

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5071346号
(P5071346)

(45) 発行日 平成24年11月14日(2012.11.14)

(24) 登録日 平成24年8月31日(2012.8.31)

(51) Int.Cl. F I
G 1 O L 21/02 (2006.01) G 1 O L 21/02 1 O 2 Z
 G 1 O L 21/02 1 O 1 B

請求項の数 8 (全 20 頁)

<p>(21) 出願番号 特願2008-274939 (P2008-274939) (22) 出願日 平成20年10月24日(2008.10.24) (65) 公開番号 特開2010-102204 (P2010-102204A) (43) 公開日 平成22年5月6日(2010.5.6) 審査請求日 平成23年8月23日(2011.8.23)</p>	<p>(73) 特許権者 000004075 ヤマハ株式会社 静岡県浜松市中区中沢町10番1号 (74) 代理人 100125689 弁理士 大林 章 (74) 代理人 100125335 弁理士 矢代 仁 (74) 代理人 100121108 弁理士 高橋 太朗 (72) 発明者 劉 恩彩 静岡県浜松市中区中沢町10番1号 ヤマ ハ株式会社内 審査官 山下 剛史</p>
---	---

最終頁に続く

(54) 【発明の名称】 雑音抑圧装置及び雑音抑圧方法

(57) 【特許請求の範囲】

【請求項1】

入力信号に含まれる音声の有無を検出する音声検出手段と、
 前記音声検出手段による検出結果及び前記入力信号に基づいて、当該入力信号に含まれる雑音スペクトルを推定する第1雑音スペクトル推定手段と、
 前記音声検出手段による検出結果の如何に関わらず、前記入力信号に基づいて前記雑音スペクトルを推定する第2雑音スペクトル推定手段と、
 前記音声検出手段が前記音声を継続的に検出する時間の長さに応じ、かつ、前記第1及び第2雑音スペクトル推定手段それぞれによる推定結果たる第1及び第2雑音スペクトル推定値に基づく、雑音スペクトル最終推定値を算出する雑音スペクトル算出手段と、
 前記雑音スペクトル最終推定値に基づいて、雑音抑圧ゲインを算出するゲイン演算手段と、
 前記入力信号に前記雑音抑圧ゲインを適用することで、当該入力信号に含まれる雑音を抑圧する雑音抑圧手段と
 を備えることを特徴とする雑音抑圧装置。

【請求項2】

前記雑音スペクトル算出手段は、
 前記第1及び第2雑音スペクトル推定値に乘算するための、前記時間の長さに応じて値の異なる重み付け係数を算出する係数算出部を含み、
 当該雑音スペクトル算出手段は、

前記重み付け係数を用いた、前記第 1 及び第 2 雑音スペクトル推定値についての加重平均値を算出することで、前記雑音スペクトル最終推定値を算出する、
ことを特徴とする請求項 1 に記載の雑音抑圧装置。

【請求項 3】

前記雑音スペクトル算出手段は、
前記時間の長さに応じて動作し、かつ、前記第 1 雑音スペクトル推定値を初期値とし前記第 2 雑音スペクトル推定値が入力されることで動作する積分器を含む、
ことを特徴とする請求項 1 又は 2 に記載の雑音抑圧装置。

【請求項 4】

前記雑音スペクトル算出手段は、
前記時間の長さが第 1 所定値以下である場合は、
前記第 1 雑音スペクトル推定値をそのまま、前記雑音スペクトル最終推定値として算出し、
前記時間の長さが第 2 所定値以上である場合は、
前記第 2 雑音スペクトル推定値をそのまま、前記雑音スペクトル最終推定値として算出し、
前記時間の長さが、第 1 所定値を上回り、かつ、第 2 所定値を下回る場合は、
前記第 1 及び第 2 雑音スペクトル推定値間の中間値を、前記雑音スペクトル最終推定値として算出する、
ことを特徴とする請求項 1 乃至 3 のいずれか一項に記載の雑音抑圧装置。

10

20

【請求項 5】

前記中間値は、
前記第 1 及び第 2 雑音スペクトル推定値間を繋ぐように、前記時間の長さに応じて異なる値をとる、
ことを特徴とする請求項 4 に記載の雑音抑圧装置。

【請求項 6】

前記音声検出手段は、
前記入力信号のうち所定の周波数帯域に属する部分に関する前記音声の有無、を検出する、
ことを特徴とする請求項 1 乃至 5 のいずれか一項に記載の雑音抑圧装置。

30

【請求項 7】

入力信号に含まれる音声の有無を検出する音声検出工程と、
前記音声検出工程による検出結果及び前記入力信号に基づいて、当該入力信号に含まれる雑音スペクトルを推定する第 1 雑音スペクトル推定工程と、
前記音声検出工程による検出結果の如何に関わらず、前記入力信号に基づいて前記雑音スペクトルを推定する第 2 雑音スペクトル推定工程と、
前記音声検出工程における前記音声の継続的な検出時間の長さに応じ、かつ、前記第 1 及び第 2 雑音スペクトル推定工程それぞれによる推定結果たる第 1 及び第 2 雑音スペクトル推定値に基づく、雑音スペクトル最終推定値を算出する雑音スペクトル算出工程と、
前記雑音スペクトル最終推定値に基づいて、雑音抑圧ゲインを算出するゲイン演算工程と、
前記入力信号に前記雑音抑圧ゲインを適用することで、当該入力信号に含まれる雑音を抑圧する雑音抑圧工程と
を含むことを特徴とする雑音抑圧方法。

40

【請求項 8】

前記雑音スペクトル算出工程は、
前記第 1 及び第 2 雑音スペクトル推定値に乗算するための、前記時間の長さに応じて値の異なる重み付け係数を算出する係数算出工程を含み、
当該雑音スペクトル算出工程では、
前記重み付け係数を用いた、前記第 1 及び第 2 雑音スペクトル推定値についての加重平

50

均値が算出されることで、前記雑音スペクトル最終推定値が算出される、
ことを特徴とする請求項 7 に記載の雑音抑圧方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、雑音抑圧装置及び雑音抑圧方法に関する。

【背景技術】

【0002】

従来、入力信号に応じてスピーカ等の負荷を駆動する音声再生装置、遠隔地間で音声を
伝達する音声通信装置、音声の種別等を区別・認識することでその意味等を理解する音声
認識装置、等々が提案されている。これら各装置では、音声を正確に再生し、伝達し、あ
るいは認識等するため、そこに含まれる雑音の影響が除去されるのが好適である。

10

そのような雑音抑圧の技術としては、例えば以下の特許文献 1 及び 2 並びに非特許文献
1 及び 2 に開示されているようなものが知られている。

【特許文献 1】特開 2006 - 126859 号公報

【特許文献 2】米国特許第 5572621 号

【非特許文献 1】Boll, S., "Suppression of acoustic noise in speech using spectral
subtraction", IEEE Trans. Vol. ASSP-27, No2, pp.113-120, 1979.

【非特許文献 2】Doblinger G., "Computationally efficient speech enhancement by
spectral minima tracking in subbands", Proc. of Eurospeech EUROSPEECH '95, pp.15
13-1516

20

【発明の開示】

【発明が解決しようとする課題】

【0003】

これら各文献に開示される技術は、基本的に、周波数領域の振幅スペクトルのレベルを
適当な手法によって差し引くことによって雑音を抑圧する手法、即ちいわゆるスペクトラ
ル・サブトラクション法 (Spectral Subtraction) に関連し、これらいずれの技術によっ
ても、一定の雑音抑圧効果が享受され得る。

【0004】

30

しかしながら、これらの文献においては未開示・未解決の課題はなおある。

例えば、前記のスペクトラル・サブトラクション法は、入力信号中に含まれる雑音スペ
クトルを推定し、その雑音スペクトル推定値を、振幅スペクトルから差し引く手法に基づ
くが、この場合、その雑音スペクトル推定値の正確性を確保することが困難だという問題
がある。すなわち、ここでいう雑音スペクトル推定値は、それがあくまでも“推定値”で
ある以上、必ずしも、実際の雑音スペクトルを反映するとは限らないのである。そうす
ると、その結果、より差し引くべきところで差し引き控えが生じ、あるいはその逆に、差
引いてはならないところで差し引きの過剰が生じる、などといった事象を発生させること
になるので、結局、正確な雑音抑圧ができないという不具合が生じることになる。

【0005】

40

前述の特許文献 1 は、入力信号に音声が含まれているかどうかを検出し、音声のない期
間のスペクトルの平均値を雑音スペクトルとして推定する手法を提案し、特許文献 2 及び
非特許文献 2 は、音声検出を行わず、入力信号から直接的に雑音スペクトルを推定する手
法を提案する。

