

(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2015-14665

(P2015-14665A)

(43) 公開日 平成27年1月22日(2015.1.22)

(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	2 0 0 V
G 1 0 L 15/00	(2013.01)	G 1 0 L	15/00	2 0 0 K
G 1 0 L 15/28	(2013.01)	G 1 0 L	15/22	2 0 0 H
		G 1 0 L	15/28	2 3 0 J

審査請求 未請求 請求項の数 7 O L (全 21 頁)

(21) 出願番号 特願2013-140426 (P2013-140426)
 (22) 出願日 平成25年7月4日(2013.7.4)

(71) 出願人 000002369
 セイコーエプソン株式会社
 東京都新宿区西新宿2丁目4番1号
 (74) 代理人 100095728
 弁理士 上柳 雅誉
 (74) 代理人 100127661
 弁理士 宮坂 一彦
 (74) 代理人 100116665
 弁理士 渡辺 和昭
 (72) 発明者 野中 勉
 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

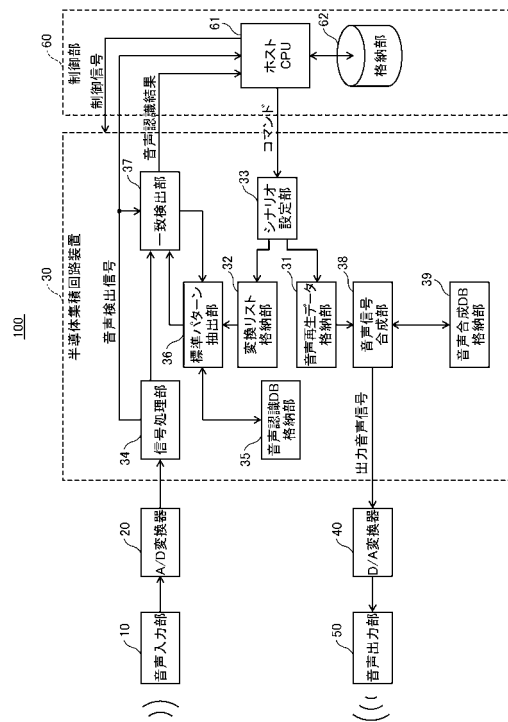
(54) 【発明の名称】 音声認識装置及び方法、並びに、半導体集積回路装置

(57) 【要約】

【課題】 音声認識におけるシナリオの設定や変更を容易に実現できるようにする。

【解決手段】 音声認識用の半導体集積回路装置は、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信し、シナリオフロー情報に従って、音声再生データ格納部において所定の音声再生データを選択すると共に、変換リスト格納部において所定の変換リストを選択するシナリオ設定部と、所定の変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを音声認識データベースから抽出する標準パターン抽出部と、所定の音声再生データに基づいて出力音声信号を合成する音声信号合成部と、入力された音声信号の周波数成分の分布状態を表す特徴パターンを生成する信号処理部と、特徴パターンと標準パターンとの比較を行い、音声認識結果を出力する一致検出部とを含む。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、前記ユーザーの音声に対する音声認識の処理を行う音声認識装置において用いられる半導体集積回路装置であって、

複数の前記音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信し、前記シナリオフロー情報に従って、音声再生データ格納部に格納されている複数の前記音声再生データの内から所定の音声再生データを選択すると共に、変換リスト格納部に格納されている複数の前記変換リストの内から所定の変換リストを選択するシナリオ設定部と、

所定の言語において用いられる複数の音素の周波数成分の分布状態を表す標準パターンを含む音声認識データベースから、前記所定の変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出する標準パターン抽出部と、

前記所定の音声再生データに基づいて出力音声信号を合成する音声信号合成部と、

入力された音声信号にフーリエ変換を施すことにより前記音声信号の周波数成分を抽出し、前記音声信号の周波数成分の分布状態を表す特徴パターンを生成する信号処理部と、

前記音声信号の少なくとも一部から生成された前記特徴パターンと前記音声認識データベースから抽出された前記標準パターンとの比較を行い、音声認識結果を出力する一致検出部と、

を具備する半導体集積回路装置。

【請求項 2】

前記シナリオ設定部が、1回の音声認識動作において、前記音声再生データ格納部に格納されている複数の前記音声再生データの内から、前記シナリオフロー情報に従って前記所定の音声再生データを選択すると共に、前記変換リスト格納部に格納されている複数の前記変換リストの内から、前記シナリオフロー情報に設定されている前記所定の変換リストを選択する、請求項 1 記載の半導体集積回路装置。

【請求項 3】

前記音声信号合成部が、前記シナリオ設定部によって選択された第 1 の音声再生データに基づいて出力音声信号を合成した後に、前記シナリオ設定部が、前記一致検出部から出力される音声認識結果に対応して前記シナリオフロー情報に設定されている第 2 の音声再生データを選択する、請求項 1 又は 2 記載の半導体集積回路装置。

【請求項 4】

前記シナリオ設定部が、複数の前記変換リストの内から、前記シナリオフロー情報に従って、前記第 2 の音声再生データに対応する前記所定の変換リストを選択する、請求項 3 記載の半導体集積回路装置。

【請求項 5】

前記シナリオ設定部が、少なくとも 1 つの前記音声再生データ又は少なくとも 1 つの前記変換リストを設定又は変更するコマンドを受信して、前記音声再生データ格納部において前記少なくとも 1 つの前記音声再生データを設定又は変更し、又は、前記変換リスト格納部において前記少なくとも 1 つの前記変換リストを設定又は変更する、請求項 1 ~ 4 のいずれか 1 項記載の半導体集積回路装置。

【請求項 6】

請求項 1 ~ 5 のいずれか 1 項記載の半導体集積回路装置と、

前記複数の音声再生データと前記複数の変換リストとの対応関係を表す前記シナリオフロー情報を指定する前記コマンドを前記半導体集積回路装置に送信する制御部と、を具備する音声認識装置。

【請求項 7】

音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、前記ユーザーの音声に対する音声認識の処理を行う音声認識装置において用いられる音声認識方法であって、

10

20

30

40

50

複数の前記音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信するステップ(a)と、

前記シナリオフロー情報に従って、音声再生データ格納部に格納されている複数の前記音声再生データの内から所定の音声再生データを選択すると共に、変換リスト格納部に格納されている複数の前記変換リストの内から所定の变換リストを選択するステップ(b)と、

所定の言語において用いられる複数の音素の周波数成分の分布状態を表す標準パターンを含む音声認識データベースから、前記所定の变換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出するステップ(c)と、

前記所定の音声再生データに基づいて出力音声信号を合成するステップ(d)と、

入力された音声信号にフーリエ変換を施すことにより前記音声信号の周波数成分を抽出し、前記音声信号の周波数成分の分布状態を表す特徴パターンを生成するステップ(e)と、

前記音声信号の少なくとも一部から生成された前記特徴パターンと前記音声認識データベースから抽出された前記標準パターンとの比較を行い、音声認識結果を出力するステップ(f)と、

を具備する音声認識方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、音声を認識し、その認識結果に対応する応答や処理を行う音声認識装置及び音声認識方法に関する。さらに、本発明は、そのような音声認識装置及び音声認識方法において用いられる半導体集積回路装置等に関する。

【背景技術】

【0002】

音声認識は、入力される音声信号を解析し、その結果として得られる特徴パターンを、予め収録された音声信号に基づいて音声認識データベースに用意されている標準パターン(「テンプレート」ともいう)と照合することによって、認識結果を得る技術である。一般に、音声認識における選択肢情報は音声認識辞書の中に存在するが、音声認識辞書の更新には時間を要するので、音声認識処理の実行中において選択肢情報の更新が困難であった。

【0003】

関連する従来技術として、特許文献1には、音声認識サービスを中断させることなく、各音声認識処理部の認識辞書を効率よく更新できるようにする音声認識装置が開示されている。この音声認識装置は、ユーザー端末から受信した音声データを認識辞書に基づいて認識する音声認識装置であって、認識辞書が記憶された認識辞書記憶手段と、認識辞書記憶手段から認識辞書を読み出して自身の共通辞書領域に更新登録し、この認識辞書に基づいて音声データを認識する複数の音声認識処理手段と、音声認識要求に応答して、音声認識処理手段のいずれかに音声データを認識させる音声認識要求受付手段と、辞書更新要求に応答して、音声認識処理手段を複数回に分けて少なくとも1つずつ選択し、その認識辞書を順次に更新させる辞書更新要求受付手段とを含むことが記載されている。

