

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5644013号
(P5644013)

(45) 発行日 平成26年12月24日 (2014. 12. 24)

(24) 登録日 平成26年11月14日 (2014. 11. 14)

(51) Int. Cl. F I
G 1 0 L 15/20 (2006. 01) G 1 0 L 15/20 3 7 0 D
G 1 0 L 21/0208 (2013. 01) G 1 0 L 21/0208 1 0 0 B
 G 1 0 L 15/20 3 8 0

請求項の数 20 (全 12 頁)

(21) 出願番号	特願2013-513424 (P2013-513424)	(73) 特許権者	591003943 インテル・コーポレーション
(86) (22) 出願日	平成23年6月30日 (2011. 6. 30)		アメリカ合衆国 95054 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・2200
(65) 公表番号	特表2013-531275 (P2013-531275A)	(74) 代理人	110000877 龍華国際特許業務法人
(43) 公表日	平成25年8月1日 (2013. 8. 1)	(72) 発明者	ベルトマン、ウィリアム エム. アメリカ合衆国 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・2200 インテル・コーポレーション内
(86) 国際出願番号	PCT/US2011/042515		
(87) 国際公開番号	W02012/003269		
(87) 国際公開日	平成24年1月5日 (2012. 1. 5)		
審査請求日	平成24年11月29日 (2012. 11. 29)		
(31) 優先権主張番号	12/828, 195		
(32) 優先日	平成22年6月30日 (2010. 6. 30)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 発話音声処理

(57) 【特許請求の範囲】

【請求項 1】

ノイズ抑制技術を用いて処理される音声知覚経路であり、聴覚的受信者 (a u d i t o r y r e c e i v e r) へと提供される第 1 音声発話経路と、経路の先において独立した第 2 音声発話経路とを有する発話処理エンジンと、

前記第 2 音声発話経路に連結されて、音声発話信号を受信し、発話者の声門情報を含む発話者モデルに少なくとも部分的に基づいて前記音声発話信号からノイズを除去するカルマンフィルターと

を備える装置。

【請求項 2】

前記フィルターは、環境ノイズ情報を組み込んだノイズモデルにも基づいて、ノイズを除去する請求項 1 に記載の装置。

【請求項 3】

前記環境ノイズ情報は、リアルタイム情報を含む請求項 2 に記載の装置。

【請求項 4】

前記リアルタイム情報は、1 つ以上のノイズセンサからの情報を含む請求項 3 に記載の装置。

【請求項 5】

前記発話者モデルには、以前に生成された、ノイズ除去済みの、発話者についての発話信号情報が組み込まれている請求項 1 から 4 のいずれか一項に記載の装置。

【請求項 6】

前記フィルターは、フロントエンド部内に実装され、前記発話者モデルは、前記音声発話信号に発話が検出された場合に有効にされるバックエンド部内に実装される請求項 1 から 5 のいずれか一項に記載の装置。

【請求項 7】

前記発話処理エンジンは、発話認識エンジンを含む請求項 6 に記載の装置。

【請求項 8】

前記発話認識エンジンは、前記バックエンド部の一部である請求項 7 に記載の装置。

【請求項 9】

ノイズ抑制技術を用いて処理される音声知覚経路であり、聴覚的受信者へと提供される第 1 音声発話経路と、経路の先において独立した第 2 音声発話経路とを含む発話処理エンジンを有する音声処理部と、

前記第 2 音声発話経路に連結されて、音声発話信号を受信し、発話者の声門情報を含む発話者モデルに少なくとも部分的に基づいて前記音声発話信号からノイズを除去するカルマンフィルターと

を備える電子デバイス。

【請求項 10】

前記フィルターは、環境ノイズ情報を組み込んだノイズモデルにも基づいて、ノイズを除去する請求項 9 に記載の電子デバイス。

【請求項 11】

前記環境ノイズ情報は、リアルタイム情報を含む請求項 10 に記載の電子デバイス。

【請求項 12】

前記リアルタイム情報は、1 つ以上のノイズセンサからの情報を含む請求項 11 に記載の電子デバイス。

【請求項 13】

前記発話者モデルには、以前に生成された、ノイズ除去済みの、発話者についての発話信号情報が組み込まれている請求項 9 から 12 のいずれか一項に記載の電子デバイス。

【請求項 14】

前記フィルターは、フロントエンド部内に実装され、前記発話者モデルは、前記音声発話信号に発話が検出された場合に有効にされるバックエンド部内に実装される請求項 9 から 13 のいずれか一項に記載の電子デバイス。

