

(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2010-49083
(P2010-49083A)

(43) 公開日 平成22年3月4日(2010.3.4)

(51) Int.Cl.	F I	テーマコード (参考)
G10L 21/02 (2006.01)	G10L 21/02 201D	5D020
H04R 3/00 (2006.01)	H04R 3/00 320	
	G10L 21/02 103Z	

審査請求 未請求 請求項の数 10 O L (全 17 頁)

(21) 出願番号	特願2008-214065 (P2008-214065)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号
(22) 出願日	平成20年8月22日 (2008.8.22)	(74) 代理人	100121706 弁理士 中尾 直樹
		(74) 代理人	100066153 弁理士 草野 卓
		(74) 代理人	100128705 弁理士 中村 幸雄
		(72) 発明者	吉岡 拓也 東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内
		(72) 発明者	中谷 智広 東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 音響信号強調装置とその方法と、プログラムと記録媒体

(57) 【要約】

【課題】残響時間がフレーム幅を越える場合でも音源分離を可能にする。

【解決手段】この発明の音響信号強調装置は、帯域分割部と、パラメータ推定部と、残響除去部と、音源分離部と、帯域合成部とを具備する。帯域分割部は、時間領域の観測音響信号を所定数の周波数帯域毎の観測信号に分割する。パラメータ推定部は、周波数帯域毎の観測信号を入力として所定フレーム数前までの残響成分を推定するための回帰行列の推定値と、音源信号を推定するための分離行列の推定値とを生成する。残響除去部は、周波数帯域毎の観測信号に回帰行列を畳み込み演算して残響成分を求める。そして、求めた残響成分を周波数帯域毎の観測信号から減算して残響除去信号を生成する。音源分離部は、残響除去信号に分離行列を乗算して音源信号の推定値を生成する。帯域合成部は、音源信号の推定値を合成して時間領域の強調音響信号に変換する。

【選択図】 図1

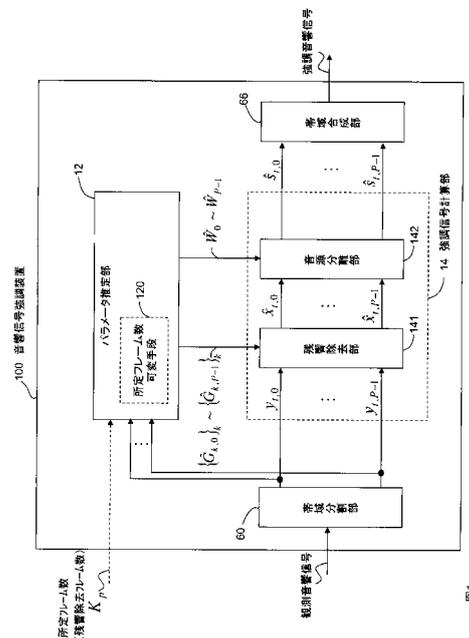


図1

【特許請求の範囲】

【請求項 1】

時間領域の観測音響信号を所定数の周波数帯域毎の観測信号に分割する帯域分割部と、
上記観測信号を入力として所定フレーム数前までの残響成分を推定するための回帰行列
の推定値と、音源信号に分離するための分離行列の推定値とを生成するパラメータ推定部
と、