【0004】

特許文献1の音声認識装置によれば、認識辞書の更新対象となる音声認識処理手段が、複数回に分けて少なくとも1つずつ選択され、その認識辞書が更新されるので、音声認識装置全体としては、音声認識サービスを中断させることなく全ての音声認識処理手段の認識辞書を更新できるようになる。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2008-15209号公報(段落0009-0012、図1)

10

20

30

40

50

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、このように、音声認識サービスを中断させることなく音声認識辞書の更新が可能であったとしても、入力される音声信号を解析して得られる特徴パターンを標準パターンと照合する際に、照合される範囲に制限が設けられていない場合には、比較すべき特徴パターンと標準パターンとの組み合わせが膨大な数となって処理時間が増大すると共に、想定する語彙が多くなることで判定が難しくなり認識率が低下してしまうことも考えられる。

【0007】

この対策としては、例えば、予め設定されたシナリオに沿って音声認識を行うことにより、比較すべき特徴パターンと標準パターンとの組み合わせの数を制限して、認識率の向上を図ることが考えられるが、シナリオを設定するためには、音声認識ソフトウェアを搭載したCPU（中央演算装置）において、そのシナリオのフロー等をプログラムコードとして組み込むか、又は、外付けの音声認識装置を制御するプログラムコードを作成する必要がある。しかしながら、シナリオを変更する場合には、CPUにおいてプログラムコードを変更しなければならないので、プログラムコードの変更や評価等に多くの時間や工数を要してしまう。そこで、本発明の目的の1つは、音声認識におけるシナリオの設定や変更を容易に実現できるようにすることである。

【課題を解決するための手段】

【0008】

上述した課題の少なくとも1つを解決するため、本発明の1つの観点に係る半導体集積回路装置は、音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、ユーザーの音声に対する音声認識の処理を行う音声認識装置において用いられる半導体集積回路装置であって、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信し、シナリオフロー情報に従って、音声再生データ格納部に格納されている複数の音声再生データの内から所定の音声再生データを選択すると共に、変換リスト格納部に格納されている複数の変換リストの内から所定の変換リストを選択するシナリオ設定部と、所定の言語において用いられる複数の音素の周波数成分の分布状態を表す標準パターンを含む音声認識データベースから、所定の変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出する標準パターン抽出部と、所定の音声再生データに基づいて出力音声信号を合成する音声信号合成部と、入力された音声信号にフーリエ変換を施すことにより音声信号の周波数成分を抽出し、音声信号の周波数成分の分布状態を表す特徴パターンを生成する信号処理部と、音声信号の少なくとも一部から生成された特徴パターンと音声認識データベースから抽出された標準パターンとの比較を行い、音声認識結果を出力する一致検出部とを具備する。

【0009】

本発明の1つの観点に係る半導体集積回路装置によれば、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを半導体集積回路装置に送信することによって、音声認識におけるシナリオの設定や変更を容易に実現することができる。

【0010】

ここで、シナリオ設定部が、1回の音声認識動作において、音声再生データ格納部に格納されている複数の音声再生データの内から、シナリオフロー情報に従って所定の音声再生データを選択すると共に、変換リスト格納部に格納されている複数の変換リストの内から、シナリオフロー情報に設定されている所定の変換リストを選択するようにしても良い。その場合には、音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答が幾つかの単語又は文章の内の1つに予測される状況を作り出し、それらの単語又は文章を含む変換リストを適用することができる。

【0011】

また、音声信号合成部が、シナリオ設定部によって選択された第1の音声再生データに基づいて出力音声信号を合成した後に、シナリオ設定部が、一致検出部から出力される音声認識結果に対応してシナリオフロー情報に設定されている第2の音声再生データを選択するようにしても良い。その場合には、第1の音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答に応じて、次の質問又はメッセージを表す第2の音声再生データを適切に選択することができる。

【0012】

さらに、シナリオ設定部は、複数の変換リストの内から、シナリオフロー情報に従って、第2の音声再生データに対応する所定の変換リストを選択するようにしても良い。その場合には、第2の音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答に対応して、適切な変換リストを選択することができる。

10

【0013】

また、シナリオ設定部は、少なくとも1つの音声再生データ又は少なくとも1つの変換リストを設定又は変更するコマンドを受信して、音声再生データ格納部において少なくとも1つの音声再生データを設定又は変更し、又は、変換リスト格納部において少なくとも1つの変換リストを設定又は変更するようにしても良い。その場合には、音声認識におけるシナリオの設定や変更に伴い、音声再生データ又は変換リストの設定又は変更を容易に実現できるようになる。

【0014】

本発明の1つの観点に係る音声認識装置は、本発明のいずれかの観点に係る半導体集積回路装置と、複数の音声再生データと複数の変換リストとの対応関係を表すシナリオフロー情報を指定するコマンドを半導体集積回路装置に送信する制御部とを具備する。本発明の1つの観点に係る音声認識装置によれば、シナリオフローを変更する場合に、制御部のCPUにおけるプログラムコードの変更が最小限で済み、音声認識におけるシナリオの変更が容易になる。また、制御部の動作が簡素化されるので、簡単な構成のCPUを制御部において用いることができる。

20

【0015】

本発明の1つの観点に係る音声認識方法は、音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、ユーザーの音声に対する音声認識の処理を行う音声認識装置において用いられる音声認識方法であって、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信するステップ(a)と、シナリオフロー情報に従って、音声再生データ格納部に格納されている複数の音声再生データの内から所定の音声再生データを選択すると共に、変換リスト格納部に格納されている複数の変換リストの内から所定の変換リストを選択するステップ(b)と、所定の言語において用いられる複数の音素の周波数成分の分布状態を表す標準パターンを含む音声認識データベースから、所定の変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出するステップ(c)と、所定の音声再生データに基づいて出力音声信号を合成するステップ(d)と、入力された音声信号にフーリエ変換を施すことにより音声信号の周波数成分を抽出し、音声信号の周波数成分の分布状態を表す特徴パターンを生成するステップ(e)と、音声信号の少なくとも一部から生成された特徴パターンと音声認識データベースから抽出された標準パターンとの比較を行い、音声認識結果を出力するステップ(f)とを具備する。

30

40

【0016】

本発明の1つの観点に係る音声認識方法によれば、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドを受信することによって、音声認識におけるシナリオの設定や変更を容易に実現することができる。

【図面の簡単な説明】

【0017】

【図1】本発明の一実施形態に係る音声認識装置の構成例を示すブロック図。

【図2】図1に示す音声認識装置において用いられる音声認識方法を示すフロー図。

50

- 【図 3】音声再生データによって表される質問又はメッセージの例を示す図。
 【図 4】変換リスト格納部に格納されている変換リストの例を示す図。
 【図 5】ホスト CPU からシナリオ設定部に送信される指定コマンドの例を示す図。
 【図 6】図 5 に示す指定コマンドによって設定される音声認識シナリオを示す図。
 【図 7】音声再生データによって表されるフレーズの例を示す図。
 【図 8】変換リスト格納部に格納されている変換リストの例を示す図。
 【図 9】第 1 の音声認識シナリオを設定する第 1 の指定コマンドの例を示す図。
 【図 10】図 9 に示す指定コマンドによって設定される音声認識シナリオを示す図。
 【図 11】第 2 の音声認識シナリオを設定する第 2 の指定コマンドの例を示す図。
 【図 12】図 11 に示す指定コマンドによって設定される音声認識シナリオを示す図。
 【発明を実施するための形態】

10

【0018】

以下、本発明の実施形態について、図面を参照しながら詳しく説明する。

図 1 は、本発明の一実施形態に係る音声認識装置の構成例を示すブロック図である。音声認識装置 100 は、例えば、自動販売機、家電製品、住宅設備、車載装置（ナビゲーション装置等）、又は、携帯端末等に搭載される。音声認識装置 100 は、音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、音声認識において変換候補となる複数の単語又は文章を含む変換リストを用いてユーザーの音声を認識し、その認識結果に対応する応答や処理を行う。

【0019】

20

図 1 に示すように、音声認識装置 100 は、音声入力部 10 と、A/D 変換器 20 と、音声認識用の半導体集積回路装置 30 と、D/A 変換器 40 と、音声出力部 50 と、制御部 60 とを含んでいる。なお、音声入力部 10、A/D 変換器 20、D/A 変換器 40、及び、音声出力部 50 の少なくとも一部を、半導体集積回路装置 30 に内蔵しても良い。

