

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2012-506073  
(P2012-506073A)

(43) 公表日 平成24年3月8日(2012.3.8)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 11/02 (2006.01)	G 1 0 L 11/02	
G 1 0 L 21/02 (2006.01)	G 1 0 L 21/02	1 O 1 A
G 1 0 L 11/00 (2006.01)	G 1 0 L 11/00	4 O 2 L

審査請求 有 予備審査請求 有 (全 31 頁)

(21) 出願番号	特願2011-532248 (P2011-532248)	(71) 出願人	595020643 クアルコム・インコーポレイテッド QUALCOMM INCORPORATED
(86) (22) 出願日	平成21年10月15日 (2009.10.15)		
(85) 翻訳文提出日	平成23年6月15日 (2011.6.15)		
(86) 国際出願番号	PCT/US2009/060828		
(87) 国際公開番号	W02010/045450		アメリカ合衆国、カリフォルニア州 92 121-1714、サン・ディエゴ、モア ハウス・ドライブ 5775
(87) 国際公開日	平成22年4月22日 (2010.4.22)		
(31) 優先権主張番号	61/105,727	(74) 代理人	100108855 弁理士 蔵田 昌俊
(32) 優先日	平成20年10月15日 (2008.10.15)		
(33) 優先権主張国	米国 (US)	(74) 代理人	100159651 弁理士 高倉 成男
(31) 優先権主張番号	12/579,322	(74) 代理人	100091351 弁理士 河野 哲
(32) 優先日	平成21年10月14日 (2009.10.14)		
(33) 優先権主張国	米国 (US)	(74) 代理人	100088683 弁理士 中村 誠

最終頁に続く

(54) 【発明の名称】 オーディオ信号における雑音推定の方法および装置

(57) 【要約】

雑音レベル/スペクトル推定および音声活動検出のためのシステムおよび方法が開示される。ある幾つかの実施形態は、雑音レベルを推定し、引き続いて音声の存在を検出するための確率モデルを含む。これらの実施形態は、標準的音声活動検出子(VAD)より性能が優れており、種々の雑音の多い環境において改善された検出を実現する。

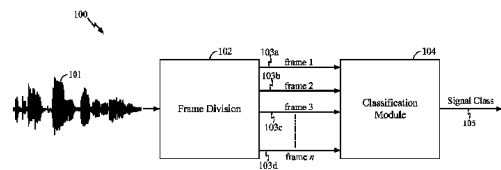


FIG. 1

## 【特許請求の範囲】

## 【請求項 1】

オーディオ信号の現在フレームにおける雑音レベルを推定する方法であって、  
 複数のオーディオフレームの雑音レベルを決定することと、  
 前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算することと、  
 前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、  
 を備える方法。

## 【請求項 2】

前記平均値からの減算に先立って前記標準偏差をスケール(scaling)することを更に備える請求項 1 に記載の方法。

## 【請求項 3】

複数の雑音レベル推定値の最小値を決定することによって前記現在雑音レベル推定値を決定することを更に備える請求項 1 に記載の方法。

## 【請求項 4】

前記複数のオーディオフレームは約 100 フレームを備える請求項 1 に記載の方法。

## 【請求項 5】

前記雑音レベル推定値を計算することは平滑化係数を使用することを備える請求項 1 に記載の方法。

## 【請求項 6】

前記雑音レベル推定値は音声活動の期間中一定に保持される請求項 5 に記載の方法。

## 【請求項 7】

前記平滑化係数は第 2 の平滑化係数を使用して前記現在フレームにおける音声の確率と 1 との間を補間することによって再帰的に平均化される請求項 5 に記載の方法。

## 【請求項 8】

前記雑音レベル推定値は複数の予め決められた雑音レベルの最小値を備える請求項 1 に記載の方法。

## 【請求項 9】

前記雑音レベルの平均値は前記雑音レベルの予め計算された平均値を現在雑音レベルによって補間することによって推定される請求項 1 に記載の方法。

## 【請求項 10】

前記計算された雑音レベル推定値を所望の信号レベル未満の 12 ~ 24 dB の間に限界設定することを更に備える請求項 1 に記載の方法。

## 【請求項 11】

前記現在フレームを、非雑音セグメントを有するものとして識別することによって音声活動を検出することを更に備える請求項 1 に記載の方法。

## 【請求項 12】

すべての  $\alpha$  が  $[0.2, 1]$  であるとして、音声の確率  $\alpha > \alpha_{th}$  であるときに音声活動が宣言される請求項 11 に記載の方法。

## 【請求項 13】

複数のオーディオフレームの雑音レベルを決定するように構成された第 1 のモジュールと、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算するように構成された第 2 のモジュールと、

現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算するように構成された第 3 のモジュールと、  
 を備える雑音決定システム。

## 【請求項 14】

前記第 3 のモジュールは前記平均値からの減算に先立って前記標準偏差をスケール(sca

10

20

30

40

50

le)するように構成される請求項 1 3 に記載の雑音決定システム。

【請求項 1 5】

前記雑音レベル推定値を計算することは平滑化係数を使用することを備える請求項 1 3 に記載の雑音決定システム。

【請求項 1 6】

前記雑音レベル推定値は音声活動の期間中一定に保持される請求項 1 5 に記載の雑音決定システム。

【請求項 1 7】

前記平滑化係数は第 2 の平滑化係数を使用して現在フレームにおける音声の確率と 1 という値との間を補間することによって再帰的に平均化される請求項 1 5 に記載の雑音決定システム。

10

【請求項 1 8】

信号の複数の時間・周波数ビンにおける前記信号の雑音レベルを推定する方法であって、前記信号の前記ビンの各々に関して、

複数のオーディオフレームの雑音レベルを決定することと、

前記時間・周波数ビンにおける雑音レベルを推定することと、

前記時間・周波数ビンにおける仮の雑音レベルを決定することと、

前記仮の雑音レベルから前記時間・周波数ビンにおける 2 次雑音レベルを決定することと、

前記時間・周波数ビンにおける 2 次雑音レベルから限界設定された雑音レベルを決定することと、

20

を備える方法。

【請求項 1 9】

前記限界設定された雑音レベルを決定することはアクティブな所望信号レベル未満の 1 2 ~ 2 4 d B の間で前記推定雑音レベルを限界設定することを備える請求項 1 8 に記載の方法。

【請求項 2 0】

現在フレームにおける各周波数に関する音声の確率の重み付け合計を取ることによって前記現在フレームにおける音声の確率を計算することを更に備える請求項 1 8 に記載の方法。

30

【請求項 2 1】

6 0 0 H z ~ 1 5 5 0 H z の範囲内の重み付けが少なくとも 0 . 0 2 という値を与えられる請求項 2 0 に記載の方法。

【請求項 2 2】

オーディオ信号の現在フレームにおける雑音レベルを推定するシステムであって、

複数のオーディオフレームの雑音レベルを決定する手段と、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算する手段と、

前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算する手段と、

40

を備えるシステム。

【請求項 2 3】

前記現在フレームの雑音レベル推定値を計算する前記手段は前記平均値からの減算に先立って前記標準偏差をスケールする請求項 2 2 に記載の雑音決定システム。

【請求項 2 4】

前記雑音レベルを決定する前記手段は信号のエネルギーレベルを決定するように構成されたモジュールを備える請求項 2 2 に記載のシステム。

【請求項 2 5】

前記雑音レベルの平均値と標準偏差とを計算する前記手段は数学的演算を実行するように構成されたモジュールを備える請求項 2 2 に記載のシステム。

50

**【請求項 26】**

雑音レベル推定値を計算する前記手段は数学的演算を実行するように構成されたモジュールを備える請求項 22 に記載のシステム。

**【請求項 27】**

プロセッサ上で実行されたときに、  
複数のオーディオフレームの雑音レベルを決定することと、  
前記複数のオーディオフレームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、  
現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、  
を備える方法を実行する命令を備えるコンピュータ可読媒体。

10

**【請求項 28】**

前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える請求項 27 に記載の方法。

**【請求項 29】**

複数のオーディオフレームの雑音レベルを決定することと、  
前記複数のオーディオフレームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、  
現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、  
を備える方法を実行するようにプログラムされたプロセッサ。

20

**【請求項 30】**

前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える請求項 29 に記載の方法。

**【発明の詳細な説明】****【関連出願への相互参照】****【0001】**

本出願は、出願の全体が引用によってここに組み込まれている 2008 年 10 月 15 日に出願された米国仮特許出願第 61/105,727 号からの優先権を主張する。

**【技術分野】**

30

**【0002】**

本開示は、概して雑音レベル/スペクトル推定および音声活動検出のための方法および装置に関し、また特に雑音レベルを推定し音声の存在を検出する確率モデルの使用に関する。