しかしながら、これらのいずれにしても、それら雑音スペクトルが“推定値”である
という事情は何ら変わらないから、原理的な観点を徹底すれば、前述の不具合を強力に解決
し得る手段を提供しているとはいえない。

【0006】

また、前記の各文献を個別的にみても、以下のような問題がある。すなわち、第 1 に、
特許文献 2 では、有限時間窓内における入力信号のスペクトルを平滑化処理した値の中か

50

ら最小値をみつけ、これを雑音スペクトルとする手法が開示されているが、これでは、当該のスペクトルを記憶しておくための記憶容量が比較的大きくなり、また、最小値を求めるための演算量も大きくなる、などといった問題がある。

また、この特許文献2では、前述したような「最小値」を用いて雑音スペクトルを推定するために、入力信号に雑音しか含まれていない場合における雑音スペクトル推定値を過小に評価してしまう傾向がある。また、同じ理由から、入力信号に音声信号が含まれていると、前記「最小値」を用いたとしても、今度は、雑音スペクトル推定値を過大に評価してしまう傾向がある。前者によれば、ほとんど抑圧されていない雑音のいわば垂れ流しとなり、後者によれば、差し引き過ぎによる音質劣化の問題が生じる。このような不具合は、前記の非特許文献2でも同様に生じ得る。

10

【0007】

さらに、前記の特許文献1は、雑音のレベルが急変する場合に必ずしもうまく対応することができないという問題がある。すなわち、特許文献1では、入力信号のレベルが急変すると、それを音声信号であるとみなして、雑音スペクトル推定値の更新が行われなくなるが、そのレベルの急変は、本当は“雑音”のレベルに係るものであったという場合、問題がある。しかも、この状況が継続すると、雑音スペクトル推定値の未更新も継続してしまうことになるため、結局、その急変したレベルをもつ雑音が例えば“再生”され続けるという問題が生じることになってしまう。

【0008】

本発明は、上述した課題の少なくとも一部を解決することの可能な雑音抑圧装置及び雑音抑圧方法を提供することを課題とする。

20

【課題を解決するための手段】

【0009】

本発明に係る雑音抑圧装置は、上述した課題を解決するため、入力信号に含まれる音声の有無を検出する音声検出手段と、前記音声検出手段による検出結果及び前記入力信号に基づいて、当該入力信号に含まれる雑音スペクトルを推定する第1雑音スペクトル推定手段と、前記音声検出手段による検出結果の如何に関わらず、前記入力信号に基づいて前記雑音スペクトルを推定する第2雑音スペクトル推定手段と、前記音声検出手段が前記音声を継続的に検出する時間の長さに応じ、かつ、前記第1及び第2雑音スペクトル推定手段それぞれによる推定結果たる第1及び第2雑音スペクトル推定値に基づく、雑音スペクトル最終推定値を算出する雑音スペクトル算出手段と、前記雑音スペクトル最終推定値に基づいて、雑音抑圧ゲインを算出するゲイン演算手段と、前記入力信号に前記雑音抑圧ゲインを適用することで、当該入力信号に含まれる雑音を抑圧する雑音抑圧手段と、を備える

30

【0010】

本発明によれば、第1及び第2雑音スペクトル推定手段によって、音声検出手段による検出結果に基づいた第1雑音スペクトル推定値と、それに関わらない第2雑音スペクトル推定値とが取得される。前者は、音声検出の結果を踏まえた推定値なので、より信頼性が高く、また、音声信号の有無に応じた、使用すべき雑音スペクトルの選択（ひいては、雑音抑圧処理態様の選択）等が可能となるので、雑音抑圧効果もより実効的になる可能性が高い。他方、後者は、そのような効果は薄い、音声検出の結果に関わらない推定値であるので、その推定の根拠は原理的には入力信号それ自体に尽きるということもでき、それゆえにこそ、そのレベルの変動に対しては非常に優れた追随性を示す可能性が高い。

40

とはいえ、これらそれぞれの推定値に関しては、上述したように、前記第1雑音スペクトル推定値に関しては“雑音レベル”が急激に変化した場合にこれを有効に抑圧することが困難である点、及び、前記第2雑音スペクトル推定値に関しては雑音スペクトルの過小評価又は過大評価が生じ得る点、において、それぞれ難がある。

【0011】

しかるに、本発明においては、雑音スペクトル算出手段が、音声を継続的に検出する時間の長さに応じて、雑音スペクトル最終推定値を算出するようになっている。

ここで、この「時間の長さ」は、「音声を継続的に検出」した場合における時間の長さ

50

であるが、實際上、人が話す際には息継ぎ等が必要となること等々より、入力信号中に音声信号が非常に長い間含まれている可能性は極めて小さいと考えられることに鑑みるに、その、「音声」を検出したものとされた「時間」中において、実は、「雑音」しか存在しない時間が含まれている可能性（言い換えると、その「音声」の検出は、単に、“雑音レベルの上昇の検出”に過ぎない可能性）が否定されない。むしろ、当該の「時間」が長ければ長いほど、そのことが強く推定される。

本発明は、このような場合に好適に対応可能である。すなわち、前述のように、「雑音スペクトル最終推定値」は、前記「時間の長さ」に応じて算出されるので、例えば、当該の時間の長さが極めて長くなるような場合は、「音声」検出の「時間」は継続しているものの、実は入力信号は雑音信号のみということがあり得るので、それを有効に抑圧すべく、前記第2雑音スペクトル推定値の影響を強めた雑音スペクトル最終推定値を算出する、

10

などといった対応をとることが可能となっているのである。いずれにしても、本発明によれば、前記「時間の長さ」に応じた第1及び第2雑音スペクトル推定値の単純な使い分け等を含む各種の利用態様が選択可能であるので、時に、第1雑音スペクトル推定値に係る前述の利点を生かし（又は、その欠点を殺し）、時に、第2雑音スペクトル推定値に係る前述の利点を生かす（又は、その欠点を殺す）、などといった配慮に基づき、雑音スペクトル最終推定値の算出を行うことが可能になる。

【0012】

この発明に係る雑音抑圧装置では、前記雑音スペクトル算出手段は、前記第1及び第2雑音スペクトル推定値に乘算するための、前記時間の長さに応じて値の異なる重み付け係数を算出する係数算出部を含み、当該雑音スペクトル算出手段は、前記重み付け係数を用いた、前記第1及び第2雑音スペクトル推定値についての加重平均値を算出することで、前記雑音スペクトル最終推定値を算出する、ように構成してもよい。

20

この態様によれば、雑音スペクトル最終推定値の算出が好適に行われる。

特に、本態様では、音声検出の継続時間の長さに応じた「重み付け係数」を用いた「加重平均値」が求められることから、前述した、本発明に係る効果、即ち第1及び第2雑音スペクトル推定値それぞれの利点を生かし、欠点を殺すという効果が、いわば重疊的に奏され得ることになる（ここで「重疊的に」とは、定性的に言えば、第1及び第2雑音スペクトル推定値のそれぞれの利点を半分ずつ程度享受するという目的で、両者それぞれの影響を被った雑音スペクトル最終推定値を算出する、などという場合を想定している。以下

30

【0013】

また、本発明に係る雑音抑圧装置では、前記雑音スペクトル算出手段は、前記時間の長さに応じて動作し、かつ、前記第1雑音スペクトル推定値を初期値とし前記第2雑音スペクトル推定値が入力されることで動作する積分器を含む、ように構成してもよい。

この態様によれば、雑音スペクトル最終推定値の算出が好適に行われる。

特に、本態様では、上述のような「積分器」が含まれていることから、時間の長さに応じた、第1雑音スペクトル推定値から第2雑音スペクトル推定値への連続的な推移、等を実現することができ、そのような態様において、前述した本発明に係る効果がやはり、いわば重疊的に奏され得ることになる。

40

【0014】

また、本発明に係る雑音抑圧装置では、前記雑音スペクトル算出手段は、前記時間の長さが第1所定値以下である場合は、前記第1雑音スペクトル推定値をそのまま、前記雑音スペクトル最終推定値として算出し、前記時間の長さが第2所定値以上である場合は、前記第2雑音スペクトル推定値をそのまま、前記雑音スペクトル最終推定値として算出し、前記時間の長さが、第1所定値を上回り、かつ、第2所定値を下回る場合は、前記第1及び第2雑音スペクトル推定値間の中間値を前記雑音スペクトル最終推定値として算出する、ように構成してもよい。

