

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5419361号
(P5419361)

(45) 発行日 平成26年2月19日 (2014. 2. 19)

(24) 登録日 平成25年11月29日 (2013. 11. 29)

(51) Int. Cl.	F 1		
G 1 O L 15/30 (2013. 01)	G 1 O L 15/28	2 1 O A	
G 1 O L 15/28 (2013. 01)	G 1 O L 15/28	4 0 0	
G 1 O L 15/00 (2013. 01)	G 1 O L 15/00	2 0 0 A	

請求項の数 14 (全 17 頁)

(21) 出願番号	特願2008-19878 (P2008-19878)	(73) 特許権者	511204038
(22) 出願日	平成20年1月30日 (2008. 1. 30)		ニュアンス コミュニケーションズ, イ
(65) 公開番号	特開2008-191662 (P2008-191662A)		ンコーポレイテッド
(43) 公開日	平成20年8月21日 (2008. 8. 21)		アメリカ合衆国 マサチューセッツ 01
審査請求日	平成23年1月20日 (2011. 1. 20)		803, バーリントン, ウェイサイド
(31) 優先権主張番号	07002342.9		ロード 1
(32) 優先日	平成19年2月2日 (2007. 2. 2)	(74) 代理人	100078282
(33) 優先権主張国	欧州特許庁 (EP)		弁理士 山本 秀策
		(74) 代理人	100062409
			弁理士 安村 高明
		(74) 代理人	100113413
			弁理士 森下 夏樹
		(72) 発明者	マルクス バック
			ドイツ国 88400 ビーベラハ, ゲ
			シュヴィステル-ロメール-ヴェーク 4
			最終頁に続く

(54) 【発明の名称】 音声制御システムおよび音声制御方法

(57) 【特許請求の範囲】

【請求項 1】

音声制御システムの音声制御方法であって、該音声制御システムは、該音声制御システムの少なくとも一人のユーザからのスピーチ信号を受信および認識するための静止ユニットを備え、該方法は、

該音声制御システムの該少なくとも一人のユーザによって携帯される少なくとも1つのモバイルデバイスを提供することと、

該少なくとも1つのモバイルデバイスから該静止ユニットに制御信号を伝送することであって、該制御信号は、該少なくとも一人のユーザのスピーチとは関係ない、ことと、

該伝送された制御信号を該静止ユニットによって受信することと、

該制御信号から該少なくとも一人のユーザを識別し、該少なくとも一人のユーザのアイデンティティに基づいて該ユーザによるスピーチのスピーチ認識のためのスピーチ設定を構成することと、

該受信した制御信号に応答して、該静止ユニットを動作させることと

を包含し、

該静止ユニットを動作させることは、該音声制御システムの該少なくとも一人のユーザを位置決定することと、該少なくとも一人のユーザの位置決定に基づいて、該静止ユニットの一部であるマイクのアレイのマイクを調整することとを包含し、

該静止ユニットに、所定の可聴信号のパラメータを格納することと、

該モバイルデバイスから該静止ユニットに該所定の可聴信号を伝送することと、

10

20

該所定の可聴信号を該静止ユニットによって受信することと、
該受信した所定の可聴信号のパラメータと、該所定の可聴信号の該格納されたパラメータとを比較することにより、該静止ユニットが設けられている部屋の音響特性を解析することと、
該部屋の該解析された音響特性に基づいて、該スピーチ信号を認識するために使用される処理パラメータを適応させることと
を特徴とする、方法。

【請求項 2】

前記静止ユニットを動作させることは、該静止ユニットを作動させること、および/またはスピーチ信号を認識することを包含する、請求項 1 に記載の方法。

10

【請求項 3】

前記少なくとも一人のユーザからのスピーチ信号は、該少なくとも一人のユーザの識別に基づいて認識される、請求項 2 に記載の方法。

【請求項 4】

前記制御信号は、定期的にまたは前記少なくとも一人のユーザによる要求に応じて伝送される、請求項 1 ~ 請求項 3 のいずれか一項に記載の方法。

【請求項 5】

前記伝送および受信される制御信号は、赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号である、請求項 1 ~ 請求項 4 のいずれか一項に記載の方法。

20

【請求項 6】

前記制御信号は、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタント、またはヘッドセット、または補聴器、特に両耳補聴器によって伝送される、請求項 1 ~ 請求項 5 のいずれか一項に記載の方法。

【請求項 7】

前記少なくとも一人のユーザからのスピーチ信号を前記モバイルデバイスによって検出する追加的なステップと、

該検出されたスピーチ信号を該モバイルデバイスから前記静止ユニットに伝送する追加的なステップと、

該伝送されたスピーチ信号を該静止ユニットによって受信する追加的なステップと、
 該受信したスピーチ信号をスピーチ認識のために処理し、認識結果を取得する追加的なステップと、

30

該認識結果に基づいて、該静止ユニットを動作させる追加的なステップと
を包含する、請求項 1 ~ 請求項 6 のいずれか一項に記載の方法。

【請求項 8】

前記少なくとも一人のユーザからのスピーチ信号を前記モバイルデバイスによって検出する追加的なステップと、

該検出されたスピーチ信号を該モバイルデバイスによって認識し、認識結果を取得する追加的なステップと

を包含し、

40

該少なくとも 1 つのモバイルデバイスから前記静止ユニットに伝送される前記制御信号は、該認識結果に基づいて生成される、請求項 1 ~ 請求項 6 のいずれか一項に記載の方法。

【請求項 9】

音声制御システムであって、該音声制御システムは、
スピーチ信号を受信することと、所定の可聴信号のパラメータを格納することとを実行するように構成された静止ユニットと、

該静止ユニットに制御信号および該所定の可聴信号を伝送するように構成された少なくとも 1 つのモバイルデバイスであって、該制御信号は、該音声制御システムの少なくとも一人のユーザのスピーチとは関係ない、モバイルデバイスと

50

を備え、

該静止ユニットはまた、該少なくとも1つのモバイルデバイスによって伝送された該制御信号および該所定の可聴信号を受信し、該受信した制御信号に応答して、動作するようにも構成されており、