【請求項 15】

前記発話処理エンジンは、発話認識エンジンを含む請求項 14 に記載の電子デバイス。

【請求項 16】

前記発話認識エンジンは、前記バックエンド部の一部である請求項 15 に記載の電子デバイス。

【請求項 17】

前記フィルターは、発話認識に用いられる請求項 1 から 8 のいずれか一項に記載の装置

【請求項 18】

フィルタリングされた前記音声発話信号は、前記第 2 音声発話経路に接続された発話認識エンジンに供給される請求項 17 に記載の装置。

【請求項 19】

前記フィルターは、パルス形の周期的成分を付加して発話ソースが発生した声紋調波を構成する請求項 1 から 8 および請求項 17 から 18 のいずれか一項に記載の装置。

【請求項 20】

前記第 1 音声発話経路は、複製された発話音声信号を入力する請求項 1 から 8 および請求項 17 から 19 のいずれか一項に記載の装置。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【0001】

本発明は、一般的に音声処理に関し、特に発話信号処理に関する。

【図面の簡単な説明】

【0002】

本発明の実施形態は、添付の図面において、限定としてではなく例示として示され、図面においては、同様の参照番号により同様の要素を示す。

【図1】いくつかの実施形態に係る発話処理エンジンの図である。

【図2】いくつかの実施形態に係るシンセサイザーの図である。

【図3】いくつかの実施形態に係る発話処理エンジンを実施するための構造を示す図である。

【図4】いくつかの実施形態に係る電子デバイスプラットフォームの図である。

【発明を実施するための形態】

【0003】

たとえば、車載機能付きのモバイルインターネットデバイスおよびキーボード機能が限定されている電話等では、音声命令および連続発話認識が用いられている。いかなる発話認識エンジンにもクリーンな入力を供給することができることが望まれるが、環境におけるバックグラウンドノイズによってこれは妨害されている。たとえば、オープンに与えられる口述命令語(open dictation word)の正確性は、自動車の騒音および喫茶店の環境等では約20%に低下する可能性があることが実験で示されており、ユーザにとっては許容範囲を超える。

【0004】

今日の発話エンジンは、バックグラウンドノイズの影響を低減させるいくつかのノイズ低減機能を有する。しかし、これらの機能では、困難な環境においてオープンに口述命令を与えることを可能にするには十分ではないかもしれない。したがって、発話信号処理を向上させるべく、カルマンフィルタリング技術を用いる場合がある。

【0005】

本明細書に提示されるいくつかの実施形態では、音声ノイズフィルタリング処理を、別々の発話認識経路と人間による受信経路とに分岐することにより、発話認識性能が強化されうる。つまり、音声経路を模造して、「知覚」(もしくは聴覚による受信)チャンネルと、発話認識エンジン用に音声の前処理するべく用いられる別個のチャンネルとを生成する。

【0006】

図1は、いくつかの実施形態に係る発話処理エンジン102のブロック図である。発話処理エンジンは、カルマン法に基づくフィルタリングエンジン104、発話者/音声モデル106、環境ノイズモデル107、自動発話認識(ASR)エンジン108、および標準ノイズ抑制ブロック110を備える。

【0007】

音声(たとえば、マイクからのデジタル化音声)は、SPE(発話処理エンジン)に入力されて2つの経路に、つまり、カルマンフィルターブロック104に至る発話認識経路と、ユーザに聴き取られるように、ブロック110において標準ノイズ抑制技術を用いて処理された音声知覚経路(複製された音声)とに分割される。カルマンフィルターは、発話者/音声モデル106と、環境ノイズモデル107とからの成分を用いて音声信号からノイズを除去し、フィルター済み信号を自動発話認識(ASR)エンジン108に供給する。

【0008】

発話者/ノイズモデル106(少なくとも、初期バージョン)は、初期バージョンが未熟であっても、SPEはそれを用いて動作するので、SPEが実行される前に生成され、SPEが実行されている間に更新される。発話者/音声エンジン106は、現在の発話者に対応付けられた特定の特性を提供する。このような特性は、ユーザ固有の基本声門周波数を含む1つ以上の声門調波(global harmonics)、およびその他の任意の適切な情報を含みうる。たとえば、以前に取得したモデル(たとえば、ユーザの訓

10

20

30

40

50

練から得られたもの)が利用可能である場合、それらを発話者/ユーザモデル106に組み込んでよい。図示のように、特定のユーザについて以前に生成された「クリーン」な音声情報(x'(n))を用いてもよい。

