

(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2010-539792

(P2010-539792A)

(43) 公表日 平成22年12月16日(2010.12.16)

(51) Int.Cl.	F I	テーマコード (参考)
<b>H04S 1/00 (2006.01)</b>	H04S 1/00 F	5D062
<b>G10L 21/02 (2006.01)</b>	G10L 21/02 302B	
<b>G10L 21/04 (2006.01)</b>	G10L 21/04 120Z	
<b>G10L 11/00 (2006.01)</b>	G10L 11/00 402G	

審査請求 有 予備審査請求 有 (全 28 頁)

(21) 出願番号 特願2010-524855 (P2010-524855)  
 (86) (22) 出願日 平成20年9月10日 (2008.9.10)  
 (85) 翻訳文提出日 平成22年4月23日 (2010.4.23)  
 (86) 国際出願番号 PCT/US2008/010591  
 (87) 国際公開番号 W02009/035615  
 (87) 国際公開日 平成21年3月19日 (2009.3.19)  
 (31) 優先権主張番号 60/993,601  
 (32) 優先日 平成19年9月12日 (2007.9.12)  
 (33) 優先権主張国 米国 (US)

(71) 出願人 591102637  
 ドルビー・ラボラトリーズ・ライセンシング・コーポレーション  
 DOLBY LABORATORIES LICENSING CORPORATION  
 アメリカ合衆国、カリフォルニア州 94103-4813、サン・フランシスコ、ポトレロ・アベニュー 100  
 (74) 代理人 100071010 弁理士 山崎 行造  
 (74) 代理人 100121762 弁理士 杉山 直人  
 (74) 代理人 100126767 弁理士 白銀 博

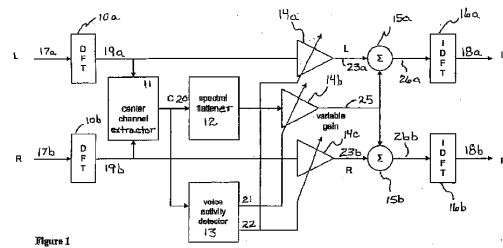
最終頁に続く

(54) 【発明の名称】 スピーチ増強

(57) 【要約】

スピーチを増強する方法であって、オーディオ信号の中央チャンネルを抽出して、中央チャンネルのスペクトルを平坦化して、平坦化されたスピーチ・チャンネルにオーディオ信号を混合することにより、オーディオ信号における任意のスピーチを増強することを含む。また、複数のチャンネルによりオーディオ信号から音の中央チャンネルを抽出する方法、オーディオ信号のスペクトルを平坦化する方法、及びオーディオ信号におけるスピーチを検出する方法が開示される。更に、スピーチエンハンサが開示され、これはオーディオ信号の中央チャンネルを抽出するための抽出器と、前記中央チャンネルのスペクトルを平坦化するためのスペクトル平坦化器と、前記中央チャンネルにおけるスピーチ検出における信頼度を生成するスピーチ信頼度生成器と、前記平坦化されたスピーチ・チャンネルを原オーディオ信号に、検出されたスピーチの信頼度に比例して混合することにより、オーディオ信号における任意のスピーチを増強するミキサとを備える。

【選択図】 図 1



**【特許請求の範囲】****【請求項 1】**

複数のチャンネルによりオーディオ信号から音の中央チャンネルを抽出する方法であって、

- (1) 候補中央チャンネルの比より小さいオーディオ信号の第1のチャンネルと、
- (2) 候補中央チャンネルの比より小さいオーディオ信号の第2のチャンネルの共役とを乗じて、

を概ね最小化して、

候補中央チャンネルを概ね最小化されたに乘じることにより、抽出された中央チャンネルを形成することを含む方法。

10

**【請求項 2】**

オーディオ信号のスペクトルを平坦化する方法であって、

推定されたスピーチ・チャンネルを知覚帯域に分離して、

知覚帯域のうちの何れが最も多くのエネルギーを有しているかを判定し、

より少ないエネルギーを有する知覚帯域の利得を増大させることにより、オーディオ信号における任意のスピーチのスペクトルを平坦化させることを含む方法。

**【請求項 3】**

請求項2の方法において、前記増大は、より少ないエネルギーを有する知覚帯域の利得を最大まで増大することを含む方法。

**【請求項 4】**

20

オーディオ信号におけるスピーチを検出する方法であって、

オーディオ信号の候補中央チャンネルにおけるスペクトル変動を測定し、

候補中央チャンネルよりも少ないオーディオ信号のスペクトル変動を測定して、

これらスペクトル変動を比較することにより、オーディオ信号におけるスピーチを検出することを含む方法。

**【請求項 5】**

スピーチを増強する方法であって、

オーディオ信号の中央チャンネルを抽出して、

中央チャンネルのスペクトルを平坦化して、

平坦化されたスピーチ・チャンネルにオーディオ信号を混合することにより、オーディオ信号における任意のスピーチを増強することを含む方法。

30

**【請求項 6】**

請求項5の方法において、

中央チャンネルにおけるスピーチ検出に信頼度を生成すると共に、

前記混合は、

平坦化されたスピーチ・チャンネルにオーディオ信号を、検出されたスピーチを有する信頼度に比例させて混合することを更に含む方法。

**【請求項 7】**

請求項6の方法において、

前記信頼度は、可能性確率が最も低いものから可能性確率が最も高いものまで変動すると共に、

40

前記生成は、

最低可能性確率よりも高く、且つ最高可能性確率よりも低い値に対して生成された信頼度を更に制限することを含む方法。

**【請求項 8】**

請求項5の方法において、前記抽出は、請求項1の方法を用いてオーディオ信号の中央チャンネルを抽出することを含む方法。

**【請求項 9】**

請求項5の方法において、前記平坦化は、請求項2の方法を用いて前記中央チャンネルのスペクトルを平坦化することを含む方法。

50

## 【請求項 10】

請求項 5 の方法において、前記生成は、請求項 3 の方法を用いて前記中央チャンネルにおけるスピーチ検出に信頼度を生成することを含む方法。

## 【請求項 11】

請求項 5 の方法において、前記抽出は、請求項 1 の方法を用いてオーディオ信号の中央チャンネルを抽出することを含み、

前記平坦化は、請求項 2 の方法を用いて前記中央チャンネルのスペクトルを平坦化することを含み、

前記生成は、請求項 3 の方法を用いて前記中央チャンネルにおけるスピーチ検出に信頼度を生成することを含む方法。

10

## 【請求項 12】

請求項 1 乃至 11 の何れか一項に記載の方法を実行するコンピュータ・プログラムが格納されたコンピュータ読み取り可能な記録媒体。

## 【請求項 13】

コンピュータ・システムであり、

CPUと、

請求項 12 の記録媒体と、

前記 CPU と前記記録媒体とを結合するバスを含むコンピュータ・システム。

## 【請求項 14】

スピーチエンハンサであって、

20

オーディオ信号の中央チャンネルを抽出するための抽出器と、

前記中央チャンネルのスペクトルを平坦化するためのスペクトル平坦化器と、

前記中央チャンネルにおけるスピーチ検出における信頼度を生成するスピーチ信頼度生成器と、

前記平坦化されたスピーチ・チャンネルを原オーディオ信号に、検出されたスピーチの信頼度に比例して混合することにより、オーディオ信号における任意のスピーチを増強するミキサとを備えるスピーチエンハンサ。

