

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6707914号
(P6707914)

(45) 発行日 令和2年6月10日(2020.6.10)

(24) 登録日 令和2年5月25日(2020.5.25)

(51) Int.Cl. F I
G 1 O L 21/0232 (2013.01) G 1 O L 21/0232

請求項の数 6 (全 17 頁)

(21) 出願番号	特願2016-44475 (P2016-44475)	(73) 特許権者	000000295
(22) 出願日	平成28年3月8日(2016.3.8)		沖電気工業株式会社
(65) 公開番号	特開2017-161635 (P2017-161635A)		東京都港区虎ノ門一丁目7番12号
(43) 公開日	平成29年9月14日(2017.9.14)	(74) 代理人	100180275
審査請求日	平成30年11月15日(2018.11.15)		弁理士 吉田 倫太郎
		(74) 代理人	100161861
			弁理士 若林 裕介
		(74) 代理人	100090620
			弁理士 工藤 宣幸
		(72) 発明者	藤枝 大
			東京都港区虎ノ門一丁目7番12号 沖電気工業株式会社内
		審査官	渡部 幸和

最終頁に続く

(54) 【発明の名称】 ゲイン処理装置及びプログラム、並びに、音響信号処理装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

入力音響信号に基づく入力スペクトルと、所定の雑音推定方法から得られる雑音スペクトルとに基づいて、第1の帯域ごとの目的音存在確率を算出する目的音存在確率算出手段と、

前記目的音存在確率に基づいて、前記第1の帯域ごとの結合係数を算出する結合係数算出手段と、

所定のゲイン算出方法から得られる第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとを、前記結合係数に基づいて凸結合したゲインを生成するゲイン結合手段とを有し、

前記結合係数算出手段は、

前記目的音存在確率を、前記第1の帯域より帯域数の少ない第2の帯域に集約して集約目的音存在確率を取得する帯域集約手段と、

前記集約目的音存在確率を時間方向に平滑化して平滑化目的音存在確率を取得する時間平滑化手段と、

前記平滑化目的音存在確率を、前記第1の帯域へと分配する帯域分配手段とを有することを特徴とするゲイン処理装置。

【請求項2】

前記帯域集約手段は、前記第2の帯域へと集約される所定の複数の前記第1の帯域の中から、前記目的音存在確率の要素の最大値を選択して前記第2の帯域の前記集約目的音存

在確率の要素の値とすることを特徴とする請求項1に記載のゲイン処理装置。

【請求項3】

前記ゲイン結合手段は、

前記第1のゲインを、前記第1の帯域より帯域数の少ない第3の帯域に集約して集約ゲインを取得するゲイン帯域集約手段と、

前記集約ゲインを時間方向に平滑化して平滑化ゲインを取得するゲイン時間平滑化手段と、

前記平滑化ゲインを、前記第1の帯域へと分配するゲイン帯域分配手段と

を有することを特徴とする、請求項1に記載のゲイン処理装置。

【請求項4】

入力音響信号を周波数変換して入力スペクトルを取得する周波数解析手段と、

前記入力スペクトルに含まれる雑音スペクトルを推定して取得する雑音推定手段と、

前記入力スペクトルと、前記雑音スペクトルとに基づいて、第1のゲインを算出する、ゲイン算出手段と、

前記入力スペクトルと、前記雑音スペクトルと、前記第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとに基づいて、前記第1のゲインを修正した第3のゲインを算出するゲイン修正手段と、

前記入力スペクトルに、前記第3のゲインを乗じて出力スペクトルを取得する乗算手段と、

前記出力スペクトルに基づく信号を出力する出力手段とを有し、

前記ゲイン修正手段として、請求項1～3のいずれかに記載のゲイン処理装置を適用したことを

特徴とする音響信号処理装置。

【請求項5】

コンピュータを、

入力音響信号に基づく入力スペクトルと、所定の雑音推定方法から得られる雑音スペクトルとに基づいて、第1の帯域ごとの目的音存在確率を算出する目的音存在確率算出手段と、

前記目的音存在確率に基づいて、前記第1の帯域ごとの結合係数を算出する結合係数算出手段と、

所定のゲイン算出方法から得られる第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとを、前記結合係数に基づいて凸結合したゲインを生成するゲイン結合手段として機能させ、

前記結合係数算出手段は、

前記目的音存在確率を、前記第1の帯域より帯域数の少ない第2の帯域に集約して集約目的音存在確率を取得する帯域集約手段と、

前記集約目的音存在確率を時間方向に平滑化して平滑化目的音存在確率を取得する時間平滑化手段と、

前記平滑化目的音存在確率を、前記第1の帯域へと分配する帯域分配手段とを有する

ことを特徴とするゲイン処理プログラム。

【請求項6】

コンピュータを、

入力音響信号を周波数変換して入力スペクトルを取得する周波数解析手段と、

前記入力スペクトルに含まれる雑音スペクトルを推定して取得する雑音推定手段と、

前記入力スペクトルと、前記雑音スペクトルとに基づいて、第1のゲインを算出する、ゲイン算出手段と、

前記入力スペクトルと、前記雑音スペクトルと、前記第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとに基づいて、前記第1のゲインを修正した第3のゲインを算出するゲイン修正手段と、

前記入力スペクトルに、前記第3のゲインを乗じる乗算手段と、

10

20

30

40

50

前記乗算手段によって得られる出力スペクトルから波形を復元する波形復元手段として機能させ、

前記ゲイン修正手段として、請求項 1 ~ 3 のいずれかに記載のゲイン処理装置を適用したこと

を特徴とする音響信号処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ゲイン処理装置及びプログラム、並びに、音響信号処理装置及びプログラムに関し、例えば、入力音響信号に含まれる雑音成分を抑圧して音声成分を強調することを必要とする通信端末、オーディオ機器、音声認識装置などに適用し得る。

10

【背景技術】

【0002】

自然環境において雑音はいたる所に存在するため、一般に実世界で収録された音声には種々の発信元からの雑音が含まれる。それらの雑音は、人が聴くにしても音声の了解性を低下させ、また、音声認識装置等の音声処理装置に入力するにしても音声処理の精度（例えば音声認識率）を低下させる。そのため、入力信号に混入した雑音成分を抑圧して音声成分を強調する技術の需要は高く、これまでに様々な音声強調方法（雑音抑圧方法と呼ばれることもある）が開発されてきた。

