(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2011-65128
(P2011-65128A)
(43)公開日　平成23年3月31日(2011.3.31)

(51)Int.Cl.	F I	テーマコード（参考）
G10L 21/02 (2006.01)	G10L 21/02 101B	

審査請求 未請求　請求項の数 9　OL　（全 17 頁）

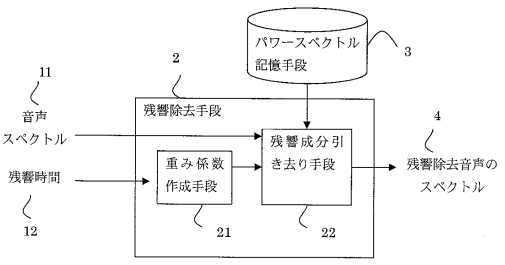
(21)出願番号　特願2010-30398(P2010-30398)
(22)出願日　平成22年2月15日(2010.2.15)
(31)優先権主張番号　特願2009-191303(P2009-191303)
(32)優先日　平成21年8月20日(2009.8.20)
(33)優先権主張国　日本国(JP)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目７番３号
(74)代理人　100110423
弁理士　曾我　道治
(74)代理人　100084010
弁理士　古川　秀利
(74)代理人　100094695
弁理士　鈴木　憲七
(74)代理人　100111648
弁理士　梶並　順
(74)代理人　100122437
弁理士　大宅　一宏
(74)代理人　100147566
弁理士　上田　俊一

最終頁に続く

(54)【発明の名称】残響除去装置

(57)【要約】
【課題】従来の残響除去は発話末尾の残響減衰特性を重視し、正しい発話終了検出を前提としているが発話終了検出は誤りやすく、使用する室の残響減衰特性のモデルを仮定していないため、本来一定であるはずの室の残響減衰特性が発話によって異なって推定されやすい問題があった。
【解決手段】観測時点以前に音源から出力された音のエネルギーが観測された音のエネルギーに占める割合を推定する重み係数を与えられた残響時間から推定する重み係数作成手段と、観測時点以前に出力された音のエネルギーを記憶するパワースペクトル記憶手段と、前記重み係数を用いて観測された音のエネルギーからパワースペクトル記憶手段に記憶された、観測時点以前に出力されたエネルギーを引き去る残響成分引き去り手段を備える。
【選択図】図１



【特許請求の範囲】
【請求項１】
観測時点以前に音源から出力された音のエネルギーが、観測された音のエネルギーに占める割合を推定する重み係数を音の残響時間から推定する重み係数作成手段と、観測時点以前に出力された音のエネルギーを記憶するパワースペクトル記憶手段と、前記重み係数を用いて観測された音のエネルギーからパワースペクトル記憶手段に記憶された、観測時点以前に出力されたエネルギーを引き去る残響成分引き去り手段を備えたことを特徴とする残響除去装置。
【請求項２】
前記重み係数作成手段は、音の発生する室の平均音響エネルギー密度の減衰特性を時間の指数関数でモデリングし、このモデリングから残響時間をパラメータとして重み係数を推定することを特徴とする請求項１記載の残響除去装置。
【請求項３】
設定された任意の残響時間から重み係数を推定し、この推定重み係数を用いて残響除去し、残響除去後のエネルギー成分がフロアリングされた第１の所定閾値を下回る割合を観測し、その割合が第２の所定閾値と比べて大きいか小さいかで設定残響時間を増減させて残響時間を推定する残響時間予測手段を備え、前記重み係数作成手段が用いる残響時間は、この残響時間予測手段が推定する残響時間とすることを特徴とする請求項１記載の残響除去装置。
【請求項４】
残響時間を事前に学習データから推定し、この推定残響時間より作成した重み係数を記憶する重み係数記憶手段を備え、残響成分引き去り手段は、この重み係数記憶手段より重み係数を入力し、残響成分引き去り処理のリアルタイム化を可能とすることを特徴とする請求項３に記載の残響除去装置。
【請求項５】
前記残響時間予測手段は、残響時間を周波数帯域ごとに計算する構成にされたことを特徴とする請求項３に記載の残響除去装置。
【請求項６】
前記残響時間の長短に応じて、残響除去を行うか否かを事後的に決定する出力スペクトル決定手段を備えたことを特徴とする請求項３に記載の残響除去装置。
【請求項７】
あらかじめ学習された観測時点以前の入力音声エネルギーにおける騒音の平均エネルギーを記憶する騒音パワースペクトル記憶手段と、
観測された音のエネルギーから、騒音パワースペクトル記憶手段に記憶された観測時点以前の入力音声エネルギーにおける騒音の平均エネルギーを引き去り、その出力を残響成分引き去り手段に入力する騒音成分引き去り手段を備えたことを特徴とする請求項１に記載の残響除去装置。
【請求項８】
あらかじめ学習された観測時点以前の入力音声エネルギーにおける騒音の平均エネルギーを記憶する騒音パワースペクトル記憶手段と、
パワースペクトル記憶手段が記憶する観測時点以前に出力された音のエネルギーから、騒音パワースペクトル記憶手段に記憶された観測時点以前の入力音声エネルギーにおける騒音の平均エネルギーを引き去り、その出力を残響成分引き去り手段に入力する騒音成分引き去り手段を備えたことを特徴とする請求項１に記載の残響除去装置。
【請求項９】
上記重み係数作成手段で、重み係数推定に用いられる残響時間を予測するため、音源からの音のエネルギーを観測された音のエネルギーから算出するために設定されるフロアリング係数を用いて残響時間を予測する残響時間予測手段と、
入力音声が音声か非音声か、直接音か残響音かを判定する音声/非音声、直接音/残響音判定手段と、

音声/非音声、直接音/残響音判定手段の判定結果を用いて上記残響時間予測手段で用いられる設定フロアリング係数を作成するフロアリング係数作成手段を備えることを特徴とする請求項１または８に記載の残響除去装置。

【発明の詳細な説明】

【技術分野】

【０００１】

この発明は、残響の重畳した音声データから演算量の少ない処理で残響を除去する残響除去装置に関するものである。

【背景技術】

【０００２】

一般に音声データに残響が重畳すると、音素重なりが起こったり、ポーズの部分に残響成分が重畳したりすることで、音声の認識性能が大きく低下する。重畳した残響成分は、残響（インパルス応答）の逆フィルターを構成して畳み込むことで効果的に除去することができる。しかし逆フィルターを作成するための計算量がタップ数の２乗に比例するため膨大になり、適用される装置に組み込まれるなど演算量に制約を受ける環境では現実的でない。

簡易的な残響除去フィルターの係数決定法に関する従来技術としては、周波数領域でパワースペクトルの減算を行うスペクトラルサブトラクション法（以下SS法）によるものがいくつか提案されている。たとえば特許文献1では発話末尾の残響減衰特性を用いることによって、簡易的な残響除去のフィルター係数を決定していた。