該静止ユニットはさらに、

(A) 処理ユニットであって、該処理ユニットは、(a) 該制御信号から該少なくとも一人のユーザを識別し、該少なくとも一人のユーザのアイデンティティに基づいて該ユーザによるスピーチのスピーチ認識のためのスピーチ設定を構成することと、(b) 該受信した制御信号に応答して、該モバイルデバイスを位置決定することと、(c) 該受信した所定の可聴信号と、該所定の可聴信号の該格納されたパラメータとを比較することにより、該静止ユニットが設けられている部屋の音響特性を解析することと、(d) 該部屋の該解析された音響特性に基づいて、該スピーチ信号を認識するために使用される処理パラメータを適応させることとを実行するように構成される、処理ユニットと、

10

(B) マイク、特に少なくとも1つの指向性マイクを含む、マイクのアレイと

を備え、

該静止ユニットは、該モバイルデバイスの位置決定に応答して、該マイクのアレイを調整するように構成される、音声制御システム。

【請求項10】

前記静止ユニットは、前記受信した制御信号に応答して、作動するように構成されており、該受信した制御信号に応答して、スピーチ信号を認識し、かつ/または前記モバイルデバイスを識別するように構成された、処理ユニットをさらに備える、請求項9に記載の音声制御システム。

20

【請求項11】

前記少なくとも1つのモバイルデバイスは、赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を、前記制御信号として伝送するように構成されており、前記静止ユニットは、該伝送された赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を受信するように構成されている、請求項9～請求項10のいずれか一項に記載の音声制御システム。

【請求項12】

前記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタント、またはヘッドセット、または補聴器、特に両耳補聴器である、請求項9～請求項11のいずれか一項に記載の音声制御システム。

30

【請求項13】

前記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、該少なくとも1つのモバイルデバイスはまた、前記少なくとも一人のユーザからのスピーチ信号を検出し、該検出されたスピーチ信号を前記静止ユニットに伝送するようにも構成されており、

該静止ユニットは、該モバイルデバイスによって伝送された該スピーチ信号を認識し、認識結果を取得し、該受信した制御信号に応答して動作するように構成されている、請求項9～請求項12のいずれか一項に記載の音声制御システム。

40

【請求項14】

前記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、該少なくとも1つのモバイルデバイスは、該少なくとも一人のユーザからのスピーチ信号を検出し、該検出されたスピーチ信号を認識し、認識結果を取得し、該認識結果に基づいて、前記制御信号を生成するように構成されている、請求項9～請求項13のいずれか一項に記載の音声制御システム。

【発明の詳細な説明】

【技術分野】

【0001】

(発明の分野)

50

本発明は、音声制御方法および音声制御システムに関する。特に、本発明は、ユーザからのスピーチ信号を認識するための静止ユニットを含む音声制御システムの音声制御方法に関する。

【背景技術】

【0002】

(発明の背景)

スピーチ認識システム、すなわち、音響信号を語 (word) または記号列 (string) の系列に変換するためのデバイスは、最近ますます普及してきている。近年では、スピーチ認識技術における劇的な進歩のおかげで、高性能のスピーチ解析、認識アルゴリズム、およびスピーチダイアログシステムが、一般的に利用可能になってきた。現在のスピーチ入力能力は、音声ダイアル、コールルーティング、文書作成等を含んでいる。スピーチダイアログシステムは、例えば、自動車に使用され得、ユーザが様々なデバイス (例えば、モバイル電話、カーラジオ、ナビゲーションシステム、および/またはエアコン) を制御することを可能にし得る。

【0003】

しかしながら、多くのアプリケーションにおいて、当該技術分野の音声制御システムの信頼性および快適性は、十分に満足できるものではない。現在の音声制御システムは、静止マイク (アレイ) による、静止している話者 (speaker) (例えば、自動車の乗客) の言葉の認識、または1つ以上のマイクを含むモバイルデバイス (例えば、ヘッドセットまたはセル電話) による、移動している話者の言葉の認識に限定されている。スピーチ認識システムの処理は、検出されたスピーチ信号の品質に多大に依存する。スピーチ信号の品質は、話者と (単数または複数の) マイクとの間の距離が増大するに伴って悪くなる。

【0004】

遠くの話のスピーチ収集における信号品質およびスピーチ認識手段の性能を向上させるための通常の方法は、マルチチャネルシステム (すなわち、例えば非特許文献1に記載されているようなマイクのアレイ) の利用である。

【0005】

現在のマルチチャネルシステムは、いわゆる「一般化サイドローブキャンセラー (GSC)」をしばしば使用する。例えば、非特許文献2を参照されたい。GSCは、2つの信号処理経路: ブロッキングマトリックスおよび適応ノイズキャンセリング手段を有する第1 (または低い方) の適応経路、および固定ビームフォーマを有する第2 (または高い方) の非適応経路から構成されている。

【0006】

固定ビームフォーマは、例えば、固定ビームパターンを用いる時間遅延補償手段によって前処理された信号を改善する。適応処理方法は、システムの動作の間の処理パラメータ (例えば、フィルタ係数) の永続的な適応を特徴とする。GSCの低い方の信号処理経路は、固定ビームフォーマの出力信号の残余ノイズを除去するために使用されるノイズ基準信号を生成するように最適化される。

【0007】

しかしながら、ビーム形成デバイスの性能は、話者の位置または所望の信号の検出のための所定の優先的方向の情報に依存する。移動している話者の場合、通常、閉ざされた部屋よりもインパルス応答がいくぶん複雑になるので、所望の信号のソースの検出の評価は、いくぶん困難になる。

【0008】

さらに、話者の位置決定は、非常に迅速に実行されなければならない。なぜならば、通常、検出されたスピーチ信号そのものが、移動している話者の位置決定のために使用されるからである。加えて、話者は話者が実際に話しているときにのみ位置決定され得る。したがって、信頼性のあるスピーチ認識のためには、高速信号処理が必要である。しかしながら、近年における工学的処理をもってしても、そのような信号処理は、依然としてエラ