## 【発明の詳細な説明】

## 【発明の詳細な説明】

## 【0001】

30

発明の開示

本明細書は、複数のチャンネルでオーディオ信号から音の中央チャンネルを抽出して、オーディオ信号のスペクトルを平坦化して、オーディオ信号におけるスピーチ（発話）を検出して、スピーチを引き立たせる方法及び装置を説明する。複数のチャンネルによりオーディオ信号から音の中央チャンネルを抽出する方法は、（1）候補中央チャンネルの比より小さいオーディオ信号の第1のチャンネル及び（2）候補中央チャンネルの比より小さいオーディオ信号の第2のチャンネルの共役を乗じて、を概ね最小化して、候補中央チャンネルにその概ね最小化された を乗じることにより、抽出された中央チャンネルを形成することを含む。

## 【0002】

40

オーディオ信号のスペクトルを平坦化する方法は、推定されたスピーチ・チャンネルを知覚帯域に分離して、知覚帯域のうちの何れが最も多くのエネルギーを有しているかを判定し、より少ないエネルギーを有する知覚帯域の利得を増大させることにより、オーディオ信号における任意のスピーチのスペクトルを平坦化させることを含んでもよい。この増大は、より少ないエネルギーを有する知覚帯域の利得を最大まで増大することを含んでもよい。

## 【0003】

オーディオ信号におけるスピーチを検出する方法は、オーディオ信号の候補中央チャンネルにおけるスペクトル変動を測定し、候補中央チャンネルよりも少ないオーディオ信号のスペクトル変動を測定して、これらスペクトル変動を比較することにより、オーディオ信号におけるスピーチを検出することを含んでもよい。

50

## 【 0 0 0 4 】

スピーチを増強する方法は、オーディオ信号の中央チャンネルを抽出して、中央チャンネルのスペクトルを平坦化して、平坦化されたスピーチ・チャンネルにオーディオ信号を混合することにより、オーディオ信号における任意のスピーチを増強することを含んでもよい。この方法は、中央チャンネルにおけるスピーチ検出に信頼度を生成することを更に含んでもよく、その混合は、平坦化されたスピーチ・チャンネルにオーディオ信号を、検出されたスピーチを有する信頼度に比例させて混合することを含んでもよい。その信頼度は、可能性確率が最も低いものから可能性確率が最も高いものまで変動し得るので、その生成は、最低可能性確率よりも高く、且つ最高可能性確率よりも低い値に対して生成された信頼度を制限することを更に含んでもよい。その抽出は、上述した方法を用いて、オーディオ信号の中央チャンネルを抽出することを含んでもよい。上述の平坦化は、上述の方法を用いて中央チャンネルのスペクトルを平坦化することを含んでもよい。上述の生成は、上述の方法を用いて中央チャンネルにおけるスピーチ検出に信頼度を生成することを含んでもよい。

10

## 【 0 0 0 5 】

上述の抽出は、上述の方法を用いてオーディオ信号の中央チャンネルを抽出することを含んでもよく、上述の平坦化は、上述の方法を用いて中央チャンネルのスペクトルを平坦化することを含んでもよく、上述の生成は、上述の方法を用いて中央チャンネルにおけるスピーチ検出に信頼度を生成することを含んでもよい。

20

## 【 0 0 0 6 】

本明細書は、上述の方法の何れかを実行するコンピュータ・プログラムが格納されたコンピュータ読み取り可能な記録媒体のみならず、CPU、該記録媒体、及びこれらCPUと記録媒体とを結合するバスを含むコンピュータ・システムを教示する。

## 【 図面の簡単な説明 】

## 【 0 0 0 7 】

【 図 1 】 本発明の一つの実施例によるスピーチ・エンハンサーの機能ブロック図である。

【 図 2 】 計 4 0 帯域をもたらす間隔 1 E R B のフィルタの適宜なセットを表す図である。

【 図 3 】 本発明の一つの実施例による混合プロセスを説明する図である。

【 図 4 】 本発明の一つの実施例によるコンピュータ・システムを例示する図である。

30

## 【 0 0 0 8 】

発明を実施するための最良の形態

図 1 は本発明の一つの実施例によるスピーチ・エンハンサーの機能ブロック図である。スピーチ・エンハンサー 1 は、入力信号 1 7、離散フーリエ変換器 1 0 a, 1 0 b、中央チャンネル抽出器 1 1、スペクトル平坦化器 1 2、発声活動検出器 1 3、可変利得増幅器 1 5, 1 5 c、逆離散フーリエ変換器 1 8 a, 1 8 b 及び出力信号 1 8 を含む。入力信号 1 7 はそれぞれ左右のチャンネル 1 7 a, 1 7 b から成り、同様に出力信号 1 8 はそれぞれ左右のチャンネル 1 8 a, 1 8 b から成る。

## 【 0 0 0 9 】

各々の離散フーリエ変換器 1 8 は、入力として入力信号 1 7 の左右チャンネル 1 7 a, 1 7 b を受け取って、出力として変換 1 9 a, 1 9 b を形成する。中央チャンネル抽出器 1 1 は、変換 1 9 を受け取って、出力として仮の中央チャンネル C 2 0 を形成する。スペクトル平坦化器 1 2 は入力として仮の中央チャンネル C 2 0 を受け取って、成形された中央チャンネル 2 4 を出力として形成し、一方、発声活動検出器 1 3 は同じ入力 C 2 0 を受け取って、一方では可変利得増幅器 1 4 a 及び 1 4 c のための制御信号 2 2 を、他方では可変利得増幅器 1 4 b のための制御信号 2 1 を、出力として形成する。

40

## 【 0 0 1 0 】

増幅器 1 4 a は、入力及び制御信号として、左チャンネル変換 1 9 a 及び発声活動検出器 1 3 の出力制御信号 2 2 をそれぞれ受け取る。同様に、増幅器 1 4 c は、入力及び制御信号として、右チャンネル変換 1 9 b 及び発声活動検出器出力制御信号 2 2 をそれぞれ受け取る。増幅器 1 4 b は、入力及び制御信号として、スペクトル的に成形された中央チャ

50

ンネル 2 4 及びスペクトル平坦化器 1 2 の出力発声活動検出器制御信号 2 1 を受け取る。

【 0 0 1 1 】

ミキサー 1 5 a は、増幅器 1 4 からの出力である利得調整された左変換 2 3 a と、利得調整されたスペクトル的に成形された中央チャンネル 2 5 とを受け取って、出力として信号 2 6 a を形成する。同様に、ミキサー 1 5 b は、増幅器 1 4 c からの利得調整された右変換 2 3 b と、利得調整されたスペクトル的に成形された中央チャンネル 2 5 とを受け取って、出力として信号 2 6 b を形成する。

【 0 0 1 2 】

逆変換器 1 8 a , 1 8 b は、各々の信号 2 6 a , 2 6 b を受け取って、それぞれ導出された左及び右チャンネル信号  $L' 1 8 a$  及び  $R' 1 8 b$  を形成する。

10

【 0 0 1 3 】

スピーチ・エンハンサー 1 の操作を以下に更に詳細に説明する。中央チャンネル抽出、スペクトル平坦化、発声活動検出及び混合の処理については、一つの実施例に沿って最初は概略的に次いでより詳細に順番に説明する。

