能に設定された登録語の設定で認識する、
請求項 1 から 3 のいずれか 1 項に記載の音声認識装置。

【請求項 5】

音声信号から特徴量を求める特徴量抽出手段をさらに備え、

前記音声認識部は、前記特徴量抽出手段で抽出された特徴量と、前記認識されうる登録語の音声データとに基づいて、音声信号を認識する、
請求項 1 から 4 のいずれか 1 項に記載の音声認識装置。

【請求項 6】

前記特徴量抽出手段で求めた特徴量が特徴ベクトル列情報を含む、
請求項 5 に記載の音声認識装置。

【請求項 7】

前記認識されうる登録語それぞれが、前記特徴量抽出手段で求めた特徴量に含まれる特徴ベクトル列を出力する尤度を算出する尤度算出手段をさらに備え、

前記音声認識部は、算出された尤度に基づいて、音声信号の認識を行う、
請求項 6 に記載の音声認識装置。

【請求項 8】

前記尤度算出手段は、遷移元の状態から遷移先の状態への遷移確率と、状態が遷移するときその遷移先の状態で前記特徴ベクトル列に係る各特徴ベクトルが出力される確率とを、前記特徴ベクトル列にわたって乗算を繰り返して得られる累積値を求める、
請求項 7 に記載の音声認識装置。

【請求項 9】

請求項 1 から 8 のいずれか 1 項に記載の音声認識装置を備え、

前記音声認識装置の音声認識の結果に基づいて動作する、
ロボット。

【請求項 10】

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、

前記認識登録語設定工程では、前記距離判別工程で判別した距離が第 2 の長さの場合における認識可能な登録語の集合が、前記判別した距離が第 2 の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する、

音声認識方法。

【請求項 11】

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、

前記認識登録語設定工程では、前記判別した距離に基づいて、第 1 の登録語を認識可能

10

20

30

40

50

に維持しつつ、該第 1 の登録語に音響的に類似する第 2 の登録語を認識可能な登録語から認識不可の登録語に変更する、

音声認識方法。

【請求項 1 2】

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、

前記認識登録語設定工程では、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、

前記音声認識工程では、音声信号を、各登録語に付されている重みに従って認識する、音声認識方法。

【請求項 1 3】

コンピュータに、

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、

を実行させ、

前記認識登録語設定工程では、前記距離判別工程で判別した距離が第 2 の長さの場合における認識可能な登録語の集合が、前記判別した距離が第 2 の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する、

プログラム。

【請求項 1 4】

コンピュータに、

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、

を実行させ、

前記認識登録語設定工程では、前記判別した距離に基づいて、第 1 の登録語を認識可能に維持しつつ、該第 1 の登録語に音響的に類似する第 2 の登録語を認識可能な登録語から認識不可の登録語に変更する、

プログラム。

【請求項 1 5】

コンピュータに、

音源までの距離を判別する距離判別工程と、

音声信号を認識処理する音声認識工程と、

前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、

を実行させ、

前記認識登録語設定工程では、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、

前記音声認識工程では、音声信号を、各登録語に付されている重みに従って認識する、プログラム。

【請求項 1 6】

前記音声認識工程では、複数の音声認識エンジンのうちの 1 つを、前記距離判別工程にて判別した距離に基づいて選択し、前記選択した音声認識エンジンによる音声信号の認識処理をコンピュータに実行させる、

請求項 1 3 から 1 5 のいずれか 1 項に記載のプログラム。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音声認識装置、音声認識方法、プログラム及びロボットに関する。

【背景技術】

【0002】

音声認識の際、背景雑音等の影響が大きいと誤認識の可能性が高くなるという問題がある。この問題を回避するため、特許文献1に開示された音声認識技術は、認識対象の各音声に複数の標準パターンを用意しておき、音声が入力されたときに、標準パターンそれぞれに対するスコア付けを行い、スコアに基づいて認識する。この音声認識技術は、背景雑音等の影響の大きさ等に基づいて、スコアを変更し、背景雑音等の影響による誤認識の可能性を低減する。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開平11-161292号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

HMM (Hidden Markov Model) などの一般的な音響モデルを用いた音声認識技術において、各単語の認識率は、その単語と音響的に類似する単語が認識対象として登録されている場合には、類似単語が登録されていない場合より低くなる。特に、音声認識装置と音源との距離が離れていると音声認識装置に到達する音声の音圧レベルが低下してしまうため、認識率は更に低くなる。また、音声認識装置と音源との距離が近すぎる場合には、音声信号が音割れを起こし、認識率は低くなってしまう。

20

【0005】

話者などの音声発生源と音声認識装置とが離れていると、音声認識装置に到達する音声の音圧レベルが低下してしまう。この場合、特許文献1に開示された音声認識技術では、複数の標準パターン全てに対するスコアが低下してしまうため、高い認識率を確保できなくなる。話者が近すぎて音割れが発生している場合にも、認識対象の音声自体の品質が低い

30

【0006】

本発明は上記事情に鑑み、音源までの距離の遠近によらず、高い認識率で音声を認識することができる音声認識装置、音声認識方法、プログラム及びロボットを提供することを目的とする。

【課題を解決するための手段】

【0007】

上記目的を達成するために、本発明に係る音声認識装置は、音源までの距離を判別する距離判別部と、音声信号を認識する音声認識部と、前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、を備え、前記認識登録語設定部は、前記距離判別部が判別した距離が第2の長さの場合における認識可能な登録語の集合が、前記判別した距離が第2の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する。

40

上記目的を達成するために、本発明に係る音声認識装置は、音源までの距離を判別する距離判別部と、音声信号を認識する音声認識部と、前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、を備え、前記認識登録語設定部は、前記判別した距離に基づいて、第1の登録語を認識可能に維持しつつ、該第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更する。

上記目的を達成するために、本発明に係る音声認識装置は、音源までの距離を判別する

50

距離判別部と、音声信号を認識する音声認識部と、前記音声認識部により認識されうる登録語の設定を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、を備え、前記認識登録語設定部は、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、前記音声認識部は、音声信号を、各登録語に付されている重みに従って認識する。

【0008】

上記目的を達成するために、本発明に係る音声認識方法は、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、前記認識登録語設定工程では、前記距離判別工程で判別した距離が第2の長さの場合における認識可能な登録語の集合が、前記判別した距離が第2の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する。

10

上記目的を達成するために、本発明に係る音声認識方法は、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、前記認識登録語設定工程では、前記判別した距離に基づいて、第1の登録語を認識可能に維持しつつ、該第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更する。

上記目的を達成するために、本発明に係る音声認識方法は、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を備え、前記認識登録語設定工程では、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、前記音声認識工程では、音声信号を、各登録語に付されている重みに従って認識する。

