

(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2007-264132
(P2007-264132A)

(43) 公開日 平成19年10月11日(2007.10.11)

(51) Int. Cl.		F I		テーマコード (参考)
G 1 0 L 15/04 (2006.01)		G 1 0 L 15/04	3 0 0 A	5 D 0 1 5
G 1 0 L 15/28 (2006.01)		G 1 0 L 15/04	3 0 0 D	
G 1 0 L 11/02 (2006.01)		G 1 0 L 15/28	4 0 0	
		G 1 0 L 11/02		

審査請求 未請求 請求項の数 6 O L (全 9 頁)

(21) 出願番号	特願2006-86607 (P2006-86607)	(71) 出願人	000003078 株式会社東芝 東京都港区芝浦一丁目1番1号
(22) 出願日	平成18年3月27日 (2006.3.27)	(74) 代理人	100059225 弁理士 葛田 璋子
		(74) 代理人	100076314 弁理士 葛田 正人
		(74) 代理人	100112612 弁理士 中村 哲士
		(74) 代理人	100112623 弁理士 富田 克幸
		(72) 発明者	金澤 博史 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

最終頁に続く

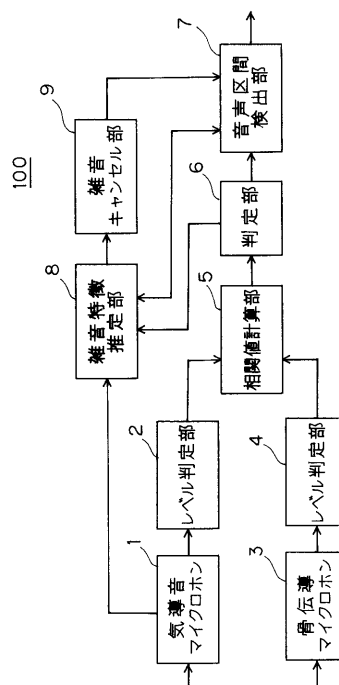
(54) 【発明の名称】 音声検出装置及びその方法

(57) 【要約】

【課題】 確実に発話の内容を検出できる音声検出装置を提供する。

【解決手段】 気導音マイクロホン1と骨伝導マイクロホン3の出力が共に単位時間毎に予め設定したレベルを超えた場合に、相関値を計算し、相関が高いと判定された場合に、その時刻情報を音声区間検出部7に送ることで、確実に音声区間と想定される区間を音声検出の際に利用できる。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

それぞれ特性の異なる複数のマイクロホンと、
前記各マイクロホンからそれぞれ入力された入力信号の間の相関値を求める相関値計算手段と、

前記相関値により前記各入力信号の間に相関があるか否かを判定し、相関があると判定された時の時刻情報を出力する判定手段と、

前記時刻情報に基づいて相関があるとされた区間における前記各入力信号の中の一つの入力信号を音声信号として出力する音声区間検出手段と、

を有する

10

ことを特徴とする音声検出装置。

【請求項 2】

前記複数のマイクロホンのうち、少なくとも 2 つは気導音マイクロホンと骨伝導マイクロホンである

ことを特徴とする請求項 1 記載の音声検出装置。

【請求項 3】

前記音声区間手段において音声信号として出力される入力信号は、前記気導音マイクロホンの入力信号である

ことを特徴とする請求項 2 記載の音声検出装置。

【請求項 4】

20

前記気導音マイクロホンの入力信号の雑音を除去し、この除去した入力信号を前記音声区間検出手段に出力する雑音キャンセル手段を有する

ことを特徴とする請求項 3 記載の音声検出装置。

【請求項 5】

前記判定手段において相関がないと判定された区間における前記気導音マイクロホンの入力信号から雑音特徴信号を推定する雑音特徴推定手段を有し、

前記雑音キャンセル手段は、前記雑音特徴信号に基づいて前記気導音マイクロホンの入力信号の雑音を除去する

ことを特徴とする請求項 4 記載の音声検出装置。

【請求項 6】

30

それぞれ特性の異なる複数のマイクロホンからそれぞれ入力された入力信号の間の相関値を求め、

前記相関値により前記各入力信号の間に相関があるか否かを判定し、相関があると判定された時の時刻情報を出力し、

前記時刻情報に基づいて相関があるとされた区間における前記各入力信号の中の一つの入力信号を音声信号として出力する

ことを特徴とする音声検出方法。

【発明の詳細な説明】**【技術分野】****【0001】**

40

本発明は、複数のマイクロホンを利用した音声検出装置及びその方法に関する。

【背景技術】**【0002】**

骨伝導マイクロホンと通常の気導音マイクロホンを併用して、音声認識のための区間検出精度を向上させる方法として、特許文献 1 の技術が提案されている。この方法は各マイクロホン出力に対し、別個の音声区間検出処理を行い、ノイズレベルに応じてどちらの検出結果を優先するかを決め、その優先度に基づき、どちらか一方の音声区間検出結果を最終結果とするという方法である。

