

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3869888号

(P3869888)

(45) 発行日 平成19年1月17日(2007.1.17)

(24) 登録日 平成18年10月20日(2006.10.20)

(51) Int. Cl.	F I
G 1 0 L 15/06 (2006.01)	G 1 0 L 15/06 5 0 0 L
G 1 0 L 15/20 (2006.01)	G 1 0 L 15/20 3 7 0 E
G 1 0 L 15/00 (2006.01)	G 1 0 L 15/00 2 0 0 A

請求項の数 13 (全 14 頁)

(21) 出願番号	特願平8-147357	(73) 特許権者	590000248
(22) 出願日	平成8年6月10日(1996.6.10)		コーニンクレッカ フィリップス エレク
(65) 公開番号	特開平9-6388		トロニクス エヌ ヴィ
(43) 公開日	平成9年1月10日(1997.1.10)		Koninklijke Philips
審査請求日	平成15年6月9日(2003.6.9)		Electronics N. V.
(31) 優先権主張番号	19521258.4		オランダ国 5621 ペーアー アイ
(32) 優先日	平成7年6月10日(1995.6.10)		ドーフエン フルーネヴァウツウェッハ
(33) 優先権主張国	ドイツ(DE)		1
			Groenewoudseweg 1, 5
			621 BA Eindhoven, T
			he Netherlands
		(74) 代理人	100061815
			弁理士 矢野 敏雄
		(74) 代理人	100094798
			弁理士 山崎 利臣

最終頁に続く

(54) 【発明の名称】 音声認識装置

(57) 【特許請求の範囲】

【請求項1】

ノイズ信号(N)の重畳された音声信号(S)を受信するための音声認識装置において、

第1のスペクトル値(xd)を送出するための第1のスペクトル分析手段(6)を有し、

前記第1のスペクトル値は、妨害された音声信号(X)から導出され、第1のスペクトルベクトル(x)に統合されるものであり、

ノイズ信号(N)の推定値(N^h)を送出するための手段(3、4)を有し、

前記ノイズ信号推定値(N^h)から第2のスペクトル値を形成するための第2のスペクトル分析手段(7)を有し、

前記第2のスペクトル値は、第2のスペクトルベクトル(n)に統合されており、

基準スペクトル値(yd)からなる基準ベクトル(y)を記憶するためのメモリユニット(8)を有し、

前記基準ベクトル(y)にはインデックス(j)が配属されており、該インデックスは基準音声信号のスペクトルに相応するものであり、

インデックス(j)に配属された比較ベクトル(y')を、基準ベクトル(y)および第2のスペクトルベクトル(n)から形成するための手段(9)を有し、

比較ベクトル(y')の少なくとも1つを第2のスペクトルベクトル(n)に連続的に適合され、

さらに、比較ベクトル (y') を第 1 のスペクトルベクトル (x) と比較することにより、インデックス (j) を選択するための手段 (10、11) が設けられおり、
ノイズ信号 (N) の推定値 (N^{\wedge}) を形成するため、既知のオーディオ信号のエコーをシミュレートする、ことを特徴とする音声認識装置。

【請求項 2】

ノイズ信号 (N) の推定値 (N^{\wedge}) を形成するため、ノイズ信号 (N) を所属のノイズ信号源の近傍で記録する、請求項 1 記載の音声認識装置。

【請求項 3】

ノイズ信号源の信号 (N) をフィルタリングして、ノイズ信号 (N) の推定値 (N^{\wedge}) を形成するための適応形フィルタ (4) が設けられており、

さらに、形成された推定値 (N^{\wedge}) を妨害を受けた音声信号 (X) から減算するための手段 (5) が設けられている、請求項 1 または 2 項記載の音声認識装置。

【請求項 4】

音声信号休止 ($j = 0$) に相応する比較ベクトル (y') を、それぞれ存在する第 2 のスペクトルベクトル (n) により連続的に置換する、請求項 1 から 3 までのいずれか 1 項記載の音声認識装置。

【請求項 5】

すべての比較ベクトル (y') をそれぞれ存在する第 2 のスペクトルベクトル (n) に適応する、請求項 1 から 3 までのいずれか 1 項記載の音声認識装置。

【請求項 6】

それぞれのインデックス (j) に基準ベクトル (y) および 1 つまたは複数の順次連続する第 2 のスペクトルベクトル ($n(i)$, $n(i - 1)$, ...) を、当該インデックス (j) に対して相応する数の比較ベクトル (y') を配属し、

第 1 のスペクトルベクトル (x) との比較の際に、個々の比較ベクトル (y') の重み付けを、重み付け係数 (g) をそれぞれのベクトル距離に加算することによって行う、請求項 1 から 3 までのいずれか 1 項記載の音声認識装置。

【請求項 7】

音響経路で通信機器まで伝送される使用者の音声信号を識別し、通信機器の部分機能を識別された音声信号に従って制御する、請求項 1 から 6 までのいずれか 1 項記載の音声認識装置。

【請求項 8】

遠距離通話網を介して通信機器に伝送される音声信号を識別し、通信機器の部分機能を制御する、請求項 1 から 7 までのいずれか 1 項記載の音声認識装置。

【請求項 9】

通信機器として携帯電話を使用する、請求項 7 または 8 記載の音声認識装置。

【請求項 10】

通信機器として呼びかけ応答機を使用する、請求項 7 または 8 記載の音声認識装置。

【請求項 11】

自動車の無線受信機の部分機能を制御する、請求項 1 から 6 までのいずれか 1 項記載の音声認識装置。

【請求項 12】

自動車の部分機能の制御に用いる、請求項 1 から 6 までのいずれか 1 項記載の音声認識装置。

【請求項 13】

当該電気機器または他の電気機器の部分機能を、音声認識装置の識別された音声信号に依存して制御する、請求項 1 から 12 までのいずれか 1 項記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ノイズ信号の重畳された音声信号を受信するための音声認識装置に関する。