10

20

30

40

50

ーを生じやすく、費用がかかる。

【0009】

関連する問題は、音声制御システムは通常、例えばプッシュアウトボタンを用いることによって、触覚的に作動しなければならないということである。このことは、多くのユーザによって不快に思われ、音声によって動作され得るシステムの応答の一部が欠けるという結果にもなる。別の関連する問題は、特定の話者の優先順位付けである。特定の話者の識別を要求する音声制御は、特定のユーザが話者の必要性および優先度にしたがって音声制御システムを動作させ得るように、個別的に提供され得る。音声制御のそのようなカスタマイズは、音声制御システムの信頼性を向上させ、消費者に受け入れられやすくするためには、望ましい。

10

【非特許文献1】Brandstein, M. および Ward, D. 編、「Microphone Arrays: Signal Processing Techniques and Applications」、Springer、Berlin 2001

【非特許文献2】Griffiths, L. J. および Jim, C. W.、「An alternative approach to linearly constrained adaptive beamforming」、IEEE Transactions on Antennas and Propagation、1982、vol. 30、p. 27

【発明の開示】

【課題を解決するための手段】

20

【0010】

(発明の記載)

上述の問題を踏まえると、本発明は、音声制御システムの音声制御方法を提供し、この音声制御システムは、音声制御システムの少なくとも一人のユーザからのスピーチ信号を受信および認識するための静止ユニットを含んでおり、この方法は、

音声制御システムの少なくとも一人のユーザによって携帯される少なくとも1つのモバイルデバイスを提供することと、

少なくとも1つのモバイルデバイスから静止ユニットに制御信号を伝送することと、

伝送された制御信号を静止ユニットによって受信することと、

受信した制御信号に 응답して、静止ユニットを動作させることと

30

を包含する。

【0011】

本発明にしたがうと、音声制御システムの使用は、ユーザによって携帯されているモバイルデバイスによって、容易化および改善される。モバイルデバイスは、制御信号(または多数の信号)を伝送することによって、静止ユニットと通信する。通信は、赤外線信号、および/または無線信号、および/または光学信号、および/または(可聴範囲内または可聴範囲外の)音響信号を伝送することによって、実行され得る。二人以上のユーザ(各々、自分自身のモバイルデバイスを携帯している)は、音声制御システムを動作させ得る。

【0012】

40

モバイルデバイスは、例えばセル電話、スマートフォン(タッチスクリーンおよびキーボードならびにプログラマブルロジックを含む)、またはパーソナルデジタルアシスタント、または遠隔制御、またはヘッドセット、または補聴器、特に両耳補聴器であり得る。したがって、ユーザによって任意の方法で携帯されるモバイルデバイスは、音声制御システムの静止ユニットとの通信を確立させるために使用され得る。非常に単純な場合において、例えば、ユーザは両耳補聴器を携帯する。無線信号は、両耳補聴器の性能を向上させるために、両耳補聴器の左ピースと右ピースとの間で伝送される。これらの信号は、音声制御システムの静止ユニットによって受信され得、その結果これは、ユーザが室内に存在しており、(例えば、スピーチ認識手段を作動させるか、または適応させることによって)しかるべく動作させていることを決定する。より複雑な実施例は、以下で与えられる。

50

【 0 0 1 3 】

音声制御システムは、ユーザの口頭のコマンドによって制御される任意のシステムであり得、例えば、H I F Iオーディオデバイス、またはテレビ、または別の電子通信デバイス、またはエンターテインメントデバイスであり得る。静止ユニットは、デバイスのハウジング内に配置され得、この動作は音声によって制御される。静止ユニットはまた、ユーザの口頭のコマンドに回答して動作する実際のデバイスから遠くに配置され得る。しかしながら、静止ユニットは、部屋（リビングルームまたはオフィス）に固定され得る。

【 0 0 1 4 】

従来技術とは異なり、スピーチ信号を受信および認識する静止ユニットは、少なくとも1つのモバイルデバイスによって伝送された制御信号に回答して動作する。その結果、静止ユニットは、例えば、受信した制御信号に回答して、ユーザの存在を認識し、スピーチ認識手段および/または静止ユニットの一部であるマイクを制御し得る。

10

【 0 0 1 5 】

受信した制御信号に回答した静止ユニットの可能な動作は、静止ユニットを作動させること、および/またはスピーチ信号を認識すること、および/または音声制御システムの少なくとも一人のユーザを位置決定すること、および/または音声制御システムの少なくとも一人のユーザを識別することを含んでいる。静止ユニットの作動は、特に、静止ユニットの一部であるスピーチ認識手段の作動を含み得る。

【 0 0 1 6 】

モバイルデバイスを携帯しているユーザが、静止ユニットが設けられている部屋に入ると、静止ユニットは、ユーザによって（例えば、定期的にもしくは要求に応じて）伝送された制御信号を受信し得、これにより、ユーザの存在を知らされ得る。その結果、スピーチ認識システムが作動され得、ユーザのごく最初の言葉が、高い信頼性で検出および認識され得る。制御信号がユーザによって、要求に応じて伝送される場合、（例えば、同じ部屋に存在する別の人物とのユーザの会話に起因する）音声制御システムの意図されていない応答は、回避され得る。

20

【 0 0 1 7 】

少なくとも1つのモバイルデバイスの各々は、その他のモバイルデバイスによって伝送された制御信号とは異なる個々の制御信号を送信し得、これにより、静止ユニットが特定のユーザを（厳密に言うと、特定のモバイルデバイスを）識別することを可能にする。したがって、音声制御のためのスピーチ認識は、識別されたユーザに依存して、静止ユニットによって実行され得る。例えば、スピーチ認識処理のために使用される語彙のリストの言葉、または個別的に訓練された（*trained*）コードブックは、識別されたユーザの言葉のスピーチ認識のために適切に使用され得る。この結果、静止ユニットのマイクのアレイによって検出される、ユーザによる言葉の認識結果の信頼性が、改善され得る。

