

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6084654号
(P6084654)

(45) 発行日 平成29年2月22日 (2017.2.22)

(24) 登録日 平成29年2月3日 (2017.2.3)

| | | |
|-------------------------|---------------|---------|
| (51) Int.Cl. | F I | |
| G 1 0 L 17/04 (2013.01) | G 1 0 L 17/04 | |
| G 1 0 L 17/22 (2013.01) | G 1 0 L 17/22 | |
| G 1 0 L 17/00 (2013.01) | G 1 0 L 17/00 | 2 0 0 D |
| G 1 0 L 13/00 (2006.01) | G 1 0 L 13/00 | 1 0 0 M |
| G 0 6 F 3/16 (2006.01) | G 0 6 F 3/16 | 6 4 0 |
| 請求項の数 8 (全 23 頁) 最終頁に続く | | |

(21) 出願番号 特願2015-113949 (P2015-113949)
 (22) 出願日 平成27年6月4日 (2015.6.4)
 (65) 公開番号 特開2017-3611 (P2017-3611A)
 (43) 公開日 平成29年1月5日 (2017.1.5)
 審査請求日 平成28年2月23日 (2016.2.23)

(73) 特許権者 000005049
 シャープ株式会社
 大阪府堺市堺区匠町 1 番地
 (74) 代理人 110001195
 特許業務法人深見特許事務所
 (72) 発明者 畠山 泰貴
 大阪府大阪市阿倍野区長池町 2 2 番 2 2 号
 シャープ株式会社内
 審査官 千本 潤介

最終頁に続く

(54) 【発明の名称】 音声認識装置、音声認識システム、当該音声認識システムで使用される端末、および、話者識別モデルを生成するための方法

(57) 【特許請求の範囲】

【請求項 1】

音声認識装置であって、
 話者を識別する情報を含む発話と、話者を識別する情報を含まない発話とをそれぞれ受け付けるための音声入力部と、
 音声認識処理を行うための音声認識部と、
 音声を出力するための音声出力部と、
 前記音声認識処理の結果に基づいて前記音声認識装置を制御するための制御部とを備え

、
前記音声出力部は、話者とゲームを行うことにより得られる話者を識別する情報を含まない発話の後に、話者を識別する情報を尋ねる問い合わせを出力し、

前記制御部は、前記問い合わせの前に発せられた前記話者を識別する情報を含まない発話と、前記問い合わせに回答する発話に含まれる話者を識別する情報とを関連付けることにより、話者を識別するための話者識別モデルを生成する、音声認識装置。

【請求項 2】

前記ゲームは、しりとりゲーム、および早口言葉ゲームのうち少なくとも一方のゲームを含む、請求項 1 に記載の音声認識装置。

【請求項 3】

前記制御部は、前記音声出力部から出力される発話に対する回答の内容に基づいて、前記音声出力部から次に出力する発話の内容を決定するように構成されている、請求項 1 基

10

20

たは 2 に記載の音声認識装置。

【請求項 4】

難易度の異なる複数の問い合わせを記憶するための記憶装置をさらに備え、前記制御部は、前記音声入力部が受け付ける前記ゲームのための発話の内容に基づいて、前記音声出力部から次に出力される前記ゲームのための発話の内容を、前記記憶装置に記憶される複数の問い合わせの中から決定するように構成されている、請求項 1 ~ 3 のいずれかに記載の音声認識装置。

【請求項 5】

生成された前記話者識別モデルを格納するための記憶部をさらに備え、前記制御部は、前記問い合わせに対する応答に基づいて、前記生成された話者識別モデルを更新するように構成されている、請求項 1 ~ 4 のいずれかに記載の音声認識装置。

【請求項 6】

音声認識システムであって、
 端末と、
 前記端末と通信可能な装置とを備え、
 前記端末は、
 話者を識別する情報を含む発話と、話者を識別する情報を含まない発話とを受け付けるための音声入力部と、
 音声を出力するための音声出力部と、
 前記音声入力部および前記音声出力部に電氣的に接続されて、前記装置と通信するための通信部とを備え、

前記音声出力部は、話者とゲームを行うことにより得られる話者を識別する情報を含まない発話の後に、話者を識別する情報を尋ねる問い合わせを出力するように構成され、

前記装置は、
 前記端末と通信するための通信部と、
 音声認識処理を行うための音声認識処理部と、
 前記音声認識処理の結果に基づいて前記装置を制御するための制御部とを備え、
 前記制御部は、前記問い合わせの前に発せられた前記話者を識別する情報を含まない発話と、前記問い合わせに回答する発話に含まれる話者を識別する情報とを関連付けることにより、話者を識別するための話者識別モデルを生成する、音声認識システム。

【請求項 7】

請求項 6 に記載の音声認識システムで使用される端末。

【請求項 8】

話者識別モデルを生成するための方法であって、
ゲームを行うことにより話者を識別する情報を含まない発話を受け付けるステップと、話者を識別する情報を尋ねる問い合わせを出力するステップと、
前記問い合わせに回答する発話を受け付けるステップと、
 音声認識処理を行うステップと、
 前記音声認識処理の結果に基づいて、前記問い合わせの前に発せられた前記話者を識別する情報を含まない発話と、前記問い合わせに回答する発話に含まれる話者を識別する情報とを関連付けることにより、話者を識別するための話者識別モデルを生成するステップとを含む、方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は音声認識に関し、より特定的には、話者を識別する技術に関する。

【背景技術】

【0002】

音声認識において話者を識別する技術が知られている。たとえば、特開 2010 - 21

10

20

30

40

50

7319号公報(特許文献1)は、「音声信号から話者の特定を行う話者特定装置において、話者特定のための精度向上を図る」ための技術を開示している([要約]参照)。特開平7-261781号公報(特許文献2)は、「話者認識精度が高い話者認識のための音素モデルを作成する学習方法」を開示している([要約]参照)。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2010-217319号公報

【特許文献2】特開平7-261781号公報

【発明の概要】

10

【発明が解決しようとする課題】

【0004】

従来の音声に基づく話者識別では、話者を識別するためのモデルは予め与えられているものとし、より短いユーザ発話で効率よいモデルを構築することが目標とされている。そのため、短いながらも1分~2分程度の発話を予めユーザに要求し、得られた音声データから話者識別のモデルの確立を行っている。

【0005】

従来の技術は、ユーザに前処理としての発話を要求するものである。しかしながら、音声コミュニケーションにおける話者識別では、より自然な対話を行うために、ユーザの発話が学習のために用いられていることをユーザに感じさせることなく学習データを取得する必要がある。そのため、話者識別のモデルが構築されていない状態でユーザに負荷を強いることなく話者識別のモデルを構築するために必要な音声データを取得する必要がある。

20

【0006】

本開示は、上述のような問題点を解決するためになされたものであって、ある局面における目的は、話者識別のモデルを構築するために必要な音声データを取得できる音声認識装置を提供することである。

【0007】

他の局面における目的は、話者識別のモデルを構築するために必要な音声データを取得できる音声認識システムを提供することである。

30

【0008】

他の局面における目的は、当該音声認識システムで使用される端末を提供することである。

【0009】

さらに他の局面における目的は、話者識別のモデルを構築するために必要な話者識別モデルを生成するための方法を提供することである。

【課題を解決するための手段】

【0010】

一実施の形態に従う音声認識装置は、話者を識別する情報を含む発話と、話者を識別する情報を含まない発話とをそれぞれ受け付けるための音声入力部と、音声認識処理を行うための音声認識部と、音声を出力するための音声出力部と、音声認識処理の結果に基づいて音声認識装置を制御するための制御部とを備える。制御部は、話者を識別する情報と、話者を識別する情報を含まない発話とを関連付けることにより、話者を識別するための話者識別モデルを生成する。