上記観測信号に上記回帰行列を畳み込み演算して残響成分を求め、上記残響成分を上記
観測信号から減算して残響除去信号を生成する残響除去部と、

上記残響除去信号に上記分離行列を乗算して音源信号の推定値を生成する音源分離部と

10

、
上記音源信号の推定値を合成して時間領域の強調音響信号を生成する帯域合成部と、
を具備する音響信号強調装置。

【請求項 2】

請求項 1 に記載した音響信号強調装置において、

上記パラメータ推定部は、

上記観測信号を記録する観測信号記録部と、

上記回帰行列の推定値と上記分離行列の推定値の初期値を設定するパラメータ推定値初
期化部と、

上記初期値又は上記回帰行列の推定値と上記分離行列の推定値と、上記観測信号とを入
力として全極パラメータ推定値を更新する全極パラメータ推定値更新部と、

20

上記全極パラメータ推定値を入力として音源信号のパワースペクトル密度の推定値を更
新する音源信号パワースペクトル推定値更新部と、

上記音源信号のパワースペクトル密度の推定値と上記分離行列の推定値と上記観測信号
とを入力として上記回帰行列の推定値を更新する回帰行列推定値更新部と、

上記回帰行列の推定値と上記音源信号のパワースペクトル密度の推定値と上記観測信号
とを入力として上記分離行列の推定値を更新する分離行列推定値更新部と、

上記回帰行列の推定値と上記分離行列の推定値が収束したか否かを判定する収束判定部
と、

を備えることを特徴とする音響信号強調装置。

【請求項 3】

30

請求項 2 に記載した音響信号強調装置において、

上記分離行列推定値更新部は、

上記分離行列の推定値を、音源分離技術における分離行列の計算法を用いて更新するも
のであることを特徴とする音響信号強調装置。

【請求項 4】

請求項 1 乃至 3 の何れかに記載した音響信号強調装置において、

上記パラメータ推定部は、上記所定フレーム数を上記観測信号の周波数帯域に応じて可
変する所定フレーム数可変手段を備えることを特徴とする音響信号強調装置。

【請求項 5】

帯域分割部が、時間領域の観測音響信号を所定数の周波数帯域毎の観測信号に分割する
帯域分割過程と、

40

パラメータ推定部が、上記観測信号を入力として所定フレーム数前までの残響成分を推
定するための回帰行列の推定値と、音源信号に分離するための分離行列の推定値とを生成
するパラメータ推定過程と、

残響除去部が、上記観測信号に上記回帰行列を畳み込み演算して残響成分を求め、上記
残響成分を上記観測信号から減算して残響除去信号を生成する残響除去過程と、

音源分離部が、上記残響除去信号に上記分離行列を乗算して音源信号の推定値を生成す
る音源分離過程と、

帯域合成部が、上記音源信号の推定値を合成して時間領域の強調音響信号を生成する帯
域合成過程と、

50

を含む音響信号強調方法。

【請求項 6】

請求項 5 に記載した音響信号強調方法において、

上記パラメータ推定過程は、

パラメータ推定値初期化部が、上記回帰行列の推定値と上記分離行列の推定値の初期値を設定するパラメータ推定値初期化ステップと、

全極パラメータ推定値更新部が、上記初期値又は上記回帰行列の推定値と上記分離行列の推定値と、上記観測信号とを入力として全極パラメータ推定値を更新する全極パラメータ推定値更新ステップと、

音源信号パワースペクトル推定値更新部が、上記全極パラメータ推定値を入力として音源信号のパワースペクトル密度の推定値を更新する音源信号パワースペクトル推定値更新ステップと、

回帰行列推定値更新部が、上記音源信号のパワースペクトル密度の推定値と上記分離行列の推定値と上記観測信号とを入力として上記回帰行列の推定値を更新する回帰行列推定値更新ステップと、

分離行列推定値更新部が、上記回帰行列の推定値と上記音源信号のパワースペクトル密度の推定値と上記観測信号とを入力として上記分離行列の推定値を更新する分離行列推定値更新ステップと、

収束判定部が、上記回帰行列の推定値と上記分離行列の推定値が収束したか否かを判定する収束判定ステップと、

を含むことを特徴とする音響信号強調方法。

【請求項 7】

請求項 6 に記載した音響信号強調方法において、

上記分離行列推定値更新過程は、

上記分離行列の推定値を、音源分離技術における分離行列の計算法を用いて更新する過程であることを特徴とする音響信号強調方法。

【請求項 8】

請求項 5 乃至 7 の何れかに記載した音響信号強調方法において、

上記パラメータ推定過程は、上記所定フレーム数を上記観測信号の周波数帯域に応じて可変する所定フレーム数可変ステップを含むことを特徴とする音響信号強調方法。

【請求項 9】

請求項 1 乃至 4 の何れかに記載した音響信号強調装置としてコンピュータを機能させるための装置プログラム。

【請求項 10】

請求項 9 に記載した何れかの装置プログラムを記録したコンピュータで読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、例えば音声認識システムやテレビ会議システム等に利用可能な音響信号強調装置とその方法と、プログラムと記録媒体に関する。

