

(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2017-138536
(P2017-138536A)

(43) 公開日 平成29年8月10日(2017.8.10)

(51) Int.Cl.			F I			テーマコード (参考)		
G 1 O L	15/30	(2013.01)	G 1 O L	15/30		5 K 1 2 7		
G 1 O L	15/20	(2006.01)	G 1 O L	15/20	3 7 O F			
G 1 O L	15/00	(2013.01)	G 1 O L	15/00	2 0 0 A			
H O 4 M	1/00	(2006.01)	H O 4 M	1/00	V			
H O 4 M	1/60	(2006.01)	H O 4 M	1/60	C			

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号 特願2016-20631 (P2016-20631)
(22) 出願日 平成28年2月5日(2016.2.5)

(71) 出願人 392026693
株式会社 N T T ドコモ
東京都千代田区永田町二丁目 1 1 番 1 号
(74) 代理人 100088155
弁理士 長谷川 芳樹
(74) 代理人 100113435
弁理士 黒木 義樹
(74) 代理人 100121980
弁理士 沖山 隆
(74) 代理人 100128107
弁理士 深石 賢治
(72) 発明者 松岡 保静
東京都千代田区永田町二丁目 1 1 番 1 号
株式会社 N T T ドコモ内
F ターム(参考) 5K127 BA03 BB24 BB33 CA27 DA15
GD09 KA04 MA02 MA34

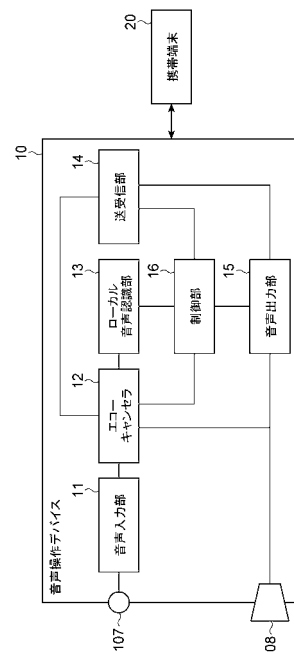
(54) 【発明の名称】 音声処理装置

(57) 【要約】

【課題】 L S R から D S R に切り替えられる場合に適切に音声認識を行う。

【解決手段】 音声操作デバイス 1 0 は、音声認識対象の音声を入力する音声入力部 1 1 と、入力された音声認識対象の音声に対して L S R を行って、当該 L S R の結果に基づいて D S R を行うか否かを判断するローカル音声認識部 1 3 と、D S R を行うと判断された場合に、音声認識対象の音声を、音声認識を行う装置に送信する送受信部 1 4 と、D S R を行うと判断された場合に、音声に関する予め設定された制御を行う制御部 1 6 とを備える。

【選択図】 図 1



【特許請求の範囲】**【請求項 1】**

音声認識対象の音声を入力する音声入力手段と、
前記音声入力手段によって入力された前記音声認識対象の音声に対して L S R を行って、当該 L S R の結果に基づいて D S R を行うか否かを判断する判断手段と、
前記判断手段によって D S R を行うと判断された場合に、前記音声認識対象の音声を、音声認識を行う装置に送信する音声送信手段と、
前記判断手段によって D S R を行うと判断された場合に、音声に関する予め設定された制御を行う制御手段と、
を備える音声処理装置。

10

【請求項 2】

前記音声送信手段は、近距離無線通信によって前記音声認識対象の音声を送信し、
前記制御手段は、前記近距離無線通信のプロトコルを予め設定されたプロトコルに設定する、請求項 1 に記載の音声処理装置。

【請求項 3】

前記音声認識対象以外の音声を入力して音声出力する音声出力手段を更に備え、
前記制御手段は、前記音声出力の音量を変更する、請求項 1 又は 2 に記載の音声処理装置。

【請求項 4】

前記音声認識対象以外の音声を入力して音声出力する音声出力手段と、
前記音声出力される音声に基づいてフィルタを更新して、当該フィルタを用いて、前記音声入力手段によって入力された前記音声認識対象の音声のエコー除去を行うエコー除去手段と、を更に備え、
前記制御手段は、前記エコー除去手段によるフィルタの更新の範囲を制限する、請求項 1 ~ 3 の何れか一項に記載の音声処理装置。