この態様によれば、雑音スペクトル最終推定値の算出が好適に行われる。

特に、本態様では、前記時間の長さとの大小関係に応じて、雑音スペクトル

50

最終推定値が算出されるようになっているので、前述したような「音声」の誤検出の場合等に好適に対応することが可能である。

なお、本態様にいう「中間値」なる概念は、1個の値しかとりえない定数を意味する場合があることは勿論、後述するように連続的な値をとる一群の数の集合や、あるいは場合によっては、段階的な値をとる一群の数の集合、等々をも意味し得る。また、「中間」というのも、厳密な意味での中点（即ち、第1及び第2雑音スペクトル推定値間を結ぶ線分上の点であって、これら各推定値から等距離にある点）を意味しない。

【0015】

この態様では、前記中間値は、前記第1及び第2雑音スペクトル推定値間を繋ぐように、前記時間の長さに応じて異なる値をとる、ように構成してもよい。

10

この態様によれば、前記「中間値」が好適に特徴付けられるが、この場合、この「中間値」は、前述のように、前記「時間の長さ」が第1所定値を上回り第2所定値を下回る場合に用いられる概念として規定されていることから、本態様において、「時間の長さに応じて異なる値をとる」という場合における「時間の長さ」に限っては、その下限が第1所定値によって、その上限が第2所定値によって画されている。

このことから、本態様の「中間値」は、時間の長さが第1所定値以下の場合の第1雑音スペクトル推定値と、第2所定値以上の場合の第2雑音スペクトル推定値との間で連続的に変化する値（の一群）として想定されることになる。そして、そのような「中間値」は、第1及び第2雑音スペクトル推定値それぞれの性質を適度に併せ持つものといえる。

以上のことから、本態様によっても、前述した本発明に係る効果が、いわば重疊的に奏され得ることになる。

20

【0016】

また、本発明に係る雑音抑圧装置では、前記音声検出手段は、前記入力信号のうち所定の周波数帯域に属する部分に関する前記音声の有無、を検出する、ように構成してもよい。

この態様によれば、音声検出をより精度高く行うことが可能となり、あるいは、音声検出に係る演算速度等の向上が図られる。これは、本態様に言う「周波数帯域」が、例えば、予め音声が含まれていると強く推測される周波数帯域を含み得ることによる。

【0017】

一方、本発明に係る雑音抑圧方法は、上記課題を解決するため、入力信号に含まれる音声の有無を検出する音声検出工程と、前記音声検出工程による検出結果及び前記入力信号に基づいて、当該入力信号に含まれる雑音スペクトルを推定する第1雑音スペクトル推定工程と、前記音声検出工程による検出結果の如何に関わらず、前記入力信号に基づいて前記雑音スペクトルを推定する第2雑音スペクトル推定工程と、前記音声検出工程における前記音声の継続的な検出時間の長さに応じ、かつ、前記第1及び第2雑音スペクトル推定工程それぞれによる推定結果たる第1及び第2雑音スペクトル推定値に基づく、雑音スペクトル最終推定値を算出する雑音スペクトル算出工程と、前記雑音スペクトル最終推定値に基づいて、雑音抑圧ゲインを算出するゲイン演算工程と、前記入力信号に前記雑音抑圧ゲインを適用することで、当該入力信号に含まれる雑音を抑圧する雑音抑圧工程と、を含む。

30

40

【0018】

本発明によれば、上述した、本発明に係る雑音抑圧装置に関して述べた作用効果と本質的に異なる作用効果が奏されることが明白である。

【0019】

この発明に係る雑音抑圧方法では、前記雑音スペクトル算出工程は、前記第1及び第2雑音スペクトル推定値に乗算するための、前記時間の長さに応じて値の異なる重み付け係数を算出する係数算出工程を含み、当該雑音スペクトル算出工程では、前記重み付け係数を用いた、前記第1及び第2雑音スペクトル推定値についての加重平均値が算出されることで、前記雑音スペクトル最終推定値が算出される、ように構成してもよい。

この態様によれば、上述した、本発明に係る雑音抑圧装置の各種態様のうち、「加重平

50

均値」を算出する態様に関して述べた作用効果と本質的に異なる作用効果が奏されることが明白である。

【0020】

以上のほか、本発明の、より具体化された態様、あるいはそれによって奏される作用効果は、すぐ後から始まる実施形態における記載において明らかにされる。

【発明を実施するための最良の形態】

【0021】

<第1実施形態>

以下では、本発明に係る第1の実施の形態について図1及び図2を参照しながら説明する。なお、ここに言及した図1及び図2に加え、以下で参照する各図面(例えば図4等のグラフをも含む。)においては、各部の寸法の比率が実際のものとは適宜に異ならせてある場合がある。

10

【0022】

雑音抑圧装置1は、図1に示すように、時間・周波数変換部10、スペクトル演算部15、第1雑音スペクトル推定部21、第2雑音スペクトル推定部23、加重平均部30、雑音抑圧ゲイン演算部60、及び周波数・時間変換部70、並びに、音声検出部80、カウンタ部85、及び加重係数演算部90からなる。

【0023】

時間・周波数変換部10は、時間領域の入力信号にフーリエ変換をかけて、周波数領域の信号に変換する。このフーリエ変換は、入力信号を経時的に所定数のフレームに分け、かつ、そのフレームに適当な窓関数をかけることを通じて行われるのが好適である。

20

【0024】

スペクトル演算部15は、前記の周波数領域の信号が分けられた振幅スペクトル及び位相スペクトル、あるいはパワースペクトルを演算する。第1実施形態では特に、以下の式(1)に基づいて、パワースペクトルを算出する。

【数1】

$$P(n) = Y_r^2(n) + Y_i^2(n) \dots\dots (1)$$

30

ここで、 $P(n)$ は入力信号のパワースペクトル、 $Y_r(n)$ は入力信号をフーリエ変換した結果得られる実部、 $Y_i(n)$ は入力信号をフーリエ変換した結果得られる虚部、 n は周波数帯域(に付けられた番号。なお、周波数帯域は N 個に分割される。)である。

【0025】

音声検出部80は、入力信号の中の音声信号の有無を検出する。前述のように、入力信号がフレームに分けられる場合には、フレームごとの音声検出が行われる(なお、第1実施形態では、このような処理を前提とする。)。ここで「音声」とは特に、会話、話し言葉、音楽、各種の信号等々、人にとって有意味な音響を意味する。つまり、入力信号を適当な再生手段によって再生するとした場合、その入力信号中の「音声信号」を再生すれば、当該音響となる、という関係が成立する(もっとも、第1実施形態は、実は雑音であるものを、このような意味における「音声」として認識してしまう場合に着目する。この点についての詳細は、以後の説明において追々明らかにされる。)

40

【0026】

この音声信号は、例えば、以下の式(2)が利用されて検出される。

【数2】

$$SNR = 10 \log \left(\frac{PS}{PN} \right) \dots\dots (2)$$

50

ただし、この式(2)中、PS、PNは以下の各式によって表現される。

【数3】

$$PS = \sum_{n=N1}^{N2} P(n) \quad \dots\dots (3)$$

【数4】

$$PN = \sum_{n=N1}^{N2} N_{t-1}(n) \quad \dots\dots (4)$$

10

ここで、N1、N2は、周波数帯域の番号(したがって、N1・N2はともに、前述したN以下の整数である。)、P(n)は入力信号のパワースペクトル(前記式(1))、N_{t-1}(n)は現時点から1時点前に処理において推定された、雑音についてのパワースペクトルである。

これにより、PSは、帯域N1から帯域N2までの入力信号の総パワースペクトル、PNは、同じく帯域N1から帯域N2までの推定された雑音の総パワースペクトルとなる。

以上の式(3)及び式(4)から、式(2)のSNRは、入力信号の総パワースペクトルと、推定した雑音の総パワースペクトルとのレベル差を表す。

20

【0027】

そして、音声検出部80は、この式(2)によって算出されるSNRが、例えば所定値TH1以上となる場合(即ち、SNR ≥ TH1)は、音声あり、そうでない場合(即ち、SNR < TH1)は、音声なし、などと判断する。なお、この所定値TH1は、音声検出のための閾値を意味し、具体的には例えば6[dB]等というように設定される。

また、音声検出部80は、このようにして判断された、音声信号の有無情報を保持する。例えば、音声検出部80は、音声信号の有無に応じて、値1又は0となる音声検出フラグsp_flgをもつ(sp_flg = 1で“音声あり”、sp_flg = 0で“音声なし”)。このように、音声検出部80は、現時点において処理中である入力信号に音声信号が含まれているかどうかを常に把握する。

30

【0028】

第1雑音スペクトル推定部21は、入力信号に基づいて第1の雑音スペクトルを推定する。第1実施形態では特に、以下の式(5)に基づいて、第1雑音スペクトルが算出される。