40

【0011】

ある局面において、ユーザは、学習のための前処理を意識せずに、通常の音声対話を行うことのみで、学習に必要な音声データが収集され得る。

【0012】

この発明の上記および他の目的、特徴、局面および利点は、添付の図面と関連して理解されるこの発明に関する次の詳細な説明から明らかとなるであろう。

50

【図面の簡単な説明】

【0013】

【図1】しりとりゲームが行われる場合におけるユーザ1と端末2とのやり取りを表わす図である。

【図2】本開示に係る第1の実施例に従う音声認識システムの構成の概要を表す図である。

【図3】本開示に係る第2の実施例に従う音声認識システムの構成の概要を表す図である。

【図4】本開示に係る第3の実施例に従う音声認識システムの構成の概要を表す図である。

【図5】本開示に係る第4の実施例に従う音声認識システムの構成の概要を表す図である。

【図6】本開示に係る音声認識システムを実現する機能の構成を表すブロック図である。

【図7】音声認識システムにおいて保持されるデータの格納の一態様を概念的に表す図である。

【図8】ユーザ1と端末2との間の対話により話者モデル80が生成される状態を表す図である。

【図9】ユーザが発話の起点となる場合におけるシーケンスを表すフローチャート(その1)である。

【図10】ユーザが発話の起点となる場合におけるシーケンスを表すフローチャート(その2)である。

【図11】ユーザが発話の起点となる場合におけるシーケンスを表すフローチャート(その3)である。

【図12】ユーザが音声認識システムに既知である場合におけるユーザ1と端末2とのやり取りのシーケンスを表す図である。

【図13】ユーザが既知である場合に行なわれる処理の流れを表すシーケンスチャート(その1)である。

【図14】ユーザが既知である場合に行なわれる処理の流れを表すシーケンスチャート(その2)である。

【図15】端末2からユーザ1に話しかけることが対話のトリガとなる場合を表す図である。

【図16】音声認識システムで行われる処理の一部を表すシーケンスチャート(その1)である。

【図17】音声認識システムで行われる処理の一部を表すシーケンスチャート(その2)である。

【発明を実施するための形態】

【0014】

以下、図面を参照しつつ、本発明の実施の形態について説明する。以下の説明では、同一の部品には同一の符号を付してある。それらの名称および機能も同じである。したがって、それらについての詳細な説明は繰り返さない。

【0015】

<技術思想>

本開示によれば、ユーザが不定の場合に音声対話内容でユーザ名を問いかけてユーザを分類することにより話者識別のためのモデル構築に必要な音声データ(たとえば声紋情報)が収集される。例えば、「しりとり」や早口言葉等のゲームのような対話では、ゲーム相手(たとえば、端末、家電機器等)に対する複数回のユーザ発話が期待される。このような場合に、ゲーム相手となる装置は、予めユーザ名を問いかけてゲームをすることにより一連のユーザ発話を学習データとすることができる。または、ある未知のユーザが発話した後にユーザ名を問いかけることで一時刻前の未知のユーザ名を確定することができる。

10

20

30

40

50

【 0 0 1 6 】

本実施の形態では、音声認識の一例として、たとえば形態素解析が用いられる。この解析手法によれば、固有名詞とそうでないものが切り分けられる。たとえば、音声認識システムは、名前の辞書をデータベースとして有し得る。音声認識は、形態素解析において辞書と抽出された固有名詞とをマッチングすることにより行なわれる。

【 0 0 1 7 】

< 構成の概要 >

(構成1) ある局面に従う音声認識装置は、話者を識別する情報を含む発話と、話者を識別する情報を含まない発話とをそれぞれ受け付けるためのマイクと、音声認識処理を行うためのプロセッサと、音声を出力するためのスピーカと、音声認識処理の結果に基づいて音声認識装置を制御するためのプロセッサとを備える。プロセッサは、話者を識別する情報と、話者を識別する情報を含まない発話とを関連付けることにより、話者を識別するための話者識別モデルを生成する。話者識別モデルは、たとえば、話者の識別ID (Identification) と、話者(音声認識装置のユーザ)の名前と、当該話者の発話から抽出された声紋情報等を含み得る。

10

【 0 0 1 8 】

本実施の形態において、話者を識別する情報としては、たとえば、名前、あだ名、住民番号、政府機関から与えられた識別番号その他の情報であって、発話に含めることが可能な語句をいう。

【 0 0 1 9 】

(構成2) 好ましくは、スピーカは、話者を識別する情報を尋ねる問い合せを出力する。話者を識別する情報と、話者を識別する情報を含まない発話とを関連付けることは、話者を識別する情報と、問い合わせの後に発せられた話者を識別する情報を含まない発話とを関連付けることを含む。

20

【 0 0 2 0 】

(構成3) 好ましくは、スピーカは、話者を識別する情報を含まない発話の後に、話者を識別する情報を尋ねる問い合せを出力する。話者を識別する情報と、話者を識別する情報を含まない発話とを関連付けることは、問い合わせの前に発せられた発話と、問い合わせに回答する発話に含まれる話者を識別する情報とを関連付けることを含む。

【 0 0 2 1 】

(構成4) プロセッサは、スピーカから出力される発話に対する回答の内容に基づいて、スピーカから次に出力する発話の内容を決定するように構成されている。たとえば、音声認識装置は、複数の問い合わせを予め保持している。各問い合わせの難易度は、階層的に異なる。ある局面において、難易度が中位である問い合わせの発話に対して、予め定められた一定時間内に回答が返ってこない場合、あるいは、回答が正しくない場合、プロセッサは、難易度が低い問い合わせ(しりとりの問題)を発話する。別の局面において、予め定められた一定時間内に早期に回答が返ってきた場合、プロセッサは、難易度が高い問い合わせ(しりとりの問題)を次の問い合わせとして発話する。

30

【 0 0 2 2 】

(構成5) 当該音声認識装置は、生成された話者識別モデルを格納するためのメモリをさらに備える。プロセッサは、問い合わせに対する回答に基づいて、生成された話者識別モデルを更新するように構成されている。

40

【 0 0 2 3 】

(構成6) 別の局面に従うと、音声認識システムが提供される。音声認識システムは、端末と、当該端末と通信可能な装置とを備える。端末は、話者を識別する情報を含む発話と、話者を識別する情報を含まない発話とを受け付けるためのマイクと、音声を出力するためのスピーカと、マイクおよびスピーカに電氣的に接続されて、当該装置と通信するための通信インターフェイスとを備える。装置は、端末と通信するための通信インターフェイスと、音声認識処理を行うためのプロセッサと、音声認識処理の結果に基づいて装置を制御するためのプロセッサとを備える。プロセッサは、話者を識別する情報と、話者を識

50

別する情報を含まない発話とを関連付けることにより、話者を識別するための話者識別モデルを生成する。

【0024】

<技術思想の背景>

図1を参照して、本実施の形態に係る技術思想の背景について説明する。図1は、しりとりゲームが行われる場合におけるユーザ1と端末2とのやり取りを表わす図である。ユーザ1は、端末2に対して、メッセージ10を発する。端末2は、メッセージ10を認識すると、応答として、メッセージ11を発する。

【0025】

ユーザ1は、端末2に対して、メッセージ12を発する。端末2は、メッセージ12を認識すると、メッセージ12に含まれる名前と予め規定されたメッセージとを用いて合成されたメッセージ13を発する。

【0026】

予め定められた時間が経過すると、端末2は、メッセージ14を発する。ユーザ1は、メッセージ14を認識すると、予め規定された時間内に、応答として、メッセージ14に続く言葉を考える。ユーザ1が、端末2に対して、メッセージ15を発する。端末2は、メッセージ15を認識すると、予め準備された国語辞書を参照して、メッセージ15に続く言葉を考える。端末2は、予め規定された時間内に、メッセージ15に対する言葉としてメッセージ16を発する。このようにして、ユーザ1と端末2とは、しりとりゲームを続ける。

【0027】

端末2の発話に対して、ユーザ1が予め規定された時間内に次の言葉を返せる場合は、同様にしりとりが続く。たとえば、ユーザ1が端末2に対してメッセージ17を発する。端末2は、メッセージ17を認識すると、メッセージ18を発する。

【0028】

一方、ユーザ1が次の言葉を返せない場合がある。この場合、ユーザ1は沈黙を続けるか、分からない旨のメッセージ19を発することになる。端末2は、予め定められた一定の待ち時間内にユーザ1からの応答がないと判断した場合、あるいは、メッセージ19を認識した場合には、その内容について予め規定されていたメッセージ20を発する。

【0029】

このような場合、端末2は、ユーザ1との間のメッセージのやり取りを通じて、ユーザ1が「たろう」であることを認識し、ユーザ情報として「たろう」を各データに関連付ける。

【0030】

図2～図5を参照して本開示に係る音声認識システムの構成について説明する。

[端末]

図2は、本開示に係る第1の実施例に従う音声認識システムの構成の概要を表す図である。当該音声認識システムでは、ひとつの端末200が音声認識システムとして機能する。

【0031】

端末200は、制御部30と、音声入力部31と、音声出力部32と、話者識別部33と、話者識別学習部34と、ユーザ管理部35と、音声認識部36と、対話分析・生成部37とを備える。端末200は、たとえば、音声入出力機能と音声認識機能とを備える端末であればよい。当該端末は、たとえば、スマートフォン、テレビ、スタンドアロンで動作し得るお掃除ロボットその他の機器を含み得る。