【背景技術】

【0002】

従来、音源分離技術による音響信号強調装置の一般的な構成について、図 6 を参照して説明する。以後の説明では、各音源から発せられた音の信号を音源信号、マイクロホンで観測された音の信号を観測信号と呼ぶ。特に断らない限り、これらは周波数帯域毎の狭帯域信号を表すものとする。時間領域の信号は、観測音響信号、などのように「音響」という語を付記して表す。音響信号強調装置 600 は、帯域分割部 60、パラメータ推定部 62、音源分離部 64、帯域合成部 66 を備える。帯域分割部 60 は、マイクロホンや音声ファイル等から入力される標本化かつ量子化された時間領域の観測音響信号を、周波数帯

10

20

30

40

50

域毎の観測信号 $y_{t,p}^{(m)}$ に分割する。 m, t, p はそれぞれマイクロホン、時間、周波数帯域のインデックスを示す。全てのマイクロホンによる観測信号を式(1)のようにベクトル形式で表す。

【0003】

【数1】

$$y_{t,p} = [y_{t,p}^{(1)}, \dots, y_{t,p}^{(M)}]^T \quad (1)$$

【0004】

ここで M はマイクロホンの個数、 \cdot^T は非共役転置を表す。従来の音源分離技術は、観測信号ベクトル $y_{t,p}$ を式(2)により生成されたと仮定する。

【数2】

$$y_{t,p} = W_p^{-H} s_{t,p} \quad (2)$$

【0005】

ここで H は共役転置、 $-H$ は逆行列の共役転置、 $s_{t,p}$ は真の音源信号からなるベクトル、 W_p は真の分離行列である。音源信号ベクトルは、 $s_{t,p}^{(m)}$ を m 番目の音源信号として式(3)で定義する。

【数3】

$$s_{t,p} = [s_{t,p}^{(1)}, \dots, s_{t,p}^{(M)}]^T \quad (3)$$

【0006】

パラメータ推定部62は、全ての時間、周波数帯域に渡る観測信号ベクトル $\{y_{t,p}\}_{0 \leq t \leq T-1, 0 \leq p \leq P-1}$ から、各周波数帯域の分離行列 W_p の推定値 W_p^\wedge を計算する。音源分離部64は、各周波数帯域について、各周波数帯域の分離行列の推定値 W_p^\wedge を式(4)に示すように各周波数帯域の観測音響信号ベクトル $y_{t,p}$ に適用することで、各周波数帯域の音源信号ベクトル $s_{t,p}$ の推定値 $s_{t,p}^\wedge$ を計算する。

【数4】

$$\hat{s}_{t,p} = \hat{W}_p^H y_{t,p} \quad (4)$$

$$\hat{s}_{t,p} = [\hat{s}_{t,p}^{(1)}, \dots, \hat{s}_{t,p}^{(M)}]^T$$

【0007】

帯域合成部66は、音源信号ベクトルの推定値を合成して時間領域の強調音響信号を計算する。帯域分割部60は、短時間フーリエ変換やポリフェーズフィルタバンク分析を用いて観測音響信号を周波数帯域毎の観測信号 $y_{t,p}^{(m)}$ に分割する。帯域合成部66は、帯域分割部60が短時間フーリエ変換を用いた場合はオーバーラップ加算合成を実行し、帯域分割部60がポリフェーズフィルタバンク分析を用いた場合はポリフェーズフィルタバンク合成を実行する。以後、帯域分割部60を短時間フーリエ変換で行う場合を例に説明するが、帯域分割部60には他の時間周波数解析方法を用いても良い。

【0008】

例えば、非特許文献1に開示された従来の音源分離技術では、パラメータ推定部62が各音源信号 $s_{t,p}^{(1)}, \dots, s_{t,p}^{(m)}$ を互いに統計的に独立であると仮定して、 $s_{t,p}^{(1)\wedge}, \dots, s_{t,p}^{(m)\wedge}$ の間の独立性が最大になるように各周波数帯域の分離行列 W_p の推定値 W_p^\wedge を計算する。

【非特許文献1】牧野昭二、荒木章子、向井良、澤田宏「畳み込み混合のブラインド音源分離」システム/制御/情報、Vol.48, No.10, pp.401-408, 2004