【0003】

20

従来の音声強調の標準的な方法として、非特許文献 1、2 に記載されている技術を挙げることができる。非特許文献 1 の記載技術は、入力信号の振幅スペクトルから、別途推定した雑音成分の振幅スペクトルを減じる方法であり、スペクトルサブトラクション（Spectral Subtraction: SS）法と呼ばれる。また、非特許文献 2 の記載技術は、音声成分が強調されるように選定されたゲインを入力信号の周波数スペクトル（以下、単に入力スペクトルと呼ぶ）に乗算する方法で、MMSE-STSA（Minimum Mean-Square Error Short-Time Spectral Amplitude）法と呼ばれる。

【0004】

非特許文献 1、2 に記載された方法は、いずれも、事前に雑音成分の振幅スペクトル又はパワースペクトル（以下、単に「雑音スペクトル」と呼ぶ）を推定しておく必要がある。推定された雑音スペクトルは、必然的に推定誤差を含む。この推定誤差は、SS法又はMMSE-STSA法等により音声成分が強調された信号に、歪みを生じさせる。特に、SS法は、ミュージカルノイズと呼ばれる耳障りな歪みを生じることが知られている。

30

【0005】

このような背景から、従来、ミュージカルノイズの軽減を目的とした、特許文献 1 の記載技術がある。

【0006】

特許文献 1 に記載の雑音抑圧装置は、音響信号（スペクトル）の強度の度数分布における尖度が、雑音抑圧処理の前後で変化した度合いを示す尖度指標値を算出する尖度指標値算出部と、SS法を用いる第 1 の雑音抑圧部と、MMSE-STSA法を用いる第 2 の雑音抑圧部とを具備している。尖度指標値は第 1 の雑音抑圧部と第 2 の雑音抑圧部との両方に対して算出される。そして、尖度指標値が雑音成分の抑圧後に生じるミュージカルノイズの量と正の相関を有することを利用して、尖度指標値をより小さくできる方の雑音抑圧部を選択する。以上のように、特許文献 1 の記載技術は、尖度指標値に応じて 2 つの雑音抑圧部を切り替えることで、ミュージカルノイズの発生を軽減しようとしている。

40

【先行技術文献】

【特許文献】

【0007】

【特許文献 1】特開 2010 - 160246 号公報

50

【非特許文献】

【0008】

【非特許文献1】S. F. Boll, "Suppression of acoustic noise in using spectral subtraction," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-27, no. 2, pp. 113-120, April 1979.

【非特許文献2】Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-32, no. 6, pp. 1103-1121, December 1984.

10

【発明の概要】

【発明が解決しようとする課題】

【0009】

しかしながら、特許文献1の記載技術では、全ての周波数帯域で同時にSS法とMMSE-STSA法とを切り替えるため、切り替わった瞬間に出力音声の特性が急に変化し、不自然な音響信号として受聴者に知覚されるという問題が生じ得る。

【0010】

20

そのため、入力音響信号に含まれる雑音を抑圧して目的音成分を強調する際に不自然な変化や歪みを抑制するゲイン処理装置及びプログラム、並びに、音響信号処理装置及びプログラムが望まれている。

【課題を解決するための手段】

【0011】

第1の本発明のゲイン処理装置は、(1)入力音響信号に基づく入力スペクトルと、所定の雑音推定方法から得られる雑音スペクトルとに基づいて、第1の帯域ごとの目的音存在確率を算出する目的音存在確率算出手段と、(2)前記目的音存在確率に基づいて、前記第1の帯域ごとの結合係数を算出する結合係数算出手段と、(3)所定のゲイン算出方法から得られる第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとを、前記結合係数に基づいて凸結合したゲインを生成するゲイン結合手段とを有し、(4)前記結合係数算出手段は、(4-1)前記目的音存在確率を、前記第1の帯域より帯域数の少ない第2の帯域に集約して集約目的音存在確率を取得する帯域集約手段と、(4-2)前記集約目的音存在確率を時間方向に平滑化して平滑化目的音存在確率を取得する時間平滑化手段と、(4-3)前記平滑化目的音存在確率を、前記第1の帯域へと分配する帯域分配手段とを有することを特徴とする。

30

【0012】

第2の本発明の音響信号処理装置は、(1)入力音響信号を周波数変換して入力スペクトルを取得する周波数解析手段と、(2)前記入力スペクトルに含まれる雑音スペクトルを推定して取得する雑音推定手段と、(3)前記入力スペクトルと、前記雑音スペクトルとに基づいて、第1のゲインを算出する、ゲイン算出手段と、(4)前記入力スペクトルと、前記雑音スペクトルと、前記第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとに基づいて、前記第1のゲインを修正した第3のゲインを算出するゲイン修正手段と、(5)前記入力スペクトルに、前記第3のゲインを乗じて出力スペクトルを取得する乗算手段と、(6)前記出力スペクトルに基づく信号を出力する出力手段とを有し、前記ゲイン修正手段として第1の本発明の、ゲイン処理装置を適用したことを特徴とする。

40

【0013】

第3の本発明のゲイン処理プログラムは、コンピュータを、(1)入力音響信号に基づく入力スペクトルと、所定の雑音推定方法から得られる雑音スペクトルとに基づいて、第

50

1の帯域ごとの目的音存在確率を算出する目的音存在確率算出手段と、(2)前記目的音存在確率に基づいて、前記第1の帯域ごとの結合係数を算出する結合係数算出手段と、(3)所定のゲイン算出方法から得られる第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとを、前記結合係数に基づいて凸結合したゲインを生成するゲイン結合手段と、(4)前記結合係数算出手段は、(4-1)前記目的音存在確率を、前記第1の帯域より帯域数の少ない第2の帯域に集約して集約目的音存在確率を取得する帯域集約手段と、(4-2)前記集約目的音存在確率を時間方向に平滑化して平滑化目的音存在確率を取得する時間平滑化手段と、(4-3)前記平滑化目的音存在確率を、前記第1の帯域へと分配する帯域分配手段とを有することを特徴とする。