**【背景技術】****【0003】**

通信技術は、しばしば新しい課題を提示しながら、多くの分野で発展し続けている。携帯電話および無線ヘッドセットの出現によって現在、人々は極めて耳障りな環境すなわち低い信号対雑音比(SNR)を有する環境において真の全二重通信会話を行うことができる。信号増強および雑音抑制は、これらの状況において極めて重要になる。所望される音声の了解度は、他端部における聞き手に信号を送信する前に不要な雑音の多い信号を抑制することによって改善される。雑音の多い背景(バックグラウンド)内で音声の存在を検出することは、信号増強および雑音抑制の 1 つの重要な構成要素である。改善された音声検出を達成するために、ある幾つかのシステムは、着信信号を複数の異なる時間/周波数フレームに分割し、各フレームにおける音声の存在の確率を推定する。

40

**【0004】**

音声の存在を検出する際の最も大きな課題の 1 つは、ノイズフロア(noise floor)を追跡すること、特に単一のマイクロホン/センサーを使用して非定常的雑音レベルを追跡することである。音声活動検出は、最近の通信デバイスにおいて、特に携帯電話および無線ハンドセット・デバイスといった低い信号対雑音比の下で動作する最近の移動デバイスで

50

広く使用されている。これらのデバイスの大部分において信号増強および雑音抑制は、他端部における聞き手に雑音の多い信号を送信する前にこの雑音の多い信号に対して実行される。これは所望の音声の了解度を改善するために行われる。信号増強/雑音抑制において音声またはボイス活動検出子(VAD)は、雑音で汚染された信号における所望の音声の存在を検出するために使用される。この検出子は、音声の存在または不在の2値的決定を生成でき、あるいはまた音声存在の確率を生成することもできる。

#### 【0005】

音声の存在を検出する際の1つの課題は、ノイズ「シーリング」および「フロア」としても知られる信号内の背景雑音のレベルの上限および下限を決定することである。これは特に、単一のマイクロホン入力を使用する非定常雑音について真である。更にデバイスの、またはデバイスを使用する人の物理的移動に起因する雑音レベルの急速な変化を追跡することは、なおさら困難である。

#### 【発明の概要】

#### 【0006】

ある幾つかの実施形態においてオーディオ信号の現在フレームにおける雑音レベルを推定する方法が開示される。この方法は、複数のオーディオフレームの雑音レベルを決定すること、ならびにこれら複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算することを備える。現在フレームの雑音レベル推定は、平均値から減算される標準偏差の値を使用して計算される。

#### 【0007】

ある幾つかの実施形態において、雑音決定システムが開示される。このシステムは、複数のオーディオフレームの雑音レベルを決定するように構成された1つのモジュールと、これら複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算するように構成された1つ以上のモジュールと、を備える。このシステムはまた、現在フレームの雑音レベル推定値を上記平均値から減算される標準偏差の値として計算するように構成されたモジュールも含み得る。

#### 【0008】

ある幾つかの実施形態において、複数の時間・周波数ビン(bin)における信号の雑音レベルを推定する方法であって、1つ以上のコンピュータシステム上で実現され得る方法が開示される。信号の各ビンに関してこの方法は、複数のオーディオフレームの雑音レベルを決定し、時間・周波数ビンにおける雑音レベルを推定し、時間・周波数ビンにおける仮の雑音レベルを決定し、この仮の雑音レベルから時間・周波数ビンにおける2次雑音レベルを決定し、そして限界設定された雑音レベルを時間・周波数ビンにおける2次雑音レベルから決定する。

#### 【0009】

ある幾つかの実施形態は、オーディオ信号の現在フレームにおける雑音レベルを推定するシステムを開示する。このシステムは、複数のオーディオフレームの雑音レベルを決定する手段と、これら複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算する手段と、現在フレームの雑音レベル推定値を上記平均値から減算される標準偏差の値として計算する手段と、を備え得る。

#### 【0010】

ある幾つかの実施形態において、ある方法を実行するためにプロセッサ上で実行される命令を備えるコンピュータ可読媒体が開示される。この方法は、複数のオーディオフレームの雑音レベルを決定することと、これら複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算することと、現在フレームの雑音レベル推定値を上記平均値から減算される標準偏差の値として計算することと、を備える。

#### 【図面の簡単な説明】

#### 【0011】

下記の添付図面において種々の構成が限定としてではなく例として図示される。

【図1】本発明の原理によるVADの単純化されたブロック図である。

10

20

30

40

50

【図2】周波数ドメインVADのための周波数選択性重み付けベクトルを示すグラフである。

【図3】ピンク雑音環境下における提案された時間ドメインVADの性能を示すグラフである。

【図4】バブル雑音環境下における提案された時間ドメインVADの性能を示すグラフである。

【図5】トラヒック雑音環境下における提案された時間ドメインVADの性能を示すグラフである。

【図6】パーティ雑音環境下における提案された時間ドメインVADの性能を示すグラフである。

10

【発明を実施するための形態】

【0012】

本実施形態は、信号における雑音レベルを決定し、またある幾つかの事例で、引き続き音声を検出する方法およびシステムを備える。これらの実施形態は、従来技術に対して多数の重要な進歩を備える。1つの改善点は、従前および現在のオーディオフレームからの背景雑音の平均値に基づいて音声信号における背景雑音の推定を実行することに関する。これは、より前および現在のオーディオフレームからの最小雑音値に基づいて音声のフレームに関する現在の背景雑音レベルを計算した他のシステムとは異なる。伝統的に研究者等は、現在雑音レベルを推定するために前の雑音値の最小値を見てきた。しかしながら一実施形態では、推定された雑音信号レベルが幾つかの過去のフレームから計算され、最小値よりむしろこの集合体(アンサンブル(ensemble))の平均値が計算され、調整された(scaled)標準偏差がこの集合体から減算される。結果として得られた値は好都合にも、典型的に集合体の最小値を使用して与えられるより、現在オーディオフレームの雑音レベルの正確な推定を与える。

20

【0013】

更にこの推定された雑音レベルは、雑音のより正確な推定を維持するように着信信号レベルに基づいて動的に限定され得る。この推定された雑音レベルは、不連続性を最小にするために前の値によって更に「平滑化」または「平均化」され得る。推定雑音レベルはそれから、この雑音レベルより高いエネルギーレベルを有するフレーム内の音声を識別するために使用され得る。これは、音声の存在の較正された確率を生成するために非線形シグモイド活性化関数によって使用され得る帰納的(a posteriori)信号対雑音比(SNR)を計算することによって決定され得る。

30

【0014】

図1を参照すると従来の音声活動検出(VAD)システム100は、背景雑音を有するセグメントと背景雑音および音声の両者を有するセグメントとを備える着信信号101を受信する。このVADシステム100は、時間信号101をフレーム103a~103dに分割する。それからこれらのフレーム103a~103dの各々は、(雑音または音声)に所定のフレームを配置すべきクラスを決定する類別モジュール104に渡される。

【0015】

類別モジュール104は、所定の信号のエネルギーを計算して、このエネルギーをノイズフロアの推定値に対応する時間的に変化する閾値と比較する。このノイズフロア推定値は、各着信フレームによって更新され得る。ある幾つかの実施形態ではフレームは、フレーム信号の推定エネルギーレベルがこの特定のフレーム内の測定されたノイズフロアより高い場合に音声活動として類別される。したがってこのモジュールでは雑音スペクトル推定は、音声認識の、そして所望であればこれに続く改善の基本的構成要素である。特に、低SNRの、および非定常雑音の環境下におけるこのようなシステムのローバスト(堅牢)性は、雑音統計における急速な変化を信頼度高く追跡するための能力によって最大限に影響される。

40

【0016】

VADに基づく従来の雑音推定方法は、雑音推定値の更新を音声不在の期間に限定する

50

。しかしながらこれらのVADの信頼度は、弱い音声成分と低入力SNRとのために大幅に劣化する。パワースペクトル密度ヒストグラムに基づく他の技法は、コンピュータ的に高価であり、高価なメモリ資源を必要とし、低SNR条件下でうまく働かず、したがって携帯電話およびブルートゥースヘッドセット・アプリケーションのためには適さない。最小統計は、過去の複数のフレームの最小値を雑音推定値であると見なすことによって動作する雑音スペクトル推定のために使用されるもう1つの方法である。不都合なことにこの方法は、定常雑音に対しては有利に働くが、非定常的環境を取り扱うときには極めて不利になる。