10

20

30

40

50

【 0 0 0 2 】

【従来の技術】

発声された音声の自動認識は多数の適用事例に利用できる。その際、受信した音声信号に重畳された妨害ノイズによって、音声認識に悪影響の及ぼされることが多い。このことで音声認識におけるエラーレートが高まってしまう。妨害ノイズはたとえば、音声認識装置を有する機器の周囲に存在する妨害源によって発生される。

【 0 0 0 3 】

また、電話回線と接続されている通信ターミナルに音声認識装置を用いた場合には、ラインエコーによってもこの種のノイズ信号が形成される可能性がある。ドイツ連邦共和国特許出願第 4 4 2 7 1 2 4 . 7 号明細書によれば、音声アナウンスを発生させかつ音声信号を認識する手段を有する通信装置が公知である。音声アナウンスおよび音声認識手段によって、遠隔地の加入者との通信が行われる。この場合、遠隔地の加入者への伝送に際して音声アナウンスが反射することにより不所望なエコーが発生し、これによって遠隔地の加入者からの音声信号の認識に悪影響が及ぼされる。音声認識においてエラーレートを低減するためのエコー補償はスペクトル領域で行われ、その際、通信機器における受信信号の短期間のスペクトル値と推定されたエコーの短期間のスペクトル値との差が形成される。エコーの短期間のスペクトル値を推定するために、通信機器の送信信号（音声アナウンス）の短期間のスペクトル値が求められて、それに対し適応形の重み付けが加えられる。そしてエコー作用の抑圧された差分スペクトル値が音声認識手段により評価される。

【 0 0 0 4 】

【発明が解決しようとする課題】

本発明の課題は、エラーが最小限に抑えられた音声認識を確実に行わせる音声認識装置を提供することにある。

【 0 0 0 5 】

【課題を解決するための手段】

上記課題は、第 1 のスペクトル値を送出するための第 1 のスペクトル分析手段を有し、前記第 1 のスペクトル値は、妨害された音声信号から導出され、第 1 のスペクトルベクトルに統合されるものであり、

ノイズ信号の推定値を送出するための手段を有し、

前記ノイズ信号推定値から第 2 のスペクトル値を形成するための第 2 のスペクトル分析手段を有し、

前記第 2 のスペクトル値は、第 2 のスペクトルベクトルに統合されており、

基準スペクトル値からなる基準ベクトルを記憶するためのメモリユニットを有し、

前記基準ベクトルにはインデックスが配属されており、該インデックスは基準音声信号のスペクトルに相応するものであり、

インデックスに配属された比較ベクトルを、基準ベクトルおよび第 2 のスペクトルベクトルから形成するための手段を有し、

比較ベクトルの少なくとも 1 つを第 2 のスペクトルベクトルに連続的に適合され、

さらに、比較ベクトルを第 1 のスペクトルベクトルと比較することにより、インデックスを選択するための手段が設けられており、

ノイズ信号 (N) の推定値 (N[^]) を形成するため、既知のオーディオ信号のエコーをシミュレートするように構成して解決される。

【 0 0 0 6 】

【発明の実施の形態】

ノイズ信号の推定値には、目下のノイズ信号に関する付加的な情報が含まれている。これはたとえば、ノイズ信号を発するノイズ源の近くに配置された付加的なマイクロフォンを用いることで捕捉され、このことで音声信号から十分に分離されたノイズ信号を用いることができ、推定することができる。ノイズ信号が既知の音声信号（たとえば留守番電話機に記憶されている音声アナウンスまたは車両内の音楽再生）のエコーであるならば、エコーを発生させたもとの音声信号から、伝達関数の推定によりノイズ信号を求めること

ができる。基準ベクトルと第2のスペクトルベクトル(ノイズベクトル)から導出された比較ベクトルの連続的な整合により、エラーが最小限に抑えられた音声認識が確実に行われるようになり、その際、順次連続して選択される指標は音声信号休止を伴う音声信号波形に対応し、したがって個々の音声認識の結果を表す。この音声認識装置は信号プロセッサを用いて実現すると有利であり、それによって音声認識装置のための信号処理がプログラミング制御されて実行される。

【0007】

比較はたとえば次のようにして行われる。すなわち、第1のスペクトルベクトルと比較ベクトルとの間の距離が形成され、これは第1のスペクトルベクトルと比較ベクトルとの類似性を表す尺度を成すものであって、その際、距離が小さければ類似性が高いことになる。この場合、複数の比較ベクトルが1つの指標に対応づけられるものとすれば、比較ベクトルのための所定の重み付け係数を導入することで音声認識結果が改善される。指標の選択に際して、結果として距離が最小化されるような指標が選択され、この場合、たとえば総和距離の最小化が行われるようにして選択される。基準ベクトルは、音声認識装置の稼働開始前に求められメモリユニットに格納された基準音声信号のスペクトルに対応する。本発明による特徴を備えた音声認識装置は音声認識に関して、音声信号に重畳されるノイズ信号に対しロバストな特性を有している。その際、システムの周囲からの妨害ノイズも効果的に補償されるし、たとえば電話回線中の反射により生じるような持続的ないしは大きな遅延を伴って到来しノイズ信号として作用するエコー信号も、効果的に補償される。