【0014】

10

第4の本発明の音響信号処理プログラムは、コンピュータを、(1)入力音響信号を周波数変換して入力スペクトルを取得する周波数解析手段と、(2)前記入力スペクトルに含まれる雑音スペクトルを推定して取得する雑音推定手段と、(3)前記入力スペクトルと、前記雑音スペクトルとに基づいて、第1のゲインを算出する、ゲイン算出手段と、(4)前記入力スペクトルと、前記雑音スペクトルと、前記第1のゲインと、前記第1のゲインを周波数方向と時間方向に平滑化した第2のゲインとに基づいて、前記第1のゲインを修正した第3のゲインを算出するゲイン修正手段と、(5)前記入力スペクトルに、前記第3のゲインを乗じる乗算手段と、(6)前記乗算手段によって得られる出力スペクトルから波形を復元する波形復元手段として機能させ、(7)前記ゲイン修正手段として、第1の本発明のゲイン処理装置を適用したことを特徴とする。

20

【発明の効果】

【0015】

本発明によれば、入力音響信号に含まれる雑音を抑圧して目的音成分を強調する際に不自然な変化や歪みを抑制することができる。

【図面の簡単な説明】

【0016】

【図1】第1の実施形態に係るゲイン処理手段の機能的構成について示したブロック図である。

【図2】第1の実施形態に係る音声処理装置の機能的構成について示したブロック図である。

30

【図3】第1の実施形態に係る結合係数算出手段の機能的構成について示した説明図である。

【図4】第1の実施形態に係る音声処理装置(ゲイン修正手段)で、処理される第1の帯域と第2の帯域との関係について示したグラフである。

【図5】第1の実施形態に係る音声処理装置(ゲイン修正手段)で、サンプルデータに基づいて算出された音声存在確率Lの例について示したグラフである。

【図6】第1の実施形態に係る音声処理装置(ゲイン修正手段)で、サンプルデータに基づいて算出された集約音声存在確率Pの例について示したグラフである。

【図7】第1の実施形態に係る音声処理装置(ゲイン修正手段)で、サンプルデータに基づいて算出された平滑化音声存在確率Mの例について示したグラフである。

40

【図8】第1の実施形態に係る音声処理装置(ゲイン修正手段)で、サンプルデータに基づいて算出された結合係数Cの例について示したグラフである。

【図9】第2の実施形態に係る音声処理装置の機能的構成について示したブロック図である。

【図10】第2の実施形態に係るゲイン処理手段の機能的構成について示したブロック図である。

【図11】第2の実施形態に係るゲイン平滑化手段の機能的構成について示したブロック図である。

【発明を実施するための形態】

50

【 0 0 1 7 】

(A) 第 1 の実施形態

以下、本発明によるゲイン処理装置及びプログラム、並びに、音響信号処理装置及びプログラムの第 1 の実施形態を、図面を参照しながら詳述する。以下では、本発明の音響信号処理装置及びゲイン処理装置を、音声処理装置及びゲイン修正手段に適用した例について説明する。

【 0 0 1 8 】

(A - 1) 第 1 の実施形態の構成

図 2 は、この実施形態に係る音声処理装置 3 0 0 の全体構成について示したブロック図である。

10

【 0 0 1 9 】

音声処理装置 3 0 0 は、例えば、会議端末、携帯電話端末（スマートフォン）、IP 電話端末等の電話端末のマイクロホンで捕捉した音響信号に対して目的音（例えば、近端話者の音声）を強調（雑音成分を抑制）する処理等に好適に用いられる。

【 0 0 2 0 】

音声処理装置 3 0 0 は、音声を含む音声信号（時間領域の音声信号）である入力信号 x （時間領域の入力信号）について音声強調を行って、出力信号 y （時間領域の出力信号）を生成するものである。

【 0 0 2 1 】

図 2 に示すように、第 1 の実施形態の音声処理装置 3 0 0 は、周波数解析手段 3 0 1、雑音推定手段 3 0 2、ゲイン算出手段 3 0 3、ゲイン修正手段 3 0 4、乗算手段 3 0 5、及び波形復元手段 3 0 6 を有している。

20

【 0 0 2 2 】

音声処理装置 3 0 0 は、例えば、プロセッサ及びメモリを有するコンピュータにプログラム（実施形態に係る音響信号処理プログラムを含む）をインストールすることにより実現するようにしてもよい。また、ゲイン修正手段 3 0 4 についても、例えば、プロセッサ及びメモリを有するコンピュータにプログラム（実施形態に係るゲイン処理プログラム）をインストールすることにより実現するようにしてもよい。

【 0 0 2 3 】

周波数解析手段 3 0 1 は、入力信号 x （時間領域の信号）を周波数変換して入力スペクトル X （周波数領域の信号）を算出するものである。

30

【 0 0 2 4 】

雑音推定手段 3 0 2 は、入力スペクトル X に基づいて雑音成分を推定し、推定した雑音成分である雑音スペクトル N を算出するものである。

【 0 0 2 5 】

ゲイン算出手段 3 0 3 は、入力スペクトル X と雑音スペクトル N とに基づいて、雑音推定手段 3 0 2 が推定した雑音成分を抑圧する原ゲイン G を算出するものである。

【 0 0 2 6 】

ゲイン修正手段 3 0 4 は、入力スペクトル X と雑音スペクトル N と原ゲイン G と所定の定数に基づく無歪ゲイン F とに基づいて、原ゲイン G を修正した修正ゲイン H を算出するものである。

40

【 0 0 2 7 】

乗算手段 3 0 5 は、入力スペクトル X に修正ゲイン H を乗じて出力スペクトル Y を算出する。

【 0 0 2 8 】

波形復元手段 3 0 6 は、出力スペクトル Y （周波数領域の信号）から波形を復元し、得られた復元波形を出力信号 y （時間領域の信号）として出力するものである。波形復元手段 3 0 6 が出力信号 y を出力する方式や信号の形式については限定されないものであり、例えば、音声データ（例えば、PCM 形式のデータ）として所定のデータ記録媒体に出力するようにしてもよいし、パケット形式に変換して通信により他の通信装置に向けて送出

50

するようにしてもよい。また、音声処理装置 300 では、時間領域の出力信号 y ではなく、周波数領域の出力スペクトル Y のまま出力するようにしてもよい。

【0029】

次に、ゲイン修正手段 304 内部の機能的構成について図 1 を用いて説明する。

【0030】

図 1 に示すように、第 1 の実施形態のゲイン修正手段 304 は、音声存在確率算出手段 101、結合係数算出手段 102、及びゲイン結合手段 103 を有している。