【0017】

一実施形態は、「パーティ雑音」または「バブル雑音」といった極めて非定常的雑音の環境を含む多種類の望ましくないオーディオ信号を追跡する際に極めて有効である雑音スペクトル推定システムおよび方法を備える。このシステムは、このような推定に助けとならない環境においても正確なノイズフロアを生成する。この推定されたノイズフロアは、音声の存在の確率を決定するためにシグモイド関数「ロジスティック関数」において使用される帰納的(a posteriori)SNRを計算する際に使用される。ある幾つかの実施形態では、この関数のために音声決定モジュールが使用される。

10

【0018】

$x[n]$ および $d[n]$ がそれぞれ所望の音声および無相関の付加的雑音信号を表すとする。観測された信号または汚染された信号 $y[n]$ は、単に(数式1)によって与えられるこれらの追加分である：

20

【数1】

$$y[n] = x[n] + d[n] \quad (1)$$

【0019】

2つの仮定 $H_0[n]$ および $H_1[n]$ はそれぞれ、 $n$ 番目の( $n^{\text{th}}$ )時間フレームにおける音声の不在および存在を示す。ある幾つかの実施形態では雑音の多い測定の過去のエネルギーレベル値は、音声不在の期間中に再帰的に平均化され得る。これとは対照的に推定値は音声存在の期間中、一定に保持され得る。具体的には、

30

【数2】

$$H_0[n]: \lambda_d[n] = \alpha_d \lambda_d[n-1] + (1 - \alpha_d) \sigma_y^2[n] \quad (2)$$

【数3】

$$H_1[n]: \lambda_d[n] = \lambda_d[n-1] \quad (3)$$

40

【0020】

ここで

【数 4】

$$\sigma_y^2[n] = \sum_{i=n-100}^n |y[i]|^2$$

【0021】

は時間フレーム  $n$  における雑音の多い信号のエネルギーであり、 $\alpha_d$  は 0 と 1 の間の平滑化パラメータを示す。しかしながら音声が存在するときは必ずしも明らかでないので、方法  $H_0$  または  $H_1$  の各々を適用すべきときは明らかでない可能性がある。その代わりに時間の経過と共に平滑化係数  $\alpha_s$  を更新することによって再帰的平均を推定する「条件付き音声存在確率」を使用することができる：

【数 5】

$$\lambda_d[n] = \alpha_s[n] \lambda_d[n-1] + (1 - \alpha_s[n]) \sigma_y^2[n] \quad (4)$$

20

【0022】

ここで

【数 6】

$$\alpha_s[n] = \alpha_d + (1 - \alpha_d) \text{prob}[n] \quad (5)$$

【0023】

この仕方で、音声の存在が分からないときに、より正確な推定値が取得され得る。

【0024】

他には、雑音レベル推定のために前に考えられた最小統計に基づく方法がある。例えば過去の 100 フレームに関して推定された雑音の多い信号レベル  $\lambda_d$  を見て、アンサンブルの最小値を計算し、これを推定雑音レベルとして宣言することができる。すなわち：

【数 7】

$$\hat{\sigma}_n^2[n] = \min[\lambda_d(n-100:n)] \quad (6)$$

40

【0025】

ここで  $\min[x]$  はベクトル  $x$  のエントリーの最小値を示し、

【数 8】

$$\hat{\sigma}_n^2[n]$$

50

## 【 0 0 2 6 】

は時間フレーム  $n$  における推定雑音レベルである。100 より多い、または少ないフレームに関して演算を実行することができるが、100 はここおよび本明細書全体を通して単なる例示的範囲として提示されている。このアプローチは定常的雑音環境には有効に働くが、非定常的環境ではうまく行かない。

## 【 0 0 2 7 】

中でもこの問題に取り組むために本実施形態は、システムの全体的検出効率を改善するために以下に説明される技法を使用する。

## 【 0 0 2 8 】

[平均統計値]

一実施形態では本発明のシステムおよび方法は、ノイズフロアを計算するために最小統計値よりむしろ平均統計値を使用する。具体的に信号エネルギー  $\sigma_1^2$  は、過去のフレーム値のスケール(scaled)標準偏差  $\mu$  から減算することによって計算される。それから現在のエネルギーレベル  $\sigma_2^2$  は過去のフレームから前に計算されたすべての信号エネルギー  $\sigma_1^2$  の最小値として選択される：

## 【数 9】

$$\hat{\sigma}_1^2[n] = [\bar{\lambda}_d[n-100:n] - \alpha * \sigma(\lambda_d[n-100:n])] \quad (7)$$

10

20

## 【数 10】

$$\hat{\sigma}_2^2[n] = \min(\hat{\sigma}_1^2[n-100:n]) \quad (8)$$

## 【 0 0 2 9 】

ここで  $\mu$  はベクトル  $x$  のエントリーの平均値である。本実施形態は、過去の100フレームにわたる推定雑音レベルのスケール(scaled)標準偏差を同数のフレームにわたる推定雑音レベルの平均値から減算することを考えている。

## 【 0 0 3 0 】

[雑音推定値を使用する音声検出]

いったん雑音推定値  $\sigma_1^2$  が計算されると、音声は高いSNRの領域を識別することによって推測され得る。特にロジスティック回帰ベースの類別子に基づいて音声の存在の較正された確率を正確に推定する数学的モデルが開発され得る。ある幾つかの実施形態では特徴要素ベースの類別子が使用され得る。音声の短期間スペクトルは対数(log)分布によってうまくモデル化されるので、特徴要素の集合としてSNR自身よりむしろ推定された帰納的SNRの対数を使用できる。すなわち

## 【数 11】

$$\chi[n] = 10 \left\{ \log_{10} \left( \sum_{i=n-100}^n |y[i]|^2 \right) - \log_{10} (\sigma_{noise}^2[n]) \right\} \quad (9)$$

30

40

## 【 0 0 3 1 】

安定のために上記の量の時間的平滑化をすることもできる。

50

【数 1 2】

$$\hat{\chi}[n] = \beta_1 \hat{\chi}[n-1] + (1 - \beta_1) \chi[n]$$

$$\beta_1 \in [0.75, 0.85]$$
(10)

【0032】

それから所望の音声検出のために、ロジスティック関数として知られる非線形・メモリなし活性化関数を使用され得る。時間フレーム  $n$  における音声の存在の確率は下記によって与えられえる。

10

【数 1 3】

$$prob[n] = \frac{1}{1 + \exp(-\hat{\chi}[n])}$$
(11)

【0033】

所望であれば推定確率  $prob[n]$  はまた、音声におけるサドンバースト (sudden burst) を追跡するために小さな忘却因子 (forgetting factor) を使用して時間的平滑化されることも可能である。音声の不在および存在の 2 値的決定を取得するために推定確率 ( $prob[0, 1]$ ) は、事前選択された閾値と比較され得る。  $prob$  の高い値は音声の存在の、より高い確率を示す。例えばもし  $prob[n] > 0.7$  であれば、時間フレーム  $n$  における音声の存在が宣言され得る。そうでなければこのフレームは単に非音声活動だけを有すると考えられ得る。提案された実施形態は、より正確な雑音レベル決定の結果として、より正確な音声検出を実現する。

20

【0034】

[雑音推定の改善]

平均値および標準偏差の計算は、過去のフレーム推定値を記憶するために十分なメモリを必要とする。この要件は、限定されたメモリを有するある幾つかのアプリケーション/デバイス (ある幾つかの極めて小さな携帯型デバイスといった) に関して禁止的であり得る。このような場合、上記の計算に取って代わるために下記の近似式が使用され得る。平均推定値の近似値は、平滑定数  $\alpha_M$  を用いてパワー推定値  $x(n)$  を指数関数的に平均化することによって計算され得る。同様に分散推定値の近似値は、  $n$  がフレームインデックスを表すとして平滑化定数  $\alpha_V$  を用いてパワー推定値の平方を指数関数的に平均化することによって計算され得る：

30

【数 1 4】

$$\hat{x}(n) = \alpha_M \hat{x}(n-1) + (1 - \alpha_M) x(n)$$
(12)

40

【数 1 5】

$$\hat{v}(n) = \alpha_V \hat{v}(n-1) + (1 - \alpha_V) x^2(n)$$
(13)

50

## 【 0 0 3 5 】

代替として標準偏差推定値の近似値は、分散推定値

## 【 数 1 6 】

$$\hat{v}(n)$$

## 【 0 0 3 6 】

の平方根を取ることによって取得され得る。平滑化定数  $\alpha_m$  &  $\alpha_v$  は、20 ~ 100 フレームにわたる平均化に対応するように範囲 [0.95, 0.99] 内で選択され得る。更に

## 【 数 1 7 】

$$\hat{\sigma}_1^2[n]$$

## 【 0 0 3 7 】

の近似値は、平均およびスケール(scaled)標準偏差推定値の間の差異を計算することによって取得され得る。いったん平均マイナススケールされた標準偏差推定値(mean-minus-scaled standard deviation)が取得されると、1セットの例えば100フレームにわたる差異に対する最小統計が実行され得る。