【0009】

発話者/音声モデルと同じく、環境ノイズモデル107も、想定したノイズ環境、または特定の環境もしくは以前に特性化した環境(たとえば、オフィス、自動車、飛行機等)についての初期設定データ/想定に基づいたものでよい。環境ノイズモデルは、環境に対応付けられた静的データ(たとえば、想定したバックグラウンドノイズ成分)、および/または、リアルタイムセンサ等から得られる動的データであってよい。たとえば、環境ノイズモデルは、ノイズモデル推定器の性能を強化するべく、自動車速度、バックグラウンドノイズをマイクで拾ったデータ、空調情報等のセンサ入力を含んでよい。いくつかの実施形態では、たとえば単一のチャンネルについて、音声活動検出器アルゴリズムを用いて無発話期間を検出することによるノイズ推定方法を採用してよい。ノイズモデルとカルマンフィルタリングとの間に反復ループを用いることにより、ノイズモデルをさらに強化してよい。

10

【0010】

フィルター104は、発話者モデルおよびノイズモデルの一方もしくは両方を用いて、受信した音声信号をフィルタリングしてよい。フィルターは、再び発話者モデルから、パルス形の周期的成分をカルマンフィルタリングに付加して、発話ソース(たとえば、人間、または、たとえば口述命令装置、音声制御装置、もしくは翻訳装置を用いるその他の発話エンティティ)が発生した声門調波を構成する拡張(extension)を用いてよい。カルマンフィルタリングは白色ノイズ入力に通常用いられてきたが、人間の発話の場合、周期的入力の付加により発話発生をより良く真似ることができるであろう。所定のモデル情報と声門調波パラメータとを含む発話者モデルを用いて、発話者モデルについて、所定の、もしくは以前に決定した、係数のセットをロードしてよい。カルマンフィルタリングによって、必ずしも人間による聞こえ方が顕著には改善しない音声生成されるが、通常、発話認識エンジンの性能は向上する。人間による知覚と、カルマン前処理フィルタリングを用いる発話認識入力との両方を最大限に高めるべく、音声経路を模造(2つの経路を生成)する。

20

【0011】

独立入力および駆動ノイズ、並びに付加的な有色ノイズを構成するノイズ観測を用いて、声道応答をARもしくはARMAシステムとしてモデル化するべく、カルマン技術を利用して実装されたフィルター104を用いることができる。

30

【0012】

従来のカルマン法の用法では、周期的な駆動入力通常無視され、簡略化のため、駆動白色ノイズだけが用いられる。この想定が示唆しているのは、(理想的な性能下では)フィルターによって、生理値を持たず、自然な聞こえ方もしない、クリーンであるが無声の発話信号が生成されるということである。しかし、この想定は、フィルターパラメータだけが必要である場合には十分であろう。

【0013】

他方、線形カルマンフィルターでは、音声生成において観察される基本的な双方向特性が捕らえられ、したがって、ノイズ条件下において、クリーンな入力について、より良好な推定値が生成されると我々は判断した。たとえば、CP分析およびソースモデリングと組み合わせると、線形カルマンフィルターは、発話処理で、さらに良好な性能を発揮する。この種のスキームにおけるエラーは、生理学的/音響学的な発声の失敗(misrepresentation)の結果にではなく、そのスキームのパラメータ推定エラーに対応付けられる。したがって、本明細書に開示される発話増強スキームは、線形カルマンフィルターに基づいており、その構成は、以下の表の「線形」という項目名の下に示されている。

40

【表 1】

(a) 非線形	(b) 線形
$x_{k+1} = f(x_k, u_k, w_k)$ $y_k = h(x_k, v_k)$	$x_{k+1} = \Phi x_k + B u_k + w_k$ $y_k = H x_k + v_k$

【 0 0 1 4 】

10

状態 x_k は、声門ソース u_k および環境ノイズ w_k により生成されるクリーンな発話入力に対応する。（ x は、S P E への実際の入力ではない）。測定された信号 y_k は、観測ノイズ v_k だけ損失している。上記したように、以前のカルマン法アプローチでは、簡略化のために周期的入力 u_k を無視し、白色ノイズが増加した発話を生成していた。しかし、この周期的入力を含め、状態遷移行列を C P 表記することにより、クリーンな入力 x_k について、より良好な推定値が得られ、したがって発話認識性能が良好になる。以下の項では、本明細書において用いられるカルマンフィルタリングをより詳細に記載する。

