

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6162254号  
(P6162254)

(45) 発行日 平成29年7月12日(2017.7.12)

(24) 登録日 平成29年6月23日(2017.6.23)

(51) Int. Cl.

F I

G 1 O L 21/0364 (2013.01)

G 1 O L 21/0364

請求項の数 20 (全 30 頁)

(21) 出願番号	特願2015-551143 (P2015-551143)	(73) 特許権者	500341779
(86) (22) 出願日	平成25年8月23日 (2013. 8. 23)		フラウンホーファー・ゲゼルシャフト・ツール・フェルデルング・デル・アンゲヴァンテン・フォルシュング・アインゲトラー
(65) 公表番号	特表2016-505896 (P2016-505896A)		ゲネル・フェライン
(43) 公表日	平成28年2月25日 (2016. 2. 25)		ドイツ連邦共和国, 80686 ミュンヘン, ハンザシュトラッセ 27ツェー
(86) 国際出願番号	PCT/EP2013/067574	(74) 代理人	100085497
(87) 国際公開番号	W02014/108222		弁理士 筒井 秀隆
(87) 国際公開日	平成26年7月17日 (2014. 7. 17)	(72) 発明者	シェプケル, ヘニング
審査請求日	平成27年8月18日 (2015. 8. 18)		ドイツ連邦共和国 26121 オルデンブルク ナードルステル ストラッセ 38アー
(31) 優先権主張番号	61/750, 228		
(32) 優先日	平成25年1月8日 (2013. 1. 8)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 背景ノイズにおけるスピーチ了解度を増幅及び圧縮により向上させる装置と方法

(57) 【特許請求の範囲】

【請求項 1】

スピーチ入力信号から修正済みスピーチ信号を生成する装置であって、前記スピーチ入力信号は複数のスピーチサブバンド信号を含み、前記修正済みスピーチ信号は複数の修正済みサブバンド信号を含み、前記装置は、

前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 ( $s_n[k]$ ) のための重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

を、前記スピーチサブバンド信号 ( $s_n[k]$ ) の信号パワー ( $P_n[l]$ ) に依存して生成する、重み付け情報生成部 (110) と、

10

前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 ( $s_n[k]$ ) に対して当該スピーチサブバンド信号 ( $s_n[k]$ ) の前記重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

を適用することにより、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 ( $s_n[k]$ ) を修正し、前記複数の修正済みサブバンド信号の1つの修正済みサブバンド信号を取得する、信号修正部 (120) と、を含み、

前記複数のスピーチサブバンド信号のうちの第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、前記複数のスピーチサブバンド信号のうちの

20

第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅されるように、前記重み付け情報生成部(110)は前記複数のスピーチサブバンド信号の各々のための前記重み付け情報を生成し、かつ前記信号修正部(120)は前記スピーチサブバンド信号の各々を修正するよう構成され、

前記第1信号パワーが前記第2信号パワーよりも大きいとき、前記第1増幅度は前記第2増幅度よりも低く設定されている、装置。

【請求項2】

請求項1に記載の装置であって、

ノイズ入力信号の複数のノイズサブバンド信号のうちの1つのノイズサブバンド信号( $r_n[k]$ )が前記複数のスピーチサブバンド信号の各スピーチサブバンド信号( $s_n[k]$ )に対して割り当てられ、

前記重み付け情報生成部(110)は、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号( $s_n[k]$ )の前記重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

を、前記スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )のノイズスペクトルレベル( $d_n[l]$ )に依存して生成するよう構成されており、

前記重み付け情報生成部(110)は、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号( $s_n[k]$ )の前記重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

を、前記スピーチサブバンド信号のスピーチスペクトルレベル( $e_n[l]$ )に依存して生成するよう構成されている、装置。

【請求項3】

請求項2に記載の装置であって、前記重み付け情報生成部(110)は、前記スピーチサブバンド信号( $s_n[k]$ )の前記スピーチスペクトルレベル( $e_n[l]$ )と、前記スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )の前記ノイズスペクトルレベル( $d_n[l]$ )との信号対ノイズ比( $q(e_n, d_n)$ )を決定することにより、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号( $s_n[k]$ )の重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

を生成するよう構成されている、装置。

【請求項4】

請求項3に記載の装置であって、前記スピーチサブバンド信号( $s_n[k]$ )の前記スピーチスペクトルレベル( $e_n[l]$ )と、前記スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )の前記ノイズスペクトルレベル( $d_n[l]$ )との前記信号対ノイズ比( $q(e_n, d_n)$ )は次式により定義され、

【数8】

$$q(e_n, d_n) = \begin{cases} 0 & \text{if } e_n \leq d_n - 15 \text{ dB} \\ \frac{e_n - d_n + 15 \text{ dB}}{30 \text{ dB}} & \text{if } d_n - 15 \text{ dB} < e_n \leq d_n + 15 \text{ dB} \\ 1 & \text{if } e_n > d_n + 15 \text{ dB} \end{cases}$$

ここで、 $e_n$ は前記スピーチサブバンド信号( $s_n[k]$ )の前記スピーチスペクトルレベルであり、 $d_n$ は前記スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )の前記ノイズスペクトルレベルである、装置。

【請求項5】

請求項3又は4に記載の装置であって、

前記重み付け情報生成部 ( 1 1 0 ) は、スピーチ了解度指数

$$(\tilde{SII}[l])$$

を決定すること、及び、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 (  $s_n[k]$  ) について、当該スピーチサブバンド信号 (  $s_n[k]$  ) の前記スピーチスペクトルレベル (  $e_n[l]$  ) と、当該スピーチサブバンド信号 (  $s_n[k]$  ) の前記ノイズサブバンド信号 (  $r_n[k]$  ) の前記ノイズスペクトルレベル (  $d_n[l]$  ) との信号対ノイズ比 (  $q(e_n, d_n)$  ) を決定することによって、前記スピーチ入力信号の前記複数のスピーチサブバンド信号の前記重み付け情報

$$(W_n, W_{n,comp}, W_{n,lin}, \bar{W}_n)$$

10

を、生成するよう構成されており、

前記スピーチ了解度指数 ( S I I ) は前記スピーチ入力信号のスピーチ了解度を示す、装置。

【請求項 6】

請求項 5 に記載の装置であって、

前記重み付け情報生成部 ( 1 1 0 ) は、前記スピーチ了解度指数

$$\tilde{SII}[l]$$

を次式に従って決定するよう構成されており、

20

【数 9】

$$\tilde{SII}[l] = \sum_{n=1}^N i_n \cdot q(e_n[l], d_n[l]) \cdot \min \left\{ 1 - \frac{d_n[l] + 15 \text{ dB} - u_n - 10 \text{ dB}}{160 \text{ dB}}, 1 \right\},$$

ここで、 $n$  は前記複数のスピーチサブバンド信号の  $n$  番目のスピーチサブバンド信号を示し、 $N$  は前記スピーチサブバンド信号の総数を示し、 $l$  はブロックを示し、 $q(e_n, d_n)$  は、前記  $n$  番目のスピーチサブバンド信号 (  $s_n[k]$  ) のスピーチスペクトルレベル (  $e_n[l]$  ) と、前記  $n$  番目のスピーチサブバンド信号 (  $s_n[k]$  ) の前記ノイズサブバンド信号 (  $r_n[k]$  ) のノイズスペクトルレベル (  $d_n[l]$  ) との信号対ノイズ比を示し、 $u_n$  は固定値であるスピーチスペクトルレベルを示し、 $i_n$  は帯域重要度を示す、装置。

30

【請求項 7】

請求項 5 又は 6 に記載の装置であって、前記重み付け情報生成部 ( 1 1 0 ) は、前記スピーチ了解度指数

$$(\tilde{SII}[l])$$

に依存し、前記スピーチサブバンド信号 (  $s_n[k]$  ) の信号パワー (  $p_n[l]$  ) に依存し、かつ前記複数のスピーチサブバンド信号の全てのスピーチサブバンド信号の信号パワーの合計 (  $p_{(max)}[l]$  ) に依存して、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 (  $s_n[k]$  ) について線形ゲイン (  $w_{n,(lin)}$  ) を決定することにより、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 (  $s_n[k]$  ) の前記重み付け情報を生成するよう構成されている、装置。

40

【請求項 8】

請求項 7 に記載の装置であって、前記重み付け情報生成部 ( 1 1 0 ) は、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 (  $s_n[k]$  ) についての線形ゲイン  $w_{n,(lin)}$  を、次式に従って生成するよう構成され、

【数 1 0】

$$w_{n,(lin)} [l] = \sqrt{\frac{\phi_n^{\hat{S}II} [l] \cdot \phi_{(max)} [l]}{\sum_{\lambda=1}^N \phi_{\lambda}^{\hat{S}II} [l]} \cdot \phi_n [l]}$$

ここで、 $n$  は前記複数のスピーチサブバンド信号の  $n$  番目のスピーチサブバンド信号を示し、 $N$  は前記スピーチサブバンド信号の総数を示し、 $l$  はブロックを示し、 $\phi_n [l]$  は  $n$  番目のスピーチサブバンド信号の信号パワーを示し、 $\phi_{(max)} [l]$  は前記複数のスピーチサブバンド信号の全てのスピーチサブバンド信号の信号パワーの合計を示す、装置

10

【請求項 9】

請求項 3 乃至 6 のいずれか一項に記載の装置であって、  
前記重み付け情報生成部 (110) は、圧縮比  $cr_n [l]$  を次式に従って決定するよう構成されており、

【数 1 1】

$$cr_n [l] = \max \left\{ cr_{(max)} \cdot (1 - q(e_n [l], d_n [l])), 1 \right\}$$

ここで、 $q(e_n [l], d_n [l])$  は、前記スピーチスペクトルレベルと前記ノイズスペクトルレベルとの信号対ノイズ比であり、前記信号対ノイズ比  $q(e_n [l], d_n [l])$  は 0 と 1 の間のある数値を示し、 $cr_{(max)}$  はある固定値を示し、 $l$  はあるブロックを示す、装置。

20

【請求項 10】

請求項 7 又は 8 に記載の装置であって、  
前記重み付け情報生成部 (110) は、圧縮比  $cr_n [l]$  を次式に従って決定するよう構成されており、

【数 1 2】

$$cr_n [l] = \max \left\{ cr_{(max)} \cdot (1 - q(e_n [l], d_n [l])), 1 \right\}$$

30

ここで、 $q(e_n [l], d_n [l])$  は、前記スピーチスペクトルレベルと前記ノイズスペクトルレベルとの信号対ノイズ比であり、前記信号対ノイズ比  $q(e_n [l], d_n [l])$  は 0 と 1 の間のある数値を示し、 $cr_{(max)}$  はある固定値を示し、 $l$  はあるブロックを示す、装置。

【請求項 11】

請求項 9 又は 10 に記載の装置であって、  
前記重み付け情報生成部 (110) は、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 ( $s_n [k]$ ) の前記重み付け情報を、当該サブバンド信号 ( $s_n [k]$ ) の圧縮ゲイン  $w_{n,(comp)}$  を次式に従って決定することにより、生成するよう構成され、

40

【数 1 3】

$$w_{n,(comp)} [l \cdot M - m] = \sqrt{\left( \frac{\phi_n [l]}{\hat{s}_n^2 [l \cdot M - m]} \right)^{(cr_n[l]-1)/cr_n[l]}, \quad m = 0, \dots, M-1}$$

ここで、 $M$  はブロック  $l$  の長さを示し、 $s_n [k]$  は前記スピーチサブバンド信号 ( $s_n [k]$ ) の信号パワーを示し、

$$\hat{s}_n^2 [l \cdot M - m]$$

50

は前記スピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定の二乗を示す、装置。

【請求項 1 2】

請求項 1 1 に記載の装置であって、

前記重み付け情報生成部 ( 1 1 0 ) は、前記スピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定

$$\hat{s}[k]$$

を次式に従って決定するよう構成され、

【数 1 4】

$$\hat{s}_n[k] = \begin{cases} \hat{s}_n[k-1] \cdot \alpha_a + (1 - \alpha_a) \cdot |s_n[k]| & \text{if } |s_n[k]| \geq \hat{s}_n[k-1] \\ \hat{s}_n[k-1] \cdot \alpha_r + (1 - \alpha_r) \cdot |s_n[k]| & \text{if } |s_n[k]| < \hat{s}_n[k-1] \end{cases}$$