【0031】

音声存在確率算出手段 101 は、入力スペクトル X と雑音スペクトル N とに基づいて音声存在確率 L (詳細については後述) を算出するものである。

10

【0032】

結合係数算出手段 102 は、音声存在確率 L に基づいて結合係数 C (詳細については後述) を算出するものである。

【0033】

ゲイン結合手段 103 は、原ゲイン G と無歪ゲイン F と音声存在確率 L に基づいて修正ゲイン H を算出するものである。修正ゲイン H の詳細については後述する。

【0034】

次に、結合係数算出手段 102 の内部構成について図 3 を用いて説明する。

【0035】

結合係数算出手段 102 は、帯域集約手段 201、時間平滑化手段 202、及び帯域分配手段 203 を有している。

20

【0036】

帯域集約手段 201 は、音声存在確率 L の帯域を第 2 の帯域 (詳細については後述) へと集約し、得られた集約音声存在確率 P (詳細については後述) を時間平滑化手段 202 に供給するものである。

【0037】

時間平滑化手段 202 は、帯域集約手段 201 から与えられた集約音声存在確率 P を時間平滑化して、得られた平滑化音声存在確率 M (詳細については後述) を帯域分配手段 203 に与えるものである。

【0038】

帯域分配手段 203 は、時間平滑化手段 202 から与えられた平滑化音声存在確率 M を第 1 の帯域 (詳細については後述) へと分配し、得られた結合係数 C を後段の処理に与えるものである。

30

【0039】

(A-2) 第 1 の実施形態の動作

次に、以上のような構成を有する第 1 の実施形態の音声処理装置 300 (ゲイン修正手段 304) の動作を説明する。

【0040】

まず、音声処理装置 300 が、入力信号 x (時間領域の入力信号) について音声強調を行って、出力信号 y (時間領域の出力信号) を生成する処理の概要について図 2 を用いて説明する。

40

【0041】

周波数解析手段 301 は、例えば、フーリエ変換に代表される任意の周波数解析手法、またはフィルタバンクに代表される任意の帯域分割手法によって、入力信号 x (入力音声) を K 個の第 1 の帯域に分割する。以下では、 k 番目の第 1 の帯域に分割された信号を「 X_k 」と表し、 K 個の帯域すべてをまとめた入力スペクトルを「 X 」と表す。そして、周波数解析手段 301 は、得られた入力スペクトル X を雑音推定手段 302、ゲイン算出手段 303、ゲイン修正手段 304、及び乗算手段 305 に与える。

【0042】

雑音推定手段 302 は、周波数解析手段 301 から与えられた入力スペクトル X に基づ

50

いて、所定の雑音推定方法によって入力スペクトル X に含まれる雑音成分を推定して雑音スペクトル N を算出し、得られた雑音スペクトル N をゲイン算出手段303及びゲイン修正手段304に与える。

【0043】

雑音推定手段302で用いられる雑音推定方法には、任意の雑音推定方法を適用することができ、例えば、参考文献1(R. Martin, "Spectral Subtraction Based on Minimum Statistics," in Proceedings of 7th European Signal Processing Conference, pp. 1182-1185, 1994)に記載の技術や、参考文献2(特開2015-152627)に記載の技術等を適用することができる。

10

【0044】

ゲイン算出手段303は、周波数解析手段301から与えられた入力スペクトル X と、雑音推定手段302から与えられた雑音スペクトル N とに基づいて、所定の音声強調方法によって入力スペクトル X 中に含まれる雑音成分を抑圧して音声成分を強調する音声強調ゲインを算出し、得られた音声強調ゲインを原ゲイン G としてゲイン修正手段304に与える。所定の音声強調方法(雑音抑圧方法)には、任意の音声強調方法(雑音抑圧方法)を適用することができ、例えば、非特許文献1に記載の技術や、非特許文献2に記載の技術等を適用することができる。

【0045】

20

ゲイン修正手段304は、周波数解析手段301から与えられた入力スペクトル X と、雑音推定手段302から与えられた雑音スペクトル N と、ゲイン算出手段303から与えられた原ゲイン G と、所定の定数値で与えられる無歪ゲイン F とに基づいて、修正ゲイン H を算出し、得られた修正ゲイン H を乗算手段305に与える。

【0046】

乗算手段305は、周波数解析手段301から与えられた入力スペクトル X に、ゲイン修正手段304から与えられた修正ゲイン H を乗じて出力スペクトル Y を算出し、得られた出力スペクトル Y を波形復元手段306に与える。

【0047】

波形復元手段306は、周波数解析手段301で用いた周波数解析手法または帯域分割手法に対応する波形復元手法を用いて、乗算手段305から与えられた出力スペクトル Y に基づいて時間波形を再構成し、得られた出力信号 y を出力する。

30

【0048】

次に、ゲイン修正手段304の動作について、図1を用いて説明する。

【0049】

音声存在確率算出手段101は、入力スペクトル X と雑音スペクトル N とに基づいて音声存在確率 L を算出し、得られた音声存在確率 L を結合係数算出手段102に与える。

【0050】

入力スペクトル X と雑音スペクトル N は、第1の帯域数 K 個に帯域分割された帯域信号である。以下では、 k 番目の第1の帯域の要素を、 X_k 、 N_k (但し、 $k = 0, 1, \dots, K - 1$)と表す。また、以下では、音声存在確率 L は、第1の帯域それぞれに個別の値を有し、これを同様に L_k と表す。 L_k は、 X_k と N_k を変数とする所定の関数の出力として与えられる。上述の所定の関数は、 X_k に関して広義単調増加であり、 N_k に関して広義単調減少であり、例えば、値域としては $0.0 \sim 1.0$ の実数値とすることができる。 X_k が N_k より十分大きければ、 L_k は 1.0 となり、確率 1.0 で音声が存在することを示す。一方、 X_k が N_k より十分小さければ、 L_k は 0.0 となり、確率 0.0 で音声が存在する(すなわち音声が存在しない)ことを示す。 L_k を求めるための所定の関数として、例えば、(1)式が好適に用いられる。