【0032】

制御部30は、端末200の動作を制御する。音声入力部31は、音声の入力を受け付けて信号を制御部30に出力する。音声出力部32は、制御部30から出力された信号を音声に変換して、端末200の外部に音声を出力する。音声出力部32は、たとえばスピーカ、端子等を含む。話者識別部33は、制御部30から送られる信号に基づいて、端末

10

20

30

40

50

200に対する発話を行なった話者を識別する。別の局面において、話者識別部33は、当該信号と端末200に保存されているデータとに基づいて話者を識別する。当該データは、たとえば、端末200のユーザとして予め登録された声紋情報等を含み得る。

【0033】

話者識別学習部34は、話者識別部33により識別された話者の情報(ユーザID等)を用いて、話者毎のデータ(ユーザプロファイル)を作成する。ユーザ管理部35は、端末200のユーザ情報を保存する。ユーザ情報は、ユーザプロファイル等を含み得る。音声認識部36は、制御部30から送られる音声信号を用いて音声認識処理を実行する。たとえば、音声認識部36は、発話に含まれている文字を抽出する。

【0034】

対話分析・生成部37は、音声認識部36による認識の結果に基づいて端末200に対するメッセージを分析する。さらに、対話分析・生成部37は、分析の結果に応じて、当該発話に応じた応答を生成する。別の局面において、対話分析・生成部37は、端末200における設定に基づいて、端末200のユーザに対する働きかけのための発話を生成する。当該設定は、たとえば、端末200が、自己の近傍にユーザの存在を検知したこと、予め設定された時刻が到来したこと等を含み得る。

【0035】

[端末+サーバ]

図3は、本開示に係る第2の実施例に従う音声認識システムの構成の概要を表す図である。当該音声認識システムは、端末300と、サーバ350とを備える。端末300は、音声入力部31と、音声出力部32とを備える。端末300は、プロセッサ(図示しない)によって制御される。サーバ350は、制御部30と、話者識別部33と、話者識別学習部34と、ユーザ管理部35と、音声認識部36と、対話分析・生成部37とを備える。端末300は、たとえば、音声入出力機能と通信機能とを備える端末として実現される。そのような端末は、たとえば、携帯電話その他の情報通信端末、音声認識機能と通信機能とを備えるお掃除ロボットその他の機器等を含み得る。

【0036】

端末300は、ユーザの発話を受け付けると、その発話に応じた音声信号を、通信インターフェイス(図示しない)を介してサーバ350に送信する。サーバ350は、その音声信号を受信すると、話者識別処理、音声認識処理、対話分析、応答生成等の処理を実行する。各処理は、図2に示される構成によって実現される処理と同様なので、詳細な説明は繰り返さない。

【0037】

サーバ350は、生成された応答を通信インターフェイス(図示しない)を介して端末300に送信する。端末300がその応答を受信すると、音声出力部32は、その応答に応じた音声を出力する。

【0038】

[端末+サーバ+話者識別サーバ]

図4は、本開示に係る第3の実施例に従う音声認識システムの構成の概要を表す図である。当該音声認識システムは、端末300と、サーバ400と、話者識別サーバ410とを備える。サーバ400は、制御部30と、ユーザ管理部35と、音声認識部36と、対話分析・生成部37とを備える。話者識別サーバ410は、話者識別部33と、話者識別学習部34とを備える。

【0039】

サーバ400と話者識別サーバ410とは、公知の構成を有するコンピュータ装置によって実現される。当該コンピュータは、主たる構成要素として、プログラムを実行するCPU(Central Processing Unit)と、キーボードその他の入力装置と、RAM(Random Access Memory)と、ハードディスクと、光ディスク駆動装置と、モニタと、通信IF(Interface)とを備える。

【0040】

10

20

30

40

50

コンピュータにおける処理は、各ハードウェアおよびCPUにより実行されるソフトウェアによって実現される。ある局面において、当該ソフトウェアは、ハードディスクに予め格納されている。別の局面において、当該ソフトウェアは、CD-ROMその他のコンピュータ読み取り可能な不揮発性のデータ記録媒体に格納されてプログラム製品として流通している。さらに別の局面において、当該ソフトウェアは、インターネットその他のネットワークに接続されている情報提供事業者によってダウンロード可能なプログラム製品として提供される場合もある。

【0041】

コンピュータのハードウェア構成は、一般的なものである。したがって、サーバ400と話者識別サーバ410のハードウェア構成の説明は繰り返さない。本実施の形態に係る技術思想を実現する本質的な部分は、当該コンピュータに格納されたプログラムであるともいえる。

10

【0042】

サーバ400は、端末300から送られた音声信号を受信すると、通信インターフェイスを介して、その音声信号を話者識別サーバ410に送信する。

【0043】

話者識別サーバ410は、話者を認識し、また、話者を登録するためのデータを生成する。話者識別サーバ410は、生成したデータをサーバ400に送信する。

【0044】

[端末 + サーバ + 話者識別サーバ + 音声認識サーバ]

20

図5は、本開示に係る第4の実施例に従う音声認識システムの構成の概要を表す図である。当該音声認識システムは、端末300と、サーバ500と、話者識別サーバ410と、音声認識サーバ520とを備える。サーバ500は、制御部30と、ユーザ管理部35と、対話分析・生成部37とを備える。音声認識サーバ520は、音声認識部36を備える。

【0045】

サーバ500は、端末300から音声信号を受信すると、その音声信号を話者識別サーバ410および音声認識サーバ520に送信する。音声認識サーバ520は、当該音声信号を用いて音声認識処理を実行し、認識の結果をサーバ500に送信する。

【0046】

30

その他の動作は、前述の他の実施例に従う音声認識システムの構成における動作と同様である。したがって、他の動作の説明は繰り返さない。

【0047】

[機能構成]

図6は、本開示に係る音声認識システムを実現する機能の構成を表すブロック図である。音声認識システムは、端末モジュール600と、メインモジュール610と、話者識別モジュール620と、音声認識モジュール630とを備える。

【0048】

端末モジュール600は、音声入力部31と音声出力部32とを備える。端末モジュール600は、ユーザの近傍にあって発話を受け付けて、音声データと端末IDとをメインモジュール610に送信する。別の局面において、端末モジュール600は、メインモジュール610から送られた合成音声データを受信し、合成音声データに基づく音声を音声出力部32から出力する。

40

【0049】

メインモジュール610において、制御部30は、音声データと話者モデルリストとを話者識別モジュール620に送信する。話者識別モジュール620は、話者を識別すると、話者識別結果(たとえば、メッセージのID、話者が識別できたことを表すフラグ等)をメインモジュール610に送信する。

【0050】

制御部30は、端末IDまたは音声データをユーザ管理部35に送信する。ユーザ管理

50

部 35 は、端末 ID または音声データを保存する。

【 0051 】

制御部 30 は、ユーザ管理部 35 から話者モデルリストを読み出す。

制御部 30 は、対話分析・生成部 37 との間で、たとえば、テキストデータのやり取りを行なう。

【 0052 】

制御部 30 は、音声データを音声認識モジュール 630 に送信する。音声認識モジュール 630 は、音声データを用いて音声認識処理を実行すると、その結果をテキストとして制御部 30 に送る。

【 0053 】

図 6 に示される機能は、図 2 ~ 図 5 に示される構成のいずれかによって実現される。

[データ構造]

図 7 を参照して、本実施の形態に係る音声認識システムのデータ構造について説明する。図 7 は、音声認識システムにおいて保持されるデータの格納の一態様を概念的に表す図である。ある局面において、音声認識システムは、端末管理テーブルと、家庭管理テーブルと、ユーザ管理テーブルとを含む。

【 0054 】

(端末管理テーブル)

端末管理テーブルは、端末 ID と、所属ユーザ ID とを含む。端末 ID は、音声認識システムにおいて登録された端末を識別する。ある局面において、端末 ID は、音声認識システムの管理者（たとえば、制御部 30 を含むコンピュータの管理者）によって一意に付与される。別の局面において、端末 ID は、当該端末のユーザが希望する任意の文字列（たとえば、英数字、記号など）によって構成される。この場合、端末 ID の重複が生じないように、たとえば、制御部 30 は、ユーザによって入力された ID が既に使用されているか否かをチェックし、使用済みの端末 ID が入力された場合は、その旨を端末に通知する。所属ユーザ ID は、当該端末の使用者として登録されたユーザを識別する。端末の使用者の数は特に限られない。