【 0 0 1 4 】

中央チャンネル抽出

以下のように仮定する。

【 0 0 1 5 】

( 1 ) 対象 1 7 の信号はスピーチを包含する。

【 0 0 1 6 】

( 2 ) 多重チャンネル信号 ( 即ち、左及び右、又はステレオ ) の場合、スピーチは中央にパンされる。

20

【 0 0 1 7 】

( 3 ) 実際のパンされた中央は、音源左右信号の比  $\alpha$  ( ) から成る。

【 0 0 1 8 】

( 4 ) その比の減算の結果は一对の直交信号である。

【 0 0 1 9 】

これらの仮定の下に操作して、中央チャンネル抽出器 1 1 はステレオ信号 1 7 から中央にパンされたコンテンツ  $C 2 0$  を抽出する。中央にパンされたコンテンツのために、左右両方のチャンネルの同一の領域は、その中央にパンされたコンテンツを含む。中央にパンされたコンテンツは、左右両方のチャンネルから同一部分を除去することにより抽出される。

30

【 0 0 2 0 】

残りの左右信号について ( ブロックのフレーム上で、或いは新しいブロックが入る毎に連続的に更新される方法を用いて )、 $L R^* = 0$  ( ここで \* は共役を示す ) を計算し、比が零に十分に近い値になるまで調整するようにしてもよい。

【 0 0 2 1 】

スペクトル平坦化

聴覚フィルタは、推定されたスピーチ・チャンネルにおけるスピーチを知覚帯域へ分離する。最も多くのエネルギーを有する帯域、データの各々のブロックについて判定される。そのブロックについてのスピーチ・チャンネルのスペクトル形状は、残りの帯域における低エネルギーを補償するために修正される。このスペクトルは平坦化される。低エネルギーを有する帯域は、或る最大限まで増大された利得を持つ。一つの実施形態においては、全ての帯域は最大利得を共有してもよい。代替的な実施形態においては、各々の帯域は、それ自身の最大利得を有してもよい。( 全ての帯域が同じエネルギーを有するという望ましくない場合には、スペクトルは既に平坦である。スペクトル成形が生じないか、或いはスペクトル成形が同一の機能により達成されることも考慮されるであろう。 )

40

スペクトル平坦化はチャンネルのコンテンツとは無関係に生じる。非スピーチを処理してもよいが、これがシステムにおいて後で用いられることはない。非スピーチは、スピーチとは非常に異なるスペクトルを有するので、非スピーチのための平坦化は、通常はスピー

50

ーチについてのものと同じではない。

【0022】

#### 発声活動検出器

推定されたスピーチが単独のチャンネルへ分離されると、それはスピーチ・コンテンツについて分析される（それはスピーチを包含するか？）。コンテンツはスペクトル平坦化とは独立に分析される。スピーチ・コンテンツは、データの隣接するフレームにおけるスペクトル変動を測定することにより判定される。（各々のフレームはデータの多くのブロックから成り得るが、フレームは一般に48kHzサンプル・レートの2、4、又は8ブロックである。）

スピーチ・チャンネルがステレオから抽出される場所では、残りのステレオ信号がスピーチ分析に役立つであろう。この概念は、任意の多重チャンネル源における隣接するチャンネルにより一般的に適用される。

10

【0023】

#### ミキシング

スピーチが存在すると見做される時、平坦化されたスピーチ・チャンネルは、スピーチ・チャンネルが実際にスピーチを包含するという信頼度に関連する或る割合で原信号と混合される。一般に、信頼度が高いときは、より多くの平坦化スピーチ・チャンネルが用いられる。信頼度が低いときは、より少ない平坦化スピーチ・チャンネルが用いられる。

【0024】

中央チャンネル抽出、スペクトル平坦化、発声活動検出及び混合の処理について、一つの実施例によって更に詳細に順番に説明する。

20

【0025】

#### 2チャンネル源からの仮の中央及びサラウンド・チャンネル抽出

スピーチ増強によれば、中央にパンされたオーディオのみを抽出、処理、及び再挿入することが望まれる。ステレオ混合においては、スピーチは最も頻繁に中央へパンされる。

【0026】

ここで、中央にパンされたオーディオ（仮の中央チャンネル）の2チャンネル混合物からの抽出について説明する。数学的な証明は第1の部分構成する。第2の部分は、この証明を実環境ステレオ信号に適用して、仮の中央を導出する。

30

【0027】

仮の中央が原ステレオから取り去られると、直交チャンネルを有するステレオ信号が残る。類似の方法は、周辺にパンされたオーディオから仮のサラウンド・チャンネルを導出する。

【0028】

#### 中央チャンネル抽出 - 数学的証明

或る2チャンネル信号が与えられると、そのチャンネルは左(L)と右(R)に分けられるであろう。この左右のチャンネルの各々は、共通の情報のみならず、各々に固有の情報を包含する。共通の情報をC（中央にパンされている）、固有の情報を左のみ右のみについてそれぞれL及びRとして表すことができる。

40

【数1】

$$\begin{aligned} L &= L + C \\ R &= R + C \end{aligned} \quad (1)$$

【0029】

「固有」とはL及びRが互いに直交することを意味する。

【数 2】

$$\mathbf{LR}^* = 0 \quad (2)$$

【0030】

L 及び R を実数と虚数部分とへ分けると、

【数 3】

$$L_r R_r + L_i R_i = 0 \quad (3)$$

10

【0031】

ここで、 $L_r$  は L の実数部分、 $L_i$  は L の虚数部分であり、R についても同様である。いま、中央にパンされた C を L 及び R から減じることにより、非直交対 (L 及び R) から直交対 (L 及び R) が形成されたものと見做す。

【数 4】

$$\mathbf{L} = \mathbf{L} - \mathbf{C} \quad (4)$$

$$\mathbf{R} = \mathbf{R} - \mathbf{C} \quad (5)$$

20

【0032】

ここで  $C = \alpha C$  (但し、 $C$  は推定された中央チャンネルであり、 $\alpha$  は倍率である) とすると、

【数 5】

$$\mathbf{L} = \mathbf{L} - \alpha \mathbf{C} \quad (6)$$

$$\mathbf{R} = \mathbf{R} - \alpha \mathbf{C} \quad (7)$$

30

【0033】

式 (6) 及び式 (7) を式 (3) へ代入すると、

【数 6】

$$\begin{aligned} L_r R_r + L_i R_i &= (L_r - \alpha C_r)(R_r - \alpha C_r) + (L_i - \alpha C_i)(R_i - \alpha C_i) \\ &= L_r R_r - \alpha C_r (L_r + R_r) + \alpha^2 C_r^2 + L_i R_i - \alpha C_i (L_i + R_i) + \alpha^2 C_i^2 \\ &= \alpha^2 [C_r^2 + C_i^2] + \alpha [-C_r (L_r + R_r) - C_i (L_i + R_i)] + [L_r R_r + L_i R_i] \\ &= 0 \end{aligned}$$

40

(8)