【先行技術文献】

【特許文献】

【０００３】

【特許文献１】特開2008－58900号公報

【発明の概要】

【発明が解決しようとする課題】

【０００４】

特許文献1にあげられている従来の残響除去法は発話末尾の残響減衰特性を重視しており、発話終了が正しく検出できることを前提としている。しかし、発話終了を正しく検出することはそれほど簡単でなく誤りやすいという問題点があった。

また使用する室における残響減衰特性の物理的なモデルを仮定していないため、本来一定であるはずの室の残響減衰特性が発話によって異なって推定されやすいという問題点もあった。

この発明は上記のような問題点を解決するためになされたもので、観測された音のエネルギーに占める観測時点以前の音源から出力された音のエネルギーの割合を推定する重み係数を残響時間から推定し、この重み係数を用いて残響除去フィルターを作成する。

また、室の平均音響エネルギー密度が時間に関し指数的に減衰することを利用して、発話に影響されにくい残響除去フィルターを作成し残響除去性能を向上させることを目的とする。

さらに残響に加え騒音が重畳した場合にも、残響成分のみを引き去ることで音声の認識性能を向上させることを目的とする。

【課題を解決するための手段】

【０００５】

この発明に係る残響除去装置は、観測時点以前に音源から出力された音のエネルギーが、観測された音のエネルギーに占める割合を推定する重み係数を音の残響時間から推定する重み係数作成手段と、観測時点以前に出力された音のエネルギーを記憶するパワースペクトル記憶手段と、前記重み係数を用いて観測された音のエネルギーからパワースペクトル記憶手段に記憶された、観測時点以前に出力されたエネルギーを引き去る残響成分引き去り手段を備える。

【０００６】

また、この発明に係る残響除去装置は、前記重み係数作成手段が、音の発生する室の平均音響エネルギー密度の減衰特性を時間の指数関数でモデリングし、このモデリングから残響時間をパラメータとして重み係数を推定する構成にされたものである。

【発明の効果】

【０００７】

この発明に係る残響除去装置によれば、重み係数作成手段により観測された音のエネルギーに占める観測時点以前の音源から出力された音のエネルギーの割合を推定する重み係数を残響時間から推定し、この重み係数を用いて残響成分引き去り手段（残響除去フィルターに相当）がパワースペクトル記憶手段に記憶された観測時点以前に出力された音のエネルギーを観測された音のエネルギーから引き去ることで、残響除去処理をしているので、現場で推定に用いるパラメータが残響時間のみであるため非合理な重み係数が作成されにくく、環境の変化に頑健である。

【０００８】

また、この発明に係る残響除去装置によれば、重み係数作成手段は、音の発生する室の平均音響エネルギー密度の減衰特性を時間の指数関数でモデリングし、このモデリングから残響時間をパラメータとして重み係数を推定するので、発話に影響されにくい残響成分引き去り手段（残響除去フィルター）を作成し残響除去性能を向上させることができる。

【図面の簡単な説明】

【０００９】

【図１】この発明の実施の形態１による残響除去装置の構成図である。

【図２】この発明の実施の形態２による残響除去装置の構成図である。

【図３】この発明の実施の形態３による残響除去装置の構成図である。

【図４】この発明の実施の形態４による残響除去装置の構成図である。

【図５】この発明の実施の形態５による残響除去装置の構成図である。

【図６】この発明の実施の形態６による残響除去装置の構成図である。

【図７】室の残響減衰特性の説明図である。

【図８】室の残響減衰特性とフィルター係数の決定方法の関係図である。

【図９】残響時間が予測より長い場合の残響成分除去処理の状態図である。

【図１０】残響時間が予測より短い場合の残響成分除去処理の状態図である。

【図１１】実施の形態２による残響除去装置の動作フロー図である。

【図１２】室の残響時間に対する推定残響時間とフロアリングする割合との関係特性図である。

【図１３】室の残響時間に対する推定残響時間とフロアリングする割合とその閾値の関係を示す特性図である。

【図１４】この発明の実施の形態７による残響除去装置の構成図である。

【図１５】この発明の実施の形態８による残響除去装置の構成図である。

【図１６】この発明の実施の形態９による残響除去装置の構成図である。

【図１７】この発明の実施の形態１０による残響除去装置の構成図である。

【図１８】騒音存在時の室の残響減衰特性の説明図である。

【発明を実施するための形態】

【００１０】

実施の形態１．

図１はこの発明の実施の形態１による残響除去装置の構成図である。この実施の形態の場合、残響時間が既知であると仮定している。残響時間とは室の平均音響エネルギーが60dB減衰するのにかかる時間である。

図１において残響除去手段2は、入力された音のスペクトル11から既知であると仮定した所与の残響時間12を用いて残響成分を除去した音のスペクトル4を得るためのものである。

残響除去手段2は、重み係数作成手段21と残響成分引き去り手段22を備え、重み係数作成手段21は、所与の残響時間12から重み係数を作成する。残響成分引き去り手段22は、入

力された音のスペクトル11のパワースペクトルから、過去に入力されパワースペクトル記憶手段3に記憶されていたパワースペクトルに重み係数作成手段21によって作成された重み係数を掛け合わせたパワースペクトルを残響成分としてSS法により引き去り、残響除去音声のスペクトル4を得る。この残響成分引き去り手段22が、残響除去フィルターを形成する。

【００１１】

この発明の原理について説明する。反射音と直接音はインコヒーレントであるためエネルギーの加算が許される。ゆえに残響のある環境では、受音点で観測された音のパワースペクトル（ $|X_k[i]|^2$ ）はsフレーム前の音源のパワースペクトル（ $|Y_k[i-s]|^2$ ）と、現在の音源のパワースペクトル（ $|Y_k[i]|^2$ ）とによる式（1）のように重み付き和の関係にある。ここでのスペクトルは、あるフレーム幅をフレームシフト fr をもって短時間フーリエ変換を行った結果得られたものとする。

【００１２】

【数１】

$$|X_k[i]|^2 = \sum_{s=0}^{i} w[s] \cdot |Y_k[i-s]|^2 \quad \cdots \quad (1)$$

【００１３】

ここで i は現在のフレーム番号、k はフーリエ変換の次元（ 0 ≦ k ≦ N - 1 ）、w[s]は重み係数（ 0 ≦ s ≦ i ）である。また音源のスペクトルは未知であるので、過去の音源のパワースペクトル $|Y_k[i-s]|^2$ をパワースペクトル記憶手段3に記憶された過去の観測パワースペクトル（ $|X_k[i-s]|^2$ ）で式（2）のように近似する。q は残響成分が重畳することによって増加するエネルギーの割合を示しており、一般に残響時間の関数であると考えられる。

【００１４】

【数２】

$$q|Y_k[i-s]|^2 \approx |X_k[i-s]|^2 \quad \cdots \quad (2)$$

【００１５】

さらにw[0] = 1を仮定して式（1）、（2）より式（3）が導かれる。

【００１６】

【数３】

$$|Y_k[i]|^2 = |X_k[i]|^2 - \frac{1}{q}\sum_{s=1}^{i} w[s] \cdot |X_k[i-s]|^2 \quad \cdots \quad (3)$$