10

20

【0008】

ノイズ信号の推定値はたとえば、付随するノイズ信号源の近くでノイズ信号を別個に捕捉することにより導出できる。この目的で、ノイズ信号源の近くにマイクロフォンが配置される。音声信号または別の音響信号のエコーであれば、ノイズ信号として到来するエコーの構造は既知である。このことをノイズ信号の推定値の形成に利用する。また、これらの構成の組み合わせも可能である。

【0009】

本発明の1つの実施形態によれば、それぞれ1つの指標に対し、その指標のための比較ベクトルの相応の個数を形成するために、少なくとも1つの基準ベクトルと1つまたは複数の順次連続する第2のスペクトルベクトルが割り当てられ、第1のスペクトルベクトルとの比較に際して、個々の比較ベクトルの評価が重み付け係数と個々のベクトル距離との加算により行われ、このような実施形態によってきわめて信頼性の高い音声認識結果が得られる。

30

【0010】

有利にはこの音声認識装置は、音響経路を介して伝送される音声認識装置を含む機器の近くのユーザによる音声信号を認識するために用いられ、その際、この音声認識装置は機器の部分的な機能を制御するために用いられる。たとえば車両内の無線受信機の機能または車両の他の機器における機能を制御するために、この音声認識装置を用いることができる。また、通信機器の部分的な機能を制御するために用いられ電話回線を介して伝送される音声信号を認識するために、この音声認識装置を使用することができ、たとえばライン

40

【0011】

さらにこの音声認識装置を種々の複数の機器の制御にも使用できる。また、口述記録機における適用も有利である。

【0012】

次に、図面を参照して本発明の実施例について詳細に説明する。

【0013】

【実施例】

50

図1に示されている本発明による音声認識装置1は、妨害を受けた音声信号 $X(t)$ を受信する受信分岐2を有しており、この音声信号 $X(t)$ は、妨害を受けていない音声信号 $S(t)$ とノイズ信号 $N(t)$ とから加算的に合成されたものである。第2の受信分岐3を介して音声認識装置1へノイズ信号源の信号を表す信号 $N(t)$ が供給され、この信号 $N(t)$ によってノイズ信号 $N(t)$ が引き起こされる。ノイズ信号 $N(t)$ と信号 $N(t)$ との関係は1つの伝達関数により作り出されるものであり、この伝達関数によってノイズ信号源から受信分岐2の入力側に配置された受信手段への伝送経路が記述される。その際、この受信手段は妨害を受けた音声信号 $X(t)$ を供給する。

【0014】

音声認識装置1が留守番電話機内に統合されていて、ノイズ信号 $N(t)$ が、留守番電話機における遠隔問い合わせにあたり電話回線中の音声アナウンスの反射により生じるラインエコーである場合、信号 $N(t)$ は留守番電話機内に記憶されている音声アナウンスにより定まり、これを利用できる。また、音声認識装置1が車両で用いられる機器内に統合されていて、ノイズ信号 $N(t)$ が走行時の騒音により引き起こされるとみなされるならば、認識すべき音声信号 $S(t)$ とはほとんど無関係に妨害ノイズを検出する車両内の付加的なマイクロフォンを用いて走行時の騒音を捕捉することで、信号 $N(t)$ が得られる。この場合、音声アナウンスがユーザガイドに用いられるところでは、信号 $N(t)$ は機器ユーザ向けの音声アナウンスの音響エコーを表す可能性もある。ここではいずれにしても、音響エコーとその他の妨害ノイズ(たとえば車両騒音)とが別個に記録されるようにした、信号 $N(t)$ のマルチチャンネル記録が有利である。

【0015】

ノイズ信号源と受信分岐2の受信手段との間の既述の伝達関数に整合される適応形フィルタ4により、信号 $N(t)$ からノイズ信号推定値 $N^{\wedge}(t)$ が形成され、このノイズ信号推定値 $N^{\wedge}(t)$ はノイズ信号 $N(t)$ に対する近似として用いられる。適応形フィルタ4は受信分岐2と結合されていて、信号 $X(t)$ がノイズ信号 $N(t)$ だけから成る音声休止中、すでに存在している信号 $N(t)$ および $X(t) = N(t)$ からそのまま得られる既述の伝達関数に応じて、適応形フィルタ4が調整される。

【0016】

減算手段5により、妨害を受けた音声信号 $X(t)$ からノイズ信号推定値 $N^{\wedge}(t)$ が減算され、音声信号 $S(t)$ を近似的に表す音声信号推定値 $X^{\wedge}(t) = X(t) - N^{\wedge}(t)$ が形成される。このようにして、信号 $X(t)$ に含まれているノイズ信号 $N(t)$ をすでに部分的に補償することができる。

【0017】

機能ブロック6により、推定された音声信号 $X^{\wedge}(t)$ に対してかけられた短期間スペクトル分析がまとめられる。この場合、サンプル値として用いられる時間信号 $X^{\wedge}(t)$ が一定長のタイムスロットに分けられる。次に、スペクトル変換たとえばフーリエ変換によって出力密度スペクトルが算出される。このようにして得られたスペクトル値はサブサンプリングされて、それらの個数が低減される。特別な実施形態によれば、得られたサンプリング値が対数化され、続いてそれらに対し高域通過フィルタリング処理がなされる。このことはヨーロッパ特許出願公開第0508547号公報に詳細に述べられている。この場合、受信分岐2で受信された妨害を受けた音声信号 X^{\wedge} は、振幅周波数および変調とはほとんど無関係になる。このようにして機能ブロック6は音声信号推定値 $X^{\wedge}(t)$ から、 D 個の成分ないしスペクトル値 $x_d(i)$ ($d = 1, 2, \dots, D$)を有する第1のスペクトルベクトル $x(i)$ を生成する。ここで i は、短期間スペクトル分析に基づく $X^{\wedge}(t)$ のタイムスロットを表す。