20

【0009】

上記目的を達成するために、本発明に係るロボットは、本発明に係る音声認識装置を備え、前記音声認識装置の音声認識の結果に基づいて動作する、ことを特徴とする。

【0010】

上記目的を達成するために、本発明に係るプログラムは、コンピュータに、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を実行させ、前記認識登録語設定工程では、前記距離判別工程で判別した距離が第2の長さの場合における認識可能な登録語の集合が、前記判別した距離が第2の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する。

30

上記目的を達成するために、本発明に係るプログラムは、コンピュータに、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を実行させ、前記認識登録語設定工程では、前記判別した距離に基づいて、第1の登録語を認識可能に維持しつつ、該第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更する。

40

上記目的を達成するために、本発明に係るプログラムは、コンピュータに、音源までの距離を判別する距離判別工程と、音声信号を認識処理する音声認識工程と、前記音声認識工程で認識されうる登録語の設定を、前記距離判別工程で判別された距離に応じて変更する認識登録語設定工程と、を実行させ、前記認識登録語設定工程では、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、前記音声認識工程では、音声信号を、各登録語に付されている重みに従って認識する。

【発明の効果】

【0011】

50

本発明によれば、音源との距離の遠近によらず高い音声認識率を確保することができる。

【図面の簡単な説明】

【 0 0 1 2 】

【図 1】本発明の実施の形態 1 に係るロボットの外觀図である。

【図 2】本発明の実施の形態 1 に係るロボットのハードウェア構成図である。

【図 3】本発明の実施の形態 1 に係る、音声認識装置としてのロボットのブロック図である。

【図 4】(a) は本発明の実施の形態 1 に係る認識対象テーブル内の登録語と距離パラメータを説明する図である。(b) は本発明の実施の形態 1 に係る認識対象テーブル内の登録語と音響モデルを説明する図である。

10

【図 5】本発明の実施の形態 1 に係る音声認識装置で使用される音響モデル (H M M) の一例を説明する図である。

【図 6】本発明の実施の形態 1 に係る応答データテーブルを説明する図である。

【図 7】本発明の実施の形態 1 に係る音声データ記憶処理のフローチャートである。

【図 8】本発明の実施の形態 1 に係る音源の位置判別処理のフローチャートである。

【図 9】本発明の実施の形態 1 に係る音声認識処理のフローチャートである。

【図 1 0】本発明の実施の形態 1 に係る対応動作実行処理のフローチャートである。

【図 1 1】本発明の実施の形態 1 に係る特徴量抽出を説明する図である。

【図 1 2】本発明の実施の形態 2 に係る入力音声信号の音割れを説明する図である。

20

【図 1 3】本発明の実施の形態 2 に係る音声認識装置で使用される認識対象除外テーブルを説明する図である。

【図 1 4】(a) は本発明の実施の形態 3 に係る音声認識装置で使用される認識対象テーブル内でグループ分けされた登録語を説明する図である。(b) は本発明の実施の形態 3 に係る音声認識装置で使用される認識対象テーブル内でグループごとに設定された距離ごとの重み付けを説明する図である。

【図 1 5】ロボットの周辺の音響環境に応じて登録語の認識のし易さを変更する例を説明するための図である。

【図 1 6】距離に応じて音声認識エンジンを切り替える音声認識プログラムの例を説明するための図である。

30

【発明を実施するための形態】

【 0 0 1 3 】

以下、本発明の音声認識装置及び方法、さらに、該音声認識装置及び方法をコマンド入力手段として使用するロボットの実施の形態について、図面を参照して説明する。

【 0 0 1 4 】

(実施の形態 1)

図 1 は、実施の形態 1 に係る人型のロボット 1 を正面から見た外觀図である。

【 0 0 1 5 】

ロボット 1 は、例えば、住宅内を移動し、また、住人と会話したりする。あるいは、ロボット 1 は、イベント会場内を移動、来場者の問い合わせに回答して、来場者にそのイベントの情報を音声により提供したり、来場者を案内したりする。

40

【 0 0 1 6 】

ロボット 1 は、音声認識機能を備え、ユーザから提供された音声を認識し、それをコマンドとして対応する動作を行う。ロボット 1 は、図 1 に示すように頭 2 と胴体 3 とを備える。

【 0 0 1 7 】

ロボット 1 の頭 2 には、マイク 2 1 と、カメラ 2 2 と、スピーカ 2 3 と、センサ群 2 4 と、が配置されている。

【 0 0 1 8 】

マイク 2 1 は、複数個配置されており、アレーマイクを構成し、周囲の音声を入力する

50

音声入力部として機能する。

【0019】

カメラ22は、鼻の位置に設けられており、後述する制御部100の制御の下、撮像を行う。

【0020】

スピーカ23は、口の位置に設けられており、後述する制御部100の制御の下、各種の音声を出力する。

【0021】

センサ群24は、目と耳の位置に設けられており、距測計、加速度センサ、障害物検知センサ等を含み、各種物理量を検出する。

10

【0022】

頭2と胴体3とは、破線で示される首関節31によって相互に連結されている。首関節31は、複数のモータを含み、後述する制御部100がこれら複数のモータを駆動することにより、ロボット1の頭2が、ピッチ方向、ロール方向、ヨー方向に回転する。

【0023】

足回り部32は、胴体3の下部に設けられた車輪（ホイール）と駆動モータとを含む。車輪は、2つの前輪と2つの後輪を含む。車輪は、例えば、オムニホイール、メカナムホイールから構成される。後述の制御部100は、モータを駆動することにより車輪を回転させ、ロボット1を移動させる。

【0024】

20

次に、図2を参照してロボット1のハードウェア構成を説明する。

【0025】

図2に示すように、ロボット1は、上述のマイク21、カメラ22、スピーカ23、センサ群24、首関節31、足回り部32に加えて、操作ボタン33と、記憶部101と、制御部100と、を備える。

【0026】

操作ボタン33は、胴体3の背中位置に設けられた操作ボタンであり、電源ボタン、音量調節ボタン等を含む。

【0027】

記憶部101は、フラッシュメモリ等を含み、後述の制御部100によって実行されるプログラム、各種データ等を記憶する。具体的には、記憶部101は、距離判別プログラム102、音声認識プログラム103、動作制御プログラム104、認識対象テーブル105、応答データテーブル106を記憶する。

30

【0028】

距離判別プログラム102は、複数のマイク21で取得された音声信号から、MUSIC (MULTIPLE Signal Classification) 法等の波源推定法により、音源の方向と距離とを求めるプログラムである。この実施の形態では、音源はロボット1のユーザである。