【００１７】

すると式（1）と式（3）のw[s]（ 1 ≦ s ≦ i ）が同一視でき、重み係数 w[s] が残響成分引き去り手段22でSS法に用いる係数そのものとなる。したがって、前記重み係数w[s]が適切に推定できれば、式（3）によってスペクトル記憶手段3に記憶された過去の観測されたパワースペクトル（ $|X_k[i-s]|^2$ ）に重み係数w[s]を乗算した積を現在の観測されたパワースペクトル（ $|X_k[i]|^2$ ）から引き去ることで残響除去ができる。

この発明は重み係数作成手段21によって前記重み係数w[s]を決定し、式（3）に従い残響成分引き去り手段22で残響を除去するものである。

【００１８】

次に具体的な動作について説明する。残響除去を行う前に、重み係数作成手段21によって重み係数w[s]を作成する。残響は減衰の程度に応じて、図７のように反射音が疎な初期残響と反射音の密な後期残響に分けることができ、適用先が音声認識である場合には認識性能に悪影響を与えるのは主に後期残響である。初期残響はさまざまな要因が影響して複雑であるが、後期残響は拡散音場理論に基づき、その室の平均エネルギー密度の時間変化

を式（4）の指数関数の形に仮定できる（たとえば、Philip M. Morse and K. Uno Ingard, Theoretical Acoustics, Princeton University Press, 1968, pp. 576-579.を参照）。

【００１９】

【数４】

$$\bar{E}(t) = \bar{E}(0)e^{-\frac{13.82}{RT}t} \quad \cdots \quad (4)$$

ここで、$\bar{E}(t)$ は室の平均エネルギー密度、t [秒]は時間、RT は残響時間[秒]である。

【００２０】

インパルス応答2乗積分法の導出法に倣えば、室の伝達関数 h と観測点のエネルギー密度 E(t)は式（5）の関係にある。

【００２１】

【数５】

$$E(t) = \frac{\int_t^\infty h^2(s)ds}{\int_0^\infty h^2(s)ds} \quad \cdots\cdots \quad (5)$$

ゆえに部屋の伝達関数 h の 2 乗減衰特性は観測点のエネルギー密度 $E(t)$ の時間微分であり、観測点のエネルギー密度 $E(t)$ を部屋の平均エネルギー密度 $\bar{E}(t)$ と仮定すれば、これも指数関数の形に仮定することができる。

【００２２】

これらを利用すると残響時間 RT をあらかじめ決めておけば、式（6）にしたがって重み係数を定めることができる。すなわち重み係数作成手段21は、残響時間 RT を入力とし、式（6）にしたがって、重み係数 w[s] を計算し出力する。

式（6）は重み係数 w[s] を残響減衰の程度によって3区分し、各区間での係数を定めたものである。

【００２３】

【数６】

$$\begin{aligned} w[s] &= 0 && 1 \le s \le \frac{D}{fr} \\ &= \alpha e^{-\frac{13.82fr}{RTf_s}s} && \frac{D}{fr} < s \le \frac{f_s}{fr}\frac{RT}{2} \\ &= 0 && \frac{f_s}{fr}\frac{RT}{2} < s \end{aligned} \quad \cdots \quad (6)$$

【００２４】

ここでf_sはサンプリング周波数[Hz]、frはフレームシフト、α はサブトラクト係数（0 < α）である。D を後期残響域に遷移するのにかかるサンプル数とした。

【００２５】

ある点に到来する反射音の時間密度は時間の2乗に比例し、室容積に反比例するため、小規模な室では後期残響への移行は早くに起こる。式 (6)において1つ目の条件、すなわち、後期残響に移行するまでの区間は直接音もしくは初期残響であるので、SS法での処理は行わない。自己のフレームに重畳した残響成分はCMN (Cepstrum Mean Normalization)により除かれるので、これは合理的な仮定である。3つめの条件は計算量の低減のために導入したものであり、必ずしも必要なものでないが、例えば30dB減衰した時点で計算を打ち切ると式 (6) のようになる。(20dBでも60dBでもかまわない。) これを概念的に表したのが、図８である。室の残響減衰特性は指数的であるが、初期残響の部分は複雑であり音声認識装置に悪影響を与えないので無視する。計算量を削減するため、十分に減衰したと

思われる時点で残響除去を打ち切ることができる。

【００２６】

次に残響成分引き去り手段22では入力された音のスペクトル11から、パワースペクトル記憶手段3に記憶された過去に入力された音源からのパワースペクトルに重み係数 w[s] を掛け合わせた重みつき和を、式（3）にしたがい減算する。音声認識のためであれば、得られたパワースペクトル4 $|Y_k[i]|^2$ をそのまま音声認識装置に出力する。可聴化のためであれば、元の位相を付加して残響を除去した音声のスペクトル4（ $|Y_k[i]|$ ）を得る。このときSS法による残留のミュージカルノイズを、フィルターを用いて除去することで音質の向上・認識性能の向上を図ることができる。

【００２７】

上述のとおり、この発明による残響除去装置は、重み係数作成手段21によって拡散音場理論より導出される式（6）にしたがい、残響時間をパラメータとする重み係数 w[s] を求める。現場で推定するパラメータが残響時間のみであるため非合理な重み係数が作成されにくく、環境の変化に頑健である。式（6）が簡単に計算できることに加えて、ある程度減衰した時点で計算を打ち切っているため計算量も少なく抑えられている。

【００２８】

実施の形態２．

発明の実施の形態１では、残響時間は所与のものとした。残響時間が未知の場合にこれを予測して残響除去を行うために、本実施の形態では残響除去手段2は残響時間予測手段20を備えている。その構成図を図２に示す。残響時間予測手段20以外の構成は図１に示す発明の実施の形態1と同様であり説明を省略する。この実施の形態２によれば、残響時間予測手段20によって予測された残響時間を重み係数作成手段21に与えることで、残響時間が未知な場合にも残響を除去することができる。

【００２９】

次に動作について説明する。全体の動作の流れ図を図１１に示す。ステップS1において音データを収録し、発話終了を待って、収録された音データから、ステップS2において対象音声区間を切り出す。たとえば収録データ全体から、発話開始フレームi_sから発話終了フレームi_eまでを切り出す。これを短時間フーリエ変換してスペクトルを求める。（フーリエ変換の次元数は前述のとおり N である。）ステップS3では残響時間予測手段20により残響時間の初期値として実際の残響時間と比べて十分小さい値RT_{init}を設定する。（逆に十分大きい値を設定してもよい。）これを元に発明の実施の形態１と同様、ステップS4で式（4）のRTをRT_{init}として重み係数作成手段21によって重み係数を作成する。ステップS5で式（1）に従い、この作成された重み係数を用いて残響成分引き去り手段22によって残響を除去する。一般にSS法を行う場合には、引き去り後のパワースペクトル $|Y_k[i]|^2$ が式（7）の条件を満たすように残響時間の値を設定する。（満たさない場合は $|Y_k[i]|^2 = \beta \cdot |X_k[i]|^2$ とする。これをフロアリングと呼び、$\beta \cdot |X_k[i]|^2$は第１の所定閾値である。）