30

【 0 0 1 8 】

さらに、ユーザが位置決定され得るので、静止ユニットの一部であるマイクのアレイは、ユーザの位置/方向にしたがって調整され得る。静止ユニットの動作が、マイクのアレイのマイクを音声制御システムのユーザの位置に向けることを含む場合、検出されたスピーチ信号の信号対ノイズ比（*SNR*）は、改善される。信号処理は、マイクのアレイによって取得された個々のマイク信号のビーム形成（上述を参照）を含み得、音声制御システムの正しい動作のために高い信頼性で認識され得る高品質のスピーチ信号を実現する。

40

【 0 0 1 9 】

モバイルデバイスから静止ユニットへの制御信号の定期的な伝送は、ユーザの位置追跡を可能にする。スピーチ認識およびマイクのアレイの調整のためのパラメータの適応は、一定期間にわたるユーザの実際の位置に基づいて、動的に実行され得る。

【 0 0 2 0 】

一実施形態にしたがうと、音声制御システムの音声制御のための本発明の方法は、モバイルデバイスから静止ユニットに追加的なオーディオ信号を伝送することと、追加的なオーディオ信号を静止ユニットによって受信することと、

50

受信した追加的なオーディオ信号に基づいて、静止ユニットが設けられている部屋の音響特性を解析することと

をさらに含んでいる。

【 0 0 2 1 】

したがって、上述の利点に加えて、モバイルデバイスはまた、静止ユニットが設けられている部屋の音響特性を解析するためにも使用され得る。この実施形態にしたがうと、モバイルデバイスは、予め選択されたオーディオ信号（これは、静止ユニットによって事前に知られている）を送信する。すなわち、追加的なオーディオ信号のパラメータ（例えば、周波数、短時間のスペクトル、出力密度等）は、静止ユニットに格納され、検出された追加的なオーディオ信号と比較され得る。この比較に基づいて、部屋の反響特性およびラウドスピーカ - 部屋 - マイクのシステム（このシステムのラウドスピーカは、追加的なオーディオ信号を伝送するモバイルデバイスのラウドスピーカであり、このシステムのマイクは、静止ユニットの（単数または複数の）マイクである）のインパルス応答が評価され得る。

10

【 0 0 2 2 】

部屋（この部屋には、静止ユニットが設けられており、音声制御システムがユーザによって動作されている）の解析された音響特性に基づいて、ユーザの言葉のスピーチ認識のために使用される処理パラメータが、適応され得る。例えば、検出されたスピーチ信号の品質を向上させるために使用されるノイズ低減およびエコー補償フィルタリング手段が、しかるべく適応され得る。

20

【 0 0 2 3 】

さらに、本明細書中に開示されている方法の上述の実施例は、ユーザからのスピーチ信号をモバイルデバイス（例えば、セル電話）によって検出する追加的なステップと、検出されたスピーチ信号をモバイルデバイスから静止ユニットに伝送する追加的なステップと、伝送されたスピーチ信号を静止ユニットによって受信する追加的なステップと、受信したスピーチ信号をスピーチ認識のために処理し、認識結果を取得する追加的なステップと、認識結果に基づいて、静止ユニットを動作させる追加的なステップとをさらに含んでいる。

30

【 0 0 2 4 】

この実施形態にしたがうと、モバイルデバイスは、上述の制御信号を伝送するために使用されるのみならず、音声制御システムの動作のためのユーザのコマンドを含むスピーチ信号を伝送するためにも使用される。ユーザの言葉の検出のために使用されるモバイルデバイスのマイクは、ユーザの口に比較的近いので、騒々しい環境においては、静止ユニットによって検出される（そして、部屋の音響伝達によって影響される）スピーチ信号よりも、モバイルデバイスから伝送されるスピーチ信号を使用することが好適であり得る。

【 0 0 2 5 】

この実施形態において、ユーザの言葉を表すスピーチ信号の認識は、モバイルデバイスから静止ユニットに伝送されるスピーチ信号のみに基づいて実行されとは限らないことに留意されたい。現在のセル電話の比較的乏しいサンプリングレートを仮定すると、モバイルデバイスから静止ユニットに伝送された、スピーチ信号の歯擦音（sibilant）ではない部分のみが、スピーチ認識処理のために使用され得、その一方、静止ユニットのマイクのアレイによって検出された、スピーチ信号の歯擦音部分（ユーザの同じ言葉を表す）は、スピーチ認識処理のために使用され得る。

40

【 0 0 2 6 】

別の実施例にしたがうと、この方法は、ユーザからのスピーチ信号をモバイルデバイスによって（例えば、セル電話によって）検出する追加的なステップと、

50

検出されたスピーチ信号をモバイルデバイスによって認識し、認識結果を取得する追加的なステップと

を含んでおり、さらに、

少なくとも1つのモバイルデバイスから静止ユニットに伝送された制御信号は、認識結果に基づいて生成される。

【0027】

この実施形態は、モバイルデバイスに設けられた認識器を使用する。最近のセル電話またはPDAの一部は、スピーチ入力（特に、キーワードまたはキーフレーズの形式）のスピーチ認識機能を示している。モバイルデバイスは、キーワードまたはキーフレーズ（例えば、「～を開始（start）」、「停止（stop）」、「もっと大きく（louder）」）を認識し得、静止ユニットに適切な制御信号を伝送し得る。そのような信号は、静止ユニットによって検出および認識されたスピーチ信号よりも上の優先度を与えられ得る。すなわち、音声制御システムが、ユーザの意図にしたがって動作しないとき、ユーザは、モバイルデバイスに設けられた認識器によって取得された認識結果に基づいて生成された制御信号を介して、音声制御システムと直接的に相互作用し得る。このようにして、柔軟性および信頼性が、さらに向上させられる。