【0018】

時間信号を表すノイズ信号推定値 $N^{\wedge}(t)$ は機能ブロック7へ供給され、このブロックは機能ブロック6のように短期間スペクトル分析を行い、 D 個の成分 $n_d(i)$ ($d = 1, 2, \dots, D$)を有するスペクトルベクトル $n(i)$ を生成する。音声認識装置1は信号プロセッサを用いて実現するのが有利である。この場合、機能ブロック6, 7によ

10

20

30

40

50

りまとめられる両方のスペクトル変換は順次連続して実施される。

【0019】

メモリ8には基準ベクトル $y(j, m)$ が記憶されており($j = 1, 2, \dots; m = 1, 2, \dots, M$)、これらの基準ベクトルは1つの基準音声信号の短期間スペクトルによりそれぞれ形成される。付加的な指標 $j = 0$ に対し音声休止が割り当てられている。対応する基準ベクトルはやはりメモリ8に格納されている。1つの基準音声信号は、指標 j で記述される1つのワードセクションを成している。また、 M により基準ベクトルの個数つまり基準音声信号の個数が表され、これらは1つのワードセクションまたは音声休止を記述するために用いられる。種々の話し手のタイプが用いられる場合、1つの特定のワードセクションに対しそれぞれ異なる基準音声信号が生じる。種々の妨害源形式を考慮することで、音声休止に対しそれぞれ異なる基準音声信号が生じる。 M が1よりも大きければ重み付け係数 $g(j, m)$ によって、個々の比較ベクトルが j 個の付随のワードセクションまたは付随する音声休止の記述にどの程度適しているかが表される。これらの重み付け係数の適用法については後述する。基準ベクトル $y(j, m)$ を形成するために、個々の音声認識の適用事例に必要な基準音声信号が音声認識装置の稼働開始前に記録され、それらに対して機能ブロック6, 7の処理による短期間スペクトル分析がかけられる。そして平均値形成により得られたスペクトルベクトルが基準ベクトル $y(j, m)$ としてメモリユニット8に格納される。

10

【0020】

図2の場合、本発明による音声認識装置1は、車両における機器17用の対話制御インターフェースの1つの構成部分である。機器17はたとえば携帯電話機、交通情報システムまたは無線受信機である。機器17の機能の少なくとも一部分は音声認識装置の音声認識結果Eにより制御され、たとえば携帯電話機に統合されているならば音声認識結果Eによって相手先の電話番号が発せられる。音声認識装置1の受信分岐2は、車両のドライバに対し配向されたマイクロフォン12と接続されており、ドライバの音声信号を受信するために用いられる。第2の受信分岐3は3つの受信経路3a, 3b, 3cにより構成されている。受信経路3aと3bを介して、音声増幅器15からスピーカ装置14へ伝送されるステレオ音声信号が信号 $N_1(t)$ および $N_2(t)$ として音声認識装置へマルチチャンネル形式で導かれる。これらの信号 $N_1(t)$ および $N_2(t)$ は、それらがスピーカ14によって発せられた後で受信分岐2の受信信号 $X(t)$ へ妨害を及ぼし、車両のドライバの音声信号に重畳される。さらにこの場合、付加的なマイクロフォン13が設けられており、このマイクロフォンは妨害ノイズを捕捉するために用いられ、そのマイクロフォン信号は信号 $N_3(t)$ として受信経路3cへ導かれる。この特別な事例では受信経路3cにおいて3つの受信経路3a, 3b, 3cを介して、それぞれ1つのノイズ信号源の信号に対応する異なる複数の信号 $N_1(t)$, $N_2(t)$ および $N_3(t)$ が受信されるので、これらの信号をコスト的に有利に処理するために図1による音声処理システムにおいて、適応形フィルタ4および減算手段5を省略するかないしは減算手段5による減算を省略するよう変形できる。信号 $N_1(t)$, $N_2(t)$ および $N_3(t)$ の処理については、図4を説明する際に詳細に説明する。

20

30

【0021】

対話式制御装置16により、音声認識装置1の認識結果Eに依存して車両のドライバへのリターンメッセージとして音声アナウンスが発せられ、さらにこの装置16により、車両のドライバとの対話中にたとえば音声増幅器15のミュート回路も制御される。

40

【0022】

図1による音声認識装置のその他の動作モードについては多数の変形実施例がある。次に、図3を参照しながら第1の変形実施例について詳細に説明する。

【0023】

この第1の変形実施例の場合、 $M = 1$ にセットされ、つまり各指標 j に対して正確に1つの基準ベクトル $y(j)$ が割り当てられる。したがって、図3の縦座標上に表されたダイアグラムにプロットされている1つの指標 j に対し、正確に1つの基準ベクトル $y(j)$