40

【数 1】

$$L_k = \frac{\tanh\left(\frac{X_k}{2N_k} - 6\right) - 1}{2} \quad (1)$$

【0051】

結合係数算出手段102は、音声存在確率算出手段101から与えられた音声存在確率Lに基づいて結合係数Cを算出し、得られた結合係数Cをゲイン結合手段103に与える。

10

【0052】

結合係数Cは、音声存在確率Lと同様に、第1の帯域それぞれに個別の値を有する。以下では、k番目の第1の帯域の結合係数Cを C_k と示す。 C_k は例えば、0.0~1.0の実数値で与えられ、大きいほど音声らしさが高いことを意味する。結合係数算出手段102は、音声存在確率Lを周波数方向と時間方向に平滑化した値を結合係数Cとして算出する。結合係数算出手段102で適用される平滑化の方法は任意であり、例えば画像処理で用いられるような2次元の移動平均法や2次元ガウシアンフィルタなどを適用できるが、特に図3で示される構成によって平滑化する方式が好ましい。結合係数算出手段102の詳細動作(すなわち結合係数 C_k の算出方法)については後述する。

【0053】

ゲイン結合手段103は、所定のゲイン算出方法(例えば、従来の任意の音声強調方法)によって算出された原ゲインGと、所定の定数値で与えられる無歪ゲインFと、結合係数算出手段102より与えられた結合係数Cとに基づいて、修正ゲインHを算出し、得られた修正ゲインHを出力する。

20

【0054】

原ゲインG、無歪ゲインF及び修正ゲインHは、結合係数Cと同様に、第1の帯域それぞれに個別の値を有し、これをそれぞれ G_k 、 F_k 及び H_k と示す。 H_k は、例えば、(2)式のように、 G_k と F_k とを、 C_k に基づいて凸結合することで算出することができる。

【数 2】

$$H_k = C_k G_k + (1 - C_k) F_k \quad (2)$$

30

【0055】

無歪ゲインの要素 F_k は、すべての第1の帯域 $k = 0, 1, \dots, K - 1$ で同じ値としても良いが、帯域ごとに異なる値とする方がより好適である。但し、隣り合う帯域同士で F_k が大きく異なると、出力信号に特有の歪み(人工的な定常音や、音楽的な調性を持った定常音)が生じてしまうため、例えば(3)式のように、隣り合う帯域との比が0.90~1.11に収まるように決めるのが好適である。

【数 3】

$$0.90 \leq \frac{F_k}{F_{k-1}} \leq 1.11, \quad k = 1, 2, \dots, K - 1 \quad (3)$$

40

【0056】

もし無歪ゲインFをすべての第1の帯域に適用した場合、出力信号はまったく歪まないが、その理由を説明する。第1の実施形態に示すように、無歪ゲインFは、時間方向にはまったく変化せず、周波数方向には(3)式の範囲内でのみ変化する。従って、無歪ゲインFは、時間方向と周波数方向の変化が小さい。このように変化の小さいゲインは、局所的(例えば、時間幅にして数十ミリ秒、周波数幅にして100Hz前後)に見ればほとんど変化していない。変化していないゲインを適用することは、音量を変化させることと同

50

じである。従って、無歪ゲインFをすべての第1の帯域に適用しても、入力信号の音量を変化させるだけなので、出力信号に歪みをまったく生じさせない。

【0057】

しかし、時間方向と周波数方向の変化が滑らかな無歪ゲインFをすべての第1の帯域に適用してしまうと、入力信号に含まれる音声成分を強調できない。それどころか、音声成分ごと抑圧してしまう。従って、音声成分がある帯域では、原ゲインGを用いることで、音声成分を強調する必要がある。

【0058】

それぞれの帯域kにおいて、結合係数 C_k が1.0に近いほど、入力スペクトルの要素 X_k は音声優勢であることを示す。そして、結合係数 C_k は、周波数方向と時間方向とに平滑化されているので、周波数方向と時間方向とに滑らかに変化する。そこで、ゲイン結合手段103は、 $C_k = 1.0$ では $H_k = G_k$ となるようにする。一方、 $C_k = 0.0$ では $H_k = F_k$ となるようにする。そして、 $0.0 < C_k < 1.0$ の場合、 G_k と F_k とを適切な割合で混合する。以上のようにすることで、修正ゲインHが帯域ごとに滑らかに変化するので、出力信号に歪みを生じさせないゲインを得ることができる。

【0059】

次に、結合係数算出手段102の詳細動作を、図3を用いて説明する。

【0060】

帯域集約手段201は、与えられた音声存在確率Lの帯域を第2の帯域へと集約し、得られた集約音声存在確率Pを時間平滑化手段202に与える。第2の帯域は、入力信号の帯域を、第1の帯域数K個より少ない第2の帯域数J個に分割した帯域である。第2の帯域の定め方は任意であり、例えば、J個のすべての帯域幅を同じ400Hzとしても良いが、臨界帯域幅に従うのが好適である。

【0061】

臨界帯域幅とは、人間の聴覚において、ある純音に対する聴覚マスクングが有効な帯域幅のことである。臨界帯域幅は、純音の周波数に対して非線形で、500Hz以下の純音に対しては常に100Hzであり、500Hzより上の周波数では純音の周波数に対して $1/3 \sim 1/4$ オクターブの割合で増加していく特性を持つ。臨界帯域幅(単位はBark)の近似式がいくつか提案されているが、例えば(4)式が用いられる。(4)式において、 f (Hz)であり、また b (Bark)である。

【数4】

$$b = 13 \operatorname{atan}(0.00076f) + 3.5 \operatorname{atan}\left(\left(\frac{f}{7500}\right)^2\right) \quad (4)$$

【0062】

以下では、第2の帯域j($j = 1, 2, \dots, J$)に集約される第1の帯域を、 $k_{(j-1)}$ 、 $k_{(j)} - 1$ とする。また、以下では、第1の帯域kの中心周波数を f_k とし、第2の帯域jの中心周波数を f_j (ファイ__ジェー)としたとき、 $k_{(j)}$ は、(5)式を満たす最小の第1の帯域kと定義する。このように定義される第1の帯域と第2の帯域との関係を図4に示す。そして、帯域集約手段201は、(6)式により、集約音声存在確率Pのj番目の値である P_j を、音声存在確率Lの $k_{(j-1)} \sim k_{(j)} - 1$ の最大値とする。このように、最大値を選択することで、ある第1の帯域に高い確率で音声成分が存在するという情報が、後段の帯域分配手段203まで保存されるので、原ゲインGの音声を強調する性質を修正ゲインHに反映させることができる。