【数5】

$$N1_t(n) = \begin{cases} \mu N_{t-1}(n) + (1-\mu)P(n); & \text{case}\cdot A \\ N_{t-1}(n); & \text{case}\cdot B \end{cases} \quad \dots\dots (5)$$

40

ここで、N1_t(n)は、現に処理中であるフレームにおける第1雑音スペクトル推定値、N_{t-1}(n)は、その直前に処理したフレームにおける雑音スペクトル推定値(したがって、“t”は、現に処理中であるフレームそれ自体を表現する添え字である。)、μは平滑化係数である。また、式(1)中、case・Aとあるのは、前記のフラグsp_flg = 0の場合(つまり、現時点、音声なし)を表現し、case・Bとあるのは、フラグsp_flg = 1の場合(つまり、現時点、音声あり)を表現している。

このように、第1雑音スペクトル推定部21は、現時点において音声検出部80が音声信号を検出しているかどうかに応じて、第1雑音スペクトル推定値N1_t(n)を求めるために利用する式を変更する。すなわち、sp_flg = 1の時には、その直前の雑音スペクトル推定値N_t(n)をそのまま用いて第1雑音スペクトル推定値N1_t(t)を求

50

め、 $sp_{flg} = 0$ の時には、入力したパワースペクトルを時間軸上で平滑化することで、第 1 雑音スペクトル推定値 $N1_t(n)$ を求める。

【0029】

第 2 雑音スペクトル推定部 23 は、入力信号に基づいて第 2 の雑音スペクトルを推定する。第 1 実施形態では特に、以下の式 (6) 及び式 (7) に基づいて、第 2 雑音スペクトルが算出される。

【数 6】

$$PA_t(n) = \alpha PA_{t-1}(n) + (1-\alpha)P(n) \dots\dots (6)$$

10

【数 7】

$$N2_t = \begin{cases} \gamma N2_{t-1}(n) + \frac{1-\gamma}{1-\beta} (PA_t(n) - \beta PA_{t-1}(n)); & \text{case} \cdot C \\ PA_t(n); & \text{case} \cdot D \end{cases} \dots\dots (7)$$

ここで、 $PA_t(n)$ は、現に処理中であるフレームにおけるパワースペクトルであって平滑化されたもの、 $PA_{t-1}(n)$ は、その直前のフレームにおける当該パワースペクトルであって平滑化されたもの、 α は平滑化係数、である。

20

また、 $N2_t(n)$ は、現に処理中であるフレームにおける第 2 雑音スペクトル推定値、 $N2_{t-1}(n)$ は、その直前に処理したフレームにおける第 2 雑音スペクトル推定値、 γ は制御パラメータである。また、式 (7) 中、 $\text{case} \cdot C$ とあるのは、 $PA_t(n) > N2_{t-1}(n)$ が成立する場合を表現し、 $\text{case} \cdot D$ とあるのは、それ以外の場合を表現している。

【0030】

この場合、式 (7) の $\text{case} \cdot D$ として示される式及び式 (6) の組み合わせが前記の式 (5) の $\text{case} \cdot A$ として示される式と実質的にほぼ同義である。

他方、式 (7) の $\text{case} \cdot C$ として示される式は、前記式 (5) 中においては該当するものはない。もっとも、この式は、上述のように、即ち $PA_t(n) > N2_{t-1}(n)$ が成立する場合、即ち、現に処理中であるフレームにおけるパワースペクトルが、その直前のフレームにおける第 2 雑音スペクトル推定値を超える場合に発動するから、この $\text{case} \cdot C$ は、その現に処理中であるフレームが、音声信号を含んでいる可能性を示唆するものと捉えることも不可能でない（仮に、多くの $n (= 0, 1, 2, 3, \dots)$ について、かかる条件が満たされるなら、その可能性はより高まるといえる。もっとも、あくまでも“示唆”に過ぎない。）。

30

これら式 (6) 及び式 (7) は、以上のような意味における限りで、前記式 (5) と共通性をもつといえる。

【0031】

ちなみに、前記の式 (5) との対比において、これら式 (6) 及び式 (7) で重要なことは、式 (5) とは異なって、音声検出の結果を必要とすることなく、雑音スペクトル推定値を算出可能であるということにある。この点については、後に改めて触れる。

40

【0032】

加重平均部 30 は、前述した第 1 雑音スペクトル推定値 $N1_t(n)$ 、及び、第 2 雑音スペクトル推定値 $N2_t(n)$ 、並びに、後述する加重係数演算部 90 によって算出された加重係数 w を用いて、これら各推定値についての加重平均値を求める。すなわち、

【数 8】

$$N_t(n) = w \cdot N1_t(n) + (1-w)N2_t(n) \dots\dots (8)$$

50

である。この加重平均値 N_t が、最終的に適用されるべき雑音スペクトル推定値である。

【0033】

カウンタ部85は、音声検出部80において音声が発見された回数をフレームを単位としてカウントする。つまり、このカウンタ部85は、前述の音声検出フラグ sp_flag が、1となるごとに、1ずつ増加する変数をもつ。この変数を CNT とすると、 $sp_flag = 1$ ならば、 $CNT = CNT + 1$ 、 $sp_flag = 0$ ならば、 $CNT = 0$ とする。

これにより、カウンタ部85は、音声検出の継続時間を提供する。

【0034】

加重係数演算部90は、前述の加重平均部30における演算に供すべき加重係数 w を演算する。第1実施形態では特に、以下の式(9)に基づいて、加重係数 w が算出される。

【数9】

$$w = \begin{cases} 1; & caseE \\ 1 - \frac{CNT - TH2}{TH3 - TH2}; & caseF \\ 0; & caseG \end{cases} \quad \dots\dots (9)$$

ここで、 $TH2$ は加重係数調整下限時、 $TH3$ は加重係数調整上限時である。また、式(9)中、 $case \cdot E$ とあるのは、 $CNT = TH2$ の場合を表現し、 $case \cdot F$ とあるのは、 $TH2 < CNT < TH3$ の場合を表現し、 $case \cdot G$ とあるのは、それら以外の場合を表現する。前記 $TH2$ 及び $TH3$ のそれぞれが、加重係数調整下限時、及び、その上限時と名付けられているのは、この式(9)の $case \cdot F$ として示される式が、 $TH2 < CNT < TH3$ のときに発動することに由来している。

【0035】

このような式(9)によって算出される加重係数 w は、例えば図2に示されるようなグラフを描く。この図2において、前述の $TH2$ は本来「回数」を表現するが、横軸の時間軸上の $0.5 [s]$ の地点に相当するものとして描かれており、同じく $TH3$ は $1.0 [s]$ の地点に相当するものとして描かれている。 CNT は、上述のように音声検出部80による音声継続の回数を表現するが、その値に相当する時間の長さは一義的に決定可能である。いずれにしても、 CNT と、 $TH2$ 又は $TH3$ とは、同じ次元(「回数」あるいは「時間」)で比較可能である。

これらのことから、図2に示す加重係数 w は、 $0.5 [s]$ を境に、時間の増加に連れ、即ち CNT の増大に連れて、次第に小さくなっていく。そして、そのような加重係数 w の調整期限は、 $1.0 [s]$ でもって画される(以上、式(9)の $case \cdot F$)。また、 CNT が、 $0.5 [s]$ 以下に相当する時間を表現する場合は、 $w = 1.0$ (式(9)の $case \cdot E$)、 $1.0 [s]$ 以上に相当する時間を表現する場合は、 $w = 0$ (式(9)の $case \cdot G$)となる。なお、図2においては、このような加重平均 w の変動に伴う $1 - w$ の変動が破線でもって示されている。

以上のことと、前述した式(8)の表式とから、結局、加重係数演算部90は、 CNT が大きくなるに従い、第1雑音スペクトル推定値 $N_{1t}(n)$ の影響を小さくし、かつ、第2雑音スペクトル推定値 $N_{2t}(n)$ の影響を大きくするような、加重係数 w を定めることになる。

【0036】

雑音抑圧ゲイン演算部60は、前記振幅スペクトルと、式(8)で求められた雑音スペクトル推定値 $N_t(n)$ とに基づいて、雑音抑圧ゲインを算出する。第1実施形態では特に、以下の式(10)により雑音抑圧ゲインを算出する。