50

)が割り当てられる。この場合、メモリ8により基準ベクトル $y(j)$ が機能ブロック9へ伝送され、この機能ブロック9は各基準ベクトル $y(j)$ と第2のスペクトルベクトル $n(i)$ から比較ベクトル $y'(i, j)$ を形成するために用いられる。各指標 j に対し時点 i で1つの比較ベクトル $y'(i, j)$ が割り当てられる。図3による変形実施例では、音声休止を記述する比較ベクトル $y'(i, 0)$ だけが継続的に第2のスペクトルベクトル $n(i)$ に整合される。比較ベクトル $y'(i, 0)$ は、

$$y'(i, 0) = n'(i) = a \cdot n(i)$$

により求められ、ここで a は一定の所定の係数を表し、これによって減算手段5を用いたノイズ信号推定値 $N^{\wedge}(t)$ の減算によるノイズ信号 $N(t)$ の減衰が表される。したがって音声信号推定値 $X^{\wedge}(t)$ には、妨害を受けた音声信号 $X(t)$ よりも係数 a だけ減衰されたノイズ信号成分が含まれている。減算手段5による差の形成を省略した場合、係数 a は値 $a = 1$ をとる。このような省略を行っても音声認識装置1はまだ機能を果たせるが、減算手段5による差の形成により音声認識結果は改善される。係数 a は経験的に求めることができ、音声認識装置1の使用条件に依存するものである。対数化されたスペクトル値を用いる場合、 $y'(i, 0) = n'(i)$ は $\ln(a)$ を $n(i)$ の成分と加算することにより形成される。すべてについて $j > 0$ であれば、

$$y'(i, j) = y'(j) = y(j)$$

が成り立つ。

【0024】

したがって比較ベクトル $y'(i, j)$ は時間とは無関係であり、つまり時間変数 i には依存せず、基準ベクトル $y(j)$ に等しい。

【0025】

比較ユニット10において、機能ブロック9から供給された比較ベクトル $y'(i, j)$ および第2のスペクトルベクトル $x(i)$ から、

【0026】

【数1】

$$d(i, j) = \sum_{d=1}^D |x_d(i) - y_d'(i, j)|^p$$

【0027】

にしたがって距離 $d(i, j)$ が形成される。ここで $x_d(i)$ と $y_d'(i, j)$ は、第1のスペクトルベクトル $x(i)$ と比較ベクトル $y'(i, j)$ の成分を表す。値 p はこの実施例では $p = 1$ にセットされるが、 $p = 2$ とした距離算出も可能である。すべての指標 j について各サンプリング時点 i ごとに距離が形成される。これらは第1のスペクトルベクトル $x(i)$ と比較ベクトル $y'(i, j)$ との間の類似性を表す尺度を成し、この場合、各ベクトル間の類似性は距離が小さくなるにつれて増大する。図3では、比較ユニット10へ供給される第1のスペクトルベクトル $x(i)$ は、ダイアグラムの横座標にプロットされたサンプリング時点 i に対応づけられている。

【0028】

機能ブロック11により、比較ユニット10において求められた距離 $d(i, j)$ の評価が行われる。各サンプリング時点 i ごとに各指標 j に対しそれぞれ所属の距離が求められる。この評価はたとえば(ヴィタビ・アルゴリズムとも呼ばれる)ダイナミックプログラミング方式により行われ、これについてはたとえばIEEE Transaction on Acoustics, Speech, and Signal Processing, ASSP-32(1984)におけるH. Neyによる論文の第263頁~271頁に記載されている。距離 $d(i, j)$ を評価するためのさらに別の可能性として、Hidden-Markovモデルを用いることが挙げられ、そこにおいてヴィタビ・アルゴリズムまたはその他の最適化アルゴリズムが適用される。この場合、図

3によるダイアグラム $j(i)$ において、結果として最小の総和距離となる経路が求められる。

【0029】

図3による実施例では、音声認識装置1の語彙中にはたとえば単語 "stop" が含まれており、この単語の音形が基準ベクトル $y(1) \sim y(4)$ により表されている。音声認識装置は対話形システムの一部を成し、このシステムはまずはじめに単語 "stop" をアナウンス $N(t)$ として送出する。このアナウンスによりノイズ信号として作用するエコー信号 $N(t)$ が生じ、これは受信分岐2により受信される信号 $X(t)$ の構成部分となる(最初の "stop" が $X(t)$ に入る)。図3の場合、エコー遅延時間は無視できる程度に小さいものとしている。次に単語 "stop" が発声されて、音声認識装置1の受信分岐2により認識すべき音声信号 $S(t)$ として記録される。音声信号 $S(t)$ は信号 $X(t)$ の一部分を成すにすぎず、 $S(t)$ とは無関係に生成される信号 $N(t)$ の一部分ではない。図3には、個々の文字が繰り返されることで発声速度の変化している様子が示されている。ここに図示されている各単語の構成部分(文字)とスペクトルベクトル $x(i)$ との対応づけは、必ずしも行う必要はない。また、長かったり短かったりする単語の構成部分も、基本的にそれぞれ1つのスペクトルベクトル $x(i)$ によって記述できる。

10

【0030】

共に距離マトリクスを形成する距離 $d(i, j)$ を計算する際には、機能ブロック6により形成された第1のスペクトルベクトル $x(i)$ が比較ベクトル $y'(i, j)$ と比較される。ここで $y'(i, 0) = n'(i)$ はノイズエコー信号 $N(t)$ を反映する。これにより、信号 $X(t)$ の第1のワード(ノイズエコー信号に相応する)に対して、インデックス $j = 0$ に対する特に小さな距離が得られる。

20

【0031】

この実施例ではノイズエコー信号は識別語彙にも含まれるから、基準ベクトルのシーケンス $y(1), y(2), y(3), y(3), y(4)$ 、ないし適合された比較ベクトルのシーケンス $y'(i, 1), y'(i, 2), y'(i, 3), y'(i, 3), y(i, 4)$ により表すことができる。これは図3に破線で示されている。にもかかわらず、サンプリング時点 $i = 1$ から $i = 6$ に対する経路はインデックス $j = 0$ に留まる。これは、破線の経路ではワード "ストップ" の一般的表現がこのワードの固有の発音と比較され、一方 $j = 0$ に対してはこのような発音変化が比較を損ない、距離の増大することがあるからである。