【0020】

制御部 60 は、ホスト CPU（中央演算装置）61 と、格納部 62 とを含んでいる。ホスト CPU 61 は、格納部 62 の記録媒体に記録されているソフトウェア（音声認識制御プログラム）に基づいて動作する。記録媒体としては、ハードディスク、フレキシブルディスク、MO、MT、CD-ROM、又は、DVD-ROM 等を用いることができる。

【0021】

30

ホスト CPU 61 は、音声認識用の半導体集積回路装置 30 に制御信号を供給することにより、半導体集積回路装置 30 における音声認識動作を制御する。また、ホスト CPU 61 は、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定する指定コマンドを、半導体集積回路装置 30 に送信する。さらに、ホスト CPU 61 は、少なくとも 1 つの音声再生データ又は少なくとも 1 つの変換リストを設定又は変更（追加若しくは削除）する設定コマンドを、半導体集積回路装置 30 に送信する。

【0022】

音声入力部 10 は、音声を電気信号（音声信号）に変換するマイクロフォンと、マイクロフォンから出力される音声信号を増幅する増幅器と、増幅された音声信号の帯域を制限するローパスフィルタとを含んでいる。A/D 変換器 20 は、音声入力部 10 から出力されるアナログの音声信号をサンプリングすることにより、デジタルの音声信号（音声データ）に変換する。例えば、音声データにおける音声周波数帯域は 12 kHz であり、ビット数は 16 ビットである。

40

【0023】

半導体集積回路装置 30 は、音声再生データ格納部 31 と、変換リスト格納部 32 と、シナリオ設定部 33 と、信号処理部 34 と、音声認識データベース（DB）格納部 35 と、標準パターン抽出部 36 と、一致検出部 37 と、音声信号合成部 38 と、音声合成データベース（DB）格納部 39 とを含んでいる。なお、音声再生データ格納部 31、変換リスト格納部 32、音声認識データベース格納部 35、及び、音声合成データベース格納部 39 の少なくとも一部を、半導体集積回路装置 30 に外付けされるメモリー等に内蔵して

50

も良い。

【 0 0 2 4 】

音声再生データ格納部 3 1 は、ユーザーに質問又はメッセージを発するために用いられる複数の音声再生データを格納する。また、変換リスト格納部 3 2 は、音声認識において変換候補となる複数の単語又は文章のテキストデータを各々が含む複数の変換リストを格納する。シナリオ設定部 3 3 は、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定する指定コマンドを、ホスト CPU 6 1 から受信する。

【 0 0 2 5 】

シナリオ設定部 3 3 は、シナリオフロー情報に従って、1 回の音声認識動作において、音声再生データ格納部 3 1 に格納されている複数の音声再生データの内から 1 つの音声再生データを選択すると共に、変換リスト格納部 3 2 に格納されている複数の変換リストの内から 1 つの変換リストを選択する。これにより、音声認識におけるシナリオの設定や変更を容易に実現することができる。

10

【 0 0 2 6 】

また、シナリオ設定部 3 3 は、少なくとも 1 つの音声再生データ又は少なくとも 1 つの変換リストを設定又は変更する設定コマンドをホスト CPU 6 1 から受信する。シナリオ設定部 3 3 は、設定コマンドに従って、音声再生データ格納部 3 1 において少なくとも 1 つの音声再生データを設定又は変更したり、又は、変換リスト格納部 3 2 において少なくとも 1 つの変換リストを設定又は変更したりすることができる。これにより、音声認識におけるシナリオの設定や変更に伴って、音声再生データ又は変換リストの設定又は変更を容易に実現することができる。

20

【 0 0 2 7 】

信号処理部 3 4 は、入力された音声信号にフーリエ変換を施すことにより音声信号の周波数成分を抽出し、音声信号の周波数成分の分布状態を表す特徴パターンを生成する。生成された特徴パターンは、一致検出部 3 7 に出力される。また、信号処理部 3 4 は、入力された音声信号のレベルが所定の値を超えたときに、音声検出信号を活性化して一致検出部 3 7 及びホスト CPU 6 1 に出力する。これにより、ユーザーからの要求又は回答の有無を判定することができる。

【 0 0 2 8 】

ここで、音声信号から特徴パターンを求める手法の一例について説明する。信号処理部 3 4 は、入力された音声信号にフィルタ処理を施して高域成分を強調する。次に、信号処理部 3 4 は、音声信号によって表される音声波形にハミング窓をかけることにより、時系列の音声信号を所定の時間毎に区切って複数のフレームを作成する。さらに、信号処理部 3 4 は、フレーム毎に音声信号をフーリエ変換することにより、複数の周波数成分を抽出する。各々の周波数成分は複素数であるので、信号処理部 3 4 は、各々の周波数成分の絶対値を求める。

30

【 0 0 2 9 】

信号処理部 3 4 は、それらの周波数成分に、メル尺度に基づいて定められた周波数領域の窓をかけて積分することにより、窓の数に対応する数の数値を求める。さらに、信号処理部 3 4 は、それらの数値の対数をとって、対数値を離散コサイン変換する。これにより、周波数領域の窓が 2 0 個であれば、2 0 個の数値が得られる。

40

【 0 0 3 0 】

このようにして得られた数値の内低次のもの（例えば、1 2 個）が、MFCC（メル周波数ケプストラム係数）と呼ばれる。信号処理部 3 4 は、フレーム毎に MFCC を算出し、HMM（隠れマルコフモデル）に従って MFCC を連結して、時系列で入力された音声信号に含まれている各々の音素に対応する MFCC として特徴パターンを求める。

【 0 0 3 1 】

ここで、「音素」とは、ある言語において同じとみなされる音の要素を意味する。以下においては、言語として日本語が用いられる場合について説明する。日本語の音素としては、「a」、「i」、「u」、「e」、「o」の母音と、「k」、「s」、「t」、「n

50

」等の子音と、「j」、「w」の半母音と、「N」、「Q」、「H」の特殊モーラとが該当する。

【0032】

音声認識データベース格納部35は、所定の言語において用いられる各種の音素について周波数成分の分布状態を表す標準パターンを含む音声認識データベースを格納する。音声認識データベースにおいては、各種の音素を表すテキストデータと、選択肢情報としての標準パターンとが、対応付けられている。

【0033】

標準パターンは、多数（例えば、200人程度）の話者が発した音声を用いて予め作成される。標準パターンの作成においては、各々の音素を表す音声信号からMFCCが求められる。ただし、多数の話者が発した音声を用いて作成されたMFCCにおいては、それぞれの数値がばらつきを有している。

10

【0034】

従って、各々の音素についての標準パターンは、多次元空間（例えば、12次元空間）において、ばらつきを含む広がりをもっている。信号処理部34に入力された音声信号から生成された特徴パターンが標準パターンの広がり内の範囲内に入っていれば、両者の音素が一致していると判定される。

【0035】

また、1つの音声認識データベースではなく、複数の音声認識データベースを用いるようにしても良い。例えば、音声認識データベース格納部35は、年齢及び性別の異なる複数群の話者の音声を収録して得られた音声信号に基づいて生成された複数の音声認識データベースを格納しても良い。その場合に、一致検出部37は、複数の音声認識データベースの内から、音素の一致を良好に検出できる音声認識データベースを選択して使用することができる。

20

【0036】

あるいは、音声認識装置を使用するユーザーの年齢及び性別を特定できる場合には、音声認識データベース格納部35は、年齢及び性別の異なる複数群の話者の音声を収録して得られた音声データに基づいて生成された複数の音声認識データベースを、年齢及び性別を特定する情報に対応付けて格納しても良い。その場合に、一致検出部37は、音声認識データベース格納部35に格納されている複数の音声認識データベースの内から、音声認識装置を使用するユーザーの年齢及び性別を特定する情報に従って1つの音声認識データベースを選択して使用することができる。

30

【0037】

標準パターン抽出部36は、音声認識データベース格納部35に格納されている音声認識データベースから、シナリオ設定部33によって選択された変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出する。

【0038】

一致検出部37は、音声検出信号が活性化されているときに動作し、入力された音声信号の少なくとも一部から生成された特徴パターンを、音声認識データベースから抽出された標準パターンと比較することによって、両者の一致を検出する。