【0003】

また、特許文献 2 には、単一指向性マイクロホンで受けた音信号と無指向性マイクロホ

50

ンで受けた音信号との相関度を得て、その相関結果に基づいて検出対象音を検出する技術も提案されている。

【特許文献1】特開平4 - 276799号公報

【特許文献2】特開2005 - 227511公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

特許文献1の技術では、背景雑音レベルが低い場合、気導音マイクロホンの音声区間検出結果が優先されることになる。しかし、背景雑音レベルが低い場合でも、例えば隣の話し声など非定常雑音がターゲット音声の前後に付加した場合には、その付加した区間も含めた区間が音声区間として検出されるといった誤った音声区間検出がなされる場合がある。その場合、誤りの修正ができず、誤った音声区間の音声信号が音声認識装置に入力され、ひいては認識誤りを生じさせるといった問題点がある。

10

【0005】

また、骨伝導マイクロホン及び気導音マイクロホンから得られる音声信号に対する両方の音声区間検出処理が、全て終了して、初めてどの時刻の音声区間を音声認識装置に供するかが確定する。そのため、音声認識の実時間処理が出来ないという問題点もある。

【0006】

さらに、特許文献2の技術では、無指向性マイクロホンによって発話を受音するため、例えば隣の話し声なども検出対象音として検出され、音声認識処理に入力されることになり、正確な認識を行えないという問題点がある。

20

【0007】

そこで、本発明は、確実に検出すべき対象音のみを検出できる音声検出装置及びその方法を提供する。

【課題を解決するための手段】

【0008】

本発明は、それぞれ特性の異なる複数のマイクロホンと、前記各マイクロホンからそれぞれ入力された入力信号の間の相関値を求める相関値計算手段と、前記相関値により前記各入力信号の間に相関があるか否かを判定し、相関があると判定された時の時刻情報を入力する判定手段と、前記時刻情報に基づいて相関があるとされた区間における前記各入力信号の中の一つの入力信号を音声信号として出力する音声区間検出手段と、を有することを特徴とする音声検出装置である。

30

【発明の効果】

【0009】

本発明であると、確実に発話の内容を検出できる。

【発明を実施するための最良の形態】

【0010】

本発明の一実施形態である音声検出装置100について図1から図5に基づいて説明する。

【0011】

40

(1) 音声検出装置100の構成

図1は、本実施形態の音声検出装置100の構成を示すブロック図である。

【0012】

音声検出装置100は、気導音マイクロホン1と骨伝導マイクロホン2の2本のマイクロホンを利用する場合について示す。そして、音声を認識したい者(以下、マイクロホン装着者という)が、骨伝導マイクロホン2を身体、特に顔付近に装着し、かつ、気導音マイクロホン1によってマイクロホン装着者の声を受音できる状態になっているとする。

【0013】

音声検出装置100は、気導音マイクロホン1、レベル判定部2、骨伝導マイクロホン3、レベル判定部4、相関値計算部5、判定部6、音声区間検出部7、雑音特徴推定部8

50

、雑音キャンセル部 9 とを備えている。なお、本実施例では、音声検出装置 100 に音声認識装置が接続されている状況を想定し、音声検出装置 100 の出力が音声認識処理に供されることを踏まえて説明する。

【0014】

(2) 気導音マイクロホン 1 とレベル判定部 2

気導音マイクロホン 1 には空気中を伝わる音響信号が入力され、この音響信号が電気信号に変換され、さらに A/D 変換後、デジタル信号（以下、入力音響信号という）に変換され、レベル判定部 2 に送られる。

【0015】

レベル判定部 2 では、例えば、10m 秒毎に、入力音響信号の波形の振幅の 2 乗和によりパワーが計算される。入力音響信号のパワーと、予め設定されたパワーの閾値と比較され、閾値を超えている場合には、10m 秒分の入力音響信号は相関値計算部 5 へ入力される。