【００３０】

【数７】

$$|Y_k[i]|^2 > \beta |X_k[i]|^2 \quad \cdots\cdots \quad (7)$$

【００３１】

ここで β はフロアリング係数（ 0 < β < 1)である。これはパワースペクトルが負になることはないという原理に基づいている。この考え方を利用して残響時間を自動的に推定することができる。

対象区間の時間・周波数平面での時系列パワースペクトルの総要素数 $N \times (i_e - i_s + 1)$ のうち、残響成分を除去した後のパワースペクトルが、式（7）を満たさず、フロアリングしたフレームの数cntを観測する。この観測されたフレーム数cntを式（8）のように時系列パワースペクトルの総要素数 $N \times (i_e - i_s + 1)$ で除した割合をrとする。

【００３２】

このときすべてのフレーム($i_e - i_s + 1$)を用いるのではなく、S / Nの高い音声区間のフレーム(($i_e - i_s + 1) - n_n$)のみを対象とすることで r の分散を小さくし、後の残響時間推定の精度を向上させることができる。ここで n_n はノイズと思われるフレーム数である。さらにβ・$|X_k[i]|^2$ が別途ノイズ区間より学習した背景騒音のレベルよりも低い場合には背景騒音の値に置き換えることで、エネルギーを引き去りすぎることにより起こる認識性能の低下を抑えることができる。

【００３３】

【数８】

$$r = \frac{cnt}{N \times (i_e - i_s + 1)} \quad \cdots\cdots \quad (8)$$

【００３４】

残響時間が長く推定されるほど、引き去られるエネルギーが大きくなるため、cntおよび rは大きくなり、残響時間が短く推定されるほど cnt および r は小さくなる。たとえば図９のように残響時間が予測よりも長いときにはより過去のスペクトルまで引かれることになるので、フロアリングする割合（左下残響成分除去後の特性図における横軸の太線部分）r が大きくなる。逆に図１０のように残響時間が予測よりも短いときにはフロアリングする割合（左下残響成分除去後の特性図における横軸の太線部分）r が小さくなる。

【００３５】

この指標rを用いて残響時間を推定する。ところで、残響成分が大半である場合、式（2）（3）に示した残響成分重畳による増加エネルギーの割合 q は室の残響時間 RT_0 に関して単調増加関数となる。（∵RT_0が長いほど室に残留しているエネルギーが大きくqが大きくなるため）RT_0が長いほど1/qは小さくなるので、式（6）で決定される係数を用いてSS法による処理を行うと、RT_0が長いほど本来式（3）で求められる音源のパワースペクトル$|Y_k[i]|^2$よりも引き去りすぎることになる。よってRT_0が長いほど、引き去られるエネルギーが大きくなるのでフロアリングしやすくなり、逆にRT_0が短いほどフロアリングしにくくなる。よってフロアリングしやすさの指標aとRT_0の間には式（9）の関係がある。

【００３６】

【数９】

$$RT_0 = \theta a - c \quad \cdots\cdots \quad (9)$$

【００３７】

ここで傾き θ と残響時間のオフセット c は事前に定めておく。a は以下のようにして算出する。まず図１２のように、推定残響時間RTを適当な初期値から始めて少しずつ増加させることで RT と r との関係を求める。このとき r は RT に関して単調増加関数となる。それから、たとえば最小2乗法を用いて回帰式 r = a RT + b を求める。このとき r は RT が大きくなると飽和するので、適切な範囲を選ぶことが必要である。たとえば RT を ΔRT だけ増加させたときの r の増分 Δr が定められた閾値よりも大きい範囲のみを対象とする。この回帰式の係数 a は RT_0 と対応するフロアリングしやすさの指標である。ゆえに上の式（９）よりRT_0はMAX(θa - c, 0)で求められる。MAXは引数の2値の大きい方を返す関数である。

【００３８】

ほとんど直接音であって a が小さい場合には、対象のスペクトルのエネルギー$|X_k[i]|^2$が小さいため、観測された場所の残響時間が長いほどフロアリングしにくくなり、短いほどフロアリングしやすくなる。よって推定残響時間と r との関係は、たとえば室の残響時間 RT_0 が0.1秒と0.5秒の場合で図１３のようになり、適切な閾値γ（第２の所定閾値）を事前の実験で設定すれば、以下の簡易的な手法でも残響を除去することができる。まず RT_{init} が十分小さい以上、ステップS6において、r < γ を満たす。（∵残響時間が0秒であればr = 0である。） 従って次にステップS7に進む。

【００３９】

次にステップS7によって、残響時間予測手段20が残響時間を RT_{init} よりもΔRTだけ増加させる。このときΔRTが十分小さくなるように注意する。この新しく設定した残響時間（ $RT_{init}+\Delta RT$ ）によってステップS4からS6を行う。すると今回算出されたｒは、必ず前回のｒよりも大きくなる。これを逐次的に繰り返し、iter回目の繰り返しでrが γを上回れば、残響時間 RT は式（10）のように推定される。

【００４０】

【数１０】

$$RT = RT_{init} + (iter - 1)\,\Delta RT \qquad \cdots\cdots \quad (10)$$

【００４１】

逆に残響時間の初期値として十分大きな RT_{init} からはじめた場合は、ΔRT だけ残響時間を減少させていき、ｒ が γ を下回れば正しく残響時間が推定されたことになる。ΔRTが十分小さければ両者によって求められた残響時間は一致する。このときに残響成分引き去り手段22より出力されたスペクトルは残響が除去されたものとなる。これをステップS8で音声認識装置に入力する。重み係数決定後の動作は実施の形態１と同様である。

【００４２】

上述のとおり、この発明の実施の形態２による残響除去装置は、残響時間予測手段20によって残響時間が未知な環境で残響時間を推定することができるため残響時間の変化に対応でき、環境変化に頑健な残響除去が行える。

【００４３】

実施の形態３．

発明の実施の形態２は前述のとおり、対象区間終了後に発話開始フレーム i_s から発話終了フレーム i_e までの全音声データを用いて、残響時間RTを推定し残響除去を行うためリアルタイム処理には不向きである。ここで発明の実施の形態１と発明の実施の形態２を組み合わせることで、リアルタイム処理を行うことができる実施の形態を説明する。そのための構成図を図３に示す。

