

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5077847号
(P5077847)

(45) 発行日 平成24年11月21日(2012.11.21)

(24) 登録日 平成24年9月7日(2012.9.7)

(51) Int.Cl. F I
G 1 O K 15/00 (2006.01) G 1 O K 15/00 L

請求項の数 9 (全 22 頁)

<p>(21) 出願番号 特願2008-95540 (P2008-95540) (22) 出願日 平成20年3月4日(2008.3.4) (65) 公開番号 特開2009-211021 (P2009-211021A) (43) 公開日 平成21年9月17日(2009.9.17) 審査請求日 平成23年1月14日(2011.1.14)</p> <p>特許法第30条第1項適用 平成20年2月13日に国立大学法人北陸先端科学技術大学院大学の修士論文審査会で発表</p>	<p>(73) 特許権者 304024430 国立大学法人北陸先端科学技術大学院大学 石川県能美市旭台一丁目1番地 (74) 代理人 100125645 弁理士 是枝 洋介 (72) 発明者 鶴木 祐史 石川県能美市旭台一丁目1番地 国立大学 法人北陸先端科学技術大学院大学内 (72) 発明者 平松 壮太 石川県能美市旭台一丁目1番地 国立大学 法人北陸先端科学技術大学院大学内</p> <p>審査官 富澤 直樹</p>
---	---

最終頁に続く

(54) 【発明の名称】 残響時間推定装置及び残響時間推定方法

(57) 【特許請求の範囲】

【請求項1】

残響が付加された時系列の音響信号に基づいて、前記音響信号に対応する時系列のパワーエンベロープを生成するパワーエンベロープ生成手段と、

前記パワーエンベロープ生成手段によって生成されたパワーエンベロープに基づいて、周波数系列の変調スペクトルを生成する変調スペクトル生成手段と、

前記変調スペクトル生成手段によって生成された変調スペクトルに基づいて、前記音響信号が観測された系の残響特性に関する伝達関数に対応する残響時間を推定する残響時間推定手段と

を備える、残響時間推定装置。

【請求項2】

前記周波数系列の変調スペクトルにおいて周辺の周波数領域よりも大きい変調スペクトルを示す主要変調周波数を特定する主要変調周波数特定手段を更に備え、

前記残響時間推定手段は、前記周波数系列の前記変調スペクトルに前記伝達関数の逆伝達関数を適用したときに、適用後の前記主要変調周波数における変調スペクトルが、残響が付加されていない原音を示す時系列の原音信号に対応する周波数系列の変調スペクトルの前記主要変調周波数における変調スペクトルと略一致するような前記伝達関数に対応する残響時間を推定するように構成されている、請求項1に記載の残響時間推定装置。

【請求項3】

前記主要変調周波数特定手段は、前記パワーエンベロープに対する自己相関関数を求め

、前記自己相関関数がピークを示す時間シフト量の逆数を前記主要変調周波数として特定するように構成されている請求項 2 に記載の残響時間推定装置。

【請求項 4】

前記パワーエンベロープ生成手段によって生成されたパワーエンベロープに対して適用されるローパスフィルタを更に備え、

前記主要変調周波数特定手段は、前記ローパスフィルタから出力されたパワーエンベロープに基づいて、前記主要変調周波数を特定するように構成されている請求項 2 又は 3 に記載の残響時間推定装置。

【請求項 5】

前記音響信号を複数チャンネルに帯域分割する帯域分割手段と、

前記帯域分割手段によって帯域分割された各チャンネルから、残響時間推定に用いるチャンネルを決定するチャンネル決定手段とを更に備える、請求項 1 乃至 4 の何れかに記載の残響時間推定装置。

10

【請求項 6】

前記パワーエンベロープ生成手段は、前記帯域分割手段によって帯域分割された各チャンネルについて、パワーエンベロープを生成するように構成されており、

前記パワーエンベロープ生成手段によって生成されたパワーエンベロープの中で、所定の基準値を越える高レベル部を検出する高レベル部検出手段を更に備え、

前記チャンネル決定手段は、前記高レベル部検出手段によって検出された高レベル部に基づいて、残響時間推定に用いるチャンネルを決定するように構成されている、請求項 5

20

【請求項 7】

前記チャンネル決定手段は、前記高レベル部検出手段によって検出された 2 つの高レベル部の間に、微小なピークが存在するか否かを判定し、微小なピークが存在する場合には、当該チャンネルを推定に用いるチャンネルから除外するように構成されている、請求項 6 に記載の残響時間推定装置。

【請求項 8】

前記チャンネル決定手段は、前記高レベル部検出手段によって検出された高レベル部の中に谷が存在するか否かを判定し、谷が存在する場合には、当該チャンネルを推定に用いるチャンネルから除外するように構成されている、請求項 6 又は 7 に記載の残響時間推定装置。

30

【請求項 9】

残響が付加された時系列の音響信号に基づいて、前記音響信号に対応する時系列のパワーエンベロープを生成するステップと、

生成されたパワーエンベロープに基づいて、周波数系列の変調スペクトルを生成するステップと、

生成された変調スペクトルに基づいて、前記音響信号が観測された系の残響特性に関する伝達関数に対応する残響時間を推定するステップと

を備える、残響時間推定方法。

【発明の詳細な説明】

40

【技術分野】

【0001】

本発明は、系の残響時間を推定する残響時間推定装置及びその方法に関し、特に、時系列の音響信号から求められた周波数系列の変調スペクトルにより、原音信号を用いずに残響時間をブラインド推定する残響時間推定装置及びその方法に関する。

【背景技術】

【0002】

音響の残響時間は、室の残響特性を知る上で欠かせないパラメータである。残響時間は、近年では音響信号に対する F₀ 推定、音声認識、原音信号の回復等に利用されている。かかる残響時間を求めるためには、従来、室の伝達特性を測定する必要があり、この伝達

50

特性の測定には、音源、室内の人や物の状況、室内の静寂性等の多くの制約が存在する。また、従来の残響時間の測定には比較的長時間を必要とするため、刻々と残響特性が変化する場合時には測定が不可能な場合もあった。このような観点から、残響時間のリアルタイム測定が可能な手法の開発が望まれている。

【0003】

残響時間をリアルタイムに求めるためには、系の伝達特性を測定することなく、観測された音響信号のみから残響時間を推定することができるブラインド推定を行うことが必要である。ここで「系」とは、種々の事象の解析のために想定された空間のことをいう。非特許文献1には、残響の影響を受けた音響信号のパワーエンベロープから、元の音源信号のパワーエンベロープを回復する回復方法が開示されている。この非特許文献1に開示されている回復方法では、室内に伝送される音響のパワーエンベロープの入出力の関係（時間領域では畳込み、周波数領域では積）、すなわち変調度を、変調周波数を変数とした関数として表した変調伝達関数を利用している。更に詳しく説明すると、観測した音響信号のパワーエンベロープに、系の変調伝達関数の逆関数（逆フィルタ）を適用することで、元音響（残響が付加されていない音源からの音響）のパワーエンベロープを回復する。そして、非特許文献1では、残響が付加された音響信号のパワーエンベロープが、逆フィルタにより元音響のパワーエンベロープと同じ形状に回復されるときに、逆フィルタの残響時間パラメータが系の残響時間と等しい値となることを前提として、逆フィルタを算出する処理で求められる逆フィルタの残響時間パラメータを、残響時間として推定することが開示されている。

【非特許文献1】 古川、鶴木、赤木、「MTFに基づいた残響音声パワーエンベロープの回復方法」、信学技報、社団法人電子情報通信学会、平成14年4月、EA2002-15、SP2002-15、p.49-54

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、上述した非特許文献1に開示されている残響音声パワーエンベロープの回復方法において、逆フィルタを算出する処理で求められる残響時間は、残響時間が0.5秒付近までは実際の残響時間と推定残響時間とが十分に一致しているが、それ以降は徐々に両者の間の差が大きくなり、十分に一致しているとはいえないものであった。

【0005】

本発明は斯かる事情に鑑みてなされたものであり、その目的は系の伝達特性を測定することなく、正確な残響時間を推定することが可能な残響時間推定装置及び残響時間推定方法を提供することにある。

【課題を解決するための手段】

【0006】

上述した課題を解決するために、本発明の残響時間推定装置は、残響が付加された時系列の音響信号に基づいて、前記音響信号に対応する時系列のパワーエンベロープを生成するパワーエンベロープ生成手段と、前記パワーエンベロープ生成手段によって生成されたパワーエンベロープに基づいて、周波数系列の変調スペクトルを生成する変調スペクトル生成手段と、前記変調スペクトル生成手段によって生成された変調スペクトルに基づいて、前記音響信号が観測された系の残響特性に関する伝達関数に対応する残響時間を推定する残響時間推定手段とを備える。

【0007】

上記発明においては、前記周波数系列の変調スペクトルにおいて高い変調スペクトルを示す主要変調周波数を特定する主要変調周波数特定手段を更に備え、前記残響時間推定手段が、前記周波数系列の前記変調スペクトルに前記伝達関数の逆伝達関数を適用したときに、適用後の前記主要変調周波数における変調スペクトルが、残響が付加されていない原音を示す時系列の原音信号に対応する周波数系列の変調スペクトルの前記主要変調周波数における変調スペクトルと略一致するような前記伝達関数に対応する残響時間を推定する

