

(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2020-118803

(P2020-118803A)

(43) 公開日 令和2年8月6日(2020.8.6)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/22 4 5 3	
G 1 0 L 15/18 (2013.01)	G 1 0 L 15/18 3 0 0 H	
G 1 0 L 15/19 (2013.01)	G 1 0 L 15/19	

審査請求 未請求 請求項の数 9 O L (全 34 頁)

(21) 出願番号 特願2019-8679 (P2019-8679)
 (22) 出願日 平成31年1月22日 (2019.1.22)

(71) 出願人 391021684
 菱洋エレクトロ株式会社
 東京都中央区築地1丁目12番22号
 (74) 代理人 100120868
 弁理士 安彦 元
 (72) 発明者 菊田 敦
 東京都中央区築地一丁目12-22 コン
 ワビル 菱洋エレクトロ株式会社内
 (72) 発明者 越田 高広
 東京都中央区築地一丁目12-22 コン
 ワビル 菱洋エレクトロ株式会社内

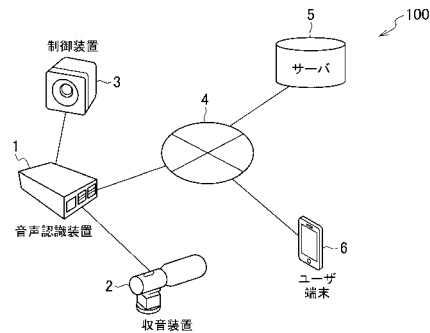
(54) 【発明の名称】 音声認識システム、及び音声認識装置

(57) 【要約】

【課題】認識精度の向上を可能とする音声認識システム、及び音声認識装置を提供する。

【解決手段】利用者に報知する音声に関する発話情報が記憶された発話データベースと、発話情報に含まれる第1発話情報を出力する第1出力手段と、第1発話情報を認識した利用者から発せられた音声に基づき、音声データを取得する取得手段と、音素認識により、音声データに含まれる開始無音区間及び終了無音区間を抽出し、開始無音区間と終了無音区間との間に挟まれた音素及び休止区間の配列を、認識対象データとして抽出する抽出手段と、文字列データベースを参照し、音素情報を選択し、選択された音素情報に紐づく文字列情報及びクラスIDを、候補データとして複数検出する検出手段と、算出手段と、選択手段と、生成手段とを備えることを特徴とする。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

音声に対応する認識情報を生成する音声認識システムであって、
 利用者に報知する音声に関する発話情報が記憶された発話データベースと、
 前記発話情報に含まれる第 1 発話情報を出力する第 1 出力手段と、
 前記第 1 発話情報を認識した前記利用者から発せられた前記音声に基づき、音声データを取得する取得手段と、

音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、
 前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記音素認識により認識対象データとして抽出する抽出手段と、

予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラス ID とが記憶された文字列データベースと、

前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラス ID を、候補データとして複数検出する検出手段と、

予め取得された前記クラス ID の配列順序を示す文法情報が記憶された文法データベースと、

前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせ合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出手段と、

前記信頼度に基づき、複数の前記候補データから評価データを選択する選択手段と、

前記評価データに基づき、前記認識情報を生成する生成手段と

を備えることを特徴とする音声認識システム。

【請求項 2】

前記生成手段のあと、前記認識情報に基づき、前記発話情報に含まれる前記第 1 発話情報、又は前記第 1 発話情報とは異なる第 2 発話情報を出力する第 2 出力手段をさらに備えること

を特徴とする請求項 1 記載の音声認識システム。

【請求項 3】

前記発話データベースには、前記発話情報毎に紐づく返答情報が記憶され、

前記第 2 出力手段は、前記認識情報に対し、前記第 1 発話情報に紐づく第 1 返答情報と同等であると判定した場合に、前記第 2 発話情報を出力すること

を特徴とする請求項 2 記載の音声認識システム。

【請求項 4】

前記発話データベースには、前記発話情報毎に紐づく返答情報が記憶され、

前記第 2 出力手段は、前記認識情報に対し、前記第 1 発話情報に紐づく第 1 返答情報と同等ではないと判定した場合に、前記第 1 発話情報を出力すること

を特徴とする請求項 2 記載の音声認識システム。

【請求項 5】

予め取得された前記文字列情報と、前記文字列情報を組み合わせた参照センテンスと、
 前記文字列情報毎に付与された閾値とが記憶された参照データベースをさらに備え、

前記生成手段は、

前記参照データベースを参照し、前記参照センテンスのうち、前記評価データに対応する第 1 参照センテンスを指定する指定手段と、

前記評価データに対応する前記信頼度と、前記第 1 参照センテンスに含まれる第 1 文字列情報に付与された第 1 閾値とを比較する比較手段と、

を有し、前記比較手段の比較結果に基づき、前記認識情報を生成すること

を特徴とする請求項 1 ~ 4 の何れか 1 項記載の音声認識システム。

【請求項 6】

前記認識情報に基づき、参照する前記文字列データベースの内容を選択する設定手段を

10

20

30

40

50

さらに備えること

を特徴とする請求項 1 ~ 5 の何れか 1 項記載の音声認識システム。

【請求項 7】

音声に対応する認識情報を生成する音声認識システムであって、
 利用者から発せられた前記音声に基づき、音声データを取得する取得手段と、
 音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、
 前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記
 音素認識により認識対象データとして抽出する抽出手段と、
 予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に
 付与されたクラス ID とが記憶された文字列データベースと、
 前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前
 記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラス ID
 を、候補データとして複数検出する検出手段と、
 予め取得された前記クラス ID の配列順序を示す文法情報が記憶された文法データベー
 スと、
 前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合
 あわせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列
 情報に対する信頼度を、前記文法データベースを用いて算出する算出手段と、
 前記信頼度に基づき、複数の前記候補データから評価データを選択する選択手段と、
 前記評価データに基づき、前記認識情報を生成する生成手段と、
 前記利用者に報知する音声に関する発話情報が記憶された発話データベースと、
 前記認識情報に基づき、前記発話情報に含まれる第 1 発話情報を出力する出力手段と、
 を備えることを特徴とする音声認識システム。

10

20

【請求項 8】

音声に対応する認識情報を生成する音声認識装置であって、
 利用者に報知する音声に関する発話情報が記憶された発話データベースと、
 前記発話情報に含まれる第 1 発話情報を出力する出力部と、
 前記第 1 発話情報を認識した前記利用者から発せられた前記音声に基づき、音声データ
 を取得する取得部と、
 音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、
 前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記
 音素認識により認識対象データとして抽出する抽出部と、
 予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に
 付与されたクラス ID とが記憶された文字列データベースと、
 前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前
 記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラス ID
 を、候補データとして複数検出する検出部と、
 予め取得された前記クラス ID の配列順序を示す文法情報が記憶された文法データベー
 スと、
 前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合
 あわせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列
 情報に対する信頼度を、前記文法データベースを用いて算出する算出部と、
 前記信頼度に基づき、複数の前記候補データから評価データを選択する選択部と、
 前記評価データに基づき、前記認識情報を生成する生成部と
 を備えることを特徴とする音声認識装置。

30

40

【請求項 9】

音声に対応する認識情報を生成する音声認識装置であって、
 利用者から発せられた前記音声に基づき、音声データを取得する取得部と、
 音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、
 前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記

50

音素認識により認識対象データとして抽出する抽出部と、

予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラスIDとが記憶された文字列データベースと、

前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラスIDを、候補データとして複数検出する検出部と、

予め取得された前記クラスIDの配列順序を示す文法情報が記憶された文法データベースと、

前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出部と、

前記信頼度に基づき、複数の前記候補データから評価データを選択する選択部と、

前記評価データに基づき、前記認識情報を生成する生成部と、

前記利用者に報知する音声に関する発話情報が記憶された発話データベースと、

前記認識情報に基づき、前記発話情報に含まれる第1発話情報を出力する出力部と、
を備えることを特徴とする音声認識装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声に対応する認識情報を生成する音声認識システム、及び音声認識装置に関する。

【背景技術】

【0002】

従来、音声認識に関する技術として、例えば特許文献1の認知機能評価装置や、特許文献2の発話内容の把握システム等が提案されている。

【0003】

特許文献1の認知機能評価装置では、フォルマント解析部は、対象者の音声に含まれる特定の音素の瞬時音圧の時間変動を対象期間に亘って表している対象データを受け取る。そして、フォルマント解析部は、対象期間を複数のフレームに分割し、特定のフォルマントの周波数を、2つ以上の対象フレームのそれぞれについて求める。特徴解析部は、対象フレーム毎に求められた特定のフォルマントの周波数について特徴量を求める。評価部は、特徴量に基づいて対象者の認知機能の評価する。

【0004】

特許文献2では、録取された音声データに対して音素基準の音声認識を行ってインデクシングされたデータを保存し、これを用いて核心語に基づく発話内容を把握することにより、発話内容の把握が正確に、手軽に且つ速やかに行われる、録取された音声データに対する核心語の取出に基づく発話内容の把握システムと、このシステムを用いたインデクシング方法及び発話内容の把握方法等が開示されている。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2018-50847号公報

【特許文献2】特開2015-539364号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

ここで、音声認識に関する技術では、様々な分野での応用が期待される一方で、認識精度の向上が課題として挙げられている。認識精度を向上させるために、音素を用いる方法が注目を集めているが、音声データから音素の配列を取得する際のバラつき等により、依然として認識精度の向上が課題として挙げられている。

10

20

30

40

50

【0007】

この点、特許文献1では、対象者の音声に基づく特定のフォルマント周波数について特徴量を求め、特徴量に基づいて対象者の認知機能を評価することで、精度の向上を図っている。しかしながら、特許文献1の開示技術では、対象者の発する音声の内容までを認識することができない。

【0008】

また、特許文献2では、核心語に基づく発話内容を把握することにより、発話内容の把握を実現する技術が開示されている。しかしながら、特許文献2の開示技術では、音素の類似する核心語が発話内容に含まれる場合、認識精度が悪くなる恐れがある。このような状況により、認識精度の向上を可能とする音声認識に関する技術が望まれている。

10

【0009】

そこで本発明は、上述した問題に鑑みて案出されたものであり、その目的とするところは、認識精度の向上を可能とする音声認識システム、及び音声認識装置を提供することにある。

【課題を解決するための手段】

【0010】

第1発明に係る音声認識システムは、音声に対応する認識情報を生成する音声認識システムであって、利用者に報知する音声に関する発話情報が記憶された発話データベースと、前記発話情報に含まれる第1発話情報を出力する第1出力手段と、前記第1発話情報を認識した前記利用者から発せられた前記音声に基づき、音声データを取得する取得手段と、音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記音素認識により認識対象データとして抽出する抽出手段と、予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラスIDとが記憶された文字列データベースと、前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラスIDを、候補データとして複数検出する検出手段と、予め取得された前記クラスIDの配列順序を示す文法情報が記憶された文法データベースと、前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出手段と、前記信頼度に基づき、複数の前記候補データから評価データを選択する選択手段と、前記評価データに基づき、前記認識情報を生成する生成手段とを備えることを特徴とする。

20

30

【0011】

第2発明に係る音声認識システムは、第1発明において、前記生成手段のあと、前記認識情報に基づき、前記発話情報に含まれる前記第1発話情報、又は前記第1発話情報とは異なる第2発話情報を出力する第2出力手段をさらに備えることを特徴とする。

【0012】

第3発明に係る音声認識システムは、第2発明において、前記発話データベースには、前記発話情報毎に紐づく返答情報が記憶され、前記第2出力手段は、前記認識情報に対し、前記第1発話情報に紐づく第1返答情報と同等であると判定した場合に、前記第2発話情報を出力することを特徴とする。

40

【0013】

第4発明に係る音声認識システムは、第2発明において、前記発話データベースには、前記発話情報毎に紐づく返答情報が記憶され、前記第2出力手段は、前記認識情報に対し、前記第1発話情報に紐づく第1返答情報と同等ではないと判定した場合に、前記第1発話情報を出力することを特徴とする。

【0014】

第5発明に係る音声認識システムは、第1発明～第4発明の何れかにおいて、予め取得された前記文字列情報と、前記文字列情報を組み合わせた参照センテンスと、前記文字列

50

情報毎に付与された閾値とが記憶された参照データベースをさらに備え、前記生成手段は、前記参照データベースを参照し、前記参照センテンスのうち、前記評価データに対応する第1参照センテンスを指定する指定手段と、前記評価データに対応する前記信頼度と、前記第1参照センテンスに含まれる第1文字列情報に付与された第1閾値とを比較する比較手段と、を有し、前記比較手段の比較結果に基づき、前記認識情報を生成することを特徴とする。

【0015】

第6発明に係る音声認識システムは、第1発明～第5発明の何れかにおいて、前記認識情報に基づき、参照する前記文字列データベースの内容を選択する設定手段をさらに備えることを特徴とする。

10

【0016】

第7発明に係る音声認識システムは、音声に対応する認識情報を生成する音声認識システムであって、利用者から発せられた前記音声に基づき、音声データを取得する取得手段と、音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記音素認識により認識対象データとして抽出する抽出手段と、予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラスIDとが記憶された文字列データベースと、前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラスIDを、候補データとして複数検出する検出手段と、予め取得された前記クラスIDの配列順序を示す文法情報が記憶された文法データベースと、前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出手段と、前記信頼度に基づき、複数の前記候補データから評価データを選択する選択手段と、前記評価データに基づき、前記認識情報を生成する生成手段と、前記利用者に報知する音声に関する発話情報が記憶された発話データベースと、前記認識情報に基づき、前記発話情報に含まれる第1発話情報を出力する出力手段と、を備えることを特徴とする。