10

【0028】

本発明はまた、音声制御システムを提供し、この音声制御システムは、

スピーチ信号を受信するように構成された静止ユニットと、

静止ユニットに制御信号を伝送するように構成された少なくとも1つのモバイルデバイスと

20

を備えており、

静止ユニットはまた、少なくとも1つのモバイルデバイスによって伝送された制御信号を受信し、受信した制御信号に応答して動作するようにも構成されている。

【0029】

一実施形態において、静止ユニットは、受信した制御信号に応答して作動するように構成されており、受信した制御信号にしたがって、スピーチ信号を認識し、かつ/またはモバイルデバイスを位置決定および/または識別するように構成された、処理ユニットをさらに含んでいる。

【0030】

30

好適には、静止ユニットは、マイク、特に少なくとも1つの指向性マイクを含む、マイクのアレイをさらに含み得、静止ユニットは、受信した制御信号に応答して、マイクのアレイを調整するように構成され得る。特に、マイクのアレイのマイクは、受信した制御信号に応答して、ユーザ/スピーカの位置に向けられ得る。

【0031】

少なくとも1つのモバイルデバイスは、特に、赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を、制御信号として伝送するように構成され得、静止ユニットは、伝送された赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を、適切な受信器によって受信するように構成され得る。

40

【0032】

本発明の音声制御システムの少なくとも1つのモバイルデバイスは、例えば、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタント、または遠隔制御、または電話ヘッドセット、または補聴器、特に両耳補聴器であり得る。

【0033】

一実施例にしたがうと、少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、これはまた、ユーザからのスピーチ信号を（モバイルデバイスに組み込まれた1つ以上のマイクを用いて）検出し、検出されたスピーチ信号を静止ユニットに伝送するようにも構成されており、静止ユニットは、モバイルデバイスによって伝送されたスピーチ信号を認識し、認識結果を取得し、

50

受信した制御信号に応答して動作するように構成されている。

【0034】

別の実施例にしたがうと、少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、これはまた、ユーザからのスピーチ信号を検出し、検出されたスピーチ信号を認識し、認識結果を取得し、認識結果に基づいて制御信号を生成するようにも構成されている。当然ながら、最後に言及された2つの実施例の特徴は、ユーザによる音声制御システムの制御の可変性をさらに向上させるために、互いに組み合わせられ得る。

【0035】

本発明の追加的な特徴および利点は、図面を参照して記載される。記載中、添付の図面に対する参照がなされるが、これらの図面は、本発明の好適な実施形態を示すように意図されている。そのような実施形態は、本発明の全範囲を表しているわけではないことに留意されたい。

【0036】

本発明は、さらに以下の手段を提供する。

【0037】

(項目1)

音声制御システムの音声制御方法であって、該音声制御システムは、該音声制御システムの少なくとも一人のユーザからのスピーチ信号を受信および認識するための静止ユニットを含んでおり、該方法は、

該音声制御システムの該少なくとも一人のユーザによって携帯される少なくとも1つのモバイルデバイスを提供することと、

該少なくとも1つのモバイルデバイスから該静止ユニットに制御信号を伝送することと、

該伝送された制御信号を該静止ユニットによって受信することと、

該受信した制御信号に応答して、該静止ユニットを動作させることと

を包含する、方法。

【0038】

(項目2)

上記静止ユニットを動作させることは、該静止ユニットを作動させること、および/またはスピーチ信号を認識すること、および/または上記音声制御システムの上記少なくとも一人のユーザを位置決定すること、および/または該音声制御システムの該少なくとも一人のユーザを識別することを含んでいる、項目1に記載の方法。

【0039】

(項目3)

上記少なくとも一人のユーザからのスピーチ信号は、該少なくとも一人のユーザの識別に基づいて認識される、項目2に記載の方法。

【0040】

(項目4)

上記静止ユニットを動作させることは、該静止ユニットの一部であるマイクのアレイのマイクを調整することを含んでいる、項目1～項目3のいずれか一項に記載の方法。

【0041】

(項目5)

上記制御信号は、定期的にもまたは上記少なくとも一人のユーザによる要求に応じて伝送される、項目1～項目4のいずれか一項に記載の方法。

【0042】

(項目6)

上記伝送および受信される制御信号は、赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号である、項目1～項目5のいずれか一項に記載の方法。

10

20

30

40

50

【 0 0 4 3 】

(項 目 7)

上記制御信号は、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタント、または遠隔制御、またはヘッドセット、または補聴器、特に両耳補聴器によって伝送される、項目 1 ~ 項目 6 のいずれか一項に記載の方法。

【 0 0 4 4 】

(項 目 8)

上記モバイルデバイスから上記静止ユニットに追加的なオーディオ信号を伝送することと、

該追加的なオーディオ信号を該静止ユニットによって受信することと、

該受信した追加的なオーディオ信号に基づいて、該静止ユニットが設けられている部屋の音響特性を解析することと

をさらに含んでいる、項目 1 ~ 項目 7 のいずれか一項に記載の方法。

10

【 0 0 4 5 】

(項 目 9)

上記少なくとも一人のユーザからのスピーチ信号を上記モバイルデバイスによって検出する追加的なステップと、

該検出されたスピーチ信号を該モバイルデバイスから上記静止ユニットに伝送する追加的なステップと、

該伝送されたスピーチ信号を該静止ユニットによって受信する追加的なステップと、

該受信したスピーチ信号をスピーチ認識のために処理し、認識結果を取得する追加的なステップと、

該認識結果に基づいて、該静止ユニットを動作させる追加的なステップと

を含んでいる、項目 1 ~ 項目 8 のいずれか一項に記載の方法。

20

【 0 0 4 6 】

(項 目 1 0)

上記少なくとも一人のユーザからのスピーチ信号を上記モバイルデバイスによって検出する追加的なステップと、

該検出されたスピーチ信号を該モバイルデバイスによって認識し、認識結果を取得する追加的なステップと

を含んでおり、

該少なくとも 1 つのモバイルデバイスから上記静止ユニットに伝送される上記制御信号は、該認識結果に基づいて生成される、項目 1 ~ 項目 8 のいずれか一項に記載の方法。

30

【 0 0 4 7 】

(項 目 1 1)