【 0055 】

(家庭管理テーブル)

家庭管理テーブルは、家庭 ID と、当該家庭に所属する端末の端末 ID とを含む。家庭 ID は、音声認識システムのサービスを利用するユーザのグループとして家庭を識別する。ユーザのグループの単位は家庭に限られない。複数のユーザが一つのグループに関連付けられるものであればよい。家庭 ID には、1 つ以上の端末の各端末 ID が関連付けられている。家庭に関連付けられる端末の数は特に限られない。

【 0056 】

(ユーザ管理テーブル)

ユーザ管理テーブルは、ユーザ ID と、ユーザ名と、話者モデルデータと、音声データリストとを含む。

【 0057 】

ユーザ ID は、端末を使用するユーザを識別する。ユーザ名は、当該ユーザ ID が割り当てられたユーザを識別する。話者モデルデータは、当該ユーザを識別するためのデータである。話者モデルデータは、たとえば、声紋情報を含み得る。

【 0058 】

音声データリストは、当該ユーザを識別するための音声データを含む。当該音声データは、ユーザから端末に対する発話、端末の発話に対するユーザの応答、端末に表示された文字列のユーザによる発話等を含み得る。

【 0059 】

[話者モデルの生成]

図 8 を参照して、話者モデルの生成について説明する。図 8 は、ユーザ 1 と端末 2 との間の対話により話者モデル 80 が生成される状態を表す図である。なお、図 1 における状

10

20

30

40

50

態と同様の状態の説明は繰り返さない。

【0060】

ユーザ1と端末2との対話において、ユーザ1が未登録の場合には、端末2は、まず最初にユーザ名を聞いて、以降の一定区間(たとえば、ゲーム終了等)までをそのユーザの発話として音声データをデータベースに登録する。音声データは声紋情報を含み得る。

【0061】

ユーザ発話毎に、話者識別学習部は、対象の音声DB(Database)からこれまでの音声データ全てを学習データとして話者識別の学習を行う。

【0062】

IDが端末ごとに割り当てられる。端末とユーザ名とによってユーザを管理することにより他端末で同名のユーザがいるばあいでも対応可能となる。

10

【0063】

ユーザ1が自身の名前を発すると(メッセージ12)、端末2はメッセージ12を認識する。端末2は、メッセージ12からユーザ名(=たろう)を抽出すると、当該ユーザ名と端末2の端末IDとをユーザ管理部35に送信する。その後も、ユーザ1が発話すると、各メッセージ15、メッセージ17は、端末2を通してユーザ管理部35に蓄積される。

【0064】

話者識別学習部34は、ユーザ管理部35に保存されている端末IDとユーザ名とを読み出して、話者モデル80を生成する。話者モデル80は、当該ユーザ名と端末IDとを含む。したがって、以降は、端末2がユーザ1と対話することによりユーザ名が特定されると、当該ユーザに関連付けられた話者モデル80が利用可能となる。

20

【0065】

[制御構造]

図9~図11を参照して、本実施の形態に係る音声認識システムの制御構造について説明する。図9から図11は、それぞれ、ユーザが発話の起点となる場合におけるシーケンスを表すフローチャートである。

【0066】

ステップ910にて、ユーザによる話者識別学習用のシーケンスを開始するための発話が行なわれる。たとえば、ユーザは「しりとりにしようよ」というメッセージ911を発する。音声入力部31は、メッセージ911を受け付けると、メッセージ911に応じた音声信号を制御部30に送信する。

30

【0067】

ステップ915にて、制御部30は、当該音声信号を受信したことを検知すると、音声認識リクエストを音声認識部36に送信する。

【0068】

ステップ920にて、制御部30は、当該音声信号を受信したことを検知すると、話者モデルリスト取得リクエストをユーザ管理部35に送信する。話者モデルリスト取得リクエストは、当該発話を与えたユーザに関連付けられている話者モデルリストにアクセスすることを要求する。

40

【0069】

ステップ925にて、制御部30は、当該話者モデルリスト取得リクエストに回答して、話者モデルリストレスポンスを制御部30に送信する。話者モデルリストレスポンスは、当該ユーザに関連付けられている話者モデルリストの取得結果を含む。

【0070】

ステップ930にて、制御部30は、話者識別部33に対して、話者識別リクエストを送信する。話者識別部33は、話者識別リクエストの受信を検知すると、ユーザ管理部35に保存されているデータを参照して、ステップ910にて発話を行なったユーザ(話者)の識別を試みる。

【0071】

50

ステップ935にて、音声認識部36は、ステップ915における音声認識リクエストに
応答して、音声認識レスポンスを制御部30に送信する。音声認識レスポンスは、音声
認識が成功したか否かを含む。

【0072】

ステップ940にて、話者識別部33は、話者識別失敗レスポンスを話者識別部33に
送信する。すなわち、ユーザが音声認識システムに登録されていないため、話者識別部3
3は、当該発話を与えたユーザ（話者）を識別することができない。そこで、話者の識別
が失敗したことを通知する話者識別失敗レスポンスが、話者識別部33から制御部30に
送られる。

【0073】

ステップ945にて、制御部30は、話者識別失敗レスポンスの受信に応答して、対話
分析・生成リクエストを対話分析・生成部37に送信する。対話分析・生成リクエストは
、音声識別結果および話者識別結果を含み得る。対話分析・生成部37は、対話分析・生
成リクエストを受信すると、当該発話を与えたユーザの名前を取得するためのメッセー
ジを生成する。たとえば、対話分析・生成部37は、音声認識システムにおいて予め準備さ
れているテンプレートと、メッセージ911に含まれる用語「しりとり」とを用いて、メ
ッセージ946（しりとりをはじめよ。それじゃ、名前を教えてね。）を作成する。

【0074】

ステップ950にて、対話分析・生成部37は、生成したメッセージ946を制御部3
0に送信する。制御部30は、当該メッセージの受信を検知すると、当該発話を与えた端
末の端末IDと当該メッセージとを含む音声レスポンスを生成する。

【0075】

ステップ955にて、制御部30は、音声出力部32に対して、当該音声レスポンスを
送信する。音声出力部32は、当該音声レスポンスの信号を受信すると、当該信号に基
づく音声を出力する。ユーザが当該音声を認識すると、その音声に対する発話を行なう。そ
の発話は、音声入力部31によって受け付けられる。

【0076】

ステップ960にて、音声入力部31は、受け付けたメッセージ961（名前登録発話
）の内容を制御部30に送信する。メッセージ961は、たとえば「たろうだよ」のよう
に、メッセージ946に対する回答（名前）を含む。制御部30は、メッセージ961の
受信を検知すると、音声認識リクエストを生成する。

【0077】

ステップ965にて、制御部30は、音声認識部36に対して音声認識リクエストを送
信する。音声認識部36は、音声認識リクエストの受信を検知すると、メッセージ961
の音声認識処理を実行する。

【0078】

ステップ970にて、制御部30は、ユーザ管理部35に対して、話者モデルリスト取
得リクエストを送信する。ユーザ管理部35は、話者モデルリクエストの受信を検知する
と、話者モデルリストの取得を試みる。ユーザ管理部35は、取得を試みた結果を話者モ
デルリストレスポンスとして生成する。

【0079】

ステップ975にて、ユーザ管理部35は、制御部30に対して、話者モデルリストレ
スポンスを送信する。

【0080】

ステップ980にて、制御部30は、話者モデルリストレスポンスの受信に応答して、
話者識別リクエストを話者識別部33に送信する。話者識別部33は、話者識別リクエス
トの受信を検知すると、話者の識別を開始し、識別結果を生成する。

【0081】

図10を参照して、ステップ1010にて、音声認識部36は、話者識別リクエストに
対する応答として、音声認識レスポンスを制御部30に送信する。当該音声認識レスポ
ンス

10

20

30

40

50

スは、メッセージ 961 の内容を認識できた旨を含み得る。

【0082】

ステップ 1015 にて、話者識別部 33 は、話者識別失敗レスポンスを制御部 30 に送信する。すなわち、話者（たろう）は、音声認識システムにおいて登録されていない。そこで、話者識別部 33 は、話者を識別する試みが失敗したことを表すレスポンスを生成する。

【0083】

ステップ 1020 にて、制御部 30 は、対話分析・生成リクエストを対話分析・生成部 37 に送信する。対話分析・生成部 37 は、対話分析・生成リクエストの受信に回答して、対話のためのメッセージ 1031 を生成する。メッセージ 1031 は、たとえば「たろうさんだね。それじゃはじめよ。最初はりんご。」のように、発話の内容および話者を識別する情報を含むメッセージとして生成される。