ように構成されていることが好ましい。

【0008】

また、この場合においては、前記主要変調周波数特定手段が、前記パワーエンベローブに対する自己相関関数を求め、前記自己相関関数がピークを示す時間シフト量の逆数を前記主要変調周波数として特定するように構成されていることが好ましい。

【0009】

上記発明においては、前記パワーエンベローブ生成手段によって生成されたパワーエンベローブに対して適用されるローパスフィルタを更に備え、前記主要変調周波数特定手段が、前記ローパスフィルタから出力されたパワーエンベローブに基づいて、前記主要変調周波数を特定するように構成されていることが好ましい。

10

【0010】

上記発明においては、前記音響信号を複数チャンネルに帯域分割する帯域分割手段と、前記帯域分割手段によって帯域分割された各チャンネルから、残響時間推定に用いるチャンネルを決定するチャンネル決定手段とを更に備えることが好ましい。

【0011】

この場合においては、前記パワーエンベローブ生成手段は、前記帯域分割手段によって帯域分割された各チャンネルについて、パワーエンベローブを生成するように構成されており、前記パワーエンベローブ生成手段によって生成されたパワーエンベローブの中で、所定の基準値を越える高レベル部を検出する高レベル部検出手段を更に備え、前記チャンネル決定手段は、前記高レベル部検出手段によって検出された高レベル部に基づいて、残響時間推定に用いるチャンネルを決定するように構成されていることが好ましい。

20

【0012】

また、この場合においては、前記チャンネル決定手段が、前記高レベル部検出手段によって検出された2つの高レベル部の間に、微小なピークが存在するか否かを判定し、微小なピークが存在する場合には、当該チャンネルを推定に用いるチャンネルから除外するように構成されていることが好ましい。

【0013】

また、上記発明においては、前記チャンネル決定手段が、前記高レベル部検出手段によって検出された高レベル部の中に谷が存在するか否かを判定し、谷が存在する場合には、当該チャンネルを推定に用いるチャンネルから除外するように構成されていることが好ましい。

30

【0014】

本発明の残響時間推定方法は、残響が付加された時系列の音響信号に基づいて、前記音響信号に対応する時系列のパワーエンベローブを生成するステップと、生成されたパワーエンベローブに基づいて、周波数系列の変調スペクトルを生成するステップと、生成された変調スペクトルに基づいて、前記音響信号が観測された系の残響特性に関する伝達関数に対応する残響時間を推定するステップとを備える。

【発明の効果】

【0015】

本発明の残響時間推定装置及び残響時間推定方法によれば、系の伝達特性を測定することなく、正確な残響時間を推定することができる。

40

【発明を実施するための最良の形態】

【0016】

以下、本発明の好ましい実施の形態を、図面を参照しながら説明する。

【0017】

(実施の形態1)

実施の形態1は、ハードウェアのデジタル信号処理回路により主として構成された残響時間推定装置である。

【0018】

[残響時間推定装置の構成]

50

図1は、本発明の実施の形態1に係る残響時間推定装置の構成を示すブロック図である。図1に示すように、本実施の形態に係る残響時間推定装置1は、室の音響を入力するためのマイクロフォン2と、マイクロフォン2で取り込んだアナログ音響信号に対してA/D変換を行うA/D変換器3と、A/D変換器3から出力されるデジタル音響信号に対して信号処理を行うデジタル信号処理回路4と、デジタル信号処理回路4の処理結果を受け付け、残響時間の推定処理を実行する演算回路5と、メモリ6と、演算回路5により推定された残響時間を表示する液晶表示部7とを備えている。

【0019】

図1に示すように、デジタル信号処理回路4は、パワーエンベロープ生成部41、ローパスフィルタ42、主要変調周波数取得部43、及び正規化変調スペクトル生成部44の機能ブロックを有している。パワーエンベロープ生成部41は、A/D変換器3から出力されたデジタル音響信号から、パワーエンベロープ信号を生成する。このパワーエンベロープ信号は、音響信号の振幅の時間包絡線情報を2乗したものである。パワーエンベロープ信号生成部41は、次式により表される信号処理を実施するように構成されている。

【数1】

$$\hat{e}_v(t)^2 := \text{LPF} [|y(t) + j\text{Hilbert}(y(t))|^2] \quad (1)$$

但し、LPF[・]は低域通過フィルタを示しており、Hilbert(・)はヒルベルト変換を示している。また、会話音声の変動リズムを考慮すると、主要な音響情報は変調周波数20Hzまでに存在することから、本実施の形態では、低域通過フィルタのカットオフ周波数を20Hzとしている。

【0020】

主要変調周波数取得部43の前段には、ローパスフィルタ42が設けられており、ローパスフィルタ42によって高周波成分が除去されたパワーエンベロープ信号が主要変調周波数取得部43に与えられる。この主要変調周波数取得部43は、パワーエンベロープ信号の自己相関関数を求め、パワーエンベロープと、その自己相関関数との時間シフト量(ずれ時間)を主要変調周波数 f_{md} として決定する。また、本実施の形態では、主要変調周波数 f_{md} は比較的低い変調周波数であるので、2Hzまでに現れるものと考えられることから、ローパスフィルタ42のカットオフ周波数を2Hzとしている。このように、主要変調周波数取得部43の前段にローパスフィルタ42を設けることにより、主要変調周波数 f_{md} の誤検出を防止することができる。なお、本実施の形態においては、主要変調周波数として取得する変調周波数の数を1つとしている。ただし、これに限定されるものではなく、主要変調周波数の取得数を複数としてもよい。しかし、回路の複雑化、コスト増大等の観点より、主要変調周波数の取得数は少ない方がよく、1つ又は2つとすることが特に好ましい。なお、本実施の形態においては、ローパスフィルタ42のカットオフ周波数を2Hzとしたが、これに限定されるものではなく、5Hz又は10Hz等の他のカットオフ周波数を採用することもできる。ただし、カットオフ周波数を高い値に設定すると、パワーエンベロープにノイズの高周波成分が多く含まれるおそれがあることから、10Hz以下の値にすることが好ましい。

【0021】

正規化変調スペクトル生成部44は、FFT回路を有しており、時系列のデジタル信号であるパワーエンベロープ信号から、周波数系列のデジタル信号である変調スペクトル信号を生成する。また、この正規化変調スペクトル生成部44では、変調スペクトルの大きさを、DC成分の変調スペクトル(変調周波数が0Hzにおける変調スペクトル)で正規化した変調スペクトル信号を生成するように構成されている。

【0022】

上記のデジタル信号処理回路4から出力される主要変調周波数及び変調スペクトルのデータは、演算回路5に与えられる。演算回路5は、FPGA又はASIC等により構成された残響時間推定処理専用のプロセッサである。かかる演算回路5は、室における変調度(室内に伝送される音響のパワーエンベロープの入出力の関係)を、変調周波数を変数と

10

20

30

40

50

した関数として表した変調伝達関数（以下、MTFという）を利用して、室の残響時間を推定する。

【0023】

ここでMTFについて詳細に説明する。MTFは、伝送路の特性が線形受動システムとして記述することができる場合には、インパルス応答の自乗のフーリエ変換として次式で定義される。

【数2】

$$M(f_m) = \frac{\int_0^{\infty} h^2(t) \exp(-j2\pi f_m t) dt}{\int_0^{\infty} h^2(t) dt} \quad (2)$$

但し、 f_m は変調周波数を、 $h(t)$ は室内インパルス応答を示す。また、室内音響学でよく使われる統計近似のインパルス応答は、次式で表される。

【数3】

$$h(t) = \exp\left(-\frac{6.9t}{T_R}\right) n(t) \quad (3)$$

但し、 T_R は残響時間を、 $n(t)$ は白色雑音を示す。上記の式(3)を用いれば、MTFは、ある残響時間 T_R において、変調周波数 f_m に対する入出力パワーの割合として表される。室内が拡散音場である場合、減衰過程が指数関数的であることを、波動論を用いて示すことができる。この場合、MTFの大きさ(変調度)は、次式で示される。

【数4】

$$m(f_m) = |M(f_m)| = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-1/2} \quad (4)$$

図2は、式(4)のMTF(変調度)を示すグラフである。図2において、横軸は変調周波数を、縦軸はMTF(変調度)を示している。例えば、変調周波数 $f_m = 10\text{ Hz}$ のときのMTFは、図2に示すように0.402である。これは、入力パワーエンベロープにある変調周波数 $f_m = 10\text{ Hz}$ の成分の変調度が1であるとする、その出力パワーエンベロープの同変調周波数における成分の変調度は0.402に減衰するという意味する。この減衰特性は、ある意味で変調周波数に対する低域通過フィルタ特性と解釈することができる。