【発明の開示】

【発明が解決しようとする課題】

【0009】

10

20

30

40

50

従来の音響信号強調装置では、残響時間が短時間フーリエ変換のフレーム幅を超える場合に、干渉信号の残響成分が除去されない問題点があった。干渉信号の残響成分を除去する目的で、フレーム内に残響成分が含まれるように、フレーム幅を増やす方法が考えられる。しかし、単純にフレーム幅を増やしても、帯域分割部60における分割数が増えることで音源信号間の独立性の仮定が成り立たなくなるため、各分離行列の推定が困難になることが知られている。この現象は、例えば参考文献「Araki, S., Mukai, R., Makino, S., Nishikawa, T and Saruwatari, H.: The fundamental limitation of frequency domain blind source separation for convolutivemixtures of speech, IEEE Trans. Speech, Audio Process., Vol. 11, No. 2, pp. 109-1166, 2003」に報告されている。

【0010】

10

この発明は、このような問題点に鑑みてなされたものであり、残響時間が短時間フーリエ変換のフレーム幅を越える場合でも音源分離が可能な音響信号強調装置とその方法と、プログラムと記録媒体を提供することを目的とする。

【課題を解決するための手段】

【0011】

20

この発明の音響信号強調装置は、帯域分割部と、パラメータ推定部と、残響除去部と、音源分離部と、帯域合成部とを具備する。帯域分割部は、時間領域の観測音響信号を所定数の周波数帯域毎の観測信号に分割する。パラメータ推定部は、周波数帯域毎の観測信号を入力として所定フレーム数前までの残響成分を推定するための回帰行列の推定値と、音源信号に分離するための分離行列の推定値とを生成する。残響除去部は、周波数帯域毎の観測信号に回帰行列を畳み込み演算して残響成分を求める。そして、求めた残響成分を周波数帯域毎の観測信号から減算して残響除去信号を生成する。音源分離部は、周波数帯域毎の残響除去信号に分離行列を乗算して音源信号の推定値を生成する。帯域合成部は、音源信号の推定値を合成して時間領域の強調音響信号を生成する。

【発明の効果】

【0012】

30

この発明の音響信号強調装置は、残響除去部で残響を除去した後の残響除去信号に、分離行列を乗算して音源分離を実施する。したがって、残響時間がフレーム幅を越える場合でも音源分離することが可能になる。また、強調後の音響信号は残響成分が抑圧されているため、より高品質な強調音響信号を得ることが出来る。

【発明を実施するための最良の形態】

【0013】

以下、この発明の実施の形態を図面を参照して説明する。複数の図面中同一のものには同じ参照符号を付し、説明は繰り返さない。

〔この発明の基本的な考え〕

実施例の説明の前に、この発明の音響信号強調方法の基本的な考えについて説明する。この発明では、観測信号 $y_{t,p}$ を式(2)ではなく式(5)で生成されたと仮定する。

【数5】

$$y_{t,p} = \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p} + x_{t,p} \quad (5)$$

40

$$x_{t,p} = W_p^{-H} s_{t,p} \quad (6)$$

【0014】

つまり、観測信号 $y_{t,p}$ を未知のパラメータ $G_{k,p}$ と W_p と音源信号 $s_{t,p}$ とから生成されるものとモデル化している。このモデルは、音源信号 $s_{t,p}$ が混合行列 W_p^{-H} によって混合されて $x_{t,p}$ となり、さらにパラメータ $G_{k,p}$ を k 次の回帰行列にもつ M 入力 M 出力の自己回帰系によって残響が重畳されて観測信号 $y_{t,p}$ となることを意味している。この発明のパラメータ推定部は、観測信号 $y_{t,p}$ から全ての周波数帯域 p ($0 \leq p \leq P-1$) に渡って、 $G_{k,p}$ と W_p とを推定する。 K_p は、 p 番目の周波数帯

50

域における残響の長さに対応し、これを残響除去フレーム数と称する。

【 0 0 1 5 】

パラメータ推定部で推定された各周波数帯域の回帰行列の推定値を $G_{k,p}^{\wedge}$ 、分離行列の推定値を W_p^{\wedge} とおくと、この発明は、式 (7) と式 (8) に示すようにして音源信号を推定する。