【0063】

(6)式によって音声存在確率Lが集約される様子を、図5及び図6に示す。

【0064】

図5は、サンプルデータに基づく音声存在確率Lである。図5では、第1の帯域の各値 L_k 間を実線で結んでいる。図5では、音声存在確率の値 L_k が密に並んだ状態となって

10

20

30

40

50

いる。

【0065】

図6は、サンプルデータに基づく図5の音声存在確率Lから得られた集約音声存在確率Pである。図6では、第2の帯域の各値 P_j を黒丸で描き、各値 P_j 間を点線で結んでいる。図6では、集約音声存在確率の値 P_j が疎に並んだ状態となっている。

【数5】

$$\phi_{j+1} - f_{k(j)} \leq f_{k(j)} - \phi_j \quad (5)$$

$$P_j = \max_{i=k(j-1), \dots, k(j)-1} \{L_i\} \quad (6)$$

10

【0066】

なお、以上では、第2の帯域に集約される第1の帯域の範囲は、隣り合う第2の帯域で重ならないように記載したが、重なるようにしても良い。また、以上では、第2の帯域に集約する方法として、第1の帯域の所定の範囲の最大値を取るとしたが、平均値を取るようにしても良いし、重み付き平均値を取るようにしても良い。

【0067】

時間平滑化手段202は、帯域集約手段201から与えられた集約音声存在確率Pを時間平滑化して、得られた平滑化音声存在確率Mを帯域分配手段203に与える。平滑化音声存在確率Mは、集約音声存在確率Pと同様に、第2の帯域それぞれに個別の値を有し、これを M_j と示す。時間平滑化手段202による時間平滑化の方法は限定されるものではなく、任意の方法を適用することができる。時間平滑化の代表的な方法には、移動平均法と時定数フィルタ（リーク積分とも呼ばれる）があるが、ここでは、例として時定数フィルタを使うこととする。以下では、第2の帯域jにおける前回の平滑化音声存在確率を M'_j と示すものとする。そうすると、 M_j は、 P_j 、 M'_j 、及び時定数（アルファ）に基づいて、(7)式によって算出することができる。

20

【0068】

(7)式によって集約音声存在確率Pが時間平滑化される様子を、図7に示す。

30

【0069】

図7はサンプルデータに基づく図6の集約音声存在確率Pから得られた平滑化音声存在確率Mの例である、図7では、第2の帯域の各値 M_j を黒丸で描き、各値 M_j 間を点線で結び、平滑化音声存在確率の値 M_j が疎に並んでいる。なお、ここでは、時定数は、例えば0.6とする。

【数6】

$$M_j = \alpha M'_j + (1 - \alpha) P_j \quad (7)$$

【0070】

帯域分配手段203は、時間平滑化手段202から与えられた平滑化音声存在確率Mを第1の帯域へと分配し、得られた結合係数Cを後段の処理に与える。帯域分配手段203で用いる分配の方法としては、例えば、任意の補間法（いわゆる「内挿法」）が好適に用いられる。すなわち、第2の帯域に従って周波数軸上に疎に定義されている平滑化音声存在確率Mを、任意の補間法によってより密な第1の帯域へと値を補間し、得られた第1の帯域に従って周波数軸上に密に定義された値を結合係数Cとする。また、帯域分配手段203で用いる補間法としては、例えば、多項式補間、キュービック補間、スプライン補間などを用いることができるが、特に線形補間が好適に用いられる。

40

【0071】

第1の実施形態における帯域分配手段203は、(8)式及び(9)式によって、平滑

50

化音声存在確率Mを第1の帯域へと分配し、結合係数Cを算出する。

【0072】

図8はサンプルデータに基づく図7の平滑化音声存在確率Mから得られた結合係数Cの例である、図8では、第1の帯域の各値 C_k 間を実線で結んでいる。

【数7】

$$w_{k,j} = \begin{cases} \frac{\phi_j - f_k}{\phi_j - f_{k(j-1)}} & (f_{k(j-1)} < f_k \leq \phi_j) \\ \frac{f_k - \phi_j}{f_{k(j)} - \phi_j} & (\phi_j < f_k < f_{k(j)}) \\ 0 & (\text{otherwise}) \end{cases} \quad (8) \quad 10$$

$$C_k = \sum_{j=1}^J w_{k,j} M_j \quad (9)$$

【0073】

20

(A-3) 第1の実施形態の効果

第1の実施形態によれば、以下のような効果を奏することができる。

【0074】

ゲイン修正手段304では、従来の音声強調方法によって算出された、音声成分を強調できるが歪みを生じてしまう原ゲインと、所定の定数値で与えられた、歪みを生じないが音声成分を強調できない無歪ゲインとを、音声存在確率に従って周波数方向と時間方向とに滑らかに変化する結合係数に基づいて凸結合することによって修正ゲインを得る。そして、音声処理装置300では、ゲイン修正手段304によって得られた修正ゲインを用いて音声強調を行うので、従来の音声強調方法と同等の音声強調性能を維持しながら歪みを生じない音声強調方法を提供できる。言い換えると、音声処理装置300(音声処理装置300)では、SS法やMMSE-STSA法をはじめとした従来の目的音強調方法と同等の目的音強調効果を維持しながら、目的音区間では歪みを軽減し、さらに雑音区間では歪みを生じずに(特性を不自然に変化させずに)、雑音成分を抑圧して音声成分を強調することができる。

30

【0075】

(B) 第2の実施形態

以下、本発明によるゲイン処理装置及びプログラム、並びに、音響信号処理装置及びプログラムの第2の実施形態を、図面を参照しながら詳述する。

【0076】

(B-1) 第2の実施形態の構成

40

図9は、第2の実施形態の音声処理装置600の構成を示すブロック図である。図9では、上述の図2と同一部分又は対応部分には、同一符号又は対応符号を付している。

【0077】

以下では、第2の実施形態について第1の実施形態との差異を説明する。

【0078】

第2の実施形態の音声処理装置600では、ゲイン修正手段304がゲイン修正手段604に置き換わっている点で第1の実施形態と異なっている。ゲイン修正手段604は、入力スペクトルXと雑音スペクトルNと原ゲインGとに基づいて修正ゲインHを算出する。