【0024】

次に、本実施の形態における残響時間推定の原理について説明する。図3Aは、残響が付加されていない音響信号のパワーエンベロープを示すグラフであり、図3Bは、その変調スペクトルを示すグラフである。また、図4Aは、図3Aで示す音響に残響が付加した場合の音響信号のパワーエンベロープを示すグラフであり、図4Bは、その変調スペクトルを示すグラフである。図3A、図4Aの音響信号の主要変調周波数は共に5 Hzであり、図4Aの音響信号に付加されている残響の残響時間 T_R は2.0秒である。図4Bの変調スペクトルに対してMTFの逆フィルタを適用することで、当該変調スペクトルの主要変調周波数(5 Hz)における値を、図3Bの変調スペクトルの同主要変調周波数の値に戻すことができれば、その逆フィルタのパラメータによって適切な残響時間を推定することが可能である。

【0025】

図3B及び図4Bに示すように、変調周波数0 Hz(DC成分)付近においては、残響の影響を受けていないことが分かる。また、MTFが入出力パワーの割合(比率)を示すこと、系のパワーエンベロープの正規化されたFourier変換がMTFであることから、元音響の変調スペクトルは残響が付加されるとその系のMTFに従い減少することがわかる。そこで、本実施の形態においては、残響付加のない音響の変調スペクトルにおいて、主要変調周波数のパワーが0 Hz付近のパワーと十分に近い場合が多いという事実に基づき、観測された音響の変調スペクトルの0 Hzにおけるパワーが、残響付加のない元音響の変調スペクトルの0 Hz及び主要変調周波数におけるパワーと一致すると仮定し、

10

20

30

40

50

元音響の変調スペクトルの主要変調周波数におけるパワーを、観測された音響の変調スペクトルの主要周波数におけるパワーまで減少させるようなM T Fを求め、そのM T Fの残響時間パラメータを系の残響時間として推定する。このことを換言すると、本実施の形態に係る残響時間推定装置は、観測された音響の変調スペクトルに対してM T Fの逆フィルタを適用したときに、当該変調スペクトルの主要変調周波数におけるパワーが、元音響の変調スペクトルの主要変調周波数におけるパワー（すなわち、観測された音響の変調スペクトルの0 H zにおけるパワー）と略一致するようなM T Fに対応する逆特性を推定する。図4 Bの破線は、元音響の変調スペクトルの主要変調周波数におけるパワーを、観測された音響の変調スペクトルの主要周波数におけるパワーまで減少させるようなM T Fを示しており、このM T Fに対応する残響時間が推定される。

10

【0026】

このようなコンセプトに基づいて、演算回路5は具体的には以下のような処理を実行するように構成されている。メモリ6には、残響時間をパラメータとして有するM T Fの逆特性が、演算回路5によって処理可能な関数データとして記憶されている。このメモリ6に記憶されている逆フィルタ（逆特性）の関数データは、推定候補として予め与えられた複数の残響時間の分だけ用意される。ここで、かかる逆フィルタについて更に詳しく説明する。M T Fの概念に基づいて、残響が付加されていない音源の音響信号（以下、音源信号という） $x(t)$ 、室内インパルス応答 $h(t)$ 、及びそれらの畳み込みとして得られる観測された音響信号（以下、残響信号という） $y(t)$ の関係を次のようにモデル化する。

20

【数5】

$$y(t) = x(t) * h(t), \quad (5)$$

$$x(t) = e_x(t) n_1(t), \quad (6)$$

$$h(t) = e_h(t) n_2(t), \quad (7)$$

$$e_x(t) = a \exp(-0.9t/T_R), \quad (8)$$

$$(n_{1k}(t) n_{2k}(t - \tau)) = \delta(\tau) \quad (9)$$

ここで、“*”は畳み込みを、 $e_x(t)$ 及び $e_h(t)$ はそれぞれ $x(t)$ 及び $h(t)$ のエンベロープを、 $n_1(t)$ 及び $n_2(t)$ は相互に無関係な白色雑音を、 a は振幅項を示している。 $n_1(t)$ と $n_2(t)$ との相互独立性から、残響信号とそのパワーエンベロープとの間には次の関係があることが知られている。

30

【数6】

$$\begin{aligned} \langle y^2(t) \rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} x(\tau) h(t - \tau, T_R) d\tau \right\}^2 \right\rangle \\ &= \int_{-\infty}^{\infty} e_x^2(\tau) e_h^2(t - \tau, T_R) d\tau \\ &= e_y^2(t) \end{aligned} \quad (10)$$

但し、 $\langle \cdot \rangle$ は集合平均を表す。この関係から、 $e^2_y(t)$ は $e^2_x(t)$ と $e^2_h(t)$ との畳み込みで得られることが分かる。

【0027】

次に、周波数領域における関係を考える。式(5)～(10)の関係式は、実際には離散時間で利用されるため、ここではその周波数変換としてz変換を利用する。z変換をZ[・]としたときに、式(8)のz領域の変調伝達関数は、次式のように表される。

40

【数7】

$$Z[e_x^2(\tau)] = \frac{a^2}{1 - \exp(-\frac{0.9}{T_R})z^{-1}} \quad (11)$$

次に、パワーエンベロープに関する入出力の関係は、 $Z[e^2_y(t)] / Z[e^2_x(t)] = Z[e^2_h(t)]$ となることから、音源信号の変調スペクトル $Z[e^2_x(t)]$ は、次式のように求めることができる。

【数 8】

$$Z[e_s^2[n]] = \frac{Z[e_y^2[n]]}{\sigma^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) \right\} z^{-1} \quad (12)$$

但し、 f_s はサンプリング周波数である。ここで、式(12)の第2項(式(11))がちょうど逆フィルタの特性(残響が積分処理であるのに対して、回復は微分処理に相当する)を表しており、1次のIIRフィルタで実現できることが分かる。この式(12)の逆フィルタを表す関数データが、メモリ6に複数の残響時間の分だけ記憶されている。演算回路5は、メモリ6から各残響時間の逆フィルタを読み出し、デジタル信号処理回路4から受け付けた主要変調周波数及び変調スペクトルに対して各残響時間の逆フィルタを適用して、各残響時間に対応する音源信号の主要変調周波数における変調スペクトルを算出するように構成されている。

10

【0028】

また、演算回路5は、このようにして求めた各残響時間に対応する音源信号の主要変調周波数における変調スペクトルのうち、大きさが0dB(変調度で表すと1)に最も近い1つを選択し、その残響時間を、室の残響時間として推定するように構成されている。この原理を、以下に説明する。

【0029】

各変調スペクトルを $E(f_m)$ とする。ここで、 $m(f_m) = E_h(f_m)$ を考慮して、式(10)の右辺の畳み込みを(対数)変調スペクトル表現で表すと、次式が得られる。

20

【数 9】

$$\log(E_y(f_m)) = \log(E_x(f_m)) + \log(m(f_m)) \quad (13)$$

ここで、 $f_m = f_{m_d}$ のときの変調スペクトル及び変調伝達関数MTFの関係を考える。残響時間 T_R を完全に推定することができたかすると、主要変調周波数 f_{m_d} における音源信号の変調スペクトルと回復された信号の変調スペクトルとは一致する。また、 $f_m = 0$ のときの変調伝達関数 $m(0)$ はあらゆる T_R に対して1であるから、 $f_m = 0$ における音源信号と残響信号の変調スペクトルも常に一致する。これらの条件から、次式が導かれる。

30

【数 10】

$$E_x(f_{m_d}) = E_x(0) = E_y(0) \quad (14)$$

これを式(13)に代入すると、次式が導かれる。

【数 11】

$$\log(E_y(f_{m_d})) = \log(E_y(0)) + \log(m(f_{m_d})) \quad (15)$$

ここで、変調伝達関数 $m(f_m)$ は f_m の関数であるが、残響時間パラメータ T_R も見掛け上変数とみることができる。ブラインド推定では、 T_R が不明であるため、式(4)の $m(f_m)$ を $m(f_m, T_R)$ と表せば、求めるべき残響時間 T_R は、次式で表すことができる。

40

【数 12】

$$\hat{T}_R = \arg \min_{T_R} \left(\left| \log(\hat{E}_y(0)) + \log(m(f_{m_d}, T_R)) - \log(\hat{E}_y(f_{m_d})) \right| \right) \quad (16)$$

この式により、残響時間のブラインド推定が可能となる。

【0030】

〔残響時間推定装置の動作〕

次に、残響時間推定装置の動作について説明する。作業者は、残響時間を測定する室内において、残響時間測定装置1のマイクロフォン2で音響をサンプリングする。マイクロフォン2から出力されたアナログ音響信号は、A/D変換器3によりデジタル音響信号へ

50

変換され、このデジタル音響信号（残響信号）がデジタル信号処理回路4へと与えられる。デジタル信号処理回路4に与えられたデジタル音響信号は、パワーエンベロープ生成部41によってパワーエンベロープへと変換され、このパワーエンベロープ信号が、ローパスフィルタ42と正規化変調スペクトル生成部44とにそれぞれ与えられる。