20

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、音声を入力して処理する音声処理装置に関する。

【背景技術】

30

【0002】

従来から、ユーザからの音声をマイク等で入力した装置が、音声をサーバに送信し、サーバが音声認識を行うシステムが知られている（例えば、特許文献 1 参照）。

【先行技術文献】**【特許文献】****【0003】**

【特許文献 1】特開 2014 - 60657 号公報

【発明の概要】**【発明が解決しようとする課題】****【0004】**

40

上記のようなシステムの一形態として、音声を入力した装置が、自装置において L S R（ローカル型音声認識）を行って、認識された音声に基づいて D S R（分散型音声認識）を行うか否かを判断するというものがある。当該装置では、L S R が行われる状態と、D S R が行われる状態とが存在する。当該装置における L S R が行われる状態は、D S R を行うことを考慮したものではない場合がある。そのため、装置の振る舞いがそれぞれの状態で同様であるとする D S R での音声認識ができない、あるいはしにくくなるという問題がある。例えば、D S R のための音声の送信や入力が適切に行われぬおそれがある。

【0005】

本発明は、上記に鑑みてなされたものであり、L S R から D S R に切り替えられる場合に適切に音声認識を行うことができる音声処理装置を提供することを目的とする。

50

【課題を解決するための手段】**【0006】**

上記目的を達成するために、本発明に係る音声処理装置は、音声認識対象の音声を入力する音声入力手段と、音声入力手段によって入力された音声認識対象の音声に対してLSRを行って、当該LSRの結果に基づいてDSRを行うか否かを判断する判断手段と、判断手段によってDSRを行うと判断された場合に、音声認識対象の音声を、音声認識を行う装置に送信する音声送信手段と、判断手段によってDSRを行うと判断された場合に、音声に関する予め設定された制御を行う制御手段と、を備える。

【0007】

本発明に係る音声処理装置では、LSRからDSRに切り替えられる場合に音声に関する予め設定された制御が行われたため、DSRが行いやすい状態（音声認識しやすい状態）にすることができる。これにより、本発明に係る音声処理装置によれば、LSRからDSRに切り替えられる場合に適切に音声認識を行うことができる。

10

【0008】

音声送信手段は、近距離無線通信によって音声認識対象の音声を送信し、制御手段は、近距離無線通信のプロトコルを予め設定されたプロトコルに設定する、こととしてもよい。この構成によれば、LSRが行われる場合と、DSRが行われる場合とで、プロトコルを切り替えることができる。これにより、各場合に合ったプロトコルにすることができる。従って、それぞれの場合に適切な近距離無線通信を行うことができ、LSRからDSRに切り替えられる場合に適切に音声を送信することができる。

20

【0009】

音声処理装置は、音声認識対象以外の音声を入力して音声出力する音声出力手段を更に備え、制御手段は、音声出力の音量を変更する、こととしてもよい。この構成によれば、例えば、音楽の音声出力を行いつつ、DSRが行われる場合に適切に音声を入力することができる。

【0010】

音声処理装置は、音声認識対象以外の音声を入力して音声出力する音声出力手段と、音声出力される音声に基づいてフィルタを更新して、当該フィルタを用いて、音声入力手段によって入力された音声認識対象の音声のエコー除去を行うエコー除去手段と、を更に備え、制御手段は、エコー除去手段によるフィルタの更新の範囲を制限する、こととしてもよい。この構成によれば、例えば、音楽の音声出力を行いつつ、DSRが行われる場合にエコーの除去を行うことができ、適切に音声認識のための音声を取得することができる。

30

【発明の効果】**【0011】**

本発明では、LSRからDSRに切り替えられる場合に音声に関する予め設定された制御が行われたため、DSRが行いやすい状態（音声認識しやすい状態）にすることができる。これにより、本発明によれば、LSRからDSRに切り替えられる場合に適切に音声認識を行うことができる。