【0034】

式 (8) は二次方程式の形になり、

【数 7】

$$\alpha^2 X + \alpha Y + Z = 0 \quad (9)$$

50

【 0 0 3 5 】

ここで累乗根は以下のように得られる。

【 数 8 】

$$\alpha = \frac{-Y \pm \sqrt{Y^2 - 4XZ}}{2X} \quad (10)$$

【 0 0 3 6 】

ここで式 ( 6 ) 及び式 ( 7 ) における C を、

【 数 9 】

$$C = L + R \quad (11)$$

【 0 0 3 7 】

として、実数と虚数とに分けると、

【 数 1 0 】

$$C_r = L_r + R_r \quad (12)$$

$$C_i = L_i + R_i \quad (13)$$

【 0 0 3 8 】

すると、二次方程式 ( 9 ) においては、

【 数 1 1 】

$$X = C_r^2 + C_i^2 = (L_r + R_r)^2 + (L_i + R_i)^2 \quad (14)$$

$$Y = -C_r(L_r + R_r) - C_i(L_i + R_i) = -(L_r + R_r)^2 - (L_i + R_i)^2 = -X \quad (15)$$

$$Z = L_r R_r + L_i R_i \quad (16)$$

【 0 0 3 9 】

式 ( 1 4 ) 、式 ( 1 5 ) 及び式 ( 1 6 ) を式 ( 1 0 ) へ代入して、 について解くと、

10

20

30



【数 1 2】

$$\begin{aligned} \alpha &= \frac{-Y \pm \sqrt{Y^2 - 4XZ}}{2X} = \frac{X \pm \sqrt{X^2 - 4XZ}}{2X} = \frac{1 \pm \sqrt{1 - 4\frac{Z}{X}}}{2} \\ &= \frac{1 \pm \sqrt{1 - 4\frac{L_r R_r + L_i R_i}{(L_r + R_r)^2 + (L_i + R_i)^2}}}{2} = \frac{1}{2} \times \left[ 1 \pm \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right] \end{aligned} \quad (17)$$

10

【0040】

に対する解について負の根を選び、周辺にパンされた情報による混乱を避けるため  
を範囲  $\{0, 0.5\}$  に限定する（但し、その値は本発明には重要ではない）。仮の中央  
チャンネル式は以下ようになる。

【数 1 3】

$$\begin{aligned} C &= \alpha C = \alpha(L + R) \\ &= \alpha \left[ (L_r + R_r) + \sqrt{-1}(L_i + R_i) \right] \end{aligned} \quad (18)$$

20

【0041】

ここで、

【数 1 4】

$$\alpha = \min \left\{ \max \left\{ 0, \frac{1}{2} \times \left[ 1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right] \right\}, 0.5 \right\} \quad (19)$$

30

【0042】

である。（ $\min \{ \}$  及び  $\max \{ \}$  関数は、 を範囲  $\{0, 0.5\}$  に制限する  
が、その値は本発明には重要ではない。）

仮のサラウンド・チャンネルは同様に以下のように導ける。

【0043】

【数 1 5】

$$\begin{aligned} S &= \beta S = \beta(L - R) \\ &= \beta \left[ (L_r - R_r) + \sqrt{-1}(L_i - R_i) \right] \end{aligned} \quad (20)$$

40

$$\beta = \min \left\{ \max \left\{ 0, \frac{1}{2} \times \left[ 1 - \sqrt{\frac{(L_r + R_r)^2 + (L_i + R_i)^2}{(L_r - R_r)^2 + (L_i - R_i)^2}} \right] \right\}, 0.5 \right\} \quad (21)$$

【0044】

ここで  $S$  は、原ステレオ対  $(L, R)$  において周辺にパンされたオーディオであり、  
且つ  $S$  は  $(L - R)$  になるものと仮定する。この場合も、 に対する解について負の根を

50

選び、周辺にパンされた情報による混乱を避けるため を範囲  $\{0, 0.5\}$  に限定する (但し、その値は本発明には重要ではない)。

【0045】

いまやC及びSが導出されたので、これらを原ステレオ対(L及びR)から除去して、二つの原チャンネルからオーディオの四つのチャンネルを形成することができる。即ち、  
【数16】

$$L' = L - C - S \quad (22)$$

$$R' = R - C + S \quad (23)$$

10

【0046】

ここでL'は導出された左チャンネル、Cは導出された中央チャンネル、R'は導出された右チャンネル、Sは導出されたサラウンド・チャンネルである。

【0047】

中央チャンネル抽出 - 適用

上述のように、スピーチ増強方法にとって、その主要な懸念は中央チャンネルの抽出である。この部分において、上述の技術は、オーディオ信号の複雑な周波数領域表現に適用される。

20

【0048】

仮の中央チャンネル抽出の第1段階は、オーディオ・サンプルのブロックでDFTを実行し、その結果として生じる変換係数を得ることである。DFTのブロック・サイズはサンプリング・レートに依存する。例えば48kHzのサンプリング・レート $f_s$ においては、 $N = 512$ サンプルのブロック・サイズが可能である。ハミング・ウィンドウのようなウィンドーイング関数 $w[n]$ により、変換の適用に先立ってサンプルのブロックを重み付けする。

【0049】

【数17】

$$w[n] = 0.5 \left( 1 - \cos \left( \frac{2\pi n}{N-1} \right) \right) \quad 0 \leq n < N \quad (24)$$

30

【0050】

ここでnは整数であり、Nはブロックにおけるサンプルの数である。

【0051】

DFT係数を次式(25)で以下のように計算する。

【0052】

【数18】

$$X_m[k, c] = \sum_{n=0}^{N-1} x[mN+n, c] w[n] e^{\frac{-j2\pi kn}{N}} \quad \begin{array}{l} 0 \leq k < N \\ 1 \leq c \leq 3 \end{array} \quad (25)$$

40

【0053】

ここで $x[n, c]$ はブロックmのチャンネルcにおけるサンプル番号nであり、jは虚数単位( $j^2 = -1$ )であり、 $X_m[k, c]$ はブロックmにおけるサンプルについてのチャンネルcにおける変換係数kである。チャンネルの数は三つ、即ち、左、右、及び仮の中央( $x[n, c]$ の場合においては、左及び右のみ)であることに留意されたい

50

。以下の方程式において、左チャンネルは  $c = 1$  として表され、仮の中央チャンネルは  $c = 2$ （未だ導出されていない）、右チャンネルは  $c = 3$  として表される。また、高速フーリエ変換（FFT）はDFTを効率的に実行する。

【0054】

左と右との和及び差は、原則として周波数ビン毎に求めた。実数及び虚数部分はグループ分けして二乗した。各ビンは、 $B_m(k)$  を計算するのに先立ってブロック間で平滑化した。この平滑化は、可聴なアーチファクト（これは、ビンにおけるパワーがデータのブロック間で急激に変化したときに生じる）を低減させる。平滑化は、例えば、漏れ積分回路（leaky integrator）、非線形スモザー、線形且つ多極のローパス・スモザー、或いは更に精巧なスモザーで実行してもよい。

10

【数19】

$$B_m(k)_{diff} = (\text{Re}\{X_m[k,1]\} - \text{Re}\{X_m[k,3]\})^2 + (\text{Im}\{X_m[k,1]\} - \text{Im}\{X_m[k,3]\})^2$$

(26a)

$$B_m(k)_{sum} = (\text{Re}\{X_m[k,1]\} + \text{Re}\{X_m[k,3]\})^2 + (\text{Im}\{X_m[k,1]\} + \text{Im}\{X_m[k,3]\})^2$$

(26b)