【0031】

ローパスフィルタ42によって高周波成分が除去されたパワーエンベロープ信号は、主要変調周波数取得部43に与えられる。主要変調周波数取得部43は、自己相関関数との時間シフト量を求めることにより、主要変調周波数を取得する。ここで取得される主要変調周波数は、変調周波数の全域のうち、特に高い変調スペクトルを示す変調周波数となる。例えば、主要変調周波数は、パワーエンベロープが正弦波の場合、その周波数と一致する。

10

【0032】

正規化変調スペクトル生成部44では、パワーエンベロープ信号に対してフーリエ変換が適用され、周波数領域の変調スペクトルが得られる。この変調スペクトルは、DC成分の変調スペクトルの値により正規化されたものとなる。

【0033】

デジタル信号処理回路4により出力された主要変調周波数及び変調スペクトルのデータは、演算回路5に与えられる。演算回路5は、メモリ6から複数の残響時間に対応するMTFの逆フィルタのデータを読み出し、デジタル信号処理回路4から受け付けた主要変調周波数及び変調スペクトルに対して各残響時間の逆フィルタを適用する。これにより、各残響時間に対応する音源信号の主要変調周波数における変調スペクトルが算出される。

20

【0034】

次に、演算回路5は、このようにして求めた各残響時間に対応する音源信号の主要変調周波数における変調スペクトルのうち、その大きさが0dBに最も近い1つを選択し、その残響時間を、室の残響時間として推定する。そして、演算回路5は、液晶表示部7を駆動制御し、推定した残響時間を表示させる。

【0035】

[評価実験]

評価用の音源信号として、式(6)の人工的なAM信号を利用した。また、パワーエンベロープを変調周波数5Hzの正弦波（変調度は1）とし、これに白色雑音キャリアを乗じた信号を用いた。このパワーエンベロープに対する変調スペクトルは、0Hzの変調スペクトルと5Hzの変調スペクトルが同一の値となった。次に、式(7)で定義される室内残響インパルス応答を利用した。ここでは、一つの残響時間 T_R につき、100種類の白色雑音キャリアを用意した。また、本評価実験で利用した残響時間 T_R は、0.1秒、0.3秒、0.5秒、1.0秒、及び2.0秒の5種類とした。そのため、本評価実験では、合計500個のパルス応答が用意され、これらを人工的な信号に畳み込むことで、残響信号を作成した。

30

【0036】

これらの500個の残響信号に対して、本実施の形態に係る残響時間推定装置1が実施する残響時間推定と同一の残響時間推定方法（以下、本法という）と、非特許文献1に記載されている残響音声パワーエンベロープの回復方法における残響時間推定方法（以下、従来法という）とにより、残響時間を推定した。図5は、評価実験の結果を示すグラフである。図中の実線は、本実施の形態に係る残響時間推定装置1による推定結果を示し、図中の破線は、従来法による推定結果を示している。また、図中の直線（点線）は、理想推定値を示している。本法及び従来法により推定された値は、残響時間 T_R 毎に100種類の残響信号から得られた推定値の平均値を示している。図中のエラーバーはそれぞれの残響時間における標準偏差を示している。

40

【0037】

理想直線と比較すると、従来法による推定結果は、残響時間が長くなるにつれ過小推定になる傾向があり、特に残響時間が1秒を超えたところから飽和する傾向がみられる。こ

50

のような現象が生じる原因が、残響信号から生成したパワーエンベロープを回復する際に、ローパスフィルタで除去しきれなかったパワーエンベロープ上の高周波成分が逆フィルタリング（微分処理）で強調され、それらの位相の状態によっては、残響時間推定の精度を左右する谷の形成（変調度を定義するもの）に影響を与えることにある。これに対し、本法の推定結果は、ほとんど理想直線と一致しており、正確に残響時間を推定していることが分かる。両手法では、残響時間の増加にともない、パワーエンベロープ上の高周波成分が逆フィルタにより異常に回復される。その際、従来法では時間領域で残響時間を推定するため高周波数の影響を直接受けてしまうが、本法では変調周波数領域で主要な周波数成分を利用して残響時間を推定するため高周波数の影響を受けない。これが、本法がうまく機能した理由である。

10

【0038】

（実施の形態2）

実施の形態2は、人の音声から残響時間を推定する場合に特に適した構成の残響時間推定装置である。

【0039】

〔人の音声信号に対する残響時間推定方法の適用〕

図6A～図6Dは、10Hzの正弦波1周期分を2セット用いて作ったパワーエンベロープと、その変調スペクトルを示すグラフである。図6Aに示したパワーエンベロープの2つの正弦波のセット間の距離は0.1秒であり、同様に図6Bでは0.2秒、図6Cでは0.5秒、図6Dでは1.0秒である。このように、パワーエンベロープの2つの正弦波のセット間の平坦な区間を長くしていくと、変調スペクトル上で0Hzの近傍の変調周波数 f_{m0} Hzに最初のピークが存在し、そのピークにおけるパワーの値 $E_x(f_{m0})$ （図6A～Dにおいて示した箇所）と主要変調周波数でのパワーの値 $E_x(f_{md})$ が近付いていく傾向があることがわかる。ここで、 f_{m0} は、0Hzを除いて最も0Hzに近い変調周波数である。このような f_{m0} におけるパワーは、図2を参照すると残響により減衰しないことがわかる。そこで、この f_{m0} を主要変調周波数 f_{md} のパワー回復の「基準値」として、残響時間の推定を実施すれば、かかる形状のパワーエンベロープに対しても残響時間を正確に推定することが可能となると考えられる。

20

【0040】

図7は、図6A～Dに示したような正弦波2セットの間の距離（時間間隔）と、基準周波数 f_{m0} におけるパワー値 $E(f_{m0})$ と主要変調周波数 f_{md} におけるパワー値 $E(f_{md})$ との差との関係を示したグラフである。図中縦軸は、 $E(f_{m0}) - E(f_{md})$ を示し、横軸は正弦波2セット間の距離を示している。図7では、パワーエンベロープに用いる1周期分の正弦波の周波数を5、10、20Hzとした。この図から、 $E(f_{m0}) = E(f_{md})$ となるケースが存在することが分かる。また、 f_{md} は、1周期分の正弦波の時間差に対応した変調周波数として現れるので、この図から $E(f_{m0}) = E(f_{md})$ となるような場合には、 f_{md} は比較適低い変調周波数であることが分かる。

30

【0041】

f_{m0} Hz上の基準値（パワー）は、残響により全く減衰を受けないわけではない。残響により基準値がどの程度影響を受けるかは重要な問題である。残響が付くことによって基準値が大きく減衰してしまえば、推定される残響時間の値が、理論値よりも短くなってしまふ。

40

【0042】

図8は、クリーンな状態での基準値における変調度を1としたMTF($M(f_m, T_R)$):実線)と、残響により減衰した基準値における変調度を1としたMTF($M'(f_m, T_R)$):破線)を示すグラフである。図に示すように、基準値が減衰すると、変調スペクトルの主要変調周波数 f_{md} におけるパワー $E_y(f_{md})$ において一致する $M(f_m, T_R)$ と $M'(f_m, T_R)$ とは異なる関数であるため、 $M(f_m, T_{Rc}) = M'(f_m, T_{Rw})$ となるときの T_{Rc} と T_{Rw} の値も異なってくる。

【0043】

50

f_{m0} の値は、FFT長に依存している。図9Aは、基準周波数 f_{m0} とMTFの値との関係を、複数の残響時間について示したグラフであり、図9Bは、基準周波数 f_{m0} のパワー $E_y(f_{m0})$ を基準値とした場合の残響時間の誤差と変調周波数との関係を示すグラフである。図9Aは、残響時間 T_R が、0.1、0.3、0.5、1.0、2.0秒のそれぞれの場合に対する、 $f_{m0} = 0.025、0.05、0.1$ Hz上でのMTFの値を示している。サンプリング周波数 f_s を20 kHzとすると、この図は、周波数分解能を高めるために、FFT長が30、20、10秒のときの基準値の減衰を示している。また、図9Bは、基準周波数 $f_{m0} = 0.025、0.05、0.1$ 秒の各ケースについて、残響時間の誤差と変調周波数との関係を示している。この図から、FFT長が短い程、残響時間が長い程、変調周波数が低い程、誤差が大きくなることが分かる。しかし、 $f_{m0} = 0.025、0.05$ Hz上に基準値をおいた場合には、低い変調周波数域で誤差が最も大きくなる $T_R = 2.0$ 秒の場合でも、残響時間の誤差は高々 10^{-3} 桁である。すなわち、FFT長が20秒程度であれば、0 Hzを除いて0 Hzに最も近い変調周波数成分におけるパワーを基準値としても、残響時間の推定値にはほとんど誤差の影響がないと考えられる。