20

【0017】

第8発明に係る音声認識装置は、音声に対応する認識情報を生成する音声認識装置であって、利用者に報知する音声に関する発話情報が記憶された発話データベースと、前記発話情報に含まれる第1発話情報を出力する出力部と、前記第1発話情報を認識した前記利用者から発せられた前記音声に基づき、音声データを取得する取得部と、音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、前記開始無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記音素認識により認識対象データとして抽出する抽出部と、予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラスIDとが記憶された文字列データベースと、前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラスIDを、候補データとして複数検出する検出部と、予め取得された前記クラスIDの配列順序を示す文法情報が記憶された文法データベースと、前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出部と、前記信頼度に基づき、複数の前記候補データから評価データを選択する選択部と、前記評価データに基づき、前記認識情報を生成する生成部とを備えることを特徴とする。

30

40

【0018】

第9発明に係る音声認識装置は、音声に対応する認識情報を生成する音声認識装置であって、利用者から発せられた前記音声に基づき、音声データを取得する取得部と、音素認識により、前記音声データに含まれる開始無音区間及び終了無音区間を抽出し、前記開始

50

無音区間と前記終了無音区間との間に挟まれた音素及び休止区間の配列を、前記音素認識により認識対象データとして抽出する抽出部と、予め取得された文字列情報と、前記文字列情報に紐づく音素情報と、前記文字列情報に付与されたクラスIDとが記憶された文字列データベースと、前記文字列データベースを参照し、前記認識対象データの有する前記配列に対応する前記音素情報を選択し、選択された前記音素情報に紐づく前記文字列情報及び前記クラスIDを、候補データとして複数検出する検出部と、予め取得された前記クラスIDの配列順序を示す文法情報が記憶された文法データベースと、前記文法データベースを参照し、複数の前記候補データを前記文法情報に基づき組み合わせたセンテンスを生成し、前記センテンスに含まれる前記候補データ毎の前記文字列情報に対する信頼度を、前記文法データベースを用いて算出する算出部と、前記信頼度に基づき、複数の前記候補データから評価データを選択する選択部と、前記評価データに基づき、前記認識情報を生成する生成部と、前記利用者に報知する音声に関する発話情報が記憶された発話データベースと、前記認識情報に基づき、前記発話情報に含まれる第1発話情報を出力する出力部と、を備えることを特徴とする。

10

【発明の効果】**【0019】**

第1発明～第7発明によれば、抽出手段は、音素及び休止区間の配列を認識対象データとして抽出する。また、検出手段は、認識対象データの有する配列に対応する音素情報を選択し、候補データを検出する。このため、認識対象データにおける音素のみを考慮した配列に対して候補データを検出する場合に比べ、誤認識を低減させることができる。これにより、認識精度の向上を図ることが可能となる。

20

【0020】

また、第1発明～第7発明によれば、文字列データベースには、音素と休止区間との配列に対応する音素情報、及び音素情報に紐づく文字列情報が記憶される。このため、音素全体に対してパターンマッチングするために記憶するデータに比べて、データ容量の削減や、データ蓄積の簡易化を実現することができる。

【0021】

また、第1発明～第6発明によれば、第1出力手段は、第1発話情報を出力する。このため、第1発話情報に対する利用者から発せられた音声を、認識情報として生成することができる。これにより、発話情報に対する利用者の認識力を高精度に評価することが可能となる。

30

【0022】

特に、第2発明によれば、第2出力手段は、認識情報に基づき、第2発話情報、又は第1発話情報を出力する。このため、利用者の認識力を評価した結果を、発話として報知することができる。これにより、対話形式による音声認識を実現することが可能となる。

【0023】

特に、第3発明によれば、第2出力手段は、認識情報に対し、第1返答情報と同等であると判定した場合に、第2発話情報を出力する。このため、第1発話情報に対して利用者が発した音声、が、所望の内容である場合、第1発話情報とは異なる第2発話情報を報知することができる。これにより、対話形式のバリエーションを増やすことが可能となる。

40

【0024】

特に、第4発明によれば、第2出力手段は、認識情報に対し、第1返答情報と同等ではないと判定した場合に、第1発話情報を出力する。このため、第1発話情報に対して利用者が発した音声、が、所望の内容ではない場合、第1発話情報を再度報知することができる。これにより、発話練習等のような正確な発話情報を復唱する必要がある場合や、暗号認識等のような正確な認識を得る必要がある場合にも利用することが可能となる。

【0025】

特に、第5発明によれば、比較手段は、信頼度と、第1閾値とを比較する。このため、複数の候補データから相対的に選択された評価データに対し、閾値による判定も行うことで、誤認識をさらに低減させることができる。これにより、認識精度のさらなる向上が可

50

能となる。

【0026】

特に、第6発明によれば、設定手段は、認識情報に基づき、参照する文字列データベースの内容を選択する。このため、特定の状況に応じた音声のみを認識することができる。これにより、認識精度のさらなる向上を図ることが可能となる。

【0027】

第8、9発明によれば、抽出部は、音素及び休止区間の配列を認識対象データとして抽出する。また、検出部は、認識対象データの有する配列に対応する音素情報を選択し、候補データを検出する。このため、認識対象データにおける音素のみを考慮した配列に対して候補データを検出する場合に比べ、誤認識を低減させることができる。これにより、認識精度の向上を図ることが可能となる。

10

【0028】

また、第8、9発明によれば、文字列データベースには、音素と休止区間との配列に対応する音素情報、及び音素情報に紐づく文字列情報が記憶される。このため、音素全体に対してパターンマッチングするために記憶するデータに比べて、データ容量の削減や、データ蓄積の簡易化を実現することができる。

【0029】

また、第8発明によれば、出力部は、第1発話情報を出力する。このため、第1発話情報に対する利用者から発せられた音声を、認識情報として出力することができる。これにより、発話情報に対する利用者の認識力を高精度に評価することが可能となる。

20

【図面の簡単な説明】

【0030】

【図1】図1は、本実施形態における音声認識システムの構成の一例を示す模式図である。

【図2】図2は、本実施形態における音声認識システムの動作の一例を示す模式図である。

【図3】図3(a)は、本実施形態における音声認識装置の構成の一例を示す模式図であり、図3(b)は、本実施形態における音声認識装置の機能の一例を示す模式図であり、図3(c)は、本実施形態における生成部の一例を示す模式図である。

【図4】図4は、本実施形態における音声認識装置の各機能の一例を示す模式図である。

30

【図5】図5は、文字列データベース、文法データベース、発話データベース、及び参照データベースの一例を示す模式図である。

【図6】図6(a)は、本実施形態における音声認識システムの動作の一例を示すフローチャートであり、図6(b)は、生成手段の一例を示すフローチャートであり、図6(c)は、反映手段の一例を示すフローチャートである。

【図7】図7は、更新手段の一例を示す模式図である。

【図8】図8(a)は、更新手段の一例を示すフローチャートであり、図8(b)は、設定手段の一例を示すフローチャートである。

【図9】図9は、設定手段の一例を示す模式図である。

【図10】図10は、条件情報の一例を示す模式図である。

40

【図11】図11は、文字列データベース及び参照データベースの変形例を示す模式図である。

【図12】図12は、参照データベースの変形例を示す模式図である。

【発明を実施するための形態】

【0031】

以下、本発明の実施形態における音声認識システム及び音声認識装置の一例について、図面を参照しながら説明する。

【0032】

(音声認識システム100の構成)

図1～図5を参照して、本実施形態における音声認識システム100の構成の一例につ

50

いて説明する。図 1 は、本実施形態における音声認識システム 100 の全体の構成を示す模式図である。

【0033】

音声認識システム 100 は、利用者の用途に応じて構築された文字列データベース及び文法データベースを参照し、利用者の音声に対応する認識情報を生成する。文字列データベースには、利用者が発すると想定される文字列（文字列情報）と、文字列に対応する音素（音素情報）が記憶される。このため、上記文字列及び音素を蓄積することで用途に応じた認識情報を生成でき、様々な用途に展開することが可能となる。

【0034】

特に、文字列データベースに記憶される音素の配列（音素情報）は、音声に含まれる休止区間を踏まえて分類することで、音声に対する認識情報の精度を飛躍的に向上させることが可能となることを、発明者が発見した。また、音声認識システム 100 では、例えば文字列データベースに記憶された文字列情報に限定して、音声に対する認識情報を生成する。このため、用途毎に文法データベースの内容を変更することで、用途に適した認識情報を高精度に生成することが可能となる。

【0035】

文法データベースには、文字列情報を組み合わせたセンテンスを生成するために必要な文法情報が記憶される。文法情報は、文字列情報毎に紐づくクラス ID の配列順序を示す情報を複数含む。文法データベースを参照することで、休止区間を踏まえて分類された音素の配列に基づいて文字列情報を検出したあと、容易に各文字列情報を組み合わせることができる。これにより、音声に対する文法を考慮した認識情報を生成することができる。この結果、利用者等の発する音声の内容を踏まえた音声認識を高精度に実現することが可能となる。

【0036】

図 1 に示すように、音声認識システム 100 は、音声認識装置 1 を備え、例えば收音装置 2、制御装置 3、及びサーバ 5 の少なくとも何れかを備えてもよい。音声認識システム 100 では、例えば收音装置 2 等を用いて利用者等の音声を收音し、音声認識装置 1 を用いて音声に対する認識情報を生成する。認識情報は、音声を文字列に変換したテキストデータ等のほか、例えば制御装置 3 等を制御する情報を含み、例えば利用者に報知（返答）するための音声に関する発話情報を含んでもよい。

【0037】

音声認識システム 100 では、音声認識装置 1 に対して、收音装置 2 や制御装置 3 が直接接続されるほか、例えば公衆通信網 4 を介して接続されてもよい。また、音声認識装置 1 に対して、例えば公衆通信網 4 を介して、サーバ 5 や利用者等の保有するユーザ端末 6 が、公衆通信網 4 を介して接続されてもよい。なお、音声認識装置 1 には、收音装置 2 及び制御装置 3 の少なくとも何れかが一体に形成されてもよい。

【0038】

音声認識システム 100 の用途として、例えば図 2 に示すように、利用者との疑似対話や、発声訓練等が、一例として挙げられる。例えば利用者に報知する音声に関する発話情報に基づき、スピーカ等の制御装置 3 から音声が出力される（発話情報出力 1）。その後、利用者が音声を認識し、利用者から発せられた音声、マイク等の收音装置 2 を用いて收音される（音声収集）。そして、收音された音声に基づく音声データを音声認識装置 1 が取得し、音声データに対する認識情報を生成する（認識情報生成）。生成された認識情報は、例えば利用者に再度報知するための発話情報を選択するときに用いられ、選択された発話情報に基づき、再度制御装置 3 から音声が出力される（発話情報出力 2）。なお、発話情報出力 2 では、例えば生成された認識情報に基づき、制御装置 3 から音声が出力されてもよい。

【0039】

< 音声認識装置 1 >

図 3 (a) は、音声認識装置 1 の構成の一例を示す模式図である。音声認識装置 1 とし

10

20

30

40

50

て、Raspberry Pi（登録商標）等のシングルボードコンピュータが用いられるほか、例えばパーソナルコンピュータ（PC）等の電子機器が用いられてもよい。音声認識装置1は、筐体10と、CPU（Central Processing Unit）101と、ROM（Read Only Memory）102と、RAM（Random Access Memory）103と、保存部104と、I/F105～107とを備える。各構成101～107は、内部バス110により接続される。

【0040】

CPU101は、音声認識装置1全体を制御する。ROM102は、CPU101の動作コードを格納する。RAM103は、CPU101の動作時に使用される作業領域である。保存部104は、文字列データベース等の各種情報が保存される。保存部104として、例えばSDメモリーカードのほか、例えばHDD（Hard Disk Drive）、SSD（Solid State Drive）等が用いられる。

10

【0041】

I/F105は、收音装置2、制御装置3、公衆通信網4等との各種情報の送受信を行うためのインターフェースである。I/F106は、用途に応じて接続される入力部分108との各種情報の送受信を行うためのインターフェースである。入力部分108として、例えばキーボードが用いられ、音声認識システム100の管理等を行う利用者等は、入力部分108を介して、各種情報又は音声認識装置1の制御コマンド等を入力又は選択する。I/F107は、用途に応じて接続される出力部分109との各種情報の送受信を行うためのインターフェースである。出力部分109は、保存部104に保存された各種情報、認識情報、音声認識装置1の処理状況等を入力する。出力部分109として、ディスプレイが用いられ、例えばタッチパネル式でもよい。この場合、出力部分109が入力部分108を含む構成としてもよい。出力部分109として、例えばスピーカ機能を備えたディスプレイ等の機器が用いられてもよい。なお、I/F105～I/F107として、例えば同一のものが用いられてもよく、各I/F105～I/F107として、例えばそれぞれ複数ものが用いられてもよい。また、入力部分108及び出力部分109の少なくとも何れかは、状況に応じて取り外されてもよい。