20

$$B_{temp} = \lambda_1 B_{m-1}(k)_{diff} + (1 - \lambda_1) B_m(k)_{diff} \quad 0 << \lambda_1 < 1 \quad (26c)$$

$$B_m(k)_{diff} = B_{temp}$$

$$B_{temp} = \lambda_1 B_{m-1}(k)_{sum} + (1 - \lambda_1) B_m(k)_{sum} \quad 0 << \lambda_1 < 1 \quad (26d)$$

$$B_m(k)_{diff} = B_{temp}$$

30

【0055】

ここで  $\text{Re}\{\quad\}$  は実数部分であり、 $\text{Im}\{\quad\}$  は虚数部分であり、 $\lambda_1$  は漏れ積分回路係数である。漏れ積分回路はローパス・フィルタリング効果を有し、 $\lambda_1$  についての代表的な値は0.9である。次に、ブロック  $m$  についての抽出係数  $\alpha_m(k)$  は式(19)を用いて以下のように導かれる。

【数20】

$$\alpha_m(k) = \min \left\{ \max \left\{ 0, \frac{1}{2} \times \left[ 1 - \sqrt{\frac{E_m(k)_{diff}}{E_m(k)_{sum}}} \right] \right\}, 0.5 \right\} \quad (27)$$

40

【0056】

そして、ブロック  $m$  についての仮の中央チャンネルは式(18)を用いて以下のように導かれる。

【数 2 1】

$$X_m[k,2] = \alpha_m(k)(X_m[k,1] + X_m[k,3]) \quad (28)$$

【0057】

スペクトル平坦化

以下、本発明のスペクトル平坦化の実施例を説明する。大部分がスピーチである単独のチャンネルを仮定し、そのスピーチ信号を離散フーリエ変換（DFT）又は関連した変換によって周波数領域へ変換する。振幅スペクトルは、変換周波数ピンを二乗することによってパワースペクトルへ変換する。

10

【0058】

次いで、周波数ピンは臨界若しくは聴覚フィルタ・スケールで可能な帯域へ分類する。スピーチ信号を臨界帯域へ分割することは、人間の聴覚系（特に蝸牛）によく似ている。これらのフィルタは、概ね丸められた指数形を示して、等価長矩形帯域幅（ERB）スケールで均一に間隔をあけられる。このERBスケールは、音響心理学で用いられる単なる尺度であって、聴覚フィルタの帯域幅及び間隔を概算する。図2は1ERBの間隔を有するフィルタの適宜なセットを表しており、合計40の帯域がもたらされる。オーディオ・データの帯域化も可聴なアーチファクト（これは、原則としてピン毎に処理するとき生じる）を除去するのに役立つ。次いで、臨界帯域パワーを時間に対して平滑化する。即ち、隣接するブロックに亘って平滑化する。

20

【0059】

平滑化された臨界帯域のうちの最大出力を求めて、対応する利得を残りの（非最大）帯域について計算して、それらの出力を最大出力へ近似させる。利得補償は、基底膜の圧縮（非線形）特性に類似する。これらの利得は、飽和を避けるために、最大値へ制限される。これらの利得を原信号へ適用するためには、これらを変換してDFTフォーマットへ戻さねばならない。従って、帯域毎出力利得は最初に周波数ピン出力利得へ変換して戻し、次いでピン毎出力利得を各ピンの平方根を採ることにより振幅利得へ変換する。かくして原信号変換ピンには、計算されたピン毎振幅利得を乗じることができる。次いでスペクトル平坦化信号を変換して周波数領域から時間領域へ戻す。仮の中央の場合、これは時間領域へ復帰させるのに先立って、先ず原信号と混合する。図3はその処理を説明している。

30

上述のスペクトル平坦化システムは、入力された信号の特性を考慮していない。非スピーチ信号が平坦化されるならば、音質における知覚可能な変化は深刻なものとなる。非スピーチ信号の処理を避けるために、上述の方法は、発声活動検出器13に結び付けることができる。発声活動検出器13がスピーチの存在を示すとき、平坦化スピーチが用いられる。

【0060】

平坦化すべき信号は、上述のようにして周波数領域へ既に変換されていると仮定する。単純化のために、上記に用いられたチャンネル表記法は省略した。DFT係数を出力へ変換して、次いでDFT領域から臨界帯域へ変換する。

40

【数 2 2】

$$C_m[p] = \sum_{k=0}^{N-1} H[k,p] |X_m[k]|^2 \quad (29)$$

$$0 \leq p < P$$

【0061】

ここでH[k, p]はP臨界帯域フィルタである。

50

## 【 0 0 6 2 】

次いで各帯域における出力を、脳の皮質レベルで生じる時間積分と同様に、ブロック間で平滑化する。平滑化は、例えば、漏れ積分回路、非線形スムーザー、線形且つ多極ローパス・スムーザー、或いは更に精巧なスムーザーにより実行してもよい。この平滑化も遷移挙動（これは、利得にブロック間の急激な変動を引き起こし、可聴なポンピングをもたらす）を除去するのに役立つ。次にピーク出力は以下のように求められる。

## 【数 2 3】

$$E_m[p] = \lambda_2 E_{m-1}[p] + (1 - \lambda_2) C_m[p] \quad 0 \ll \lambda_2 < 1 \quad (30a)$$

10

$$E_{\max} = \max_p \{E_m[p]\} \quad (30b)$$

## 【 0 0 6 3 】

ここで  $E_m[p]$  は平滑化された臨界帯域出力、 $\lambda_2$  は漏れ積分回路係数、及び  $E_{\max}$  はピーク出力である。漏れ積分回路はローパス・フィルタリング効果を有しており、その利得は  $\lambda_2$  についての代表的な値が 0.9 である。

## 【 0 0 6 4 】

次に帯域毎出力利得を求め、最大利得を過度な補償を避けるように制限すると、

## 【数 2 4】

$$G_m[p] = \min \left\{ \left( \frac{E_{\max}}{E[p]} \right)^\gamma, G_{\max} \right\} \quad (31a)$$

$$0 < \gamma < 1 \quad (31b)$$

20

30

## 【 0 0 6 5 】

を得る。ここで  $G_m[p]$  は各帯域へ適用すべき出力利得、 $G_{\max}$  は許容できる最大出力利得であり、 $\gamma$  はスペクトルの平坦化の度を決定する。実際には、 $\gamma$  は 1 に近似する。指定された利得の量に対する他の汎用制限のみならず、システムが処理を実行するならば、 $G_{\max}$  はダイナミック・レンジ（又は無歪限界）に依存する。 $G_{\max}$  についての代表的な値は 20 dB である。

## 【 0 0 6 6 】

次に帯域毎出力利得をピン毎出力に変換して、平方根を採ってピン毎振幅利得を得る。

## 【数 2 5】

$$Y_m[k] = \sum_{p=0}^{P-1} [G_m[p] H[k, p]]^{1/2} \quad (32)$$

$$0 \leq k < K$$

40

## 【 0 0 6 7 】

ここで  $Y_m[k]$  はピン毎振幅利得である。

## 【 0 0 6 8 】

次に振幅利得を発声活動検出器出力  $2_1$ ,  $2_2$  に基づいて修正する。発声活動検出のた

50

めの方法を本発明の一つの実施例によって以下に説明する。

【0069】

発声活動検出

スペクトル束は、信号の出力スペクトルが変化する速度を測定し、オーディオの隣接するフレームの間の出力を比較する。(フレームは、オーディオ・データの複数のブロックである。)スペクトル束は、発声活動検出、或いは「スピーチ対オーディオ分類における他の判定がなされたもの」を示す。多くの場合、付加的な指標が用いられ、その結果は、オーディオが本当にスピーチであるか否かの判定をなすために集積される。