10

20

30

40

【数 10】

$$G(n) = \max\left(0, \frac{Y(n) - N_t(n)}{Y(n)}\right) \dots\dots (10)$$

ここで、 $\max(a, b)$ は、 a 及び b のうちいずれか大きい値を返す関数を意味する。

この式(10)により、入力された振幅スペクトル $Y(n)$ に対する雑音スペクトル推定値 $N_t(n)$ との間において、 $Y(n) < N_t(n)$ が成立する場合は、 $G(n) = 0$ となり、 $Y(n) > N_t(n)$ が成立する場合は、 $G(n) = (Y(n) - N_t(n)) / Y(n)$ となる。 10

【0037】

図1に示す乗算器11は、以上のようにして求められた雑音抑圧ゲイン $G(n)$ を、振幅スペクトル $Y(n)$ にかける。すなわち、 $S(n) = G(n) \cdot Y(n)$ なる演算が行われ、その結果、最終的に得るべき雑音抑圧後の振幅スペクトル $S(n)$ が得られる。

最後に、周波数・時間変換部70は、以上のようにして求められた雑音抑圧後の振幅スペクトル $S(n)$ と、時間・周波数変換部10から直接的に供給される位相スペクトルに基づいて、時間領域の出力信号を生成する。第1実施形態では、時間・周波数変換部10においてフーリエ変換がかけられているので、周波数・時間変換部70は、逆フーリエ変換を実施する。 20

【0038】

次に、以上に述べた第1実施形態に係る雑音抑圧装置1の作用ないし動作及び効果について、既に参照した図1及び図2に加えて、図3乃至図8を参照しながら説明する。

まず、時間・周波数変換部10は、入力信号に対して、フーリエ変換を施し、これを周波数領域の信号に変換する(図3のステップS101)。この際、時間・周波数変換部10は、前述のように、フレームごとの処理を実施する。そして、スペクトル演算部15は、前記式(1)に従って、入力信号のパワースペクトル $P(n)$ を算出する(図3のステップS102)。 30

【0039】

続いて、音声検出部80は、前述で求められたパワースペクトル $P(n)$ 及び式(2)~式(4)に基づいて、入力信号中に音声信号が含まれるか否かを判断する(図3のステップS201)。

この場合、例えば音声検出の精度を高めること等その他の目的をもって、式(3)及び式(4)中の $N1$ 及び $N2$ が適当に設定され得る。すなわち、音声信号が含まれているであろう周波数帯域が所定の帯域に限定されることが予め推測されるときは、全周波数帯域について音声信号の有無を検出するよりも、その所定の帯域をカバーするように $N1$ 及び $N2$ を設定して、音声信号の有無を検出するほうが合理的である。この場合、それにより音声検出の精度がより高まり、あるいは演算速度が向上する、等といった利点が享受され得る(ただし、 $N1$ 及び $N2$ が、全周波数帯域をカバーするように設定されることを排除するわけではない。) 40

いずれにせよ、以上のようにして帯域 $N1$ 及び $N2$ 間に音声信号が含まれるかどうか判断され、これが肯定されるときは、音声検出フラグ sp_flag が立てられ(即ち、 $sp_flag = 1$)、否定されるときは寝かされる(即ち、 $sp_flag = 0$)。

そして、カウンタ部85は、この音声検出部80による音声検出作業において、 $sp_flag = 1$ が成立する限りは、前記変数 CNT を増加させ続ける(図3のステップS202及びS203参照)。

【0040】

このような音声検出処理に並行して、雑音抑圧装置1では、第1雑音スペクトル推定部21及び第2雑音スペクトル推定部23が、第1雑音スペクトル推定値 $N1_t(n)$ 及び 50

第2雑音スペクトル推定値 $N_{2t}(n)$ を算出する（図3のステップS103及びS104）。

まず、第1雑音スペクトル推定部21は、前述の式(5)に基づいて、第1雑音スペクトル推定値 $N_{1t}(n)$ を算出する。この算出処理は、前述のように、 sp_flg の値の別に応じて異なる。すなわち、 $sp_flg = 0$ が成立する場合は、式(5)の $case \cdot A$ として示される式により時間軸上の平滑化処理が行われることによって、 $sp_flg = 1$ が成立する場合は、式(5)の $case \cdot B$ として示される式により前フレームの雑音スペクトル推定値 $N_t(n)$ がそのまま、今フレームの第1雑音スペクトル推定値 $N_{1t}(n)$ にあてられる。

また、第2雑音スペクトル推定部23は、前述の式(6)及び式(7)に基づいて、第2雑音スペクトル推定値 $N_{2t}(n)$ を算出する。この算出処理は、前述のように、現に処理中であるフレームにおけるパワースペクトルであってその平滑化処理を受けたものたる $PA_t(n)$ と、直前に処理したフレームにおける第2雑音スペクトル推定値 $N_{2t-1}(n)$ との大小関係に応じて、異なる。すなわち、 $PA_t(n) > N_{2t-1}(n)$ が成立する場合は、式(7)の $case \cdot C$ として示される式により、そうでない場合は $case \cdot D$ として示される式により、それぞれ第2雑音スペクトル推定値 $N_{2t}(n)$ は求められる。

【0041】

このようにして算出される第1及び第2雑音スペクトル推定値 $N_{1t}(n)$ 及び $N_{2t}(n)$ は、例えば図4及び図5に示されるように表現される。

まず、これら各図のうち図4は、入力信号波形の一例を表し、図5は、その図4に示される入力信号に基づいて算出されたパワースペクトルの一例を表している。図4は、いわば生の信号であり、図5は、前記式(1)による算出後（あるいは、図3のステップS102の処理後）のパワースペクトル $P(n)$ をそれぞれ表現しているともいえる。なお、図5は、周波数250〔Hz〕前後の周波数帯域を特に示している。また、図5の左上端付近に示された破線矢印は、表示限界を超えてなおレベルが上昇したが、その図示が省略化されていることを表現している。

このような図4及び図5において、第1雑音スペクトル推定値 $N_{1t}(n)$ は、図5中に示す曲線S1として、第2雑音スペクトル推定値 $N_{2t}(n)$ は、同じく曲線S2として、それぞれ示されている。この場合、注意すべき点が2つある。

【0042】

その1つは、例えば図4中の符号M1で示すような雑音信号のレベルが急変している部分（以下、「雑音レベル急変部M1」という。）が存在しているにもかかわらず、第1雑音スペクトル推定値 $N_{1t}(n)$ ないし曲線S1は、それに追従するような挙動を示していない。これは、前記式(5)において、 $case \cdot B$ として示される式が利用されてしまった結果である。

本来、当該の式は、 $sp_flg = 1$ 、即ち“音声あり”の場合に発動するはずであるが、音声検出部80、あるいは前記式(2)から式(4)に基づく音声検出方法によると、図4の雑音レベル急変部M1のように、本当は単に雑音信号のレベルが急変しただけのポイントを、音声信号ありと誤って認定してしまう可能性は完全には捨てきれない。図5に描かれる曲線S1の結果は、そのためである。つまり、雑音レベル急変部M1が音声信号ありと誤認定された結果、第1雑音スペクトル推定値 $N_{1t}(n)$ は、従前の雑音スペクトル推定値 $N_t(n)$ に維持されてしまい、本当は抑圧すべき雑音信号であるのに、そのレベルに追従しなくなってしまうのである。

【0043】

他方、注意すべき点の2つめは、以下のようなものである。すなわち、第2雑音スペクトル推定値 $N_{2t}(n)$ に関しては、前述したような第1雑音スペクトル推定値 $N_{1t}(n)$ についてみられる不都合はなく、図5に示すように、パワースペクトル $P(n)$ の変動によく追従している。しかし他方、例えば図5中の符号M2で示すような音声信号ありの部分（以下、「音声信号部M2」という。）において推定された第2雑音スペクトル推定値 $N_{2t}(n)$

10

20

30

40

50

$2_t(n)$ の値は、若干過大となってしまう。

これらの結果はいずれも、前記式(6)及び式(7)に内在する特徴に因っている。つまり、第2雑音スペクトル推定値 $N_{2_t}(n)$ は、音声検出結果を必要としない算出原理に基づいて求められており、その算出根拠は殆どパワースペクトル $P(n)$ のみにあるということになっているから、パワースペクトル $P(n)$ の変動によく追従する反面、パワースペクトル $P(n)$ の値によっては、過大評価を行ってしまうおそれがあるのである。

このような過大評価された第2雑音スペクトル推定値 $N_{2_t}(n)$ をそのまま用いるかたちで、前記音声信号部 M2 に関する雑音抑圧処理を実行してしまうと、必要以上の雑音抑圧を行ってしまうこととなり、また、音質の劣化が懸念されることになる。