20

【0042】

図3（b）は、音声認識装置1の機能の一例を示す模式図である。音声認識装置1は、取得部11と、抽出部12と、記憶部13と、検出部14と、算出部15と、選択部16と、生成部17と、出力部18とを備える。音声認識装置1は、例えば反映部19を備えてもよい。なお、図3（b）に示した各機能は、CPU101が、RAM103を作業領域として、保存部104等に記憶されたプログラムを実行することにより実現される。また、各機能の一部は、例えばJulius等の公知の音素認識技術を含む音声認識エンジンや、Python等のような公知の汎用プログラミング言語を用いて実現し、各種データの抽出や生成等の処理を行ってもよい。また、各機能の一部は、人工知能により制御されてもよい。ここで、「人工知能」は、いかなる周知の人工知能技術に基づくものであってもよい。

30

【0043】

<取得部11>

取得部11は、少なくとも1つの音声データを取得する。取得部11は、例えば收音装置2等を用いて收音した音声（音声信号）に対し、PCM（pulse code modulation）等のパルス変調したデータを、音声データとして取得する。取得部11は、收音装置2の種類に応じて、例えば複数の音声データを一度に取得してもよい。なお、音声信号をパルス変調する処理は、收音装置2等により行われるほか、例えば取得部11により行われてもよい。

40

【0044】

取得部11は、例えば同時に複数の音声データを取得してもよい。この場合、音声認識装置1に対して、收音装置2が複数接続されるほか、複数の音声を同時に收音できる收音装置2が接続されてもよい。なお、取得部11は、音声データのほか、例えばI/F10

50

5、I/F106を介して各種情報(データ)を収音装置2等から取得する。

【0045】

<抽出部12>

抽出部12は、音素認識により、音声データに含まれる開始無音区間及び終了無音区間を抽出する。また、抽出部12は、開始無音区間と終了無音区間との間に挟まれた音素及び休止区間の配列を、音素認識により認識対象データとして抽出する。すなわち、抽出部12の基本的な機能は、音素認識技術を用いることにより実現することができる。

【0046】

抽出部12は、例えば100ミリ秒以上1秒以下の非発話状態(無音区間)を、開始無音区間及び終了無音区間として抽出する。抽出部12は、開始無音区間と終了無音区間との間に挟まれた区間(音声区間)に対し、音素及び休止区間を割り当てる。抽出部12は、それぞれ割り当てられた音素及び休止区間の配列を、認識対象データとして抽出する。

【0047】

音素は、母音と、子音とを含む公知のものである。休止区間は、開始無音区間及び終了無音区間よりも短い区間を示し、例えば音素の区間と同程度の区間(長さ)を示す。抽出部12は、例えば各音素の長さ又は認識対象データ全体の長さを判定したあと、休止区間の長さを設定した上で、音素及び休止区間を割り当てた配列を、認識対象データとして抽出してもよい。すなわち、抽出部12は、音素の長さ又は認識対象データ全体の長さに応じて、休止区間の長さを設定してもよい。

【0048】

抽出部12は、例えば図4に示すように、開始無音区間「silB」及び終了無音区間「siIE」を抽出し、音声区間における配列「a/k/a/r/i/*w/o*/ts/u/k/e/t/e」(*は休止区間を示す)を、対象認識データとして抽出する。抽出部12は、例えば1つの音声データからそれぞれ異なる配列の対象認識データを複数抽出してもよい。この場合、抽出部12における音素及び休止区間の割り当てに伴うバラつきを考慮した音声認識を実施することができる。例えば抽出部12は、1つ以上5つ以下の対象認識データを抽出することで、処理時間を抑えた上で、認識精度を高めることができる。なお、抽出部12は、例えば開始無音区間及び終了無音区間の少なくとも何れかを含む配列を、対象認識データとして抽出してもよい。

【0049】

休止区間は、例えば呼吸音及びリップノイズの少なくとも何れかを含んでもよい。すなわち、抽出部12は、例えば休止区間に含まれる呼吸音及びリップノイズの少なくとも何れかを、認識対象データとして抽出してもよい。この場合、後述する文字列データベースに記憶された音素情報に、呼吸音及びリップノイズの少なくとも何れかを含ませることで、より精度の高い認識情報を生成することが可能となる。

【0050】

<記憶部13、データベース>

記憶部13は、各種データを保存部104に記憶させ、又は各種データを保存部104から取出す。記憶部13は、必要に応じて保存部104に記憶された各種データベースを取出す。

【0051】

保存部104には、例えば図5に示すように、文字列データベース、文法データベース、及び発話データベースが記憶され、例えば参照データベースが記憶されてもよい。

【0052】

文字列データベースには、予め取得された文字列情報と、文字列情報に紐づく音素情報と、文字列情報に付与されたクラスIDとが記憶される。文字列データベースは、検出部14によって候補データを検出するときに用いられる。

【0053】

音素情報は、利用者が発すると想定される音素の配列(例えば第1音素情報「a/k/a/r/i」等)を複数含む。音素の配列は、休止区間により分離される区間に対応するほか、例

10

20

30

40

50

例えば「h/i/*/i/t/e」のように休止区間を含んでもよく、利用条件に応じて任意に設定される。なお、音素情報は、例えば開始無音区間及び終了無音区間の少なくとも何れかを含んでもよい。

【0054】

文字列情報は、各音素の配列に紐づく文字列（例えば第1文字列情報「明かり」等）を含む。このため、文字列情報には、単語や形態素等の意味を持つ表現要素が用いられるほか、意味を持たない文字列が用いられてもよい。なお、文字列情報は、日本語のほか、例えば2カ国以上の言語を含んでもよく、数字や利用箇所を用いられる略称等の文字列を含んでもよい。また、同一の文字列情報に対して、異なる音素の配列が紐づけられてもよい。

10

【0055】

クラスIDは、文字列情報に紐づき、文字列情報の単語等が文法上用いられると想定される配列箇所（例えば第1クラスID「1」等）を示す。例えば音声の文法（センテンス）が「対象」+「助詞」+「アクション」として表すことができる場合、クラスIDとして、音声の「対象」となる文字列情報に対して「1」が用いられ、音声の「助詞」となる文字列情報に対して「2」が用いられ、音声の「アクション」となる文字列情報に対して「3」が用いられる。

【0056】

文法データベースには、予め取得された複数のクラスIDの配列順序を示す文法情報が記憶される。文法データベースは、算出部15によって信頼度を算出するときに用いられる。文法情報として、例えば第1文法情報「1、2、3」が用いられる場合、音声の候補として「対象」+「助詞」+「アクション」を示すセンテンスを生成することができる。文法情報は、例えば第1文法情報「1、2、3」、第2文法情報「4、5、6」、第3文法情報「2、1、3」等のクラスIDの配列順序を複数含む。

20

【0057】

発話データベースには、予め取得された発話情報が記憶され、例えば返答情報が記憶されてもよい。発話情報は、利用者に報知する音声に関する情報を示し、用途に応じて利用者が認識できる言語に対応して記憶される。発話情報は、例えば文字列、数値、音素等の公知のデータ形式で記憶される。発話情報は、例えば制御装置3から発する音声を指定する数値やID等のような、制御装置3を制御するためのデータ形式で記憶されてもよい。発話情報は、例えばPCM等のデータ形式で記憶されてもよい。

30

【0058】

返答情報は、発話情報に紐づく情報を示す。返答情報は、認識情報を評価するために用いられ、例えば認識情報と等しいデータ形式で記憶される。返答情報として、発話情報に対して利用者の返答として相応しい内容であるか否かを判定できる情報が含まれ、例えば文字列や音素等のデータ形式が含まれてもよい。

【0059】

例えば図5に示すように、発話情報として「おはようございます」（第1発話情報）という内容に対し、返答情報「おはようございます」（第1返答情報）という文字列が紐づいている。この場合、第1返答情報が、認識情報と一致するか否かを評価し、利用者から発せられた音声に対する判定をしてもよい。また、返答情報として、例えば閾値等の数値を含んでもよい。この場合、返答情報の数値（閾値）を用いて、認識情報に含まれる数値を評価し、利用者から発せられた音声に対する評価をしてもよい。これにより、音声認識システム100では、疑似対話形式の音声認識を実現することができる。

40

【0060】

参照データベースには、予め取得された文字列情報と、文字列を組み合わせた参照センテンスと、文字列情報毎に付与された閾値とが記憶され、例えば文字列情報に紐づく音素情報が記憶されてもよい。参照データベースは、生成部17によって認識情報を生成するときに、必要に応じて用いられる。なお、参照データベースに記憶される文字列情報及び音素情報は、例えば文字列データベースに記憶される文字列情報及び音素情報と等しくす

50

ることで、データ容量を少なくすることができる。

【0061】

< 検出部 14 >

検出部 14 は、文字列データベースを参照し、認識対象データの有する音素の配列に対応する音素情報を選択する。また、検出部 14 は、選択された音素情報に紐づく文字列情報及びクラス ID を候補データとして複数検出する。

【0062】

検出部 14 は、例えば図 4 に示すように、認識対象データに対応する音素情報「a/k/a/r/i」、「w/o」、「ts/u/k/e/t/e」を選択し、各音素情報に紐づく文字列情報及びクラス ID 「明かり/1」、「を/2」、「つけて/3」を、それぞれ候補データとして検出する。このとき、認識対象データの数に応じて、候補データの数が増加する。なお、各音素の配列は、予め休止区間毎に区切られて分類されるほか、音素及び休止区間を含む音素情報に基づいて分類されてもよい。

10

【0063】

< 算出部 15 >

算出部 15 は、文法データベースを参照し、複数の候補データを文法情報に基づき組み合わせたセンテンスを生成する。また、算出部 15 は、センテンスに含まれる候補データ毎の文字列情報に対する信頼度を、文法データベースを用いて算出する。信頼度の算出には、例えば文法データベースを参照したスタックデコーディングサーチが行われる。

【0064】

算出部 15 は、例えば図 4 に示すように、第 1 文法情報「1、2、3」に含まれるクラス ID 毎に、各候補データ「明かり/1」、「を/2」、「つけて/3」のクラス ID を対応させ、センテンス「明かり/1」「を/2」「つけて/3」を生成する。このとき、例えば文法情報が「3、1、2」の場合、センテンスとして「つけて/3」「明かり/1」「を/2」が生成される。

20

【0065】

算出部 15 は、センテンスに含まれる各候補データ「明かり/1」、「を/2」、「つけて/3」、に対応する信頼度「0.982」、「1.000」、「0.990」を算出する。算出部 15 は、各候補データの文字列情報に対して 0.000 以上 1.000 以下の範囲で信頼度を算出する。

30

【0066】

算出部 15 は、例えば各センテンスに対して、優先度を示すランクを設定（図 4 ではランク 1～ランク 5）してもよい。ランクを設定することで、任意のランク下位にランク付けされたセンテンス（例えばランク 6 以下）を、評価対象から除外することができる。このため、後述する評価データとして選択される候補データの数を減らすことができ、処理速度の向上を図ることが可能となる。

【0067】

算出部 15 は、例えば内容の異なるセンテンスに同一の候補データが含まれる場合、各候補データにはそれぞれ異なる信頼度を算出してもよい。例えば、第 1 センテンスに含まれる各候補データ「明かり/1」、「を/2」、「つけて/3」に対応する信頼度「0.982」、「1.000」、「0.990」が算出された場合、第 2 センテンスに含まれる各候補データ「明かり/1」、「を/2」、「弾いて/3」に対応する信頼度「0.942」、「1.000」、「0.023」が算出される。すなわち、同一の候補データ「明かり」であっても、センテンスの内容や組み合わせの順序によって、異なる信頼度が算出されてもよい。

40

【0068】

例えば信頼度として、各文字列情報に対して予め設定された値が用いられてもよい。この場合、予め設定された値は、例えば文法データベースに記憶される。上記のほか、例えば信頼度として、検出部 14 において検出された候補データ（又は候補データの文字列情報）の種類及び数に応じた相対値が用いられてもよい。例えば検出部 14 において検出さ

50

れた複数の候補データにおいて、1つのクラスIDが付与された文字列情報の種類が多くなるにつれて、各文字列情報に対して低い信頼度を算出するようにしてもよい。

【0069】

< 選択部16 >

選択部16は、信頼度に基づき、複数の候補データから評価データを選択する。選択部16は、例えば複数の候補データのうち、クラスID毎に最も高い信頼度が算出された候補データを、評価データとして選択する。例えば選択部16は、同じクラスID「3」における候補データ「つけて/3/0.990」、「弾いて/3/0.023」のうち、最も高い信頼度を有する候補データ「つけて/3/0.990」を評価データとして選択する。なお、選択部16は、例えば1つのクラスIDに対して複数の候補データを、評価データとして選択してもよい。この場合、後述する生成部17において、複数の候補データから1つ選択するようにしてもよい。

10

【0070】

< 生成部17 >

生成部17は、評価データに基づき、認識情報を生成する。生成部17は、例えば評価データをテキストや数値等のデータ形式に変換し、認識情報として生成するほか、例えば評価データを音声データ形式や、制御装置3から音声を発生させるための制御データ形式に変換し、認識情報として生成してもよい。なお、評価データに基づくテキストや数値等のデータ形式、音声データ形式、又は制御データ形式に変換する方法は、公知の技術を用いることができ、必要に応じて各データ形式を蓄積したデータベース等を用いてもよい。