【0070】

一般に、スピーチのスペクトル束は音楽のそれよりも若干高い。即ち、音楽スペクトルは、フレーム間でスピーチ・スペクトルよりも安定する傾向にある。

10

【0071】

ステレオの場合、スペクトルの中央チャンネルが抽出されるところで、DFT係数は先ず中央と横のオーディオ(原ステレオから仮の中央を減じたもの)に分けられる。これは、伝統的な中間/横ステレオ処理とは異なっており、伝統的な中間/横ステレオ処理が一般に $(L+R)/2$ 、 $(L-R)/2$ であるのに対し、中央/横処理は $C$ 、 $L+R-2C$ である。

【0072】

上述したように周波数領域へ変換された信号によれば、DFT係数は出力へ変換されて、次いでDFT領域から臨界帯域領域へ変換される。臨界帯域出力は次いで中央と横との両方のスペクトル束を計算するのに用いられる。

20

【数26】

$$\tilde{X}_m[p] = \sum_{k=0}^{N-1} \left[ H[k,p] |X_m[k,2]|^2 \right]^{\frac{1}{2}} \quad (33a)$$

$$0 \leq p < P$$

$$\tilde{S}_m[p] = \sum_{k=0}^{N-1} \left[ H[k,p] |X_m[k,1] + X_m[k,3] - 2X_m[k,2]|^2 \right]^{\frac{1}{2}} \quad (33b)$$

30

$$0 \leq p < P$$

【0073】

ここで $X_m[p]$ は仮の中央の臨界帯域表現、 $S_m[p]$ は残りの信号(左と右との和から中央を減じたもの)の臨界帯域表現であり、 $H[k,p]$ は上述したようにP臨界帯域フィルタである。

【0074】

データの穿孔する2Jブロックから(中央及び横振幅について)二つのフレーム・バッファを形成する。

40

【数27】

$$\bar{X}_{new}(m,p) = \frac{1}{J} \sum_{l=m}^{m-J} \tilde{X}_l[p] \quad (34a) \quad \bar{X}_{old}(m,p) = \frac{1}{J} \sum_{l=m-J-1}^{m-2J} \tilde{X}_l[p] \quad (34b)$$

$$\bar{S}_{new}(m,p) = \frac{1}{J} \sum_{l=m}^{m-J} \tilde{S}_l[p] \quad (34c) \quad \bar{S}_{old}(m,p) = \frac{1}{J} \sum_{l=m-J-1}^{m-2J} \tilde{S}_l[p] \quad (34d)$$

50

【 0 0 7 5 】

次の段階は、現在のフレームと先行するフレームとの平均出力から中央チャンネルについての加重  $W$  を計算する。これは帯域の限られた範囲に亘ってなされる。

【 数 2 8 】

$$W(m) = \sum_{p=P_{start}}^{P_{end}} \frac{|\bar{X}_{new}(m, p)|^2 + |\bar{X}_{old}(m, p)|^2}{P_{end} - P_{start}} \quad 1 \leq P_{start} < P_{end} \leq P \quad (35)$$

【 0 0 7 6 】

10

帯域バンドの範囲は、スピーチの主要な帯域幅約 1 0 0 - 8 0 0 0 H z に限定される。中央と側方との両方についての非加重スペクトル束は次のように計算される。

【 0 0 7 7 】

【 数 2 9 】

$$F_X(m) = \sum_{p=P_{start}}^{P_{end}} |\bar{X}_{new}(m, p) - \bar{X}_{old}(m, p)|^2 \quad (36a)$$

$$F_S(m) = \sum_{p=P_{start}}^{P_{end}} |\bar{S}_{new}(m, p) - \bar{S}_{old}(m, p)|^2 \quad (36b)$$

20

【 0 0 7 8 】

ここで  $F_X(m)$  は中央の非加重スペクトル束であり、 $F_S(m)$  は側面の非加重スペクトル束である。

【 0 0 7 9 】

従ってスペクトル束の偏った推定値は以下のように計算される。

【 0 0 8 0 】

【 数 3 0 】

$$F_{\bar{X}}(m) > F_{\bar{S}}(m) \text{ and } W(m) > W_{min} \quad (37a)$$

30

【 0 0 8 1 】

であるならば、

【 数 3 1 】

$$F_{Tot}(m) = \frac{F_X(m) - F_S(m)}{2L \times W(m)} \quad (37b)$$

40

【 0 0 8 2 】

さもなければ、

【 数 3 2 】

$$F_{Tot}(m) = 0 \quad (37c)$$

【 0 0 8 3 】

である。ここで  $F_{Tot}(m)$  は全束推定値であり、 $W_{min}$  は許容される最小加重である。 $W_{min}$  はダイナミック・レンジに依存するが、代表的な値は  $W_{min} = 60 \text{ dB}$

50

Bである。

【 0 0 8 4 】

スペクトル束についての最終的な平滑化値は、単純な一次 I I R ローパス・フィルタにより  $F_{T o t}(m)$  の値をローパス・フィルタリングすることにより計算される。このフィルタは信号のサンプル・レート及びブロックの大きさに依存するが、一実施形態においては、 $f_s = 48 \text{ kHz}$  について  $0.025 * f_s$  の正規化カットオフを有する一次ローパス・フィルタにより規定できる。ここで  $f_s$  はデジタル・システムのサンプル・レートである。

【 0 0 8 5 】

$F_{T o t}(m)$  はかくして次の範囲に短縮される。即ち、

【 数 3 3 】

10

$$0 \leq F_{T o t}(m) \leq 1 :$$

【 0 0 8 6 】

であるから、

【 数 3 4 】

$$F_{T o t}(m) = \min\{\max\{0.0, F_{T o t}(m)\}, 1.0\} \quad (38)$$

20

【 0 0 8 7 】

(  $\min\{\}$  及び  $\max\{\}$  関数は、本実施例によれば  $F_{T o t}(m)$  を  $\{0, 1\}$  に制限する。 )

混合

平坦化された中央チャンネルは、発声活動検出器の出力に基づいて原オーディオン信号と混合される。

【 0 0 8 8 】

( 上述に示す通りの ) スペクトル平坦化についてのピン当りの振幅利得  $Y_m[k]$  は、

( 上述のように導かれるように ) 仮の中央チャンネル  $X_m[k, 2]$  へ適用される。

【 数 3 5 】

$$\begin{aligned} X_{temp} &= Y_m[k] X_m[k, 2] \\ X_m[k, 2] &= X_{temp} \end{aligned} \quad (39)$$

30

【 0 0 8 9 】

発声活動検出器 13 が、スピーチを検知するときは  $F_{T o t}(m) = 1$  とし、非スピーチを検知するときは  $F_{T o t}(m) = 0$  とする。0 と 1 との間の値が可能であり、これは発声活動検出器 13 がスピーチの存在について軟判定をなす場合に得られる。

40

【 0 0 9 0 】

左チャンネルについて、



【数 3 6】

$$\begin{aligned}
 X_{temp} &= (1 - F_{Tot}(m))X_m[k,1] + F_{Tot}(m)X_m[k,2] \\
 X_m[k,1] &= X_{temp} \\
 0 \leq F_{Tot}(m) &\leq 1
 \end{aligned}
 \tag{40a}$$

【0091】

同様に、右チャンネルについて、

【数 3 7】

$$\begin{aligned}
 X_{temp}' &= (1 - F_{Tot}(m))X_m[k,3] + F_{Tot}(m)X_m[k,2] \\
 X_m[k,3] &= X_{temp}' \\
 0 \leq F_{Tot}(m) &\leq 1
 \end{aligned}
 \tag{40b}$$