【 0 0 1 5 】

いくつかの実施形態では、発話の増強にカルマンフィルタリングモデルに基づくアプローチを用いる。このアプローチでは、クリーンな発話は、バックグラウンドノイズにより線形に損失した特定の信号（representation）の後に続く想定する。通常、標準的カルマンフィルタリングでは、入力として通常、白色ガウスノイズを有する自己回帰（AR）モデルを用いて、クリーンな発話を表現する。クリーンな発話は、離散時間方程式 1 で表される。

20

【数 1】

$$x[n] = \sum_{k=1}^p \alpha_k x[n-k] + w[n] \quad (1)$$

ここで、 $x[n]$ はクリーンな発話であり、 α_n は AR もしくは線形予測符号化（LPC）の係数であり、 $w[n]$ は白色ノイズ入力であり、 p は AR モデルの次数である（通常、経験則として $p = f s / 1000 + 2$ に従うと想定され、 $f s$ は kHz 単位のサンプリングレートである）。このモデルを書き換えて、式（2）および（3）で示されるような、カルマンフィルタに必要とされる所望の構成を生成することができる。したがって、以下ようになる。

30

【数 2】

$$x_{k+1} = \Phi x_k + G w_k \quad (2)$$

【数 3】

$$y_k = H x_k + v_k \quad (3)$$

ここで、 x_{k+1} および x_k は、今後および現在のクリーンな発話についての p 個のサンプルを含むベクトルであり、 Φ は制御可能な正準形の最後の行に LPC 係数を含む状態遷移行列であり、 w_k は白色ノイズ入力を表し、この白色ノイズ入力は、ベクトルゲイン G を介して現在のサンプルに影響を与えるベクトルに変換される。クリーンな発話は、現在のサンプルを取得するべく射影ベクトル H を介して射影され、取得された現在のサンプルはバックグラウンドノイズ v_k に線形に加算されて、損失した観測もしくはノイズを含む発話 y_k が生成される。

40

【 0 0 1 6 】

カルマンフィルタリングには、伝達段階および更新段階の 2 つの基本的な段階が含まれる。伝達段階では、モデルを用いて、以前の推定値（したがって、表記は $n | n - 1$ ）に基づき、現在のサンプルを予測する。これは、式（4）で表される。以前の p 個の点を含

50

む1つのベクトルの1つのバッファだけが必要であることに注意すべきである。更新段階は、式(5)から(7)に表され、まず、予測したサンプルを、予測値と推測値との誤差を考慮して補正する。この誤差は、式(6)および(7)に定義されるカルマンゲイン K_n により制御される。これらのパラメータの全ては、各フレーム内で一回算出してよく、つまり、発話は、各フレーム(通常、25ms以下の持続期間)内における静止したプロセスであると考えられる。

【数4】

$$\hat{x}_{n|n-1} = \Phi \hat{x}_{n-1|n-1} \quad (4)$$

【数5】

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n (y_n - H_n \hat{x}_{n|n-1}) \quad (5)$$

【数6】

$$K_n = P_{n|n-1} H_n^T (H_n P_{n|n-1} H_n^T + R_n)^{-1} \quad (6)$$

【数7】

$$P_{n|n} = I - (K_n H_n) P_{n|n-1} \quad (7)$$

【0017】

本プロジェクトで提案する「修正カルマンフィルター」は、システムにおける2つの基本的なノイズについての想定、つまり、有声区間では声門パルスによってもARモデルが駆動されるという想定、および、バックグラウンドノイズは自身に対応付けられた共鳴音(非白色プロセス)を有するという想定を一般化することにより、標準フィルターを拡張したものである。声門パルスは、 $u[n]$ で表され、声帯振動がある場合に存在する。バックグラウンドノイズは、次数が q (推定すると、経験則的に、たとえば $q = fs / 2000$ と取得される)のARモデルに従うと想定される。したがって、システムの新しい構成を表す2つの式は、以下のようなになる。

【数8】

$$x[n] = \sum_{k=1}^p \alpha_k x[n-k] + u[n] + w_s[n] \quad (8)$$

【数9】

$$v[n] = \sum_{k=1}^q \beta_k v[n-k] + w_v[n] \quad (9)$$

【0018】

発話およびノイズのモデルは似た構成を有するので、カルマンフィルターに必要な状態方程式は、より大きい対角行列に埋め込まれる2つのサブシステムを生成することにより、拡張することができる。式(10)から(13)に示す通りに、同じシステム構成を用いて発話およびノイズを追跡する。これらの式において、下付き文字 s は発話を表し、下付き文字 v はバックグラウンドノイズを表す。声門パルスは、ベクトル B が G と同じ構成を有する現在のサンプルにだけ導入される。