【0044】

以上のようにして第1及び第2雑音スペクトル推定値 $N_{1_t}(n)$ 及び $N_{2_t}(n)$ が求められたら、次に、前記変数 CNT の値に応じた、加重係数 w の設定が行われる(図3のステップ S106 乃至 S110 参照)。すなわち、図1の加重係数演算部 90 は、CNT が、加重係数調整下限時 TH2 以下となるときは、 $w = 1$ に設定され(図3のステップ S108)、加重係数調整上限時 TH3 以上となるときは、 $w = 0$ に設定される(図3のステップ S109)。上記いずれの場合にも該当しない場合、即ち $TH2 < CNT < TH3$ が成立するときは、 $w = p$ に設定される(図3のステップ S110)。ここで p は、前述の式(9)の $case \cdot F$ として示される式の右辺に一致し、より具体的には例えば、図2に従って定められる値である。

【0045】

次に、加重平均部 30 は、この加重係数 w 及び前述の式(8)に従って、雑音スペクトル推定値 $N_t(n)$ を求める(図3のステップ S301)。

この場合、重要なのは、前記加重係数 w が CNT の値の大きさに応じて定められていることから、雑音スペクトル推定値 $N_t(n)$ もまた、その CNT の値の大きさに応じて定められることにある。すなわち、CNT が小さい、即ち音声継続の時間がさほど長くないときには、雑音スペクトル推定値 $N_t(n)$ は、第1雑音スペクトル推定値 $N_{1_t}(n)$ の値に等しくなり、CNT が大きい、即ち音声継続の時間が極めて長いときには、雑音スペクトル推定値 $N_t(n)$ は、第2雑音スペクトル推定値 $N_{2_t}(n)$ の値に等しくなる。

そして、これらの中間の場合には、雑音スペクトル推定値 $N_t(n)$ は、第1及び第2雑音スペクトル推定値 $N_{1_t}(n)$ 及び $N_{2_t}(n)$ が適度に調合された値に等しくなる。なお、この場合においては、CNT の値の増加に連れて、雑音スペクトル推定値 $N_t(n)$ を決定するに当たっての、第1雑音スペクトル推定値 $N_{1_t}(n)$ の影響が次第に低下していく反面、第2雑音スペクトル推定値 $N_{2_t}(n)$ の影響が次第に強まっていく。

【0046】

以上の処理を経ると、最終的に適用すべき雑音スペクトル推定値 $N_t(n)$ が得られるが、上で参照した図5においては、その結果も表現されている。すなわち、図5において、この雑音スペクトル推定値 $N_t(n)$ は、曲線 Sf として示されている。この雑音スペクトル推定値 $N_t(n)$ ないし曲線 Sf は、前述した加重平均化処理を経ているために、図5に示すように、パワースペクトル $P(n)$ の変動に対する追従性を一定程度保持しながら、また、音声信号部 M2 において過大なスペクトル値を推定するという事もない。なお、図5に示す曲線 S1、S2 及び Sf を見やすくするため、それらの各々のみを抜き出した図面を、図6乃至図8として示しておいた。

【0047】

最後に、雑音抑圧ゲイン演算部 60 は、このような雑音スペクトル推定値 $N_t(n)$ を用い、かつ、前記式(10)に従い、雑音抑圧ゲイン $G(n)$ を算出する。そして、この雑音抑圧ゲイン $G(n)$ に、元の振幅スペクトル $Y(n)$ をかければ(図1に示す乗算器 11 参照)、雑音抑圧後の振幅スペクトル $S(n)$ が算出される(図3のステップ S302)。

【0048】

10

20

30

40

50

以上に述べたような構成及び作用をもつ雑音抑圧装置 1 によれば、以下の効果が奏される。

(1) 第 1 実施形態の雑音抑圧装置 1 によれば、雑音スペクトル推定値 $N_t(n)$ が、現実に入力信号に含まれているであろう雑音スペクトルをよりよく反映するから、そこに含まれる雑音が極めて好適に抑圧される。これは既に述べたように、CNT の大きさに応じた、雑音スペクトル推定値 $N_t(n)$ の設定が行われるようになっているからである。

【 0 0 4 9 】

一般に、入力信号中に音声信号が非常に長い間含まれている可能性は極めて小さい。このような推定は、例えば、人の会話が行われる場合においては話者の交代等の場面ではしばしば無音の合間が介在すること、あるいはまた、一人の人間が話し続けるとしても途中で必ず息継ぎ等が必要であること、等から合理的な根拠をもつ。

10

そうすると、前記 CNT があまりに大きな値をもつことは不自然ということになるから、その CNT の増加途上においては、本来音声信号と判定してはならない場面があったのではないかと、換言すると、雑音信号を音声信号と誤認定してしまっている場合があるのではないかと、ということが強く推定されることになる。

【 0 0 5 0 】

図 3 のステップ S 1 0 7 ; Y E S からステップ S 1 0 9 へと至る処理は、このような事情を酌んでいる。つまり、CNT が、加重係数調整上限時 TH 3 を越えるような場合には、前述したような不都合な事態が発生している可能性が大きいので、加重係数 w を 0 にする、つまり、図 2 あるいは式 (8) に従って、雑音スペクトル推定値 $N_t(n)$ を、第 2 雑音スペクトル推定値 $N_{2t}(n)$ に等しくしてしまうことで、音声検出の結果をいわば無視するようにするのである。これにより、パワースペクトル $P(n)$ によく追従する第 2 雑音スペクトル推定値 $N_{2t}(n)$ に従った雑音抑圧処理が行われることになるので、雑音信号に対する好適な対応 (要するに、その効果的な抑圧) が可能となる。

20

【 0 0 5 1 】

なお、このような処理は、いわば異常な事態への好適な対応という側面をもつが、その意味では、図 3 のステップ S 1 0 6 ; Y E S からステップ S 1 0 8 へと至る処理 (加重係数 w を 1 とする処理) の流れは、“通常の処理” という意味合いをもつ。実際、音声検出結果を利用した雑音スペクトルの推定の方が、そうではない推定よりも信頼性が高く、また、第 1 実施形態では、前述の式 (5) のように、音声信号の有無に応じて第 1 雑音スペクトル推定値 $N_{1t}(n)$ の値の定め方が異なり、かつ、そのそれぞれに応じた適切な雑音抑圧処理が行われるようになっているから、雑音抑圧効果もより実効的になる可能性が大きい。したがって、このような処理を、“通常の処理” として認識することは正当といえる (もっとも、ここでいう “異常” ないし “通常” とは、その実質が「音声の誤検出の可能性」の有無にのみ関連するのであって、何かそれ以上の重大な事態を想定しているわけでないことは言うまでもない。) 。

30

【 0 0 5 2 】

(2) 一方、図 3 のステップ S 1 0 7 ; N O からステップ S 1 1 0 へと至る処理の流れは、上の説明との対比から明らかなように、上記 2 つの場合の中間的な場合の処理という意味合いを持つことになる。すなわち、第 1 実施形態によれば、CNT が、非常に大きいともいえないが、上述した観点からすると誤認定のおそれがあり得る値をもつ領域においては、上記 2 つの場合における処理のいわば折衷的な処理が行われる。しかも、第 1 実施形態では、そのような場合であっても、当該領域において機械的に求められた一定の中間値を一律に用いるというのではなくて、CNT の値に応じた、加重係数 w の調整が行われ (図 2 等参照) 、その都度、最適な雑音スペクトル推定値 $N_t(n)$ の算出が試みられる。

40

このような処理の仕方は、ドラスティックな処理の変更を伴わない分、より現実に対応した対応であるといえ、前述した雑音レベルの急変に対応した雑音抑圧、及び、雑音スペクトルの過大評価の回避による音質劣化の回避等の 2 つの効果バランスよく享受するためには極めて有効ということが出来る。このことは、第 1 実施形態によって奏される大きな

50

効果の1つである。

【0053】

<第2実施形態>

以下では、本発明に係る第2の実施の形態について図9を参照しながら説明する。なお、この第2実施形態は、上記第1実施形態との対比において、第1及び第2雑音スペクトル推定値 $N1_t(n)$ 及び $N2_t(n)$ 間の混合処理に関連する相違点があり、その他の点については、特に断りがない限り上記第1実施形態と全く同じである。したがって、以下では、前記相違点に関する説明を主に行い、その他の点についての説明は簡略化ないし省略する。また、図面上の符号についても前記相違点以外については流用する。