ここで、 $s_n[k]$  は前記スピーチサブバンド信号を示し、 $|s_n[k]|$  は当該スピーチサブバンド信号の振幅を示し、 $\alpha_a$  は第 1 平滑化定数であり、 $\alpha_r$  は第 2 平滑化定数である、装置。

【請求項 1 3】

請求項 1 乃至 1 0 のいずれか一項に記載の装置であって、前記重み付け情報生成部 ( 1 1 0 ) は、次式を適用することで、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号 ( $s_n[k]$ ) の前記重み付け情報

$$\bar{w}_n$$

を生成するよう構成されており、

【数 1 5】

$$\bar{w}_n[l \cdot M - m] = \alpha_p \bar{w}_n[l \cdot M - m - 1] + (1 - \alpha_p) p_{\bar{\lambda}_n[l]} (\hat{s}_n^2[l \cdot M - m])$$

ここで、 $n$  は前記複数のスピーチサブバンド信号の  $n$  番目のスピーチサブバンド信号を示し、 $N$  はスピーチサブバンド信号の総数を示し、 $l$  はブロックを示し、 $\alpha_p$  は平滑化定数であり、

$$\hat{s}_n^2[l \cdot M - m]$$

は前記スピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定の二乗を示し、

$$p_{\bar{\lambda}_n[l]} (\hat{s}_n^2[l \cdot M - m])$$

は

$$\bar{\lambda}_n[l]$$

の線形のインターポレーションおよびエクストラポレーションを実行する関数を示し、

$$\bar{\lambda}_n[l]$$

は平滑化された入力 - 出力 - 特性を示す、装置。

【請求項 1 4】

請求項 1 乃至 1 3 のいずれか一項に記載の装置であって、全てのスピーチサブバンド信号の全てのスピーチ信号パワー ( $\bar{w}_n[1]$ ) の第 1 合計と、全ての修正済みサブバンド信号の全てのスピーチ信号パワーの第 2 合計との違いが 2 0 % よりも小さくなるように、前記重み付け情報生成部 ( 1 1 0 ) は前記複数のスピーチサブバンド信号の各々について

10

20

30

40

50

前記重み付け情報を生成し、かつ前記信号修正部(120)は前記スピーチサブバンド信号の各々を修正するよう構成されている、装置。

【請求項15】

請求項2に記載の装置であって、前記重み付け情報生成部(110)は、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号( $s_n[k]$ )の前記重み付け情報を、重み付きの加算( $a_n[l]$ )を決定することで生成するよう構成されており、前記重み付きの加算は、当該スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )の前記ノイズスペクトルレベル( $d_n[l]$ )に依存し、かつ残響スペクトルレベル( $z_n[l]$ )に依存している、装置。

【請求項16】

請求項15に記載の装置であって、前記重み付け情報生成部(110)は、ラウドスピーカとマイクロホンとの間の室内インパルス応答に依存し、残響時間 $T60$ に依存し、又は直接対残響のエネルギー比に依存して、前記残響スペクトルレベル( $z_n[l]$ )を生成するよう構成されている、装置。

【請求項17】

請求項15又は16に記載の装置であって、前記重み付け情報生成部(110)は、前記重み付きの加算 $a_n[l]$ を次式に従って決定するよう構成されており、

【数16】

$$a_n[l] = \beta z_n[l] + d_n[l]$$

ここで、 $d_n[l]$ は前記スピーチサブバンド信号( $s_n[k]$ )の前記ノイズサブバンド信号( $r_n[k]$ )のノイズスペクトルレベルであり、 $z_n[l]$ は前記残響スペクトルレベルを示し、 $\beta$ は実数である、装置。

【請求項18】

請求項1乃至17のいずれか一項に記載の装置であって、前記装置は、第1フィルタバンク(105)と第2フィルタバンク(125)とを更に備え、

前記第1フィルタバンク(105)は、時間ドメインで表現されている未処理のスピーチ信号を、時間ドメインからサブバンドドメインへと変換して、前記複数のスピーチサブバンド信号を含む前記スピーチ入力信号を取得するよう構成されており、

前記第2フィルタバンク(125)は、前記サブバンドドメインで表現され且つ前記複数の修正済みサブバンド信号を含む前記修正済みスピーチ信号を、前記サブバンドドメインから前記時間ドメインへと変換して、時間ドメインの出力信号を取得するよう構成されている、装置。

【請求項19】

スピーチ入力信号から修正済みスピーチ信号を生成する方法であって、前記スピーチ入力信号は複数のスピーチサブバンド信号を含み、前記修正済みスピーチ信号は複数の修正済みサブバンド信号を含み、前記方法は、

前記複数のスピーチサブバンド信号の各スピーチサブバンド信号のための重み付け情報を、当該スピーチサブバンド信号の信号パワーに依存して生成するステップと、

前記各スピーチサブバンド信号に対して当該スピーチサブバンド信号の前記重み付け情報を適用することで、前記複数のスピーチサブバンド信号の各スピーチサブバンド信号を修正して、複数の修正済みサブバンド信号の1つの修正済みサブバンド信号を取得するステップと、を含み、

前記複数のスピーチサブバンド信号の各々のための前記重み付け情報を生成するステップと、前記スピーチサブバンド信号の各々を修正するステップとは、前記複数のスピーチサブバンド信号のうちの第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、前記複数のスピーチサブバンド信号のうちの第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅されるように実行され、

前記第1信号パワーが前記第2信号パワーよりも大きいとき、前記第1増幅度は前記第2増幅度よりも低く設定されている、方法。

10

20

30

40

50

## 【請求項 20】

コンピュータ又はプロセッサ上で操作されたとき、請求項 19 に記載の方法を実行する、コンピュータプログラム。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、オーディオ信号処理に関し、特に、背景ノイズにおけるスピーチ了解度(speech intelligibility)を増幅及び圧縮により向上させる装置と方法に関するものである。

## 【背景技術】

## 【0002】

多くのスピーチ伝達アプリケーション(例えば鉄道駅の構内放送システムや携帯電話など)においては、付加的ノイズ及びノ又は残響によって妨害される状況下でも高いスピーチ了解度を維持することが非常に重要である。その目標を維持する一つの簡易な手法は、出力する前にスピーチ信号を増幅して、良好な信号対ノイズ比(SNR)を達成することである。しかし、そのような簡易な増幅は、増幅システムの技術的な制限や不快なほど高い音響レベルにより、不可能である場合が多い。従って、スピーチ了解度を向上させる一方で、入力時のパワーと比較して同等の出力パワーを維持するようなアルゴリズムが望まれる。本発明は、全体のスピーチレベルを増大させることなく、付加的なノイズを有するシナリオ内でスピーチ了解度を向上させることが可能なアルゴリズムを含む。

## 【0003】

簡易な増幅器を超える信号処理の指針については、非特許文献 1、非特許文献 2、非特許文献 3、非特許文献 5 及び非特許文献 6 に開示されている。

## 【0004】

しかしながら、スピーチ伝達アプリケーションのための改善された信号処理概念が提供されるとすれば、大きな意義があるであろう。

## 【先行技術文献】

## 【非特許文献】

## 【0005】

【非特許文献 1】ANSI (1997). Methods for calculation of the speech intelligibility index. American National Standard ANSI S3.5-1997 (American National Standards Institute, Inc.), New York, USA.

【非特許文献 2】Sauert, B. and Vary, P. (2010). Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement. In Proc. of ITG-Fachtagung Sprachkommunikation.(Bochum, Germany, Oct. 6-8, 2010), volume 9.

【非特許文献 3】Sauert, B. and Vary, P. (2012). Near-end listening enhancement in the presence of bandpass noises. In Proc. of ITG-Fachtagung Sprachkommunikation. (Braunschweig, Germany, Sept. 26-28, 2012).

【非特許文献 4】Vaidyanathan, P., Mitra, S., and Neuvo, Y. (1986). A new approach to the realization of low-sensitivity iir digital filters. Acoustics, Speech and Signal Processing, IEEE Transactions on, 34(2):350 - 361.

【非特許文献 5】Zorila, T.-C., Kandia, V., and Stylianou, Y. (2012a). Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression. In 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest Romania.

【非特許文献 6】Zorila, T.-C., Kandia, V., and Stylianou, Y. (2012b). Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. In Proceedings of Interspeech 2012 (Portland, USA).

## 【発明の概要】

## 【発明が解決しようとする課題】

10

20

30

40

50

## 【0006】

本発明の目的は、スピーチ伝達アプリケーションのための改善された信号処理概念を提供することである。本発明のこの目的は、請求項1に記載の装置と、請求項19に記載の方法と、請求項20に記載のコンピュータプログラムにより達成される。

## 【課題を解決するための手段】

## 【0007】

スピーチ入力信号から修正済みスピーチ信号を生成する装置が提供される。そのスピーチ入力信号は、複数のスピーチサブバンド信号を含む。修正済みスピーチ信号は複数の修正済みサブバンド信号を含む。その装置は、複数のスピーチサブバンド信号の各スピーチサブバンド信号のための重み付け情報を、当該スピーチサブバンド信号の信号パワーに依存して生成する重み付け情報生成部を含む。さらにその装置は、スピーチサブバンド信号に対して当該スピーチサブバンド信号の重み付け情報を適用することで、複数のスピーチサブバンド信号の各スピーチサブバンド信号を修正して、複数の修正済みサブバンド信号の各修正済みサブバンド信号を取得する信号修正部を含む。複数のスピーチサブバンド信号のうち第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度(first degree)を用いて増幅され、複数のスピーチサブバンド信号のうち第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度(second degree)を用いて増幅されるように、重み付け情報生成部は複数のスピーチサブバンド信号の各々のための重み付け情報を生成し、かつ信号修正部はスピーチサブバンド信号の各々を修正するよう構成され、第1信号パワーが第2信号パワーよりも大きいとき、第1増幅度は第2増幅度よりも低く設定されている。

## 【0008】

複数のスピーチサブバンド信号のうち第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、また、複数のスピーチサブバンド信号のうち第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅される。第1増幅度が第2増幅度よりも低い場合には、例えば第1スピーチサブバンド信号を増幅した結果として得られる第1修正済みサブバンド信号の信号パワーと第1スピーチサブバンド信号の信号パワーとの比が、第2スピーチサブバンド信号を増幅した結果として得られる第2修正済みサブバンド信号の信号パワーと第2スピーチサブバンド信号の信号パワーとの比よりも低いことを意味している。

## 【0009】

本提案の概念を用いた実施形態は、時間及び周波数依存のゲイン特性と時間及び周波数依存の圧縮特性とを組み合わせてもよく、両特性は共に推定されたスピーチ了解度指数(SII)の関数である。そのゲインは、スピーチレベルが一定に保持される一方で了解度が最大化されるように、現在のノイズ信号に依存してスピーチ信号を適応的に前処理するために使用されてもよい。

## 【0010】

この概念が使用される技術システム、例えば対応するアルゴリズムが作動している技術システムに依るが、スピーチレベルを更に変化させる目的で、この概念(例えばアルゴリズム)を一般的な音量制御と組み合わせてもよく、又は組み合わせなくてもよい。以下に、アルゴリズムの一つの可能な実現例の詳細な説明を提供する。

## 【0011】

個々のステップの具体的なパラメータ又は機能は修正可能であり、当業者であればそのような修正を認識することができるであろう。

## 【0012】

スピーチ入力信号から修正済みスピーチ信号を生成する方法が提供される。そのスピーチ入力信号は複数のスピーチサブバンド信号を含む。修正済みスピーチ信号は複数の修正済みサブバンド信号を含む。その方法は、  
- 複数のスピーチサブバンド信号の各スピーチサブバンド信号のための重み付け情報を、当該スピーチサブバンド信号の信号パワーに依存して生成するステップと、



- 前記スピーチサブバンド信号に対して当該スピーチサブバンド信号の重み付け情報を適用することで、複数のスピーチサブバンド信号の各スピーチサブバンド信号を修正して、複数の修正済みサブバンド信号の1つの修正済みサブバンド信号を取得するステップと、を含む。