【００４４】

残響時間はそれほど時変性がなく発話によってあまり変化しないと考えられる。よってリアルタイム処理が必要ない場合に、発明の実施の形態２と同様、図３に破線で囲まれた「学習時」と示されている手段により破線の流れで処理を行い、残響時間予測手段20によってあらかじめ残響時間を得る。作成された重み係数を重み係数記憶手段23に記憶する。リアルタイム処理が必要な場合には、図３一点鎖線で囲まれた「リアルタイム処理時」の手段により実線の流れに従って、重み係数記憶手段23に記憶された重み係数をそのまま用いて残響成分引き去り手段22によって残響除去を行う。

【００４５】

上述のとおり、この実施の形態による残響除去装置は、学習データより求めた重み係数を重み係数記憶手段23に記憶することでリアルタイム処理が可能である。これは発話長が長い場合などリアルタイム処理が特に重要な場合において効果がある。

【００４６】

実施の形態４．

発明の実施の形態２では時間・周波数平面上のパワースペクトルの要素数$N\times(i_e - i_s + 1)$のうちフロアリングする数cntを数えたが、これを周波数帯域ごとに分けることもできる。すなわち図１１の残響除去動作のフローを周波数帯域ごとに並列的に行う。

【００４７】

実施の形態４の構成は発明の実施の形態２の残響除去手段２を周波数帯域ごとに並列に複数備えたもので、図４のようになる。たとえば、残響除去処理をM個の周波数帯域に等分割して実施する場合を考える。（もちろん等分割でなくてもかまわない。）入力音のスペクトルを周波数帯域ごとに分割し、それぞれ 1_1，1_2，…，1_{M-1} とする。第m番目（$0 \leqq m \leqq M - 1$）の周波数帯域における重み係数作成手段 21_m は、各帯域独立の残響時間によって独立の重み係数を作成する。その後残響成分引き去り手段 22_m によって周波数帯域ごとにパワースペクトルを引き去り、残響時間予測手段 20_m によりそれぞれの周波数帯域

でフロアリングした数 cnt_m を数える。このとき式 (11) で表される r_m を γと比較し、$iter_m$回目の繰り返しで γを上回った場合(十分小さな RT_{init} からはじめた場合)、m番目の帯域の残響時間 RT_m を式 (12) にしたがい残響時間予測手段20_mによって推定するものである。

【００４８】

【数１１】

$$r_m = \frac{cnt_m}{N/M \times (i_e - i_s + 1)} \quad \cdots\cdots \quad (11)$$

$$RT_m = RT_{init} + (iter_m - 1)\,\Delta RT \quad \cdots\cdots \quad (12)$$

【００４９】

このように周波数帯域ごとに残響時間を算出することによって、m_1番目の周波数帯域が騒音の影響が顕著であった場合、ほかの周波数帯域（たとえばm_1-1番目）で推定された残響時間RT_{m1-1}でRT_{m1}を置き換えることが考えられる。もしくは RT_1 から RT_{M-1} の平均の残響時間を全体の残響時間とすることも考えられる。求められた残響時間から、発明の実施の形態１を用いて残響時間推定後にあらためて残響除去を行うことができる。

【００５０】

この実施の形態によれば、フラッターエコーなど音場に依存する何らかの原因で特定の周波数帯域のみ残響時間が長い場合にも適切な残響時間を設定することができる。

残響除去後の出力の音声スペクトル4_1，4_2，…，4_{M-1}も周波数帯域ごとに別々に出力されるので、これらを合成して例えば次の処理である音声認識装置に入力する。

【００５１】

上述のとおり、発明の実施の形態４による残響除去装置は、周波数帯域ごとに設けた残響時間予測手段20_mにより帯域別の残響時間を算出することができるため、騒音の影響や音場の特異性の影響を軽減することができる。

【００５２】

実施の形態５．

音源が狭帯域であった場合には、ほとんど音源の成分が含まれていない周波数帯域のパワースペクトルを用いて残響時間推定を行うと、推定の精度が低下する。そのような場合には、音源に含まれている可能性の高い周波数成分から求められた残響時間を全体の残響時間とすることもできる。この発明の実施の形態５の構成図を図５に示す。たとえば人の声は特に1 kHz以下に表われるので、16 kHzサンプリングで8 kHzまでの周波数帯域において、1 kHzから8 kHzの周波数帯域（その他の帯域）の残響時間は0 Hzから1 kHzの周波数帯域（対象帯域）の残響時間で代用する。すなわち1番目の周波数帯域の残響時間予測手段20_1によって推定された残響時間から、重み係数作成手段21_1によって重み係数を作成する。この重み係数を各帯域で共通のものとして、残響成分引き去り手段22_1，22_2に入力し、それぞれの帯域での残響除去されたスペクトルを得る。

【００５３】

上述のとおり、発明の実施の形態５による残響除去装置は、信頼度の高い周波数帯域の残響時間予測手段20_1によって推定された残響時間を全体の残響時間とすることで、主たる音源が狭帯域であった場合に残響除去の精度を向上させることができる。

【００５４】

実施の形態６．

残響の影響が少ない環境で収録された音データから残響除去を行うと、元の音のパワーを減耗させる。これによって音声認識性能が低下する可能性がある。そこで残響の影響が少ない場合には残響除去を行わないことも考えられる。この実施の形態６の構成図を図６に示す。この実施の形態は図２に示す実施の形態２の構成に新たに出力スペクトル決定手段5が備えられている。出力スペクトル決定手段5は残響時間予測手段20からの確定した残響時間 RT が、ある閾値 RT_{min} よりも小さい場合には、入力の音声スペクトル1をそのま

ま出力の音声スペクトル4とする。残響時間 RT が閾値 RT_{min} よりも大きければ、残響を除去したスペクトルを出力の音声スペクトル4とする。これによって、残響の影響が小さい場合にも認識性能を維持できる。

【００５５】

上述のとおり、発明の実施の形態６による残響除去装置は残響時間の閾値 RT_{min} の設定を適切に行うことで、出力スペクトル決定手段5によって残響除去を行ったスペクトルと行わないスペクトルから音声認識に有用な情報がより多く含まれている方を選択でき、それらの良いほうを用いた場合の認識性能を得ることができる。

【００５６】

実施の形態７．

定常騒音の場合、SS法により騒音成分のみを除去してから残響除去を行うことで騒音がない場合と同様に扱うことができる。この実施の形態７の構成図を図１４に示す。この実施の形態７は騒音がない場合の実施の形態２と比べて、騒音成分引き去り手段8と騒音パワースペクトル記憶手段6が新たに備えられている。騒音成分引き去り手段8は残響成分引き去り手段22の前段に備えられる。騒音成分引き去り手段8により騒音パワースペクトル記憶手段6に記憶された、あらかじめ学習された過去の騒音の平均パワースペクトルを用いて入力された音声スペクトル1から騒音成分を引き去ることで、騒音がない場合の実施の形態２と同じようにして残響除去を行う。

【００５７】

実施の形態８．

通常の騒音はそれほど定常ではなく、SS法による騒音除去はそれほど有効に働かない。有効に働かない場合には、騒音のスペクトルが引かれた部分と引かれていない部分の不連続性によって音声認識性能が低下する。よって騒音成分は除去せず、騒音重畳モデルで対応し、残響成分のみを除去することが有効であると考えられる。