【 数 6 】

$$\hat{x}_{t,p} = y_{t,p} - \sum_{k=1}^{K_p} \hat{G}_{k,p}^H y_{t-k,p} \quad (7)$$

$$\hat{s}_{t,p} = \hat{W}_p^H \hat{x}_{t,p} \quad (8)$$

10

【 0 0 1 6 】

この発明は、式 (7) に示すように観測信号 $y_{t,p}$ に回帰行列の推定値 $G_{k,p}^{\wedge}$ を畳み込み演算して残響成分を求め、その残響成分を観測信号から減算して残響除去信号 $x_{t,p}^{\wedge}$ を求める。そして残響除去信号 $x_{t,p}^{\wedge}$ に分離行列の推定値 W_p^{\wedge} を乗算して音源分離を実施する。よって、残響時間がフレーム幅を越える場合でも音源分離することが可能になる。

【 実施例 1 】

【 0 0 1 7 】

図 1 にこの発明の音響信号強調装置 1 0 0 の機能構成例を示す。その動作フローを図 2 に示す。音響信号強調装置 1 0 0 は、帯域分割部 6 0 と、パラメータ推定部 1 2 と、強調信号計算部 1 4 と、帯域合成部 6 6 とを備える。帯域分割部 6 0 と帯域合成部 6 6 とは、従来の音響信号強調装置 6 0 0 と同じものである。音響信号強調装置 1 0 0 は、例えば R O M、R A M、C P U 等で構成されるコンピュータに所定のプログラムが読み込まれて、C P U がそのプログラムを実行することで実現される。

20

【 0 0 1 8 】

帯域分割部 6 0 は、時間領域の観測音響信号を所定数の周波数帯域毎の観測信号 $y_{t,p}$ に分割する (ステップ S 6 0)。パラメータ推定部 1 2 は、観測信号 $y_{t,p}$ を入力として所定フレーム数前までの残響成分を推定するための回帰行列の推定値 $G_{k,p}^{\wedge}$ と、音源信号を推定するための分離行列の推定値 W_p^{\wedge} とを生成する (ステップ S 1 2)。ここで、所定フレーム数前とは、上記した残響除去フレーム数 K_p のことであり、この値は外部から設定されるようにしても良いし、パラメータ推定部 1 2 の内部に予め固定値として設定しておいても良い。

30

【 0 0 1 9 】

回帰行列 $G_{k,p}^{\wedge}$ と分離行列 W_p^{\wedge} の推定値は、式 (9) に示す対数尤度関数を最大化する $G_{k,p}$ 及び W_p の値として求められる。

【 数 7 】

$$L(\Theta; y) = \sum_{p=0}^{P-1} \sum_{t=0}^{T-1} \left\{ \log \det \left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P} \right)^{-1} W_p^H \right) - \left(y_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p} \right)^H W_{ps} \Lambda_t \left(\frac{2\pi p}{P} \right)^{-1} \times W_p^H \left(y_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p} \right) \right\} \quad (9)$$

40

【 0 0 2 0 】

50

ただし、 Θ は全ての未知パラメータの集合であり、それぞれ次のように定義される。

【数 8】

$$\Theta = \{s, w, g\} \quad (10)$$

$$s = \{a_{t,1}^{(m)}, \dots, a_{t,Q}^{(m)}, v_t^{(m)}\}_{1 \leq m \leq M, 0 \leq t \leq T-1} \quad (11)$$

$$w = \{W_p\}_{0 \leq p \leq P-1} \quad (12)$$

$$g = \left\{ \left\{ G_{k,p} \right\}_{1 \leq k \leq K_p} \right\}_{0 \leq p \leq P-1} \quad (13)$$

10

式(11)の s は、以後に説明する音源信号の全極パラメータの集合である。また、式(12)の w と式(13)の g は、それぞれ分離行列と回帰行列の集合である。また、式(9)に現れる s_t については後述する。

【0021】

式(9)の導出過程を説明する。まず、推定する音源信号 $S_{t,p}^{(m)}$ の確率密度関数を求める。この発明では各音源信号の短時間パワースペクトル密度は、 Q 次の全極型の関数で与えられると仮定する。すなわち、 ω を角周波数、第 t フレームにおける m 番目の音源信号のパワースペクトル密度を $s_t^{(m)}(\omega)$ と書くと、音源信号のパワースペクトル密度 $s_t^{(m)}(\omega)$ は式(14)で与えられる。