10

【0044】

一方、人の音声の立上りと立ち下がりには、高い周波数のエネルギーが集中するため、帯域分割した音声のパワーエンベロープには、図6Dに示したような、ピーク間の平坦区間が長い形状がしばしばみられる。図10A及び図10Bは、帯域分割したときの人の音声のパワーエンベロープのグラフ及び変調スペクトルのグラフを示している。図10Aは残響が付加されていない音源信号の場合であり、図10Bは残響が付加された残響信号の場合である。音源信号の変調スペクトルは、0 Hzにごく近い変調周波数のパワーと主要な成分のパワーが等しくなっている。よって、音声信号に帯域分割を行うことにより、残響時間の推定が可能であると考えられる。

20

【0045】

[残響時間推定装置の構成]

上記のような原理に基づいて、本実施の形態では、音響信号(音声信号)を帯域分割し、帯域分割された各チャンネルのパワーエンベロープに基づいて残響時間の推定値を求め、これとともに各チャンネルのパワーエンベロープに基づいて処理に適したチャンネルを選定し、選定されたチャンネルによる残響時間の推定値の平均値を最終的な残響時間の推定値として採用する構成としている。また、本実施の形態に係る残響時間推定装置は、0 Hzを除いて0 Hzに最も近い変調周波数における変調スペクトル値を基準値とし、この基準値で正規化した変調スペクトルを生成するように構成されている。

30

【0046】

図11は、本発明の実施の形態2に係る残響時間推定装置の構成を示すブロック図である。図11に示すように、本実施の形態に係る残響時間推定装置201は、パワーエンベロープ生成部41、ローパスフィルタ42、主要変調周波数取得部43、及び正規化変調スペクトル生成部48の機能ブロックに加えて、帯域分割部45、及びチャンネル選定部46の機能ブロックを有するデジタル信号処理回路204を備えている。

40

【0047】

帯域分割部45は、100 Hzの帯域幅の定帯域通過フィルタバンクを備えており(図示せず)、A/D変換器3から出力されたデジタル音響信号を100チャンネルに帯域分割するように構成されている。この帯域分割部45によって帯域分割された各チャンネル信号は、パワーエンベロープ生成部41に与えられ、各チャンネル信号のパワーエンベロープ信号が生成されるようになっている。

【0048】

パワーエンベロープ生成部41から出力されるパワーエンベロープ信号は、ローパスフィルタ42及び正規化変調スペクトル生成部48に与えられると共に、チャンネル選定部46にも与えられる。本実施の形態の正規化変調スペクトル生成部48は、0 Hzを除いて0 Hzに最も近い変調周波数である基準周波数 f_{m0} におけるパワー値を基準値とし、

50

この基準値で正規化された変調スペクトル信号を出力するように構成されている。

【 0 0 4 9 】

また、チャンネル選定部 4 6 は、高レベル部検出部 4 6 a と、第 1 選定部 4 6 b と、第 2 選定部 4 6 c と、第 3 選定部 4 6 d とを有しており、帯域分割されたチャンネルの内、残響時間推定に用いるチャンネルを選定するように構成されている。

【 0 0 5 0 】

図 1 2 A ~ C は、チャンネル選定部 4 6 による残響時間推定に用いるチャンネルの選定過程を示す模式図である。図 1 2 A に示すように、まず高レベル部検出部 4 6 a により、所定の基準レベル（図中破線で示す）に基づいてパワーエンベロープ信号の高レベル部の立上り時点 $b(n)$ と立ち下がり時点 $e(n)$ とが検出される。次に、立上り時点 $b(n)$ から立ち下がり時点 $e(n)$ までの区間で、パワーが最大の時点を検出し、ピーク時点 $p(n)$ とされる。検出された立上り時点 $b(n)$ 、ピーク時点 $p(n)$ 、及び立ち下がり時点 $e(n)$ は 1 つのセットとされる。高レベル部検出部 4 6 a は、このようにしてパワーエンベロープ信号に含まれる全ての高レベル部のセットを検出するように構成されている。

10

【 0 0 5 1 】

第 1 選定部 4 6 b は、図 1 2 B に示すように、立上り時点 $b(n)$ から一つ前のセットの立ち下がり時点 $e(n-1)$ までの間に、微小なピークが存在するか否かを判定するように構成されている。微小なピークが存在する場合には、そのセットを残響時間推定に用いる対象から除外する。

20

【 0 0 5 2 】

第 2 選定部 4 6 c は、図 1 2 B に示すように、このような微小なピークが存在しない場合に、セットに大きな谷が含まれているか否かを判定するように構成されている。第 2 選定部 4 6 c は、図示しないローパスフィルタを備えており、このローパスフィルタによって谷が検出される。このような谷が存在するセットは、残響時間推定に用いる対象から除外される。

【 0 0 5 3 】

第 3 選定部 4 6 d は、上記のように、微小なピークが存在せず、且つ、谷が存在しないセットが 1 つのパワーエンベロープ中に連続して 2 つ存在する場合には、図 1 2 C に示すように、その 2 つのピークのパワーの差と時間差とを求め、これらが所定の基準値以上であるか否かを判定し、パワーの差及び時間差のいずれも基準値以上である場合には、当該チャンネルを残響時間推定に用いる対象として選定する。このようにしてチャンネル選定部 4 6 によって得られた選定結果は、演算回路 2 0 5 に与えられる。

30

【 0 0 5 4 】

演算回路 2 0 5 は、実施の形態 1 の場合と同様にして、主要変調周波数取得部 4 3 及び正規化変調スペクトル生成部 4 8 からの出力データを受け付け、各チャンネルにおける残響時間の推定値を演算する。また、演算回路 2 0 5 は、チャンネル選定部 4 6 からの出力データを受け付け、残響時間推定に用いる対象のチャンネルの残響時間推定結果の平均値を演算し、この平均値を残響時間の推定値とする。かかる推定された残響時間は、演算回路 2 0 5 が液晶表示部 7 を駆動制御することにより、液晶表示部 7 に表示される。

40

【 0 0 5 5 】

なお、本実施の形態に係る残響時間推定装置 2 0 1 のその他の構成は、実施の形態 1 に係る残響時間推定装置 1 の構成と同様であるので、同一の構成要素に付いては同符号を付し、その説明を省略する。

【 0 0 5 6 】

[評価実験]

女性話者が発話した 8 文章を評価用音響信号として利用し、本実施の形態に係る残響時間推定装置が実施する残響時間推定と同一の残響時間推定方法（以下、本法という）の評価実験を実施した。評価用音響信号の発話内容は以下の通りである。

(1) 「第一回通訳国際会議に参加のご登録をご希望される方は、所定の申込用紙に住所

50

- ・氏名と発表・聴講の別を明記して、国際会議事務局までお申し込みください。」
- (2) 「はい。こちらは第一回通話電話国際会議事務局です。」
- (3) 「もしも。通話国際会議への参加を申し込みたいのですけれども、どのような手続をすればよろしいのでしょうか。」
- (4) 「通話電話の国際会議に参加するためには、所定の申し込み用紙を用いて参加登録することが必要です。」
- (5) 「会議に発表するのではなくて聴講するだけだと、費用はいくらがかかりますか。」
- (6) 「ご発表を希望される場合には、予稿集代・登録料を含めた参加費用は4万円です。」
- (7) 「聴講のみの場合は当日の受け付けも可能で、予稿集代を含めた費用は3万5千円がかかります。」
- (8) 「参加登録の申し込み用紙はどのようにして手に入れればよろしいのでしょうか。」

10

利用した残響時間 T_R は、0.1、0.3、0.5、1.0、及び2.0秒の5種類とした。インパルス応答は、式(7)のものを残響時間毎に50種類用意した。なお、総刺激数は2000個(8個×5条件×50キャリア)である。本実験では、本法と従来法とのそれぞれについて、全ての発話に対して得られた残響時間の推定値の平均値を求め、この平均値について評価した。

【0057】

図13は、評価実験の結果を示すグラフである。図の記載方法は図5と同様である。本法では、残響時間 T_R が0.1、0.3、0.5、及び1.0秒のときに、概ね理想値に近い推定結果を得た。また、残響時間 T_R が2.0秒のときも、僅かに過大に推定されているものの、良好な結果を得ていることが分かる。

20

【0058】

(実施の形態3)

実施の形態3は、残響時間を推定するために用いられるコンピュータプログラムをコンピュータが実行することにより実現された残響時間推定装置である。なお、本実施の形態に係る残響時間推定装置301は、実施の形態2に係る残響時間推定装置201と実質的に同一の処理を、ソフトウェアにより実現したものである。

【0059】

[残響時間推定装置の構成]

図14は、本発明の実施の形態3に係る残響時間推定装置の構成を示すブロック図である。図14に示すように、コンピュータ301aは、本体311と、画像表示部312と、入力部313とを備えている。本体311は、CPU311aと、ROM311b、RAM311c、ハードディスク311d、読出装置311e、入出力インタフェース311f、及び画像出力インタフェース311gを備えており、CPU311a、ROM311b、RAM311c、ハードディスク311d、読出装置311e、入出力インタフェース311f、および画像出力インタフェース311gは、バス311iによって接続されている。

30

【0060】

CPU311aは、RAM311cにロードされたコンピュータプログラムを実行することが可能である。そして、後述するような残響時間推定プログラム314aを当該CPU311aが実行することにより、コンピュータ301aが残響時間推定装置301として機能する。