騒音が存在する場合には、式(1)の 受音点で観測された音のパワースペクトル $|X_k[i]|^2$ は背景騒音のパワースペクトル $|N_k|^2$ を考慮して式(13)のようになる。ここで $|N_k|^2$ は定常的であると仮定する。

【００５８】

【数１２】

$$|X_k[i]|^2 = \sum_{s=0}^{i} w[s] \cdot |Y_k[i-s]|^2 + |N_k|^2 \qquad \cdots\cdots \quad (13)$$

【００５９】

さらに式(2)は式(14)のようになる。

【００６０】

【数１３】

$$q|Y_k[i-s]|^2 + |N_k|^2 \approx |X_k[i-s]|^2 \qquad \cdots\cdots \quad (14)$$

【００６１】

式(13)，(14)より式(15)が導かれる。

【００６２】

【数１４】

$$|Y_k[i]|^2 + |N_k|^2 = |X_k[i]|^2 - \frac{1}{q}\sum_{s=1}^{i} w[s] \cdot \left(|X_k[i-s]|^2 - |N_k|^2\right) \qquad \cdots\cdots \quad (15)$$

【００６３】

これによって発明の実施の形態1と同様、重み係数w[s]を決定すれば、式(15)にしたがいSS法により残響除去ができる。ここで背景騒音のパワースペクトル $|N_k|^2$ は発話前の数フレームの平均から求めることとする。理想的には現在の音源のパワースペクトル $|Y_k[i]|^2$ を求めたいが、背景騒音成分は完全に定常ではないのでSS法によって完全に背景騒音

を除去することはできない。よって上記に述べたとおり、背景騒音は除去せずに残響のみを除去することを考え、現在の音源のパワースペクトル$|Y_k[i]|^2$ではなく、現在の音源のパワースペクトル$|Y_k[i]|^2$に背景騒音のパワースペクトル$|N_k|^2$を加算した式(15)の左辺$|Y_k[i]|^2+|N_k|^2$を求めることとする。

【００６４】

これを概念的に述べると、騒音成分が十分定常であれば図１８のように常に存在するため、パワースペクトル記憶手段3に記憶された過去の入力音声のパワースペクトルにも入力音声データ1と同程度のレベルの騒音が含まれる。図１８では、背景騒音に埋もれた残響音を後記減衰残響としている。ゆえに、騒音がない場合と同じように、図７のような減衰を仮定して引き去ると、入力音声データ1に含まれる騒音のエネルギーが引き去られすぎることになる。

よってパワースペクトル記憶手段3に記憶された過去の入力音声のパワースペクトルから騒音成分を取り除いておく必要がある。この実施の形態８の構成図を図１５に示す。実施の形態７のように入力音声データ1から騒音成分を除去するのではなく、騒音パワースペクトル記憶手段6に記憶された、あらかじめ学習された過去の騒音の平均パワースペクトルを参照して騒音成分引き去り手段8が、パワースペクトル記憶手段3に記憶された過去の入力音声のパワースペクトルから騒音成分を除去し、残響成分引き去り手段22に入力することで、入力音声データ1に含まれている騒音成分に悪影響を与えることなく、それに含まれている残響成分のみを取り去ることができる。

さらに重み係数作成手段21にも改良を加えることができる。

【００６５】

【数１５】

後期残響（図１８(b), (c)）は拡散音場理論に基づき、室の平均エネルギー密度$\bar{E}(t)$ の時間変化を指数関数の形に仮定できる。騒音が存在することを考慮すると$\bar{E}(t)$は式(16)のように表せる。

$$\begin{aligned}\bar{E}(t) &= \bar{E}(0)\exp\left(-\frac{13.82}{RT_0}t\right) \quad for(b)\\ &= \sum_{k=0}^{N-1}|N_k|^2 \qquad for(c)\end{aligned} \qquad \cdots\cdots \quad (16)$$

【００６６】

これより、図１８の初期残響(a)は無視し、後期残響(b)は式(16)の指数的減衰を仮定し、直接音より1発話中の最良のS/N（SN_{max}）だけ減衰した(c)の部分は背景騒音と等しいとすることで重み係数 w[s]を式(17)のように決定できる。

【００６７】

【数１６】

$$\begin{aligned}w[s] &= 0 && 1 \le s \le \frac{D}{fr}\\ &= \sigma\alpha\exp\left(-\frac{13.82fr}{RT_0 f_s}s\right) && \frac{D}{fr} < s \le \frac{f_s}{fr}\frac{SN_{max}RT_0}{60}\\ &= 0 && \frac{f_s}{fr}\frac{SN_{max}RT_0}{60} < s\end{aligned} \qquad \cdots\cdots \quad (17)$$

【００６８】

σは1としてもよいが、引き去りエネルギーの正規化係数として用いることができる。SN_{max}の違いによって引き去り量に差が出てしまうが、音声の指数減衰を仮定してσを乗じることで引き去り量を正規化することができる。たとえばSN_{max} の上限を30dBに設定した場合には、SN_{max}がそれ以下の場合には式(18)に示すようにσを設定すればよい。

【００６９】

【数１７】

$$\sigma = \frac{\int_0^{\frac{30RT}{60}} \exp(-t)dt}{\int_0^{\frac{SN_{max}RT}{60}} \exp(-t)dt} = \frac{1-\exp(-30RT/60)}{1-\exp(-SN_{max}RT/60)} \quad \cdots\cdots (18)$$

【００７０】

実施の形態９．

音声認識に悪影響を与えるのは、音声でかつ残響成分が主体的な場合であるので、別の枠組みによって音声/非音声、直接音/残響音を判断し、音声でかつ残響成分の場合のみ残響除去を行うことも考えられる。このような機能を有する実施の形態９の構成図を図１６に示す。この実施の形態９は実施の形態２に加えて、音声/非音声、直接音/残響音判定手段7、フロアリング係数作成手段24を備えている。音声/非音声、直接音/残響音判定手段7によって入力音声スペクトル１が、音声で残響音であると判定された場合にはフロアリング係数作成手段24によって式(7)のフロアリング係数βを小さく設定することで残響除去の効果を大きくし、それ以外の場合にはβを大きく設定することで、もとの音声データのスペクトルに必要以上のひずみを生じさせなくすることができる。

【００７１】

音声/非音声、直接音/残響音判定手段7の判定は、騒音パワースペクトル記憶手段6に記憶された過去の騒音の平均スペクトルと現在の音声スペクトル1を比較することによって行う。判定の基準には、たとえばS/Nやそれらの相関を利用することができる。ここではS/Nを用いる手法に関して述べるが、相関を用いた場合も同様である。S/Nの水準を4段階設け、mp以上では直接音が優位、cp付近では背景騒音と音声が混在、np以下では背景騒音である可能性が高いと考えた。すなわちmp > cp > npの順を仮定している。cp < SN < mpとなるフレームが残響を最も引き去りたい図１８(b)の遷移域に対応しており、そこでのβが小さくなるように式(19)に従いフロアリング係数作成手段24でβを設定した。