【0029】

音声認識プログラム103は、マイク21で取得された音声信号を、認識対象テーブル105を参照して認識し、テキストデータに変換するプログラムである。認識結果に相当するテキストデータは、ロボット1への一種のコマンドとなる。音声認識プログラム103は、音源までの距離が遠くなるに従って、認識対象テーブル105の使用範囲、即ち、認識対象の登録語の数を減少させるという特徴的な処理を併せて行う。音声認識処理の詳細は、図9を参照して後述する。

40

【0030】

動作制御プログラム104は、応答データテーブル106を参照して、音声認識プログラム103から出力された認識結果に相当するテキストデータに対応する制御を、スピーカ23と首関節31及び足回り部32のモータに対して実行することにより、ロボット1にユーザの呼びかけに応答する動作を実行させる。

【0031】

50

認識対象テーブル105は、音声認識プログラム103により認識される登録語を、音源までの距離に応じて変更する、認識登録語設定部としての機能を有する部分であり、音声認識プログラム103で認識する対象となる登録語（認識可能な登録語）が距離別に定義されている。なお、認識対象とする各登録語は、音声認識の単位としての語であり、文法上の最小の言語単位としての単語のみではなく語、句、語句、フレーズなどを広く含む。図4(a)に示すように、各登録語は、音源までの距離に対応付けられており、音源までの距離が長くなるに従って、段階的に、認識対象の登録語（認識可能な登録語）が減少し、認識不可な登録語が増加するように設定されている。例えば、図4(a)の例では、音源までの距離Lが、10mより大きければ、認識対象の登録語を0として、音声認識を行わない、また、音源までの距離Lが5.0mより大きく10m以下であれば、「おーい」10、「ロボット1の名前」、「こっち来て」の3つの登録語のみを認識の対象とする。また、音源までの距離Lが、1.5mより大きく5.0m以下では、直接対応付けられている「ねえねえ」等の4つの登録語と、5.0mより大きく10m以下に対応付けられている「おーい」等の3つの登録語の計7つの登録語が認識の対象となる。同様に、音源までの距離Lが、0.5m以下では、登録されている全ての登録語が認識対象となる。

【0032】

別の観点から説明すると、認識対象テーブル105は、音源までの距離が第1の長さの場合における認識可能な登録語の集合が、音源までの距離が第1の長さより短い場合における認識可能な登録語の集合に含まれるように認識可能範囲が設定されている。

【0033】

換言すると、認識対象テーブル105は、音声認識プログラム103が、入力音声を確認する登録語を距離に応じて変更する。或いは、認識対象テーブル105は、音源までの距離が長くなるに従って、音声認識プログラム103が、入力音声を認識する登録語を少なくする。或いは、認識対象テーブル105は、音源までの距離に応じて、音声認識プログラム103が、入力音声を認識する登録語の認識のし易さを音源までの距離に応じて変更する。

【0034】

さらに、認識対象テーブル105には、図4(b)に示すように、各登録語にその音響モデルのデータがリンクされている。ここでは、音響モデルはHMMであるとする。HMMは、図5に例示するように、音声がある遷移元の状態から次の遷移先の状態に遷移する確率を表す周知のモデルである。なお、図5においては、 i と j と t を自然数とすると、 S_i は各状態、 a_{ij} は状態 S_i から状態 S_j へ遷移する確率、 o_t は状態遷移時に所定の確率で出力される特徴ベクトル、 $b_j(o_t)$ は状態 S_j に遷移したとき o_t が出力される確率を表す。音響モデル内のこれらのパラメータは、あらかじめ登録語を音声学習することにより設定されている。

【0035】

図2に示す応答データテーブル106は、図6に例示するように、音声認識プログラム103による認識結果と、その認識結果が得られたときに制御部100が実行する処理とを定義するデータである。なお、認識結果は、認識対象テーブル105に認識対象として登録されている登録語のいずれかである。

【0036】

制御部100は、CPU(Central Processing Unit)、RAM(Random Access Memory)等を備え、ロボット1の上述の各部に接続されている。制御部100は、記憶部101に記憶されている距離判別プログラム102と音声認識プログラム103と動作制御プログラム104とを実行することにより、マイク21からの音声データを処理し、スピーカ23と、首関節31及び足回り部32内のモータを駆動することにより、ロボット1に音声への反応或いは応答に相当する動作を実行させる。制御部100の動作の詳細については後述する。

【0037】

制御部100は、記憶部101に記憶された距離判別プログラム102と音声認識プロ

10

20

30

40

50

グラム１０３と動作制御プログラム１０４とを実行することにより、機能的に、図３に示す距離判別部２０１と音声認識部２０２と駆動部２０８として機能する。

【００３８】

距離判別部２０１は、複数のマイク２１から入力された音声信号を処理し、ロボット１と音源との距離を判別し、判別した距離を示す距離データを、音声認識部２０２内の尤度算出部２０７に出力する。この機能は距離判別プログラム１０２と制御部１００によって実現される。

【００３９】

音声認識部２０２は特徴量抽出部２０３と尤度算出部２０７を備える。特徴量抽出部２０３は、記憶部１０１に記憶された音声信号から特徴量を抽出し、抽出した特徴量を特徴ベクトル列として尤度算出部２０７に出力する。

10

【００４０】

尤度算出部２０７は、特徴量抽出部２０３から取得した特徴ベクトル列と、距離判別部２０１から取得した距離データと、認識対象テーブル１０５に登録されている登録語に基づいて音声認識を行い、認識結果を出力する。より詳細には、尤度算出部２０７は、i) 距離データが示す音源までの距離に基づいて、認識対象テーブル１０５に登録されている登録語のうちの認識の対象とする登録語を特定し、ii) 特定した登録語の音響データと特徴量抽出部２０３から取得した特徴ベクトル列とを比較することにより、iii) 各登録語の尤度を求め、iv) 尤度が最も高く且つ閾値以上の登録語を認識結果として出力する。音声認識部２０２内のこれらの機能は音声認識プログラム１０３と制御部１００によって実現される。

20

【００４１】

駆動部２０８は、応答データテーブル１０６を参照し、音声認識部２０２が出力した認識結果を一種のコマンドとして、対応する制御を行うことにより、ロボット１にユーザに向けて音声を発しさせたりユーザの方向を向いたり、ユーザに駆け寄る等の動作を実行させる。駆動部２０８の機能は、動作制御プログラム１０４と制御部１００によって実現される。

【００４２】

次に、上記構成を有するロボット１の動作を説明する。

【００４３】

30

ロボット１の動作は、大きく分けて、i) ユーザからの音声を入力して記憶し、ii) 記憶した音声データからユーザまでの距離を求め、iii) 求めた距離に従って、認識の対象とする登録語を制限しつつ音声データを認識し、iv) 認識結果に対応する動作を行う、という動作に大別される。