スピーチ信号を受信するように構成された静止ユニットと、

該静止ユニットに制御信号を伝送するように構成された少なくとも 1 つのモバイルデバイスと

を備えた音声制御システムであって、

該静止ユニットはまた、該少なくとも 1 つのモバイルデバイスによって伝送された該制御信号を受信し、該受信した制御信号に応答して、動作するようにも構成されている、音声制御システム。

40

【 0 0 4 8 】

(項 目 1 2)

上記静止ユニットは、上記受信した制御信号に応答して、作動するように構成されており、該受信した制御信号に応答して、スピーチ信号を認識し、かつ/または上記モバイルデバイスを位置決定および/または識別するように構成された、処理ユニットをさらに含んでいる、項目 1 1 に記載の音声制御システム。

【 0 0 4 9 】

(項 目 1 3)

50

上記静止ユニットは、マイク、特に少なくとも1つの指向性マイクを含む、マイクのアレイを含んでおり、該静止ユニットは、上記受信した制御信号にตอบสนองして、該マイクのアレイを調整するように構成されている、項目11または項目12に記載の音声制御システム。

【0050】

(項目14)

上記少なくとも1つのモバイルデバイスは、赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を、制御信号として伝送するように構成されており、上記静止ユニットは、該伝送された赤外線信号、および/または無線信号、および/または光学信号、および/または音響信号を受信するように構成されている、項目11～項目13のいずれか一項に記載の音声制御システム。

10

【0051】

(項目15)

上記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタント、または遠隔制御、またはヘッドセット、または補聴器、特に両耳補聴器である、項目11～項目14のいずれか一項に記載の音声制御システム。

【0052】

(項目16)

上記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、該少なくとも1つのモバイルデバイスはまた、上記少なくとも一人のユーザからのスピーチ信号を検出し、該検出されたスピーチ信号を上記静止ユニットに伝送するようにも構成されており、

20

該静止ユニットは、該モバイルデバイスによって伝送された該スピーチ信号を認識し、認識結果を取得し、該受信した制御信号にตอบสนองして動作するように構成されている、項目11～項目15のいずれか一項に記載の音声制御システム。

【0053】

(項目17)

上記少なくとも1つのモバイルデバイスは、セル電話、またはスマートフォン、またはパーソナルデジタルアシスタントであり、該少なくとも1つのモバイルデバイスは、該少なくとも一人のユーザからのスピーチ信号を検出し、該検出されたスピーチ信号を認識し、認識結果を取得し、該認識結果に基づいて、上記制御信号を生成するように構成されている、項目11～項目16のいずれか一項に記載の音声制御システム。

30

【0054】

(摘要)

本発明は、音声制御システムの音声制御方法に関し、この音声制御システムは、音声制御システムの少なくとも一人のユーザからのスピーチ信号を受信および認識するための静止ユニットを含んでおり、この方法は、音声制御システムの少なくとも一人のユーザによって携帯される少なくとも1つのモバイルデバイスを提供することと；少なくとも1つのモバイルデバイスから静止ユニットに制御信号を伝送することと；伝送された制御信号を静止ユニットによって受信し、受信した静止信号にตอบสนองして、静止ユニットを動作させることとを含む。本発明は、音声制御システムに関し、この音声制御システムは、スピーチ信号を受信するように構成された静止ユニットと；静止ユニットに制御信号を伝送するように構成された少なくとも1つのモバイルデバイスとを含んでおり、この静止ユニットはまた、少なくとも1つのモバイルデバイスによって伝送された制御信号を受信し、受信した制御信号にตอบสนองして、動作するようにも構成されている。

40

【発明を実施するための最良の形態】

【0055】

図1は、本明細書中に開示される音声制御システムの実施例を示しており、この音声制御システムは、部屋(例えば、リビングルームまたはオフィス)に設けられた静止ユニッ

50

ト1を含んでいる。静止ユニット1は、H I F Iオーディオシステムの一部であり得、これには、無線受信器2が備えられている。無線受信器2は、B l u e t o o t h受信器であり得る。複数のアンテナ2は、B l u e t o o t hによって媒介される位置決定(すなわち、B l u e t o o t h信号を介して人物の位置を決定すること)のために設けられている。静止ユニット1はまた、スピーチ信号を検出するためのマイクのアレイ3をも含んでいる。検出されたスピーチ信号は、静止ユニット1の一部であるスピーチ認識手段によって認識される。

【0056】

B l u e t o o t hは、ますます普及してきている、短距離(=10m)の無線通信技術の代表である。B l u e t o o t hは、費用および電力消費の両方が比較的安く、このことが、B l u e t o o t hをモバイルデバイスに対して非常に適したものにしている。B l u e t o o t h技術は、I S M帯域(2,402~2,480GHz)を使用し、この帯域は、全世界で利用可能であり、免許を必要としない。使用される高速の周波数ホッピング方法は、毎秒1600回の周波数の変化を伴い、比較的高い耐干渉性を保証する。当然ながら、無線通信は、B l u e t o o t h接続に限定されるわけではない。例えば、無線ローカルエリアネットワーク(WLAN)は、1つの代替的な無線リンクを提供する。

10

【0057】

室内には、多数の人物(静止ユニット1のユーザ)が存在する。各ユーザは、モバイルデバイス4を携帯しており、このモバイルデバイスは、静止ユニットの受信器2に無線信号を伝送するように構成されている。ユーザによって携帯されているモバイルデバイス4は、音声制御システムの一部であり、様々なもの(例えば、セル電話、PDA、または補聴器)であり得る。モバイルデバイス4は、信号を定期的に伝送し得、この信号は、受信器2によって受信される。さらに、モバイルデバイス4には、動作(特に、静止ユニット1との通信)にいくらかの柔軟性を提供するプログラマブルロジックが提供され得る。

20

【0058】

両耳補聴器の場合、無線信号は、両耳補聴器の左部分と右部分との間で、任意の方法で伝送される。静止ユニット1の受信器2は、そのような両耳補聴器を用いているユーザが実際に室内に居ることをこの静止ユニットが決定することを可能にするために、そのような信号を受信および識別するように適応されている。当然ながら、そのような両耳補聴器には、静止ユニット1に制御信号を伝送するための追加的な送信器が提供され得る。