【0079】

50

図10は、第2の実施形態のゲイン修正手段604内部の機能的構成について示したブロック図である。図10では、上述の図1と同一部分又は対応部分には、同一符号又は対応符号を付している。

【0080】

第2の実施形態のゲイン修正手段604は、原ゲインGに基づいて無歪ゲインFを算出するゲイン平滑化手段404が追加されている点で第1の実施形態と異なっている。

【0081】

次に、ゲイン平滑化手段404の詳細構成について図11を用いて説明する。

【0082】

図11に示すように、ゲイン平滑化手段404は、帯域集約手段501、時間平滑化手段502、及び帯域分配手段503を有している。

【0083】

帯域集約手段501は、与えられた原ゲインGの帯域を第3の帯域へと集約して、集約ゲイン（大文字のガンマ）を得るものである。

【0084】

時間平滑化手段502は、帯域集約手段501から与えられた集約ゲインを時間平滑化して、平滑化ゲイン（大文字のファイ）を得るものである。

【0085】

帯域分配手段503は、時間平滑化手段502から与えられた平滑化ゲインを第1の帯域へと分配し、得られた無歪ゲインFを後段の処理に与えるものである。

【0086】

（B-2）第2の実施形態の動作

次に、以上のような構成を有する第2の実施形態の音声処理装置600の動作を、図9を用いて説明する。なお、第2の実施形態の音声処理装置600では、第1の実施形態と異なる部分はゲイン修正手段604だけであるため、以下では、ゲイン修正手段604を中心とした動作についてのみ説明し、他の部分の動作については説明省略する。

【0087】

ゲイン修正手段604は、周波数解析手段301から与えられた入力スペクトルXと、雑音推定手段302から与えられた雑音スペクトルNと、ゲイン算出手段303から与えられた原ゲインGとに基づいて、修正ゲインHを算出し、得られた修正ゲインHを乗算手段305に与える。

【0088】

次に、上述した構成を有する第2の実施形態のゲイン修正手段604の動作について、図10を用いて説明する。なお、第2の実施形態のゲイン修正手段604では、第1の実施形態と異なる部分はゲイン平滑化手段404だけであるため、以下では、ゲイン平滑化手段404を中心とした動作についてのみ説明し、他の部分の動作については説明省略する。

【0089】

ゲイン平滑化手段404は、与えられた原ゲインGを周波数方向と時間方向に平滑化した値を無歪ゲインとして算出する。ゲイン平滑化手段404が行う平滑化の方法は任意であり、例えば画像処理で用いられるような2次元の移動平均法や2次元ガウシアンフィルタなどを適用できるが、特に図11で示される構成を適用することが望ましい。

【0090】

次に、ゲイン平滑化手段404の詳細動作を、図11を参照しながら説明する。

【0091】

帯域集約手段501は、与えられた原ゲインGの帯域を第3の帯域へと集約し、得られた集約ゲイン（大文字のガンマ）を時間平滑化手段502に与える。第2の帯域は、入力信号の帯域を、第1の帯域数K個より少ない第3の帯域数J個に分割した帯域である。第3の帯域の定め方は任意であるが、第1の実施形態における第2の帯域と同様とするのが好適である。

10

20

30

40

50

【0092】

帯域集約手段501は、(10)式のように、集約ゲイン Γ_j のj番目の値を、原ゲインGの $k_{(j-1)} \sim k_{(j)} - 1$ の平均値とする。

【数8】

$$\Gamma_j = \frac{1}{k_{(j)} - k_{(j-1)}} \sum_{l=k_{(j-1)}}^{k_{(j)}-1} G_l \quad (10)$$

10

【0093】

なお、以上では、第3の帯域に集約する方法として、第1の帯域の所定の範囲の平均値を取るとしたが、最大値、最小値、中央値、重み付き平均値のいずれかを取るようにしても良い。

【0094】

時間平滑化手段502は、帯域集約手段501から与えられた集約ゲイン Γ_j を時間平滑化して、得られた平滑化ゲイン Φ_j (大文字のファイ) を帯域分配手段503に与える。平滑化ゲイン Φ_j は、集約ゲイン Γ_j と同様に、第3の帯域それぞれに個別の値を有し、これを Φ_j と示す。時間平滑化手段502による時間平滑化の方法は限定されるものではなく、任意の方法を適用することができる。時間平滑化手段502に適用される時間平滑化の代表的な方法には、例えば、移動平均法と時定数フィルタがある。ここでは、例として、時間平滑化手段502に時定数フィルタを用いるものとして説明する。例えば、第3の帯域jにおける前回の平滑化音声存在確率を Φ_j' と示すことにすると、 Φ_j は、 Φ_j' 、 Γ_j 及び時定数 β に基づいて、(11)式によって算出することができる。(11)式において、時定数 β は、例えば0.99とするようにしてもよい。

20

【数9】

$$\Phi_j = \beta \Phi_j' + (1 - \beta) \Gamma_j \quad (11)$$

【0095】

帯域分配手段503は、時間平滑化手段502から与えられた平滑化ゲイン Φ_j を第1の帯域へと分配し、得られた無歪ゲインFを後段の処理に与える。帯域分配手段503が行う分配の方法としては、例えば、任意の補間法(いわゆる「内挿法」)が好適に用いられる。すなわち、帯域分配手段503は、第3の帯域に従って周波数軸上に疎に定義されている平滑化ゲイン Φ_j を、任意の補間法によってより密な第1の帯域へと値を補間し、得られた第1の帯域に従って周波数軸上に密に定義された値を無歪ゲインFとする。また、帯域分配手段503で適用される補間法としては、例えば、多項式補間、キュービック補間、スプライン補間などが挙げられるが、特に線形補間が好適に用いられる。

30

【0096】

第2の実施形態における帯域分配手段503は、例えば、(12)式によって、平滑化ゲイン Φ_j を第1の帯域へと分配し、無歪ゲインFを算出する。ここで、(12)式の $w_{k,j}$ は、第1の実施形態における帯域分配手段203と同様に、(8)式で求めることができる。