30

【0032】

これに対して、音声信号 $S(t)$ 、すなわち識別すべき有用信号を表す振動 $X(t)$ の第2のワードは、シーケンス $y(1), y(1), y(2), y(3), y(4)$ との比較で小さなディスタンス $d(i, j)$ しか示さない。これはベクトル $n'(i)$ がこの領域では音声休止を表すからである。

【0033】

このようにして機能ブロック11は、図3に示された所属経路 $j(i)$ を送出する。この所属経路は、信号 $X(t)$ のエコー(第1の "ストップ")の間は信号 $X(t)$ における音声休止を表し、実際に発声されたワード(第2の "ストップ")の間は識別されたワードを表す。この所属経路 $j(i)$ は機能ブロック11から送出された識別結果 E に相応する。

40

【0034】

ノイズ信号 $N(t)$ としての音声信号は、例えば対話システムでのエコーとして発生する。この音声信号は通常の音声認識装置に対しては特に問題となる。なぜならこの信号は有用信号 $S(t)$ がない場合では強く減衰しても、ワードを機能ブロック11から送出された識別結果 E に簡単にワードを間違っ挿入してしまい、エラー経路 $j(i)$ となるからである。このことは本発明の音声認識装置により回避される。

【0035】

50

ノイズ信号 $N(t)$ および有用信号 (音声信号) $S(t)$ が同時に発生する場合、有用信号休止に相応する基準ベクトル $y(0)$ だけを置換するのでは十分でない。むしろすべての基準ベクトル $y(j)$ を適合しなければならない。このことは図3には示されていない。

【0036】

適合された比較ベクトルは次式にしたがって形成される。

【0037】

$$y'(i, j) = y(j) + n'(i)$$

対数スペクトル値を使用する場合には、適合された比較ベクトルの成分は次式にしたがって計算される。

【0038】

$$y d'(i, j) = \ln(\exp(y d(j)) + \exp(n d'(i)))$$

計算コストを低減するために、この成分は次式により近似される。

【0039】

$$y d'(i, j) = \max(y d(j), n d'(i))$$

この手段により、音声情報の隠蔽がモデル化される。なぜなら、インデックス j および周波数バンド d に対して記憶されたスペクトル値 $y d(j)$ が瞬時のノイズ信号スペクトル $n d'(i)$ により、後者の方が大きい値を有する場合に置換されるからである。このようにして、音声情報の瞬時の隠蔽部分がノイズ信号 $N(t)$ ないし $N(t)$ により比較の際に考慮され、これにより個々の発音が完全にバックグラウンドノイズの中で行われても、ワードの識別が可能となる。

【0040】

ハイパスフィルタリングされたスペクトル値が基準ベクトルの成分として使用される場合には、隠蔽を上述べたようにパワー付加により計算することはできない。この場合に対しては本発明は変形される。以下それを説明する。

【0041】

音声信号推定値 $X^{\wedge}(t)$ におけるノイズ信号パワーがノイズ信号成分 $N(t)$ のパワーに対して、すべての周波数帯域で均一に一定の係数 a だけ減衰されていない場合には、音声認識装置1は最適の音声識別結果を得るために変形しなければならない。とくに、時間領域で動作する適合係数4が省略される場合には、 $n'(i)$ はスペクトルベクトル $n(i)$ に依存するだけでなく、ノイズ信号源と受信経路2との間の伝送関数に所属するパルス応答の持続時間に相応して、複数のスペクトルベクトル $n(i-r)$ にも依存する。ここで $r = 0, 1, \dots, R$ 。

【0042】

この考察から本発明は次のように変形される。すなわち、機能ブロック9により比較ベクトル $y'(i, j)$ 次のように形成されるのである。すなわち、各サンプリング時点 i に対する1つのインデックス j に、少なくとも1つの基準ベクトル $y(j, m)$ と、付加的に1つまたは複数の順次連続する第2のスペクトルベクトル $n(i-r)$ を配属するのである。図4の所属の実施例では、各インデックス j に対して1つの基準ベクトル $y(i)$ と2つの順次連続する第2のスペクトルベクトル $n(i-1)$ および $n(i)$ が比較ベクトル $y'(i, j, m)$ として設けられている。ただし、 $m = -R, \dots, M = -1, 0, 1, (M = 1, R = 1)$ 。

【0043】

以下のことが当てはまる。

【0044】

$$y'(i, j, -1) = n(i-1)$$

$$y'(i, j, 0) = n(i)$$

$$y'(i, j, 1) = y(j)$$

このために、機能ブロック9にはスペクトルベクトル $n(i-1)$ が中間記憶される。重み付け係数 $g(i, j, m)$ (ただし $m = -R, \dots, M$) により、 $n(i)$ ないし $n(i)$

10

20

30

40

50

$i - r$) により表されるノイズの隠蔽がインデックス j に相応する基準語の部分によって行われることが予想されるか否かが考慮される。この予想は、2つの信号が各時点で付加されるなら、一般的に一方の信号は他方の信号をほぼ隠蔽し、したがってスペクトルベクトルは実質的に2つの信号のどちらかに相応する、という過程に基づくものである。 $M > 1$ に対しては(すなわち複数の基準ベクトルが1つのインデックス j に配属される)、相応の重み付け係数 $g(i, j, m)$ が、相応する適用例において同じインデックス j を有する一方の基準ベクトルが他方の基準ベクトルと比較して、音声信号を表すのにより適切であるか否かを示す。