【0013】

複数のスピーチサブバンド信号の各々のための重み付け情報を生成するステップとスピーチサブバンド信号の各々を修正するステップとは、複数のスピーチサブバンド信号のうちの第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、複数のスピーチサブバンド信号のうちの第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅されるように実行され、第1信号パワーは第2信号パワーよりも大きく、第1増幅度は第2増幅度よりも低い。

10

【0014】

更に、コンピュータ又は信号プロセッサ上で作動するとき上述の方法を実行するコンピュータプログラムが提供される。

【0015】

好ましい実施形態は従属請求項において提供される。

【0016】

以下に、本発明の実施形態について図面を参照しながら更に詳細に説明する。

【図面の簡単な説明】

【0017】

20

【図1】修正済みスピーチ信号を生成する一実施形態に係る装置を示す。

【図2】修正済みスピーチ信号を生成する他の実施形態に係る装置を示す。

【図3a】スピーチサブバンド信号の増幅が行われる前のスピーチサブバンド信号のスピーチ信号パワーを示す。

【図3b】スピーチサブバンド信号の増幅の結果として得られた修正済みサブバンド信号のスピーチ信号パワーを示す。

【図4a】修正済みスピーチ信号を生成する更なる実施形態に係る装置を示す。

【図4b】修正済みスピーチ信号を生成する他の実施形態に係る装置を示す。

【図5a】記述されたアルゴリズムの一実施形態に係るフローチャートを示す。

【図5b】記述されたアルゴリズムの他の実施形態に係るフローチャートを示す。

30

【図6】一実施形態に係る近端(near-end)リスニング強化が備えられた信号モデルを示す。

【図7】1~16000Hzの中心周波数についての長期間スピーチレベルを示す。

【図8】主観的評価から得られた結果を示す。

【図9】主観的結果に関する相関分析を示す。

【発明を実施するための形態】

【0018】

図1は、スピーチ入力信号から修正済みスピーチ信号を生成する、一実施形態に係る装置を示す。スピーチ入力信号は複数のスピーチサブバンド信号を含む。修正済みスピーチ信号は複数の修正済みサブバンド信号を含む。

40

【0019】

その装置は、複数のスピーチサブバンド信号の各スピーチサブバンド信号のための重み付け情報を、当該スピーチサブバンド信号の信号パワーに依存して生成する重み付け情報生成部110を含む。

【0020】

さらにその装置は、各スピーチサブバンド信号に対してそのスピーチサブバンド信号の重み付け情報を適用することで、複数のスピーチサブバンド信号の各スピーチサブバンド信号を修正して、複数の修正済みサブバンド信号の各修正済みサブバンド信号を取得する信号修正部120を含む。

【0021】

50

複数のスピーチサブバンド信号のうちの第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、複数のスピーチサブバンド信号のうちの第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅されるように、重み付け情報生成部110は複数のスピーチサブバンド信号の各々のための重み付け情報を生成し、かつ信号修正部120はスピーチサブバンド信号の各々を修正するよう構成され、第1信号パワーは第2信号パワーよりも大きく、第1増幅度は第2増幅度よりも低い。

【0022】

図3aと図3bとは、この状態を更に詳細に示す。特に、図3aはスピーチサブバンド信号の増幅が行われる前のスピーチサブバンド信号のスピーチ信号パワーを示す。図3bはスピーチサブバンド信号の増幅の結果として得られた修正済みサブバンド信号のスピーチ信号パワーを示す。

10

【0023】

図3aと図3bとは、第1スピーチサブバンド信号のオリジナルの第1信号パワー311が増幅され、その増幅によって低減され、その結果として、第1スピーチサブバンド信号のより小さい第1信号パワー321が得られる実施形態を示す。第2スピーチサブバンド信号のオリジナルの第2信号パワー312は増幅され、その増幅によって増強され、その結果として、第2スピーチサブバンド信号のより大きな第2信号パワー322が得られる。このように、第1スピーチサブバンド信号は第1増幅度を用いて増幅されており、第2スピーチサブバンド信号は第2増幅度を用いて増幅されており、第1増幅度は第2増幅度よりも低い。第1スピーチサブバンド信号の第1のオリジナルの信号パワーは、第2スピーチサブバンド信号の第2のオリジナルの信号パワーよりも大きいものであった。

20

【0024】

図3aと図3bとにおいて、第1及び第3のスピーチサブバンド信号の各信号パワー311及び313は増幅によって低減されており、第2、第4及び第5のスピーチサブバンド信号の各信号パワー312、314及び315は増幅によって増強されている。このように、第1及び第3のスピーチサブバンド信号の各信号パワー311及び313は、第2、第4及び第5のスピーチサブバンド信号を増幅するのに用いられる増幅度よりも低い増幅度を用いてそれぞれ増幅されている。第1及び第3のスピーチサブバンド信号のオリジナルの各信号パワー311及び313は、第2、第4及び第5のスピーチサブバンド信号のオリジナルの各信号パワー312、314及び315よりも大きいものであった。

30

【0025】

更に、図3aと図3bとにおいて、第2スピーチサブバンド信号のオリジナル信号パワー312は、第4スピーチサブバンド信号のオリジナル信号パワー314よりも大きいことが分かる。第2スピーチサブバンド信号と第4スピーチサブバンド信号との両方が増幅によって増強されているが、第2サブバンド信号は、第4サブバンド信号を増幅するのに用いられた増幅度よりも低い増幅度を用いて増幅されている。なぜなら、第2スピーチサブバンド信号における修正済み(増幅済み)信号パワー322とオリジナル信号パワー312との比が、第4スピーチサブバンド信号における修正済み(増幅済み)信号パワー324とオリジナル信号パワー314との比よりも低いからである。

【0026】

40

例えば、第2スピーチサブバンド信号の修正済み(増幅済み)信号パワー322は、第2スピーチサブバンド信号のオリジナル信号パワー312の大きさの2倍であり、従って、第2スピーチサブバンド信号における修正済み信号パワー322とオリジナル信号パワー312との比は2である。第4スピーチサブバンド信号の修正済み(増幅済み)信号パワー324は、第4スピーチサブバンド信号のオリジナル信号パワー314の大きさの3倍であり、従って、第4スピーチサブバンド信号における修正済み信号パワー324とオリジナル信号パワー314との比は3である。

【0027】

更に、図3aと図3bとにおいて、第3スピーチサブバンド信号のオリジナル信号パワー313は、第1スピーチサブバンド信号のオリジナル信号パワー311よりも大きいこ

50

とが分かる。第3スピーチサブバンド信号と第1スピーチサブバンド信号との両方が増幅によって低減されているが、第3のサブバンド信号は、第1のサブバンド信号を増幅するのに用いられた増幅度よりも低い増幅度を用いて増幅されている。なぜなら、第3スピーチサブバンド信号における修正済み（増幅済み）信号パワー323とオリジナル信号パワー313との比が、第1スピーチサブバンド信号における修正済み（増幅済み）信号パワー321とオリジナル信号パワー311との比よりも低いからである。

【0028】

例えば、第3スピーチサブバンド信号の修正済み（増幅済み）信号パワー323は、第3スピーチサブバンド信号のオリジナル信号パワー313の大きさの67%であり、従って、第3スピーチサブバンド信号における修正済み信号パワー323とオリジナル信号パワー313との比は0.67である。第1スピーチサブバンド信号の修正済み（増幅済み）信号パワー321は、第1スピーチサブバンド信号のオリジナル信号パワー311の大きさの71%であり、従って、第1スピーチサブバンド信号における修正済み信号パワー321とオリジナル信号パワー311との比は0.71である。

10

【0029】

例えば、修正済みサブバンド信号を得るべくスピーチサブバンド信号を増幅するために用いられた増幅度は、修正済みサブバンド信号の信号パワーとスピーチサブバンド信号の信号パワーとの比である。

【0030】

複数のスピーチサブバンド信号のうちの第1信号パワーを有する第1スピーチサブバンド信号が第1増幅度を用いて増幅され、かつ、複数のスピーチサブバンド信号のうちの第2信号パワーを有する第2スピーチサブバンド信号が第2増幅度を用いて増幅され、第1増幅度が第2増幅度よりも低い場合には、例えばこれは、第1スピーチサブバンド信号の増幅の結果として得られた第1修正済みサブバンド信号の信号パワーと第1スピーチサブバンド信号の信号パワーとの比が、第2スピーチサブバンド信号の増幅の結果として得られた第2修正済みサブバンド信号の信号パワーと第2スピーチサブバンド信号の信号パワーとの比よりも低いことを意味する。

20

【0031】

一実施形態によれば、全てのスピーチサブバンド信号の全てのスピーチ信号パワー（ $\sum_{n=1}^N P_n$ ）の第1合計と、全ての修正済みサブバンド信号の全てのスピーチ信号パワーの第2合計との違いが20%よりも小さくなるように、重み付け情報生成部110は複数のスピーチサブバンド信号の各々について重み付け情報を生成し、かつ信号修正部120はスピーチサブバンド信号の各々を修正するよう構成されてもよい。

30

【0032】

換言すれば、全てのスピーチサブバンド信号の全てのスピーチ信号パワー（ $\sum_{n=1}^N P_n$ ）の第1合計を、全ての修正済みサブバンド信号の全てのスピーチ信号パワーの第2合計で除算すると、その結果が値dとなり、ここで、 $0.8 \leq d \leq 1.2$ が真となる。

【0033】

図2は、修正済みスピーチ信号を生成する、他の実施形態にかかる装置を示す。

【0034】

図2の装置と図1の装置との相違点は、図2の装置が第1フィルタバンク105と第2フィルタバンク125とを更に備える点である。

40

【0035】

第1フィルタバンク105は、時間ドメインで表現されている未処理のスピーチ信号を、時間ドメインからサブバンドドメインへと変換して、複数のスピーチサブバンド信号を含むスピーチ入力信号を取得するよう構成されている。

【0036】

第2フィルタバンク125は、サブバンドドメインで表現され且つ複数の修正済みサブバンド信号を含む修正済みスピーチ信号を、サブバンドドメインから時間ドメインへと変換して、時間ドメインの出力信号を取得するよう構成されている。

50

## 【 0 0 3 7 】

図 4 a は、修正済みスピーチ信号を生成する、更なる実施形態にかかる装置を示す。

## 【 0 0 3 8 】

図 2 の実施形態とは対照的に、図 4 a の装置は第 3 フィルタバンク 1 0 8 を更に含み、この第 3 フィルタバンク 1 0 8 は、時間ドメインのノイズ参照 (noise reference)  $r [ k ]$  を、時間ドメインからサブバンドドメインへと変換して、ノイズ入力信号の複数のノイズサブバンド信号  $r_n [ k ]$  を取得する。

## 【 0 0 3 9 】

更に、この実施形態にかかる重み付け情報生成部 1 1 0 をより詳細に説明する。この重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々についてのスピーチ信号パワーを以下のように計算する、スピーチ信号パワー計算部 1 3 1 を含む。更に、重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々についてのスピーチスペクトルレベルを以下のように計算する、スピーチスペクトルレベル計算部 1 3 2 を含む。更に、重み付け情報生成部 1 1 0 は、ノイズ入力信号のノイズサブバンド信号の各々についてのノイズスペクトルレベルを以下のように計算する、ノイズスペクトルレベル計算部 1 3 3 を含む。

## 【 0 0 4 0 】

一実施形態において、ノイズ入力信号の複数のノイズサブバンド信号のうちの 1 つのノイズサブバンド信号  $r_n [ k ]$  が、複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n [ k ]$  に割り当てられている。例えば、各ノイズサブバンド信号が同じサブバンドのスピーチサブバンド信号に割り当てられている。重み付け情報生成部 1 1 0 は、複数のスピーチサブバンド信号のうちの各スピーチサブバンド信号  $s_n [ k ]$  の重み付け情報を、当該スピーチサブバンド信号 ( $s_n [ k ]$ ) のノイズサブバンド信号  $r_n [ k ]$  のノイズスペクトルレベル  $d_n [ l ]$  に依存して生成するよう構成されている。更に、重み付け情報生成部 1 1 0 は、複数のスピーチサブバンド信号のうちの各スピーチサブバンド信号  $s_n [ k ]$  の重み付け情報を、当該スピーチサブバンド信号のスピーチスペクトルレベル  $e_n [ l ]$  に依存して生成するよう構成されている。