【図面の簡単な説明】**【0012】**

【図1】本発明の実施形態に係る音声処理装置である音声操作デバイスの構成を示す図である。

40

【図2】本発明の実施形態に係る音声処理装置である音声操作デバイスのハードウェア構成を示す図である。

【図3】本発明の実施形態に係る音声処理装置である音声操作デバイスで実行される処理を示すフローチャートである。

【発明を実施するための形態】**【0013】**

以下、図面と共に本発明に係る音声処理装置の実施形態について詳細に説明する。なお、図面の説明においては同一要素には同一符号を付し、重複する説明を省略する。

50

【 0 0 1 4 】

図 1 に本実施形態に係る音声処理装置である音声操作デバイス 10 を示す。音声操作デバイス 10 は、ユーザの音声によって操作を行うことができる装置である。例えば、音声操作デバイス 10 は、音楽を再生することができる機器である。音声操作デバイス 10 は、ユーザの居室に配置して用いられる。音声操作デバイス 10 は、近距離無線通信により、携帯端末 20 から再生される音楽のデータを受信して音声出力を行う。携帯端末 20 は、ユーザによって携帯される装置であり、例えば、スマートフォン又は携帯電話機に相当する。音声操作デバイス 10 と、携帯端末 20 とを含むシステムにより、携帯端末 20 に格納されている音楽を、音声操作デバイス 10 が備える（携帯端末 20 が備えるものよりも上質の）スピーカで再生することができる。ユーザの音声による音声操作デバイス 10 の操作は、例えば、音楽の再生開始及び再生停止等である。

10

【 0 0 1 5 】

ユーザの音声による音声操作デバイス 10 の操作は、次のように行われる。まず、ユーザは、音声操作デバイス 10 の操作を開始するトリガとなるキーワード（例えば、「操作開始」といったキーワード）を発話する。当該キーワードは、予め音声操作デバイス 10 に設定されている。音声操作デバイス 10 では、ユーザの音声を受け付けられて、当該音声に対して L S R が行われる。続いて、L S R の結果に基づいてユーザがキーワードを発話したか否かが判断される。ユーザがキーワードを発話したと判断された場合、音声操作デバイス 10 は、ユーザの音声による音声操作デバイス 10 の操作を受け付ける状態となる。

20

【 0 0 1 6 】

続いて、ユーザは、音声操作デバイス 10 の操作に係る内容を発話する。音声操作デバイス 10 では、ユーザの音声を受け付けられて、当該音声に対して D S R が行われる。具体的には、音声操作デバイス 10 から携帯端末 20 に対して、音声（音声のデータ）が送信される。続いて、音声認識サーバによって、当該音声に対して音声認識が行われる。当該音声認識の結果に基づいて、音声操作デバイス 10 の操作が行われる。なお、音声認識サーバは、携帯端末 20 であってもよいし、携帯端末 20 から音声を送信される別の装置であってもよい。以上のようにユーザの音声による音声操作デバイス 10 の操作が行われる。上記のように音声操作デバイス 10 では、L S R をトリガとして、D S R（D S R のための音声の送信）が行われる。

30

【 0 0 1 7 】

引き続き、本実施形態に係る音声操作デバイス 10 の機能について説明する。図 1 に示すように音声操作デバイス 10 は、音声入力部 11 と、エコーキャンセラ 12 と、ローカル音声認識部 13 と、送受信部 14 と、音声出力部 15 と、制御部 16 とを備える。また、音声操作デバイス 10 は、音声の入出力に係るハードウェアとして、マイク 107 とスピーカ 108 とを備える。

【 0 0 1 8 】

音声入力部 11 は、音声認識対象の音声を入力する音声入力手段である。音声入力部 11 は、音声認識対象の音声として、マイク 107 を介してユーザから発せられた音声を入力する（入力して録音する）。音声入力部 11 は、入力した音声をエコーキャンセラ 12 へ出力する。

40

【 0 0 1 9 】

エコーキャンセラ 12 は、音声入力部 11 から入力した音声のエコー除去を行うエコー除去手段である。上述したように音声操作デバイス 10 は、スピーカ 108 によって音声出力を行われるため、エコーキャンセラ 12 は、その音声出力によるエコーを除去する。エコーキャンセラ 12 によるエコー除去は、従来のエコーキャンセラと同様に行われる。即ち、エコーキャンセラ 12 は、スピーカ 108 によって音声出力される音声を、音声出力部 15 から入力して、当該音声に基づいてフィルタ（フィルタ係数）を更新して、当該フィルタを用いてエコー除去を行う。エコーキャンセラ 12 は、エコー除去後の音声を、L S R が行われる場合にはローカル音声認識部 13 に、D S R が行われる場合には送受信