【0045】

このようにして形成された、比較ベクトルにおける時間依存量により、またこれを機能ブロック9から比較ユニット10に送出することにより、比較ユニット10により検出すべきディスタンスとして次のものが得られる。

【0046】

【数2】

$$d(i, j) = \min_{m=-R}^M \{ g(i, j, m) + \sum_{d=1}^D | x_d(i) - y_d'(i, j, m) |^p \}$$

$$= \min_{m=1}^M \{ \min_{d=1}^D | x_d(i) - y_d(j, m) |^p \},$$

$$\left\{ \min_{r=0}^R \{ g(i, j, -r) + \sum_{d=1}^D | x_d(i) - n_d(i-r) |^p \} \right\}$$

【0047】

重み付け係数 $g(i, j, m)$ はまず粗く評価される。上にすでに述べたように、 p は1に設定することができ、 $p = 2$ も可能である。別の距離計算ではわずかな付加コストしか必要としない。なぜなら、付加項が重み付け係数 $g(i, j, -r)$ により j だけに依存することができ、 $x(i)$ とノイズベクトル $n(i)$ から $n(i - N)$ までの距離は各時点 i ごとに一度計算すればよいからである。付加項とは、 r が伝搬変数として現れる項をいう。

【0048】

距離 $d(i, j)$ の検出は、変形実施例にしたがって次のように行われる。すなわち、比較ユニット10によりスペクトルベクトル $x(i)$ とこれら3つのベクトル間の部分距離がそれぞれ重み付け係数 $g(i, j, m)$ により加算され、引き続き最小値が選択されるようにして行われる。この実施例では、メモリ8にファイルされた基準ベクトル $y(j)$ に相応するすべての時間不変比較ベクトル $y'(i, j, 1)$ に対して、重み付け係数 $g(i, j, 1)$ として値ゼロが設定される。ノイズ信号 $N(t)$ にエコー遅延に関する知識が仮定されなければ $g(i, j, -1) = g(i, j, 0)$ が良好な音声識別結果を得るための式である。

【0049】

重み付け係数 $g(i, j, m)$ はとくに、純粋なノイズ信号 ($X(t) = N(t)$) の場合に、休止状態 $j = 0$ が優先されることを保証しなければならない。このことは、 $j = 1$ に対して $g(i, j, 0) > g(i, 0, 0)$ の重み付け係数が設定される場合に達成される。このことは図4に、相応する比較ベクトル $n(i)$ または $n(i - 1)$ の下線に

10

20

30

40

50

よって示されている。

【0050】

相対的に信号パワーが大きく、確実に識別できる言語音、例えば母音に対しては、重み付け係数 $g(i, j, 0)$ を $j = 1$ に対して特に大きく選択することができる。なぜなら、所属のワードの識別はこのような言語音がノイズ信号 $N(t)$ により隠蔽されない場合だけ考慮されるからである。このようにして重み付け係数 $g(i, j, m)$ のインデックス j への依存性により、個々のワード区間の隠蔽またはノイズ信号 $N(t)$ による隠蔽を勘案すべきか否かが表される。このことは特に休止状態に対して当てはまる。

【0051】

図4では、図3と同じ信号が基礎となっている。ここでは特にエコー信号、第1の“ストップ”のエコー遅延が信号 $X(t)$ において信号 $N(t)$ に対して考慮される。ほぼ1つのサンプリング間隔のエコー遅延が示されている。このエコー遅延にかかわらず、音声認識装置1は正しい識別結果 E を送出する。音声識別の際の脱落エラーの危険性は非常に小さい。なぜなら、それぞれ個々のスペクトルベクトル $x(i)$ は $R+1$ ノイズベクトル(第2のスペクトルベクトル)と、すなわちノイズ信号 $N(t)$ ないし $N(t)$ の非常に短い区間と付加的に比較されるだけであるからである。

【0052】

図2に示したマルチチャネル・ノイズ信号の場合は、信号 $N_1(t)$ 、 $N_2(t)$ および $N_3(t)$ はこの場合、直接ノイズ信号推定値として機能ブロック7のスペクトル分析手段により第2のスペクトルベクトル $n_1(i)$ 、 $n_2(i)$ および $n_3(i)$ にさらに処理される。ここで時間変数 i の値に対して、3つのスペクトルベクトル $n_1(i)$ 、 $n_2(i)$ および $n_3(i)$ からのセット、例えば時間的に順次連続する2つのスペクトルベクトル ($n(i-1)$ および $n(i)$) のセットが、受信分岐路3が1つだけの受信経路からなる(図4参照)場合に、各インデックス j に配属される。数式の組み合わせは、複数の受信経路も、各インデックスに対して時間的に順次連続する複数のスペクトルベクトルも考慮されるように行われる。この組み合わせも有利な実施例あり、例えば時間変数 i の値に対するインデックス j に、2つのスペクトルベクトル $n_1(i-1)$ 、 $n_1(i)$ 、 $n_2(i-1)$ 、 $n_2(i)$ 、 $n_3(i-1)$ および $n_3(i)$ からのそれぞれ1つのセットが配属される。

【0053】

とくにエコー伝搬時間が大きい場合は、信号 $X^{\wedge}(t)$ におけるノイズ信号の抑圧は適応形フィルタ4の精度に依存する。この要求は比較ベクトルを機能ブロック9で連続的に適合することにより格段に低減される。ここに述べたすべての変形実施例で、適応形フィルタ4を省略することができる。減算手段5による差形成はすでに上に述べたのと同じように必ずしも必要ではない。