【００４４】

そこで、以下、場面を分けて順に説明する。

【００４５】

まず、ユーザが発した音声ロボット１が取り込んで記憶する動作を説明する。

【００４６】

制御部１００は、電源がオンの間、図７に示す音声データ記憶処理を継続して実行している。

40

【００４７】

この処理を開始すると、制御部１００はマイク２１に音声が入力されたか否かを判別する(ステップＳ１１)。音声が入力されていないと判別された場合(ステップＳ１１：No)、ステップＳ１１に戻る。

【００４８】

一方、音声が入力されていると判別した場合(ステップＳ１１：Yes)、全てのマイク２１からの音声信号をデジタル化し(ステップＳ１２)、内部RAMに記憶し(ステップＳ１３)、ステップＳ１１に戻る。こうして、ユーザの発した音声は、デジタル音声信号の形式で内部RAMに順次格納される。

50

【 0 0 4 9 】

制御部 1 0 0 の距離判別部 2 0 1 は、電源がオンの間、図 8 に示す音源位置判別処理を実行している。

【 0 0 5 0 】

処理を開始すると、距離判別部 2 0 1 は、内部 R A M に格納されている複数のマイク 2 1 で取得された音声データを、M U S I C 法音源定位手法を用いて処理し、ユーザの位置を判別する（ステップ S 2 1 ）。

【 0 0 5 1 】

次に、距離判別部 2 0 1 は、判別した位置までの距離を示す距離データと方向を示す方向データを内部 R A M に記憶し（ステップ S 2 2 ）、ステップ S 2 1 に戻る。

10

【 0 0 5 2 】

さらに、制御部 1 0 0 の音声認識部 2 0 2 は、電源がオンの間、図 9 に示す音声認識処理を継続して実行している。

【 0 0 5 3 】

処理を開始すると、音声認識部 2 0 2 は、内部 R A M に新規の音声データが記憶されているか否かを判別する（ステップ S 3 1 ）。新規の音声データが記憶されていないと判別された場合（ステップ S 3 1 : N o ）、ステップ S 3 1 に戻る。

【 0 0 5 4 】

一方、新規の音声データが記憶されていると判別した場合（ステップ S 3 1 : Y e s ）、特徴量抽出部 2 0 3 は、複数のマイク 2 1 から得られた複数の音声信号のうちから、最も振幅の大きい音声データを特定し、特定した音声データから特徴量を抽出する（ステップ S 3 2 ）。この際、特徴量抽出部 2 0 3 は、図 1 1 に示すように、一定のフレーム幅の処理フレームを、一定のフレームシフトずつずらして音声データを取得し、取得した音声データを単位として特徴量を抽出する。特徴量として、例えば、1 2 個の値の組からなる M F C C (Mel-Frequency Cepstral Coefficient メル周波数ケプストラム係数) が用いられる。特徴量抽出部 2 0 3 は、これら 1 2 個の値の組を 1 つの 1 2 次元の特徴ベクトルとし、特徴ベクトルをフレームの時系列順に並べたものを特徴ベクトル列とする。特徴量抽出部 2 0 3 は音声信号から得られた特徴ベクトル列を尤度算出部 2 0 7 に出力する。

20

【 0 0 5 5 】

次に、尤度算出部 2 0 7 は、距離判別部 2 0 1 から音源であるユーザまでの距離を示す距離データを取得する（ステップ S 3 3 ）。そして、距離データに基づいて図 4 (a) に示す認識対象テーブル 1 0 5 を参照し、音声認識の対象とする登録語を特定する（ステップ S 3 4 ）。例えば、距離データが 7 m だった場合、尤度算出部 2 0 7 は、距離が 5 より大きく 1 0 . 0 m 以下用の 3 つの登録語のみを認識対象として特定する。一方、距離データが 1 m だった場合は、距離が 0 . 5 ~ 1 . 5 、 1 . 5 ~ 5 . 0 、 5 . 0 ~ 1 0 . 0 の登録語を認識対象として特定する。

30

【 0 0 5 6 】

次に、尤度算出部 2 0 7 は、認識対象として特定した登録語のうちの 1 つを選択する（ステップ S 3 5 ）。

【 0 0 5 7 】

40

続いて、尤度算出部 2 0 7 は、図 4 (b) に示す認識対象テーブル 1 0 5 を参照し、選択した登録語の音響モデルを求め、求めた音響モデルが、特徴量抽出部 2 0 3 から取得した特徴ベクトル o の列 O を出力する尤度を算出する（ステップ S 3 6 ）。図 5 を例にとると、特徴ベクトル列 O が

$$O = (o_1, o_2, o_3, o_4, o_5)$$

と表される場合、音響モデル M において特徴ベクトル列 O を出力する確率 $P(O, M)$ は、次の累積値で表される。

$$P(O, M) = a_{12} b_2(o_1) \times a_{22} b_2(o_2) \times a_{23} b_2(o_3) \times \dots$$

これが尤度となる。なお、 a_{ij} は状態 S_i から状態 S_j へ遷移する確率、 o_t は状態遷移時に所定の確率で出力される t 番目の特徴ベクトル、 $b_j(o_t)$ は状態 S_j に遷移し

50

たとき o_t が出力される確率を表す。

【0058】

続いて、ステップ S 3 4 で特定した登録語の全てについて、尤度の算出が終了したか否かを判別する（ステップ S 3 7）。尤度を算出していない登録語が残っていると判別すれば（ステップ S 3 7：No）、ステップ S 3 5 に戻り、未処理の登録語を 1 つ選択し、同様の処理を続行する。尤度算出をしていない登録語が残っていなければ（ステップ S 3 7：Yes）、算出された尤度を相互に比較し、最大で且つ閾値以上の尤度が計算された登録語を認識結果として出力し（ステップ S 3 8）、ステップ S 3 1 に戻り次の音声信号の入力を待機する。

【0059】

制御部 100 の駆動部 208 は、電源がオンの間、継続して図 10 に示す対応動作実行処理を行っている。

【0060】

駆動部 208 は、音声認識部 202 による認識結果の出力の有無を判別する（ステップ S 4 1）。識別結果の出力がないと判別した場合（ステップ S 4 1：No）、ステップ S 4 1 に戻る。

【0061】

一方、認識結果の出力があると判別した場合（ステップ S 4 1：Yes）、駆動部 208 は応答データテーブル 106 を参照して、認識結果に対応する動作を、スピーカ 23、首関節 31 内のモータ、足回り部 32 内のモータを駆動して行う（ステップ S 4 2）。