【数10】

$$x_{k+1} = \Phi x_k + B u_k + G w_k \quad (10)$$

【数11】

$$y_k = H x_k + v_k \quad (11)$$

【数12】

$$\Phi = \begin{bmatrix} \phi_s & 0 \\ 0 & \phi_v \end{bmatrix} \quad (12)$$

10

20

30

40

50

【数 1 3】

$$H = [H_s \quad H_v] \quad (13)$$

【0 0 1 9】

カルマン法の伝達および更新を計算する式が標準カルマンフィルタと異なるのは、とりわけ、声門パルスが含まれ、ノイズはフィルタ自身により追跡されているのでノイズ共分散行列 R_n が含まれないという点である。これらの違いは、式(4)を式(14)で修正し、式(6)を式(15)で補正することにより表される。したがって、以下のようになる。

【数 1 4】

$$\hat{x}_{n|n-1} = \Phi \hat{x}_{n-1|n-1} + Bu_k \quad (14)$$

【数 1 5】

$$K_n = P_{n|n-1} H_n^T (H_n P_{n|n-1} H_n^T)^{-1} \quad (15)$$

【0 0 2 0】

これらの補正により、フィルタによって、発話信号およびバックグラウンドノイズの状況がより良く表されるようになり、より良好なノイズ除去とASR性能とが得られる。

【0 0 2 1】

新しいカルマンフィルタリング技術は、発話認識の強化にだけでなく、発話合成の向上にも用いることができる。図2を参照すると、時間領域に基づいたシンセサイザの図が示されている。提案するスキームは、入力信号に適用される3つの相互連結されたプロセスを組み合わせた構成を有する。第1分岐では、ソースコンポーネントの特徴を識別し、ソース信号を生成する。第2分岐では、フィルタ構成を検索し、CP(閉フェーズ)分析およびフルフレーム分析のいずれかを適用してフィルタの線形予測係数(LPC)を定義する。第3分岐では、包絡線を検出し、合成音の安定性を確保する。これらの分岐における計算は、シリアルもしくはパラレルに実行することができ、相互作用のレベルが適切に処理される限り、それぞれ異なるフレームおよびウィンドウの構造(frame and windowing structures)を用いてよい(たとえば、いくつかの実施例では、第1分岐で長方形のウィンドウおよび互いに重ならないフレームを用い、第2分岐でたとえば50%重なるハミングを用いることができる)。

【0 0 2 2】

図3は、電力消費を削減するべく、たとえばモバイルデバイスにおいて、音声処理エンジンのフロントエンドを実装するための一般的な構成を示す。本図では、たとえば図1のSPE102におけるそれぞれのブロックを構成するための電力効率的な態様が示されている。この構成は、計算集約的ブロック301と、メモリアクセス集約的なバックエンド305とに分割されている。計算集約的フロントエンド301は、フィルタ処理部302と、入力音声が発話を含むかを判定する判定ブロック304とを有する。メモリ集約的バックエンド305は、発話者モデルを生成かつ更新する発話者モデルブロック306と、ASRを実施する発話認識ブロック308とを有する。発話者モデルブロック306は、ノイズモデルの全部もしくは一部を生成するノイズモデル部を有してもよい。音声はフロントエンド301に入力され、フィルタ302で処理され、判定ブロック304が発話を含むと判定された場合、フィルタ302からのフィルタ済み発話信号を処理するべく、発話者モデルブロック306および発話認識ブロック308が起動される。

【0 0 2 3】

ハードウェアのフロントエンドでのメモリ要件を低下させることにより、より低電力での動作の実行が可能となり、ワット当たりの動作数が増大する。フロントエンド301における発話増強アルゴリズムのハードウェア実装により、低電力を達成するための機会が与えられ、プロセッサハードウェアのバックエンドに覚醒信号を供給する閾値検出器304の利用が可能となる。バックエンド305は、通常はメモリ集約的かつ高性能な(たと

10

20

30

40

50

えば、隠れマルコフモデルおよび/またはニューラルネットワークに基づいた) 発話認識アルゴリズムをハードウェアで実装したものである。したがって、ハードウェア(たとえば、SPEハードウェア)を計算集約的なフロントエンドと、高性能なバックエンドとに分割することによって、発話の増強および認識のための、「ボイスウェイクアップ」機能および「オールウェイズリスニング」機能も実装してよい。