【0054】

本発明の音声認識装置1により、次のような適用例に対して信頼性の高い音声識別が実現される。すなわち、例えば自動車で使用される携帯電話または無線受信機での使用、および「電話を介して通信する呼びかけ応答または自動案内装置での使用の際に、機器の音声制御が行われるような適用例で実現される。

【0055】

自動車で使用される携帯電話または無線受信機では、使用者に音響ユーザーインターフェースを介して出力される音声の音響エコーがノイズ信号として発声する。自動車の無線受信機では、使用者の音声信号の識別が無線送信の出力(例えば音楽)により障害を受けることもある。遠距離通話中に音声信号により制御可能な呼びかけ応答機の場合、ラインエコーがノイズ信号として発生する。このラインエコーは呼びかけ応答機の声の反射によって使用者への伝送の際に生じるものである。自動案内装置の場合は、呼びかけ応答機と同じように音声に起因するラインエコーがノイズ信号として発生する。このようなラインエコーがこのような機器での音声識別に障害的な影響を与えることが本発明の音声認識装置によって最小の程度となる。

10

20

30

40

50

【0056】

さらに音声認識装置1により、自動車の部分機能の音声制御を実現することができる。例えば、窓開閉を制御することができる。また口述記録機の音声識別の際のエラー率を低減することができる。この口述記録機は発声された音声を直接テキストに変換するものであり、本発明の音声認識装置を使用することができる。

【0057】

前記の音声認識装置ではノイズ成分および有用信号成分の推定を省略しており、有用信号またはノイズ信号としてそれぞれ受信された信号の個々の時間間隔の解釈を最終的な識別決定と関連して行っているため、音声認識装置はノイズ信号に対して非常に頑強である。

10

【0058】

音声認識装置で付加的に必要な計算コストはわずかである。したがって、音声認識装置は、識別可能な語数の大きな語彙を有する適用例にも、相応に語彙容量が小さい適用例にも使用することができる。

【0059】

【数3】

なお、明細書中の N は、図面では \tilde{N} を表し、

N^{\wedge} は、図面では \hat{N} を表すものである。

20

【図面の簡単な説明】

【図1】 本発明の音声認識装置のブロック回路図である。

【図2】 対話型ユーザーインターフェースを拡張し、本発明の音声認識装置を有する構成のブロック回路図である。

【図3】 音声識別の時間経過を示す線図である。

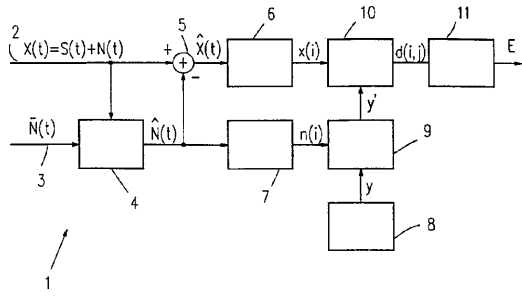
【図4】 音声識別の時間経過を示す線図である。

【符号の説明】

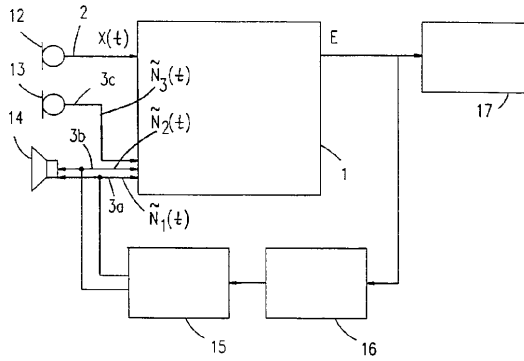
- 1 音声認識装置
- 2 第1の受信分岐
- 3 第2の受信分岐
- 4 適応形フィルタ

30

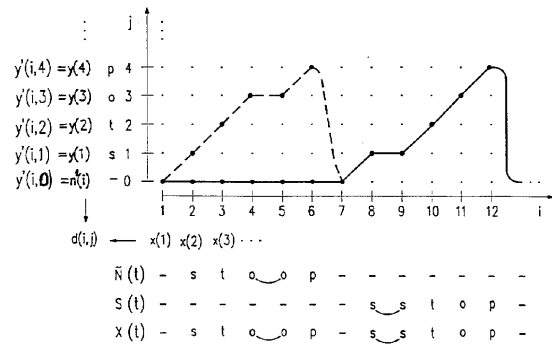
【 図 1 】



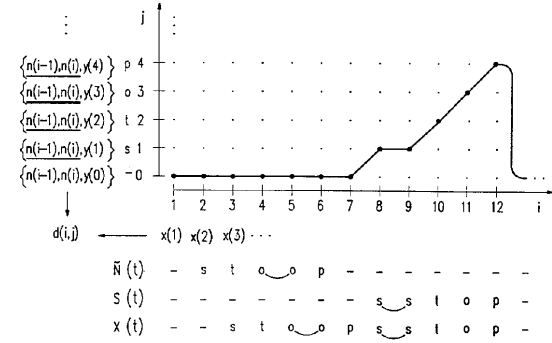
【 図 2 】



【 図 3 】



【 図 4 】



フロントページの続き

(74)代理人 230100044

弁護士 ラインハルト・アインゼル

(72)発明者 ハンス カルヴェラム

ドイツ連邦共和国 ニュルンベルク プファラー - グリュック - シュトラーセ 38

審査官 榎本 剛

(56)参考文献 特開平01 - 255000 (JP, A)

特開平02 - 083593 (JP, A)

特開平04 - 295897 (JP, A)

特開平08 - 320698 (JP, A)

特開平02 - 083597 (JP, A)

特開平04 - 057097 (JP, A)

特開平04 - 347898 (JP, A)

特開平05 - 066790 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00-15/28