## 【 0 0 4 1 】

更に、重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々について信号対ノイズ比を以下のように計算する、SNR 計算部 1 3 4 を含む。

## 【 0 0 4 2 】

例えば、一実施形態によれば、重み付け情報生成部 1 1 0 は、複数のスピーチサブバンド信号のうちの各スピーチサブバンド信号  $s_n [ k ]$  の重み付け情報を、当該スピーチサブバンド信号  $s_n [ k ]$  のスピーチスペクトルレベル  $e_n [ l ]$  と、スピーチサブバンド信号  $s_n [ k ]$  のノイズサブバンド信号  $r_n [ k ]$  のノイズスペクトルレベル  $d_n [ l ]$  との信号対ノイズ比を決定することで、生成するよう構成されている。例えば、スピーチサブバンド信号  $s_n [ k ]$  のスピーチスペクトルレベル  $e_n [ l ]$  と、スピーチサブバンド信号  $s_n [ k ]$  のノイズサブバンド信号  $r_n [ k ]$  のノイズスペクトルレベル  $d_n [ l ]$  との信号対ノイズ比  $q ( e_n , d_n )$  は、次式により定義されてもよい。

## 【 0 0 4 3 】

## 【 数 1 】

$$q(e_n, d_n) = \begin{cases} 0 & \text{if } e_n \leq d_n - 15 \text{ dB} \\ \frac{e_n - d_n + 15 \text{ dB}}{30 \text{ dB}} & \text{if } d_n - 15 \text{ dB} < e_n \leq d_n + 15 \text{ dB} \\ 1 & \text{if } e_n > d_n + 15 \text{ dB} \end{cases}$$

## 【 0 0 4 4 】

ここで、 $e_n$  は前記スピーチサブバンド信号  $s_n [ k ]$  のスピーチスペクトルレベルであり、 $d_n$  は前記スピーチサブバンド信号  $s_n [ k ]$  のノイズサブバンド信号  $r_n [ k ]$  のノ

10

20

30

40

50

イズスペクトルレベルである。

【 0 0 4 5 】

更に、重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々についての圧縮比を以下のように計算する、圧縮比計算部 1 3 5 を含む。

【 0 0 4 6 】

例えば、一実施形態によれば、重み付け情報生成部 1 1 0、例えば圧縮比計算部 1 3 5 は、圧縮比  $cr_n[l]$  を次式に従って決定するよう構成されている。

【 0 0 4 7 】

【 数 2 】

$$cr_n[l] = \max \left\{ cr_{(\max)} \cdot (1 - q(e_n[l], d_n[l])), 1 \right\}$$

10

【 0 0 4 8 】

ここで、 $q(e_n[l], d_n[l])$  は、前記スピーチスペクトルレベルの信号対ノイズ比であり、その信号対ノイズ比  $q(e_n[l], d_n[l])$  は 0 と 1 の間のある数値を示し、 $cr_{(\max)}$  はある固定値を示し、 $l$  はあるブロックを示す。 $n$  はスピーチサブバンド信号のうちの一つ ( $n$  番目のスピーチサブバンド信号) を示す。

【 0 0 4 9 】

スピーチサブバンド信号の各々が複数のブロックを含み得ることに留意すべきである。ここで、 $l$  は  $n$  番目のスピーチサブバンド信号の複数のブロックのうちの一つのブロックを示す。複数のブロックの各ブロックは、スピーチサブバンド信号の複数のサンプルを含んでもよい。

20

【 0 0 5 0 】

更に、重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々について、スピーチ信号振幅の包絡の平滑化された推定(smoothed estimate)を以下のように計算する、平滑化済み信号振幅計算部 1 3 6 を含む。

【 0 0 5 1 】

例えば、一実施形態において、重み付け情報生成部 1 1 0、例えば平滑化済み信号振幅計算部 1 3 6 が、スピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定

$$\hat{s}[k]$$

30

を次式に従って決定するよう構成されてもよい。

【 0 0 5 2 】

【 数 3 】

$$\hat{s}_n[k] = \begin{cases} \hat{s}_n[k-1] \cdot \alpha_a + (1 - \alpha_a) \cdot |s_n[k]| & \text{if } |s_n[k]| \geq \hat{s}_n[k-1] \\ \hat{s}_n[k-1] \cdot \alpha_r + (1 - \alpha_r) \cdot |s_n[k]| & \text{if } |s_n[k]| < \hat{s}_n[k-1] \end{cases}$$

【 0 0 5 3 】

ここで、 $s_n[k]$  は前記スピーチサブバンド信号を示し、 $|s_n[k]|$  は前記スピーチサブバンド信号の振幅を示し、 $\alpha_a$  は第 1 平滑化定数であり、 $\alpha_r$  は第 2 平滑化定数である。

40

【 0 0 5 4 】

更に、重み付け情報生成部 1 1 0 は、スピーチサブバンド信号の各々について圧縮ゲインを以下のように計算する、圧縮ゲイン計算部 1 3 7 を含む。

【 0 0 5 5 】

例えば、重み付け情報生成部 1 1 0 は、複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  の重み付け情報を、例えば圧縮ゲイン計算部 1 3 7 を使用して、前記サブバンド信号 ( $s_n[k]$ ) の圧縮ゲイン  $w_{n,(\text{comp})}$  を次式に従って決定することにより、生成するよう構成されている。

50

【 0 0 5 6 】

【 数 4 】

$$w_{n,(comp)}[l \cdot M - m] = \sqrt{\left(\frac{\Phi_n[l]}{\hat{s}_n^2[l \cdot M - m]}\right)^{(cr_n[l]-1)/cr_n[l]}}, \quad m = 0, \dots, M-1$$

【 0 0 5 7 】

ここで、Mはブロック1の長さを示し、 $\hat{s}_n[l]$ は前記スピーチサブバンド信号 $s_n[k]$ の信号パワーを示し、

$$\hat{s}_n^2[l \cdot M - m]$$

は前記スピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定の二乗を示す。

【 0 0 5 8 】

$\hat{s}_n[l]$ は、長さMの(完全な)ブロック1についてのスピーチサブバンド信号 $s_n[k]$ のスピーチ信号パワーを示してもよく、

$$\hat{s}_n^2[l \cdot M - m]$$

はそのブロックの特定サンプルのスピーチ信号振幅の包絡の平滑化された推定の二乗を示してもよい。圧縮、例えば音量の大きなサンプルが低減される一方で、静かなサンプルが増大する。

【 0 0 5 9 】

更に、重み付け情報生成部110は、スピーチ了解度指数を以下のように計算する、スピーチ了解度指数計算部138を含む。

【 0 0 6 0 】

例えば、一実施形態において、重み付け情報生成部110、例えばスピーチ了解度指数計算部138は、スピーチ了解度指数

$$\hat{S}ll[l]$$

を次式に従って決定するよう構成されてもよい。

【 0 0 6 1 】

【 数 5 】

$$\hat{S}ll[l] = \sum_{n=1}^N i_n \cdot q(e_n[l], d_n[l]) \cdot \min\left\{1 - \frac{d_n[l] + 15 \text{ dB} - u_n - 10 \text{ dB}}{160 \text{ dB}}, 1\right\},$$

【 0 0 6 2 】

ここで、nは複数のスピーチサブバンド信号のうちのn番目のスピーチサブバンド信号を示し、Nはスピーチサブバンド信号の総数を示し、lはブロックを示し、 $q(e_n, d_n)$ は、n番目のスピーチサブバンド信号 $s_n[k]$ のスピーチスペクトルレベル $e_n[l]$ と、n番目のスピーチサブバンド信号 $s_n[k]$ のノイズサブバンド信号 $r_n[k]$ のノイズスペクトルレベル $d_n[l]$ との信号対ノイズ比を示し、 $u_n$ は固定値であるスピーチスペクトルレベルを示し、 $i_n$ は帯域重要度(band importance)を示す。

【 0 0 6 3 】

更に、重み付け情報生成部110は、スピーチサブバンド信号の各々について以下に説明するように線形ゲインを計算する、線形ゲイン計算部139を含む。

【 0 0 6 4 】

例えば、一実施形態によれば、重み付け情報生成部110はスピーチ了解度指数

10

20

30

40

$\hat{S}_{II}[l]$ 

を決定し、更に、複数のスピーチサブバンド信号の各スピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  について、スピーチサブバンド信号  $s_n[k]$  のスピーチスペクトルレベル  $e_n[l]$  と、スピーチサブバンド信号  $s_n[k]$  のノイズサブバンド信号  $r_n[k]$  のノイズスペクトルレベル  $d_n[l]$  との信号対ノイズ比を決定することにより、スピーチ入力信号の複数のスピーチサブバンド信号の重み付け情報を生成するよう構成されてもよい。スピーチ了解度指数  $S_{II}$  はスピーチ入力信号のスピーチ了解度を示している。

【0065】

例えば、重み付け情報生成部 110 は、例えば線形ゲイン計算部 139 を使用して、スピーチ了解度指数

 $\hat{S}_{II}[l]$ 

と、スピーチサブバンド信号  $s_n[k]$  の信号パワー  $\phi_n[l]$  と、複数のスピーチサブバンド信号の全てのスピーチサブバンド信号の信号パワーの合計 ( $\phi_{(max)}[l]$ ) とに依存して、複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  について線形ゲイン  $w_{n,(lin)}$  を決定することにより、複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  の重み付け情報を生成するよう構成されてもよい。

【0066】

例えば、重み付け情報生成部 110 は、複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  についての線形ゲイン  $w_{n,(lin)}$  を、次式に従って生成するよう構成されてもよい。

【0067】

【数6】

$$w_{n,(lin)}[l] = \sqrt{\frac{\phi_n^{\hat{S}_{II}}[l] \cdot \phi_{(max)}[l]}{\sum_{\lambda=1}^N \phi_{\lambda}^{\hat{S}_{II}}[l] \cdot \phi_n[l]}}$$

【0068】

ここで、 $n$  は複数のスピーチサブバンド信号のうちの  $n$  番目のスピーチサブバンド信号を示し、 $N$  はスピーチサブバンド信号の総数を示し、 $l$  はブロックを示し、 $\phi_n[l]$  は  $n$  番目のスピーチサブバンド信号の信号パワーを示し、 $\phi_{(max)}[l]$  は複数のスピーチサブバンド信号の全てのスピーチサブバンド信号の信号パワーの合計を示す。例えば、 $\phi_{(max)}[l]$  はブロック  $l$  内のスピーチ信号の広帯域パワーを示す。

【0069】

上述した式を理解し易くするために、ブロック  $l$  に対する

 $\hat{S}_{II}[l]$ 

の依存性は明示されていない。しかしながら、

 $\hat{S}_{II}[l]$ 

はブロック  $l$  に依存している点に留意すべきである。

【0070】

 $\hat{S}_{II}[l]$ 

は 0 (了解度なし) と 1 (完全な了解度) との間のある指数であってもよい。上述した  $w_{n,(lin)}$  の式について、

10

20

30

40

$$\tilde{S}_{11}[l] = 0 \text{ 及び } \tilde{S}_{11}[l] = 1$$

のような極端な場合を以下に考察する。

【 0 0 7 1 】

$$\tilde{S}_{11}[l] = 1$$

の場合、第 1 項の分子と第 2 項の分母とが同一になり、従って上述の  $w_{n,(lin)}$  の式から除去され得る。更に、

$$\tilde{S}_{11}[l] = 1$$

の場合、第 2 項の分子と第 1 項の分母とが同一になり、従ってまた上述の  $w_{n,(lin)}$  の式から除去され得る。このようにスピーチ了解度が完全である場合、 $w_{n,(lin)}$  が 1 となり、例えばその信号は修正されない。