## 【 0 0 3 8 】

この特徴要素は単に、最小統計値と比較して非定常的雑音のピークの優れた追跡を与える。ある幾つかの実施形態では雑音レベル推定に影響を与える所望の音声ピークを補償するために、雑音レベルの標準偏差が減算される。しかしながら方程式7における過剰な減算は、過小評価雑音レベルという結果をもたらす可能性がある。この問題に取り組むために音声不在時における長期間平均が実行され得る。すなわち：

## 【 数 1 8 】

$$H_0[n]: \lambda_{d_i}[n] = \alpha_i \lambda_{d_i}[n-1] + (1 - \alpha_i) \sigma_y^2[n] \quad (14)$$

## 【 数 1 9 】

$$H_1[n]: \lambda_{d_i}[n] = \lambda_{d_i}[n-1] \quad (15)$$

## 【 0 0 3 9 】

ここで  $\alpha_1 = 0.9999$  は平滑化係数であり、雑音レベルは下記のように推定される：

【数 2 0】

$$\hat{\sigma}_n^2[n] = \max(\hat{\sigma}_2^2[n], \lambda_{d_i}[n]) \quad (16)$$

【0040】

(雑音の限界設定)

典型的には着信信号が極めて清浄(高いSNR)であるとき、雑音レベルは典型的に過小評価される。この問題を解決する1つの方法は、雑音レベルを所望信号レベル<sup>2</sup><sub>desired</sub>未満の例えば少なくとも18dBになるように低く限界設定することである。低く限界設定することは下記のフロアリング演算を使用して達成され得る：

【数 2 1】

$$\sigma_{desired}^2[n] = \alpha_2 \sigma_{desired}^2[n-1] + (1-\alpha_2) \sum_{i=n-100}^n |y[n]|^2 \quad (17)$$

$$SNR\_diff[n] = SNR\_estimate[n] - Longterm\_Avg\_SNR[n] \quad 20$$

【0041】

もし

【数 2 2】

$$\text{If } \sum_{i=n-100}^n |y[n]|^2 > \Delta_1$$

30

【0042】

であって

もし<sup>2</sup><sub>noise</sub>[n-1] > <sup>2</sup><sub>desired</sub>であればfloor<sub>1</sub>[n] = <sup>2</sup><sub>desired</sub>[n] / <sup>3</sup>であり、もしfloor[n-1] < floor<sub>1</sub>[n]であればfloor[n] = floor<sub>1</sub>[n]であり、そうでなく、もしSNR\_diff[n-1] > <sup>4</sup>であってもし<sup>2</sup><sub>noise</sub>[n-1] < <sup>5</sup>であればfloor[n] = floor<sub>1</sub>[n]である

40

終了

終了

終了

終了

## 【数 2 3】

$$\sigma_{noise}^2[n] = \max(\hat{\sigma}_n^2[n], floor[n])$$

## 【0043】

ここで因数  $\alpha_1 \sim \alpha_5$  は調整可能であり、SNR\_Estimate および Longterm\_Avg\_SNR はそれぞれ雑音推定値  $\sigma_{noise}^2[n]$  および  $\lambda_{d1}[n]$  を使用して取得される帰納的 SNR 推定値および長期間 SNR 推定値である。この仕方で雑音レベルは必要とされるアクティブな所望信号レベル未満の 12 ~ 24 dB の間に限界設定され得る。

10

## 【0044】

[周波数ベースの雑音推定]

実施形態は更に、他で使用され得る周波数ドメイン・サブバンド・ベースのコンピュータ関連音声検出子を含む。ここで各時間フレームは、時間フレームのフーリエ変換で表される 1 群の成分周波数に分割される。これらの周波数は、「時間・周波数」ビン内のそれぞれのフレーム関連付けられた状態に留まる。この説明されている実施形態はそれから、各時間・周波数ビン ( $k, n$ ) における、すなわち  $k$  番目の周波数ビンおよび  $n$  番目の時間フレームにおける音声の存在の確率を推定する。ある幾つかのアプリケーションは、音声存在の確率が時間・周波数アトムレベルおよび時間フレームレベルの両者において推定されることを必要とする。

20

## 【0045】

各時間・周波数ビン (bin) における音声検出子の動作は、各周波数ビンにおいて実行されることを除いて上記に説明された時間ドメイン実現形態に類似している。特に各時間・周波数ビン ( $k, n$ ) における雑音レベル  $\lambda_d$  は、平滑化係数  $\alpha_s$  を使用して過去のフレーム  $\lambda_d[k, n-1]$  における雑音レベルとこの周波数における過去の 100 フレームに関する信号エネルギー

30

## 【数 2 4】

$$\sum_{i=n-100}^n |Y(k, i)|^2$$

## 【0046】

との間を補間することによって推定される：

## 【数 2 5】

$$\lambda_d[k, n] = \alpha_s[k, n] \lambda_d[k, n-1] + (1 - \alpha_s[k, n]) \sum_{i=n-100}^n |Y(k, i)|^2 \quad (18)$$

40

## 【0047】

平滑化係数  $\alpha_s$  はそれ自身、音声の存在確率と 1 との間の補間 (すなわち音声が存在することがどれほどの頻度で想定され得るか) に依存している可能性がある。

## 【0048】

50

Error! Objects cannot be created from editing field codes.(19)

(誤り! オブジェクトは編集フィールドコードから作成できない)。

【0049】

上記の方程式において  $Y(k, i)$  は、 $k$  番目の周波数ビンおよび  $i$  番目の時間フレームにおける汚染された信号である。各ビンにおける仮の雑音レベルは下記のように推定され得る：

【数26】

$$\hat{\sigma}_1^2[k, n] = \left[ \bar{\lambda}_d[k, n-100:n] - \sigma(\lambda_d[k, n-100:n]) \right] \quad (20)$$

10

【数27】

$$\hat{\sigma}_2^2[k, n] = \min(\hat{\sigma}_1^2[k, n-100:n]) \quad (21)$$

20

【0050】

時間ドメインVADと同様に、音声存在  $H_0$  および不在  $H_1$  の期間中の長期間平均は下記の方程式にしたがって実行され得る：

【数28】

$$H_0[k, n]: \lambda_{d_1}[k, n] = \alpha_l \lambda_{d_1}[k, n-1] + (1 - \alpha_l) \sum_{i=n-100}^n |Y(k, i)|^2 \quad (22)$$

30

【数29】

$$H_1[k, n]: \lambda_{d_1}[k, n] = \lambda_{d_1}[k, n-1] \quad (23)$$

【0051】

それから各時間・周波数ビンにおける2次雑音レベルは下記のように推定される：

【数30】

$$\hat{\sigma}_n^2[k, n] = \max(\hat{\sigma}_2^2[k, n], \lambda_{d_1}[k, n]) \quad (24)$$

40

【0052】

ある幾つかの高SNRビンに関する雑音レベルにおける過小評価の問題に取り組むために、下記の限界設定条件および方程式が使用され得る：

【数 3 1】

$$\sigma_{desired}^2[k, n] = \alpha_2 \sigma_{desired}^2[k, n-1] + (1 - \alpha_2) \sum_{i=n-100}^n |y[k, n]|^2 \quad (25)$$

$$SNR\_diff[k, n] = SNR\_estimate[k, n] - Longterm\_Avg\_SNR[k, n]$$

【0053】

10

もし

【数 3 2】

$$\text{If } \sum_{i=n-100}^n |y[k, n]|^2 > \Delta_1$$

【0054】

20

であって

もし  $\sigma_{noise}^2[n-1] > \sigma_{desired}^2$  であれば
 $floor_1[k, n] = \sigma_{desired}^2[k, n] / \sigma_{noise}^2$  であり、
もし  $floor[k, n-1] < floor_1[k, n]$  であれば
 $floor[k, n] = floor_1[k, n]$  であり、
そうでなく、もし  $SNR\_diff[k, n-1] > \Delta_4$  であってもし  $\sigma_{noise}^2[k, n-1] < \sigma_{desired}^2$  であれば
 $floor[k, n] = floor_1[k, n]$  である

終了

終了

終了

30

終了

【数 3 3】

$$\sigma_{noise}^2[k, n] = \max(\hat{\sigma}_n^2[k, n], floor[k, n])$$

【0055】

40

ここで因数  $\alpha_1 \sim \alpha_5$  は調整可能であり、 $SNR\_Estimate$  および  $Longterm\_Avg\_SNR$  はそれぞれ雑音推定値  $\sigma_{noise}^2[k, n]$  および  $\sigma_{desired}^2[k, n]$  を使用して取得される帰納的  $SNR$  推定値および長期間  $SNR$  推定値である。 $\sigma_{noise}^2(k, n)$  は各時間・周波数ビンにおける最終的雑音レベルを表す。