20

【数 9】

$$s_t^{(m)}(\omega) = \frac{s v_t^{(m)}}{|A_t^{(m)}(e^{j\omega})|^2} \quad (14)$$

$$A_t^{(m)}(z) = 1 - a_{t,1}^{(m)} z^{-1} - \dots - a_{t,Q}^{(m)} z^{-Q} \quad (15)$$

【0022】

ここで、 $a_{t,k}^{(m)}$ は線形予測係数、 $s v_t^{(m)}$ は予測残差パワーである。また、 $a_{t,1}^{(m)}, \dots, a_{t,Q}^{(m)}$ と $s v_t^{(m)}$ をまとめて全極パラメータと呼ぶ。更に、音源信号ベクトル $S_{t,p}$ の共分散行列を M 次元対角行列 s_t を用いて式(16)のように表す。つまり、複数の音源信号間に相関が無いと仮定している。

30

【数 10】

$$s_t = \begin{bmatrix} s \lambda_t^{(1)}(\omega) & & 0 \\ & \ddots & \\ 0 & & s \lambda_t^{(M)}(\omega) \end{bmatrix} \quad (16)$$

40

【0023】

音源信号 $S_{t,p}^{(m)}$ は、平均0、分散 $s_t^{(m)}(2\pi p/P)$ の複素正規分布に従うので、音源信号 $S_{t,p}^{(m)}$ の確率密度関数は、式(17)で与えられる。

【数 11】

$$p(S_{t,p}^{(m)}; a_{t,1}^{(m)}, \dots, a_{t,Q}^{(m)}, s v_t^{(m)}) = N\{S_{t,p}^{(m)}; 0, s \lambda_t^{(m)}(2\pi p/P)\} \quad (17)$$

【0024】

$N\{x; \mu, \Sigma\}$ は、平均 μ 、共分散行列 Σ の複素正規分布に従う確率変数 x の確率密度関数を表す。例えば、 x が D 次元ベクトルである場合、 $N\{x; \mu, \Sigma\}$ は具体的には式(18)で与えられる。

50

【数 1 2】

$$N\{x; \mu, \Sigma\} = \frac{1}{\pi^D \det \Sigma} \exp\left\{-x^H \Sigma^{-1} x\right\} \quad (18)$$

ここで、任意のマイクロホン番号 m について、 $(t_1, p_1) = (t_2, p_2)$ でなければ、 $S_{t_1, p_1}^{(m)}$ と $S_{t_2, p_2}^{(m)}$ は統計的に独立である。また、 m_1, m_2 、すなわちマイクロホン番号が異なるならば $S_{t_1, p_1}^{(m_1)}$ と $S_{t_2, p_2}^{(m_2)}$ は常に統計的に独立である。

【0025】

式(6)と式(17)とから、 $x_{t, p}$ の確率密度関数は式(19)のように書ける。 10

【数 1 3】

$$\begin{aligned} p(x_{t,p} | \Theta) &= N\left\{x_{t,p}; 0, \left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} W_p^H\right)^{-1}\right\} \\ &\propto \det\left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} W_p^H\right) \\ &\times \exp\left\{-x_{t,p}^H W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} W_p^H x_{t,p}\right\} \quad (19) \end{aligned}$$

20

【0026】

よって式(5)により、過去の観測信号が与えられた下での観測信号 $y_{t, p}$ の条件付確率密度関数は、式(20)で与えられることが分かる。

【数 1 4】

$$\begin{aligned} p(y_{t,p} | y_{t-1,p}, \dots, y_{t-K_p,p}, \Theta) &= \\ &N\left\{y_{t,p}; \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p}, \left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} W_p^H\right)^{-1}\right\} \\ &\propto \det\left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} W_p^H\right) \\ &\times \exp\left\{-\left(y_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p}\right)^H W_{ps} \Lambda_t \left(\frac{2\pi p}{P}\right)^{-1} \right. \\ &\quad \left. \times W_p^H \left(y_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H y_{t-k,p}\right)\right\} \quad (20) \end{aligned}$$