50

部 1 4 に出力する。

【 0 0 2 0 】

ローカル音声認識部 1 3 は、エコーキャンセラ 1 2 から入力した音声に対して L S R を行って、当該 L S R の結果に基づいて D S R を行うか否かを判断する判断手段である。L S R は、従来の音声認識と同様に行われる。ローカル音声認識部 1 3 は、入力した音声に対して L S R を行って、L S R の結果に基づいて当該音声（に対応する文字列）が、予め記憶したキーワードを含む、あるいはキーワードと一致しているか否かを判断する（当該音声のキーワードの検出を行う）。ローカル音声認識部 1 3 は、当該音声の予め記憶したキーワードを含む、あるいはキーワードと一致していると判断した場合、D S R を行うと判断する。ローカル音声認識部 1 3 は、当該音声の予め記憶したキーワードを含まない、あるいはキーワードと一致していないと判断した場合、L S R を引き続き行うと判断する（この場合、特段の処理は行わない）。

10

【 0 0 2 1 】

ローカル音声認識部 1 3 は、D S R を行うと判断したら、その旨をエコーキャンセラ 1 2 及び制御部 1 6 に通知する。その旨を受けたエコーキャンセラ 1 2 は、それ以降、エコー除去後の音声を送受信部 1 4 に出力する。上記のように、音声操作デバイス 1 0 では、常時待ち受けの L S R で待ち受けのキーワードが検出されると、キーワードウェイクアップで D S R が起動される。

【 0 0 2 2 】

送受信部 1 4 は、近距離無線通信によって携帯端末 2 0 の間で情報の送受信を行う手段である。近距離無線通信は、例えば、B l u e t o o t h（登録商標）で行われる。近距離無線通信は、後述するように所定のプロトコルで行われる。送受信部 1 4 は、音楽のデータを携帯端末 2 0 から受信する。送受信部 1 4 は、受信した音楽のデータを音声出力部 1 5 に出力する。

20

【 0 0 2 3 】

また、送受信部 1 4 は、ローカル音声認識部 1 3 によって D S R を行うと判断された場合に、エコーキャンセラ 1 2 から入力した音声を、音声認識を行う装置に送信する音声送信手段である。具体的には、送受信部 1 4 は、エコーキャンセラ 1 2 から音声を入力すると、当該音声を近距離無線通信によって携帯端末 2 0 に送信する。

【 0 0 2 4 】

携帯端末 2 0 は、音声操作デバイス 1 0 から音声を受信し、受信した音声に対して音声認識を行う。携帯端末 2 0 は、音声認識の結果である文字列を音声操作デバイス 1 0 に送信する。送受信部 1 4 は、携帯端末 2 0 から送信された文字列を受信する。送受信部 1 4 は、受信した文字列を音声出力部 1 5 に出力する。

30

【 0 0 2 5 】

なお、音声認識は、必ずしも携帯端末 2 0 によって行われる必要はなく、携帯端末 2 0 以外の音声認識サーバによって行われてもよい。その場合、携帯端末 2 0 は、音声操作デバイス 1 0 から受信した音声を、例えば、移動体通信によって音声認識サーバに送信する。音声認識サーバは、携帯端末 2 0 から送信された音声を受信して、受信した音声に対して音声認識を行う。音声認識サーバは、音声認識の結果である文字列を携帯端末 2 0 に送信する。

40

【 0 0 2 6 】

ローカル音声認識部 1 3 による音声認識（L S R）、並びに携帯端末 2 0 又は音声認識サーバによる音声認識（D S R、サーバ型音声認識）は、例えば、音声認識エンジンが用いられて行われる。音声認識自体は、従来の任意の音声認識方法を利用することができる。

【 0 0 2 7 】

L S R は、D S R を起動とするトリガとするため、上記の通り、ユーザの発話にキーワードが含まれるか否か、あるいはユーザの発話とキーワードが一致するかを判断するためのものである。そのため、L S R は、当該キーワードの検出に特化したものであり、

50

通常、D S Rよりも、特定のキーワード以外の音声認識の精度が低い。

【0028】

音声出力部15は、音声認識対象以外の音声を入力して音声出力する音声出力手段である。具体的には、音声出力部15は、送受信部14から入力した音楽のデータに対して音声合成等を行って、スピーカ108から出力される形式にして、スピーカ108を介して音声出力する。即ち、音声出力部15は、音楽のデータを再生する。また、音声出力部15は、音声出力する音楽のデータをエコーキャンセラ12に出力して、エコー除去に用いられるようにする。