10

【0016】

(3) 骨伝導マイクロホン 3 とレベル判定部 4

一方、時間的に同期して、骨伝導マイクロホン 3 へも身体を伝わる振動音が入力され、この振動音が電気信号に変換され、さらに A/D 変換後、デジタル信号（以下、入力振動信号という）に変換され、レベル判定部 4 に送られる。

【0017】

レベル判定部 4 では、例えば、入力振動信号の 10m 秒の長さを 1 フレームと定義して、フレーム毎に入力振動信号の波形の振幅の 2 乗和によりパワーが計算される。入力振動信号のパワーと、予め設定されたパワーの閾値と比較され、閾値を超えている場合には、10m 秒分の入力振動信号は相関値計算部 5 へ入力される。

20

【0018】

なお、各レベル判定部 2、4 での閾値は入力音響信号と入力振動信号のレベルが異なることから別々の値に設定される。

【0019】

(4) 相関値計算部 5 と判定部 6

相関値計算部 5 では、レベル判定処理を通過した気導音マイクロホン 1 からの入力音響信号及び骨伝導マイクロホン 3 からの入力振動信号が入力される。双方の信号が入力された場合に相関値が計算される。

30

【0020】

相関には、例えば各信号を周波数分解し得られた周波数パターン間のユークリッド距離を利用しても良いし、相関係数を求めてもよい。

【0021】

ここで、実際にマイクロホン装着者が発声した音声の場合、レベル判定部 2 から得られる入力音響信号は空気中を伝わる当該音声となり、レベル判定部 4 から得られる入力振動信号は発声された音声により骨が振動した信号となる。双方とも信号源は同じであるので、それぞれの相関値は高い値を示すと考えられる。

【0022】

図 2 に示すように、マイクロホン装着者でない人の発声になされた場合を考える。レベル判定部 2 から得られる入力音響信号は空気中を伝わる当該音声となり、一方、骨伝導マイクロホンへの入力はないため、レベル判定部 4 から出力されない。

40

【0023】

図 3 に示すように、マイクロホン装着者が発声せず、身体の摩擦などにより骨伝導マイクロホン 3 3 にのみ入力振動信号が入力された場合を考える。気導音マイクロホン 1 への入力はないか、あるいはあっても非常にレベルの低い信号となり、レベル判定部 2 からの出力はない。

【0024】

図 4 に示すように、マイクロホン装着者ではない人の発声と、マイクロホン装着者の身

50

体から発生された振動の両方が時間的に重なってしまった場合を考える。レベル判定部 2, 4 からはそれぞれ人の声の入力音響信号と、入力振動信号が出力される。それらの信号源は異なるので相関値は非常に小さな値になると考えられる。

【0025】

相関値計算部 5 で計算された相関値が、判定部 6 で予め設定された閾値を超えた場合に両者のマイクロホン 1, 3 に入力された信号には相関があると判定され、その時の時刻情報が音声区間検出部 7 及び雑音特徴推定部 8 に出力される。

【0026】

以上の処理は、背景雑音のない場合に考えられる現象について説明したものである。いずれの場合にも、マイクロホン装着者の発声以外の信号が音声区間検出部 7 へ入力されるのを妨げることができ、確実にマイクロホン装着者が発声したと考えられる区間の時刻情報を音声区間検出部 7 へ知らせることができる。逆にいうと、当該時刻の時点ではマイクロホン装着者が発声していると考えてよいということになる。

10

【0027】

但し、マイクロホン装着者により発声された音声のうち、特に語頭の子音などは骨等がほとんど振動せず、骨伝導マイクロホン 3 では拾うことが難しいため、実際には相関値が高いという結果が得られた時刻より以前の時刻で音声を開始されている場合がある。また、これらの現象は語頭だけでなく、語中や語尾などでも発生する場合がある。これらの問題は、音声区間検出部 7 及び本音声検出装置 100 に接続される音声認識装置で対処することになる。

20

【0028】

一方、背景雑音がある場合を考えると、マイクロホン装着者の発声がない場合でも、気導音マイクロホン 1 を通過した入力音響信号は雑音の影響で高いレベルの信号となっていると想定されるので、レベル判定部 2 を通過して相関値計算部 5 へ送付される。しかし、骨伝導マイクロホン 3 は背景雑音の影響は受けないので、マイクロホン装着者の発声がない背景雑音のみの区間ではレベル判定部 4 からの出力はない。したがって、背景雑音の影響は排除することができる。このように、背景雑音がある場合でも、ない場合でも、雑音に影響されることなく、レベル判定処理及び相関値による判定処理を行うことにより、マイクロホン装着者が確実に発声した時刻の情報を音声区間検出部 7 に出力することができる。