20

【0071】

生成部17は、例えば指定部17aと、比較部17bとを有してもよい。指定部17aは、参照データベースを参照し、参照センテンスのうち、評価データに対応する第1参照センテンスを指定する。指定部17aは、例えば評価データとして「明かり/1」、「を/2」、「つけて/3」が選択された場合、図5に示す第1参照センテンスを指定する。この場合、第1参照センテンスに含まれる各文字列情報(第1文字列情報)として、評価データに含まれる候補データと等しい文字列が指定される。

【0072】

比較部17bは、評価データに対応する信頼度と、第1文字列情報に付与された閾値(第1閾値)とを比較する。比較部17bは、例えば評価データ「明かり」、「を」、「つけて」の信頼度「0.982」、「1.000」、「0.990」が、第1文字列情報「明かり」、「を」、「つけて」の第1閾値「0.800」、「0.900」、「0.880」以上か否かを比較する。この場合、生成部17は、比較結果に基づいて認識情報を生成する。例えば信頼度が第1閾値以上の場合に、生成部17が認識情報を生成してもよい。例えば信頼度が第1閾値以上の場合と、第1閾値未満の場合とに応じて、生成部17が異なる生成情報を生成してもよい。

30

【0073】

< 出力部18 >

出力部18は、発話情報を出力し、例えば認識情報を出力してもよい。出力部18は、I/F105を介して制御装置3等に発話情報や認識情報を出力する。出力部18は、例えばI/F107を介して出力部分109に発話情報や認識情報を出力してもよい。出力部18は、発話情報や認識情報のほか、例えばI/F105、I/F107を介して各種情報(データ)を制御装置3等に出力する。

40

【0074】

出力部18は、例えば認識情報に基づき、発話情報(例えば発話情報に含まれる第1発話情報)を出力する。出力部18は、例えば出力した第1発話情報に対して生成された認識情報と、出力した第1発話情報に紐づく第1返答情報とを比較する。出力部18は、比較結果に基づき、認識情報に対し、第1返答情報と同等であるか否かを判定し、判定結果に応じて出力する発話情報を選択する。

【0075】

50

出力部 18 は、公知の技術を用いて、認識情報と返答情報とを比較する。比較方法として、例えば認識情報と、返答情報とが完全一致するか否かを比較するほか、一部一致、又は類似するか否かを比較してもよい。また、認識情報が数値の場合は、返答情報を閾値として、認識情報と返答情報との比較が行われてもよい。なお、比較結果と、判定結果（認識情報が第 1 返答情報と同等であるか否か）との紐づけは、予め任意に設定することができる。

【0076】

< 反映部 19 >

反映部 19 は、認識情報又は発話情報を評価した利用者等の評価結果を取得し、参照データベースの閾値に反映させる。反映部 19 は、例えば認識情報に対して評価結果が悪い場合（すなわち、音声データに対して得られる認識情報が、利用者等の発する音声や認識と乖離している場合）、閾値を変更させることで、認識情報の改善を図る。このとき、例えば公知の機械学習方法等を用いて、評価結果を閾値に反映させてもよい。

10

【0077】

< 收音装置 2 >

收音装置 2 は、公知のマイクに加え、例えば DSP (digital signal processor) を有してもよい。收音装置 2 が DSP を有する場合、收音装置 2 は、マイクによって收音した音声信号に対し PCM 等のパルス変調したデータ（音声データ）を生成し、音声認識装置 1 に送信する。

【0078】

收音装置 2 は、例えば音声認識装置 1 と直接接続されるほか、例えば公衆通信網 4 を介して接続されてもよい。なお、收音装置 2 が收音機能のみを有する場合、音声認識装置 1 がパルス変調したデータを生成してもよい。

20

【0079】

< 制御装置 3 >

制御装置 3 は、発話情報に基づき、音声を出力する。制御装置 3 として、スピーカ等の出力装置が用いられ、例えば收音装置 2 と一体に設けられた装置が用いられてもよいほか、スピーカ機能を搭載した任意の端末が用いられてもよい。制御装置 3 は、例えば音声認識装置 1 と直接接続されるほか、例えば公衆通信網 4 を介して接続されてもよい。

【0080】

制御装置 3 は、例えば発話情報に紐づく音声データが記憶されたデータベースを備えてもよい。この場合、発話データベースに記憶された発話情報と、音声データとの紐づけが、予め設定される。制御装置 3 は、例えば認識情報に基づき、音声を出力してもよい。この場合、制御装置 3 は、例えば上述した発話データベースが記憶された記憶部を備えてもよい。

30

【0081】

< 公衆通信網 4 >

公衆通信網 4 は、音声認識装置 1 が通信回路を介して接続されるインターネット網等である。公衆通信網 4 は、いわゆる光ファイバ通信網で構成されてもよい。また、公衆通信網 4 は、有線通信網には限定されず、無線通信網等の公知の通信網で実現してもよい。

40

【0082】

< サーバ 5 >

サーバ 5 には、上述した各種情報が記憶される。サーバ 5 には、例えば公衆通信網 4 を介して送られてきた各種情報が蓄積される。サーバ 5 には、例えば保存部 104 と同様の情報が記憶され、公衆通信網 4 を介して音声認識装置 1 と各種情報の送受信が行われてもよい。すなわち、音声認識装置 1 は、保存部 104 の代わりにサーバ 5 を用いてもよい。特に、サーバ 5 に上述した各データベースが保存され、音声認識装置 1 には、サーバ 5 に記憶された各データベースの少なくとも一部が保存されるようにしてもよい。この場合、サーバ 5 を用いて音声認識装置 1 に保存された各データベースを適宜更新することで、音声認識装置 1 における更新機能や蓄積するデータ容量を最小限に抑えることができる。こ

50

のため、音声認識装置 1 を公衆通信網 4 に常時接続しない状態で利用することができ、更新が必要な場合のみ公衆通信網 4 に接続するように用いることができる。これにより、音声認識装置 1 の利用先を大幅に拡大させることができる。

【 0 0 8 3 】

< ユーザ端末 6 >

ユーザ端末 6 は、例えば音声認識システム 1 0 0 の利用者等が保有する端末を示す。ユーザ端末 6 として、主に携帯電話（携帯端末）が用いられ、それ以外ではスマートフォン、タブレット型端末、ウェアラブル端末、パーソナルコンピュータ、I o T（Internet of Things）デバイス等の電子機器のほか、あらゆる電子機器で具現化されたものが用いられてもよい。ユーザ端末 6 は、例えば公衆通信網 4 を介して音声認識装置 1 と接続されるほか、例えば音声認識装置 1 と直接接続されてもよい。利用者等は、例えばユーザ端末 6 を介して音声認識装置 1 から認識情報を取得するほか、例えば収音装置 2 の代わりにユーザ端末 6 を用いて音声を収音させてもよい。

10

【 0 0 8 4 】

（音声認識システム 1 0 0 の動作の一例）

次に、本実施形態における音声認識システム 1 0 0 の動作の一例について説明する。図 6（a）は、本実施形態における音声認識システム 1 0 0 の動作の一例を示すフローチャートである。

【 0 0 8 5 】

< 第 1 出力手段 S 1 0 0 >

まず、発話情報に含まれる第 1 発話情報を入力する（第 1 出力手段 S 1 0 0）。出力部 1 8 は、例えば記憶部 1 3 を介して保存部 1 0 4 から発話データベースに記憶された発話情報を取り出す。出力部 1 8 は、発話情報から特定の発話情報（以下、第 1 発話情報とする）を選択し、制御装置 3 に出力する。制御装置 3 は、第 1 発話情報に基づき音声を出力する。

20

【 0 0 8 6 】

出力部 1 8 が選択する第 1 発話情報は、予め利用者等により設定されてもよい。第 1 出力手段 S 1 0 0 は、例えば制御装置 3 や入力部分 1 0 8 等から入力された起動指示情報（例えばデータや音声等）に基づき実行されてもよく、実行のタイミングや起動指示情報については任意に設定でき、起動指示情報に基づいて第 1 発話情報が選択されてもよい。

30

【 0 0 8 7 】

なお、第 1 出力手段 S 1 0 0 では、出力部 1 8 から第 1 発話情報が出力されるほか、例えば制御装置 3 が第 1 発話情報を入力してもよい。この場合、第 1 発話情報として、利用者に報知される音声データが用いられ、音声認識装置 1 を介さずに、第 1 出力手段 S 1 0 0 が行われる。

【 0 0 8 8 】

< 取得手段 S 1 1 0 >

次に、第 1 発話情報を認識した利用者から発せられた音声に基づき、音声データを取得する（取得手段 S 1 1 0）。取得部 1 1 は、収音装置 2 等により収音された音声に基づき、音声データを取得する。取得部 1 1 は、例えば記憶部 1 3 を介して保存部 1 0 4 に音声データを保存する。

40

【 0 0 8 9 】

< 抽出手段 S 1 2 0 >

次に、認識対象データを抽出する（抽出手段 S 1 2 0）。抽出部 1 2 は、例えば記憶部 1 3 を介して保存部 1 0 4 から音声データを取り出し、音声データに含まれる開始無音区間及び終了無音区間を、音素認識により抽出する。また、抽出部 1 2 は、開始無音区間と終了無音区間との間に挟まれた音素及び休止区間の配列を、音素認識により認識対象データとして抽出する。抽出部 1 2 は、例えば記憶部 1 3 を介して保存部 1 0 4 に認識対象データを保存する。なお、抽出部 1 2 は、一度に複数の音声データを取得してもよい。

【 0 0 9 0 】

50

抽出部 1 2 は、例えば 1 つの音声データから複数の認識データを抽出する。このとき、複数の認識データは、それぞれ異なる音素及び休止区間の配列を有する（例えば図 4 の配列 A ~ 配列 C）。抽出部 1 2 は、例えばそれぞれ異なる条件を設定するほか、例えば同一条件で設定したときにおけるパラツキの範囲内で、複数の認識データを抽出する。

【 0 0 9 1 】

なお、例えば休止区間が呼吸音及びリップノイズの少なくとも何れかを含むとき、抽出部 1 2 は、呼吸音及びリップノイズの少なくとも何れかを含む配列を、認識対象データとして抽出してもよい。

【 0 0 9 2 】

< 検出手段 S 1 3 0 >

次に、認識対象データに基づき、候補データを検出する（検出手段 S 1 3 0）。検出部 1 4 は、例えば記憶部 1 3 を介して保存部 1 0 4 から認識対象データを取出す。検出部 1 4 は、文字列データベースを参照し、認識対象データの有する配列に対応する音素情報を選択する。また、検出部 1 4 は、選択された音素情報に紐づく文字列情報及びクラス ID を候補データとして複数検出する。検出部 1 4 は、例えば記憶部 1 3 を介して保存部 1 0 4 に候補データを保存する。なお、認識対象データの有する配列は、例えば一对の休止区間の間における音素の配列を示し、一对の休止区間の間に他の休止区間が配列されてもよい。

【 0 0 9 3 】

< 算出手段 S 1 4 0 >

次に、各候補データに対応する信頼度を算出する（算出手段 S 1 4 0）。算出部 1 5 は、例えば記憶部 1 3 を介して保存部 1 0 4 から候補データを取出す。算出部 1 5 は、文法データベースを参照し、複数の候補データを文法情報に基づき組み合わせたセンテンスを生成する。また、算出部 1 5 は、センテンスに含まれる候補データ毎に対応する信頼度を算出する。算出部 1 5 は、例えば記憶部 1 3 を介して保存部 1 0 4 に各候補データ及び信頼度を保存する。算出部 1 5 として、例えば J u l i u s 等の公知の音声認識エンジンが用いられることで、センテンスの生成及び信頼度の算出が実現されてもよい。

【 0 0 9 4 】

算出部 1 5 は、文法データベースの文法情報の種類に応じて、複数のセンテンスを生成することができる。また、算出部 1 5 は、文法情報の種類を選択することで、状況に適した音声認識を高精度で実施することができる。

【 0 0 9 5 】

< 選択手段 S 1 5 0 >

次に、信頼度に基づき、評価データを選択する（選択手段 S 1 5 0）。選択部 1 6 は、例えば記憶部 1 3 を介して保存部 1 0 4 から候補データ及び信頼度を取出す。選択部 1 6 は、例えば複数の候補データのうち、クラス ID 毎に最も高い信頼度が算出された候補データを、評価データとして選択する。選択部 1 6 は、例えば記憶部 1 3 を介して保存部 1 0 4 に評価データを保存する。

【 0 0 9 6 】

< 生成手段 S 1 6 0 >

次に、評価データに基づき、認識情報を生成する（生成手段 S 1 6 0）。生成部 1 7 は、例えば記憶部 1 3 を介して保存部 1 0 4 から評価データを取出す。生成部 1 7 は、例えば上述した公知の技術を用いて評価データを任意のデータに変換し、認識情報として生成する。

【 0 0 9 7 】

生成手段 S 1 6 0 は、例えば図 6 (b) に示すように、指定手段 S 1 6 1 と、比較手段 S 1 6 2 とを有してもよい。

【 0 0 9 8 】

指定手段 S 1 6 1 は、評価データに対応する第 1 参照センテンスを指定する。指定部 1 7 a は、参照データベースを参照し、参照センテンスのうち、評価データに対応する第 1