40

【0039】

例えば、一致検出部37は、入力された音声信号の先頭の音節から生成された特徴パターンを、変換リストに含まれているテキストデータによって表される各々の単語又は文章の先頭の音節に対応する標準パターンと比較する。変換リストにおいて、一致が検出された音節を先頭に有する変換候補が1つだけ存在する場合には、一致検出部37は、その変換候補が変換後の単語又は文章であると判定しても良い。一方、変換リストにおいて、一致が検出された音節を先頭に有する複数の変換候補が存在する場合には、一致検出部37は、変換候補が1つに絞られるまで、一致を検出すべき音節の範囲を拡大しても良い。

【0040】

ここで、「音節」とは、1個の母音を主音とし、その母音単独で、あるいは、その母音

50

の前後に1つ又は複数の子音を伴って構成される音のまとまりを意味する。また、半母音や特殊モーラも、音節を構成することができる。即ち、1つの音節は、1つ又は複数の音素によって構成される。日本語の音節としては、「あ」、「い」、「う」、「え」、「お」、「か」、「き」、「く」、「け」、「こ」等が該当する。

【0041】

例えば、音節「あ」に対応する標準パターンとは、音節「あ」を構成する音素「a」を表す標準パターンのことである。また、音節「か」に対応する標準パターンとは、音節「か」を構成する第1番目の音素「k」を表す標準パターンと、音節「か」を構成する第2番目の音素「a」を表す標準パターンとの組み合わせのことである。

【0042】

入力された音声信号の1つの音節が1つの音素で構成されている場合には、その音素の一致が検出されれば、音節の一致が検出されたことになる。一方、入力された音声信号の1つの音節が複数の音素で構成されている場合には、それらの音素の一致が検出されれば、音節の一致が検出されたことになる。

【0043】

特徴パターンと標準パターンとの間で上述したような一致が検出されると、一致検出部37は、変換候補となる複数の単語又は文章の中で一致が検出された音節を有する単語又は文章を特定する情報、例えば、その単語又は文章を表すテキストデータを、音声認識結果として出力する。これにより、ホストCPU61は、半導体集積回路装置30に入力された音声信号の少なくとも一部に対応する単語又は文章を認識することができる。

【0044】

音声信号合成部38は、シナリオ設定部33によって選択された音声再生データに基づいて、出力すべき音声を表す出力音声信号を合成する。音声再生データは、テキストデータであっても良い。その場合には、出力音声信号を合成するために、音声合成データベース格納部39に格納されている音声合成データベースが用いられる。音声合成データベースには、各種の音素について音声波形を表す音声データが含まれており、音声信号合成部38は、テキストデータによって表される単語又は文章に含まれている複数の音素について音声データを繋ぎ合わせることににより、出力音声信号を合成する。

【0045】

D/A変換器40は、音声信号合成部38から出力されるデジタルの音声信号を、アナログの音声信号に変換する。音声出力部50は、D/A変換器40から出力されるアナログの音声信号を電力増幅する電力増幅器と、電力増幅された音声信号に応じて音声を発するスピーカーとを含んでいる。スピーカーは、シナリオ設定部33によって選択された音声再生データによって表される質問又はメッセージを、音声として出力する。

【0046】

シナリオ設定部33は、1回の音声認識動作において、音声再生データ格納部31に格納されている複数の音声再生データの内から、シナリオフロー情報に従って1つの音声再生データを選択すると共に、変換リスト格納部32に格納されている複数の変換リストの内から、選択された音声再生データに対応してシナリオフロー情報に設定されている1つの変換リストを選択する。

【0047】

シナリオフロー情報には、各々の音声再生データに基づいて発せられる質問又はメッセージに対する回答として変換候補となる複数の単語又は文章を含む変換リストが設定されている。これにより、音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答が幾つかの単語又は文章の内の1つに予測される状況を作り出し、それらの単語又は文章を含む変換リストを適用することができる。

【0048】

また、音声信号合成部38が、シナリオ設定部33によって選択された第1の音声再生データに基づいて出力音声信号を合成し、第1の音声再生データに基づいて質問又はメッセージが発せられた後に、シナリオ設定部33は、一致検出部37から出力される音声認

10

20

30

40

50

識結果に対応して第2の音声再生データを選択する。これにより、第1の音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答に応じて、次の質問又はメッセージを表す第2の音声再生データを適切に選択することができる。

【0049】

さらに、シナリオ設定部33は、変換リスト格納部32に格納されている複数の変換リストの内から、シナリオフロー情報に従って、第2の音声再生データに対応する変換リストを選択する。これにより、第2の音声再生データに基づいて発せられる質問又はメッセージに対するユーザーの回答に対応して、適切な変換リストを選択することができる。

【0050】

(音声認識方法)

次に、本発明の一実施形態に係る音声認識方法について、図1及び図2を参照しながら説明する。この音声認識方法は、音声再生データに基づいてユーザーに質問又はメッセージを発すると共に、音声認識において変換候補となる複数の単語又は文章を含む変換リストを用いてユーザーの音声を認識する音声認識装置において用いられる。

【0051】

図2は、図1に示す音声認識装置において用いられる音声認識方法を示すフローチャートである。図2のステップS1において、ホストCPU61が、音声認識用の半導体集積回路装置30の電源投入時又はリセット後に、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定する指定コマンドを、半導体集積回路装置30に送信する。

【0052】

ステップS2において、半導体集積回路装置30のシナリオ設定部33が、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定する指定コマンドを、ホストCPU61から受信する。

【0053】

ステップS3において、シナリオ設定部33が、シナリオフロー情報に従って、1回の音声認識動作において、音声再生データ格納部31に格納されている複数の音声再生データの内から1つの音声再生データを選択すると共に、変換リスト格納部32に格納されている複数の変換リストの内から1つの変換リストを選択する。

【0054】

ステップS4において、標準パターン抽出部36が、所定の言語において用いられる複数の音素の周波数成分の分布状態を表す標準パターンを含む音声認識データベースから、ステップS3において選択された変換リストに含まれている各々の単語又は文章の少なくとも一部に対応する標準パターンを抽出する。

【0055】

ステップS5において、音声信号合成部38が、ステップS3において選択された音声再生データに基づいて出力音声信号を合成することにより、音声出力部50から質問又はメッセージが発せられる。

【0056】

これに回答してユーザーが音声を発すると、ステップS6において、信号処理部34が、入力された音声信号にフーリエ変換を施すことにより音声信号の周波数成分を抽出し、入力された音声信号の周波数成分の分布状態を表す特徴パターンを生成する。また、信号処理部34は、音声検出信号を活性化する。

【0057】

音声検出信号が活性化されると、ステップS7において、一致検出部37が、入力された音声信号の少なくとも一部から生成された特徴パターンと音声認識データベースから抽出された標準パターンとの一致を検出して、変換候補となる複数の単語又は文章の内一致が検出された単語又は文章を特定する音声認識結果を出力する。

【0058】

音声検出信号が活性化されてから所定の期間内に特徴パターンと標準パターンとの一致

10

20

30

40

50

を表す音声認識結果が得られると、ステップS8において、ホストCPU61が、一連の音声認識動作が完了したか否かを判定する。一連の音声認識動作が完了していれば、処理が終了する。一方、一連の音声認識動作が完了していなければ、処理がステップS3に戻る。

【0059】

本発明の一実施形態によれば、音声認識におけるシナリオに従った変換リストを用いることにより、入力された音声信号の特徴パターンと比較される標準パターンを、変換リストに設定されたテキストデータによって表される各々の単語又は文章の少なくとも一部に対応する標準パターンに絞り込むことができる。

【0060】

その際に、複数の音声再生データと複数の変換リストとの関係を表すシナリオフロー情報を指定するコマンドをホストCPU61から受信することによって、音声認識におけるシナリオの設定や変更を容易に実現することができる。従って、シナリオフローを変更する場合に、制御部60のホストCPU61におけるプログラムコードの変更が最小限で済み、音声認識におけるシナリオの変更が容易になる。また、制御部60の動作が簡素化されるので、簡単な構成のCPUを制御部60において用いることができる。

【0061】

(第1の具体例)

次に、本発明の一実施形態に係る音声認識装置における音声認識動作の第1の具体例について説明する。以下においては、図1に示す音声認識装置が食堂における食券の自動販売機に適用される場合について説明する。