40

【0061】

ROM311bは、マスクROM、PROM、EPROM、又はEEPROM等によって構成されており、CPU311aに実行されるコンピュータプログラムおよびこれに用いるデータ等が記録されている。

【0062】

RAM311cは、SRAMまたはDRAM等によって構成されている。RAM311

50

cは、ハードディスク311dに記録されている残響時間推定プログラム314aの読み出しに用いられる。また、CPU311aがコンピュータプログラムを実行するときに、CPU311aの作業領域として利用される。

【0063】

ハードディスク311dは、オペレーティングシステムおよびアプリケーションプログラム等、CPU311aに実行させるための種々のコンピュータプログラムおよび当該コンピュータプログラムの実行に用いられるデータがインストールされている。後述する残響時間推定プログラム314aも、このハードディスク311dにインストールされている。

【0064】

ハードディスク311dには、残響時間をパラメータとして有するMTFの逆特性が、CPU311aによって処理可能な関数データとして記憶されている。このハードディスク311dに記憶されている逆フィルタ（逆特性）の関数データは、推定候補として予め与えられた複数の残響時間の分だけ用意される。なお、逆フィルタの関数データの詳細は、実施の形態1と同様である。

【0065】

読出装置311eは、フレキシブルディスクドライブ、CD-ROMドライブ、またはDVD-ROMドライブ等によって構成されており、可搬型記録媒体314に記録されたコンピュータプログラムまたはデータを読み出すことができる。また、可搬型記録媒体314には、コンピュータを残響時間推定装置として機能させるための残響時間推定プログラム314aが格納されており、コンピュータ301aが当該可搬型記録媒体314から残響時間推定プログラム314aを読み出し、当該残響時間推定プログラム314aをハードディスク311dにインストールすることが可能である。

【0066】

なお、前記残響時間推定プログラム314aは、可搬型記録媒体314によって提供されるのみならず、電気通信回線（有線、無線を問わない）によってコンピュータ301aと通信可能に接続された外部の機器から前記電気通信回線を通じて提供することも可能である。例えば、前記残響時間推定プログラム314aがインターネット上のサーバコンピュータのハードディスク内に格納されており、このサーバコンピュータにコンピュータ301aがアクセスして、当該コンピュータプログラムをダウンロードし、これをハードディスク311dにインストールすることも可能である。

【0067】

また、ハードディスク311dには、例えば米マイクロソフト社が製造販売するWindows（登録商標）等のマルチタスクオペレーティングシステムがインストールされている。

以下の説明においては、本実施の形態1に係る残響時間推定プログラム314aは当該オペレーティングシステム上で動作するものとしている。

【0068】

入出力インタフェース311fは、例えばUSB、IEEE1394、又はRS-232C等のシリアルインタフェース、SCSI、IDE、又はIEEE1284等のパラレルインタフェース、およびD/A変換器、A/D変換器等からなるアナログインタフェース等から構成されている。入出力インタフェース311fには、キーボードおよびマウスを有する入力部313が接続されており、ユーザが当該入力部313を使用することにより、コンピュータ301aにデータを入力することが可能である。また、入力部313には、マイクロフォンが設けられており、音響を電気信号に変換し、これを入出力インタフェース311fによってA/D変換し、デジタル信号の音響信号を取得するように構成されている。

【0069】

画像出力インタフェース311gは、LCDまたはCRT等で構成された画像表示部312に接続されており、CPU311aから与えられた画像データに応じた映像信号を画

10

20

30

40

50

像表示部 3 1 2 に出力するようになっている。画像表示部 3 1 2 は、入力された映像信号にしたがって、画像（画面）を表示する。

【 0 0 7 0 】

[残響時間推定装置の動作]

次に、残響時間推定装置 3 0 1 の動作について説明する。図 1 5 は、本実施の形態に係る残響時間推定装置 3 0 1 の動作の流れを示すフローチャートである。まず、作業者は、残響時間を測定する室内において、入力部 3 1 3 のマイクロフォンで音響（音声）をサンプリングする。マイクロフォンから出力されたアナログ音響信号は、入出力インタフェース 3 1 1 f が備える A / D 変換器により P C M 等のデジタル音響データへ変換され、この音響データが C P U 3 1 1 a に与えられる。

10

【 0 0 7 1 】

まず、C P U 3 1 1 a は、音響データを複数の帯域に分割する（ステップ S 1）。このとき、音響データは 1 0 0 H z の帯域幅毎にチャンネル分割される。次に、C P U 3 1 1 a は、各チャンネルの音響データからパワーエンベロープをそれぞれ生成し、各パワーエンベロープを R A M 3 1 1 c に記憶する（ステップ S 2）。C P U 3 1 1 a は、チャンネルを 1 つ選択し（ステップ S 3）、R A M 3 1 1 c からこのチャンネルのパワーエンベロープを読み出し、所定のカットオフ周波数により高周波成分を除去する（ステップ S 4）。そして、C P U 3 1 1 a は、この高周波成分を除去したパワーエンベロープに対して、自己相関関数を演算し、主要変調周波数を算出する（ステップ S 5）。

【 0 0 7 2 】

次に C P U 3 1 1 a は、再度前記チャンネルのパワーエンベロープを R A M 3 1 1 c から読み出し、フーリエ変換処理を実行し（ステップ S 6）、0 H z を除いて 0 H z に最も近い変調周波数におけるパワーを基準値とし、ステップ S 6 で得られた変調スペクトルを前記基準値で正規化した正規化変調スペクトルを演算する（ステップ S 7）。次に、C P U 3 1 1 a は、ステップ S 5 で算出した主要変調周波数及びステップ S 7 で算出した正規化変調スペクトルに基づいて、残響時間を推定する（ステップ S 8）。この残響時間の推定は、次のようにして行われる。まず、C P U 3 1 1 a がハードディスク 3 1 1 d から各残響時間の逆フィルタを読み出し、ステップ S 5 で算出した主要変調周波数及びステップ S 7 で算出した変調スペクトルに対して各残響時間の逆フィルタを適用して、各残響時間に対応する音源信号の主要変調周波数における変調スペクトルを算出する。次に、C P U 3 1 1 a が、このようにして求めた各残響時間に対応する音源信号の主要変調周波数における変調スペクトルのうち、大きさが 0 d B（変調度で表すと 1）に最も近い 1 つを選択し、その残響時間を、室の残響時間として推定する。C P U 3 1 1 a は、このようにして求めた残響時間の推定値を、そのとき処理を行ったチャンネルと対応づけて R A M 3 1 1 c に記憶する。

20

30

【 0 0 7 3 】

次に C P U 3 1 1 a は、当該チャンネルのパワーエンベロープを再度 R A M 3 1 1 c から読み出し、このパワーエンベロープに基づいて、当該チャンネルが残響時間の推定に用いるのに適したものであるか否かを判定する（ステップ S 9）。この処理は、（ 1 ）パワーエンベロープの高レベル部（立上り時点、ピーク時点、立ち下がり時点のセット）の検出、（ 2 ）隣り合う高レベル部の間に微小ピークが存在するか否かの判定、（ 3 ）高レベル部に谷が存在するか否かの判定、微小ピーク及び谷の両方が存在しない場合に、（ 4 ）隣り合う高レベル部のパワーの差及び時間差が所定の基準値以上であるか否かの判定、の各処理により構成される。なお、これらの処理の詳細は、実施の形態 2 で説明したものと同様であるので、その説明を省略する。

40

【 0 0 7 4 】

次に C P U 3 1 1 a は、全てのチャンネルに対して上記のステップ S 4 ~ S 9 の処理を行ったか否かを判定し（ステップ S 1 0）、ステップ S 1 0 において上記ステップ S 4 ~ S 9 の処理が終了していないチャンネルが存在する場合（ステップ S 1 0 において N O）、未処理のチャンネルの 1 つを選択し（ステップ S 1 1）、ステップ S 4 へと処理を戻す

50

。ステップ S 9 において全チャンネルについて上記ステップ S 3 ~ S 8 の処理が終了している場合には (ステップ S 9 において Y E S)、C P U 3 1 1 a は、ステップ S 9 において、残響時間の推定に用いるのに適していると判定されたチャンネルに対応する残響時間の推定値の平均値を算出し (ステップ S 1 2)、画像出力インタフェース 3 1 1 g を駆動して、画像表示部 3 1 2 に、この平均値を本音響信号に対する残響時間の推定値として表示させ (ステップ S 1 3)、処理を終了する。

【 0 0 7 5 】

(その他の実施の形態)

上記の実施の形態 1 においては、変調スペクトルの大きさを、D C 成分の変調スペクトル (変調周波数が 0 H z における変調スペクトル) で正規化した変調スペクトル信号を生成するように正規化変調スペクトル生成部 4 4 が構成されている場合について説明したが、これに限定されるものではなく、例えば、0 H z を除いて 0 H z に最も近い変調周波数におけるパワー値で、変調スペクトルの大きさを正規化する構成であってもよい。