30

【0059】

上記の実施例ならびに以下の実施例において、モバイルデバイス4と静止ユニット1との間の通信は、無線信号の伝送に限定されないということに留意されたい。本発明の別の実施形態にしたがうと、赤外線データ伝送が使用され得る。この場合、115kBit/sまでの転送速度を提供する赤外線データ協会(Infrared Data Association)の規格IrDa1.0、または4MBit/sまでの転送速度を提供するIrDa1.1が使用され得ることが好適である。高周波数(すなわち20kHz以上)の音響信号またはUV信号もまた、使用され得る。

40

【0060】

図1に示されているデバイスの特定の例示的な動作について、図2に関連して示されている。モバイルデバイス4を携帯している人物が部屋に入ると(10)、静止ユニット1の受信器2は、ユーザによって携帯されているモバイルデバイス4から定期的に伝送された(11)信号を受信し得(12)、その結果、静止ユニット1は、スピーチ認識のために作動され得る(13)。あるいはユーザが、モバイルデバイス4を(例えば、ボタンを用いて)手で動作させ得、静止ユニット1の受信器2への信号の伝送を開始させ得る。

【0061】

マイクのアレイ3の複数のマイクによって人物からのスピーチ信号を検出することによって取得されたマイク信号をビーム形成するために、マイクのアレイ3は、この人物の決定された位置/方向にしたがって、調整され得る(14)。そして、人物の位置およびマ

50

イクのアレイからこの人物までの方向が、決定され得る(15)。ビーム形成されたマイク信号の信号対ノイズ比(SNR)は、通常、ビーム形成されていないマイク信号のSNRよりも優れているので、静止ユニットのスピーチ認識手段によるスピーチ認識は、より信頼性を有したものになる。

【0062】

さらに、本実施例にしたがうと、モバイルデバイス4は、異なる信号(例えば、異なるように変調された無線信号、または異なる周波数の無線信号)を受信器2に伝送する。その結果、個々のユーザが、静止ユニット1によって識別され得る(15)。ユーザ/話者の識別は、特に、話者を登録するように適応されたスピーチ認識手段の場合に、スピーチ認識の役に立つ。ユーザからのスピーチ信号が認識され、その後、認識結果に基づいて、音声制御動作が実行され得る(17)。これらのユーザの優先順位付けは、個々のユーザが識別される場合に、容易に導入され得る。その結果、例えば、室内に居る何人かのユーザのうちの一によって与えられた口頭のコマンドは、別の話者によって先にまたは同時に与えられた別のコマンドを無効化し得る。さらに、静止ユニット1は、特定の識別されたユーザ/話者に対する優先度を用いて、マイクのアレイ3の調整を制御するように適応され得る。

10

【0063】

モバイルデバイス4はまた、(例えば、加速度センサの形式の)運動センサをも含む得る。モバイルデバイス4の運動がそのような運動センサによって検出されたときのみ、静止デバイス1の受信器2に位置決定信号が伝送され、これにより、定期的に伝送される信号の場合と比べて、電力を節約する。そのように伝送された位置決定信号(例えば、無線信号、またはユーザに聴こえない高周波数の信号)に応答して、静止ユニットの一部であるスピーチ認識手段が作動され得、かつ/またはマイクのアレイ3のマイクが、ユーザの実際の位置に向けられ得る。

20

【0064】

別の実施例において(図3参照)、図1に示されているシステムは、以下のように機能する。モバイルデバイス4は、可聴周波数範囲内のオーディオ信号を伝送するように構成される。ユーザは、モバイルデバイス4のボタンを押し(20)、このボタンの手動の動作に応答して、オーディオ信号ならびに無線信号が、モバイルデバイス4から伝送される(21)。伝送されたオーディオ信号および無線信号は、静止ユニット1によって受信される(22)。特に、オーディオ信号は、マイクのアレイ3によって検出され、無線信号は、図1に示されている受信器2によって検出される。オーディオ信号または無線信号が受信されると、静止ユニット1におけるスピーチ認識手段が作動され、マイクのアレイのマイクが、オーディオ信号を伝送したモバイルデバイス4の位置に向けられる。

30

【0065】

静止ユニット1は、同時に伝送されたオーディオ信号および無線信号の伝送時間の差を決定することによって、モバイルデバイス4を携帯しているユーザまでの距離を決定し得る(23)。静止ユニット1におけるスピーチ認識手段は、その後、決定された距離にしたがって適応され(24)、例えば、信号処理手段に対するパラメータが調整される。ビーム形成および信号フィルタリングは、オーディオ信号および無線信号が検出される、決定された距離および方向に基づいて、制御および調整され得る。

40

【0066】

モバイルデバイス4によって伝送されたオーディオ信号は、アプリアリ(a priori)に知られる。オーディオ信号は、特に、音声制御が作動したことをユーザに示すゾングルを表し得る。信号特性(例えば、周波数、短時間の出力密度等)は、静止ユニット1に格納され、部屋の音響特性の解析(25)のために、モバイルデバイス4によって伝送された、検出および処理されるオーディオ信号と比較される。部屋の反響特性が評価され得る。

【0067】

認識処理の前に、ユーザのスピーチ信号の品質を向上させるために使用されるノイズ低

50

減フィルタおよびエコー補償フィルタリング手段は、モバイルデバイス4によって伝送されたオーディオ信号の解析によって取得された、ラウドスピーカ-部屋-マイクのシステム(このシステムのラウドスピーカは、モバイルデバイスのラウドスピーカである)のインパルス応答の評価に基づいて、適応され得る(26)。評価された部屋の音響特性に基づく上記フィルタリング手段の適応後のユーザの言葉のスピーチ認識の信頼性は、大幅に改善され得る。