30

【0029】**(5) 音声区間検出部 7**

音声区間検出部 7 での処理は特に限定するものではないが、図 5 に基づいてその一例を示す。

【0030】

音声区間検出部 7 には、上記した音声区間と想定される時刻情報の他に、気導音マイクロホン 1 から出力される入力音響信号が雑音特徴推定部 8 を経て入力される。

【0031】

また、雑音特徴推定部 8 を経由して雑音キャンセル部 1 で雑音キャンセル処理された後の信号も入力される。

40

【0032】

音声区間検出部 7 では、これら 3 種類の情報及び信号を用いて、マイクロホン装着者の正しい音声区間を検出する。

【0033】

上記したように、判定部 6 から出力される音声区間の時刻情報は、語頭の子音などの区間を含まない可能性があることから、当該時刻よりも例えば 30 フレーム (= 300 ms) 前の時点から音声区間検出処理を開始するなどの制御を行う。したがって、音声区間検出部 7 内に、入力音響信号を例えば 30 フレーム分格納する記憶領域を内蔵している。音声区間検出部 7 では、開始フレーム毎に雑音キャンセル前の入力音響信号と雑音キャンセル後の入力音響信号の両方を用い、例えばフレーム毎のパワー、周波数パターンに基づき

50

、当該フレームがマイクロホン装着者の発声した音声に該当するかどうかを判定する。

【0034】

音声に該当すると判定されたフレームは、音声開始フレームから逐次、音声終了フレームまで音声認識装置に送られ、音声認識のための特徴抽出及び音声認識処理が実行される。

【0035】

(6) 雑音特徴推定部 8

雑音特徴推定部 8 では、判定部 6 で当該フレームが音声区間と判定されるまで、雑音区間であるという判断のもとに、当該フレームの信号を雑音特徴推定に利用する。

【0036】

具体的には、連続する雑音フレームが複数連続する続く場合、当該フレームより前の雑音特徴信号 $N(n-1)$ を、当該フレームの周波数特徴信号 $F(n)$ を用いてある更新係数 α で更新することにより、当該フレームまでの雑音特徴信号 $N(n)$ とする。次のフレームが雑音フレームと判定された場合には、同様に更新する。つまり更新式は以下のようになる。

【0037】

$$N(n) = (1 - \alpha) \cdot N(n-1) + \alpha \cdot F(n)$$

この雑音特徴信号は雑音キャンセル部 9 で利用される。

【0038】

(7) 雑音キャンセル部 9

雑音キャンセル部 9 では、雑音特徴推定部 8 から送られる推定された雑音特徴信号と、当該フレームが雑音なのか音声なのかの判定情報に基づき、音声フレームである場合には、気導音マイクロホン 1 から出力される入力信号から音声成分だけを抽出する処理を行う。

【0039】

例えば、スペクトルサブトラクション法などを用い、当該フレームの周波数特徴信号から、雑音特徴信号を引き算するなどの処理を行い、雑音成分をキャンセルし、音声成分だけを抽出する処理を行う。

【0040】

この音声成分信号は音声区間検出で利用されると共に、音声区間検出部 7 を経由して、音声認識装置にも送られ、音声認識処理に利用される。

【0041】

(8) 効果

本実施形態の音声検出装置 100 であると、周囲の背景雑音などの定常雑音や、人の話し声、電話音やクラクションなどの非定常雑音の影響を受けず、正確な音声区間検出が可能となり、ひいては音声認識性能を高めることが可能となる。

【0042】

(9) 変更例

本発明は上記各実施形態に限らず、その主旨を逸脱しない限り種々に変更することができる。

【0043】

例えば、上記実施形態では、気導音マイクロホン 1 と骨伝導マイクロホン 2 の 2 本のマイクロホンを利用する場合について説明したが、これに代えて、無指向性の気導音マイクロホンと指向性の気導音マイクロホンと骨伝導マイクロホンの 3 本のマイクロホンを使用してもよい。この場合に音声か雑音かの判定は、複数の入力信号間の組み合わせの内、少なくとも過半数の組み合わせで相関があったときに音声であると判定する。また、過半数の組み合わせで相関がないときに雑音であると判定する。