【 0 0 7 2 】

$$\tilde{S}_{11}[l] = 0$$

の場合、第 1 項は  $1/N$  となり、その結果、例えば全体のパワーは  $N$  個の全ての周波数帯域の中で均等に分配される。

【 0 0 7 3 】

図 5 a は一実施形態に従うアルゴリズムのフローチャートを示す。

【 0 0 7 4 】

ステップ 1 4 1 では、時間ドメインで表現された未処理のスピーチ信号  $s_n[k]$  が時間ドメインからサブバンドドメインへと変換されて、サブバンドドメインで表現されたスピーチ入力信号が取得され、そのスピーチ入力信号は複数のスピーチサブバンド信号  $s_n[k]$  を含む。

【 0 0 7 5 】

ステップ 1 4 2 では、時間ドメインで表現された時間ドメイン・ノイズ参照  $r[k]$  が時間ドメインからサブバンドドメインへと変換されて、複数のノイズサブバンド信号  $r_n[k]$  が取得される。

【 0 0 7 6 】

ステップ 1 5 1 においては、以下に説明するように、スピーチサブバンド信号の各々についてスピーチ信号パワーが計算される。更に、ステップ 1 5 2 では、以下に説明するように、スピーチサブバンド信号の各々についてスピーチスペクトルレベルが計算される。更に、ステップ 1 5 3 では、以下に説明するように、スピーチサブバンド信号の各々についてノイズスペクトルレベルが計算される。更に、ステップ 1 5 4 では、以下に説明するように、スピーチサブバンド信号の各々について信号対ノイズ比が計算される。更に、ステップ 1 5 5 では、以下に説明するように、スピーチサブバンド信号の各々について圧縮比が計算される。更に、ステップ 1 5 6 では、以下に説明するように、スピーチサブバンド信号の各々についてスピーチ信号振幅の包絡の平滑化された推定が計算される。更に、ステップ 1 5 7 では、以下に説明するように、スピーチサブバンド信号の各々について圧縮ゲインが計算される。更に、ステップ 1 5 8 では、以下に説明するように、スピーチ了解度指数が計算される。更に、ステップ 1 5 9 では、以下に説明するように、スピーチサブバンド信号の各々について線形ゲインが計算される。

【 0 0 7 7 】

ステップ 1 6 1 において、以下に説明するように、それぞれのスピーチサブバンド信号に対して、スピーチサブバンド信号の圧縮ゲインを適用し、かつスピーチサブバンド信号の線形ゲインを適用することで、複数のスピーチサブバンド信号が増幅される。

【 0 0 7 8 】

ステップ 1 6 2 において、複数の修正済みサブバンド信号を有する修正済みスピーチ信号がサブバンドドメインから時間ドメインへと変換され、時間ドメインの出力信号

10

20

30

40

50



$\tilde{s}[k]$ 

が取得される。

【 0 0 7 9 】

図 4 b は、修正済みスピーチ信号を生成する、他の実施形態にかかる装置を示す。

【 0 0 8 0 】

図 4 b に示す実施形態においては、室内音響的情報が提案のアルゴリズム内で考慮されてもよい。スピーチ信号はラウドスピーカによって再生され、妨害されたスピーチ信号がマイクロホンによって集音される。録音された信号は、ノイズ  $r[k]$  と残響スピーチ信号とを含む。残響スピーチ信号の中に含まれた残響のいくつかの部分は有害であると認められ得る一方で、他の部分はスピーチ了解度にとって有益であると考えられ得る。例えばラウドスピーカとマイクロホンとの間の室内インパルス応答をモデル化するフィルタである室内音響的情報生成部 (R I G)、残響時間  $T_{60}$  (60 db 減衰する時間として定義される)、又は直接対残響エネルギー比 (D R R) を使用して、残響スペクトルレベル  $z_n[l]$  が、重み付け情報生成部 110、例えば残響スペクトルレベル計算部 163 により、室内音響的情報生成部から提供された情報と各サブバンドにおけるサブバンドスピーチ信号  $s_n[k]$  とを用いて計算されてもよい。重み付けファクタ を用いて重み付きの加算  $a_n[l]$  が決定され得る。

【 0 0 8 1 】

【 数 7 】

$$a_n[l] = \beta z_n[l] + d_n[l]$$

【 0 0 8 2 】

この加算は、重み付け情報生成部 110、例えば重み付き加算部 164 により決定されてもよく、また重み付き加算  $a_n[l]$  は後続の計算において使用されてもよく、その他の場合にはノイズスペクトルレベル  $d_n[l]$  だけが使用される。

【 0 0 8 3 】

$d_n$  について定義された全ての式は、 $d_n$  を  $a_n$  に置き換えることで、 $a_n$  についても適用可能である。例えば、幾つかの実施形態によれば、式 (4)、式 (5) 及び / 又は式 (8) 内において、 $d_n$  は  $a_n$  で置き換えることができ、その置き換えにより、これらの式は重み付きの加算  $a_n$  を考慮に入れることができる。

【 0 0 8 4 】

例えば、 は実数であってもよく、ここで、例えば 0 1 が適用されてもよい。

【 0 0 8 5 】

本質的に、 $a_n$  は残響に関する追加的情報 (例えば室内インパルス応答、 $T_{60}$ 、D R R など) を考慮してもよい。

【 0 0 8 6 】

以下に、実施形態の概念、とりわけ図 1、図 2、図 4 a、図 4 b、図 5 a 及び図 5 b の実施形態で用いられる概念について、より詳細に説明する。

【 0 0 8 7 】

本アルゴリズムの入力における明瞭なスピーチ信号 (「未処理のスピーチ信号」とも称される) は、離散的な時間指数  $k$  において  $s[k]$  により表される。

【 0 0 8 8 】

ノイズ参照 (例えば時間ドメインで表現されている) は  $r[k]$  により表され、参照マイクロホンを用いて録音され得る。

【 0 0 8 9 】

両方の信号は、あるフィルタバンク、例えばデシメーションなしの I I R フィルタバンクにより、オクターブ帯域に分割される。この点に関しては、例えば V a i d y a n a t h a n らによる非特許文献 4 を参照されたい。結果として得られるサブバンド信号は、 $s[k]$  と  $r[k]$  とについて、それぞれ  $s_n[k]$  と  $r_n[k]$  とで表される。

10

20

30

40

50

## 【 0 0 9 0 】

長さMのブロックlについてサブバンドスピーチ信号パワー  $\phi_n[l]$  は、次式(1)により計算される。

$$\phi_n[l] = \frac{1}{M} \sum_{k=lM-M+1}^{lM} s_n^2[k] \quad (1)$$

## 【 0 0 9 1 】

式(1)と中心周波数  $f_n$  を有するオクターブ帯域の帯域幅  $\Delta f_n$  とを用いて、等価スピーチスペクトルレベルが計算され得る。

$$e_n[l] = 10 \cdot \log_{10} \left( \frac{\phi_n[l]}{\Delta f_n} \right) \quad (2)$$

## 【 0 0 9 2 】

同様の計算がノイズサブバンド信号  $r_n[k]$  (「ノイズ参照信号」とも称される)についても実行され、等価ノイズスペクトルレベルが導かれる。

$$d_n[l] = 10 \cdot \log_{10} \left( \frac{1}{M \cdot \Delta f_n} \sum_{k=lM-M+1}^{lM} r_n^2[k] \right) \quad (3)$$

## 【 0 0 9 3 】

次に、各ブロックについて信号対ノイズ比(SNR)のマッピングが演算され得る。

$$q(e_n, d_n) = \begin{cases} 0 & \text{if } e_n \leq d_n - 15 \text{ dB} \\ \frac{e_n - d_n + 15 \text{ dB}}{30 \text{ dB}} & \text{if } d_n - 15 \text{ dB} < e_n \leq d_n + 15 \text{ dB} \\ 1 & \text{if } e_n > d_n + 15 \text{ dB} \end{cases} \quad (4)$$

## 【 0 0 9 4 】

この式(4)からのマッピング関数を使用して、各周波数チャンネル内の圧縮比が予め決定された最大圧縮比  $cr_{(\max)}$  を使用して計算され得る。最大圧縮比は、典型的には  $cr_{(\max)} = 8$  の値に設定される。

$$cr_n[l] = \max \left\{ cr_{(\max)} \cdot (1 - q(e_n[l], d_n[l])), 1 \right\}. \quad (5)$$

## 【 0 0 9 5 】

更に、スピーチ信号振幅の瞬時包絡の平滑化された推定が以下のように計算される。

$$\hat{s}_n[k] = \begin{cases} \hat{s}_n[k-1] \cdot \alpha_a + (1 - \alpha_a) \cdot |s_n[k]| & \text{if } |s_n[k]| \geq \hat{s}_n[k-1] \\ \hat{s}_n[k-1] \cdot \alpha_r + (1 - \alpha_r) \cdot |s_n[k]| & \text{if } |s_n[k]| < \hat{s}_n[k-1] \end{cases} \quad (6)$$

ここで、 $\alpha_a$  と  $\alpha_r$  とは、増大する信号振幅の場合と減少する信号振幅の場合とにおける平滑化定数をそれぞれ表している。

## 【 0 0 9 6 】

$\phi_n[l]$ 、 $cr_n[l]$ 、及び

$$\hat{s}_n[k]$$

を使用して、圧縮ゲイン  $w_{n,(\text{comp})}[k]$  が以下のように計算される。

10

20

30

40

50

$$w_{n,(comp)} [l \cdot M - m] = \sqrt{\left(\frac{\Phi_n [l]}{\hat{s}_n^2 [l \cdot M - m]}\right)^{(cr_n[l]-1)/cr_n[l]}}, \quad m = 0, \dots, M-1, \quad (7)$$

ここで、 $l \cdot M - m = k$ である。

【0097】

更に、スピーチ了解度指数 (SII) の推定が以下のように計算される。

$$\tilde{SII} [l] = \sum_{n=1}^N i_n \cdot q(e_n [l], d_n [l]) \cdot \min \left\{ 1 - \frac{d_n [l] + 15 \text{ dB} - u_n - 10 \text{ dB}}{160 \text{ dB}}, 1 \right\}, \quad (8) \quad 10$$

ここで、 $u_n$ は、ANSI (1997) に従って標準的等価スピーチスペクトルレベルとして定義される。例えば、 $u_n$ は固定値であってもよい。

【0098】

ここで、 $N$ は例えばサブバンドの総数を示す。 $i_n$ は例えば帯域重要度の関数であり、例えば $n$ 番目のサブバンドについての帯域重要度を示しており、ここで、 $i_n$ は例えば0~1の間の値であり、 $N$ 個の全てのサブバンドの $i_n$ の値は合計して1となる。

【0099】

以下の項目は、SauertとVary (2010) (非特許文献2を参照) から採用したものである。 20

$$\min \left\{ 1 - \frac{d_n [l] + 15 \text{ dB} - u_n - 10 \text{ dB}}{160 \text{ dB}}, 1 \right\}$$

【0100】

SII値は例えば0~1の間の値であり、1は非常に良好なスピーチ了解度を示し、0は非常に劣悪なスピーチ了解度を示す。

【0101】

この推定されたSIIを使用して、いわゆる線形ゲイン関数が計算される。

$$w_{n,(lin)} [l] = \sqrt{\frac{\Phi_n^{\tilde{SII}} [l]}{\sum_{\lambda=1}^N \Phi_{\lambda}^{\tilde{SII}} [l]} \cdot \frac{\Phi_{(\max)} [l]}{\Phi_n [l]}}. \quad (9) \quad 30$$

【0102】

上述の式(9)が理解し易くなるように、ブロック1に対する

$\tilde{SII}$

の依存性は明示されていない。しかしながら、

$\tilde{SII}$

はブロック1に依存している点に留意すべきである。 40

【0103】

$\Phi_{(\max)} [l]$ は、複数のスピーチサブバンド信号の全てのサブバンド信号の信号パワーの合計を示す。例えば、 $\Phi_{(\max)} [l]$ は、ブロック1の中のスピーチ信号の広帯域パワーを示す。