【0024】

図4は、携帯コンピューティングデバイス、スマートホン等用の電子デバイスプラットフォーム402の例を示す。図示された部分には、1つ以上の処理コア404、グラフィックスプロセッサ(GPX)406、メモリコントローラハブ(MCH)408、I/O部410、および電力管理部416が含まれる。GPX406は、映像コンテンツを供給するべくディスプレイ407とインターフェースしている。MCH408は、プラットフォームに追加的なメモリ(たとえば、揮発性または不揮発性)を設けるべくメモリ409とインターフェースしている。電力管理部416は、電源(たとえば、バッテリー、アダプタ変換機、VR等)を制御してプラットフォームの様々な部に電力を供給し、かつ、実行可能である場合は、様々な活動状態を管理して電力消費を低減させる。

【0025】

I/O部410は、音声処理部412および周辺インターフェース414を含む。周辺インターフェースによって、通信を行ったり、多様な周辺装置415(キーボード、無線インターフェース、プリンタ等)を有効にしたりするためのインターフェース(たとえば、PCIおよびUSB)が提供される。音声処理部412は、多様な音声入力/出力(アナログおよび/またはデジタル)を受信して、ユーザとの間で音声コンテンツを授受する。音声処理部は、内部のモジュールとも通信して、たとえば、ユーザとネットワーク(たとえば、セル、インターネット等)との間で音声を通信させる。音声処理部412は、プラットフォーム402の機能からの指示を受けて音声処理するための多様なコンポーネント(たとえば、A/D/A変換機、コーデック等)を含む。特に、音声処理部412は、発話処理を実施するための本明細書に記載のSPE413を含む。特に、SPEは、図3に示す電力効率的な構造を有する。

【0026】

これまでの記載では、数多くの特定の詳細事項を述べた。しかし、本発明の実施形態は、これらの特定の詳細事項がなくても実施されることは理解されよう。別の例では、記載の理解が不明瞭とならないように、周知の回路、構造、および技術については詳細に記載しなかった。このことを念頭に置いた場合、「一実施形態」、「ある実施形態」、「例示的な実施形態」、「多様な実施形態」等の言及は、このように言及された本発明の実施形態は、特定の特性、構造、もしくは特徴を含むが、実施形態の全てが必ずしも当該特定の特性、構造、もしくは特徴を含むのではないことを示す。さらに、いくつかの実施形態は、その他の実施形態について記載された特徴のうち、いくつか、もしくは全てを含む場合があり、いずれも含まない場合もある。

【0027】

これまでの記載および以下の特許請求の範囲では、下記の文言は、以下の通りに解釈されるべきである:「連結される(coupled)」および「接続される(connected)」等の文言、およびこれらの変化形が用いられている。これらの文言は、互いに同義語であることは意図されていないと理解されるべきである。むしろ、特定の実施形態では、「接続される(connected)」は、2つ以上の要素が、互いに、物理的もしくは電氣的に直接接触していることを表すべく用いられる。「連結される(coupled)」は、2つ以上の要素が互いに協働もしくは作用するが、物理的もしくは電氣的に直接接触している場合もしていない場合もあることを表すべく用いられる。

【0028】

「PMOSトランジスタ」という文言は、P型金属-酸化物-半導体接合電界効果トランジスタを表す。同様に、「NMOSトランジスタ」という文言は、N型金属-酸化物-半導体接合電界効果トランジスタを表す。特に別途明記される場合、もしくはその文言の

10

20

30

40

50

使用の特質上要請される場合以外は、「MOSトランジスタ」、「NMOSトランジスタ」、もしくは「PMOSトランジスタ」という文言が使用される場合はいつでも、典型例を示す意味で用いられていると理解されるべきである。これらの文言は、例をいくつか挙げると、VT、材料種、絶縁体厚さ、ゲート構成等が異なるデバイスを含む多様なMOSデバイスを包含する。さらに、特段MOS等と呼称されない限り、トランジスタという文言は、たとえば、接合型電界効果トランジスタ、バイポーラ接合トランジスタ、金属半導体FET、多様な種類の三次元トランジスタ、MOS、または、その他の今日知られているもの、もしくは今後開発されるものを含むことができる。

【0029】

本発明は、記載された実施形態に限定されず、添付の特許請求の範囲の趣旨および範囲内で修正もしくは変更しても実施可能である。たとえば、本発明は、あらゆる種類の半導体集積回路（「IC」）チップに適用できる。これらのICチップの例としては、プロセッサ、コントローラ、チップセットコンポーネント、プログラマブルロジックアレイ（PLA）、メモリチップ、ネットワークチップ等が挙げられる。