【図面の簡単な説明】

10

20

30

40

50

【0044】

【図1】本発明の一実施形態の音声検出装置のブロック図である。

【図2】気導音マイクロホンにマイクロホン装着者の発声以外の音声が入力された場合の例を示す図である。

【図3】骨伝導マイクロホンにマイクロホン装着者の発声以外の振動が入力された場合の例を示す図である。

【図4】気導音マイクロホンと骨伝導マイクロホンの両方にマイクロホン装着者の発声以外の音及び振動が同時に入力された場合の例を示す図である。

【図5】音声区間検出処理の一例を示す図である。

【符号の説明】

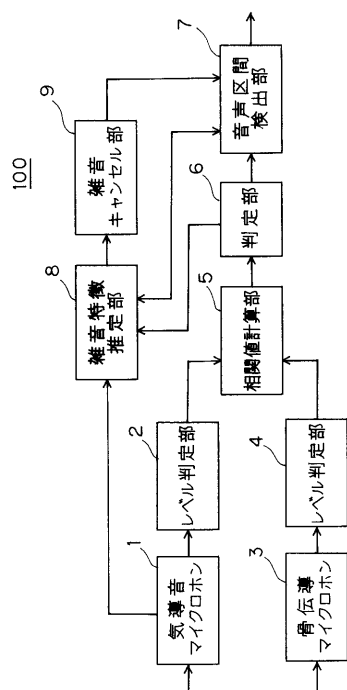
10

【0045】

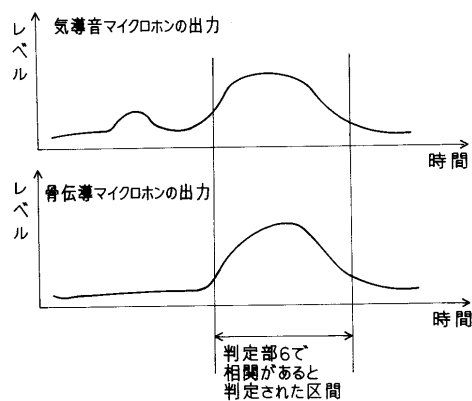
- 1 気導音マイクロホン
- 2 レベル判定部
- 3 骨伝導マイクロホン
- 4 レベル判定部
- 5 相関値計算部
- 6 判定部
- 7 音声区間検出部
- 8 雑音特徴推定部
- 9 雑音キャンセル部

20

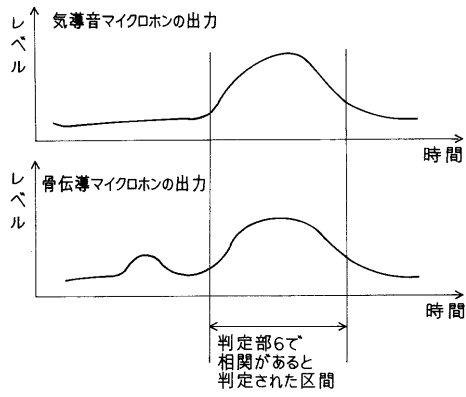
【図1】



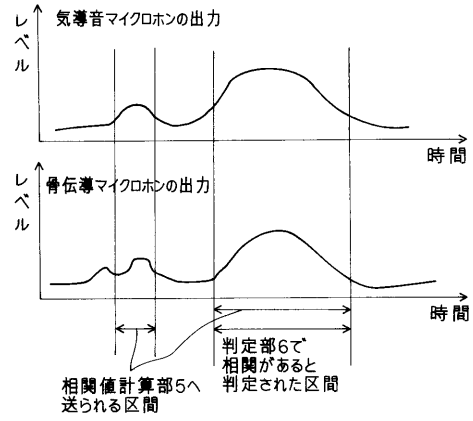
【図2】



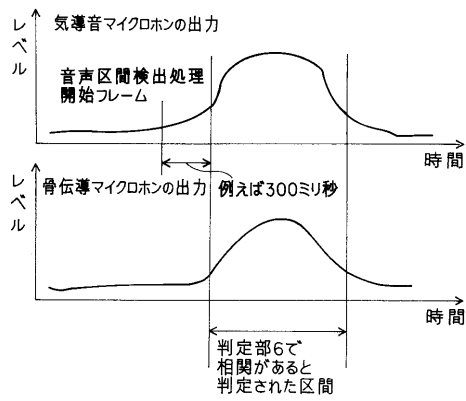
【 図 3 】



【 図 4 】



【 図 5 】



フロントページの続き

(72)発明者 平川 秀樹

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

Fターム(参考) 5D015 DD02 DD03