40

【0027】

したがって、全ての観測信号 y についての確率密度関数は式(21)で表せる。

【0028】

50

【数 15】

$$\begin{aligned}
 p(\mathbf{y}|\Theta) &= \prod_{p=0}^{P-1} \prod_{t=0}^{T-1} p(\mathbf{y}_{t,p} | \mathbf{y}_{t-1,p}, \dots, \mathbf{y}_{t-K_p,p}, \Theta) \\
 &= \prod_{p=0}^{P-1} \prod_{t=0}^{T-1} \det \left(W_{ps} \Lambda_t \left(\frac{2\pi p}{P} \right)^{-1} W_p^H \right) \\
 &\quad \times \exp \left\{ - \left(\mathbf{y}_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H \mathbf{y}_{t-k,p} \right)^H W_{ps} \Lambda_t \left(\frac{2\pi p}{P} \right)^{-1} \right. \\
 &\quad \left. \times W_p^H \left(\mathbf{y}_{t,p} - \sum_{k=1}^{K_p} G_{k,p}^H \mathbf{y}_{t-k,p} \right) \right\} \quad (21)
 \end{aligned}$$

【0029】

式(21)の両辺の対数を取ることで対数尤度関数、式(9)が導かれる。式(9)の対数尤度関数は全極パラメータ s にも依存するため、パラメータ推定部12は、分離行列 w と回帰行列 g だけではなく、全てのパラメータ $\Theta = \{s, w, g\}$ を推定する。すなわち、式(9)を最大化する $\hat{\Theta} = \{s, w, g\}$ を未知パラメータの推定値として求める。式(9)を最大化する未知パラメータの推定値 $\hat{\Theta}$ の求め方には様々な方法を採用することが出来る。例えば、最急降下法や準ニュートン法等の任意の最適化法を用いる。これらの最適化法は周知であるので、その説明は省略する。パラメータ推定部12は、未知パラメータ $\hat{\Theta}$ の内の分離行列の推定値 w と、回帰行列の推定値 g とを強調信号計算部14に供給する(ステップS12)。

【0030】

強調信号計算部14を構成する残響除去部141は、帯域分割部60が出力する周波数帯域毎に分割された観測信号 $y_{t,p}$ に回帰行列の推定値 $G_{k,p}$ を畳み込み演算して残響成分を求め、観測信号 $y_{t,p}$ から残響成分を減算して残響除去信号 $x_{t,p}$ を生成する(式(7))(ステップS141)。音源分離部142は、残響除去信号 $x_{t,p}$ に分離行列の推定値 W_p を乗算(式(8))して音源信号の推定値を生成する(ステップS142)。帯域合成部66は、音源信号の推定値を合成して時間領域の強調音響信号を生成する(ステップS66)。このようにして生成された強調音響信号は、観測信号から残響成分を除去した後の残響除去信号を音源分離したものであるため、残響時間がフレーム幅を越える場合でも音源分離された音響信号になる。また、強調後の音響信号は残響成分が抑圧されているため、より高品質な強調音響信号を得ることが出来る。

【0031】

なお、残響を除去する所定フレーム数(残響除去フレーム数 K_p)を一つの固定値として説明したが、この値を周波数帯域の中心周波数に対応させて可変しても良い。周波数帯域によって残響時間が異なることは良く知られている。室内音響の分野においては、500Hz以下の周波数帯域の残響時間が長いので、その周波数帯域については所定フレーム数を大きく設定し、それ以外の周波数帯域の所定フレーム数を小さく設定する方法が考えられる。パラメータ推定部12内に所定フレーム数可変手段120を備え、所定フレーム数可変手段120が、周波数帯域に応じて所定フレーム数を可変して設定することで、残響を効率的に除去することが可能になる。つまり、計算量を削減する効果が得られる。

【実施例2】

【0032】

次に、最急降下法等の従来一般的な最適化アルゴリズムではない方法で、未知パラメータ Θ を推定する実施例2を説明する。図3に実施例2のパラメータ推定部30の機能構

成例を示す。その動作フローを図4に示す。パラメータ推定部30は、未知パラメータの推定に特化した効率的なアルゴリズムを用いて、分離行列の推定値 $w \wedge$ と、回帰行列の推定値 $g \wedge$ とを生成するものである。