【0029】

また、音声出力部15は、送受信部14から入力した文字列に基づき、機能を実行する。この機能は、ユーザの音声による音声操作デバイス10の操作に係るものであり、例えば、上述したように音楽の再生開始及び再生停止等である。入力した文字列からの機能の実行は、従来と同様に行うこととすればよい。例えば、文字列と機能とを予め対応付けておき、当該対応付けに基づいて機能を実行することとすればよい。また、認識結果に基づいて実行される機能の決定は、携帯端末20等の音声操作デバイス10以外の装置で行われてもよい。その場合、実行される機能として決定された機能を実行させる指示（例えば、コマンド）が、携帯端末20等の音声操作デバイス10以外の装置から、音声操作デバイス10に送信されて、音声操作デバイス10において当該機能が実行される。

10

【0030】

なお、音声認識の結果の文字列に基づく機能の実行は、必ずしも音声出力部15に係る機能だけでなく、それ以外の機能に対して行われてもよい。あるいは、携帯端末20又はその他のサーバが、ユーザの発話（音声認識によって得られた文字列）に対しての回答となる文字列を生成して、音声操作デバイス10が、当該回答が音声出力されることとしてもよい。即ち、対話システムを構成することとしてもよい。

20

【0031】

制御部16は、ローカル音声認識部13からD S Rを行う旨を通知された場合に、音声に関する予め設定された制御を行う制御手段である。制御部16は、D S Rを行う場合に、適切に音声認識を行うことができるように制御を行う。具体的には、制御部16は、以下のような制御を行う。

【0032】

制御部16は、当該制御として、送受信部14と携帯端末20との間の近距離無線通信のプロトコルを予め設定されたプロトコルに設定する。上述したように携帯端末20から送受信部14には音楽のデータが送信される。そのため、D S Rを行う前（ユーザが、予め設定されたキーワードを発話する前）は、高品質の音楽データが送信できるように当該プロトコルとして、例えば、BluetoothのA2DP（Advanced Audio Distribution Profile）を設定しておく。A2DPでは、携帯端末20から音声操作デバイス10に対して、44.1kHzサンプリングの音声のデータ（音楽用高音質のデータ）を送信することができる。しかしながら、A2DPでは、音声のデータの双方向の送信を行うことができず、音声操作デバイス10から携帯端末20への音声の送信を行うことができない。

30

40

【0033】

ローカル音声認識部13からD S Rを行う旨を通知されると、制御部16は、当該プロトコルをA2DPからHFP（Hands-Free Profile）に変更するように送受信部14に指示する。HFPでは、音声操作デバイス10と携帯端末20との間で、双方向に16kHzサンプリングの音声のデータ（通話用低音質のデータ）を送信することができる。送受信部14は、当該指示を受け付けて、携帯端末20との間でHFPに従った近距離無線通信を行う。

【0034】

上記のようにプロトコルが設定されることで、D S Rが行われる場合に、音声操作デバイス10から携帯端末20に対して音声のデータを送信できるようにし、それ以外の場合

50

には、携帯端末 20 から音声操作デバイス 10 に高音質のデータを送信できるようになる。これにより、DSRが行われる場合には、音声は確実に音声認識されるようにし、それ以外の場合には音声操作デバイス 10 から高音質の音楽を再生させることができる。

【0035】

あるいは、制御部 16 は、当該制御として、スピーカ 108 からの音声出力の音量（ボリューム）、即ち、再生中の音声の音量を変更する。スピーカ 108 からの音声出力が行われていると、ユーザからの発話が入力されにくくなる。そこで、ローカル音声認識部 13 から DSR を行う旨を通知されると、制御部 16 は、音量を下げるように音声出力部 15 に指示する。音声出力部 15 は、当該指示を受け付けて、スピーカ 108 からの音声出力の音量を下げる。下げられる音量は、予め設定されている。