【0056】

次に上記に説明された時間ドメイン数学モデルに基づく方程式(方程式 2 ~ 17)は、各時間・周波数ビンにおける音声の存在の確率を推定するために使用され得る。特に各時間・周波数アトムにおける帰納的  $SNR$  は下記によって与えられる：

【数 3 4】

$$\chi[k, n] = 10 \left\{ \log_{10} \left( \sum_{i=n-100}^n |Y[k, i]|^2 \right) - \log_{10} \left( \sigma_{noise}^2[k, n] \right) \right\} \quad (26)$$

【0057】

安定のために上記の量の時間的平滑化をすることもできる：

【数 3 5】

10

$$\begin{aligned} \hat{\chi}[k, n] &= \beta_1 \hat{\chi}[k, n-1] + (1 - \beta_1) \chi[k, n] \\ \beta_1 &\in [0.75, 0.85] \end{aligned} \quad (27)$$

【0058】

そして各時間・周波数アトムにおける音声の存在の確率は

【数 3 6】

20

$$prob[k, n] = \frac{1}{1 + \exp(-\hat{\chi}[k, n])} \quad (28)$$

【0059】

によって与えられる。

【0060】

ここで  $prob[k, n]$  は  $k$  番目の周波数ビンと  $n$  番目の時間フレームとにおける音声の存在の確率を表す。

30

【0061】

[2レベル・アーキテクチャ]

上記の数学モデルは、各時間フレームにおける音声存在の確率の改善された推定値を取得するために任意選択的に各時間・周波数ビンにおける出力確率を柔軟に組み合わせることを可能にする。一実施形態は例えば、検出子の第1のレベルが時間・周波数ビンレベルで動作し、出力が第2の時間フレームレベル音声検出子に入力される2レベル・アーキテクチャを考えている。

【0062】

2レベル・アーキテクチャは、各時間フレームにおける音声の存在の確率のより良好な推定値を取得するために各時間・周波数ビンにおける推定された確率を組み合わせる。このアプローチは、音声がある周波数帯域(600 Hz ~ 1550 Hz)において支配的であるという事実を利用し得る。図2は、ある幾つかの実施形態で使用された複数の周波数重み付けのプロット203を示す。ある幾つかの実施形態ではこれらの重み付けは、下記に示されたようなビンレベルの確率の重み付け平均値を決定するために使用される：

40

【数 3 7】

$$\begin{aligned}
 \text{prob}[n] &= \sum_{i=1}^N W_i \left( \frac{1}{1 + \exp(-\hat{\chi}[i, n])} \right) \\
 \sum_{i=1}^N W_i &= 1
 \end{aligned}
 \tag{29}$$

10

【0063】

ここで重み付けベクトル  $W$  は、図 2 に示された値を備える。最後に、各フレームにおける音声の存在または不在の 2 値的決定は、時間ドメイン・アプローチと同様に、推定された確率を事前選択された閾値と比較することによって行われ得る。

【0064】

[例]

上記の実施形態の利点を評価するために音声検出は、上記に説明された時間および周波数実施形態ならびに 2 つの先行する VAD システムを使用して実行された。種々の雑音環境下におけるこれらの実例の各々に関する ROC カーブは、図 3 ~ 6 に示されている。上記の実施形態の時間および周波数バージョンの各々は、標準的 VAD よりかなり良好に機能した。これらの例の各々に関して使用された雑音データベースは、勧告された規格 ETSI E G 2 0 2 3 9 6 - 1 に基づいていた。このデータベースは、音声品質および雑音抑制評価目的のために車両雑音、街路雑音、バブル雑音などの標準的記録を提供する。VAD 性能を評価するために更なる現実世界の記録も使用された。これらの雑音環境は、定常的雑音および非定常的雑音の両者を含んでおり、テストすべき挑戦的な集大成 (コーパス (corpus)) を提供している。更に、検出を極めて困難にするために 5 dB の SNR が選択された (典型的なオフィス雑音はおよそ 30 dB であろう)。

20

【0065】

[例 1]

提案された時間ドメイン音声検出子を評価するために、種々の雑音環境下で 5 dB の SNR における受信機動作特性 (ROC) がプロットされている。図 2 に示されているように ROC カーブは、検出の確率 (音声が存在するときに音声の存在を検出する) 301 対誤警報の確率 (音声が存在しないときに音声の存在を宣言する) 302 をプロットしている。適切な検出率で極めて低い誤警報を有することが望ましい。所定の誤警報に関するより高い検出確率の値はより良好な性能を示すので、一般により高いカーブはより良好な検出子である。

30

【0066】

ROC は、4 つの異なる雑音 - ピンク雑音、バブル雑音、トラヒック雑音およびパーティ雑音に関して示されている。ピンク雑音は、周波数に反比例するパワースペクトル密度を有する定常雑音である。これは、一般に自然の物理システムにおいて観測され、しばしばオーディオ信号処理ソリューション (解決策) をテストするために使用される。バブル雑音およびトラヒック雑音は事実上、準定常的であって、一般に移動体通信環境において遭遇する雑音源である。バブル雑音信号およびトラヒック雑音信号は、ETSI E G 2 0 2 3 9 6 - 1 規格勧告によって与えられる雑音データベースにおいて利用可能である。パーティ雑音は、高度に非定常的な雑音であって、VAD の性能を評価するための極端な場合の例として使用される。大抵の単一マイクロホン音声活動検出子は、雑音の高度に非定常的な性質のためにパーティ雑音の存在において高い誤警報を発生させる。しかしながら本発明において提案された方法は、パーティ雑音の場合でも低い誤警報を発生させる。

40

【0067】

図 3 は、第 1 の標準的 VAD の ROC カーブ 303 c と第 2 の標準的 VAD の ROC カ

50

ーブ303bと本時間ベース実施形態のROCカーブ303aと本周波数ベース実施形態のROCカーブ303dとを示しており、これらはピンク雑音環境においてプロットされている。図示のように本実施形態303a、303dは、第1のVAD303bおよび第2のVAD303cの各々より大幅に優れた性能を示しており、誤警報制約302が緩和されたという理由から常により高い検出301を記録した。

【0068】

[例2]

図4は、第1の標準的VADのROCカーブ403cと第2の標準的VADのROCカーブ403bと本時間ベース実施形態のROCカーブ403aと本周波数ベース実施形態のROCカーブ403dとを示しており、これらはバブル雑音環境においてプロットされている。図示のように本実施形態403a、403dは、第1のVAD403bおよび第2のVAD403cの各々より大幅に優れた性能を示しており、誤警報制約402が緩和されたという理由から常により高い検出401を記録した。

10

【0069】

[例3]

図5は、第1の標準的VADのROCカーブ503cと第2の標準的VADのROCカーブ503bと本時間ベース実施形態のROCカーブ503aと本周波数ベース実施形態のROCカーブ503dとを示しており、これらはトラヒック雑音環境においてプロットされている。図示のように本実施形態503a、503dは、第1のVAD503bおよび第2のVAD503cの各々より大幅に優れた性能を示しており、誤警報制約502が緩和されたという理由から常により高い検出501を記録した。

20

【0070】

[例4]

図6は、第1の標準的VADのROCカーブ603cと第2の標準的VADのROCカーブ603bと本時間ベース実施形態のROCカーブ603aと本周波数ベース実施形態のROCカーブ603dとを示しており、これらはROC-I-C-A-S-S-Pオーディトリウム(聴衆席)雑音環境においてプロットされている。図示のように本実施形態603a、603dは、第1のVAD603bおよび第2のVAD603cの各々より大幅に優れた性能を示しており、誤警報制約602が緩和されたという理由から常により高い検出601を記録した。

30

【0071】

本開示において説明された技法は、ハードウェア、ソフトウェア、ファームウェア、またはこれらの任意の組合せで実現され得る。ユニットまたはコンポーネントとして説明された如何なる特徴要素も、集積された論理デバイスまたは個別ではあるが相互動作可能な論理デバイスと一緒に実現され得る。ソフトウェアで実現された場合にはこれらの技法は少なくともある程度は、実行されたときに上記の方法の1つ以上を実行する命令を備えたコンピュータ可読媒体によって実現され得る。コンピュータ可読媒体は、パッケージ資料を含み得るコンピュータプログラム製品の一部を形成し得る。コンピュータ可読媒体は、同期ダイナミック・ランダム・アクセス・メモリ(SDRAM)といったランダムアクセスメモリ(RAM)、読み出し専用メモリ(ROM)、不揮発性ランダムアクセスメモリ(NVRAM)、電氣的消去可能プログラム可能読み出し専用メモリ(EEPROM)、FLASHメモリ、磁気または光データ記憶媒体などを備え得る。これらの技法は更に、または代替として少なくともある程度は、命令またはデータ構造体の形をしたコードを搬送または伝達し、コンピュータによってアクセス、読み取り、および/または実行され得るコンピュータ可読通信媒体によって実現され得る。