【0104】

次に、両方のゲイン関数が組み合わされて、サブバンド信号がそれぞれのゲイン関数で乗算される。即ち、

$$\tilde{s}_n[lM - m] = s_n[lM - m] w_{n,lin}[l] w_{n,comp}[lM - m] \quad (10)$$

$$w_n[lM - m] = w_{n,lin}[l] w_{n,comp}[lM - m] \quad (11)$$

従って、式(10)は次式と等価となる。

$$\tilde{s}_n[lM - m] = s_n[lM - m] w_n[lM - m] \quad (12)$$

10

【0105】

一実施例によれば、ここで逆フィルタバンクが適用され、修正済みスピーチ信号が再構築される。

【0106】

しかし他の実施形態によれば、逆フィルタバンクを適用して修正済みスピーチ信号を生成する前に、特にブロックの境界におけるゲイン関数内の急速な変化を防止するために、平滑化処理が  $w_n[lM - m]$  に対して適用される。

【0107】

一実施形態において、重み付け情報生成部110は、次式を適用することにより複数のスピーチサブバンド信号の各スピーチサブバンド信号  $s_n[k]$  の重み付け情報

20

$\bar{w}_n$

を生成するよう構成されている。

$$\bar{w}_n[l \cdot M - m] = \alpha_p \bar{w}_n[l \cdot M - m - 1] + (1 - \alpha_p) p_{\lambda_n[l]} (\hat{s}_n^2[l \cdot M - m])$$

【0108】

ここで、 $n$ は複数のスピーチサブバンド信号の  $n$ 番目のスピーチサブバンド信号を示し、 $N$ はスピーチサブバンド信号の総数を示し、 $l$ はブロックを示し、 $p$ は平滑化定数であり、

30

$$\hat{s}_n^2[l \cdot M - m]$$

はスピーチサブバンド信号のスピーチ信号振幅の包絡の平滑化された推定の二乗を示す。

【0109】

以下に、一実施形態にかかる平滑化について説明する。

【0110】

平滑化は、 $w_n[lM - m]$ の基底にある入力-出力-特性(IOC)に対して適用される。その入力-出力-特性は、パラメータベクトル  $\lambda_n[l]$ の一部である入力パワー  $\gamma_{n,i}[l]$ および出力パワー  $\xi_{n,i}[l]$ の集合により定義される。即ち、

$$\lambda_n[l] = [\gamma_{n,1}[l] \quad \gamma_{n,2}[l] \quad \gamma_{n,3}[l] \quad \xi_{n,1}[l] \quad \xi_{n,2}[l] \quad \xi_{n,3}[l]] \quad (13)$$

40

【0111】

次に、入力-出力-特性は以下の式で定義される。

$$\gamma_{n,1}[l] = 1 \quad (14)$$

$$\gamma_{n,2}[l] = \Phi_n[l] \quad (15)$$

$$\gamma_{n,3}[l] = v \quad (16)$$

及び

50

$$\xi_{n,1}[l] = w_{n,(lin)}[l] (\phi_n[l])^{(1-1/cr_n[l])} \quad (17)$$

$$\xi_{n,2}[l] = w_{n,(lin)}[l] \phi_n[l] \quad (18)$$

$$\xi_{n,3}[l] = w_{n,(lin)}[l] (\phi_n[l])^{(1-1/cr_n[l])} \nu^{1/cr_n[l]} \quad (19)$$

## 【 0 1 1 2 】

ここで、は dB FS を dB SPL へと変換する。例えば、0 dB FS が 100 dB SPL と等しいと仮定すると、 $\nu = 10^{(100/10)}$  となる。I O C の線形のインターポーレーションおよびエクストラポーレーションを実行する関数

$$p\lambda_n[l] (\hat{s}_n^2[l \cdot M - m])$$

は、例えば、現在の入力パワー

$$\hat{s}_n^2[l \cdot M - m]$$

に依存してデシベルドメインで上述のパラメータによって定義され、前記現在の入力パワーとしては、例えば式(6)に従って定義されるように、スピーチ信号振幅の包絡の平滑化された推定がある。従って、次式のように記述できる。

$$w_n[l \cdot M - m] = p\lambda_n[l] (\hat{s}_n^2[l \cdot M - m]) \quad (20)$$

## 【 0 1 1 3 】

次に、パラメータベクトル  $\lambda_n[l]$  の各要素  $\lambda_{n,j}[l]$  に対し、再帰的な平滑化が適用されて、次式が導かれる。

$$\bar{\lambda}_{n,j}[l] = \alpha_\lambda \bar{\lambda}_{n,j}[l-1] + (1 - \alpha_\lambda) \lambda_{n,j}[l] \quad (21)$$

$$\bar{\lambda}_n[l]$$

は平滑化されたパラメータベクトルであり、 $\alpha_\lambda$  は平滑化定数である。

## 【 0 1 1 4 】

次に、平滑化されたゲインが次式のように計算される。

$$\bar{w}_n[l \cdot M - m] = \alpha_p \bar{w}_n[l \cdot M - m - 1] + (1 - \alpha_p) p\bar{\lambda}_n[l] (\hat{s}_n^2[l \cdot M - m]) \quad (22)$$

ここで  $\alpha_p$  は、ゲイン関数を時間にわたって更に平滑化するための平滑化定数である。

## 【 0 1 1 5 】

$$p\bar{\lambda}_n[l] (\hat{s}_n^2[l \cdot M - m])$$

は、平滑化された入力 - 出力 - 特性

$$\bar{\lambda}_n[l]$$

の線形のインターポーレーションおよびエクストラポーレーションを実行する関数として定義され、

$$\bar{\lambda}_n[l]$$

は例えば式(13)と式(21)とにより定義されている通りである。

## 【 0 1 1 6 】

次に、出力信号は次式をもたらす。

10

20

30

40

50

$$\tilde{s}_n[lM - m] = s_n[lM - m] \bar{w}_n[lM - m] \quad (23)$$

【 0 1 1 7 】

最後に、逆フィルタバンクが適用されて、修正済みスピーチ信号

$$\tilde{s}[k]$$

が再構築される。

【 0 1 1 8 】

入力パワーと出力パワーとの差を低減するために、アルゴリズムの出力と入力において、各ブロック内のパワーが平滑化されたパワー推定を用いて正規化される。従って、平滑化された入力パワーは次式のように定義される。

10

$$\tilde{\phi}_s[l] = \alpha_L \tilde{\phi}_s[l-1] + (1 - \alpha_L) \phi_s[l], \quad (24)$$

【 0 1 1 9 】

ここで、 $\alpha_L$ は平滑化定数であり、 $s[l]$ は、サブバンド信号ではなく広帯域入力信号  $s[k]$  を用いて式(1)に従って計算されたものである。次に、平滑化された出力パワー

$$\tilde{\phi}_{\tilde{s}}[l]$$

20

が本アルゴリズムの出力信号

$$\tilde{s}[k]$$

を使用して計算される。

【 0 1 2 0 】

次に、再生されるべき信号が次式のように計算される。

$$\tilde{s}'[lM - m] = \sqrt{\frac{\tilde{\phi}_s[l]}{\tilde{\phi}_{\tilde{s}}[l]}} \tilde{s}[lM - m] \quad (25)$$

30

【 0 1 2 1 】

これらの実施形態は従来技術と比べて幾つかの点で異なっている。

【 0 1 2 2 】

例えば、幾つかの実施形態では、多帯域スペクトル整形アルゴリズムと多帯域圧縮スキームとが組み合わされており、これらは、多帯域スペクトル整形アルゴリズムと単一帯域圧縮スキームとが組み合わせられた Z o r i l a らの手法(2012a, b)(非特許文献5、非特許文献6を参照)とは対照的である。

【 0 1 2 3 】

本提案の概念は、従来技術とは対照的に、線形ゲインと圧縮ゲインとを組み合わせ、その線形ゲインと圧縮ゲインとの両方が時間変化し、瞬時スピーチ信号とノイズ信号とに適応する。

40

【 0 1 2 4 】

更に、幾つかの実施形態は、静的な圧縮スキームを使用した Z o r i l a らの手法(2012a, b)(非特許文献5、非特許文献6を参照)とは対照的に、各周波数帯域内で適応的な圧縮比を適用する。

【 0 1 2 5 】

更に、幾つかの実施形態によれば、その圧縮比は、S I I を計算するために使用され従ってスピーチ知覚に関係する、関数に基づいて選択される。

【 0 1 2 6 】

50

更に、幾つかの実施形態では、周波数帯域の均等な重み付けが線形ゲイン関数の中で使用されており、他方、他の関連するアルゴリズムは異なる重み付けを使用する。S a u e r t と V a r y , 2 0 1 2 ( 非特許文献 3 ) を参照されたい。

【 0 1 2 7 】

更に、幾つかの実施形態は、スピーチ知覚に係る S I I ( の推定 ) を使用して、重み付けなしと全ての帯域の均等な重み付けとの間でクロスオーバーさせている。

【 0 1 2 8 】

本提案の実施形態は、ノイズの多い環境においてスピーチを聞くときの了解度を改善させることに役立つ。その改善度は、既存の方法を用いた場合よりも有意に高くなり得る。本提案の概念は、上述したように種々の点において先行技術とは異なっている。

10

【 0 1 2 9 】

現状の技術によるアルゴリズム、例えば上述したアルゴリズムも、了解度を改善することができるが、本提案の実施形態の特別な特徴は、現在使用可能な方法よりも一層の効率化を可能とする。

【 0 1 3 0 】

本提案の実施形態、例えば本提案の方法は、オーディオ再生を用いた多くの技術的アプリケーションにおける信号プロセッサの一部、又は信号処理ソフトウェアとして使用可能である。例えば、

- 鉄道駅、公共交通機関、学校における構内放送システム
- 携帯電話、ヘッドセットなどの通信装置
- 自動車内のインフォテインメント・システム、機内エンターテインメント・システム
- 信号ミキシングの前 ( 例えば映画のオーディオ材料のミキシングの間 ) に複数のオーディオシステムを含むメディアファイル内のスピーチの了解度を改善するためのツールとして。

20

【 0 1 3 1 】

更に、本提案の実施形態はまた、残響のような他のタイプの信号妨害に対しても使用可能である。それらの信号妨害は、上述したアルゴリズムの形態ではノイズと同様に取り扱うことができる。

【 0 1 3 2 】

図 5 b は、説明したアルゴリズムの、他の実施形態に係るフローチャートを示す。

30

【 0 1 3 3 】

図 5 b で示す実施形態においては、室内音響的情報が提案のアルゴリズム内で考慮されてもよい。スピーチ信号はラウドスピーカによって再生され、妨害されたスピーチ信号がマイクロホンによって集音される。録音された信号は、ノイズ  $r[k]$  と残響スピーチ信号とを含む。残響スピーチ信号の中に含まれた残響のいくらかの部分は有害であると認められ得る一方で、他の部分はスピーチ了解度にとって有益であると考えられ得る。例えばラウドスピーカとマイクロホンとの間の室内インパルス応答をモデル化するフィルタである室内音響的情報生成部 ( R I G )、残響時間  $T_{60}$  (  $60$  db 減衰する時間として定義される )、又は直接対残響エネルギー比 ( D R R ) を使用して、残響スペクトルレベル  $z_n[l]$  が、室内音響的情報生成部から提供された情報と各サブバンドにおけるサブバンドスピーチ信号  $s_n[k]$  とを用いて計算されてもよい ( 符号 1 6 5 を参照 )。重み付けファクタを用いて次式のような重み付きの加算  $a_n[l]$  が決定されてもよい ( 符号 1 6 6 を参照 )。

40

$$a_n[l] = \beta z_n[l] + d_n[l]$$

また、重み付きの加算  $a_n[l]$  は後続の計算で使用されてもよく、その他の場合にはノイズスペクトルレベル  $d_n[l]$  だけが使用される。

【 0 1 3 4 】

$d_n$  について定義された全ての式は、 $d_n$  を  $a_n$  に置き換えることで、 $a_n$  についても適用可能である。例えば、式 ( 4 )、式 ( 5 ) 及び / 又は式 ( 8 ) 内において、 $d_n$  は  $a_n$  で置