【0062】

自動販売機には、複数の食品名を含む食品メニューが表示されている。食品メニューには、「うどん」、「そば」、「ラーメン」の文字が表示されているものとする。その場合には、ユーザーが発する最初の言葉が、食品メニューに表示されている「うどん」、「そば」、「ラーメン」の内のいずれかになることが予測される。

【0063】

図3は、音声再生データ格納部に格納されている音声再生データによって表される質問又はメッセージの例を示す図である。この例においては、4つの音声再生番号に対応して、4つの質問又はメッセージの内容を表す音声再生データが、音声再生データ格納部31に格納されている。音声再生データは、質問又はメッセージに含まれている音素を特定できるローマ字表記又はカナ表記を表すデータを含んでいる。

【0064】

図4は、変換リスト格納部に格納されている変換リストの例を示す図である。この例においては、3つの変換リスト番号に対応して、3つの変換リストが、変換リスト格納部32に格納されている。各々の変換リストにおいて、複数の選択肢番号に対応して、複数の音声認識選択肢を表すテキストデータが設定されている。テキストデータは、音声認識選択肢に含まれている音素を特定できるローマ字表記又はカナ表記を表すデータを含んでいる。

【0065】

図5は、ホストCPUからシナリオ設定部に送信される指定コマンドの例を示す図である。図5に示す指定コマンドは、図3に示す複数の音声再生データと、図4に示す複数の変換リストとの関係を表すシナリオフロー情報を指定するものである。この例においては、指定コマンドが、音声再生データ数と、複数の音声再生番号に対応する変換リスト番号と、変換リスト数と、各々の変換リストについて選択肢数及びそれらの選択肢が認識された後の音声再生番号と、シナリオ開始時の音声再生番号とを含んでいる。なお、変換リスト番号「0」は、シナリオの終了を表している。

【0066】

図6は、図5に示す指定コマンドによって指定されるシナリオフロー情報に従って実行される音声認識シナリオを示す図である。例えば、ホストCPU61は、自動販売機の電

10

20

30

40

50

源投入時又はリセット後に、指定コマンドを半導体集積回路装置 30 に送信する。半導体集積回路装置 30 のシナリオ設定部 33 は、受信された指定コマンドによって指定されるシナリオフロー情報に従って、音声認識動作を開始する。

【0067】

まず、シナリオ設定部 33 は、図 5 に示す指定コマンドによって指定されているシナリオ開始時の音声再生番号「1」に従って、音声再生データ格納部 31 に格納されている複数の音声再生データの中から音声再生番号「1」に対応する音声再生データ Q1 を選択する。

【0068】

また、シナリオ設定部 33 は、図 5 に示す指定コマンドにおいて音声再生番号 1 に対応する変換リスト番号「1」に従って、変換リスト格納部 32 に格納されている複数の変換リストの中から変換リスト番号「1」に対応する変換リスト L1 を選択する。

10

【0069】

標準パターン抽出部 36 は、変換リスト L1 に含まれている音声認識選択肢 A1「うどん」、音声認識選択肢 A2「そば」、音声認識選択肢 A3「ラーメン」の先頭の音節「う」、「そ」、「ら」に含まれている音素「u」、「s・o」、「r・a」のそれぞれについて、対応する標準パターンを音声認識データベースから抽出する。

【0070】

一方、音声信号合成部 38 は、音声再生データ Q1 に基づいて、音声信号を合成して D/A 変換器 40 に出力する。また、D/A 変換器 40 は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部 50 に出力する。これにより、音声出力部 50 から、「どの食品にしますか？ 食品名を言って下さい。」という質問又はメッセージが発せられる。

20

【0071】

音声出力部 50 から発せられた質問又はメッセージに対して、ユーザーが、表示された食品メニューを見て、「そばを下さい。」と言うと、信号処理部 34 は、音素「s・o・b・a・・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

【0072】

一致検出部 37 は、信号処理部 34 によって生成された先頭の音節の第 1 番目の音素「s」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第 1 番目の音素「u」、「s」、「r」の標準パターンとを比較することにより、音素「s」の一致を検出する。

30

【0073】

一致が検出された音素が子音を表している場合には、さらに、一致検出部 37 が、先頭の音節の第 2 番目の音素を比較する。一致検出部 37 は、信号処理部 34 によって生成された先頭の音節の第 2 番目の音素「o」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第 2 番目の音素「o」、「a」の標準パターンとを比較することにより、音素「o」の一致を検出する。

【0074】

これにより、先頭の音節「そ」の一致が検出される。一致が検出された音声認識選択肢が 1 つであれば、ここで音声認識結果が得られる。一方、一致が検出された音声認識選択肢が複数であれば、いずれが該当するかを認識することができないので、標準パターン抽出部 36 は、次の音節に含まれている音素のそれぞれについて、対応する標準パターンを音声認識データベースから抽出し、一致検出部 37 は、一致を検出すべき音節の範囲を拡大する。

40

【0075】

一致検出部 37 は、一致が検出された先頭の音節「そ」を有する音声認識選択肢「そば」を特定する音声認識結果をホスト CPU 61 に出力する。音声認識選択肢「そば」を特定する情報としては、例えば、図 4 に示す変換リスト番号「1」及び選択肢番号「2」、

50

音声認識選択肢の日本語表記「そば」又はその一部「そ」、音声認識選択肢に含まれている音素のローマ字表記「s o b a」又はその一部「s o」等が該当する。これにより、ホストCPU61は、入力された音声信号の少なくとも一部に対応する音声認識選択肢「そば」を認識することができる。

【0076】

このようにして第1回目の音声認識動作が終了すると、シナリオ設定部33は、図5に示す指定コマンドにおける変換リスト1の選択肢2が選択された後の音声再生番号「2」に従って、音声再生データ格納部31に格納されている複数の音声再生データのの中から音声再生番号「2」に対応する音声再生データQ2を選択する。

【0077】

また、シナリオ設定部33は、図5に示す指定コマンドにおいて音声再生番号2に対応する変換リスト番号「2」に従って、変換リスト格納部32に格納されている複数の変換リストの中から変換リスト番号「2」に対応する変換リストL2を選択する。

【0078】

標準パターン抽出部36は、変換リストL2に含まれている音声認識選択肢A1「温かい」、音声認識選択肢A2「冷たい」の先頭の音節「あ」、「つ」に含まれている音素「a」、「t・u」のそれぞれについて、対応する標準パターンを音声認識データベースから抽出する。

【0079】

一方、音声信号合成部38は、音声再生データQ2に基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「温かい、冷たい、どちらにしますか?」という質問が発せられる。

【0080】

音声出力部50から発せられた質問に対して、ユーザーが、「温かいのを下さい。」と言うと、信号処理部34は、音素「a・t・a・t・a・k・a・i・・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

【0081】

一致検出部37は、信号処理部34によって生成された先頭の音節の第1番目の音素「a」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第1番目の音素「a」、「t」の標準パターンとを比較することにより、音素「a」の一致を検出する。これにより、先頭の音節「あ」の一致が検出される。

【0082】

一致検出部37は、一致が検出された先頭の音節「あ」を有する音声認識選択肢「温かい」を特定する音声認識結果をホストCPU61に出力する。これにより、ホストCPU61は、入力された音声信号の少なくとも一部に対応する音声認識選択肢「温かい」を認識することができる。

【0083】

このようにして第2回目の音声認識動作が終了すると、シナリオ設定部33は、図5に示す指定コマンドにおける変換リスト2の選択肢1が選択された後の音声再生番号「3」に従って、音声再生データ格納部31に格納されている複数の音声再生データのの中から音声再生番号「3」に対応する音声再生データQ3を選択する。

【0084】

また、シナリオ設定部33は、図5に示す指定コマンドにおいて音声再生番号3に対応する変換リスト番号「3」に従って、変換リスト格納部32に格納されている複数の変換リストの中から変換リスト番号「3」に対応する変換リストL3を選択する。

【0085】

標準パターン抽出部36は、変換リストL3に含まれている音声認識選択肢A1「大盛り」、音声認識選択肢A2「普通」の先頭の音節「お」、「ふ」に含まれている音素「o

10

20

30

40

50

」、 「h・u」のそれぞれについて、対応する標準パターンを音声認識データベースから抽出する。

【0086】

一方、音声信号合成部38は、音声再生データQ3に基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「大盛り、普通、どちらにしますか?」という質問が発せられる。