【0062】

以上、ロボット 1 の動作を個別に説明した。これらの動作を並行して行うことにより、ロボット 1 は、ユーザからの音声に対して、スピーカ 23 から音声で返答し、首関節 31 を駆動して顔をユーザに向け、ユーザに向かって移動をといった動作を行うことができる。ユーザの遠距離からの話しかけに対し、多くの登録語は認識できなくなるものの、遠距離から話しかけられると想定される登録語については他の登録語と誤認識する可能性が低減し、結果として背景雑音等の影響による誤認識率が低減する。こうして、背景雑音によらず高い認識率を有す音声認識装置及び方法と、高い認識率で音声コマンドを認識できるロボットが得られる。

【0063】

このように、本実施の形態でも、認識対象テーブル 105 は、音声認識プログラム 103 が入力音声を検出する登録語を距離に応じて変更する。或いは、認識対象テーブル 105 は、各登録語の音声認識部 202 での認識のし易さを、音源までの距離に応じて変更する。

【0064】

（実施の形態 2）

実施の形態 1 では、ユーザから遠くなる程、認識の対象とする登録語を限定することにより、環境ノイズの影響を低減する例を示した。この発明はこれに限定されない。例えば、ユーザからの距離が近い場合に、マイクで収集する音声が大きすぎて、図 12 に示すように、マイクの出力する音声データがクリッピングし、歪んでしまう場合がある。このような場合も、多数の登録語を認識対象として認識処理を行うと、認識率が低下してしまう。そこで、本実施形態では、ユーザまでの距離が基準以下の場合にも、認識対象とする登録語の数を制限し、誤認識を低減する。

【0065】

本実施の形態のロボット 1 は、小型のペットロボットなど、話者との距離がある程度近い状態で使用されうることを前提とする。また、記憶部 101 は、認識対象テーブル 105 と共に図 13 に例示する認識対象除外テーブル 107 を記憶する。認識対象除外テーブル 107 は、音源との距離が設定されている距離以下、例では、0.2 m 以下となったときに、登録されている、「こっちきて」等の登録語を音声認識の対象から除外することを指示するテーブルである。

10

20

30

40

50

【 0 0 6 6 】

このテーブルは、音源までの距離が第2の長さの場合における認識可能な登録語の集合が、音源までの距離が第2の長さより長い場合における認識可能な登録語の集合に含ませるように認識可能範囲が設定されている。

【 0 0 6 7 】

本実施形態では、図9に示す音声認識処理のステップS34で、距離データと図4(a)に示す認識対象テーブル105とから、音声認識の対象とする登録語が特定される。続いて、特定した登録語のうちから、距離データと図13に示す認識対象除外テーブル107で除外対象とされるものを除外して、音声認識対象の登録語と特定する。例えば、音源までの距離が0.2m以下の場合には、認識対象テーブル105によれば登録されている全ての登録語を認識の対象とするが、認識対象除外テーブル107により、「こっち来て」、等の登録語を認識対象から除外する。認識対象除外テーブルに登録されている登録語は、例えば、比較的距離が遠いときにのみ使用されると予想される登録語である。

10

【 0 0 6 8 】

なお、音源までの距離が近距離の場合に、一部の登録語を認識対象から除外する例を示したが、距離に応じて、全登録語のうちから、認識対象から除外する登録語を対応付けて登録しておいてもよい。

【 0 0 6 9 】

このように、実施の形態2においても、ユーザとロボット1との距離に応じて、認識対象の登録語を変化させることにより、音声認識率の低下を防いでいる。

20

【 0 0 7 0 】

(実施の形態3)

上記実施の形態においては、ユーザまでの距離が遠い時と近いときに、認識の対象とする登録語を制限し、その数を限定した。換言すれば、認識の対象から排除した。

【 0 0 7 1 】

しかしながら、遠い位置のロボットに、あたかも近くにいるときのような言葉を、大きな音で声をかけることがあり得る。このような場合に、ロボットがユーザを完全に無視することはかえって不自然になることがある。

【 0 0 7 2 】

そこで、本実施の形態においては、登録語を認識の対象から除外せず、代わりに、ユーザまでの距離に応じて、認識対象の登録語の認識の容易度を示す重みを変化させる。

30

【 0 0 7 3 】

なお、本実施の形態のロボット1の基本構成は、実施の形態1の構成と同様である。ただし、認識対象テーブル105、尤度算出部207には、実施の形態1のとは異なる点がある。

【 0 0 7 4 】

認識対象テーブル105は図14(a)に示すように、登録語が使用されると想定される距離ごとに各登録語がグループ分けされ、図14(b)に示すように、各グループについて距離ごとに重み付けパラメータが設定されている。たとえば中距離のグループに属する登録語である「こっち見て」について、距離Lが0.5mだった場合の重み付けパラメータは2となる。

40

【 0 0 7 5 】

尤度算出部207は各登録語について、図9のステップS36で尤度を算出し、さらに、算出した尤度と距離Lに対応する重み付けパラメータとを乗算して積を求める。以下、この積を補正尤度と呼ぶ。尤度算出部207は、この補正尤度が閾値以上で且つ最も大きい登録語を認識結果として出力する。つまり、重み付けパラメータは遠近に応じて認識すべき登録語の認識優先度を変動させ、認識可能な登録語を変更する働きをする。

【 0 0 7 6 】

より具体的に説明する。第1の例として、ユーザが7m離れたロボット1に「こっちきて」と話しかけた場合を考える。「こっちきて」は至近以外グループに属する登録語であ

50

り、中距離グループに属する「こっち見て」と音響的に類似する。距離データは7mとなるので、「こっちきて」の重み付けパラメータは1であり、「こっち見て」の重み付けパラメータは0.3である。よって「こっちきて」のほうが「こっち見て」よりも補正尤度が大きくなりやすく、誤認識を減らすことができる。

【0077】

第2の例として、同じくユーザが7m離れたロボット1に「天気おしえて」と話しかけた場合を考える。この場合、実施の形態1の場合においては「天気おしえて」を認識できる可能性はゼロである。一方、この実施の形態3の場合において、「天気おしえて」は中距離グループに属する登録語であり、距離データが7mとなるので、重み付けパラメータは0.3である。よって、背景雑音の影響が少ない状態だったり、「天気おしえて」と音響的に類似する登録語が他に存在しない状態であれば「天気おしえて」の補正尤度が最大になると考えられる。よって「天気おしえて」を認識できる。

10

【0078】

このように、登録語を、使用が想定される距離ごとにグループ分けをして重み付けをすることにより、距離に応じて誤認識の可能性を低減しつつも、ユーザから発せられた言葉が想定距離外からのものであっても認識できる可能性を残すことができる。