10

20

30

40

50

参照センテンスを指定する。

【0099】

比較手段S162は、評価データに対応する信頼度と、第1参照センテンスに含まれる第1文字列情報に付与された第1閾値とを比較する。比較部17bは、例えば図4に示すように、評価データの信頼度が第1閾値以上の場合に、認識が正しいと判断してもよい。この後、比較部17bの判断（比較結果）に基づき、認識情報が生成される。なお、比較部17bにおいて評価データの信頼度が第1閾値未満となり、認識が誤っていると判断した場合、そのまま終了するか、抽出手段S120から再度実施するほか、例えば利用者等に再度音声を発するように促す認識情報を生成し、第2出力手段S170を実施してもよい。

10

【0100】

<第2出力手段S170>

その後、必要に応じて、認識情報に基づく発話情報を出力する（第2出力手段S170）。出力部18は、例えば発話データベースを参照し、第1出力手段S100で出力した第1発話情報に紐づく第1返答情報と、認識情報とを比較し、比較結果に基づき、認識情報に対して第1返答情報と同等であるか否かを判定する。出力部18は、判定結果に応じて発話データベースに記憶された発話情報から特定の発話情報を選択し、I/F105を介して、制御装置3に選択された発話情報を出力する。その後、制御装置3は、選択された発話情報に基づく音声を出力する。これにより、例えば利用者は、自ら発した音声に対する認識結果を知ることができる。

20

【0101】

第2出力手段S170は、例えば認識情報に対し、第1返答情報と同等であると判定した場合に、第1発話情報とは異なる第2発話情報を選択し、制御装置3に出力してもよい。この場合、制御装置3は、第1発話情報に基づく音声とは異なり、第2発話情報に基づく音声を出力する。

【0102】

また、第2出力手段S170は、例えば認識情報に対し、第1返答情報と同等ではないと判定した場合に、第1発話情報を選択し、制御装置3に出力してもよい。この場合、制御装置3は、第1出力手段S100と同様に、第1発話情報に基づく音声を出力する。

【0103】

また、第2出力手段S170では、例えば出力部18から制御装置3に対して認識情報が出力されてもよい。この場合、制御装置3は、取得した認識情報に基づき、発話データベースに記憶された第2発話情報等を選択し、選択された発話情報に紐づく音声を出力する。

30

【0104】

<反映手段S180>

なお、例えば認識情報又は選択された発話情報を評価した利用者等の評価結果を取得し、参照データベースの閾値に反映させてもよい（反映手段S180）。この場合、反映部19は、取得部11を介して利用者等が作成した評価結果を取得する。反映部19は、評価結果に含まれる評価値等に基づき、比較手段S162における比較の結果が改善（認識精度が向上）するように、閾値を変更する。

40

【0105】

なお、反映部19は、例えば参照データベースのほか、文字列データベース及び文法データベースの少なくとも何れかに評価結果を反映させてもよい。また、算出部15が評価結果に基づき、信頼度の算出に反映させてもよい。

【0106】

これにより、本実施形態における音声認識システム100の動作が終了する。

【0107】

なお、第1出力手段S100、及び第2出力手段S170では、例えば制御装置3の代わりに、出力部分109又はユーザ端末6に対して、発話情報を出力してもよい。この場

50

合、出力部分 109 又はユーザ端末 6 は、発話情報に基づき、音声を出力する。

【0108】

なお、例えば第 1 出力手段 S100 を実施せずに、取得手段 S110 から第 2 出力手段 S170 までを実施してもよい。この場合、第 2 出力手段 S170 では、例えば出力部 18 は、認識情報に基づき、発話データベースに記憶された返答情報から特定の第 1 返答情報を選択し、第 1 返答情報に紐づく第 1 発話情報を制御装置 3 等に出力する。このとき、出力部 18 は、例えば返答情報と、認識情報とを比較し、予め設定された条件（最も近いもの等）によって、第 1 発話情報を選択する。

【0109】

例えば音声認識システム 100 では、第 1 出力手段 S100、又は第 2 出力手段 S170 を実施したあと、任意の期間利用者から音声が発せられなかった場合、再度第 1 出力手段 S100、又は第 2 出力手段 S170 を実施するほか、音声認識システム 100 を終了してもよい。再度第 1 出力手段 S100、又は第 2 出力手段 S170 を実施するとき、例えば他の返答情報を変更するように設定することができ、例えば任意の回数を超えたら他の返答情報に変更するように設定することもできる。

【0110】

例えば音声認識システム 100 では、第 2 出力手段 S170 を実施したあと、取得手段 S110 ~ 第 2 出力手段 S170 を繰り返し実施してもよい。この場合、例えば第 2 出力手段 S170 において、特定の返答情報を繰り返し選択したとき、任意の回数を超えたら強制的に他の返答情報を選択してもよい。

【0111】

本実施形態における音声認識システム 100 によれば、抽出手段 S120 は、音素及び休止区間の配列を認識対象データとして抽出する。また、検出手段 S130 は、認識対象データの有する配列に対応する音素情報を選択し、候補データを検出する。このため、認識対象データにおける音素のみを考慮した配列に対して候補データを検出する場合に比べ、誤認識を低減させることができる。これにより、認識精度の向上を図ることが可能となる。

【0112】

また、本実施形態における音声認識システム 100 によれば、文字列データベースには、音素と休止区間との配列に対応する音素情報、及び音素情報に紐づく文字列情報が記憶される。このため、音素全体に対してパターンマッチングするために記憶するデータに比べて、データ容量の削減や、データ蓄積の簡易化を実現することができる。

【0113】

特に、音声認識システム 100 の利用される環境を踏まえて、文字列データベースに記憶される文字列情報を選択することで、データ容量の削減ができ、例えば公衆通信網 4 に接続する必要がなく、利用の幅を広げることができる。また、音声データの取得から認識情報を生成するまでの時間を大幅に短縮することができる。

【0114】

また、本実施形態における音声認識システム 100 によれば、第 1 出力手段 S100 は、第 1 発話情報を出力する。このため、第 1 発話情報に対する利用者から発せられた音声を、認識情報として生成することができる。これにより、発話情報に対する利用者の認識力を高精度に評価することが可能となる。

【0115】

また、本実施形態における音声認識システム 100 によれば、第 2 出力手段 S170 は、認識情報に基づき、第 2 発話情報、又は第 1 発話情報を出力する。このため、利用者の認識力を評価した結果を、発話として報知することができる。これにより、対話形式による音声認識を実現することが可能となる。

【0116】

また、本実施形態における音声認識システム 100 によれば、第 2 出力手段 S170 は、認識情報に対し、第 1 返答情報と同等であると判定した場合に、第 2 発話情報を出力す

10

20

30

40

50

る。このため、第1発話情報に対して利用者が発した音声、が、所望の内容である場合、第1発話情報とは異なる第2発話情報を報知することができる。これにより、対話形式のバリエーションを増やすことが可能となる。

【0117】

また、本実施形態における音声認識システム100によれば、第2出力手段S170は、認識情報に対し、第1返答情報と同等ではないと判定した場合に、第1発話情報を出力する。このため、第1発話情報に対して利用者が発した音声、が、所望の内容ではない場合、第1発話情報を再度報知することができる。これにより、発話練習等のような正確な発話情報を復唱する必要がある場合や、暗号認識等のような正確な認識を得る必要がある場合にも利用することが可能となる。

10

【0118】

また、本実施形態における音声認識システム100によれば、抽出手段S120は、1つの音声データから複数の認識対象データを抽出する。このため、音素及び休止区間の配列にバラつきが発生するような音声データを取得した場合においても、認識精度の低下を抑制することができる。これにより、認識精度のさらなる向上が可能となる。

【0119】

また、本実施形態における音声認識システム100によれば、算出手段S140は、センテンスを複数生成する。すなわち、候補データを組み合わせるパターンが複数存在する場合においても、全てのパターンに対応するセンテンスを生成することができる。このため、例えばパターンマッチングの探索方法等に比べて、誤認識を低減させることができる。これにより、認識精度のさらなる向上が可能となる。

20

【0120】

また、本実施形態における音声認識システム100によれば、比較手段S162は、信頼度と、第1閾値とを比較する。このため、複数の候補データから相対的に選択された評価データに対し、閾値による判定も行うことで、誤認識をさらに低減させることができる。これにより、認識精度のさらなる向上が可能となる。

【0121】

また、本実施形態における音声認識システム100によれば、反映手段S180は、評価結果を閾値に反映させてもよい。このため、認識情報が、利用者の認識と乖離している場合、容易に改善を実施することができる。これにより、持続的な認識精度の向上を実現することができる。

30

【0122】

また、本実施形態における音声認識システム100によれば、休止区間は、呼吸音及びリップノイズの少なくとも何れかを含んでもよい。このため、音素のみでは判断し難い音声データの差異に対しても容易に判断でき、認識対象データを抽出することができる。これにより、認識精度のさらなる向上を図ることが可能となる。

【0123】

本実施形態における音声認識装置1によれば、抽出部12は、音素及び休止区間の配列を認識対象データとして抽出してもよい。また、検出部14は、認識対象データの有する配列に対応する音素情報を選択し、候補データを検出してもよい。このため、認識対象データにおける音素のみを考慮した配列に対して候補データを検出する場合に比べ、誤認識を低減させることができる。これにより、認識精度の向上を図ることが可能となる。

40

【0124】

また、本実施形態における音声認識装置1によれば、文字列データベースには、音素と休止区間との配列に対応する音素情報、及び音素情報に紐づく文字列情報が記憶されてもよい。このため、音素全体に対してパターンマッチングするために記憶するデータに比べて、データ容量の削減や、データ蓄積の簡易化を実現することができる。

【0125】

また、本実施形態における音声認識装置1によれば、出力部18は、第1発話情報を出力する。このため、第1発話情報に対する利用者から発せられた音声を、認識情報として

50

出力することができる。これにより、発話情報に対する利用者の認識力を高精度に評価することが可能となる。

【0126】

(音声認識システム100の構成の第1変形例)

次に、本実施形態における音声認識システム100の第1変形例について説明する。上述した実施形態と、第1変形例との違いは、生成部17が更新部17cを有する点である。なお、上述した構成と同様の構成については、説明を省略する。

【0127】

生成部17の有する更新部17cは、例えば図7に示すように、候補データ及び信頼度に基づき、参照データベースに記憶された閾値を更新する。すなわち、候補データ及び信頼度の内容に応じた値に、閾値を更新することができる。

10

【0128】

更新部17cは、例えば各クラスIDに紐づく複数の信頼度の平均値を算出する。更新部17cは、算出した平均値に基づき閾値を更新する。

【0129】

閾値を更新する場合、算出された平均値が閾値として用いられるほか、予め設定された係数を平均値にかけ合わせた値が、更新後の閾値として用いられてもよい。また、更新前の閾値に対して、係数を平均値にかけ合わせた値を四則演算した結果の値を更新後の閾値として用いられてもよい。

【0130】

候補データ及び信頼度の内容に基づき閾値を更新することで、例えば音声データにノイズ等が含まれ易い場合においても、音声データの品質に応じた閾値を設定することができる。また、1つのクラスIDに紐づく文字列情報が多数検出され、各文字列情報の信頼度が低い場合においても、全ての信頼度が閾値未満になることを防ぐことができる。

20

【0131】

更新部17cは、例えば各クラスIDに紐づく複数の信頼度のうち、最も低い信頼度を除いた平均値を算出してもよい。この場合、更新後の閾値は、更新前の閾値に比べて高くなる傾向を示す。これにより、誤認識を低減させることが可能となる。

【0132】

更新部17cは、例えば各クラスIDに紐づく複数の信頼度のうち、最も低い信頼度及び最も高い信頼度を除いた平均値を算出してもよい。この場合、更新後の閾値は、更新前の閾値に比べて低くなる傾向を示す。これにより、認識率を向上させることができる。また、更新前後における閾値の変動を抑制することができる。

30

【0133】

(音声認識システム100の動作の第1変形例)

次に本実施形態における音声認識システム100の第1変形例について説明する。図8(a)は、第1変形例における更新手段S163の一例を示すフローチャートである。

【0134】

図8(a)に示すように、上述した選択手段S150を実施したあと、複数の候補データ、及び複数の信頼度に基づき、参照データベースに記憶された閾値を更新する(更新手段S163)。更新部17cは、例えば記憶部13を介して保存部104から候補データ、信頼度、及び参照データベースを取出す。

40

【0135】

更新部17cは、例えば図7に示すように、ランク1、2、4に含まれるクラスID「1」に紐づく複数の信頼度「0.982」、「0.942」、「0.897」の平均値「0.940」を算出する。その後、更新部17cは、例えば算出した平均値に係数(例えば0.9)をかけ合わせた値「0.846」を、更新後の閾値として用いる。

【0136】

その後、上述した指定手段S161等を実施し、本実施形態における音声認識システム100の動作が終了する。

50

【0137】

本変形例によれば、更新手段 S 1 6 3 における更新部 1 7 c は、候補データ及び信頼度に基づき、閾値を更新する。このため、予め設定された閾値を常に用いる場合に比べて、取得する音声データにおける品質に応じた認識情報を生成することができる。これにより、利用できる環境の幅を広げることが可能となる。

【0138】

(音声認識システム 1 0 0 の動作の第 2 変形例)