40

【0072】

コードは、1つ以上のデジタル信号プロセッサ(DPS)、汎用マイクロプロセッサ、特定用途向け集積回路(ASIC)、フィールドプログラマブル論理アレイ(FPGA)、または他の同等な集積された、または個別の論理回路といった1つ以上のプロセッサによって実行され得る。したがってここで使用される用語「プロセッサ」は、上記の構造体のい

50

ずれか、またはここで説明された技法の実現のために適した他の任意の構造体を指す可能性がある。更に、ある幾つかの態様ではここで説明された機能は、符号化および復号するように構成された、または複合符号器・復号器(CODEC)に組み込まれた専用のソフトウェアユニットまたはハードウェアユニット内に備えられ得る。ユニットまたはモジュールとしての種々の特徴要素の描写は、説明されたデバイスの種々の機能的態様を強調するように意図されており、このようなユニットが別々のハードウェアまたはソフトウェアコンポーネントによって実現されなければならないことを必ずしも意味しない。むしろ1つ以上のユニットまたはモジュールに関連した機能は、共通の、または別々のハードウェアまたはソフトウェアコンポーネント内に統合され得る。これらの実施形態は、コンピュータプロセッサおよび/または電気回路を使用して実現され得る。

【0073】

本開示の種々の実施形態が説明されてきた。これらおよび他の実施形態は下記の請求項の範囲内にある。

【図1】

図1

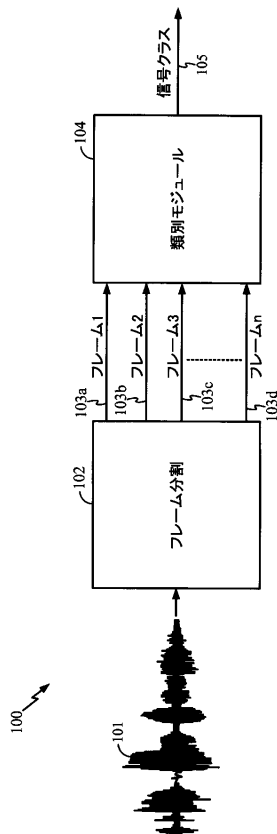


FIG. 1

【図2】

図2

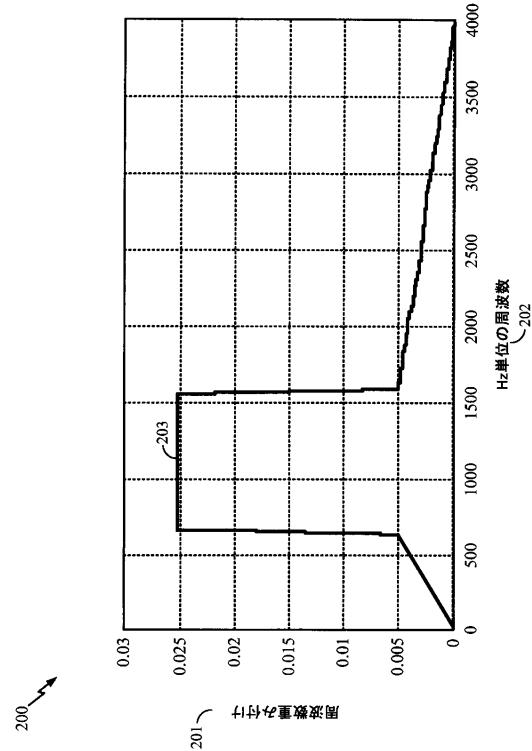
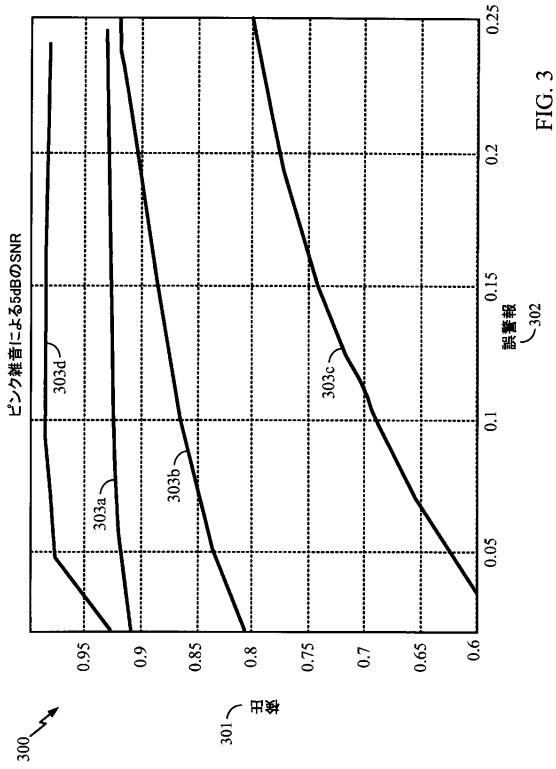


FIG. 2

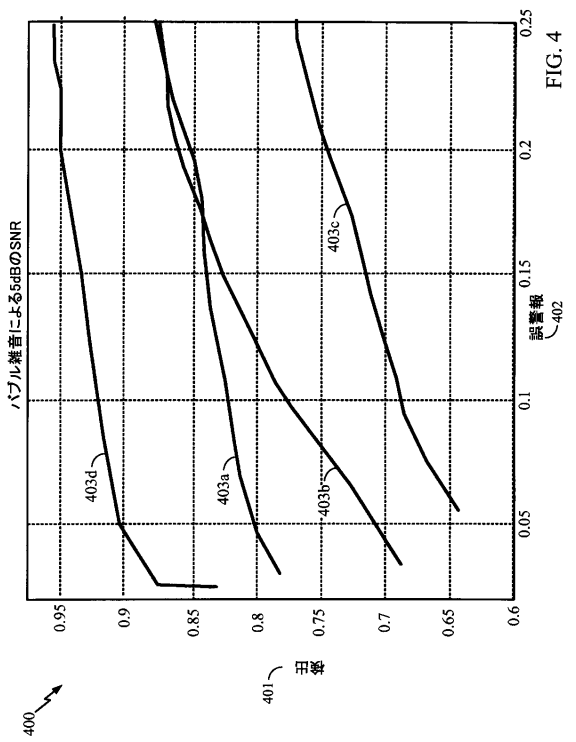
【 図 3 】

図 3



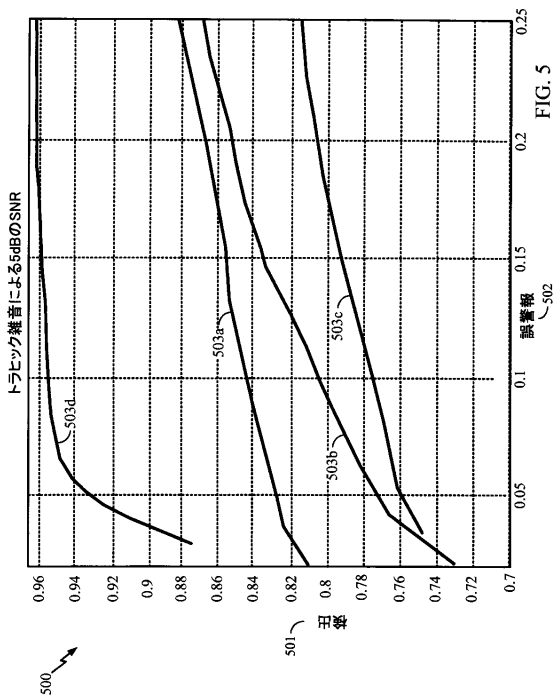
【 図 4 】

図 4



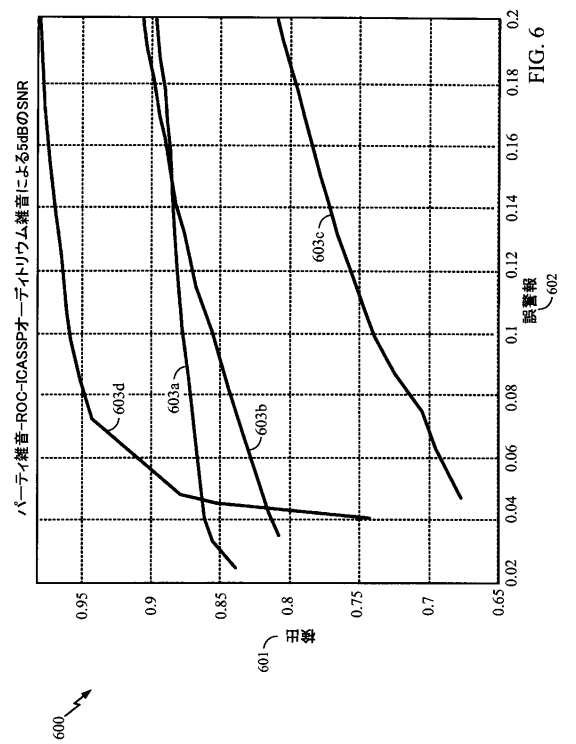
【 図 5 】

図 5



【 図 6 】

図 6



【手続補正書】

【提出日】平成23年6月15日(2011.6.15)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0048