10

【0036】

あるいは、制御部 16 は、当該制御として、エコーキャンセラ 12 によるフィルタの更新の範囲を制限する。具体的には、制御部 16 は、当該フィルタの更新を停止させる。スピーカ 108 から音声が出力され、かつ、ユーザからの音声が入力されている状態でフィルタを更新させると、適切にエコーが除去されず、音声認識しにくくなるおそれがある。そこで、ローカル音声認識部 13 から DSR を行う旨を通知されると、制御部 16 は、フィルタの更新を停止するようエコーキャンセラ 12 に指示する。エコーキャンセラ 12 は、当該指示を受け付けて、フィルタの更新を停止し、当該指示を受けた時点でのフィルタを用いてエコー除去を行う。即ち、エコーキャンセラ 12 は、バージン（ダブルトーク）モードでエコー除去を行う。なお、上記の制御として、必ずしも、フィルタの更新を停止させる必要はなく、フィルタの更新の（数値）範囲を制限することとしてもよい。制限後の範囲は、例えば、その時点のフィルタの値を中心の値とした予め設定された大きさの範囲である。

20

【0037】

制御部 16 は、所定時間経過後に DSR を終了させると共に上記の制御を終了させる。例えば、制御部 16 は、ローカル音声認識部 13 から DSR を行う旨を通知されてから、予め設定した一定時間経過したら、DSR を終了させると共に上記の制御を終了させる。あるいは、制御部 16 は、送受信部 14 によって携帯端末 20 に送信される音声を監視しておき、一定時間、一定の音量以上の音声が入力されていない場合（即ち、一定時間、ユーザからの音声が入力されていない場合）に、DSR を終了させると共に上記の制御を終了させることとしてもよい。また、制御部 16 は、DSR を終了とあわせて LSR を開始させる。

30

【0038】

具体的には、制御部 16 は、送受信部 14、ローカル音声認識部 13 及びエコーキャンセラ 12 に対して、制御前の状態に戻すよう指示する。これにより、送受信部 14 と携帯端末 20 との間の近距離無線通信のプロトコルが、HFP から A2DP に変更される。また、スピーカ 108 からの音声出力の音量が、制御前の音量にされる。また、エコーキャンセラ 12 によるフィルタの更新が行われる。また、制御部 16 は、エコーキャンセラ 12 に対して、エコー除去後の音声の出力先を送受信部 14 からローカル音声認識部 13 に変更させる。以上が、音声操作デバイス 10 の機能である。

40

【0039】

図 2 に本実施形態に係る音声操作デバイス 10 のハードウェア構成を示す。図 2 に示すように、音声操作デバイス 10 は、CPU（Central Processing Unit）101、主記憶装置である RAM（Random Access Memory）102 及び ROM 103（Read Only Memory）、操作モジュール 104、近距離無線通信モジュール 105、近距離無線通信用アンテナ 106、マイク 107 並びにスピーカ 108 等のハードウェアにより構成されている。これらの構成要素がプログラム等により動作することにより、上述した音声操作デバイス 10 の各機能が発揮される。以上が、音声操作デバイス 10 の構成である。

【0040】

引き続き、図 3 のフローチャートを用いて、本実施形態に係る音声操作デバイス 10

50

で実行される処理（音声操作デバイス10の動作方法）を説明する。本処理の開始時には、音声操作デバイス10では、携帯端末20から音楽のデータが受信されて、スピーカ108から音声出力がなされている（音楽の再生がなされている）。また、エコーキャンセラ12では、音声出力される音声に基づいてフィルタの更新が行われている。また、ユーザの音声に対して、LSRが行われる状態となっている。

【0041】

本処理では、まず、音声入力部11によって、マイク107を介してユーザから発せられた音声が入力される（S01）。入力された音声は、音声入力部11からエコーキャンセラ12に出力される。続いて、エコーキャンセラ12によって、入力された音声のエコー除去（エコーキャンセル）が行われる（S02）。エコー除去後の音声は、エコーキャンセラ12からローカル音声認識部13に出力される。続いて、ローカル音声認識部13によって、エコーキャンセラ12から入力した音声に対してLSRが行われる（S03）。続いて、ローカル音声認識部13によって、LSRの結果に基づいて、当該音声中のキーワードの検出が行われる（S04）。

10

【0042】

当該音声中にキーワードが検出されなかった場合（当該音声にキーワードを含まない、あるいはキーワードと一致していないと判断した場合）（S04のNO）、引き続き、ユーザの音声が入力されて、上記のLSRの処理が行われる（S01～S04）。