10

【0084】

ステップ 1030 にて、対話分析・生成部 37 は、メッセージ 1031 を制御部 30 に送信する。制御部 30 は、メッセージ 1031 の受信を検知すると、端末への発話に対して回答するため、メッセージ 1031 と端末 ID とを含む音声レスポンスを生成する。

【0085】

ステップ 1035 にて、制御部 30 は、当該音声レスポンスを端末に送信する。端末の音声出力部 32 は、音声レスポンスの信号を受信すると、当該信号に基づく音声を出力する。ユーザは、その音声を認識すると、次の回答を考えて、端末に発話する。音声入力部 31 は、その発話、たとえば「ゴリラ」を受け付ける。

20

【0086】

その後、しりとりのための数回のやり取りが行なわれる（ステップ 1040 以降）。

ステップ 1040 にて、音声入力部 31 は、受け付けたメッセージ 1041 を制御部 30 に送信する。制御部 30 は、メッセージ 1041 の受信を検知すると、音声認識リクエストを生成する。

【0087】

ステップ 1045 にて、制御部 30 は、音声認識リクエストを音声認識部 36 に送信する。音声認識部 36 は、当該リクエストを受信すると、音声認識処理を開始する。

【0088】

ステップ 1050 にて、制御部 30 は、話者音声保存・リスト取得リクエストをユーザ管理部 35 に送信する。ユーザ管理部 35 は、当該リクエストの受信を検知すると、話者（たろう）の識別 ID と、話者（たろう）の名前とを、互いに関連付けることにより保存する。さらに、ユーザ管理部 35 は、話者音声の保存が成功したことを表す回答を生成する。

30

【0089】

ステップ 1055 にて、ユーザ管理部 35 は、当該回答として、話者音声保存・リスト取得レスポンスを制御部 30 に送信する。

【0090】

ステップ 1060 にて、制御部 30 は、話者識別モデル学習リクエストを話者識別学習部 34 に送信する。話者識別学習部 34 は、当該リクエストの受信を検知すると、話者識別モデルとして、当該発話を与えたユーザに音声を関連付けてモデルを生成し、適宜、更新する。

40

【0091】

ステップ 1065 にて、音声認識部 36 は、音声認識リクエストに基づく処理の結果を音声認識レスポンスとして制御部 30 に送信する。

【0092】

ステップ 1070 にて、話者識別学習部 34 は、話者識別モデル学習リクエストに対する回答して、話者識別学習レスポンスを制御部 30 に送信する。

【0093】

50

ステップ1075にて、制御部30は、対話分析・生成リクエストを生成して、生成したリクエストを対話分析・生成部37に送信する。たとえば、制御部30は、話者の学習のために十分なデータがなく学習失敗であると判断した場合には、当該リクエストを生成する。対話分析・生成部37は、当該リクエストの受信を検知すると、さらに学習するためのメッセージ1081（たとえば、「ゴリラ・・・それじゃあ「ラクダ」）を生成する。

【0094】

ステップ1080にて、対話分析・生成部37は、生成したメッセージ1081を制御部30に送信する。制御部30は、メッセージ1081を受信すると、端末IDとメッセージ1081とを含む音声レスポンスを生成する。

10

【0095】

ステップ1085にて、制御部30は、生成した音声レスポンスを端末に送信する。端末は、音声レスポンスを受信すると、音声出力部32は、音声レスポンスに基づく音声を出力する。ユーザは、端末の音声出力部32から発せられた音声を認識すると、その次の応答を考える。予め定められた時間内にユーザが、当該次の応答を発すると、音声入力部31は、ユーザの発話を受け付けて、当該発話に応じた音声応答を生成する。

【0096】

図11を参照して、ステップ1110にて、音声入力部31は、メッセージ1111（たとえば、「ダイヤモンド」）を制御部30に送信する。制御部30は、メッセージ1111の受信を検知すると、音声認識リクエストと、話者音声保存・リスト取得リクエストとを生成する。

20

【0097】

ステップ1115にて、制御部30は、音声認識リクエストを音声認識部36に送信する。音声認識部36は、当該リクエストの受信を検知すると、メッセージ1111の音声認識処理を開始する。

【0098】

ステップ1120にて、制御部30は、メッセージ1111と話者音声保存・リスト取得リクエストとをユーザ管理部35に送信する。ユーザ管理部35は、当該リクエストの受信を検知すると、メッセージ1111の内容（音声データ）を、ユーザ（話者）の識別IDに関連付けて格納する。

30

【0099】

ステップ1130にて、制御部30は、話者識別モデル学習リクエストを話者識別学習部34に送信する。話者識別学習部34は、当該リクエストの受信を検知すると、話者識別モデルを学習する。より具体的には、話者識別学習部34は、ユーザの識別IDと、メッセージ1111に含まれる音声情報（たとえば、声紋情報）とを関連付けて保存する。学習が完了すると、話者識別学習部34は、話者識別モデルの学習が完了したことを表すレスポンスを生成する。

【0100】

ステップ1135にて、音声認識部36は、音声認識処理が終わったことに応答して、音声認識処理の結果を通知する音声認識レスポンスを生成し、当該レスポンスを制御部30に送信する。

40

【0101】

ステップ1140にて、話者識別学習部34は、生成したレスポンスと制御部30に送信する。制御部30は、音声認識部36からのレスポンスと話者識別学習部34からのレスポンスを受信すると、学習に十分なデータが揃い、学習が完了したか否かを判断する。たとえば、予め定められた数以上の音声データがユーザの識別IDに関連付けられた場合には、制御部30は、学習に十分なデータが揃い学習が完了したと判断する。

【0102】

制御部30は、音声認識部36からのレスポンスと話者識別学習部34からのレスポンスの受信の内容に基づいて、対話分析・生成リクエストを生成する。たとえば、制御部3

50

0 は、各レスポンスの結果に基づいて、音声認識が成功し、かつ、学習に十分なデータが揃い学習が完了したと判断すると、当該リクエストを生成する。学習に十分なデータとは、たとえば、予め定められた一定時間内に音声データから抽出された情報量（一定のデータサイズを有する声紋情報の個数など）が学習に必要であると規定された情報量を超えているものをいう。

【0103】

ステップ1145にて、制御部30は、生成したリクエストを対話分析・生成部37に送信する。対話分析・生成部37は、当該リクエストの受信を検知すると、メッセージ1111に対するメッセージ1151を生成する。

【0104】

ステップ1150にて、対話分析・生成部37は、生成したメッセージ1151を制御部30に送信する。制御部30は、メッセージ1151の受信を検知すると、端末IDとメッセージ1151とを含む音声レスポンスを生成する。

【0105】

ステップ1155にて、制御部30は、端末に音声レスポンスを生成する。端末は、音声レスポンスを受信すると、音声出力部32から音声を出力する。

【0106】

<ユーザからの発話起点のシーケンス>

図12を参照して、他の局面について説明する。図12は、ユーザが音声認識システムに既知である場合におけるユーザ1と端末2とのやり取りのシーケンスを表す図である。なお、前述の動作と同じ動作には同じ番号を付してある。したがって、同じ動作の説明は、繰り返さない。

【0107】

ユーザが既に登録されている場合には、話者モデルが適宜更新される。したがって、常に直近のユーザの音声データに基づいた話者識別が可能となる。

【0108】

ユーザ1が端末2に対して、メッセージ10を発する。端末2は、メッセージ10を受け付けると、音声認識処理と話者識別処理とを実行する。端末2は、話者識別処理の結果に基づいて、メッセージ10の話者を識別できたと判断すると、その判断の結果に応じて、メッセージ1210を発する。メッセージ1210は、メッセージ10に対する応答と、メッセージ10の話者を確認するための問いかけとを含む。ユーザ1が、メッセージ1210に対するメッセージ1220を発すると、端末2は、メッセージ1220について音声認識処理と話者識別処理とを行なう。

【0109】

端末2は、メッセージ1220の内容から、当該問いかけに対する回答が得られたと判断すると、端末2の端末IDとユーザ名（たろう）とを含むデータをユーザ管理部35に送信する。ユーザ管理部35は、当該データを蓄積する。さらに、端末2は、メッセージ1220に対するメッセージ1230を発する。