【 0 0 7 6 】

また、上記実施の形態 2 及び 3 においては、0 H z を除いて 0 H z に最も近い変調周波数における変調スペクトル値を基準値とし、この基準値で正規化した変調スペクトルを生成するよう正規化変調スペクトル生成部 4 8 が構成されている場合について説明したが、これに限定されるものではなく、変調スペクトルの大きさを、D C 成分の変調スペクトルで正規化した変調スペクトル信号を生成するように構成されていてもよい。

【 0 0 7 7 】

上記の実施の形態 3 においては、音響データを複数のチャンネルに帯域分割し、それぞれのチャンネルについて残響時間の推定値を演算し、残響時間の推定に用いるのに適したチャンネルを選定して、選定されたチャンネルの残響時間の推定値を平均することで、本音響データに対する残響時間の推定値を求める構成について述べたが、これに限定されるものではなく、帯域分割及びチャンネルの選定を行わず、音響データに対して 1 つのパワーエンベロープを演算し、それに基づいて残響時間の推定値を演算する構成としてもよい。

【 0 0 7 8 】

また、上記の実施の形態 2 及び 3 においては、全てのチャンネルのパワーエンベロープに対して主要変調周波数の取得及び正規化変調スペクトルの生成を行うことによりそれぞれのチャンネルに対応する残響時間推定値を個々に求め、残響時間の推定に用いるのに適したチャンネルを全チャンネルから選定して、残響時間の推定に用いるのに適したチャンネルのパワーエンベロープから求めた残響時間の推定値の平均を求める構成としたが、これに限定されるものではなく、全てのチャンネルのパワーエンベロープから、残響時間の推定に用いるのに適したチャンネルを選定し、選定されたチャンネルのパワーエンベロープから残響時間の推定値を求め、その平均値を算出する構成としてもよい。この場合は、選定されたチャンネルのパワーエンベロープだけに対して、主要変調周波数の取得及び正規化変調スペクトルの生成が行われるため、処理を効率的に行うことができる。

【 0 0 7 9 】

また、上記の実施の形態 1 ~ 3 においては、残響時間をパラメータとして有する M T F の逆フィルタが、演算回路 5 又は C P U 3 1 1 a によって処理可能な関数データとして、所定の複数の残響時間の分だけメモリ 6 又は R A M 3 1 1 c に記憶されており、取得された主要変調周波数及び変調スペクトルに対して各残響時間の逆フィルタを適用することにより、各残響時間に対応する音源信号の主要変調周波数における変調スペクトルを算出し、このようにして求めた各残響時間に対応する音源信号の主要変調周波数における変調スペクトルのうち、その大きさが 0 d B に最も近い 1 つを選択し、その残響時間を、室の残響時間として推定する構成とした。しかしながら、上記の構成に限定されるものではない。上記の逆フィルタが関数データとしてではなく、ルックアップテーブルとしてメモリ 6 又は R A M 3 1 1 c に記憶されており、演算回路 5 又は C P U 3 1 1 a が逆フィルタのルックアップテーブルを参照して、各残響時間に対応する音源信号の主要変調周波数にお

10

20

30

40

50

る変調スペクトルを取得する構成としてもよい。上記の各実施の形態では残響時間推定装置が1つの装置となっているが、本発明はこのような態様に限定されるわけではない。例えば、当該残響時間推定装置が備える各手段が別々の装置に設けられ、それらの装置が通信ネットワーク等を介してデータ通信等を行うことによって、上記の各実施の形態における処理が実現されるような構成であってもよい。

【産業上の利用可能性】

【0080】

本発明の残響時間推定装置及び残響時間推定方法は、時系列の音響信号から求められた周波数系列の変調スペクトルにより、原音信号を用いずに残響時間をブラインド推定する残響時間推定装置及びその方法などとして有用である。

10

【図面の簡単な説明】

【0081】

【図1】本発明の実施の形態1に係る残響時間推定装置の構成を示すブロック図。

【図2】MTF（変調度）を示すグラフ。

【図3A】残響が付加されていない音響信号のパワーエンベロープを示すグラフ。

【図3B】残響が付加されていない音響信号の変調スペクトルを示すグラフ。

【図4A】図3Aで示す音響に残響が付加した場合の音響信号のパワーエンベロープを示すグラフ。

【図4B】図4Aに示す音響信号の変調スペクトルを示すグラフ。

【図5】実施の形態1に係る残響時間推定装置が実施する残響時間推定と同一の残響時間推定方法の評価実験の結果を示すグラフ。

20

【図6A】10Hzの正弦波1周期分を2セット用いて作ったパワーエンベロープと、その変調スペクトルを示すグラフ（セット間の時間差0.1秒）。

【図6B】10Hzの正弦波1周期分を2セット用いて作ったパワーエンベロープと、その変調スペクトルを示すグラフ（セット間の時間差0.2秒）。

【図6C】10Hzの正弦波1周期分を2セット用いて作ったパワーエンベロープと、その変調スペクトルを示すグラフ（セット間の時間差0.5秒）。

【図6D】10Hzの正弦波1周期分を2セット用いて作ったパワーエンベロープと、その変調スペクトルを示すグラフ（セット間の時間差1.0秒）。

【図7】図6A～Dに示した正弦波2セットの間の時間間隔と、基準周波数と主要変調周波数とにおけるパワー値の差との関係を示したグラフ。

30

【図8】クリーンな状態での基準値における変調度を1としたMTFと、残響により減衰した基準値における変調度を1としたMTFとを示すグラフ。

【図9A】基準周波数とMTFの値との関係を、複数の残響時間について示したグラフ。

【図9B】基準周波数のパワーを基準値とした場合における残響時間の誤差と変調周波数との関係を示すグラフ。

【図10A】帯域分割したときの人の音声（残響付加なし）のパワーエンベロープのグラフ及び変調スペクトルのグラフ。

【図10B】帯域分割したときの人の音声（残響付加あり）のパワーエンベロープのグラフ及び変調スペクトルのグラフ。

40

【図11】本発明の実施の形態2に係る残響時間推定装置の構成を示すブロック図。

【図12A】チャンネル選定部による残響時間推定に用いるチャンネルの選定処理を説明するための模式図。

【図12B】チャンネル選定部による残響時間推定に用いるチャンネルの選定処理を説明するための模式図。

【図12C】チャンネル選定部による残響時間推定に用いるチャンネルの選定処理を説明するための模式図。

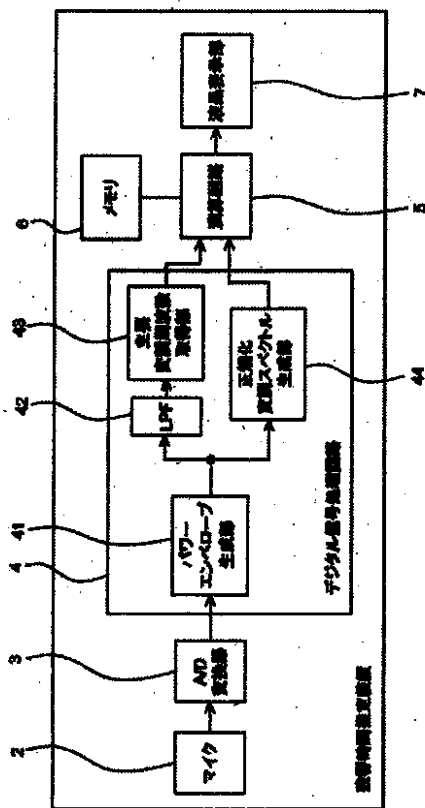
【図13】実施の形態2に係る残響時間推定装置が実施する残響時間推定と同一の残響時間推定方法の評価実験の結果を示すグラフ。