【補正方法】変更

【補正の内容】

【0048】

$$\underline{s[k,n]} = \underline{d} + (1 - \underline{d})\text{prob}[k,n] \quad (19)$$

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0072

【補正方法】変更

【補正の内容】

【0072】

コードは、1つ以上のデジタル信号プロセッサ(DPS)、汎用マイクロプロセッサ、特定用途向け集積回路(ASIC)、フィールドプログラマブル論理アレイ(FPLA)、または他の同等な集積された、または個別の論理回路といった1つ以上のプロセッサによって実行され得る。したがってここで使用される用語「プロセッサ」は、上記の構造体のいずれか、またはここで説明された技法の実現のために適した他の任意の構造体を指す可能性がある。更に、ある幾つかの態様ではここで説明された機能は、符号化および復号するように構成された、または複合符号器・復号器(CODEC)に組み込まれた専用のソフトウェアユニットまたはハードウェアユニット内に備えられ得る。ユニットまたはモジュールとしての種々の特徴要素の描写は、説明されたデバイスの種々の機能的態様を強調するように意図されており、このようなユニットが別々のハードウェアまたはソフトウェアコンポーネントによって実現されなければならないことを必ずしも意味しない。むしろ1つ以上のユニットまたはモジュールに関連した機能は、共通の、または別々のハードウェアまたはソフトウェアコンポーネント内に統合され得る。これらの実施形態は、コンピュータプロセッサおよび/または電気回路を使用して実現され得る。

以下に本件出願当初の特許請求の範囲に記載された発明を付記する。

[1] オーディオ信号の現在フレームにおける雑音レベルを推定する方法であって、複数のオーディオフィームの雑音レベルを決定することと、

前記複数のオーディオフィームにわたる雑音レベルの平均値と標準偏差とを計算することと、

前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、

を備える方法。

[2] 前記平均値からの減算に先立って前記標準偏差をスケール(scaling)することを更に備える[1]に記載の方法。

[3] 複数の雑音レベル推定値の最小値を決定することによって前記現在雑音レベル推定値を決定することを更に備える[1]に記載の方法。

[4] 前記複数のオーディオフィームは約100フレームを備える[1]に記載の方法。

[5] 前記雑音レベル推定値を計算することは平滑化係数を使用することを備える[1]に記載の方法。

[6] 前記雑音レベル推定値は音声活動の期間中に一定に保持される[5]に記載の方法。

[7] 前記平滑化係数は第2の平滑化係数を使用して前記現在フレームにおける音声の確率と1との間を補間することによって再帰的に平均化される[5]に記載の方法。

[8] 前記雑音レベル推定値は複数の予め決められた雑音レベルの最小値を備える[1]に記載の方法。

[9] 前記雑音レベルの平均値は前記雑音レベルの予め計算された平均値を現在雑音レベ

ルによって補間することによって推定される[1]に記載の方法。

[10] 前記計算された雑音レベル推定値を所望の信号レベル未満の12～24dBの間に限界設定することを更に備える[1]に記載の方法。

[11] 前記現在フレームを、非雑音セグメントを有するものとして識別することによって音声活動を検出することを更に備える[1]に記載の方法。

[12] すべての  $\theta$  が  $(0.2, 1)$  であるとして、音声の確率  $\theta >$  であるときに音声活動が宣言される[11]に記載の方法。

[13] 複数のオーディオフレームの雑音レベルを決定するように構成された第1のモジュールと、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算するように構成された第2のモジュールと、

現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算するように構成された第3のモジュールと、

を備える雑音決定システム。

[14] 前記第3のモジュールは前記平均値からの減算に先立って前記標準偏差をスケール(scale)するように構成される[13]に記載の雑音決定システム。

[15] 前記雑音レベル推定値を計算することは平滑化係数を使用することを備える[13]に記載の雑音決定システム。

[16] 前記雑音レベル推定値は音声活動の期間中一定に保持される[15]に記載の雑音決定システム。

[17] 前記平滑化係数は第2の平滑化係数を使用して現在フレームにおける音声の確率と1という値との間を補間することによって再帰的に平均化される[15]に記載の雑音決定システム。

[18] 信号の複数の時間・周波数ビンにおける前記信号の雑音レベルを推定する方法であって、前記信号の前記ビンの各々に関して、

複数のオーディオフレームの雑音レベルを決定することと、

前記時間・周波数ビンにおける雑音レベルを推定することと、

前記時間・周波数ビンにおける仮の雑音レベルを決定することと、

前記仮の雑音レベルから前記時間・周波数ビンにおける2次雑音レベルを決定することと、

前記時間・周波数ビンにおける2次雑音レベルから限界設定された雑音レベルを決定することと、

を備える方法。

[19] 前記限界設定された雑音レベルを決定することはアクティブな所望信号レベル未満の12～24dBの間で前記推定雑音レベルを限界設定することを備える[18]に記載の方法。

[20] 現在フレームにおける各周波数に関する音声の確率の重み付け合計を取ることによって前記現在フレームにおける音声の確率を計算することを更に備える[18]に記載の方法。

[21] 600Hz～1550Hzの範囲内の重み付けが少なくとも0.02という値を与えられる[20]に記載の方法。

[22] オーディオ信号の現在フレームにおける雑音レベルを推定するシステムであって、

複数のオーディオフレームの雑音レベルを決定する手段と、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算する手段と、

前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算する手段と、

を備えるシステム。

[23] 前記現在フレームの雑音レベル推定値を計算する前記手段は前記平均値からの減

算に先立って前記標準偏差をスケールする[22]に記載の雑音決定システム。

[24] 前記雑音レベルを決定する前記手段は信号のエネルギーレベルを決定するように構成されたモジュールを備える[22]に記載のシステム。

[25] 前記雑音レベルの平均値と標準偏差とを計算する前記手段は数学的演算を実行するように構成されたモジュールを備える[22]に記載のシステム。

[26] 雑音レベル推定値を計算する前記手段は数学的演算を実行するように構成されたモジュールを備える[22]に記載のシステム。

[27] プロセッサ上で実行されたときに、

複数のオーディオフレームの雑音レベルを決定することと、

前記複数のオーディオフレームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、

現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、

を備える方法を実行する命令を備えるコンピュータ可読媒体。

[28] 前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える[27]に記載の方法。

[29] 複数のオーディオフレームの雑音レベルを決定することと、

前記複数のオーディオフレームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、

現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、

を備える方法を実行するようにプログラムされたプロセッサ。

[30] 前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える[29]に記載の方法。

【手続補正3】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

オーディオ信号の現在フレームにおける雑音レベルを推定する方法であって、

複数のオーディオフレームの雑音レベルを決定することと、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算することと、

前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、

を備える方法。

【請求項2】

前記平均値からの減算に先立って前記標準偏差をスケール(scaling)することを更に備える請求項1に記載の方法。

【請求項3】

複数の雑音レベル推定値の最小値を決定することによって前記現在雑音レベル推定値を決定することを更に備える請求項1に記載の方法。

【請求項4】

前記複数のオーディオフレームは約100フレームを備える請求項1に記載の方法。

【請求項5】

前記雑音レベル推定値を計算することは平滑化係数を使用することを備える請求項1に記載の方法。

【請求項6】

前記雑音レベル推定値は音声活動の期間中一定に保持される請求項 5 に記載の方法。

【請求項 7】

前記平滑化係数は第 2 の平滑化係数を使用して前記現在フレームにおける音声の確率と 1 との間を補間することによって再帰的に平均化される請求項 5 に記載の方法。