【0110】

その後、端末2は、ユーザ1からの発話を認識するたびに、端末IDとユーザ名とを含むデータをユーザ管理部35に送信する。ユーザ管理部35は、各データを保存する。

【0111】

話者識別学習部34は、ユーザ管理部35から、端末IDとユーザ名とを参照して、蓄積されたデータから、当該ユーザに関連付けられたデータを読み出し、話者モデル80を作成する。

【0112】

図13および図14を参照して、ある局面に従う音声認識システムにおけるシーケンスについて説明する。図13および図14は、ユーザが既知である場合に行なわれる処理の流れを表すシーケンスチャートである。なお、前述の処理と同一の処理には同一のステップ番号を付してある。したがって、同一の処理の説明は繰り返さない。

10

20

30

40

50

【0113】

ステップ1340にて、話者識別部33は、話者識別が成功したことを通知するために、話者識別レスポンスを制御部30に送信する。制御部30は、当該レスポンスと、音声認識部36からのレスポンスとの受信を検知すると、対話分析・生成リクエストを生成する。当該リクエストは、音声識別結果と話者識別結果とを含む。

【0114】

ステップ1345にて、制御部30は、対話分析・生成部37に対して、対話分析・生成リクエストを送信する。対話分析・生成部37は、当該リクエストの受信を検知すると、メッセージ911に回答するためのメッセージ1351を生成する。このとき、メッセージ1351は、メッセージ911に対する応答と、メッセージ911の発話者を確認するための問いかけとを含む。

10

【0115】

ステップ1350にて、対話分析・生成部37は、生成したメッセージ1351を制御部30に送信する。制御部30がメッセージ1351と端末IDとを含む音声レスポンスを端末に送信すると、端末の音声出力部32は、音声を発話する。ユーザは、当該音声を認識して当該音声が良いと判断すると、たとえば「そうだよ」とのメッセージ1361を発する（名前登録発話）。

【0116】

ステップ1360にて、音声入力部31は、メッセージ1361の入力を受け付けると、その入力に応じた音声信号を制御部30に送信する。その後、制御部30は、音声認識リクエストを音声認識部36に送信する（ステップ965）。

20

【0117】

図14を参照して、ステップ1410にて、話者識別部33は、話者識別リクエスト（ステップ980）に対する応答を話者認識レスポンスとして話者識別部33に送信する。ユーザが音声認識システムにとって既知である場合、話者認識レスポンスは、話者が識別されたことを表す。制御部30は、当該レスポンスの受信を検知すると、対話分析・生成リクエストを生成する。

【0118】

ステップ1420にて、制御部30は、生成した対話分析・生成リクエストを対話分析・生成部37に送信する。対話分析・生成部37は、当該リクエストの受信を検知すると、メッセージ1431を生成する。メッセージ1431は、これまでのやり取りの結果に基づいて、メッセージ1351に含まれる問いかけ{たろうさんかな?}が正しかったことを踏まえた内容(やっぱり!)を含む。

30

【0119】

ステップ1430にて、対話分析・生成部37は、メッセージ1431を制御部30に送信する。制御部30は、メッセージ1431の受信を検知すると、端末IDとメッセージ1431とを含む音声レスポンスを生成する。

【0120】

ステップ1440にて、制御部30は、端末に音声レスポンスを送信する。音声出力部32は、当該音声レスポンスに基づいて、メッセージ1431を音声で出力する。

40

【0121】

その後、ステップ1040以降の処理が、前述の場合と同様に行なわれる。音声データが保存され、学習データ(たとえば、声紋情報等)は、対象ユーザの常に新しい音声データで更新される。なお、ユーザが既知の場合には、学習が完了しても、端末は、ユーザの名前を確認するための発話を行なわない。

【0122】

< 端末が発話の起点となる場合 >

図15~図17を参照して、さらに別の局面について説明する。図15は、端末2からユーザ1に話しかけることが対話のトリガとなる場合を表す図である。

【0123】

50

端末 2 からユーザに話しかけ、ユーザ発話及びユーザ名を聞き出すことによって得られた音声データをユーザ名と端末 ID とに紐付けることにより、音声データを学習する。

【 0 1 2 4 】

端末 2 は、ユーザ 1 の存在を検知すると、ユーザ 1 に対して話しかける。ユーザ 1 の存在の検知は、たとえば、赤外線センサ、人感センサ等からの出力に基づいて行なわれる。端末 2 は、たとえば、メッセージ 1 5 1 0 を発する。ユーザ 1 は、メッセージ 1 5 1 0 を認識する。

【 0 1 2 5 】

ユーザ 1 は、メッセージ 1 5 1 0 に応答して、メッセージ 1 5 2 0 を発する。端末 2 は、メッセージ 1 5 1 0 を認識すると、音声認識処理と話者識別処理とを実行する。端末 2 は、各処理の結果に基づいて、ユーザ 1 に対する発話を切り換える。たとえば、話者が既知でないと判断すると、端末 2 は、メッセージ 1 5 3 0 を生成し、音声でメッセージ 1 5 3 0 を出力する。

10

【 0 1 2 6 】

ユーザ 1 は、メッセージ 1 5 3 0 に応答してメッセージ 1 5 4 0 を端末 2 に向けて発する。端末 2 は、メッセージ 1 5 4 0 について音声認識処理および話者識別処理を実行する。さらに、端末 2 は、端末 2 のユーザ名として認識された話者「たろう」と端末 ID とを関連付け、これまで受け付けたユーザ 1 のメッセージ 1 5 2 0 , 1 5 4 0 を話者の音声データとしてユーザ管理部 3 5 に蓄積する。

【 0 1 2 7 】

さらに、端末 2 は、メッセージ 1 5 4 0 に対する応答としてメッセージ 1 5 5 0 を生成し、音声でメッセージ 1 5 5 0 を出力する。

20

【 0 1 2 8 】

ユーザ管理部 3 5 には、ユーザ「たろう」に関連付けられた音声データと、音声データから取得された識別情報（たとえば声紋情報）とが蓄積される。

【 0 1 2 9 】

図 1 6 および図 1 7 を参照して、ある局面における音声認識システムの動作について説明する。図 1 6 および図 1 7 は、音声認識システムで行われる処理の一部を表すシーケンスチャートである。

【 0 1 3 0 】

ステップ 1 6 1 0 にて、制御部 3 0 は、予め定められた条件が成立したことを検知すると、対話生成リクエストを対話分析・生成部 3 7 に送信する。当該条件は、たとえば、音声認識システムの範囲内でユーザの存在が検知されたこと、予め指定された時刻が到来したこと等である。対話生成リクエストは、たとえば、検出されたユーザに対して話しかけるためのメッセージ 1 5 1 0 の生成要求を含む。対話分析・生成部 3 7 は、当該リクエストの受信を検知すると、予め準備されたテンプレートに基づいて、メッセージ 1 5 1 0 を生成する。

30

【 0 1 3 1 】

ステップ 1 6 1 5 にて、対話分析・生成部 3 7 は、当該リクエストに応答して生成したメッセージ 1 5 1 0 を制御部 3 0 に送信する。制御部 3 0 は、メッセージ 1 5 1 0 の受信を検知すると、メッセージ 1 5 1 0 と端末 ID とを含む音声発話リクエストを端末に送信する。端末の音声入力部 3 1 は、当該リクエストを受信すると、メッセージ 1 5 1 0 を音声で出力する。ユーザは、メッセージ 1 5 1 0 を認識すると、メッセージ 1 5 1 0 に対する応答として、メッセージ 1 5 2 0 を発する。

40

【 0 1 3 2 】

ステップ 1 6 2 5 にて、音声入力部 3 1 は、メッセージ 1 5 2 0 を音声信号として制御部 3 0 に送信する。その後、ステップ 9 1 5 からステップ 1 3 4 5 まで、前述の処理と同様の処理が実行される。

【 0 1 3 3 】

ステップ 1 3 5 0 にて、対話分析・生成部 3 7 は、メッセージ 1 5 3 0 を制御部 3 0 に

50

送信する。メッセージ1530に基づく音声が出力されると、ユーザは、メッセージ1540を発する。メッセージ1540は、制御部30から音声認識部36に送られ、音声認識処理が実行される(ステップ1045)。

【0134】

図17を参照して、ステップ1050からステップ1070までの処理が、同様に実行される。その後、制御部30は、学習に十分なデータがなく、学習が失敗したと判断すると、ステップ1740の処理が実行される。より具体的には、ステップ1741にて、制御部30は、対話分析・生成リクエストを対話分析・生成部37に送信する。対話分析・生成部37は、当該リクエストの受信を検知すると、当該リクエストに応じたメッセージ1550を生成する。