【0092】

実際には、 $F_{Tot}$  は値の狭い範囲に限定される。例えば

【数 3 8】

$$0.1 \leq F_{Tot}(m) \leq 0.9$$

【0093】

は最終的混合体における平坦化信号と原信号との両方の少量を保存する。

【0094】

次にピン毎振幅利得を原入力信号へ適用し、これを逆 DFT を介して変換して時間領域へ戻す。

【数 3 9】

$$\hat{x}[mN + n, c] = \frac{1}{N} \sum_{k=0}^{N-1} X_m[k, c] e^{\frac{j2\pi kn}{N}}$$

$$0 \leq n < N$$

$$c = 1, 3 \tag{41}$$

【0095】

ここで

【数 4 0】

 $\hat{x}$ 

【0096】

は  $x$  の増強された形態であり、原ステレオ入力信号である。

【0097】

図 4 は本発明の一実施例に係るコンピュータ 4 を示す。このコンピュータ 4 はメモリ 4 1、CPU 4 2 及びバス 4 3 を含む。バス 4 3 はメモリ 4 1 及び CPU 4 2 に交信するよ

10

20

30

40

50

うに接続する。メモリ 41 は上述に説明した任意の方法を実行するためのコンピュータ・プログラムを保存する。

【0098】

本発明の幾つかの実施形態について説明した。それでもなお、当業者には本発明の要旨及び目的から逸脱することなく、説明された実施形態に如何にして様々な修正を加えるかを理解されたい。例えば説明は離散フーリエ変換器を含むが、当業者には時間領域から周波数領域及びその逆の変換の様々な代替的方法を理解される。

【0099】

#### 従来技術

Schaub, A. and P., "Spectral sharpening for speech enhancement noise reduction", Proc. ICASSP. 1991, Toronto, Canada, May 1991, pp. 993 - 996.

Sondhi,, "New methods of pitch extraction", Audio and Electroacoustics, IEEE Transactions, June 1968, Volume 16, Issue 2, pp 262 - 266.

Villchur, E., "Signal Processing to Improve Speech Intelligibility for the Hearing Impaired", 99th Audio Engineering Society Convention, September 1995.

Thomas, I. and Niederjohn, R., "Preprocessing of Speech for Added Intelligibility in High Ambient Noise", 34th Audio Engineering Society Convention, March 1968.

Moore, B. et. al., "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness", J. Audio Eng. Soc, Vol. 45, No. 4, April 1997.

Moore, B. and Oxenham, A., "Psychoacoustic consequences of compression in the peripheral auditory system", The Journal of the Acoustical Society of America - December 2002 - Volume 112, Issue 6, pp. 2962 - 2966

従来技術 スペクトル平坦化

米国特許

米国特許第 6732073 B1号 発明の名称 "Spectral enhancement of acoustic signals to provide improved recognition of speech"

米国特許第 0993480 B1号 発明の名称 "Voice intelligibi

10

20

30

40

50

lity enhancement system”

米国特許2006/026320 A1号 発明の名称” Apparatus and method for noise reduction and speech enhancement with microphones and loudspeakers”

米国特許第07191122号 発明の名称” Speech compression system and method”

10

米国特許第2007/0094017号 発明の名称” Frequency domain format enhancement”

#### 国際特許

WO 2004/013840 A1号 発明の名称” Digital Signal Processing Techniques For Improving Audio Clarity And Intelligibility”

20

WO 2003/015082号 発明の名称” Sound Intelligibility Enhancement Using A Psychoacoustic Model And An Oversampled Filterbank”

#### 論文

Sallberg, B. et. al; ” Analog Circuit Implementation for Speech Enhancement Purposes Signals”; Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference.

30

Magotra, N. and Sirivara, S.; ” Real-time digital speech processing strategies for the hearing impaired”; Acoustics, Speech, and Signal Processing. 1997. ICASSP-97., 1997 page(s): 1211-1214 vol. 2

Walker, G., Byrne, D., and Dillon, H.; ” The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise”; The Journal of the Acoustical Society of America - September 1984 - Volume 76, Issue 3, pp. 746-757

40

#### 従来技術 中央抽出

Adobe Audition has a vocal/instrument ex

50

traction function  
http://www.adobeforums.eom/cgi-bin/webx/.3bc3a3e5

winampのための「中央カット」  
http://www.hydrogenaudio.org/forums/lofi  
version/index.php/t17450.html

従来技術 スペクトル束

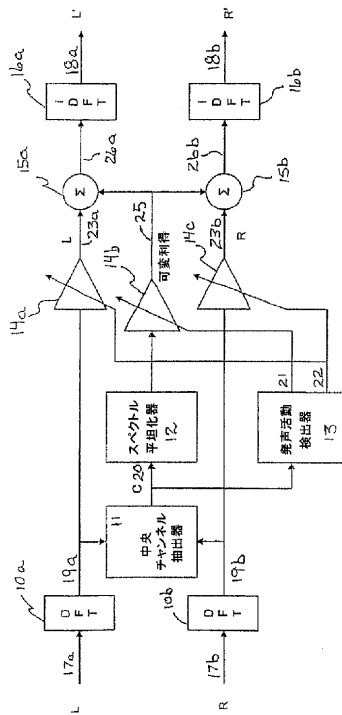
10

Vinton, M, and Robinson C; "Automated Speech/Other Discrimination for Loudness Monitoring," AES 118th Convention. 2005

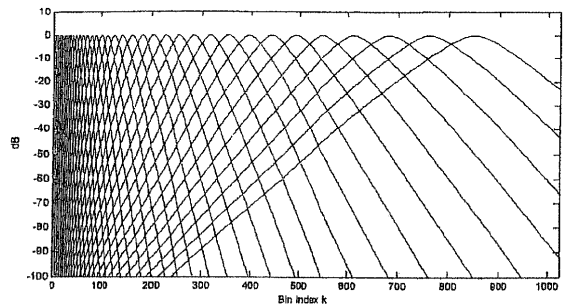
Scheirer E., and Slaney M., "Construction and evaluation of a robust multifeature speech/music discriminator", IEEE Transactions on Acoustics, Speech, and Signal Processing (ICASSP '97), 1997, pp. 1331 - 1334.