【0043】

当該音声中にキーワードが検出された場合（当該音声にキーワードを含む、あるいはキーワードと一致していると判断した場合）（S04のYES）、DSRを行うと判断される。その旨が、ローカル音声認識部13からエコーキャンセラ12及び制御部16に通知される。当該通知を受けたエコーキャンセラ12では、以降のエコー除去後の音声を送受信部14に出力される。また、当該通知を受けた制御部16によって、適切に音声認識を行うことができるよう制御が行われる。

20

【0044】

具体的には、制御部16から音声出力部15に対して、音声出力の音量（再生音のボリューム）を下げる指示が行われる。音声出力部15では、当該指示が受け付けられて、スピーカ108からの音声出力の音量が下げられる（S05）。また、制御部16からエコーキャンセラ12に対して、フィルタの更新の停止させる指示が行われる。エコーキャンセラ12では、当該指示が受け付けられて、フィルタの更新が停止される（S06）。また、制御部16から送受信部14に対して、携帯端末20との間の近距離無線通信のプロトコルをA2DPからHFPに変更させる指示が行われる。送受信部14では、当該指示が受け付けられて、携帯端末20との間の近距離無線通信のプロトコルがA2DPからHFPに変更される（S07）。なお、上記のS05～S07は、必ずしも上記の順番で行われる必要はない。

30

【0045】

続いて、音声入力部11によって、マイク107を介してユーザから発せられた音声が入力される（S08）。入力された音声は、音声入力部11からエコーキャンセラ12に出力される。続いて、エコーキャンセラ12によって、入力された音声のエコー除去（エコーキャンセル）が行われる（S09）。エコー除去後の音声は、エコーキャンセラ12から送受信部14に出力される。続いて、送受信部14によって、エコーキャンセラ12から入力した音声的近距離無線通信によって携帯端末20に送信される（S10）。

40

【0046】

携帯端末20では、当該音声を受信されて、当該音声に対して音声認識が行われる。音声認識結果は、携帯端末20から音声操作デバイス10に近距離無線通信によって送信される。音声操作デバイス10では、送受信部14によって、音声認識結果を受信される（S11）。受信された音声認識結果は、例えば、音声出力部15に出力されて、機能の実行（音声操作デバイス10の音声操作）に用いられる。

【0047】

50

続いて、制御部 16 によって、D S R を終了させるか否かが判断される (S 1 2)。D S R を終了させないと判断された場合 (S 1 2 の N O)、引き続き、ユーザの音声が入力されて、上記の D S R の処理が行われる (S 0 8 ~ S 1 2)。D S R を終了させると判断された場合 (S 1 2 の Y E S)、制御部 16 から送受信部 14、ローカル音声認識部 13 及びエコーキャンセラ 12 に対して、制御前の状態に戻すよう指示が行われる。送受信部 14、ローカル音声認識部 13 及びエコーキャンセラ 12 では、当該指示が受け付けられて、制御前の状態に戻される (S 1 3)。また、制御部 16 からの指示により、エコーキャンセラ 12 からのエコー除去後の音声の出力先が、送受信部 14 からローカル音声認識部 13 に変更される。これにより、L S R が行われる状態となり、S 0 1 からの処理が繰り返される。以上が、本実施形態に係る音声操作デバイス 10 で実行される処理である。

10

【 0 0 4 8 】

上述したように本実施形態では、L S R から D S R に切り替えられる場合に音声に関する予め設定された制御が行われたため、D S R が行いやすい状態 (音声認識しやすい状態) にすることができる。これにより、本発明に係る音声処理装置によれば、適切に音声認識を行うことができる。

【 0 0 4 9 】

具体的には、本実施形態のように近距離無線通信のプロトコルを設定 (変更) することとしてもよい。この構成によれば、L S R が行われる場合と、D S R が行われる場合とで、プロトコルを切り替えることができる。これにより、各場合に合ったプロトコルにすることができる。例えば、本実施形態のように D S R が行われない場合には、音声操作デバイス 10 に対して高品質な音楽のデータを送信できる A 2 D P としたプロトコルを用いて、D S R が行われる場合には、双方向のデータを送信できる H F P としたプロトコルを用いることができる。