50

き換えることができ、その置き換えにより、これらの式は重み付きの加算  $a_n$  を考慮に入れることができる。

【 0 1 3 5 】

例えば、 $a_n$  は実数であってもよく、ここで、例えば  $a_n = 0$  が適用されてもよい。

【 0 1 3 6 】

提案のアルゴリズムの性能は、主観的なリスニングテストを使用して、時間及び周波数依存のゲイン特性だけを使用する現状技術のアルゴリズム及び未処理の参照信号と比較された。リスニングテストは、8人の健聴被験者により、2個の異なるノイズタイプ、即ち定常的な自動車ノイズとより非定常的なカフェテリアのノイズとを用いて行われた。各ノイズタイプについて、未処理の参照状態において20%、50%及び80%の言語理解度のポイントに対応する、3つの異なる信号対ノイズ比が測定された。その結果、提案のアルゴリズムは、現状技術のアルゴリズム及び未処理の参照よりも、同等のスピーチレベルにおける両方のノイズシナリオにおいて性能が優れていることが示された。更に、客観的尺度と主観的なデータとの間の相関関係の分析は、順位の高い相関関係および高い線形相関を示しており、その点は、前処理アルゴリズムの評価において、客観的尺度を主観的なデータを予測するために部分的に使用し得ることを示唆している。

【 0 1 3 7 】

上述したように、SII依存の増幅および圧縮を用いて背景ノイズにおけるスピーチ了解度を改善する概念が提供された。

【 0 1 3 8 】

上述したように、例えば構内放送システム、カーナビゲーションシステム又は携帯電話などの通信装置において、明瞭なスピーチ信号が提供される場合も多い。しかしながら、時には近端のリスナーにとって妨害により聞き取れないことも起こり得る。上述の実施形態は、妨害を受けた所与の音響シナリオにおいて、明瞭なスピーチ信号を修正して、了解度及び/又はリスニングの快適さを高めるものである。

【 0 1 3 9 】

図6は、実施形態に係る近端リスニング強化が設けられたシナリオを示す。特に図6は、一実施形態に係る近端リスニング強化が設けられた信号モデルを示す。

【 0 1 4 0 】

図6において、次式が適用されてもよい。

$$\tilde{s}[k] = W \{s[k], \hat{r}[k], \hat{h}[k]\} \cdot s[k]$$

【 0 1 4 1 】

例えば次式のように完全なノイズ推定が可能であると想定され得る。

$$\hat{r}[k] = r[k]$$

【 0 1 4 2 】

更に、残響が全くない場合には、次式のようなになる。

$$h[k] = \delta[k]$$

【 0 1 4 3 】

残響も考慮に入れると、このようなことは全ての状態で起こりえないであろう。しかし、その代わりに、例えば室内インパルス応答  $h[k]$  のような何らかの室内情報の完全な推定が可能であると想定されてもよい。

【 0 1 4 4 】

同等のパワー制約下において、了解度  $s[k] + r[k]$  と比較して了解度

$$\tilde{s}[k] + r[k]$$

を強化する重み付け関数  $W\{\cdot\}$  を見つけることが望まれ得る。

10

20

30

40

50



## 【 0 1 4 5 】

同等のパワー制約に従って、全てのサブバンド内の全パワーが増幅前と増幅後とでほぼ同じとなるように、重み付け関数  $W\{\cdot\}$  が決定されてもよい。

## 【 0 1 4 6 】

図7は1～16000Hzの中心周波数についての長期間スピーチレベルを示す。特に、1つのスピーチ入力信号と複数の修正済みスピーチ信号とについての長期間スピーチレベルを示す。

## 【 0 1 4 7 】

一実施形態にかかるアルゴリズムは、 $s[k]$  及び

$$\hat{r}[k]$$

からSIIを推定し、2つのSII依存のステージ、特に多帯域周波数整形と多帯域圧縮スキームとを組み合わせる。

## 【 0 1 4 8 】

ある主観的な評価が行われた。処理条件は、未処理の参照(「参照」と、一実施形態に従うアルゴリズムを用いた処理の結果として得られたスピーチ信号(「DynComp」と、元来はSauertra(非特許文献3を参照)により提案された修正済みアルゴリズムを用いた処理の結果として得られたスピーチ信号(「ModSau」と)に関する主観的評価を含む。

## 【 0 1 4 9 】

その主観的評価に関し、8人の健聴被験者が参加した。2つの異なるノイズ、即ち自動車ノイズとカフェテリアノイズとがテストされた。Oldenburg Sentence Testからのスピーチ材料が使用された。信号対ノイズ比は、20%、50%及び80%の言語理解度の測定ポイントを目標として、選択された。

## 【 0 1 5 0 】

図8は主観的評価の結果を示す。

## 【 0 1 5 1 】

図9は主観的結果に関する相関分析を示す。主観的結果の予測に関して、自動車ノイズ及びカフェテリアノイズの中で、未処理の参照状態から導出されたモデル予測値の非線形変換の後に、相関分析が実施された。

$$P(SII) = \frac{m}{a + e^{-b \cdot SII}} + c$$

## 【 0 1 5 2 】

主観的評価から、実施形態に従う前処理により、スピーチ理解度の向上が達成されると結論付けられる。実施形態にかかる提案の概念は、スピーチ理解度において最高の改善を示している。更に、スピーチ理解度に関する現在のモデルは主観的データとの高い順位相関を示す。更に、変換されたモデル値に基づく予測は、高い線形相関を示すが、部分的には大きな線形のずれをも示す。

## 【 0 1 5 3 】

これまで装置を説明する文脈で幾つかの態様を示してきたが、これらの態様は対応する方法の説明でもあることは明らかであり、そのブロック又は装置が方法ステップ又は方法ステップの特徴に対応することは明らかである。同様に、方法ステップを説明する文脈で示した態様もまた、対応する装置の対応するブロックもしくは項目又は特徴を表している。

## 【 0 1 5 4 】

本発明の分解された信号は、デジタル記憶媒体に記憶されることができ、又はインターネットのような無線伝送媒体もしくは有線伝送媒体などの伝送媒体を通じて伝送されることができ。

## 【 0 1 5 5 】

所定の構成要件にも依るが、本発明の実施形態は、ハードウェア又はソフトウェアにおいて構成可能である。この構成は、その中に格納される電子的に読み取り可能な制御信号を有し、本発明の各方法が実行されるようにプログラム可能なコンピュータシステムと協働する（又は協働可能な）、デジタル記憶媒体、例えばフレキシブルディスク、DVD、CD、ROM、PROM、EPROM、EEPROM、フラッシュメモリなどのデジタル記憶媒体を使用して実行することができる。

【0156】

本発明に従う幾つかの実施形態は、上述した方法の1つを実行するようプログラム可能なコンピュータシステムと協働可能で、電子的に読み取り可能な制御信号を有する非一時的なデータキャリアを含む。

10

【0157】

一般的に、本発明の実施例は、プログラムコードを有するコンピュータプログラム製品として構成することができ、そのプログラムコードは当該コンピュータプログラム製品がコンピュータ上で作動するときに、本発明の方法の一つを実行するよう作動可能である。そのプログラムコードは例えば機械読み取り可能なキャリアに記憶されていても良い。

【0158】

本発明の他の実施形態は、上述した方法の1つを実行するための、機械読み取り可能なキャリアに記憶されたコンピュータプログラムを含む。

【0159】

換言すれば、本発明の方法のある実施形態は、そのコンピュータプログラムがコンピュータ上で作動するときに、上述した方法の1つを実行するためのプログラムコードを有するコンピュータプログラムである。

20

【0160】

本発明の他の実施形態は、上述した方法の1つを実行するために記録されたコンピュータプログラムを含む、データキャリア（又はデジタル記憶媒体又はコンピュータ読み取り可能な媒体）である。

【0161】

本発明の他の実施形態は、上述した方法の1つを実行するためのコンピュータプログラムを表現するデータストリーム又は信号列である。そのデータストリーム又は信号列は、例えばインターネットを介するデータ通信接続を介して伝送されるよう構成されても良い。

30

【0162】

他の実施形態は、上述した方法の1つを実行するよう構成又は適応された、例えばコンピュータ又はプログラム可能な論理デバイスのような処理手段を含む。

【0163】

他の実施形態は、上述した方法の1つを実行するためのコンピュータプログラムがインストールされたコンピュータを含む。

【0164】

幾つかの実施形態においては、（例えば書換え可能ゲートアレイのような）プログラム可能な論理デバイスが、上述した方法の幾つか又は全ての機能を実行するために使用されても良い。幾つかの実施形態では、書換え可能ゲートアレイは、上述した方法の1つを実行するためにマイクロプロセッサと協働しても良い。一般的に、そのような方法は、好適には任意のハードウェア装置によって実行される。

40

【0165】

上述した実施形態は、本発明の原理を単に例示的に示したにすぎない。本明細書に記載した構成及び詳細について修正及び変更が可能であることは、当業者にとって明らかである。従って、本発明は、本明細書に実施形態の説明及び解説の目的で提示した具体的詳細によって限定されるものではなく、添付した特許請求の範囲によってのみ限定されるべきである。