【0079】

以上、実施の形態3の説明について登録語を、使用が想定される距離ごとにグループ分けをして重み付けをするものとして説明したが、グループ分けをせず、各登録語について、使用が想定される距離ごとに重み付けパラメータを設定してもこの実施の形態3の目的は果たされる。

20

【0080】

本実施の形態においても、認識対象テーブル105は、音声認識プログラム103が入力音声を認識する登録語を距離に応じて変更する。或いは、認識対象テーブル105は、各登録語の音声認識部202での認識のし易さを、音源までの距離に応じて変更する。

【0081】

(実施の形態4)

上述したように、認識したい登録語と音響的に類似する登録語が認識対象の登録語として存在する場合、認識率が低下する。この問題を解決するため、距離に応じて、音響的に類似する登録語が、認識候補から除外されるように認識対象テーブルを設計することが望ましい。

30

【0082】

例えば、実施の形態1において、認識対象テーブル105が図4に示すものの場合、「こっち見て」及び「ちょっとちょっと」は距離データが5m未満の場合のみ認識候補となる。一方、距離データが5m以上の場合に使用される登録語「こっちきて」は「こっち見て」と音響的に類似するが、「ちょっとちょっと」と音響的に類似する登録語は存在しない。よって距離データが例えば7mだった場合、「こっち見て」を認識候補から除外しないと「こっちきて」の認識率が低下するが、「ちょっとちょっと」は認識候補から除外しなくとも認識率がさほど低下しない。そこで、形式的に距離に応じたグループ分けはするものの、音響的に類似するか否かで認識候補の除外を行うことにより、認識率の維持と認識可能な語数の確保を両立することができる。

40

【0083】

すなわち、認識対象テーブル105は、設計音源までの距離に基づいて、第1の登録語を認識可能に維持しつつ、この第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更するように、設定されている。

【0084】

(その他の変形例)

上記4つの実施の形態の変形例として、ロボット1の周辺の音が無音もしくはそれに近い場合、呼びかけに分類される言葉を優先的に認識するように構成してもよい。

【0085】

50

例えば、距離判別部 201 は、マイク 21 から入力される複数の音声信号から、図 15 に示すように、周囲の音源の位置を特定する。音声認識部 202 は、ロボット 1 の周辺の半径 r に音源が存在しないと判別した場合、ロボット 1 の周辺にはユーザはおらず、ユーザに話しかけられた場合は遠くから呼びかけられたものと推定できる。この点を考慮して、ロボット 1 の半径 r 内にユーザがいないと判別された場合に、認識対象テーブル 105 に登録されている登録語のうち、呼びかけに相当する登録語の認識を容易にするように設定してもよい。例えば、呼びかけ系の登録語のみを認識対象としてもよい、また、実施の形態 3 において、「至近以外」グループに属する登録語の重み付けを増加させる等してもよい。なお、ロボット 1 の周囲として半径 r の円を例示したが、周囲の形状とサイズは任意である。

10

【0086】

また、別の変形例として、音声認識プログラム 103 において複数の音声認識エンジンを使用できるようにし、距離判別部 201 が取得した距離データに基づいて複数の音声認識エンジンのうち最適なものを 1 つ選択して音声認識を行ってもよい。一般的に、ユーザとロボット 1 とが遠い場合、ユーザはロボット 1 に対して単語や短い語句を発することが多く、近い場合は長い語句を多く発する。図 4 に示す認識対象テーブル 105 はその傾向を考慮して設定されている。よって図 16 に例示するように、距離データが 1.5 m 以上の場合は単語認識用音声認識エンジンを使用し、1.5 m 未満の場合は語句認識用音声認識エンジンを使用して音声認識をすることにより、認識率の向上を図ることができる。

【0087】

20

以上、4 つの実施の形態及びそれらの変形例により本発明の実施の形態を説明した。いずれも、ユーザとロボット 1 との距離に係る距離データと、登録語に紐付けられたパラメータに基づいて、認識すべき登録語の推定をすることにより、音声認識率の向上を図っている。実施の形態 1 及び 2 では、距離データに応じて尤度算出部 207 が使用する登録語が変わり、認識できる登録語の集合が変動するので、距離データとパラメータに基づいて認識すべき登録語の推定をするものといえる。実施の形態 3 では距離データに基づいて、尤度算出部 207 によって乗算される重み付けパラメータが変わるため距離データとパラメータに基づいて認識すべき登録語の推定をするものといえる。

【0088】

また、例えば実施の形態 1 においては認識対象の登録語を距離に応じて 4 段階に切り替えたが、特段、この段階について限定をする必要はなくたとえば 2 段階でもよいし 10 段階でもよい。また、実施の形態 1 においては 1 つの認識対象テーブルから距離に応じて認識対象を制限していたが、認識対象の登録語の切り替えは特段、1 つのテーブルに制限をかける手法には限られない。例えば、認識対象テーブルを複数備えて距離に応じて切り替えるようにしてもよい。

30

【0089】

なお、本発明の技術的範囲は、上記の実施の形態及びそれらの変形例に限られず、種々の変形が可能である。例えば距離の判別について、マイク 21 からの音声信号を位置同定処理によって処理して音源の距離を求めるだけでなく、カメラ 22 のオートフォーカス機能によって得られる焦点距離を音源までの距離として利用してもよいし、センサ群 24 の 1 つにレーザー距離計を採用し、それを利用して音源までの距離を求めてもよい。なお、ユーザとロボットとの距離の判別は厳密である必要はなく、距離データに応じて登録語にパラメータを付与できる程度であればよい。

40

【0090】

以上、ロボットの入力装置として音声認識装置及び方法を利用する例を説明したが、この発明の音声認識装置及び方法は、他の装置から独立して使用される音声認識装置及び方法でもよく、また、ロボット以外の他の機械装置、コンピュータ、通信装置などの入力装置として利用することも可能である。

【0091】

以上、本発明の好ましい実施形態について説明したが、本発明に係る特定の実施形態に

50

限定されるものではなく、本発明には、特許請求の範囲に記載された発明とその均等の範囲とが含まれる。以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

【0092】

(付記)

(付記1)

音源までの距離を判別する距離判別部と、

音声信号を認識する音声認識部と、

前記音声認識部により認識されうる登録語を、前記距離判別部で判別された距離に応じて変更する認識登録語設定部と、

を備える音声認識装置。

10

【0093】

(付記2)

前記認識登録語設定部は、前記距離判別部で判別された距離が長くなるに従って、認識される登録語が減少するように、前記音声認識部により認識される登録語を変更する、ことを特徴とする付記1に記載の音声認識装置。

【0094】

(付記3)

前記認識登録語設定部は、認識可能な登録語と認識不可な登録語とを特定し、前記距離判別部で判別された距離に応じて認識可能な登録語の数を変更し、

20

前記音声認識部は、音声信号を、前記認識可能に設定された登録語で認識する、ことを特徴とする付記1または2に記載の音声認識装置。

【0095】

(付記4)