次に本実施形態における音声認識システム 1 0 0 の第 2 変形例について説明する。上述した実施形態と、第 2 変形例との違いは、設定手段 S 1 9 0 を備える点である。なお、上述した構成と同様の構成については、説明を省略する。

10

【0139】

設定手段 S 1 9 0 は、例えば図 8 (b) に示すように、生成手段 S 1 6 0 の後に実施される。設定手段 S 1 9 0 は、認識情報に基づき、検出手段 S 1 3 0、算出手段 S 1 4 0、及び指定手段 S 1 6 1 の少なくとも何れかにおいて、参照する各データベースの内容を選択する。設定手段 S 1 9 0 の実施後、上述した第 2 出力手段 S 1 7 0、第 1 出力手段 S 1 0 0、取得手段 S 1 1 0 等が適宜実施される。例えば生成手段 S 1 6 0 において、生成部 1 7 は、生成した認識情報に基づき、検出部 1 4 が参照する文字列データベースの内容を選択する。文字列データベースの内容が選択されることで、特定の状況に応じた音声のみを認識することができる。これにより、認識精度をさらに向上させることが可能となる。

20

【0140】

ここで、「文字列データベースの内容」とは、文字列データベースに記憶された文字列情報、クラス ID、及び音素情報を示す。また、「内容を選択」とは、認識情報に基づいて、文字列データベースに記憶された文字列情報、クラス ID、及び音素情報のうち、一部を選択することを示す。他のデータベースの内容についても同様である。

【0141】

なお、「特定の状況」とは、例えば音楽を再生するための音声データを認識する状況(例えば、曲名や曲番号等のみを認識するミュージックモード)等を示す。例えばミュージックモードでは、曲名や曲番号等のような特定の音声を認識する必要があり、他の音声を認識する必要がない。このため、通常、文字列データベースを用いた場合には、特定の音声は他の音声と同様に評価されるため、認識できる可能性が著しく低下する事情がある。これに対し、本実施形態によれば、設定手段 S 1 9 0 において文字列データベース等の内容を選択することができ、特定の音声に対する認識精度を向上させることが可能となる。

30

【0142】

また、「特定の状況」とは、例えば音声を認識する順序が必要な状況(例えば起動モード)等を示す。例えば起動モードでは、音声を認識する順序によって、音声認識装置 1 から得られる情報を変える必要がある。このため、通常、文字列データベースを用いた場合には、順序を認識できない、又は、順序を含む長い文字列等を記憶させる必要があり、汎用的な利用に向かない事情がある。これに対し、本実施形態によれば、設定手段 S 1 9 0 において、認識した音声の順序に応じて文字列データベース等の内容を選択することができる。このため、文字列データベースに記憶させるデータ容量を最小限に抑えることができ、汎用的な利用を実現することが可能となる。

40

【0143】

文字列データベースには、文字列情報、音素情報、及びクラス ID が、認識情報の内容に応じて選択できる状態で記憶される。また、文法データベースには、文法情報が、認識情報の内容に応じて選択できる状態で記憶される。例えば認識情報の内容が「ミュージックモードの設定」に関する情報の場合、文字列データベースに記憶された情報のうち、曲名や曲番号に限定した文字列情報等が選択されるようにでき、文法データベースに記憶された情報のうち、文字列データベースにおいて選択されたクラス ID の配列順序を示す文法情報が選択されるようにできる。また、選択された文字列データベース及び文法データベースに応じて、参照データベースに記憶された内容が選択されるようにできる。

50

【 0 1 4 4 】

例えば設定手段 S 1 9 0 において、生成部 1 7 は、評価データに基づき生成された認識情報（例えばミュージックモードに関する情報）に基づき、検出部 1 4 が参照する文字列データベースの内容を選択する。その後、適宜各出力手段 S 1 0 0、S 1 7 0 を実施し、取得手段 S 1 1 0 及び抽出手段 S 1 2 0 を実施して認識対象データを抽出した後、検出手段 S 1 3 0 において、検出部 1 4 は、文字列データベースのうち、「ミュージックモード」に特化した音素情報、文字列情報、及びクラス ID を選択して参照する。このため、設定手段 S 1 9 0 を実施しない場合に比べて、特定の内容に対する音素情報等に限定することができる。これにより、認識精度を飛躍的に向上させることが可能となる。

【 0 1 4 5 】

例えば設定手段 S 1 9 0 では、認識情報の内容毎に異なるクラス ID を選択してもよい。文字列データベースには、例えばクラス ID 「 1 ~ 1 0 」、各クラス ID が付与された文字列情報、及び文字列情報に紐づく音素情報が記憶され、そのうち曲名や曲番号に関する文字列情報には、クラス ID 「 5 ~ 1 0 」を付与して記憶される。この場合、設定手段 S 1 9 0 において、「ミュージックモードの設定」に関する認識情報が生成された場合、設定手段 S 1 9 0 においてクラス ID 「 5 ~ 1 0 」を選択することで、クラス ID 「 5 ~ 1 0 」が付与された曲名に関する文字列情報のみを、検出手段 S 1 3 0 で検出することができる。

【 0 1 4 6 】

なお、例えば設定手段 S 1 9 0 において、生成部 1 7 は、認識情報に基づき、算出部 1 5 が参照する文法データベースの内容、及び、指定部 1 7 a が参照する参照データベースの内容を選択してもよい。この場合、上述した検出手段 S 1 3 0 と同様に、算出手段 S 1 4 0 において、算出部 1 5 は、文法データベースのうち、「ミュージックモード」に特化したセンテンスの生成、及びセンテンスに含まれる候補データ毎に対応する信頼度を算出することができる。また、指定手段 S 1 6 1 において、指定部 1 7 a は、参照データベースのうち、「ミュージックモード」に特化した第 1 参照センテンスを指定することができる。これにより、特定の状態に応じた音声のみを認識することができ、認識精度を飛躍的に向上させることが可能となる。

【 0 1 4 7 】

例えば図 9 に示すように、本実施形態における音声認識システム 1 0 0 において、特定の音声データ（起動ワード）を予めデータベースに記憶させ、起動ワードを認識したときに、設定手段 S 1 9 0 が実施されるようにしてもよい。例えば特定の音声データ（図 9 では起動ワード A）を認識した場合、設定手段 S 1 9 0 において、生成部 1 7 は、認識情報に基づき、文字列データベースに含まれる文字列データベース A を選択する。これにより、その後の検出手段 S 1 3 0 において、検出部 1 4 は、文字列データベース A を参照して候補データを検出する。このため、設定手段 S 1 9 0 を実施する前に比べて、認識する音声データの種類を限定することができる。

【 0 1 4 8 】

なお、例えば図 9 の文字列データベース A に示すように、文字列データベースが起動ワードを複数記憶してもよい（図 9 では起動ワード B - 1、起動ワード B - 2）。この場合、認識した特定の音声データに応じて、選択される文字列データベースの内容を変えることができる。また、起動ワードに応じて、制御装置 3 等から出力する音声を変更してもよい。この場合、起動ワードを認識したときの認識情報に基づき、適切な発話情報が選択されるように予め設定する。

【 0 1 4 9 】

例えば図 9 では、文字列データベース A を参照して起動ワード B - 1 を認識した場合、設定手段 S 1 9 0 において、生成部 1 7 は、認識情報に基づき、文字列データベースに含まれる文字列データベース B - 1 を選択する。また、起動ワード B - 2 を認識した場合、設定手段 S 1 9 0 において、生成部 1 7 は、認識情報に基づき、文字列データベースに含まれる文字列データベース B - 2 を選択する。このように、複数の起動ワードを用いるこ

10

20

30

40

50

とにより、認識精度の低下を抑制した状態で、認識できる音声の種類を増加させることが可能となる。

【0150】

また、例えば設定手段S190において文字列データベースを選択したあと、一定期間に取得手段S110が実施されない場合、文字列データベースの選択を解除してもよい。この場合、一定期間経過後に取得手段S110が実施される際、選択前の文字列データベースを参照して検出手段S130が実施される。このため、誤って発した音声を認識した場合や、誤認識した場合におけるやり直しを円滑に実施することができる。なお、文字列データベースの選択を解除する起動ワードを設定し、予め文字列データベースに記憶してもよい。

10

【0151】

なお、設定手段S190において各データベースの内容を選択したあと、例えば第2出力手段S170において、出力部18が各データベースの内容を選択(変更)した旨を報知できる発話情報を出力するようにしてもよい。

【0152】

本変形例によれば、設定手段S190は、認識情報に基づき、参照する文字列データベースの内容を選択する。このため、特定の状況に応じた音声のみを認識することができる。これにより、認識精度のさらなる向上を図ることが可能となる。

【0153】

(取得手段S110の変形例)

次に、本実施形態における取得手段S110の変形例について説明する。上述した実施形態と、本変形例との違いは、取得部11が条件情報を取得する点である。なお、上述した構成と同様の構成については、説明を省略する。

20

【0154】

取得手段S110において取得部11は、音声データが生成される条件を示す条件情報を取得する。条件情報は、例えば図10に示すように、環境情報と、雑音情報と、收音装置情報と、利用者情報と、音特性情報とを有する。なお、上述した設定手段S190と同様に、例えば検出部14は、条件情報に基づき、参照する文字列データベース及び文法データベースの少なくとも何れかの内容を選択してもよい。また、例えば出力部18は、条件情報に基づき、参照する発話データベースから選択される発話情報の選択範囲を設定してもよい。また、例えば反映部19は、参照データベースの閾値の更新に、条件情報を用いてもよい。

30

【0155】

条件情報は、例えば收音装置2により生成されるほか、例えば利用者等が予め生成してもよい。例えば取得部11は、音声データの一部を条件情報として取得してもよい。例えば取得部11は、入力部分108等から入力された情報に基づき、条件情報を取得してもよい。この場合、例えば最初に取得手段S110が実施されたあと、上述した第1出力手段S100等が実施されてもよい。これにより、音声認識システム100において実施する各手段に、条件情報を反映させることができる。

【0156】

環境情報は、收音装置2の設置された環境に関する情報を有し、例えば屋外、屋内の広さ等を示す。環境情報を用いることで、例えば屋内における音声の反射条件等を考慮することができる。抽出される認識対象データ等の精度を高めることができる。

40

【0157】

雑音情報は、收音装置2が收音し得る雑音に関する情報を有し、例えば利用者等以外の音声、空調音等を示す。雑音情報を用いることで、音声データに含まれる不要なデータを予め除去でき、抽出される認識対象データ等の精度を高めることができる。

【0158】

收音装置情報は、收音装置2の種類、性能等に関する情報を有し、例えばマイクの数、マイクの種類等も含まれる。收音装置情報を用いることで、音声データが生成された状況

50

に対応したデータベースの選択等ができ、音声認識の精度を高めることができる。

【0159】

利用者情報は、利用者等の人数、国籍、性別等に関する情報を有する。音特性情報は、音声の音量、音圧、癖、活舌の状態等に関する情報を有する。利用者情報を用いることで、音声データの特徴を予め限定することができ、音声認識の精度を高めることができる。

【0160】

利用者情報は、例えば利用者の趣味や興味のあることに関する情報を有してもよい。例えば出力部18が、利用者情報に基づき、選択する発話情報の選択範囲を設定した場合、利用者の趣味等にあわせた音声を出力することができる。この場合、第1出力手段S100～第2出力手段S170を繰り返すことで、利用者毎に適した音声を明確に出力することができる。

10

【0161】

本変形例によれば、取得手段S110は、条件情報を取得する。すなわち、取得手段S110は、音声データを取得する際の周辺環境、音声データに含まれる雑音、音声を選択する收音装置2の種類等の各種条件を、条件情報として取得する。このため、条件情報に応じた各手段や各データベースの設定を実施することができる。これにより、利用される環境等に関わらず、認識精度の向上を図ることが可能となる。

【0162】

また、本変形例によれば、検出手段S130は、条件情報に基づき、参照する文字列データベースの内容を選択する。このため、文字列データベースには、条件情報毎に異なる文字列情報等を記憶させておくことで、条件情報毎に適した候補データを検出することができる。これにより、条件情報毎における認識精度の向上を図ることが可能となる。

20

【0163】

また、本変形例によれば、第1出力手段S100及び第2出力手段S170の少なくとも何れかは、条件情報に基づき、選択する発話情報の選択範囲を設定する。このため、利用者毎に適した音声を出力することができる。これにより、用途の拡大を実現することが可能となる。なお、条件情報を生成する取得手段S110を、上述した設定手段S190と組み合わせることで、利用者に適した音声を出力する精度をさらに向上させることができる。

【0164】

また、本変形例によれば、条件情報を取得する取得手段S110を実施したあと、第1出力手段S100～第2出力手段S170等を繰り返すことで、利用者が興味のある質問等を積み重ねることができ、利用者毎に適した内容の発話情報が選択されるようにすることができる。

30

【0165】

(文字列データベースの変形例)

次に、本実施形態における文字列データベースの変形例について説明する。上述した実施形態と、本変形例との違いは、類似文字列情報等が文字列データベースに記憶される点である。なお、上述した構成と同様の構成については、説明を省略する。

【0166】

文字列データベースには、例えば図11に示すように、予め取得された類似文字列情報(図11では類似文字列情報A～C)と、類似クラスID(図11では類似クラスID-A～C)とが記憶される。類似文字列情報には、上述した文字列情報と同様に、音素情報が紐づけられる。類似クラスIDは、類似文字列情報に付与される。