【 図 1 】

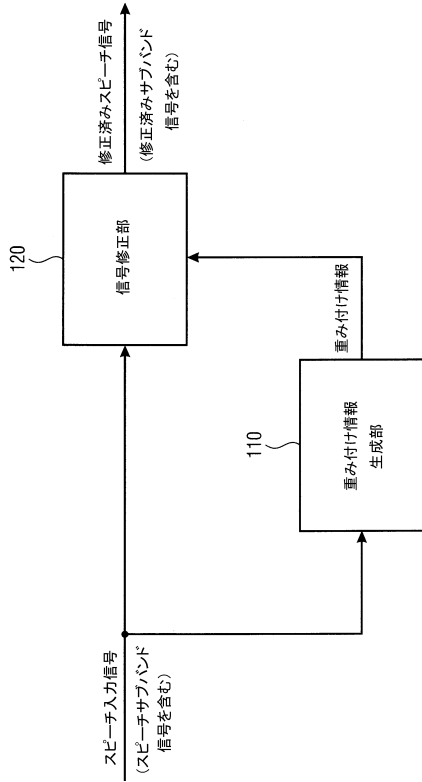


FIG 1

【 図 2 】

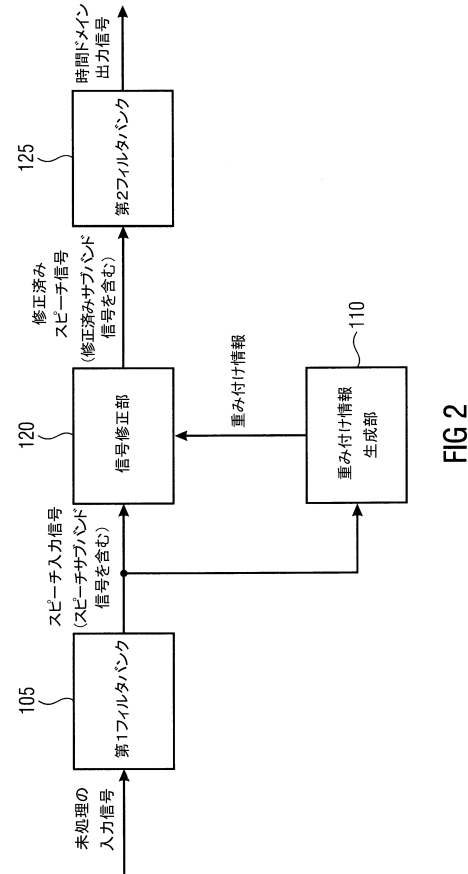


FIG 2

【 図 3 a 】

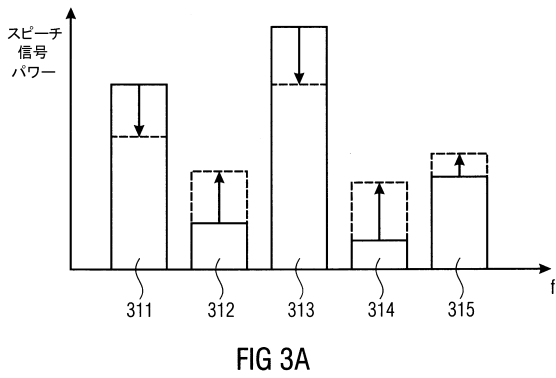


FIG 3A

【 図 3 b 】

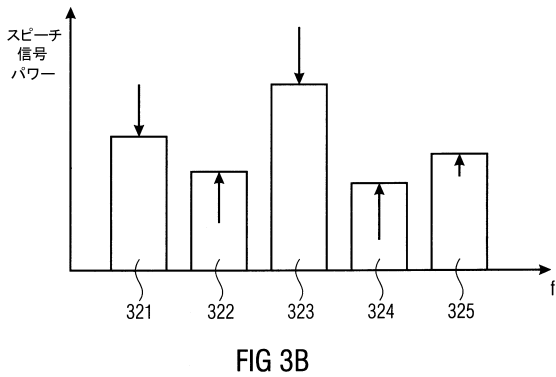


FIG 3B

【 図 4 a 】

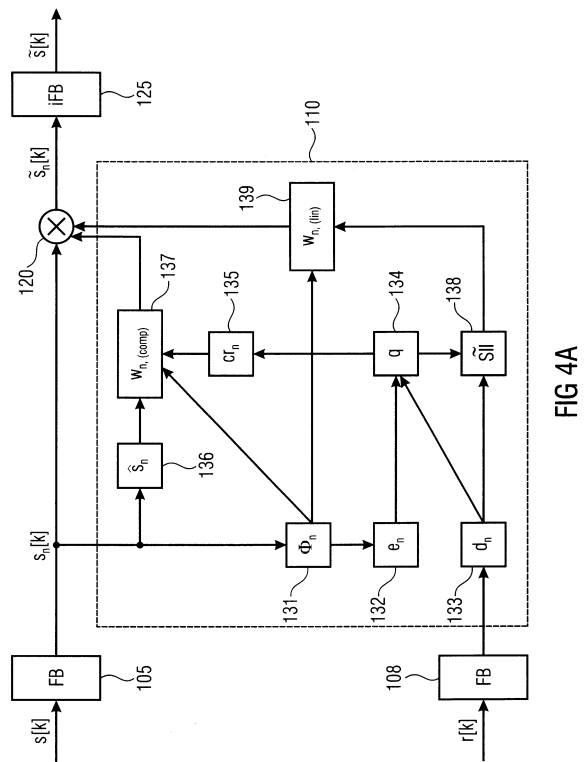


FIG 4A

【 4 b 】

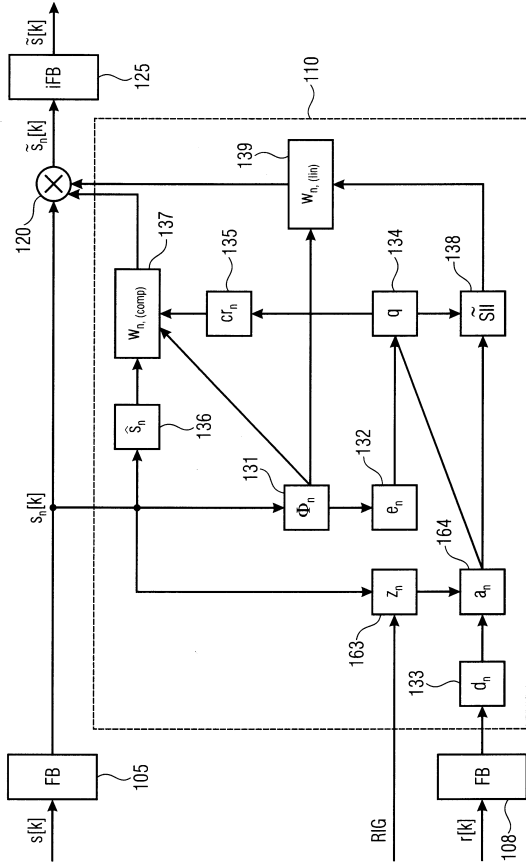


FIG 4B

【 5 a 】

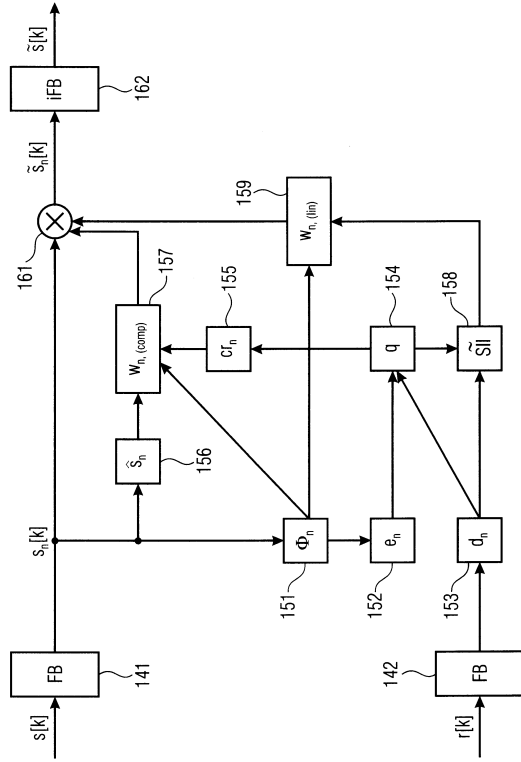


FIG 5A

【 5 b 】

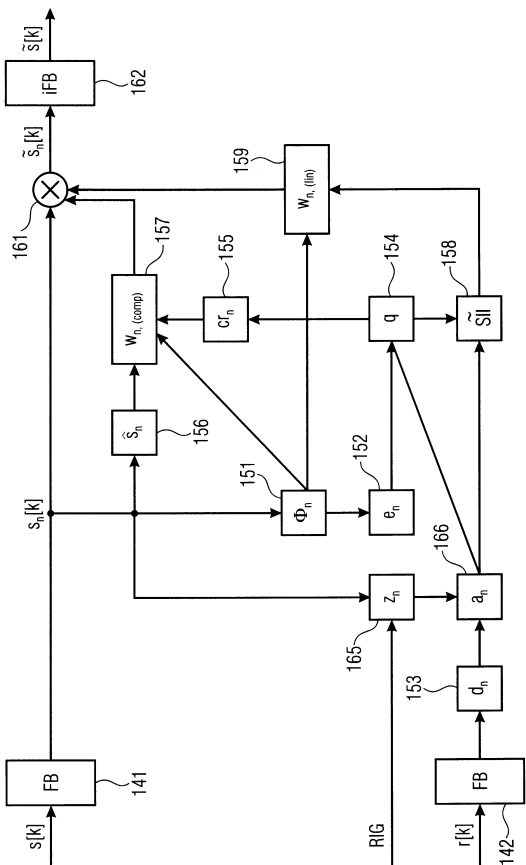


FIG 5B

【 6 】

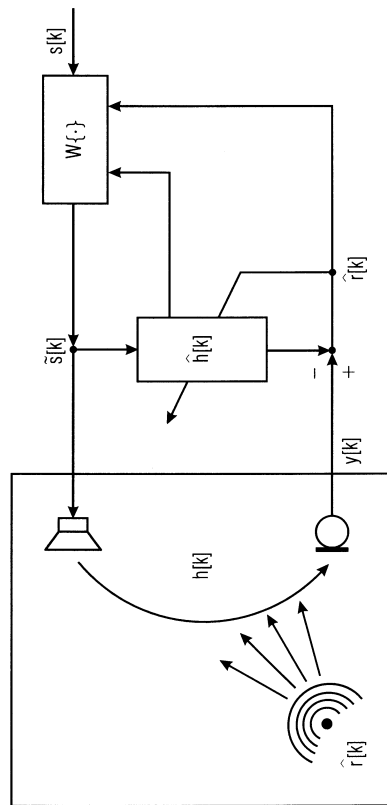


FIG 6

【 図 7 】

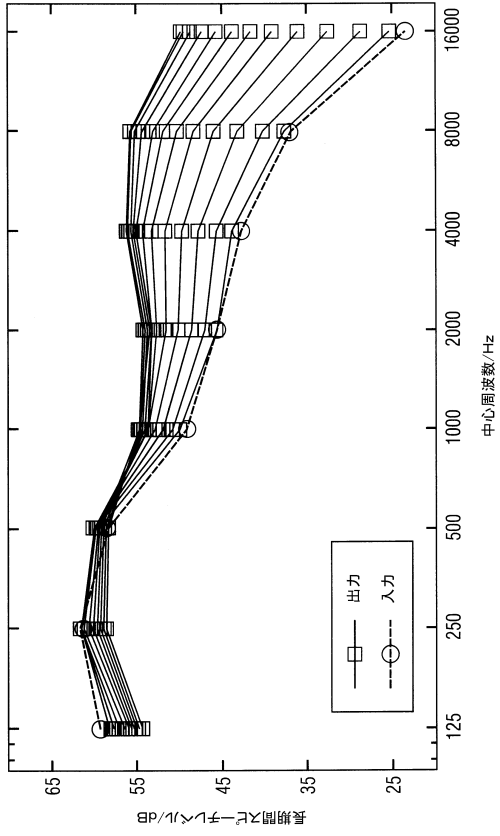


FIG 7

【 図 8 】

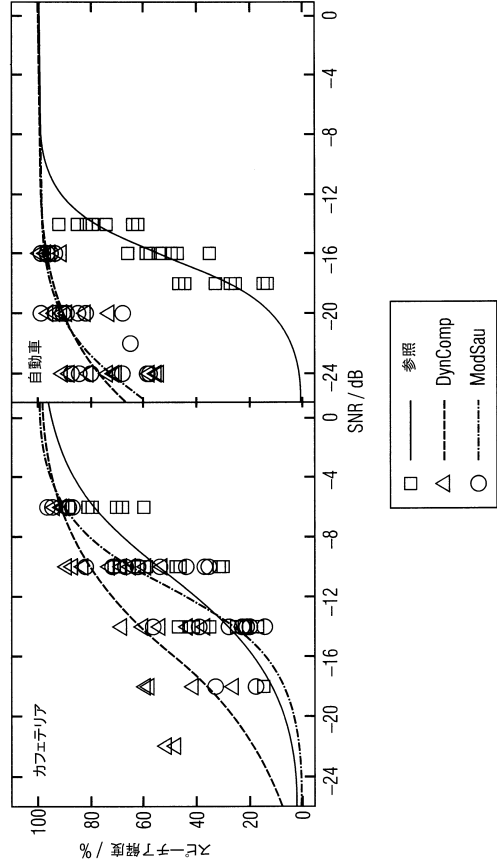


FIG 8

【 図 9 】

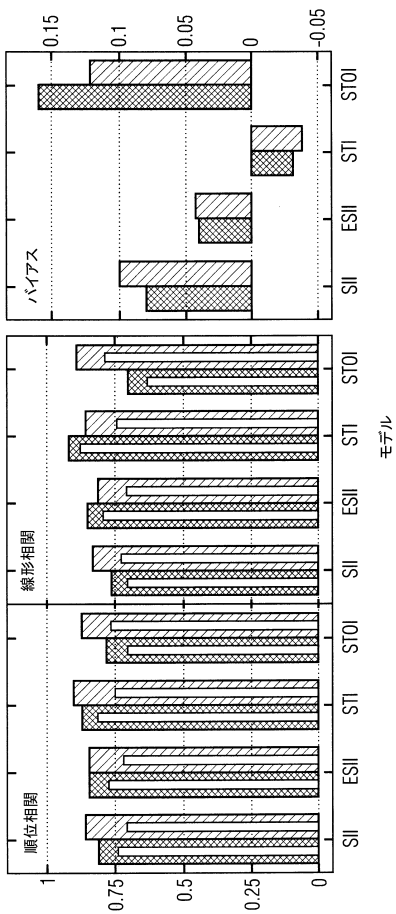


FIG 9

## フロントページの続き

- (72)発明者 レニース, ヤン  
ドイツ連邦共和国 2 6 1 2 3 オルデンブルク シュタルクロフストラーセ 5
- (72)発明者 ドクロ, シモン  
ドイツ連邦共和国 2 6 1 2 2 オルデンブルク ウンテル デン アイヘン 8
- (72)発明者 アッペル, イェンス エー  
ドイツ連邦共和国 2 6 2 0 3 ヴァルデンブルク アルテル ドルフヴェーク 3 6

審査官 安田 勇太

- (56)参考文献 特開平04 - 348000 (JP, A)  
特開2010 - 068175 (JP, A)  
特表2010 - 519601 (JP, A)  
国際公開第2011 / 048813 (WO, A1)  
特開平11 - 298990 (JP, A)

- (58)調査した分野(Int.Cl., DB名)  
G10L 21/02 - 21/0388