【0087】

音声出力部50から発せられた質問に対して、ユーザーが、「大盛りを下さい。」と言うと、信号処理部34は、音素「o・H・m・o・r・i・・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

【0088】

一致検出部37は、信号処理部34によって生成された先頭の音節の第1番目の音素「o」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第1番目の音素「o」、「h」の標準パターンとを比較することにより、音素「o」の一致を検出する。これにより、先頭の音節「お」の一致が検出される。

【0089】

一致検出部37は、一致が検出された先頭の音節「お」を有する音声認識選択肢「大盛り」を特定する音声認識結果をホストCPU61に出力する。これにより、ホストCPU61は、入力された音声信号の少なくとも一部に対応する音声認識選択肢「大盛り」を認識することができる。

【0090】

このようにして第3回目の音声認識動作が終了すると、シナリオ設定部33は、図5に示す指定コマンドにおける変換リスト3の選択肢1が選択された後の音声再生番号「4」に従って、音声再生データ格納部31に格納されている複数の音声再生データの内から音声再生番号「4」に対応する音声再生データQ4を選択する。

【0091】

音声信号合成部38は、音声再生データQ4に基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「ありがとうございます。」というメッセージが発せられる。

【0092】

シナリオ設定部33は、図5に示す指定コマンドにおいて音声再生番号4に対応する変換リスト番号「0」に従って、一連の音声認識動作を終了させる。その後、自動販売機が、食券の購入に必要な金額を表示部に表示して指定する。ユーザーが指定された金額を自動販売機に投入すると、自動販売機が食券を提供する。ユーザーが上記以外の選択をした場合にも、図6に示す音声認識シナリオに従って音声認識動作が行われる。

【0093】

(第2の具体例)

次に、本発明の一実施形態に係る音声認識装置における音声認識動作の第2の具体例について説明する。第2の具体例においては、食券の自動販売機が設置される麺専門店と多メニュー店とにおけるメニューの種類の相違に応じて、指定コマンドにより音声認識シナリオが変更される。

【0094】

図7は、音声再生データ格納部に格納されている音声再生データによって表されるフレーズの例を示す図である。複数のフレーズを組み合すことによって、1つの文章が構成される。この例においては、15個のフレーズ番号に対応して、15個のフレーズの内容を表す音声再生データが、音声再生データ格納部31に格納されている。

【0095】

10

20

30

40

50

図 8 は、変換リスト格納部に格納されている変換リストの例を示す図である。この例においては、4つの変換リスト番号に対応して、4つの変換リストが、変換リスト格納部 32 に格納されている。各々の変換リストにおいて、複数の選択肢番号に対応して、複数の音声認識選択肢を表すテキストデータが設定されている。

【0096】

(第1の音声認識シナリオ)

図 9 は、第1の音声認識シナリオを設定する第1の指定コマンドの例を示す図である。第1の指定コマンドにより、麵専門店において実行される第1の音声認識シナリオが設定される。第1の指定コマンドは、図7に示す複数の音声再生データと、図8に示す複数の変換リストとの関係を表すシナリオフロー情報を指定するものである。

10

【0097】

図9に示すように、第1の指定コマンドは、音声再生設定数と、各々の音声再生についてフレーズ数、フレーズ番号、及び、音声再生に対応する変換リスト番号と、変換リスト設定数と、各々の変換リストについて選択肢数及びそれらの選択肢が認識された後の音声再生番号と、シナリオ開始時の音声再生番号とを含んでいる。なお、変換リスト番号「0」は、シナリオの終了を表している。

【0098】

図10は、図9に示す第1の指定コマンドによって設定される第1の音声認識シナリオを示す図である。まず、シナリオ設定部33は、図9に示す指定コマンドによって指定されているシナリオ開始時の音声再生番号が「1」であるので、音声再生1の第1番目～第5番目のフレーズ番号「1」、「3」、「4」、「5」、「14」に従って、音声再生データ格納部31に格納されている複数の音声再生データの内から対応するフレーズP1、P3、P4、P5、P14を表す音声再生データを選択する。

20

【0099】

また、シナリオ設定部33は、図9に示す指定コマンドにおいて音声再生1に対応する変換リスト番号「1」に従って、変換リスト格納部32に格納されている複数の変換リストの内から変換リスト番号「1」に対応する変換リストL1を選択する。

【0100】

標準パターン抽出部36は、変換リストL1に含まれている音声認識選択肢A1「うどん」、音声認識選択肢A2「そば」、音声認識選択肢A3「ラーメン」の先頭の音節に含まれている音素のそれぞれについて、対応する標準パターンを音声認識データベースから抽出する。

30

【0101】

一方、音声信号合成部38は、フレーズP1、P3、P4、P5、P14を表す音声再生データに基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「メニューの案内です。うどん、そば、ラーメンから選択してください。」というメッセージ(音声再生1)が発せられる。

【0102】

音声出力部50から発せられたメッセージ(音声再生1)に対して、ユーザーが、「うどんを下さい。」と言うと、信号処理部34は、音素「u・d・o・N・・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

40

【0103】

一致検出部37は、信号処理部34によって生成された先頭の音節の第1番目の音素「u」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第1番目の音素の標準パターンとを比較することにより、音素「u」の一致を検出する。これにより、先頭の音節「う」の一致が検出される。一致検出部37は、一致が検出された先頭の音節「う」を有する音声認識選択肢「うどん」を特定する音声認識結果をホストCPU61に出力する。

50

【0104】

このようにして第1回目の音声認識動作が終了すると、シナリオ設定部33は、図9に示す指定コマンドにおける変換リスト1の選択肢1が選択された後の音声再生番号が「2」であるので、音声再生2の第1番目のフレーズ番号「15」に従って、音声再生データ格納部31に格納されている複数の音声再生データの内から対応するフレーズP15を表す音声再生データを選択する。

【0105】

音声信号合成部38は、フレーズP15を表す音声再生データに基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「ありがとうございます。」というメッセージ（音声再生2）が発せられる。シナリオ設定部33は、図9に示す指定コマンドにおいて音声再生2に対応する変換リスト番号「0」に従って、一連の音声認識動作を終了させる。

10

【0106】

（第2の音声認識シナリオ）

図11は、第2の音声認識シナリオを設定する第2の指定コマンドの例を示す図である。第2の指定コマンドにより、多メニュー店において実行される第2の音声認識シナリオが設定される。第2の指定コマンドも、図7に示す複数の音声再生データと、図8に示す複数の変換リストとの関係を表すシナリオフロー情報を指定するものである。

20

【0107】

図11に示すように、第2の指定コマンドは、音声再生設定数と、各々の音声再生についてフレーズ数、フレーズ番号、及び、音声再生に対応する変換リスト番号と、変換リスト設定数と、各々の変換リストについて選択肢数及びそれらの選択肢が認識された後の音声再生番号と、シナリオ開始時の音声再生番号とを含んでいる。なお、変換リスト番号「0」は、シナリオの終了を表している。

【0108】

図12は、図11に示す第2の指定コマンドによって設定される第2の音声認識シナリオを示す図である。まず、シナリオ設定部33は、図11に示す指定コマンドによって指定されているシナリオ開始時の音声再生番号が「1」であるので、音声再生1の第1番目～第5番目のフレーズ番号「1」、「2」、「6」、「10」、「14」に従って、音声再生データ格納部31に格納されている複数の音声再生データの内から対応するフレーズP1、P2、P6、P10、P14を表す音声再生データを選択する。

30

【0109】

また、シナリオ設定部33は、図11に示す指定コマンドにおいて音声再生1に対応する変換リスト番号「4」に従って、変換リスト格納部32に格納されている複数の変換リストの内から変換リスト番号「4」に対応する変換リストL4を選択する。

【0110】

標準パターン抽出部36は、変換リストL4に含まれている音声認識選択肢A1「麺類」、音声認識選択肢A2「定食」、音声認識選択肢A3「デザート」の先頭の音節に含まれている音素のそれぞれについて、対応する標準パターンを音声認識データベースから抽出する。

40

【0111】

一方、音声信号合成部38は、フレーズP1、P2、P6、P10、P14を表す音声再生データに基づいて、音声信号を合成してD/A変換器40に出力する。また、D/A変換器40は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部50に出力する。これにより、音声出力部50から、「メニューの案内です。麺類、定食、デザートから選択してください。」というメッセージ（音声再生1）が発せられる。