【図14】本発明の実施の形態3に係る残響時間推定装置の構成を示すブロック図。

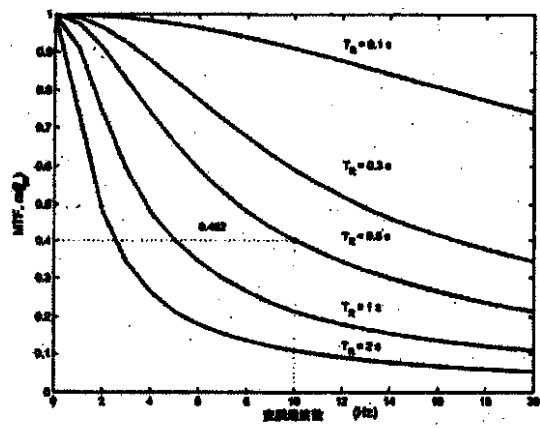
50

【図15】実施の形態3に係る残響時間推定装置の動作の流れを示すフローチャート。

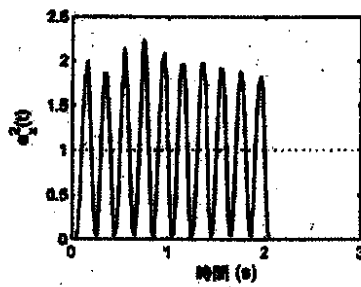
【図1】



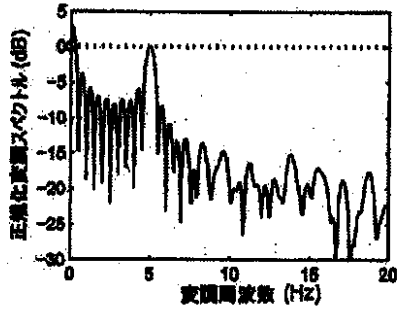
【図2】



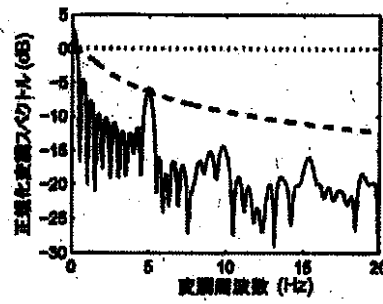
【図3A】



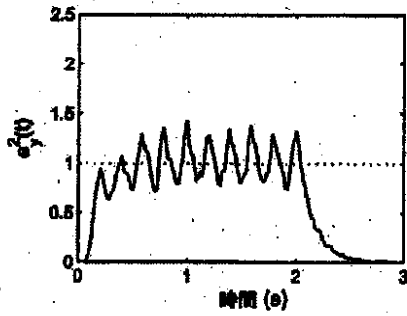
【 図 3 B 】



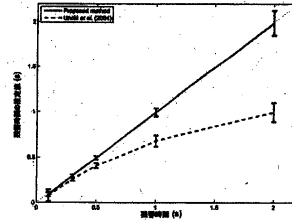
【 図 4 B 】



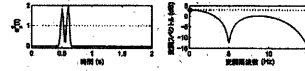
【 図 4 A 】



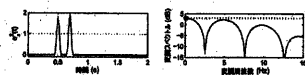
【 図 5 】



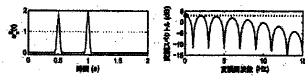
【 図 6 A 】



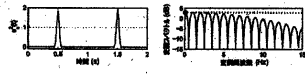
【 図 6 B 】



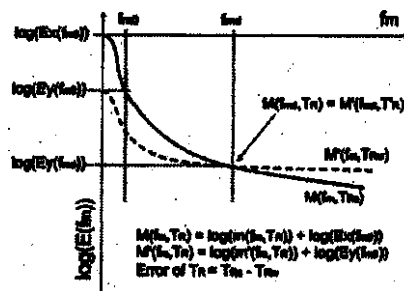
【 図 6 C 】



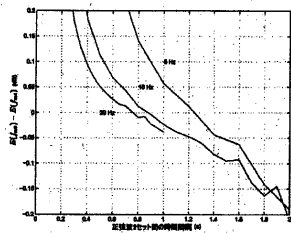
【 図 6 D 】



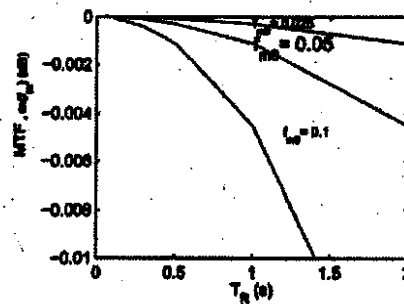
【 図 8 】



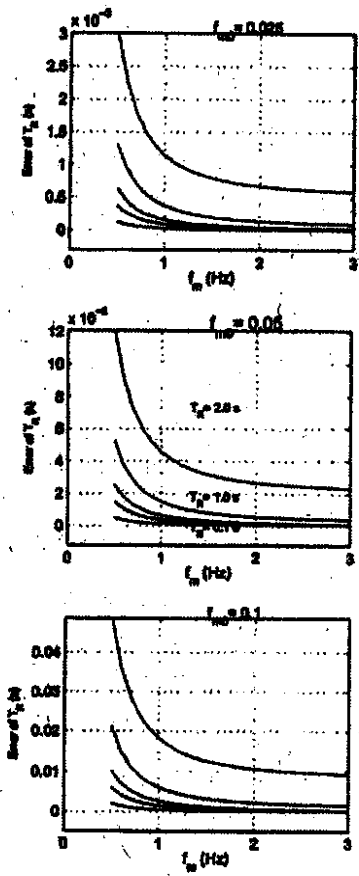
【 図 7 】



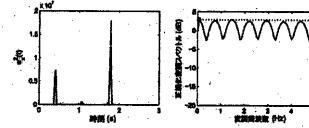
【 図 9 A 】



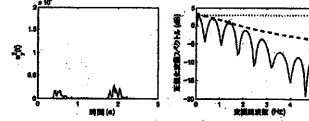
【 9 B 】



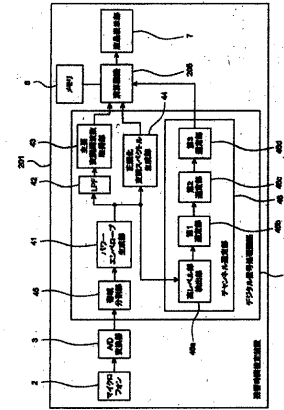
【 10 A 】



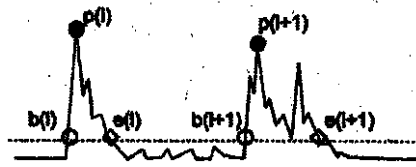
【 10 B 】



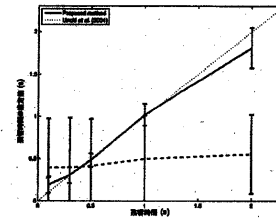
【 11 】



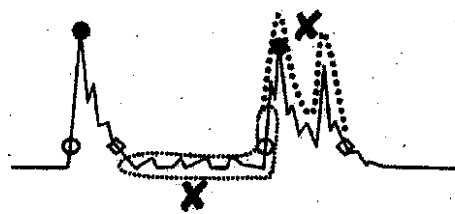
【 12 A 】



【 13 】



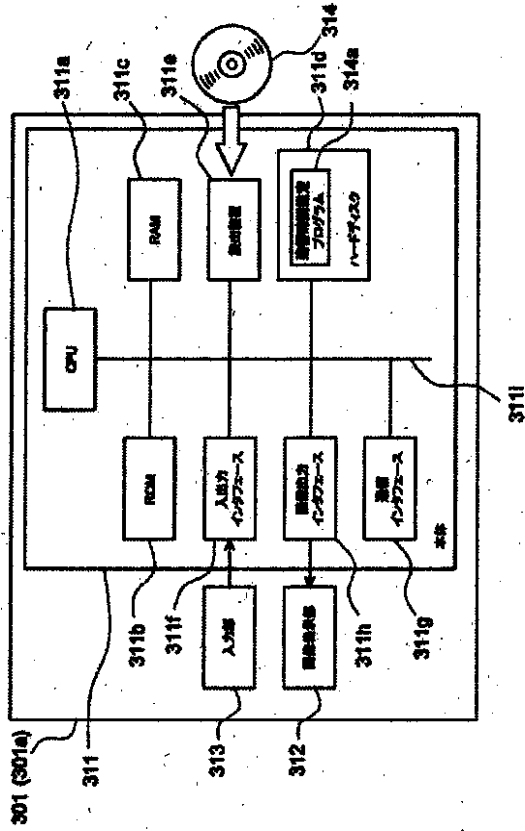
【 12 B 】



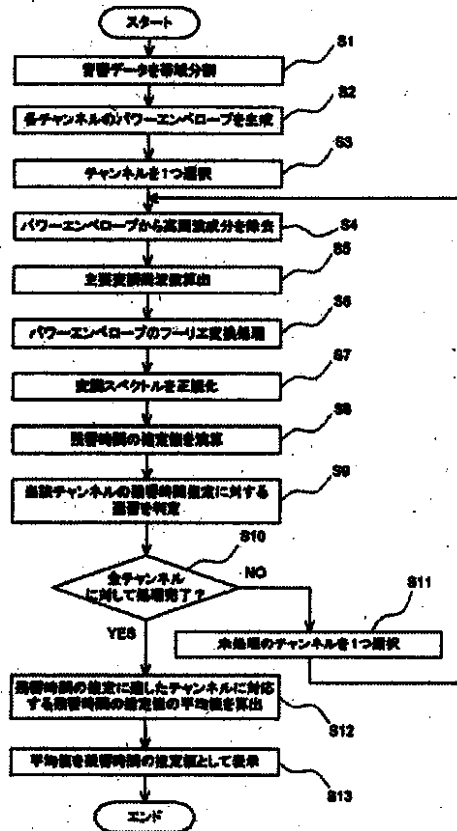
【 12 C 】



【図14】



【図15】



フロントページの続き

(56)参考文献 特開2001-100774(JP,A)

特開平03-221819(JP,A)

酒田恵吾、鵜木祐史、赤木正人、MTFに基づいた残響音声の回復法の検討、電子情報通信学会
技術研究報告、日本、社団法人 電子情報通信学会、2003年 3月20日、SP、音声 102(74
9)、p.13-18

(58)調査した分野(Int.Cl., DB名)

G10K 15/00

G10L 11/00 - 11/06

G01H 7/00