【0030】

図面のいくつかでは、信号伝達ラインが線で表されていることを理解されたい。いくつかは太く描線されて構成要素としてより重要な信号路を示し、数字表示を有して構成要素としていくつかの信号路を示し、および/または一端もしくは両端に矢印を有して一次情報の流れる方向を示す。しかし、これは、限定的な意味で解釈されるべきでない。むしろ、このような付加的な詳細は、回路のより簡単な理解を促すべく、1つ以上の例示的な実施形態に関連して用いられうる。図示されたいずれの信号線も、付加的な情報を有していようといなかろうと、実際に複数の方向に伝達されうる1つ以上の信号を有してよく、たとえば、差動対、光ファイバライン、および/またはシングルエンドラインにより実装されるデジタルもしくはアナログのライン等の、任意の適切な種類の信号スキームで実装してよい。

【0031】

例示としての大きさ/モデル/値/範囲を提示したが、本発明の実施形態はこれらに限定されないと理解されるべきである。製造技術（たとえば、フォトリソグラフィ）は時とともに発展して行くので、より小型のデバイスが製造されることが予期される。さらに、図面においては、図示および記載を簡略にするべく、また、本発明が曖昧となることを防止するべく、ICチップおよびその他のコンポーネントへの周知の電力/接地接続は図示したり、図示を省略したりした。さらに、構成をブロック図形式で示したのは、本発明が曖昧となるのを回避するためであり、また、これらブロック図構成の実施に係る特定が本発明を実施するプラットフォームに大きく左右されるという事実を考慮してのことであり、つまり、これらの特定は当業者のよく知るところである。本発明の例示的な実施形態を記載するべく特定の詳細（たとえば、回路）が記載されていたとしても、本発明は、これらの特定の詳細を省略もしくは変更しても実施することができることは当業者には明らかであろう。したがって、記載は限定としてではなく例示として見なされるべきである。

本実施形態によれば、以下の各項目もまた開示される。

(項目1)

聴覚的受信者 (auditory receiver) へと提供される第1音声発話経路と、第2音声発話経路とを有する発話処理エンジンと、

前記第2音声発話経路に連結されて、音声発話信号を受信し、発話者の声門情報を含む発話者モデルに少なくとも部分的に基づいて前記音声発話信号からノイズを除去するカルマンフィルターと

を備える装置。

(項目2)

前記フィルターは、環境ノイズ情報を組み込んだノイズモデルにも基づいて、ノイズを除去する項目1に記載の装置。

(項目3)

前記環境ノイズ情報は、リアルタイム情報を含む項目2に記載の装置。

(項目4)

前記リアルタイム情報は、1つ以上のノイズセンサからの情報を含む項目3に記載の装置。

(項目5)

前記発話者モデルには、以前に生成された、ノイズ除去済みの、発話者についての発話信号情報が組み込まれている項目1に記載の装置。

(項目6)

前記フィルターは、フロントエンド部内に実装され、前記発話者モデルは、前記音声発話信号に発話が発検出された場合に有効にされるバックエンド部内に実装される項目1に記載の装置。

(項目7)

前記発話処理エンジンは、発話認識エンジンを含む項目6に記載の装置。

(項目8)

前記発話認識エンジンは、前記バックエンド部の一部である項目7に記載の装置。

(項目9)

聴覚的受信者へと提供される第1音声発話経路と、第2音声発話経路とを含む発話処理エンジンを有する音声処理部と、

前記第2音声発話経路に連結されて、音声発話信号を受信し、発話者の声門情報を含む発話者モデルに少なくとも部分的に基づいて前記音声発話信号からノイズを除去するカルマンフィルターと

を備える電子デバイス。

(項目10)

前記フィルターは、環境ノイズ情報を組み込んだノイズモデルにも基づいて、ノイズを除去する項目9に記載の電子デバイス。

(項目11)

前記環境ノイズ情報は、リアルタイム情報を含む項目10に記載の電子デバイス。

(項目12)

前記リアルタイム情報は、1つ以上のノイズセンサからの情報を含む項目11に記載の電子デバイス。

(項目13)

前記発話者モデルには、以前に生成された、ノイズ除去済みの、発話者についての発話信号情報が組み込まれている項目9に記載の電子デバイス。

(項目14)

前記フィルターは、フロントエンド部内に実装され、前記発話者モデルは、前記音声発話信号に発話が発検出された場合に有効にされるバックエンド部内に実装される項目9に記載の電子デバイス。

(項目15)

前記発話処理エンジンは、発話認識エンジンを含む項目14に記載の電子デバイス。

(項目16)