40

【0167】

類似文字列情報は、音声認識の対象として設定された文字列情報に対して誤認識する可能性のある文字列情報を示し、利用者等が予め設定する。類似文字列情報は、文字列情報に近い音素情報を有し、例えば検出手段S130において、誤認識し易い文字列情報を示す。例えば文字列情報として「こうないかんきょう」が文字列データベースに記憶されている場合、類似文字列情報として「こうないえん」、「こうさてん」、「こうとくてん」等

50

の「こうないかんきょう」として誤認識する可能性のある文字列情報が記憶される。

【0168】

類似クラスIDは、文字列情報に付与されるクラスIDとは異なるIDが付与され、図11では、文字列情報に付与されたクラスIDが「1」に対して、類似文字列情報に付与された類似クラスIDは「9999」である。

【0169】

文字列データベースに類似文字列情報及び類似クラスIDが記憶されることで、例えば類似文字列情報が候補データに含まれた場合においても、容易に排除することが可能となる。これにより、誤認識に伴う認識情報の生成を抑制することが可能となる。

【0170】

類似文字列情報が候補データに含まれる場合、例えば比較手段S162において排除することができる。例えば図11に示すように、参照データベースに記憶された類似文字列には、信頼度の上限を超える閾値(> 1.000)が予め付与される。このため、評価データに類似文字列情報が含まれる場合においても、確実に排除することができる。

【0171】

なお、上述した閾値の設定は一例であり、類似文字列情報を排除できる値であれば任意である。また、例えば算出手段S140、選択手段S150等において、類似クラスIDを排除するように予め設定してもよい。

【0172】

上記のほか、例えば方言、数字、又は類語等のように、認識したい内容に対して複数の表現が存在する場合がある。この場合、類似文字列情報を、認識したい文字列情報と同等の閾値を設定することで、複数の文字列に対して等しい認識情報を生成することができる。

【0173】

本変形例によれば、文字列データベースには、類似文字列情報と、類似クラスIDとが記憶される。このため、類似文字列情報が候補データに含まれた場合においても、容易に排除することができる。また、方言や数字等の僅かに異なる音声であっても、同一の内容として認識させることができる。これらにより、認識精度のさらなる向上を図ることが可能となる。

【0174】

(参照データベースの変形例)

次に、本実施形態における参照データベースの変形例について説明する。上述した実施形態と、本変形例との違いは、参照データベースに記憶された情報の内容が異なる点である。なお、上述した構成と同様の構成については、説明を省略する。

【0175】

参照データベースには、例えば図12に示すように、予め取得された過去の評価データ、過去の評価データに紐づく参照センテンス、及び過去の評価データと参照センテンスとの間における連関度が記憶される。

【0176】

生成部17は、例えば参照データベースを参照し、過去の評価データのうち、評価データに対応する第1評価データ(図12の「過去の評価データ」内の破線枠)を選択する。その後、生成部17は、参照センテンスのうち、第1評価データに対応する第1参照センテンス(図12の「参照センテンス」内の破線枠)、を取得する。また、生成部17は、連関度のうち、第1評価データと第1参照センテンスとの間における第1連関度(図12の「65%」等)を取得する。なお、第1評価データ及び第1参照センテンスは、複数のデータを含んでもよい。

【0177】

生成部17は、第1連関度の値に基づき、認識情報を生成する。生成部17は、例えば第1連関度と、予め取得された閾値と比較し、閾値を上回る第1連関度に紐づく第1参照

10

20

30

40

50

センテンスを参考に、認識情報を生成する。

【0178】

過去の評価データとして、評価データと一部一致又は完全一致する情報が選択されるほか、例えば類似（同一概念等を含む）する情報が用いられる。評価データ及び過去の評価データが複数の文字列間の組み合わせで示される場合、例えば、名詞 - 動詞、名詞 - 形容詞、形容詞 - 動詞、名詞 - 名詞の何れかの組み合わせが用いられる。

【0179】

連関度（第1連関度）は、例えば百分率等の3段階以上で示される。例えば参照データベースがニューラルネットワークで構成される場合、第1連関度は、選択された過去の評価対象情報に紐づく重み変数を示す。

【0180】

上述した参照データベースを用いる場合、3段階以上に設定されている連関度に基づいて、音声認識を実現できる点に特徴がある。連関度等は、例えば0～100%までの数値で記述することができるが、これに限定されるものではなく3段階以上の数値で記述できればいかなる段階で構成されていてもよい。

【0181】

このような連関度等に基づいて、評価データに対する認識情報の候補として選ばれる第1参照センテンスにおいて、連関度等の高い又は低い順に第1参照センテンスを選択することが可能となる。このように連関度の順に選択することで、状況に見合う可能性の高い第1参照センテンスを優先的に選択することができる。他方、状況に見合う可能性の低い第1参照センテンスも除外せずに選択できるため、廃棄対象とせずに認識情報の候補として選択することが可能となる。

【0182】

上記に加え、例えば連関度等が1%のような極めて低い評価も見逃すことなく選択することができる。すなわち、連関度等が極めて低い値であっても、僅かな兆候として繋がっていることを示しており、過度の廃棄対象の選択や誤認を抑制することが可能となる。

【0183】

上述した実施形態における音声認識システム100及び音声認識装置1は、例えば以下の用途に用いることができる。

【0184】

<リハビリ>

音声認識システム100及び音声認識装置1は、例えばリハビリの分野に用いられてもよい。

【0185】

<<認知症リハビリ、脳活性トレーニング>>

例えば、上述した第1出力手段S100又は第2出力手段S170（出力部18）を備えることで、認知症リハビリや、脳活性トレーニング等に用いることができる。例えば、発話データベースに認知症リハビリや、脳活性トレーニング等に適した発話情報を記憶させることで、認知症患者や脳活性トレーニングを対象者とした音声認識システム100及び音声認識装置1の利用が可能となる。

【0186】

<メディカルチェック>

音声認識システム100及び音声認識装置1は、例えばメディカルチェックの分野に用いられてもよい。

【0187】

<<精神状態チェック>>

例えば、車両等の運転者に対して、運転前の精神状態チェックに用いることができる。運転前において、運転者の音声に対応する認識情報を生成し、通常の状態時に生成された認識情報と比較する。これにより、運転に適した精神状態が否かの評価を、定量的に実施することができる。特に、上述した第1出力手段S100又は第2出力手段S170

10

20

30

40

50

(出力部18)を備えることで、対話形式や質問形式の精神状態チェックを実現することが可能となる。なお、運転前以外にも、勤務時間前や、病院等において用いてもよい。

【0188】

<<音声リラクゼーション>>

例えば、利用者の音声に対応する認識情報に基づき、音楽を流してもよい。

【0189】

<<セキュリティチェック>>

例えば、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、銀行ATMやマンションのオートロック等のセキュリティチェックに用いることができる。特に、上述した利用者情報と組み合わせることで、セキュリティレベルを大幅に向上させることが可能となる。

10

【0190】

<<ゲーム>>

例えば、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、ゲームと連動した利用が可能となる。例えば音声認識装置1を、利用者がゲームに使用するコントローラ等に接続させることで、認識情報に基づきコントローラ等の振動や発光等を容易に実現することができる。特に、ゲーム内又は利用者が発する音声を、ゲームの進行に対応して取得できるようにすることで、タイミングを考慮したコントローラ等の駆動を実現することができる。これにより、利用者の没入感を高めることが可能となる。

20

【0191】

<<教育補助>>

例えば、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、発声練習、クイズ、演算練習等の教育補助として用いることができる。特に、利用者との対話形式で用いることで、利用者の認識力やコミュニケーション能力を養うことが可能となる。

【0192】

<接客現場>

音声認識システム100及び音声認識装置1は、例えば接客現場に用いられてもよい。

【0193】

<<挨拶の評価>>

例えば、接客業に必要とされる挨拶の評価に用いることができる。例えば模範的な挨拶に対する乖離度を、認識情報として生成することが可能となる。特に、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、接客のシミュレーション等を実現することが可能となり、接客態度等を定量的に評価することが可能となる。

30

【0194】

<<注文確認>>

例えば、飲食店等における注文確認に用いることができる。客席等に収音装置2を設置することで、顧客が注文した内容を認識情報として生成し、店員等に報知することができる。特に、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、顧客に追加の注文を提案したり、お勧め商品を紹介したりすることができる。これにより、売上向上に繋げることが可能となる。

40

【0195】

<業務効率化>

音声認識システム100及び音声認識装置1は、例えば業務効率化を図るために用いられてもよい。

【0196】

<<環境ノイズ対策>>

例えば、工事現場や、人混み等の環境ノイズ対策として用いることができる。例えば上

50

述した雑音情報等の環境情報を取得することで、環境ノイズを容易に除去することが可能となる。

【0197】

<<報告書作成>>

例えば、生成された認識情報に基づく任意の報告書形式のデータを生成して出力することで、報告書等の文書作成を容易に実現することが可能となる。この場合、例えば変換されたテキストデータを、ユーザ端末6等に送信してもよい。また、複数の認識情報をまとめてテキストデータとして変換して出力してもよい。この場合、例えばフィールドワーク本来の業務をしながらハンズフリーの操作ができるとともに、容易に報告書等を作成することができる。これにより、フィールドワーカーの負担となる報告書業務の効率化や、報告内容の品質向上を図ることが可能となる。

10

【0198】

例えば、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備えることで、報告書の内容を誘導することもできる。例えば第1出力手段S100において、出力部18は、任意の報告書の入力フォーマットに基づく第1発話情報を出力し、利用者に報告すべき内容(例えば「現在の場所は?」、「担当者は?」、「報告内容は?」等)を報知する。その後、取得手段S110において、取得部11は、報知された内容に対応する利用者の回答内容を含む音声(例えば「東京都」、「山〇男」、「トンネルの点検」等)に基づく音声データを取得する。その後、生成手段S160において、音声に対応して生成された認識情報に基づき、文字列(文章)を含む報告書のデータ(例えばワード、エクセル等)を生成する。生成された報告書のデータは、例えば監督者等の有するユーザ端末6等に送信されてもよい。

20

【0199】

なお、第1出力手段S100~生成手段S160を複数繰り返し、複数の文字列を含む1つの報告書のデータが生成されてもよい。また、報告書のデータには、例えば利用者が有するユーザ端末6、又は音声認識装置1に搭載された公知の撮像装置(カメラ)等を用いて撮像された画像データが含まれてもよい。この場合、例えば第1出力手段S100により、利用者に対して対象物等の撮像を誘導する内容(例えば「修繕前(後)の写真を撮影してください」等)を、利用者に報知できるようにしてもよい。

【0200】

上記に加え、例えば最初に取得手段S110を実施し、取得部11が上述した条件情報を取得するようにしてもよい。この場合、例えば工事現場等の騒音環境に関する環境情報を予め取得することができ、音声認識精度の低下を抑制することが可能となる。なお、例えば公知のノイズキャンセリング機能を備えた収音装置2等を用いて、騒音環境の影響を低減させてもよい。

30

【0201】

<<指差喚呼補助>>

例えば、作業現場等の指差喚呼補助として用いられてもよい。特に、指差喚呼の音声に対応する認識情報が生成されるため、適切な指差喚呼を行ったか否かを判定することができる。これにより、指差喚呼の怠りによる事故等の発生を抑制することが可能となる。また、指差喚呼の音声に対応する認識情報を生成し、保存することができるため、事故等が発生した場合における原因究明や、再発防止の検討を容易に実現することが可能となる。

40

【0202】

<<無人搬送車の制御>>

例えば、無人搬送車(AGV:Automated Guided Vehicle)の制御に用いられてもよい。この場合、上述した制御装置3をとして、無人搬送車が用いられる。これにより、パソコン等による手入力操作を行うことなく、無人搬送車の制御ができ、作業効率の向上を図ることが可能となる。特に、上述した第1出力手段S100又は第2出力手段S170(出力部18)を備える場合、例えば無人搬送車への指示確認を復唱させることで、誤動作を防ぐことができる。これにより、誤動作に伴う作業効率の低下を抑制することが可能と