20

【 0 0 5 0 】

従って、D S R が行われる場合には、音声を携帯端末 20 に送信することを可能とする一方、D S R が行われず、音声を送信する必要がない場合には、それに適したプロトコルとすることができる。従って、それぞれの場合に適切な近距離無線通信を行うことができ、L S R から D S R に切り替えられる場合に適切に音声の送信を行うことができる。

【 0 0 5 1 】

また、本実施形態のように音声出力の音量を変更することとしてもよい。この構成によれば、例えば、本実施形態のように音楽の音声出力を行いつつ (音楽の再生を行いつつ)、D S R が行われる場合に適切に音声を入力することができる。

30

【 0 0 5 2 】

また、本実施形態のようにエコーキャンセラ 12 のフィルタの更新の範囲を制限する (例えば、上述したように当該フィルタの更新の停止) こととしてもよい。この構成によれば、例えば、本実施形態のように音楽の音声出力を行いつつ、D S R が行われる場合にエコーの除去を行うことができ、適切に音声認識のための音声を取得することができる。

【 0 0 5 3 】

なお、上記の各制御は、必ずしも全て行われる必要はなく、何れか 1 つあるいは 2 つのみが行われてもよい。また、本実施形態では、音声の送信先が、携帯端末 20 であることとしたが、携帯端末 20 以外の装置であってもよい。また、本実施形態では、音楽のデータは、携帯端末 20 から受信することとしたが、音声操作デバイス 10 において予め音楽のデータを記憶しておき、それを読み出して音声出力のために入力することとしてもよい。また、本実施形態では、音声操作デバイス 10 では、音楽の再生を行う機能を有しており、音楽の再生を行いつつ L S R 及び D S R を行うものとしたが、L S R 及び D S R が行われる際に音楽の再生以外の機能が実行されてもよい。

40

【 符号の説明 】

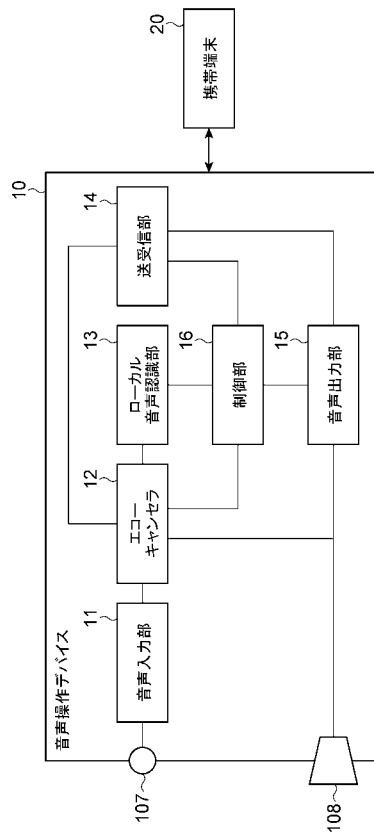
【 0 0 5 4 】

10 ... 音声操作デバイス、11 ... 音声入力部、12 ... エコーキャンセラ、13 ... ローカル音声認識部、14 ... 送受信部、15 ... 音声出力部、16 ... 制御部、101 ... C P U、1

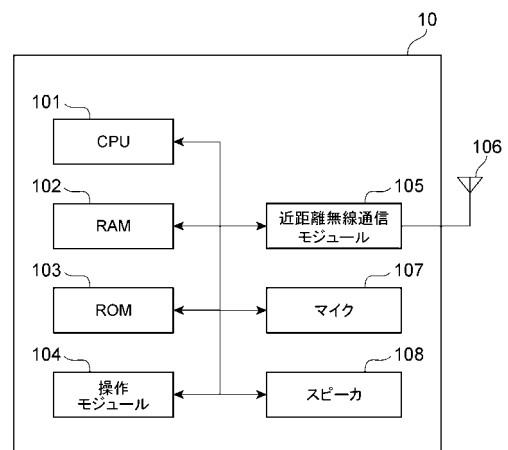
50

02 ... RAM、103 ... ROM、104 ... 操作モジュール、105 ... 近距離無線通信モジュール、106 ... 近距離無線通信アンテナ、107 ... マイク、108 ... スピーカ、20 ... 携帯端末。

【図1】



【図2】



【 図 3 】