【0112】

音声出力部50から発せられたメッセージ（音声再生1）に対して、ユーザーが、「デ

50

ザートにします。」と言うと、信号処理部 34 は、音素「d・e・z・a・H・t・o・
・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

【0113】

一致検出部 37 は、信号処理部 34 によって生成された先頭の音節の第 1 番目の音素「
d」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第 1 番目の音
素の標準パターンとを比較することにより、音素「d」の一致を検出する。一致が検出さ
れた音素が子音を表しているので、さらに、一致検出部 37 が、先頭の音節の第 2 番目の
音素を比較することにより、音素「e」の一致を検出する。これにより、先頭の音節「デ
」の一致が検出される。一致検出部 37 は、一致が検出された先頭の音節「デ」を有する
音声認識選択肢「デザート」を特定する音声認識結果をホスト CPU 61 に出力する。

10

【0114】

このようにして第 1 回目の音声認識動作が終了すると、シナリオ設定部 33 は、図 11
に示す指定コマンドにおける変換リスト 4 の選択肢 3 が選択された後の音声再生番号が「
4」であるので、音声再生 4 の第 1 番目～第 4 番目のフレーズ番号「11」、「12」、
「13」、「14」に従って、音声再生データ格納部 31 に格納されている複数の音声再
生データの内から対応するフレーズ P11、P12、P13、P14 を表す音声再生デー
タを選択する。

【0115】

また、シナリオ設定部 33 は、図 11 に示す指定コマンドにおいて音声再生 4 に対応す
る変換リスト番号「3」に従って、変換リスト格納部 32 に格納されている複数の変換リ
ストの内から変換リスト番号「3」に対応する変換リスト L3 を選択する。

20

【0116】

標準パターン抽出部 36 は、変換リスト L3 に含まれている音声認識選択肢 A1「ソフ
トクリーム」、音声認識選択肢 A2「プリン」、音声認識選択肢 A3「ショートケーキ」
の先頭の音節に含まれている音素のそれぞれについて、対応する標準パターンを音声認識
データベースから抽出する。

【0117】

一方、音声信号合成部 38 は、フレーズ P11、P12、P13、P14 を表す音声再
生データに基づいて、音声信号を合成して D/A 変換器 40 に出力する。また、D/A 変
換器 40 は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信
号を音声出力部 50 に出力する。これにより、音声出力部 50 から、「ソフトクリーム、
プリン、ショートケーキから選択してください。」というメッセージ（音声再生 4）が発
せられる。

30

【0118】

音声出力部 50 から発せられたメッセージ（音声再生 4）に対して、ユーザーが、「ソ
フトクリームにします。」と言うと、信号処理部 34 は、音素「s・o・h・u・t・o
・・・」のそれぞれについて、周波数成分の分布状態を表す特徴パターンを生成する。

【0119】

一致検出部 37 は、信号処理部 34 によって生成された先頭の音節の第 1 番目の音素「
s」の特徴パターンと、音声認識データベースから抽出された先頭の音節の第 1 番目の音
素の標準パターンとを比較することにより、音素「s」の一致を検出する。一致が検出さ
れた音素が子音を表しているので、さらに、一致検出部 37 が、先頭の音節の第 2 番目の
音素を比較することにより、音素「o」の一致を検出する。これにより、先頭の音節「ソ
」の一致が検出される。一致検出部 37 は、一致が検出された先頭の音節「ソ」を有する
音声認識選択肢「ソフトクリーム」を特定する音声認識結果をホスト CPU 61 に出力す
る。

40

【0120】

このようにして第 2 回目の音声認識動作が終了すると、シナリオ設定部 33 は、図 11
に示す指定コマンドにおける変換リスト 3 の選択肢 1 が選択された後の音声再生番号が「
5」であるので、音声再生 5 の第 1 番目のフレーズ番号「15」に従って、音声再生デー

50

タ格納部 3 1 に格納されている複数の音声再生データの内から対応するフレーズ P 1 5 を表す音声再生データを選択する。

【 0 1 2 1 】

音声信号合成部 3 8 は、フレーズ P 1 5 を表す音声再生データに基づいて、音声信号を合成して D / A 変換器 4 0 に出力する。また、D / A 変換器 4 0 は、デジタルの音声信号をアナログの音声信号に変換して、アナログの音声信号を音声出力部 5 0 に出力する。これにより、音声出力部 5 0 から、「ありがとうございます。」というメッセージ（音声再生 5 ）が発せられる。シナリオ設定部 3 3 は、図 1 1 に示す指定コマンドにおいて音声再生 5 に対応する変換リスト番号「 0 」に従って、一連の音声認識動作を終了させる。

【 0 1 2 2 】

以上の実施形態においては、本発明を自動販売機に適用した具体例について説明したが、本発明は、この実施形態に限定されるものではなく、一般的な電子機器に適用可能であると共に、当該技術分野において通常の知識を有する者によって、本発明の技術的思想内で多くの変形が可能である。

【 符号の説明 】

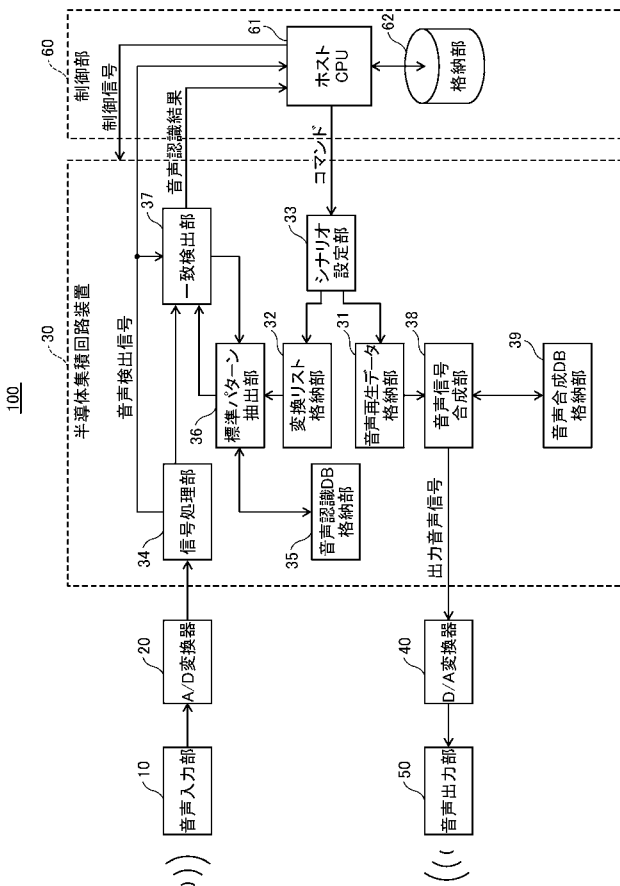
【 0 1 2 3 】

1 0 ... 音声入力部、 2 0 ... A / D 変換器、 3 0 ... 半導体集積回路装置、 3 1 ... 音声再生データ格納部、 3 2 ... 変換リスト格納部、 3 3 ... シナリオ設定部、 3 4 ... 信号処理部、 3 5 ... 音声認識データベース格納部、 3 6 ... 標準パターン抽出部、 3 7 ... 一致検出部、 3 8 ... 音声信号合成部、 3 9 ... 音声合成データベース格納部、 4 0 ... D / A 変換器、 5 0 ... 音声出力部、 6 0 ... 制御部、 6 1 ... ホスト CPU、 6 2 ... 格納部、 1 0 0 ... 音声認識装置