10

【0135】

ステップ1742にて、対話分析・生成部37は、メッセージ1550を制御部30に送信する。制御部30は、メッセージ1550の受信を検知すると、端末IDとメッセージ1550を含む音声レスポンスを生成する。

【0136】

一方、制御部30は、学習に十分なデータが揃い楽手が完了したと判断すると、ステップ1750の処理を実行する。より詳しくは、ステップ1751にて、制御部30は、対話分析・生成リクエストを対話分析・生成部37に送信する。対話分析・生成部37は、当該リクエストの受信を検知すると、当該リクエストに回答するためのメッセージ1560を生成する。

20

【0137】

ステップ1752にて、対話分析・生成部37は、メッセージ1560を制御部30に送信する。制御部30は、メッセージ1560の受信を検知すると、端末IDとメッセージ1560を含む音声レスポンスを生成する。

【0138】

ステップ1760にて、制御部30は、当該音声レスポンスを端末に送信する。音声出力部32は、音声レスポンスを受信すると、メッセージ1560を音声で出力する。

【0139】

<他の局面>

さらに他の局面について説明する。他の局面において、以下の構成が用いられてもよい。

30

【0140】

(1) 音声認識と音声認証とが並列に行なわれる。したがって、ユーザの発話内容の認識と当該ユーザの認証とが同時に行なわれる。

【0141】

(2) ユーザ毎に、対話内容のログに基づいて各ユーザの興味ある話題が推定され、推定された話題に基づく対話が生成される。

【0142】

(3) 対話数やその頻度に基づいて、ロボット(音声対話装置、あるいは音声対話システム)の発話内容が変化する。

40

【0143】

これらの要素の結果、ユーザは、ロボット(音声対話システム)に親しみを持つことができる。

【0144】

たとえば、構成(1)により、当該技術思想が適用される音声対話システムは、カメラや無線タグ等の機器からの情報を使用することなく、ユーザを特定し(音声認証)、また、当該ユーザの発言内容の取得(音声認識)が可能になる。

【0145】

次に、構成(2)により、ユーザの日々の会話が音声対話システムに記憶され、必要に応じて分析される。音声対話システムは、分析結果に基づいて、各ユーザが興味ある話題

50

(スポーツ、芸能ニュースなど)を他の情報提供装置から取得し、対話しているユーザに応じた話題を当該ユーザに提供することができる。

【0146】

さらに、構成(3)により、音声対話システムとユーザとの対話が長期にかつ定期的に行われることにより、対話内容に応じて、音声対話システムからの発話の表現(言葉づかい、語調等)が変化し得る。その結果、ユーザが音声対話システム(あるいは、音声対話システムに含まれるロボットのような音声入出力端末)に対して親近感を持ち得る。これらの各構成は、適宜組み合わせられ得る。

【0147】

<まとめ>

以上のようにして、本実施の形態に係る音声認識システムによれば、ユーザは学習のための前処理を意識せずに、通常の音声対話を行うことにより、学習に必要な音声データをシステムに与えることができる。したがって、当該システムにより提供される機能を容易に利用することができる。

【0148】

さらに他の局面において、ユーザが意識することなくユーザ認証され、当該ユーザに応じた話題が出力されるので、ユーザは音声認識システムにより提供されるサービスや機能に親近感を持ち得る。

【0149】

今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

【符号の説明】

【0150】

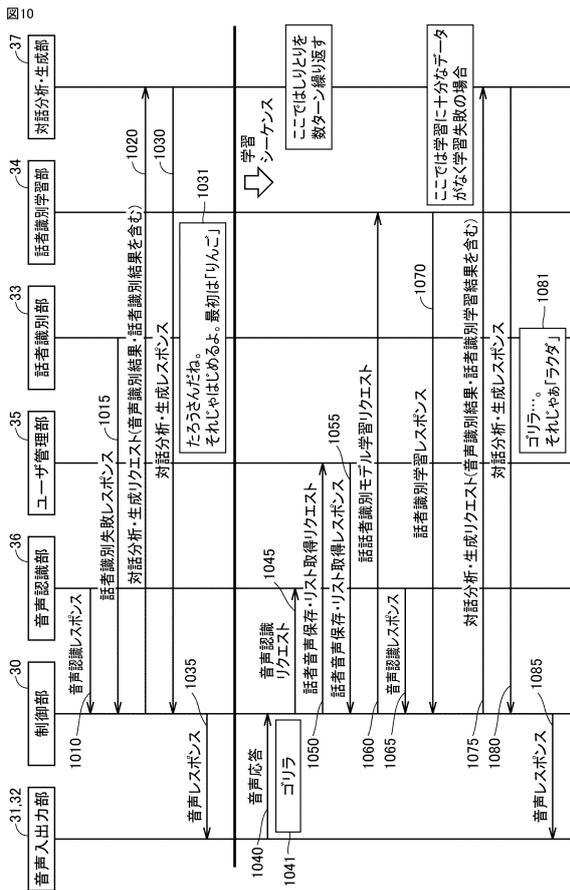
30 制御部、31 音声入力部、32 音声出力部、33 話者識別部、34 話者識別学習部、35 ユーザ管理部、36 音声認識部、37 生成部、80 話者モデル、350, 400, 500 サーバ、410 話者識別サーバ、520 音声認識サーバ、600 端末モジュール、610 メインモジュール、620 話者識別モジュール、630 音声認識モジュール。