前記発話認識エンジンは、前記バックエンド部の一部である項目15に記載の電子デバイス。

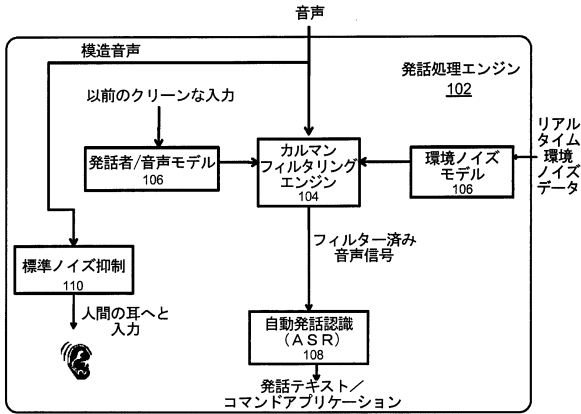
10

20

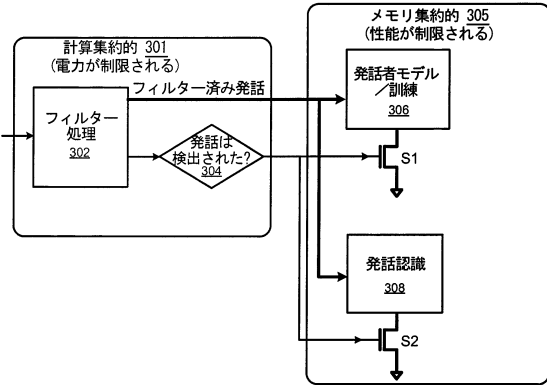
30

40

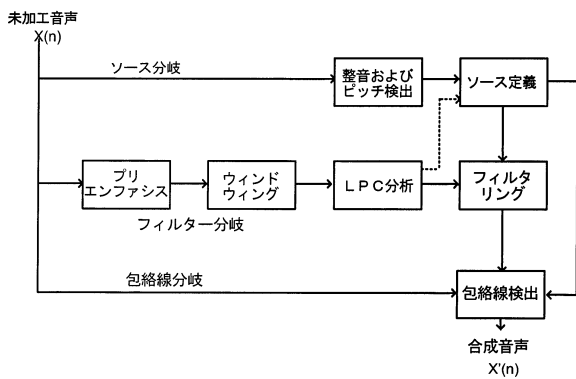
【図1】



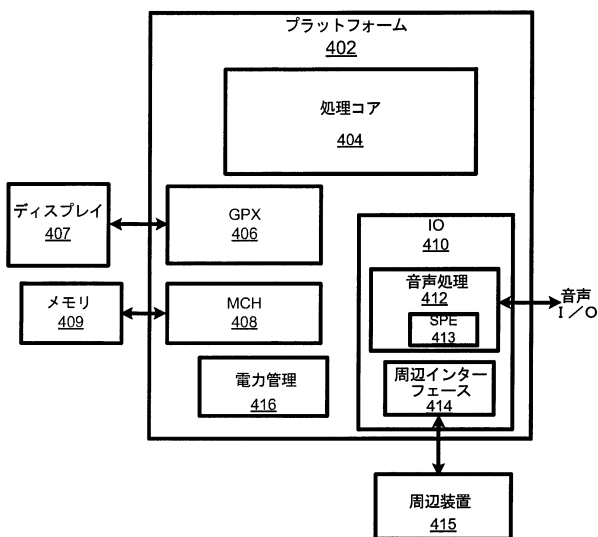
【図3】



【図2】



【図4】



フロントページの続き

- (72)発明者 サナルトウ、マティーアス
アメリカ合衆国 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・22
00 インテル・コーポレーション内
- (72)発明者 ライチョードゥリー、アリジット
アメリカ合衆国 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・22
00 インテル・コーポレーション内
- (72)発明者 ランガラジャン、アナンド ピー .
アメリカ合衆国 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・22
00 インテル・コーポレーション内
- (72)発明者 デイシャー、マイケル イー .
アメリカ合衆国 カリフォルニア州・サンタクララ・ミッション カレッジ ブレーバード・22
00 インテル・コーポレーション内

審査官 千本 潤介

- (56)参考文献 米国特許第05148488 (US, A)
米国特許第05970446 (US, A)
米国特許出願公開第2009/0222263 (US, A1)
国際公開第2009/116291 (WO, A1)
特開2002-006898 (JP, A)
米国特許出願公開第2010/0131269 (US, A1)

(58)調査した分野(Int.Cl., DB名)

G10L 15/20
G10L 21/0208