前記認識登録語設定部は、前記距離判別部が判別した距離が第1の長さの場合における認識可能な登録語の集合が、前記判別した距離が第1の長さより短い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する、ことを特徴とする付記1から3のいずれか1つに記載の音声認識装置。

【0096】

(付記5)

30

前記認識登録語設定部は、前記距離判別部が判別した距離が第2の長さの場合における認識可能な登録語の集合が、前記判別した距離が第2の長さより長い場合における認識可能な登録語の集合に含まれるように認識可能な登録語を設定する、ことを特徴とする付記1から3のいずれか1つに記載の音声認識装置。

【0097】

(付記6)

前記認識登録語設定部は、前記距離判別部が判別した距離によらず共通して認識可能な登録語を設定する、

ことを特徴とする付記1から5のいずれか1つに記載の音声認識装置。

【0098】

(付記7)

40

前記認識登録語設定部は、前記判別した距離に基づいて、第1の登録語を認識可能に維持しつつ、該第1の登録語に音響的に類似する第2の登録語を認識可能な登録語から認識不可の登録語に変更する、

ことを特徴とする付記1から6のいずれか1つに記載の音声認識装置。

【0099】

(付記8)

前記認識登録語設定部は、前記判別した距離に基づいて、認識対象の登録語に認識の容易度を示す重みを付与し、

前記音声認識部は、音声信号を、各登録語に付されている重みに従って認識する、

50

ことを特徴とする付記 1 から 7 のいずれか 1 つに記載の音声認識装置。

【0100】

(付記 9)

音声信号から特徴量を求める特徴量抽出手段をさらに備え、

前記音声認識部は、前記特徴量抽出手段で抽出された特徴量と、前記認識されうる登録語の音声データとに基づいて、音声信号を認識する、

ことを特徴とする付記 1 から 8 のいずれか 1 つに記載の音声認識装置。

【0101】

(付記 10)

前記特徴量抽出手段で求めた特徴量が特徴ベクトル列情報を含む、

ことを特徴とする付記 9 に記載の音声認識装置。

10

【0102】

(付記 11)

前記認識されうる登録語それぞれが、前記特徴量抽出手段で求めた特徴量に含まれる特徴ベクトル列を出力する尤度を算出する尤度算出手段をさらに備え、

前記音声認識部は、算出された尤度に基づいて、音声信号の認識を行う、
ことを特徴とする付記 10 に記載の音声認識装置。

【0103】

(付記 12)

前記尤度算出手段は、遷移元の状態から遷移先の状態への遷移確率と、状態が遷移するときその遷移先の状態で前記特徴ベクトル列に係る各特徴ベクトルが出力される確率とを、前記特徴ベクトル列にわたって乗算を繰り返して得られる累積値を求める、
ことを特徴とする付記 11 に記載の音声認識装置。

20

【0104】

(付記 13)

外部より音声信号を入力し、前記音声認識部に供給する音声入力部を更に備える、
ことを特徴とする付記 1 から 12 のいずれか 1 つに記載の音声認識装置。

【0105】

(付記 14)

付記 1 から 13 のいずれか 1 つに記載の音声認識装置を備え、
前記音声認識装置の音声認識の結果に基づいて動作する、
ことを特徴とするロボット。

30

【0106】

(付記 15)

音源までの距離を判別する距離判別工程と、
音声信号を認識処理する音声認識工程と、
を備え、

前記音声認識工程では、前記距離判別工程で判別した距離に対応する登録語で音声信号を認識処理する、

ことを特徴とする音声認識方法。

40

【0107】

(付記 16)

コンピュータに、
音源までの距離を判別する距離判別工程と、
音声信号を認識処理する音声認識工程と、
を実行させ、

前記音声認識工程では、コンピュータに前記距離判別工程で判別した距離に対応する登録語で音声信号を認識処理させる、
ことを特徴とするプログラム。

【0108】

50

(付記 17)

前記音声認識工程では、複数の音声認識エンジンのうちの 1 つを、前記距離判別工程にて判別した距離に基づいて選択し、前記選択した音声認識エンジンによる音声信号の認識処理をコンピュータに実行させる、ことを特徴とする付記 16 に記載のプログラム。

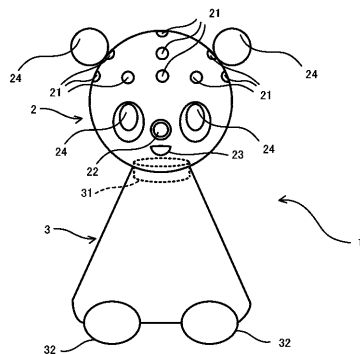
【符号の説明】

【0109】

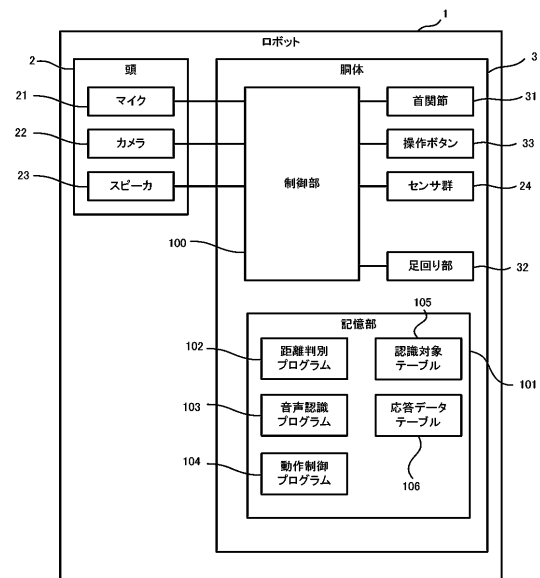
1 ロボット、2 頭、3 胴体、21 マイク、22 カメラ、23 スピーカ、24 センサ群、31 首関節、32 足回り部、33 操作ボタン、100 制御部、101 記憶部、102 距離判別プログラム、103 音声認識プログラム、104 動作制御プログラム、105 認識対象テーブル、106 応答データテーブル、107 認識対象除外テーブル、201 距離判別部、202 音声認識部、203 特徴量抽出部、207 尤度算出部、208 駆動部