【００７２】

【数１８】

$$\beta = \begin{cases} 1 & SN > mp \\ \frac{1-\beta_{min}}{mp-cp}(SN-cp)+\beta_{min} & cp < SN \leq mp \\ \frac{\beta_{min}-1}{cp-np}(SN-np)+1 & np < SN \leq cp \\ 1 & SN \leq np \end{cases} \quad \cdots\cdots (19)$$

【００７３】

実施の形態１０．

発明の実施の形態８と９を組み合わせたものも考えられる。この実施の形態１０の構成図を図１７に示す。図１７に示される各構成要素における動作は実施の形態８および実施の形態９と同様であるため説明を省略する。このような構成によってパワースペクトル記憶手段3に記憶された過去の入力音声のパワースペクトルから騒音成分を取り除きつつ、音声でかつ残響成分であると判断されたものについてのみ残響除去が行われる。このため、入力音声スペクトル1からの騒音成分の引き去りすぎが起こりにくく、精度よく残響成分のみを除去することができる。

【産業上の利用可能性】

【００７４】

家電品やさらにはロボットに適用される音声認識装置の前処理として入力音声からの残

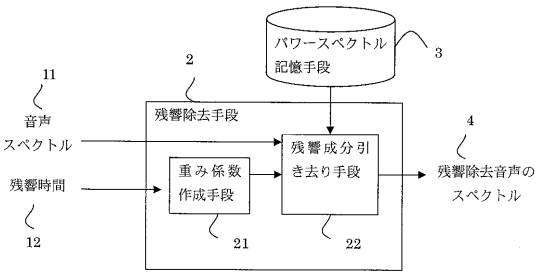
響を除去する残響除去装置や、電話会議システムにおいて、遠隔地へ音声を送信する際に会議室における残響を除去することで、聞き取り易さを改善する残響除去装置などに適用される。

【符号の説明】

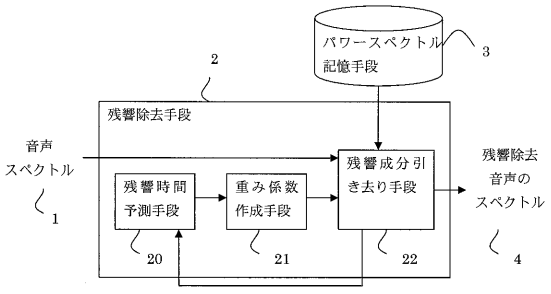
【００７５】

２；残響除去手段、３；パワースペクトル記憶手段、４；残響成分除去音のスペクトル、５；出力スペクトル決定手段、６；騒音パワースペクトル記憶手段、７；音声/非音声、直接音/残響音判定手段、８；騒音成分引き去り手段、１１；入力音スペクトル、１２；残響時間、２０；残響時間予測手段、２１；重み係数作成手段、２２；残響成分引き去り手段、２３；重み係数記憶手段、２４；フロアリング係数作成手段。