【0054】

この第2実施形態の雑音抑圧装置1'は、図9に示すように、積分器40、スイッチ45及びスイッチング信号発生部95をもつ。このうちスイッチング信号発生部95は、スイッチ45を制御するためのスイッチング信号 sw_flg を発生する。このスイッチング信号 sw_flg は、図1に示すように、カウンタ部85からの入力に依る。すなわち、スイッチング信号 sw_flg が発生するかどうかは、音声検出の継続時間を表すCNTの値に基づく。好適には例えば、CNT>TH4の場合は、スイッチング信号 sw_flg を発生させ(即ち、 $sw_flg=1$ とし)、それ以外の場合は発生させない(即ち、 $sw_flg=0$ とする)、などという基準を設けることができる(以下、これを前提とする。)

なお、図9において、スイッチ45の通常状態は、同図に示すように、第1雑音スペクトル推定部21の側と接続されている場合とし、 $sw_flg=1$ となるときは、反対側の第2スペクトル推定部23の側と接続されるものとする($sw_flg=0$ となれば、再び、第1雑音スペクトル推定部21と接続される。)

【0055】

積分器40は、例えばオペアンプ、容量素子、抵抗素子等から構成される(いずれも不図示)。第1実施形態では、図1に示すように、第1雑音スペクトル推定部21及び第2雑音スペクトル推定部23の双方からの出力を受ける。そして、この積分器40は、以下の式(11)に従って動作する。

【数11】

$$N_{i_t}(n) = \begin{cases} N1_t(n); & \text{case}\cdot H \\ \lambda N_{i_{t-1}}(n) + (1-\lambda)N2_t(n); & \text{case}\cdot I \end{cases} \dots\dots (11)$$

ここで、 $N_{i_t}(n)$ は、現時点の積分器40の出力、 $N_{i_{t-1}}(n)$ は、その1時点前の積分器40の出力、 λ は平滑化係数、である。また、式(11)中、 $\text{case}\cdot H$ とあるのは、CNT=TH4の場合を表現し、 $\text{case}\cdot I$ とあるのは、CNT>TH4の場合を表現している。この式(11)中のTH4は、積分器40を動作させるかどうかの閾値としての意味をもつが、後において改めて触れるように、実質的には、上記第1実施形態における加重係数調整下限時TH2と同じ機能をもつ。

【0056】

この式(11)において、 $N1_t(n)$ 及び $N2_t(n)$ は、上述のように、第1雑音スペクトル推定値、及び、第2雑音スペクトル推定値であるから、結局、積分器40は、 $\text{case}\cdot I$ の場合、つまりCNT>TH4の場合にのみ実質的に動作して、第1雑音スペクトル推定値 $N1_t(n)$ から第2雑音スペクトル推定値 $N2_t(n)$ への滑らかな接続を実現するような各値を、 t の増加に連れて出力することになる。

【0057】

このような第2実施形態の雑音抑圧装置1'は、次のように動作する。

まず、図3のステップS101からステップS104、また、図3のステップS201

10

20

30

40

50

からステップS203までは、上記第1実施形態と全く同じである。

異なるのは、CNTの使われ方である。すなわち、第1実施形態では、CNTは、加重係数 w の調整を通じて最終的に取得される雑音スペクトル推定値 $N_t(n)$ を決定するために用いられていたが、第2実施形態では、積分器40を動作させるかどうかの基準を提供する役割を担う。

すなわち、まず、CNTが次第に増加していき、やがてTH4に一致すると、それに伴って、スイッチング信号発生部95は、スイッチング信号 sw_flag を1とする。これにより、スイッチ45は、第2雑音スペクトル推定部23の側、即ち積分器40の側に切り替えられる。

また、これと同時に、積分器40は、式(11)の $case \cdot H$ として示される式を発動させる。ここで、 $N_{it}(n) = N_{1t}(n)$ が出力されるようになっているのは、当該積分器40に初期値をもたせるということとほぼ同義である。

【0058】

以後、CNTが、増加を続ける場合は、スイッチング信号 sw_flag は1をとり続け、スイッチ45は積分器40の側に接続されたままになり、当該積分器40は、式(11)の $case \cdot I$ として示される式を発動させる。これにより、積分器40は、当初、 $N_{1t}(n)$ を出力した後、次第に第2雑音スペクトル推定値 $N_{2t}(n)$ に近づくような値を出力し続けることになる。

【0059】

以上のことから、図9に示す雑音抑圧ゲイン演算部60には、CNTが閾値TH4に達しない間は、第1雑音スペクトル推定値 $N_{1t}(n)$ が入力され続け、CNTが閾値TH4に達した以降は、当初は、第1雑音スペクトル推定値 $N_{1t}(n)$ が、最終的には第2雑音スペクトル推定値 $N_{2t}(n)$ が、それぞれ入力されることになる(後者は、例えば、前記オペアンプ等に第2雑音スペクトル推定値 $N_{2t}(n)$ を入力することによって、その出力を $N_{2t}(n)$ において飽和させることで、好適に実現可能である。)

なお、このことから明らかなように、閾値TH4は、第1雑音スペクトル推定値 $N_{1t}(n)$ から第2雑音スペクトル推定値 $N_{2t}(n)$ への切り替えの基準点として機能する。これは、機能的にみて、第1実施形態におけるTH2と実質的に同じということがいえる。

また、これに関連して、式(11)中の平滑化係数は、例えば0.5[s]の時定数に相当する値として設定され得る。以上によると結局、第1実施形態として説明された、音声ありの継続時間0.5[s]を開始点とする加重係数調整時間0.5[s](図2参照)という処理とほぼ同様の処理が、第2実施形態に係る図9においても実現可能である。

【0060】

後の処理は、上記第1実施形態と同様である。すなわち、雑音抑圧ゲイン演算部60は、前記式(10)に従い、雑音抑圧ゲイン $G(n)$ を算出する。そして、この雑音抑圧ゲイン $G(n)$ に、元の振幅スペクトル $Y(n)$ をかけて、雑音抑圧後の振幅スペクトル $S(n)$ が算出される(図3のステップS302参照)。

【0061】

このような第2実施形態によれば、次のような効果が奏される。

まず、この第2実施形態によっても、上記第1実施形態によって奏された作用効果と本質的に異なる作用効果が奏されることは明白である。すなわち、この第2実施形態でも、上記第1実施形態に関し述べた(1)及び(2)の効果がほぼ同様に奏される。

また、この第2実施形態によれば、積分器40の利用により、第1雑音スペクトル推定値 $N_{1t}(n)$ から第2雑音スペクトル推定値 $N_{2t}(n)$ への滑らかな接続が可能になり、上記第1実施形態に関し述べた(2)の効果を、より実効的なものにする。

【0062】

以上、本発明に係る実施の形態について説明したが、本発明に係る雑音抑圧装置は、上述した形態に限定されることはなく、各種の変形が可能である。

10

20

30

40

50

(1) 上記第1及び第2実施形態では、音声検出が、式(2)から式(4)に基づいて行われるようになっているが、本発明は、かかる形態に限定されない。

例えば、単純に、入力信号のレベルが予め定めた閾値を超えるか否かを基準として音声を検出する手法も採用可能である。また、確率・統計的手法を用いて音声信号の発生確率を推定する手法等が採用されてもよいし、あるいは、検出対象としても、上記各実施形態のように、フーリエ変換後の入力信号を利用するのではなく、そのフーリエ変換前の信号を利用する手法等が採用されてもよい。

【0063】

(2) 上記第1実施形態では、加重係数 w が、線形的に増加する形態について説明しているが、本発明は、かかる形態に限定されない。例えば、加重係数 w は非線形的に、即ち曲線を描くようにして増加してもよい。

10

また、これに関連して、加重係数 w は、必ずしも連続的に増加する必要はない。例えば、 $w = 0.1, 0.2, \dots$ などというように、段階的に増加していてもよい。

あるいは、より一般的には、CNTが一定の値以上となるような状況が生じたときには、第1雑音スペクトル推定値 $N_{1t}(n)$ 及び第2雑音スペクトル推定値 $N_{2t}(n)$ に基づいて定められた、一定値たる雑音スペクトル推定値が算出されるようにしてもよい。

【0064】

(3) 上記第1及び第2実施形態では、雑音抑圧ゲイン $G(n)$ が、式(10)によって算出されているが、本発明は、かかる形態に限定されない。例えば、これ以外にも、ウィナーフィルタ法、MMSE (Minimum Mean-Square Error) 法等が用いられてよい(これらについては、例えば、Lim & Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", Proc. IEEE, Vol.67, No.12, pp.1586-1604, 1979、あるいは、Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. Vol. ASSP-32, No.6, pp.1109-1121, 1984.等を参照)。SNR (音声(信号)/ノイズ比率)を推定し、そのSNRに基づいて雑音抑圧ゲイン $G(n)$ が求められてもよい。