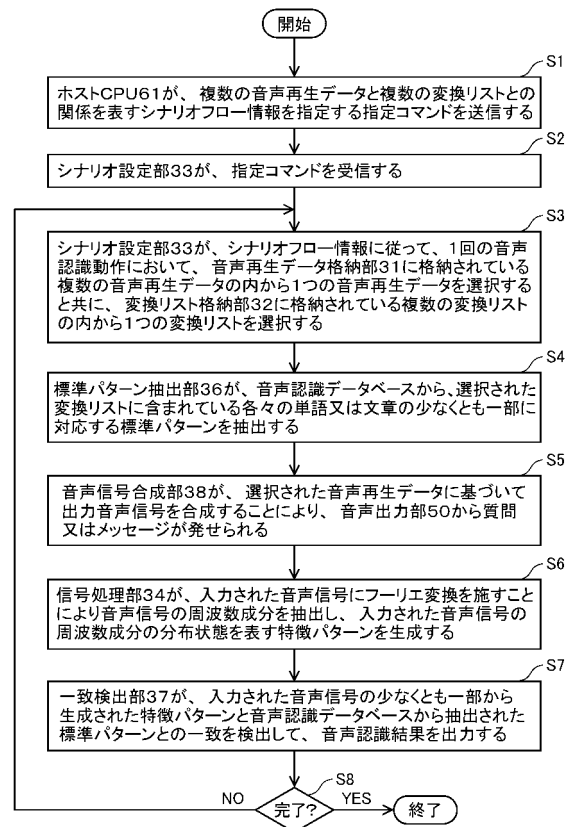
10

20

【 図 1 】



【 図 2 】



【 図 3 】

音声再生番号	質問又はメッセージの内容
1	「どの食品にしますか？食品名を言って下さい」
2	「温かい、冷たい、どちらにしますか？」
3	「大盛り、普通、どちらにしますか？」
4	「ありがとうございます」

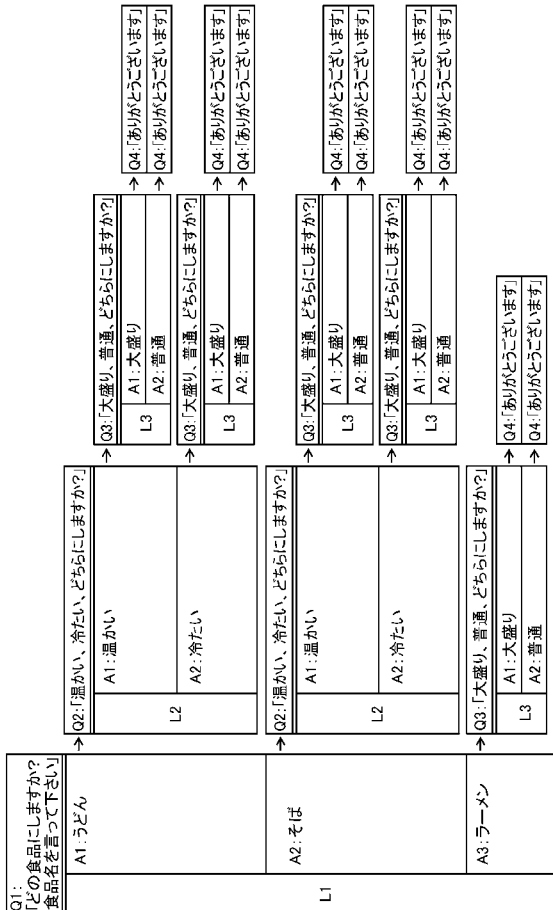
【 図 4 】

変換リスト番号	選択肢番号	音声認識選択肢
1	1	うどん
	2	そば
	3	ラーメン
2	1	温かい
	2	冷たい
3	1	大盛り
	2	普通

【 図 5 】

音声再生データ数	4
音声再生番号1に対応する変換リスト番号	1
音声再生番号2に対応する変換リスト番号	2
音声再生番号3に対応する変換リスト番号	3
音声再生番号4に対応する変換リスト番号	0
変換リスト数	3
変換リスト1の選択肢数	3
変換リスト1の選択肢1が認識された後の音声再生番号	2
変換リスト1の選択肢2が認識された後の音声再生番号	2
変換リスト1の選択肢3が認識された後の音声再生番号	3
変換リスト2の選択肢数	2
変換リスト2の選択肢1が認識された後の音声再生番号	3
変換リスト2の選択肢2が認識された後の音声再生番号	3
変換リスト3の選択肢数	2
変換リスト3の選択肢1が認識された後の音声再生番号	4
変換リスト3の選択肢2が認識された後の音声再生番号	4
シナリオ開始時の音声再生番号	1

【 図 6 】



【 図 7 】

フレーズ番号	フレーズの内容
1	「メニューの案内です」
2	「種類」
3	「うどん」
4	「そば」
5	「ラーメン」
6	「定食」
7	「日替わり定食」
8	「とんかつ定食」
9	「おさしみ定食」
10	「デザート」
11	「ソフトクリーム」
12	「プリン」
13	「ショートケーキ」
14	「から選択してください」
15	「ありがとうございます」

【 図 8 】

変換リスト番号	選択肢番号	音声認識選択肢
1	1	うどん
	2	そば
	3	ラーメン
2	1	日替わり定食
	2	とんかつ定食
	3	おさしみ定食
3	1	ソフトクリーム
	2	プリン
	3	ショートケーキ
4	1	種類
	2	定食
	3	デザート

【 図 9 】

音声再生設定数	2
音声再生1のフレーズ数	5
音声再生1の第1番目のフレーズ番号	1
音声再生1の第2番目のフレーズ番号	3
音声再生1の第3番目のフレーズ番号	4
音声再生1の第4番目のフレーズ番号	5
音声再生1の第5番目のフレーズ番号	14
音声再生1に対応する変換リスト番号	1
音声再生2のフレーズ数	1
音声再生2の第1番目のフレーズ番号	15
音声再生2に対応する変換リスト番号	0
変換リスト設定数	1
変換リスト1の選択肢数	3
変換リスト1の選択肢1が選択された後の音声再生番号	2
変換リスト1の選択肢2が選択された後の音声再生番号	2
変換リスト1の選択肢3が選択された後の音声再生番号	2
シナリオ開始時の音声再生番号	1

【 図 1 0 】

音声再生1	P1:「メニューの案内です」		
	P3:「うどん」		
	P4:「そば」		
	P5:「ラーメン」		
	P14:「から選択してください」		
L1	A1:うどん	→	音声再生2 P15:「ありがとうございます」
	A2:そば	→	音声再生2 P15:「ありがとうございます」
	A3:ラーメン	→	音声再生2 P15:「ありがとうございます」

【 図 1 1 】

音声再生設定数	5
音声再生1のフレーズ数	5
音声再生1の第1番目のフレーズ番号	1
音声再生1の第2番目のフレーズ番号	2
音声再生1の第3番目のフレーズ番号	6
音声再生1の第4番目のフレーズ番号	10
音声再生1の第5番目のフレーズ番号	14
音声再生1に対応する変換リスト番号	4
音声再生2のフレーズ数	4
音声再生2の第1番目のフレーズ番号	3
音声再生2の第2番目のフレーズ番号	4
音声再生2の第3番目のフレーズ番号	5
音声再生2の第4番目のフレーズ番号	14
音声再生2に対応する変換リスト番号	1
音声再生3のフレーズ数	4
音声再生3の第1番目のフレーズ番号	7
音声再生3の第2番目のフレーズ番号	8
音声再生3の第3番目のフレーズ番号	9
音声再生3の第4番目のフレーズ番号	14
音声再生3に対応する変換リスト番号	2
音声再生4のフレーズ数	4
音声再生4の第1番目のフレーズ番号	11
音声再生4の第2番目のフレーズ番号	12
音声再生4の第3番目のフレーズ番号	13
音声再生4の第4番目のフレーズ番号	14
音声再生4に対応する変換リスト番号	3
音声再生5のフレーズ数	1
音声再生5の第1番目のフレーズ番号	15
音声再生5に対応する変換リスト番号	0
変換リスト設定数	4
変換リスト1の選択肢数	3
変換リスト1の選択肢1が選択された後の音声再生番号	5
変換リスト1の選択肢2が選択された後の音声再生番号	5
変換リスト1の選択肢3が選択された後の音声再生番号	5
変換リスト2の選択肢数	3
変換リスト2の選択肢1が選択された後の音声再生番号	5
変換リスト2の選択肢2が選択された後の音声再生番号	5
変換リスト2の選択肢3が選択された後の音声再生番号	5
変換リスト3の選択肢数	3
変換リスト3の選択肢1が選択された後の音声再生番号	5
変換リスト3の選択肢2が選択された後の音声再生番号	5
変換リスト3の選択肢3が選択された後の音声再生番号	5
変換リスト4の選択肢数	3
変換リスト4の選択肢1が選択された後の音声再生番号	2
変換リスト4の選択肢2が選択された後の音声再生番号	3
変換リスト4の選択肢3が選択された後の音声再生番号	4
シナリオ開始時の音声再生番号	1

【 図 1 2 】