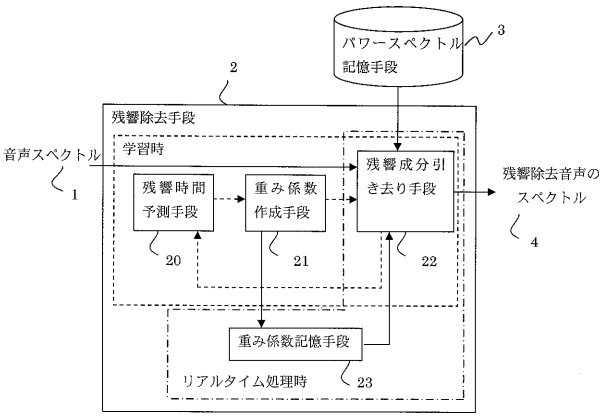
【図１】



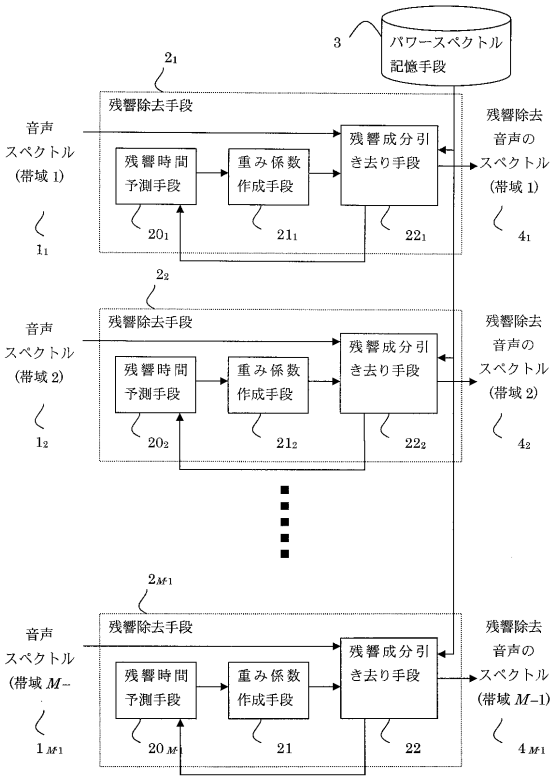
【図２】



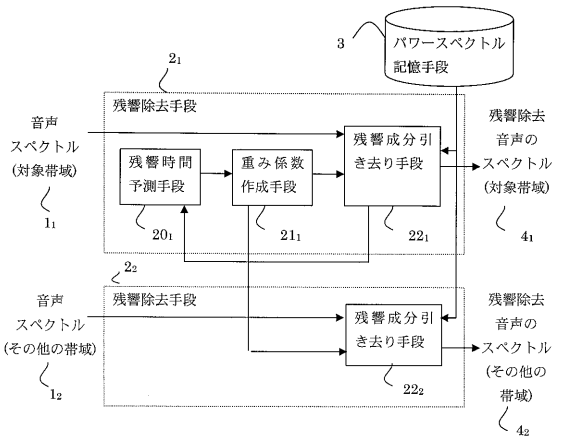
【図３】



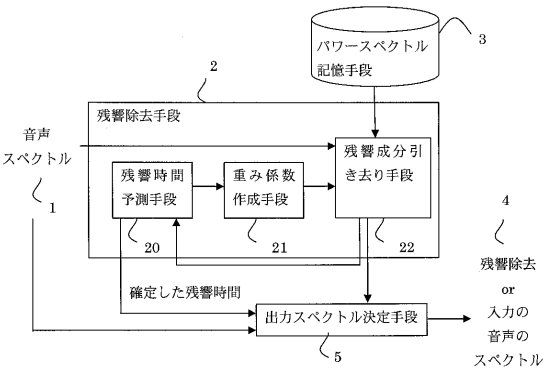
【図４】



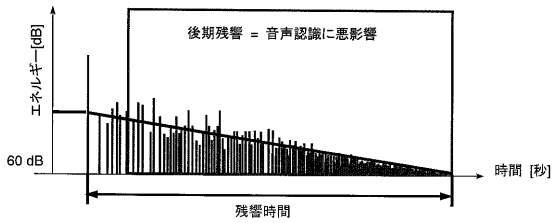
【図５】



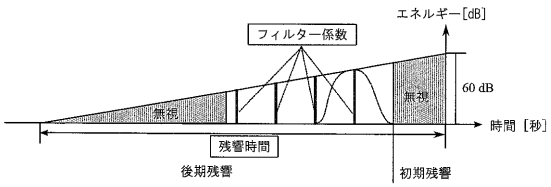
【図６】



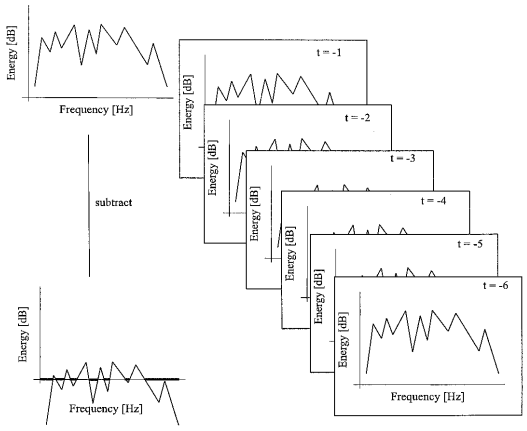
【図７】



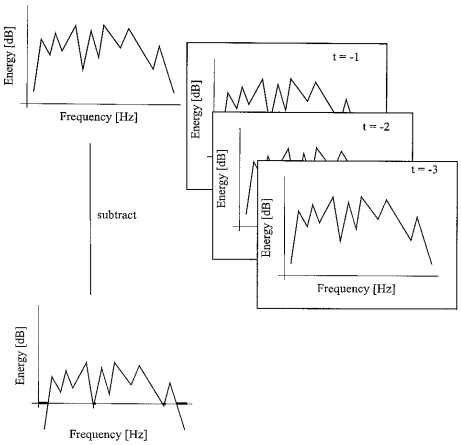
【図８】



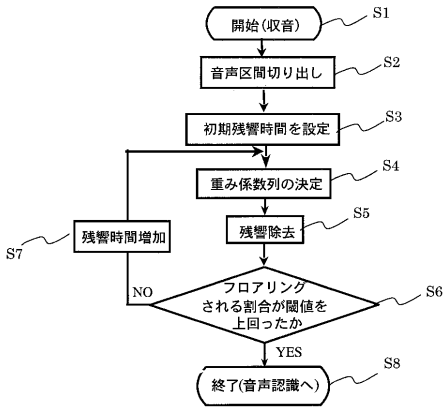
【図９】



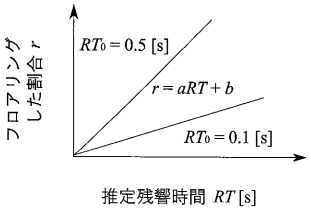
【図１０】



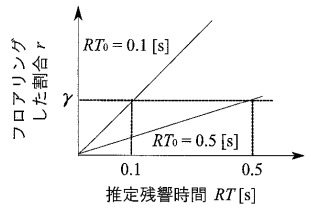
【図１１】



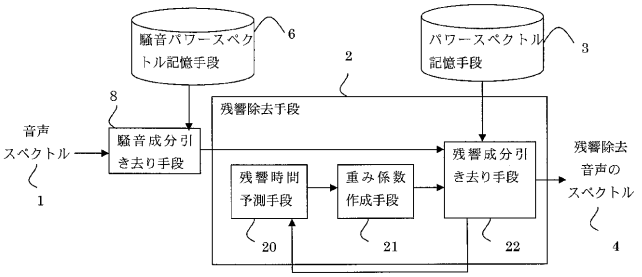
【図１２】



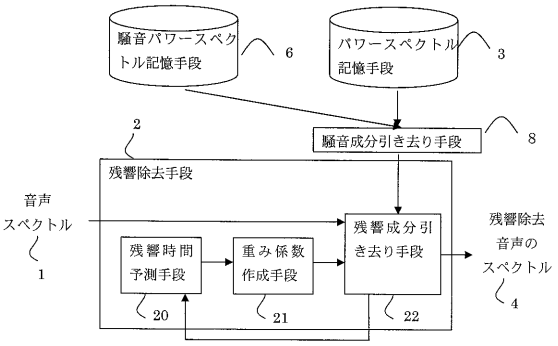
【図１３】



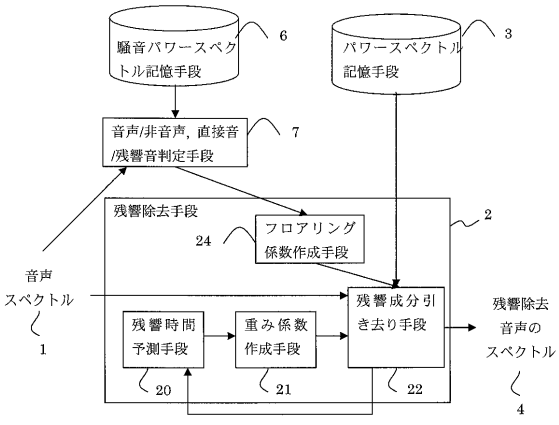
【図１４】



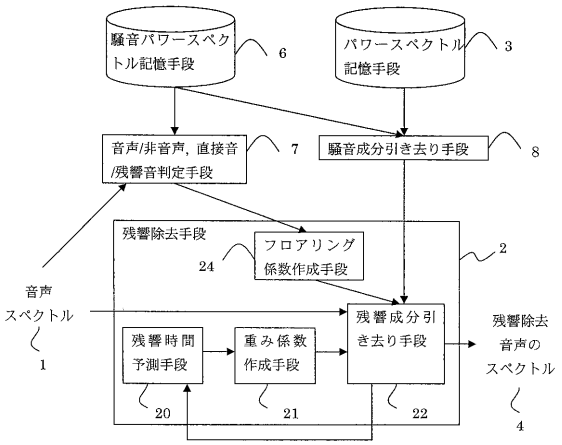
【図１５】



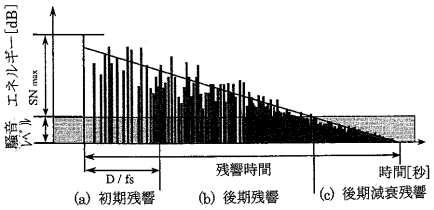
【図１６】



【図１７】



【図１８】



フロントページの続き

(74)代理人　100161171
弁理士　吉田　潤一郎
(72)発明者　太刀岡　勇気
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内