【0068】

図4は、ユーザがセル電話を携帯している場合について、図1に示されているシステムの動作の実施例を示している。セル電話は、ユーザからのスピーチ信号を検出する(30)。セル電話は、検出されたスピーチ信号を静止ユニットに伝送し(31)、この静止ユ
10
ニットは、この信号を受信し、スピーチ認識手段によるスピーチ認識のために、それを処理する(32)。認識結果に回答して、静止ユニット1は、なんらかの動作を実行する(33)。単純な実施例において、ユーザはキーワード「もっと大きく(Louder)」を発声し、この言葉の認識に回答して、静止ユニットは、音声制御によって動作されているHIFIオーディオシステムの音量を増大させる。

【0069】

しかしながら、別の実施例にしたがうと、セル電話から静止ユニット1に伝送されたスピーチ信号の一部のみが、静止ユニット1におけるスピーチ認識のために使用される。標準的なセル電話におけるサンプリングレートは、約8kHzなので、静止ユニットに伝送された歯擦音のスピーチ認識は、十分な信頼性がないことがあり得る。したがって、スピー
20
チ信号の歯擦音部分(これはまた、マイクのアレイ3によっても検出される)は、静止ユニット1におけるスピーチ認識手段によって認識されることが好適であり得る。他方、セル電話によって検出および伝送されたスピーチ信号は、話者の位置決定および識別のために使用され得る。上述のように、マイクのアレイ3ならびにスピーチ認識のために使用される処理パラメータは、話者の位置および識別にしたがって、適応され得る。

【0070】

上述の動作の代わりに、セル電話そのものが、検出されたスピーチ信号(例えば、キーワード)を認識し得る(34)。この実施例において、セル電話は、認識されたスピーチ信号に回答して、静止ユニット1にBluetooth信号を伝送する(35)。静止ユ
30
ニット1は、Bluetooth信号を受信し(36)、受信したBluetooth信号に回答して動作する。例えば、ユーザはキーセンテンス「音声制御をオンに(Voice control on)」を発声し得、このキーセンテンスは、セル電話に設けられた認識器によって認識される。対応するBluetooth信号が静止ユニット1に伝送され、これが、静止ユニット1の一部であるスピーチ認識手段を作動させる。

【0071】

上述の実施形態は、本発明を限定するものとしては意図されておらず、本発明の特徴および利点を示す実施例の役目を担っている。上述の特徴の一部または全てはまた、様々な方法で組み合わされ得るといことが理解されるべきである。

【図面の簡単な説明】

【0072】

【図1】図1は、本発明にしたがう音声制御システムの実施例を示しており、この音声制御システムは、静止ユニットと、音声制御システムのユーザによって携帯されるモバイルデバイスとを含んでいる。

【図2】図2は、本発明にしたがう音声制御方法の実施例のステップを示しており、この方法は、モバイルデバイスから音声制御システムの静止ユニットに信号を伝送するステップを含んでいる。

【図3】図3は、本発明にしたがう音声制御方法の実施例のステップを示しており、この方法は、モバイルデバイスから音声制御システムの静止ユニットにオーディオ信号を伝送し、部屋の音響特性を評価するステップを含んでいる。

【図4】図4は、本発明にしたがう音声制御方法の実施例のステップを示しており、この
50

方法は、セル電話から静止ユニットにオーディオ信号（特に、スピーチ信号）を伝送するステップを含んでいる。

【符号の説明】

【0073】

- 1 静止ユニット
- 2 受信器
- 3 マイクのアレイ
- 4 モバイルデバイス

【図1】

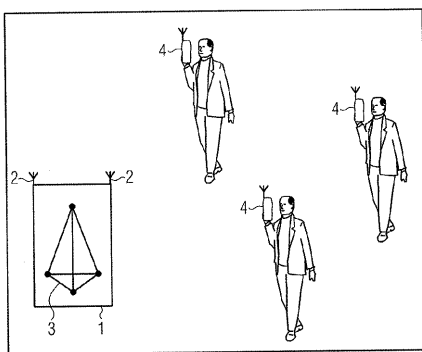


FIG. 1

【図2】

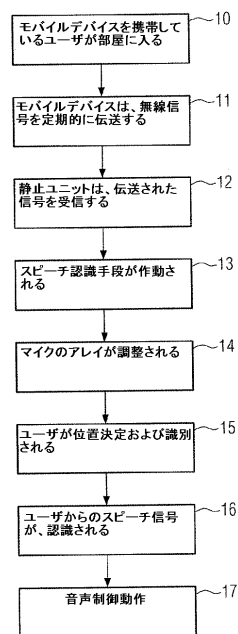


FIG. 2

【 図 3 】

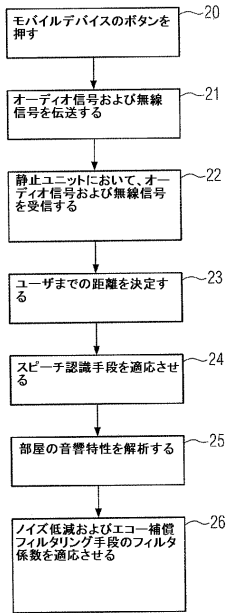


FIG. 3

【 図 4 】

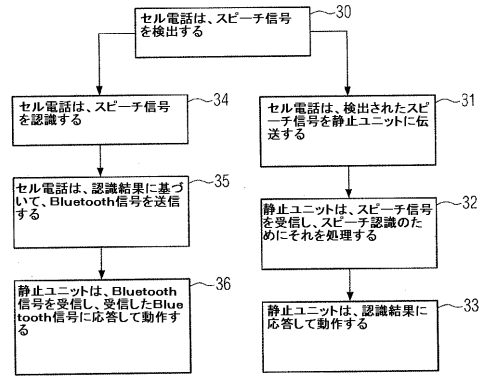


FIG. 4

フロントページの続き

(72)発明者 ラルス ケーニツヒ

ドイツ国 89075 ウルム, メリアンヴェーク 55

(72)発明者 ゲルハルト ウヴェ シュミット

ドイツ国 89081 ウルム, ハルサウゼル シュトラーセ 37

審査官 井上 健一

(56)参考文献 国際公開第03/049003(WO, A1)

国際公開第2006/121896(WO, A1)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/34