20

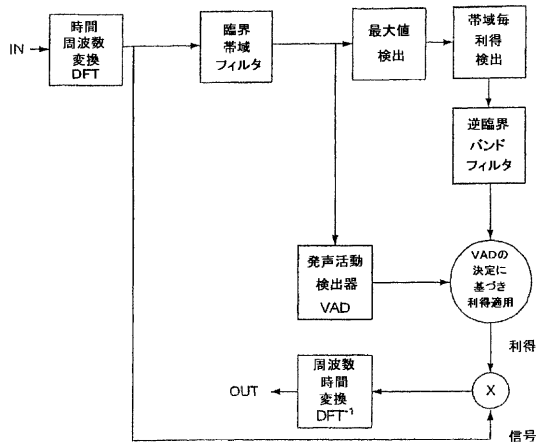
【図1】



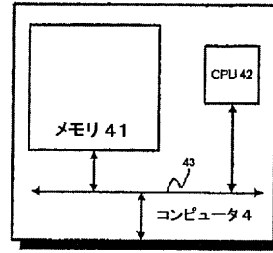
【図2】



【 図 3 】



【 図 4 】



【 手続補正書 】

【 提出日 】平成21年5月13日 (2009.5.13)

【 手続補正 1 】

【 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 】

第 1 のチャンネルと第 2 のチャンネルとを含む複数のチャンネルによりオーディオ信号から音の中央チャンネルを抽出する方法であって、

第 1 のチャンネルと第 2 のチャンネルとの合計から仮中央チャンネルを得て、

仮中央チャンネルの比より小さいオーディオ信号の第 1 のチャンネルと、仮中央チャンネルの比より小さいオーディオ信号の第 2 のチャンネルの共役とを乗じた積を計算しこの積を最小化するの値から抽出係数を得て、

前記仮中央チャンネルに前記抽出係数を乗じることにより、抽出された中央チャンネルを得ることを含む方法。

【 請求項 2 】

スピーチを増強する方法であって、

オーディオ信号の中央チャンネルを抽出して、

前記中央チャンネルにおけるスピーチ検出の信頼度を生成し、

前記中央チャンネルのスペクトルを平坦化して、

前記平坦化されたスピーチ・チャンネルに前記オーディオ信号を前記検出されたスピーチの信頼度に比例させて混合することにより、オーディオ信号における任意のスピーチを増強することを含む方法。

**【請求項 3】**

請求項 2 の方法において、

前記信頼度は、可能性確率が最も低いものから可能性確率が最も高いものまで変動すると共に、

前記生成は、

最低可能性確率よりも高く、且つ最高可能性確率よりも低い値に対して生成された信頼度を更に制限することを含む方法。

**【請求項 4】**

請求項 2 の方法において、前記抽出は、請求項 1 の方法を用いてオーディオ信号の中央チャンネルを抽出することを含む方法。

**【請求項 5】**

請求項 2 の方法において、前記抽出は、請求項 1 の方法を用いてオーディオ信号の中央チャンネルを抽出し、

前記平坦化は、オーディオ信号のスペクトルを平坦化するスペクトル平坦化方法を用いて中央チャンネルのスペクトルを平坦化することを含み、そのスペクトル平坦化方法は、推定されたスピーチ・チャンネルを知覚帯域に分離して、

知覚帯域のうちの何れが最も多くのエネルギーを有しているかを判定し、

より少ないエネルギーを有する知覚帯域の利得を増大させることにより、オーディオ信号における任意のスピーチのスペクトルを平坦化させることを含み、

前記生成は、オーディオ信号のスペクトルを平坦化するスペクトル平坦化方法を用いて中央チャンネルにおけるスピーチ検出の信頼度を生成することを含み、そのスペクトル平坦化方法は、

推定されたスピーチ・チャンネルを知覚帯域に分離して、

知覚帯域のうちの何れが最も多くのエネルギーを有しているかを判定し、

より少ないエネルギーを有する知覚帯域の利得を最大まで増大させることにより、オーディオ信号における任意のスピーチのスペクトルを平坦化させることを含む方法。

**【請求項 6】**

請求項 1 乃至 5 の何れか一項に記載の方法を実行するコンピュータ・プログラムが格納されたコンピュータ読み取り可能な記録媒体。

**【請求項 7】**

コンピュータ・システムであり、

CPUと、

請求項 6 の記録媒体と、

前記CPUと前記記録媒体とを結合するバスとを含むコンピュータ・システム。

**【請求項 8】**

スピーチエンハンサであって、

オーディオ信号の中央チャンネルを抽出するための中央チャンネル抽出器と、

前記中央チャンネルのスペクトルを平坦化するためのスペクトル平坦化器と、

前記中央チャンネルにおけるスピーチ検出における信頼度を生成するスピーチ信頼度生成器と、

前記平坦化されたスピーチ・チャンネルを原オーディオ信号に、検出されたスピーチの信頼度に比例して混合することにより、オーディオ信号における任意のスピーチを増強するミキサとを備えるスピーチエンハンサ。

## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2008/010591

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G10L21/02 H04S1/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) G10L H04S		
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, WPI Data, INSPEC		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOT J M ET AL: "Spatial Enhancement of Audio Recordings" PROCEEDINGS OF THE INTERNATIONAL AES CONFERENCE, XX, XX, 23 May 2003 (2003-05-23), pages 1-11, XP002401944 *Section 4.2, 5.2*figure 6	1, 12, 13
A	US 2007/041592 A1 (AVENDANO CARLOS M [US] ET AL) 22 February 2007 (2007-02-22) paragraphs [0036] - [0073]	1, 12, 13
X	WO 2004/049759 A (NOKIA CORP [FI]; KIRKEBY OLE [FI]) 10 June 2004 (2004-06-10) page 11, line 10 - line 24 page 12, line 2 - line 14 ----- -/-	5, 9, 10, 14
<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		*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
*E* earlier document but published on or after the international filing date		*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
*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)		*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.
*C* document referring to an oral disclosure, use, exhibition or other means		*G* document member of the same patent family
*P* document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search	Date of mailing of the international search report	
3 February 2009	11/02/2009	
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  Bensa, Julien	

## INTERNATIONAL SEARCH REPORT

International application No PCT/US2008/010591
---

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2003/161479 A1 (YANG CHINPING Q [US] ET AL) 28 August 2003 (2003-08-28) paragraph [0010]	5-11,14
A	WO 03/022003 A (KONINKL PHILIPS ELECTRONICS NV [NL]) 13 March 2003 (2003-03-13) page 3, line 19 - page 4, line 19	5-11,14



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2008/010591**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2.  As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
  
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  
1, 5-14
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

International Application No. PCT/US2008 /010591

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1,12,13

Extraction of center panned audio from 2 channel sources.  
----

2. claims: 2,3

Equalization of an audio signal spectrum  
----

3. claim: 4

Speech activity detection  
----

4. claims: 5-11,14

Speech enhancement.  
----

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No

PCT/US2008/010591

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2007041592 A1	22-02-2007	US 7257231 B1 US 2008247555 A1	14-08-2007 09-10-2008
WO 2004049759 A	10-06-2004	AU 2003282148 A1 CN 1714599 A EP 1566077 A1 FI 20022092 A KR 20050075029 A US 2004136554 A1	18-06-2004 28-12-2005 24-08-2005 23-05-2004 19-07-2005 15-07-2004
US 2003161479 A1	28-08-2003	NONE	
WO 03022003 A	13-03-2003	CN 1552171 A JP 2005502247 T US 2003044032 A1	01-12-2004 20-01-2005 06-03-2003

## 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, MT, NL, NO, PL, PT, RO, SE, SI, SK, 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, 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 100118647

弁理士 赤松 利昭

(74)代理人 100138519

弁理士 奥谷 雅子

(74)代理人 100138438

弁理士 尾首 亘聰

(74)代理人 100123892

弁理士 内藤 忠雄

(74)代理人 100131543

弁理士 常光 克明

(74)代理人 100159020

弁理士 安藤 麻子

(74)代理人 100097744

弁理士 東野 博文

(74)代理人 100161539

弁理士 武山 美子

(72)発明者 ブラウン、シー・フィリップ

アメリカ合衆国、カリフォルニア州 94103-4813、サンフランシスコ、ポトレロ・アベニュー 100、ドルビー・ラボラトリーズ・ライセンシング・コーポレーション内

Fターム(参考) 5D062 AA41