10

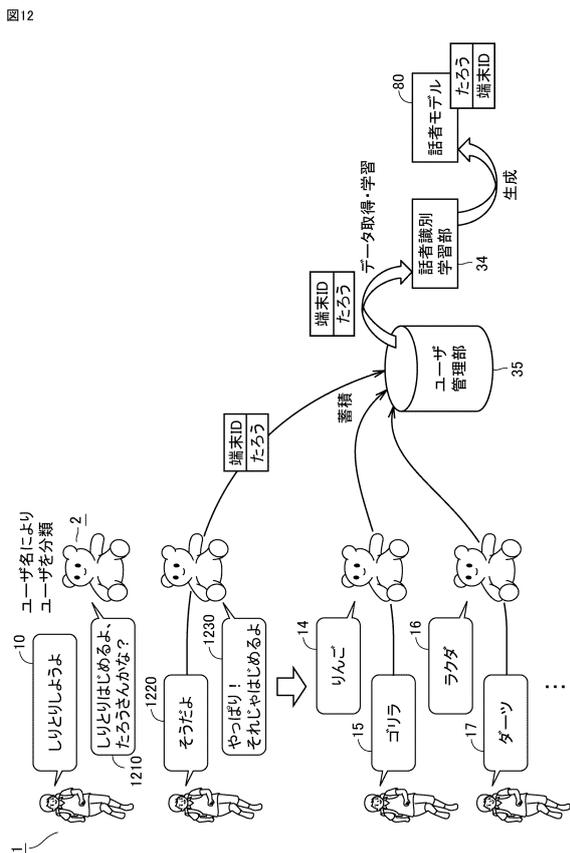
20

30

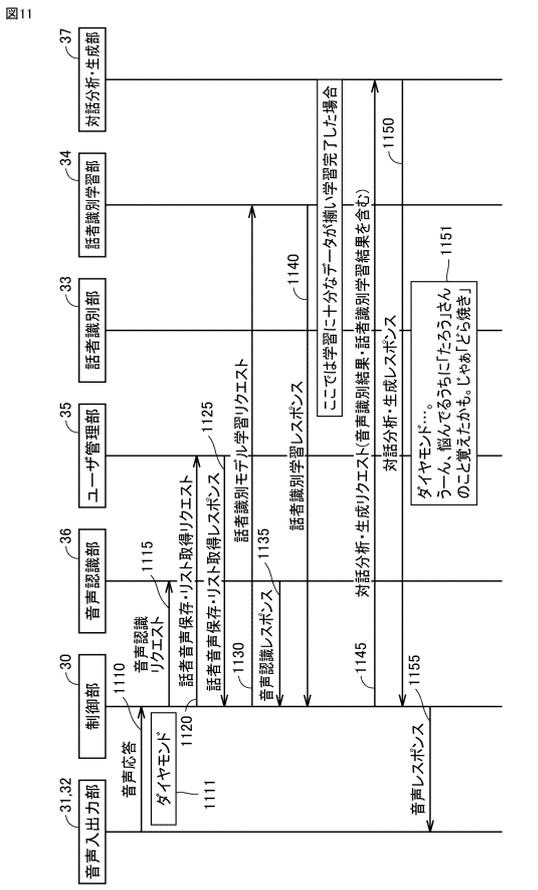
【図10】



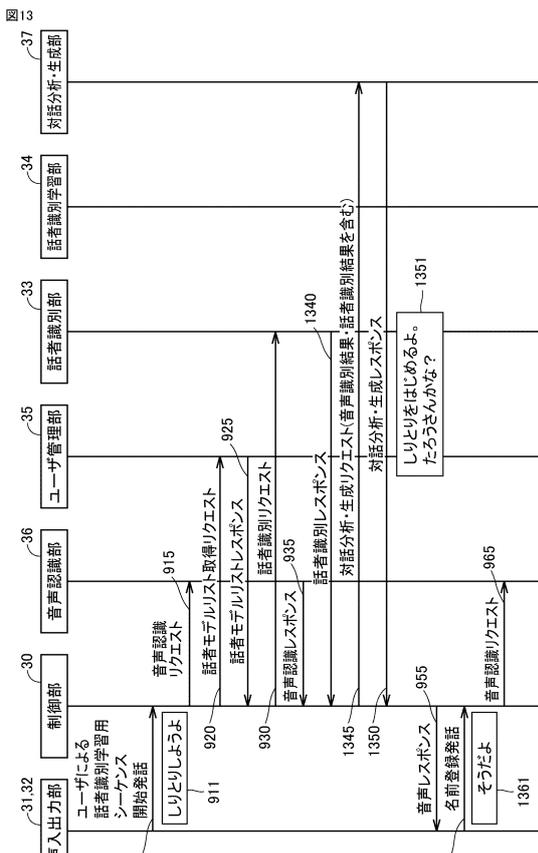
【図11】



【図12】



【図13】



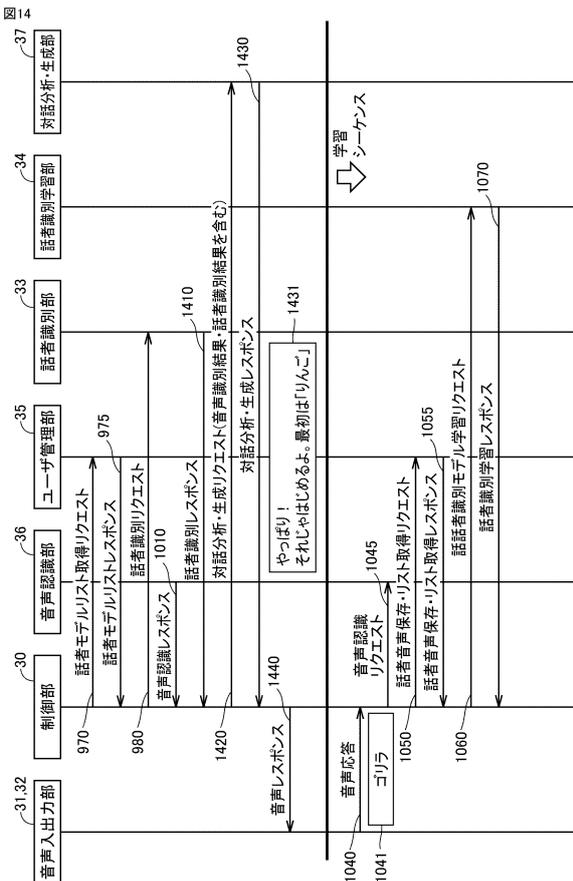
【図2】



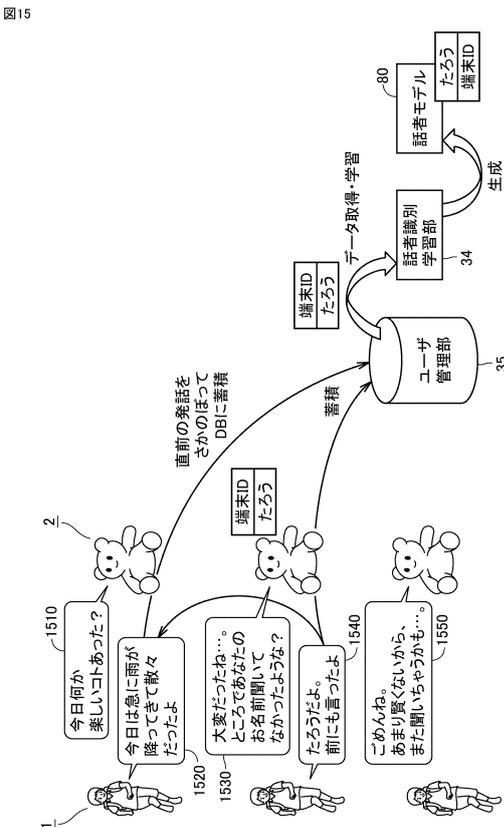
【図3】



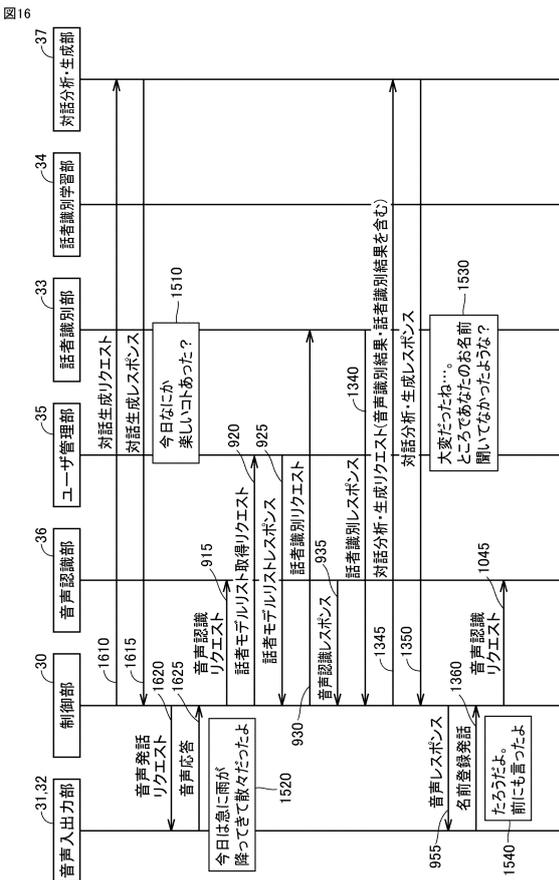
【図 14】



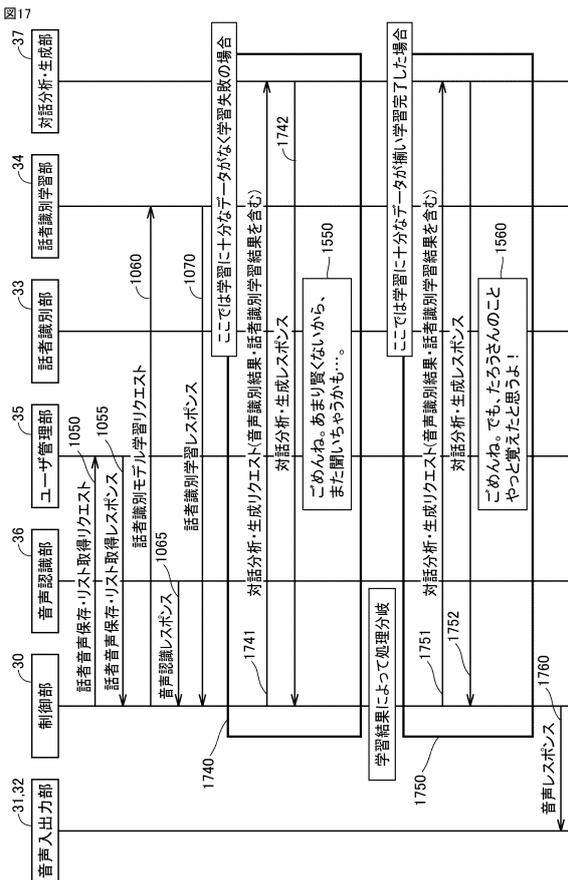
【図 15】



【図 16】



【図 17】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 3/16 6 5 0
G 0 6 F 3/16 6 9 0

(56)参考文献 特開2003-255989(JP,A)
特開2004-184788(JP,A)
特開2004-101901(JP,A)

(58)調査した分野(Int.Cl., DB名)
G 1 0 L 1 7 / 0 0 - 1 7 / 2 6
G 1 0 L 1 3 / 0 0
G 0 6 F 3 / 1 6