50

なる。

【 0 2 0 3 】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

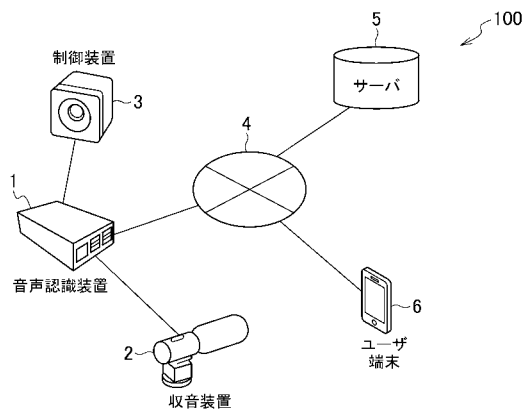
【 符号の説明 】

【 0 2 0 4 】

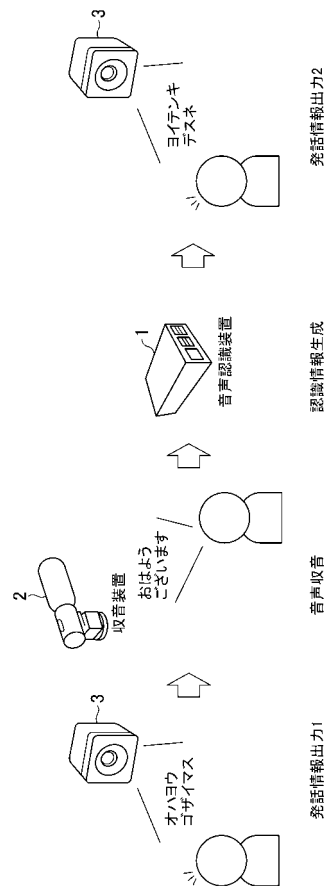
1	: 音声認識装置	
2	: 収音装置	
3	: 制御装置	
4	: 公衆通信網	
5	: サーバ	
6	: ユーザ端末	
1 0	: 筐体	
1 1	: 取得部	
1 2	: 抽出部	
1 3	: 記憶部	20
1 4	: 検出部	
1 5	: 算出部	
1 6	: 選択部	
1 7	: 生成部	
1 7 a	: 指定部	
1 7 b	: 比較部	
1 7 c	: 更新部	
1 8	: 出力部	
1 9	: 反映部	
1 0 0	: 音声認識システム	30
1 0 1	: C P U	
1 0 2	: R O M	
1 0 3	: R A M	
1 0 4	: 保存部	
1 0 5	: I / F	
1 0 6	: I / F	
1 0 7	: I / F	
1 0 8	: 入力部分	
1 0 9	: 出力部分	
1 1 0	: 内部バス	40
S 1 0 0	: 第 1 出力手段	
S 1 1 0	: 取得手段	
S 1 2 0	: 抽出手段	
S 1 3 0	: 検出手段	
S 1 4 0	: 算出手段	
S 1 5 0	: 選択手段	
S 1 6 0	: 生成手段	
S 1 6 1	: 指定手段	
S 1 6 2	: 比較手段	
S 1 6 3	: 更新手段	50

- S 1 7 0 : 第 2 出力手段
- S 1 8 0 : 反映手段
- S 1 9 0 : 設定手段

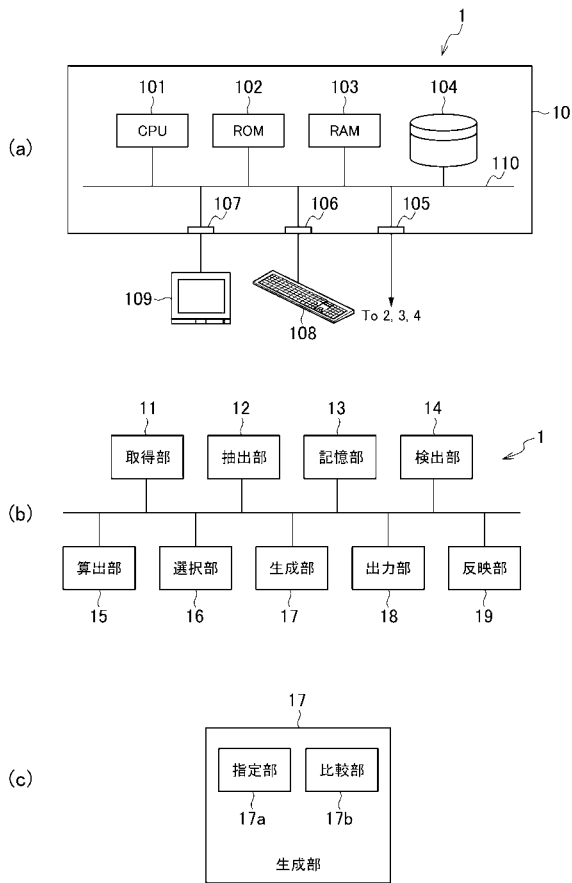
【 図 1 】



【 図 2 】



【図3】



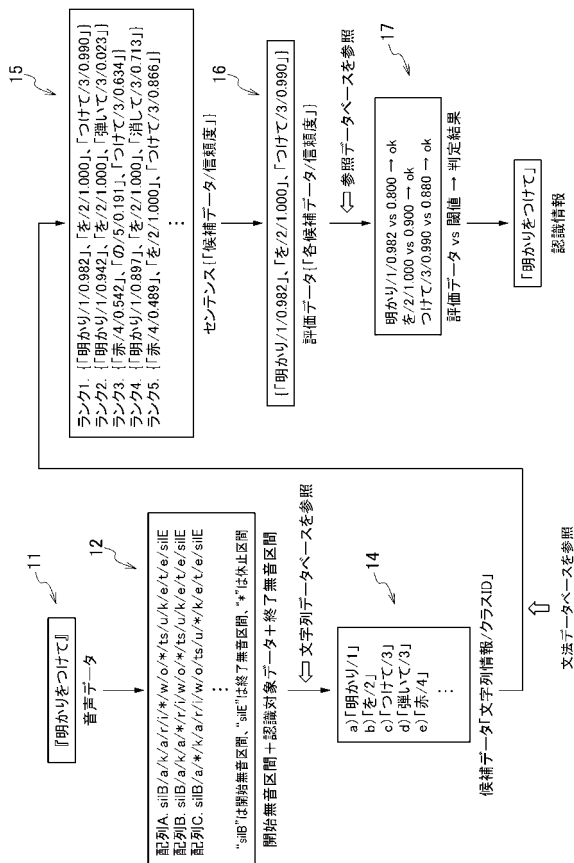
【図5】

参照センテンス	文字列情報	音素情報	閾値
第1参照センテンス	明かり	a/k/a/r/i	0.800
	を	w/o	0.900
第2参照センテンス	つけて	ts/u/k/e/t/e	0.880
	明かり	a/k/a/r/i	0.800
	を	w/o	0.900
...	弾いて	h/i/*/i/t/e	0.650
...

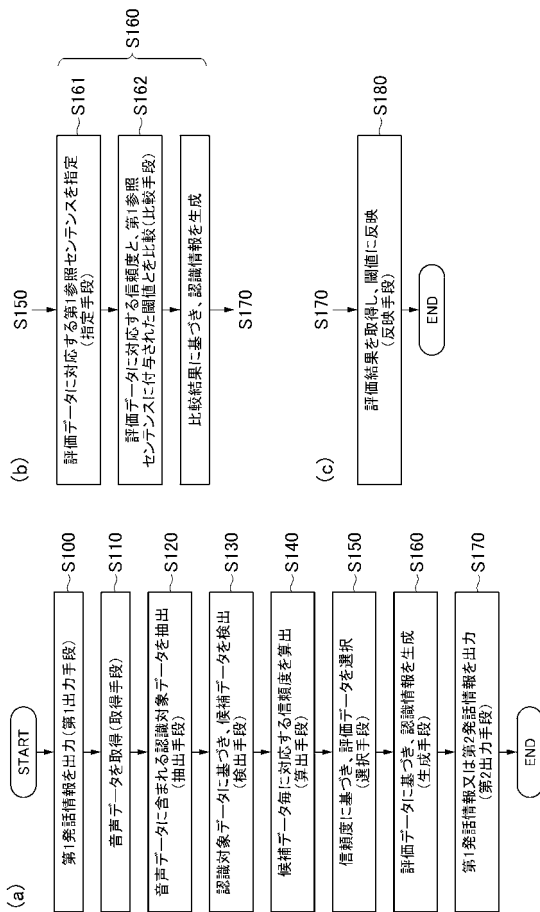
発話情報	返答情報
おはようございます (第1発話情報)	おはようございます (第1返答情報)
良い天気ですね (第2発話情報)	そうですね (第2返答情報)
よろしいでしょうか (第3発話情報)	はい (第3返答情報)
...	...

文法情報	配列順序
第1文法情報	1, 2, 3
第2文法情報	4, 5, 3
第3文法情報	4, 2, 3
...	...

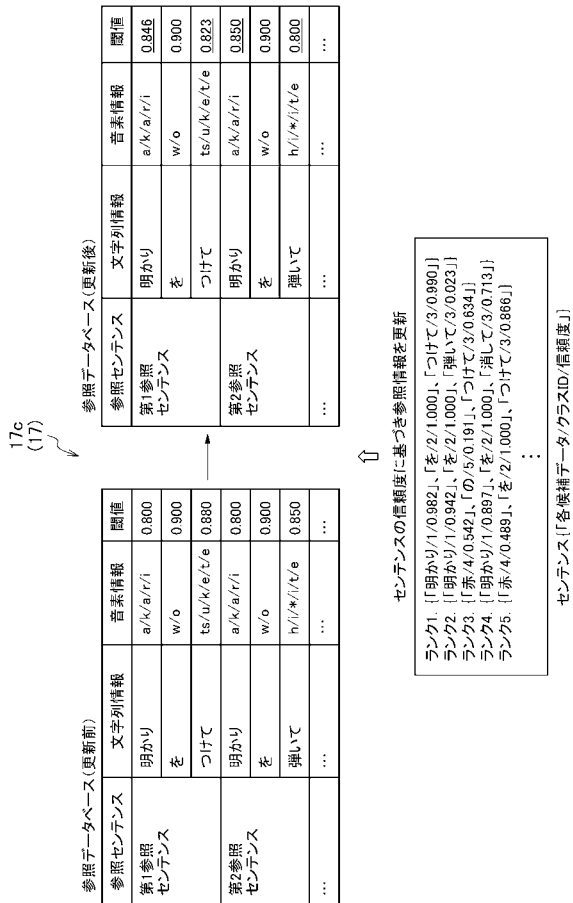
【図4】



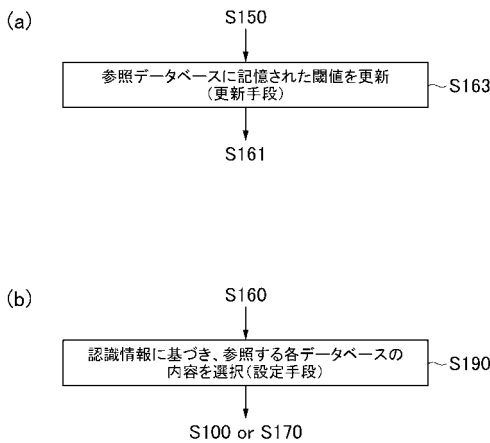
【図6】



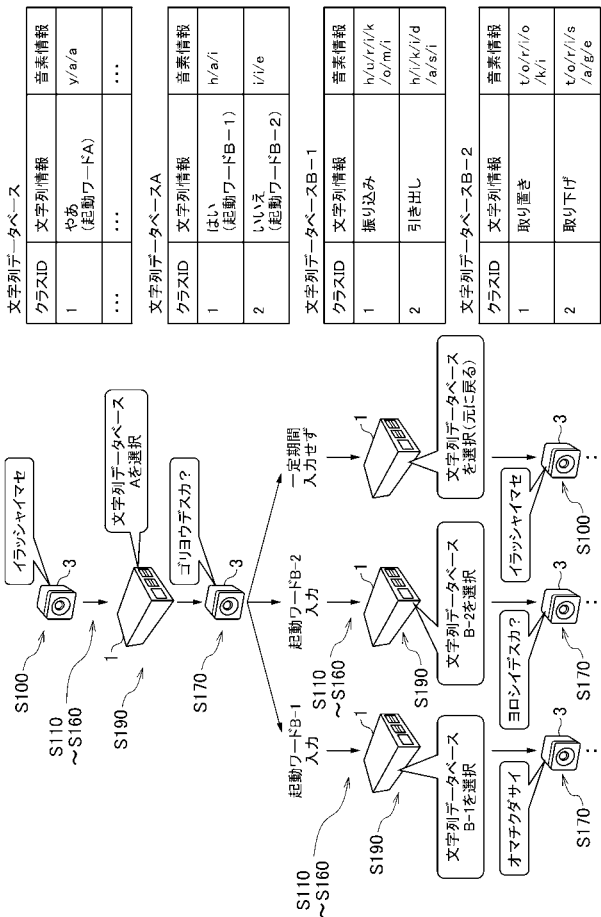
【 図 7 】



【 図 8 】



【 図 9 】



【 図 10 】

条件情報	
環境情報	收音装置2の設置された環境(屋外、屋内の広さ等)
雑音情報	收音装置2が收音し得る雑音(利用者以外の音声、空調音等)
收音装置情報	收音装置2の種類、性能等
利用者情報	利用者人数、国籍、性別等
音特性情報	音声の音量、音圧、癖、活舌の状態等
...	...

【 図 1 1 】

文字列データベース		参照データベース	
クラスID	文字列情報	文字列情報	音素情報
1	こうないかんきょう	こうないかんきょう	k/o/u/n/a/i /k/a/n/ky/o
9999 (類似クラスID-A)	こうないえん (類似文字列情報A)	を	w/o/
9999 (類似クラスID-B)	こうさてん (類似文字列情報B)	しらべる	s/i/r/a/b/e/r /u
9999 (類似クラスID-C)	こうとくてん (類似文字列情報C)	こうないえん	k/o/u/n/a/i /e/n
		を	w/o/
		しらべる	s/i/r/a/b/e/r /u
		こうさてん	k/o/u/s/a/l /e/n
		を	w/o/
		しらべる	s/i/r/a/b/e/r /u
		こうとくてん	k/o/u/t/o/k /w/r/e/n
		を	w/o/
		しらべる	s/i/r/a/b/e/r /u

【 図 1 2 】