10

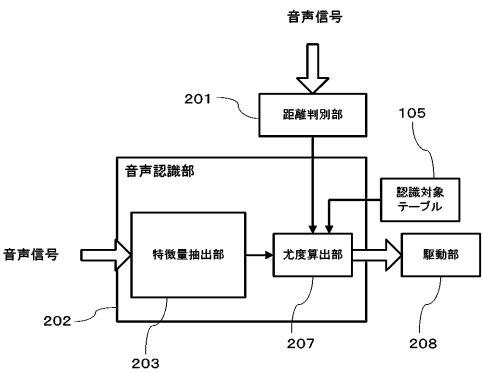
【図 1】



【図 2】



【図 3】



【図 4】

認識対象テーブル105内の登録語と距離パラメータ

距離	登録語
10.0より大	(音声認識を行わない)
5.0~10.0	おい:(ロボット1の名前);こちきて;
1.5~5.0	ねえねえ;きこえてる?;ちょっととちよつと;こちみて
0.5~1.5	天気おしえて;電気つけて;バッテリー大丈夫?;今日のニュースは? ...
0~0.5	(すべての単語を使用)

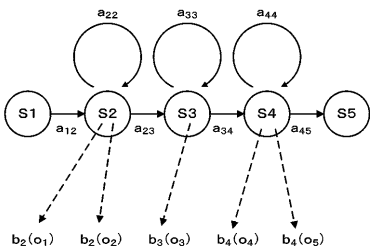
(a)

認識対象テーブル105内の登録語と音響モデル

登録語	音響モデル
おい	...
...	...

(b)

【図 5】

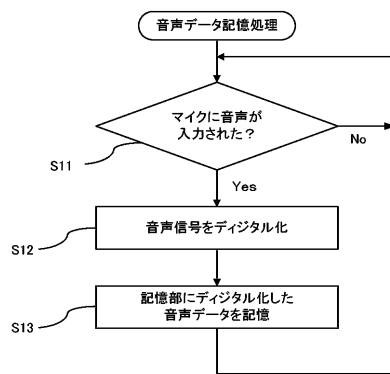


【図 6】

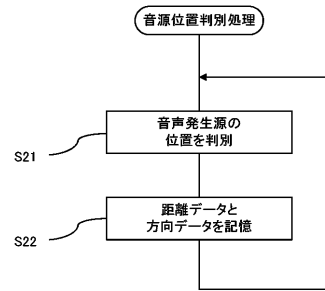
応答データテーブル106

認識結果	動作
おい	首関節内のモータを駆動回転;スピーカを駆動して返答
こちきて	足回り部内のモータを駆動させて音源の方へ移動
...	...

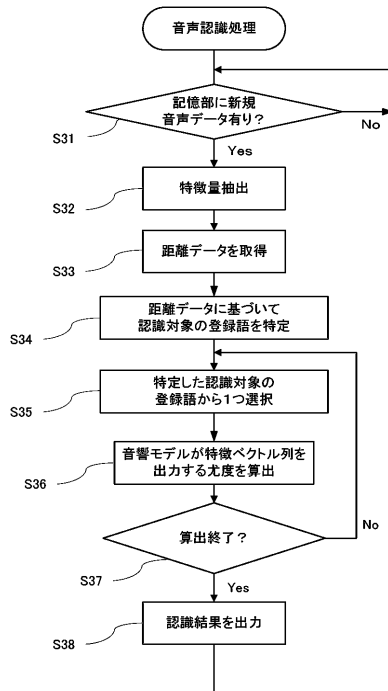
【図 7】



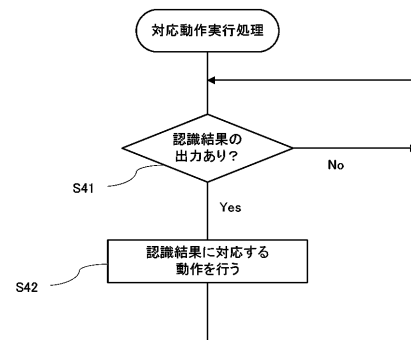
【図 8】



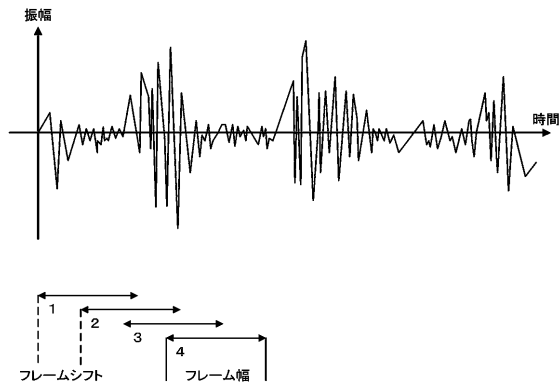
【図 9】



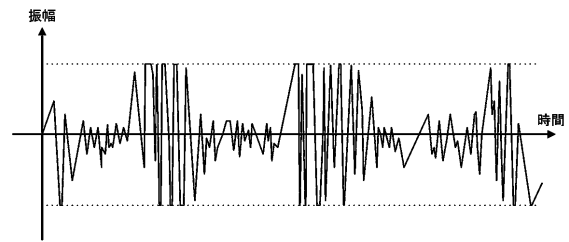
【図 10】



【図 1 1】



【図 1 2】



【図 1 3】

認識対象除外テーブル107

距離	単語
0. 2	こっちきて おいで きこえてる? ...

【図 1 4】

グループごとに分類された登録語

グループ	登録語
近距離	(具体的なやりとりに分類される単語)
中距離	天気おしえて 電気つけて こっち見て ...
全距離	おーい (ロボットの名前)くん ...
至近以外	こっちきて おいで きこえてる? ...

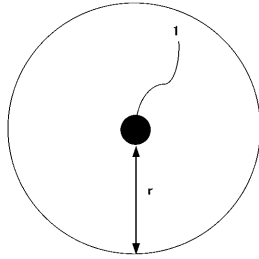
(a)

グループごとの重み付けパラメータ

グループ	$0 \leq L < 0. 2$	$0. 2 \leq L < 1. 0$	$1. 0 \leq L < 5. 0$	$5. 0 \leq L < 10. 0$
近距離	2	1	0. 3	0. 1
中距離	2	2	1	0. 3
全距離	1	1	1	1
至近以外	0. 1	1	1	1

(b)

【図 15】



【図 16】

距離パラメータと、対応して使用される音声認識エンジン

距離	音声認識エンジン
1.5~10.0	単語認識用
0~1.5	語句認識用

フロントページの続き

審査官 大野 弘

(56)参考文献 特開 2 0 1 3 - 2 5 7 4 1 8 (J P , A)
特開 2 0 0 4 - 2 2 6 6 5 6 (J P , A)
特開平 1 0 - 1 4 3 1 9 0 (J P , A)
特開 2 0 1 3 - 0 7 2 9 0 4 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)
G 1 0 L 1 5 / 2 2
G 1 0 L 1 5 / 0 0
G 1 0 L 1 5 / 1 0
G 1 0 L 1 5 / 2 4