【請求項 8】

前記雑音レベル推定値は複数の予め決められた雑音レベルの最小値を備える請求項 1 に記載の方法。

【請求項 9】

前記雑音レベルの平均値は前記雑音レベルの予め計算された平均値を現在雑音レベルによって補間することによって推定される請求項 1 に記載の方法。

【請求項 10】

前記計算された雑音レベル推定値を所望の信号レベル未満の 12 ~ 24 dB の間に限界設定することを更に備える請求項 1 に記載の方法。

【請求項 11】

前記現在フレームを、非雑音セグメントを有するものとして識別することによって音声活動を検出することを更に備える請求項 1 に記載の方法。

【請求項 12】

すべての  $\theta$  が  $[0.2, 1]$  であるとして、音声の確率  $\theta > \theta_{th}$  であるときに音声活動が宣言される請求項 11 に記載の方法。

【請求項 13】

複数のオーディオフレームの雑音レベルを決定するように構成された第 1 のモジュールと、

前記複数のオーディオフレームにわたる雑音レベルの平均値と標準偏差とを計算するように構成された第 2 のモジュールと、

現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算するように構成された第 3 のモジュールと、  
を備える雑音決定システム。

【請求項 14】

前記第 3 のモジュールは前記平均値からの減算に先立って前記標準偏差をスケール(scale)するように構成される請求項 13 に記載の雑音決定システム。

【請求項 15】

前記雑音レベル推定値を計算することは平滑化係数を使用することを備える請求項 13 に記載の雑音決定システム。

【請求項 16】

前記雑音レベル推定値は音声活動の期間中一定に保持される請求項 15 に記載の雑音決定システム。

【請求項 17】

前記平滑化係数は第 2 の平滑化係数を使用して現在フレームにおける音声の確率と 1 という値との間を補間することによって再帰的に平均化される請求項 15 に記載の雑音決定システム。

【請求項 18】

オーディオ信号の雑音レベルを前記信号の複数の時間・周波数ビンにおいて推定する方法であって、前記信号の前記ビンの各々に関して、

複数のオーディオフレームの雑音レベルを決定することと、

前記時間・周波数ビンにおける雑音レベルを推定することと、

前記時間・周波数ビンにおける仮の雑音レベルを決定することと、

前記仮の雑音レベルから前記時間・周波数ビンにおける 2 次雑音レベルを決定することと、

前記時間・周波数ビンにおける 2 次雑音レベルから限界設定された雑音レベルを決定することと、

を備える方法。

【請求項 19】

前記限界設定された雑音レベルを決定することはアクティブな所望信号レベル未満の 12 ~ 24 dB の間で前記推定雑音レベルを限界設定することを備える請求項 18 に記載の方法。

【請求項 20】

現在フレームにおける各周波数に関する音声の確率の重み付け合計を取ることによって前記現在フレームにおける音声の確率を計算することを更に備える請求項 18 に記載の方法。

【請求項 21】

600 Hz ~ 1550 Hz の範囲内の重み付けが少なくとも 0.02 という値を与えられる請求項 20 に記載の方法。

【請求項 22】

オーディオ信号の現在フレームにおける雑音レベルを推定するシステムであって、複数のオーディオフィームの雑音レベルを決定する手段と、前記複数のオーディオフィームにわたる雑音レベルの平均値と標準偏差とを計算する手段と、前記現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算する手段と、を備えるシステム。

【請求項 23】

前記現在フレームの雑音レベル推定値を計算する前記手段は前記平均値からの減算に先立って前記標準偏差をスケールする請求項 22 に記載の雑音決定システム。

【請求項 24】

前記雑音レベルを決定する前記手段は信号のエネルギーレベルを決定するように構成されたモジュールを備える請求項 22 に記載のシステム。

【請求項 25】

前記雑音レベルの平均値と標準偏差とを計算する前記手段は数学的演算を実行するように構成されたモジュールを備える請求項 22 に記載のシステム。

【請求項 26】

雑音レベル推定値を計算する前記手段は数学的演算を実行するように構成されたモジュールを備える請求項 22 に記載のシステム。

【請求項 27】

プロセッサ上で実行されたときに、複数のオーディオフィームの雑音レベルを決定することと、前記複数のオーディオフィームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、を備える方法を実行する命令を備えるコンピュータ可読媒体。

【請求項 28】

前記方法は前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える請求項 27 に記載の媒体。

【請求項 29】

複数のオーディオフィームの雑音レベルを決定することと、前記複数のオーディオフィームにわたる前記雑音レベルの平均値と標準偏差とを計算することと、現在フレームの雑音レベル推定値を前記平均値から減算される標準偏差の値として計算することと、を備える方法を実行するようにプログラムされたプロセッサ。

## 【請求項 30】

前記方法は前記平均値からの減算に先立って前記標準偏差をスケールすることを更に備える請求項 29 に記載のプロセッサ。

## 【請求項 31】

前記平滑化係数は時間の経過と共に更新される請求項 5 に記載の方法。

## 【請求項 32】

前記平滑化係数は時間の経過と共に更新される請求項 15 に記載の雑音決定システム。

## 【請求項 33】

前記雑音レベル推定値を計算する手段は時間の経過と共に更新される平滑化係数を備える請求項 22 に記載のシステム。

## 【請求項 34】

現在フレームの雑音レベル推定値を計算することは時間の経過と共に更新される平滑化係数を使用することを備える請求項 29 に記載のプロセッサ。

## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2009/060828

A. CLASSIFICATION OF SUBJECT MATTER INV. G10L11/02		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G10L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, INSPEC, COMPENDEX		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 659 570 A (LG ELECTRONICS INC [KR]) 24 May 2006 (2006-05-24)  paragraphs [0001], [0011], [0012] paragraph [0033] - paragraph [0047]	1-4, 8-14, 18-30
Y	----- -/--	5-7, 15-17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed		*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family
Date of the actual completion of the international search  8 December 2009		Date of mailing of the international search report  23/12/2009
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer  Chétry, Nicolas

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2009/060828

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>COHEN I: "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 11, no. 5, 1 September 2003 (2003-09-01), pages 466-475, XP011100006 ISSN: 1063-6676 p. 467, section II, "Time-varying recursive averaging"</p>	5-7, 15-17
Y	<p>RAINER MARTIN: "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 9, no. 5, 1 July 2001 (2001-07-01), XP011054118 ISSN: 1063-6676 p. 505, section II, "Principles of minimum statistics noise estimation"</p>	5-7, 15-17
A	<p>RIS C ET AL: "Assessing local noise level estimation methods: application to noise robust ASR" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 34, no. 1-2, 1 April 2001 (2001-04-01), pages 141-158, XP002224855 ISSN: 0167-6393 p. 146, section 2.4, "Low-energy envelope tracking" p. 146, section 2.5, "Correction of the measures"</p>	1-30
A	<p>DAVIS ET AL: "A multi-decision sub-band voice activity detector" PROCEEDINGS EUSIPCO, 6 September 2006 (2006-09-06), pages 1-5, XP002559305 Florence, Italy the whole document</p>	1-30
	----- -/--	

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2009/060828

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>JONGSEO SOHN ET AL: "A voice activity detector employing soft decision based noise spectrum adaptation" ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA 12-15 MAY 1998, NEW YORK, NY, USA, IEEE, US, vol. 1, 12 May 1998 (1998-05-12), pages 365-368, XP010279166 ISBN: 978-0-7803-4428-0 the whole document</p> <p style="text-align: center;">-----</p>	1-30

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No

PCT/US2009/060828

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 1659570	A	24-05-2006	AT 412235 T	15-11-2008
			CN 1805007 A	19-07-2006
			JP 4282659 B2	24-06-2009
			JP 2006146226 A	08-06-2006
			KR 20060056186 A	24-05-2006
			US 2006111901 A1	25-05-2006

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(74)代理人 100109830

弁理士 福原 淑弘

(74)代理人 100075672

弁理士 峰 隆司

(74)代理人 100095441

弁理士 白根 俊郎

(74)代理人 100084618

弁理士 村松 貞男

(74)代理人 100103034

弁理士 野河 信久

(74)代理人 100119976

弁理士 幸長 保次郎

(74)代理人 100153051

弁理士 河野 直樹

(74)代理人 100140176

弁理士 砂川 克

(74)代理人 100158805

弁理士 井関 守三

(74)代理人 100124394

弁理士 佐藤 立志

(74)代理人 100112807

弁理士 岡田 貴志

(74)代理人 100111073

弁理士 堀内 美保子

(74)代理人 100134290

弁理士 竹内 将訓

(72)発明者 モハンマド、アシフ・アイ .

アメリカ合衆国、カリフォルニア州 9 2 1 2 1、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5

(72)発明者 ラマクリシュナン、ディネッシュ

アメリカ合衆国、カリフォルニア州 9 2 1 2 1、サン・ディエゴ、モアハウス・ドライブ 5 7 7 5