20

【図面の簡単な説明】

【0065】

【図1】本発明の第1実施形態に係る雑音抑圧装置の構成を示すブロック図である。

【図2】図1の加重係数演算部(60)で求められる加重係数 w に関するグラフである。

30

【図3】第1実施形態に係る雑音抑圧処理の流れを示すフローチャートである。

【図4】入力信号の波形例を示すグラフである。

【図5】入力信号のパワースペクトルの演算例を示すグラフである。

【図6】図5中、曲線 S_1 (第1雑音スペクトル推定値 $N_{1t}(n)$)だけを示すグラフである。

【図7】図5中、曲線 S_2 (第2雑音スペクトル推定値 $N_{2t}(n)$)だけを示すグラフである。

【図8】図5中、曲線 S_f (最終的な雑音スペクトル推定値 $N_t(n)$)だけを示すグラフである。

【図9】本発明の第2実施形態に係る雑音抑圧装置の構成を示すブロック図である。

40

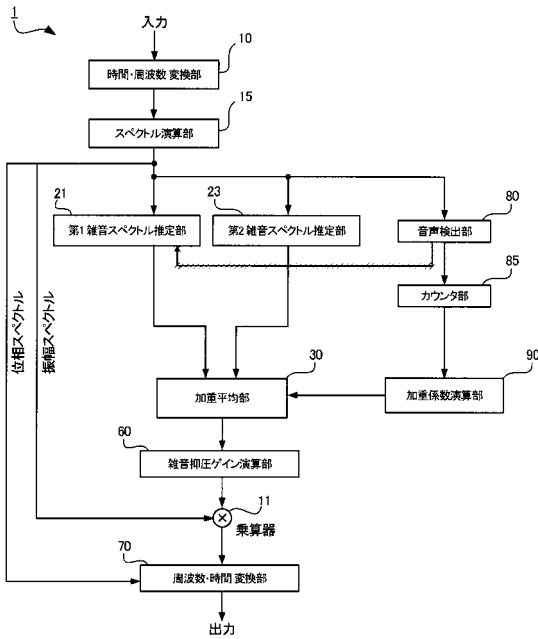
【符号の説明】

【0066】

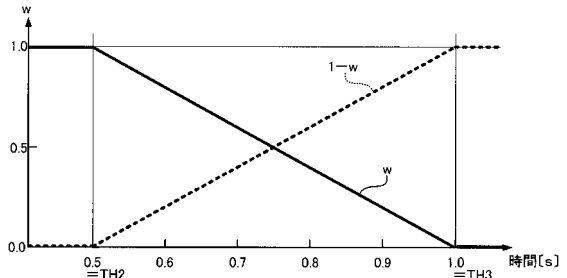
1, 1' 雑音抑圧装置、 10 時間・周波数変換部、 15 スペクトル演算部、 21 第1雑音スペクトル推定部、 23 第2雑音スペクトル推定部、 30 加重平均部、 60 雑音抑圧ゲイン演算部、 70 周波数・時間変換部、 80 音声検出部、 85 カウンタ部、 90 加重係数演算部、 11 乗算器

$N(n)$ 雑音スペクトル推定値、 $N_{1t}(n)$ 第1雑音スペクトル推定値、 $N_{2t}(n)$ 第2雑音スペクトル推定値、 $G(n)$ 雑音抑圧ゲイン

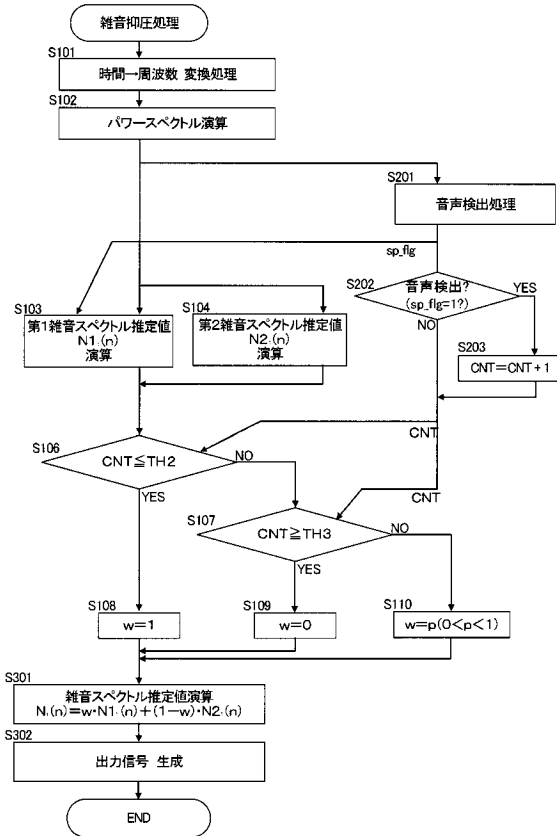
【図1】



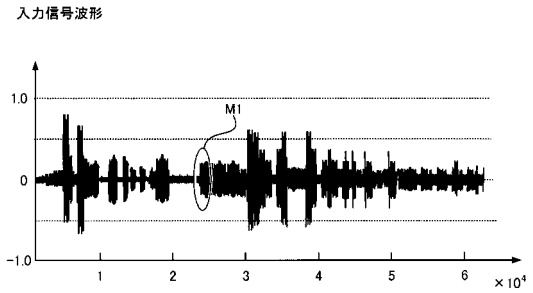
【図2】



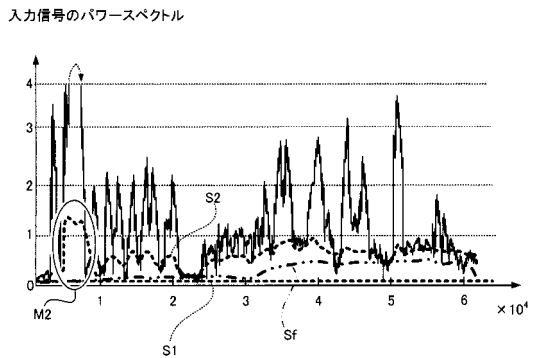
【図3】



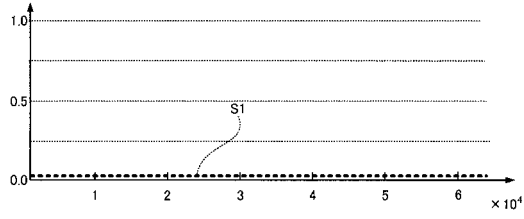
【図4】



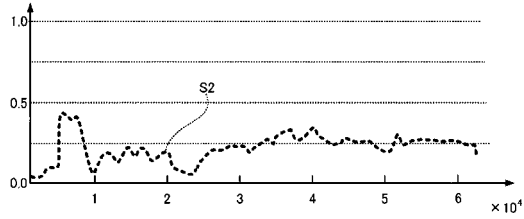
【図5】



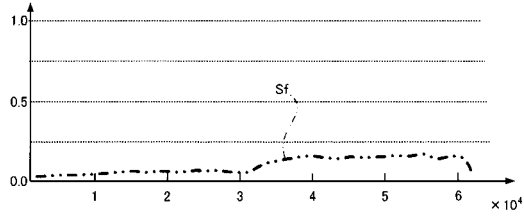
【図6】



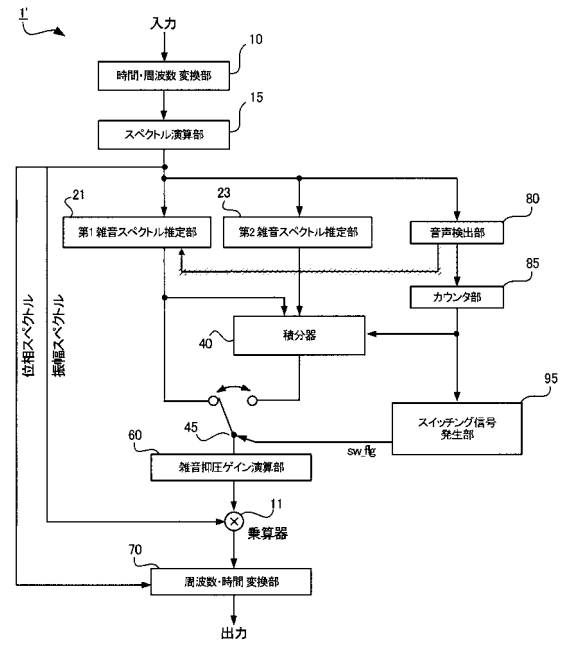
【図7】



【図8】



【図9】



フロントページの続き

- (56)参考文献 特開2004-20679(JP,A)
特開2001-159899(JP,A)
特開2006-163231(JP,A)
特開2007-212704(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G10L 15/20, 21/02