40

【数10】

$$F_k = \sum_{j=1}^J w_{k,j} \Phi_j \quad (12)$$

50

【0097】

(B-3) 第2の実施形態の効果

第2の実施形態によれば、第1の実施形態の効果に加えて以下のような効果を奏することができる。

【0098】

第2の実施形態のゲイン修正手段604では、無歪ゲインを原ゲインに基づいて算出する。これにより、ゲイン修正手段604では、入力信号の特性によって変化する原ゲインに対して適応した無歪ゲインを適用することができ、原ゲインが優勢な帯域及び時刻と無歪ゲインが優勢な帯域及び時刻との差異が適応的に調整される。これにより、第2の実施形態の音声処理装置600では、音声強調性能を維持しながら、入力信号の特性に適応してより歪みを低減させることができる。

10

【0099】

(C) 他の実施形態

本発明は、上記の各実施形態に限定されるものではなく、以下に例示するような変形実施形態も挙げることができる。

【0100】

(C-1) 上記の各実施形態では、強調する目的音を音声としているが、音声以外の音(例えば、機械音や操作音等の音)を目的音(強調対象となる音響信号)としても良い。

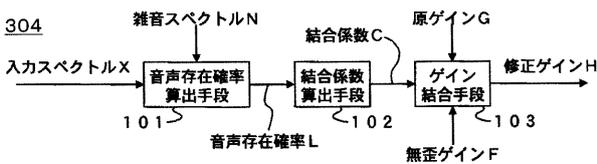
【符号の説明】

【0101】

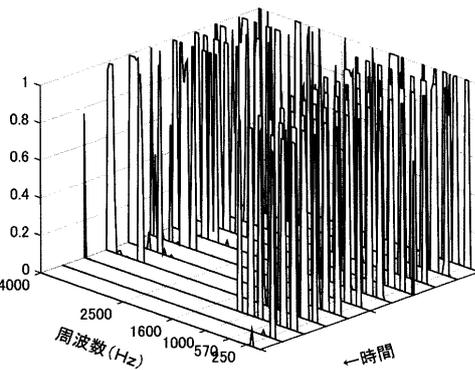
300...音声処理装置、301...周波数解析手段、302...雑音推定手段、303...ゲイン算出手段、304...ゲイン修正手段、305...乗算手段、306...波形復元手段、304...ゲイン修正手段、101...音声存在確率算出手段、102...結合係数算出手段、103...ゲイン結合手段、201...帯域集約手段、202...時間平滑化手段、203...帯域分配手段。

20

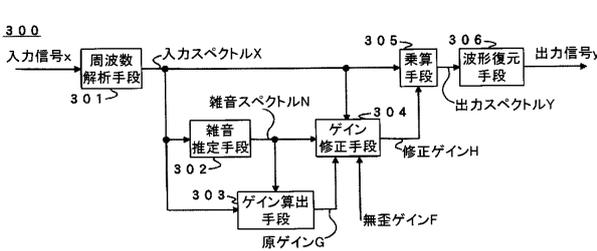
【図1】



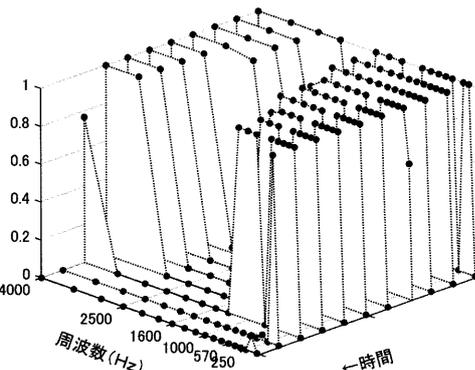
【図5】



【図2】



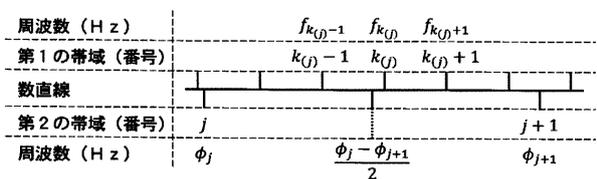
【図6】



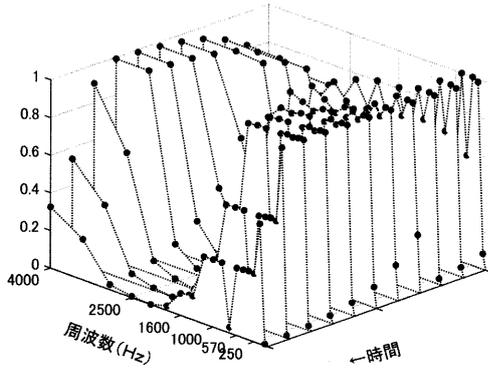
【図3】



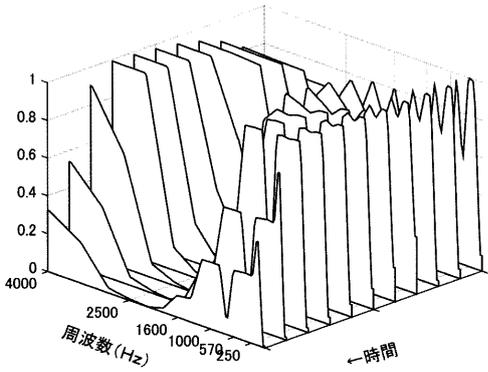
【図4】



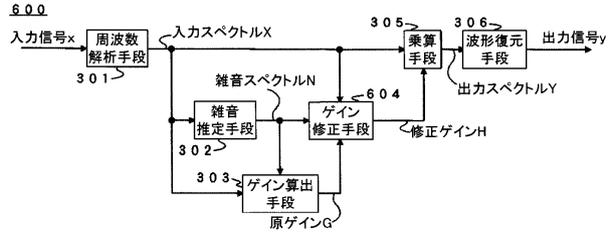
【図7】



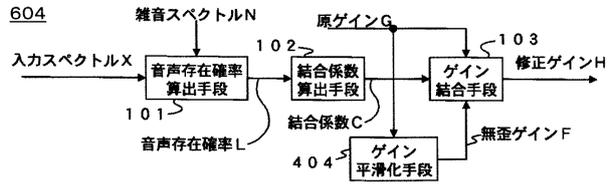
【図8】



【図9】



【図10】



【図11】



フロントページの続き

(56)参考文献 米国特許出願公開第2016/0042746 (US, A1)

特開2001-134287 (JP, A)

特開2010-102199 (JP, A)

特開2012-150227 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 21/00 - 25/93