【0033】

パラメータ推定部30は、観測信号記録部301と、パラメータ推定値初期化部302と、全極パラメータ推定値更新部303と、音源信号パワースペクトル推定値更新部304と、回帰行列推定値更新部305と、分離行列推定値更新部306と、収束判定部307とを備える。観測信号記録部301は、帯域分割部60で所定数の周波数帯域に分割された観測信号を記録する。観測信号記録部301は、観測区間中の全ての観測信号を記録すると共に記録した観測信号を、全極パラメータ推定値更新部303と回帰行列推定値更新部305と分離行列推定値更新部306とに時系列順に出力する。

10

【0034】

パラメータ推定値初期化部302は、回帰行列の推定値 $g \wedge$ と分離行列の推定値 $w \wedge$ の初期値を設定する(ステップS302)。また、繰り返し回数のインデックス i を0とする。回帰行列の推定値の初期値 $g \wedge^{(0)}$ は、零行列に設定する。分離行列の推定値 $w \wedge^{(0)}$ は、単位行列に設定する。これらは、残響と音源信号の混合がともに存在しない設定を初期値とすることを意味している。

【0035】

全極パラメータ推定値更新部303は、観測信号と、回帰行列の初期値 $g \wedge^{(0)}$ と分離行列の初期値 $w \wedge^{(0)}$ 又は更新された回帰行列の推定値 $g \wedge^{(i)}$ と分離行列の初期値 $w \wedge^{(i)}$ を入力として、全極パラメータ推定値 $s \wedge^{(i+1)}$ を式(22)に従って更新する(ステップS303)。

20

【数16】

$$s \hat{\Theta}^{(i+1)} = \arg \max_{s \Theta} L(\Theta; y) \Big|_{g \Theta = g \hat{\Theta}^{(i)}, w \Theta = w \hat{\Theta}^{(i)}} \quad (22)$$

【0036】

式(22)に示す全極パラメータの更新則について説明する。今、回帰行列の一時的な推定値 $g \wedge^{(i)}$ と分離行列の一時的な推定値 $w \wedge^{(i)}$ は既知である。その推定値を用いて式(7)と式(8)により音源信号ベクトル $s_{t,p}$ の推定値を計算する。

30

こうして得られる各音源信号の推定値を $S_{t,p} \wedge^{(m)(i+1)}$ と書く。m番目の音源信号のt番目のフレームにおける全極パラメータの更新された推定値 $a \wedge_{t,1}^{(m)(i+1)}, \dots, a \wedge_{t,Q}^{(m)(i+1)}, s \wedge_{t,p} \wedge^{(m)(i+1)}$ は、各音源信号の推定値 $S_{t,p} \wedge^{(m)(i+1)}$ に線形予測分析を適用することで求められる。

【0037】

全てのマイクロホン番号 $m(1 \leq m \leq M)$ 及びフレーム番号 $t(0 \leq t \leq T-1)$ に渡って $S_{t,p} \wedge^{(m)}$ に線形予測分析を適用することで式(23)に示す更新された全極パラメータ $s \wedge^{(i+1)}$ を得ることが出来る。

【数17】

$$s \hat{\Theta}^{(i+1)} = \left\{ \hat{a}_{t,1}^{(m)(i+1)}, \dots, \hat{a}_{t,Q}^{(m)(i+1)}, s \hat{v}_t^{(m)(i+1)} \right\}_{1 \leq m \leq M, 0 \leq t \leq T-1} \quad (23)$$

40

【0038】

更新された全極パラメータ $s \wedge^{(i+1)}$ は、音源信号パワースペクトル推定値更新部304に入力される。音源信号パワースペクトル推定値更新部304は、全極パラメータ $s \wedge^{(i+1)}$ を入力として音源信号のパワースペクトル密度の推定値 $s \wedge_{t,p} \wedge^{(i+1)}$ () を更新する(ステップS304)。音源信号のパワースペクトル密度の推定値 $s \wedge_{t,p} \wedge^{(i+1)}$ () は、入力された全極パラメータ $s \wedge^{(i+1)}$ から式(14)と式(15)と式(16)とを用いて更新される。すなわち、 $s \wedge_{t,p} \wedge^{(i+1)}$ () は、式(24)で計算される。

50

【数 18】

$${}_s \hat{\Lambda}^{(i+1)}(\omega) = \begin{bmatrix} {}_s \hat{\lambda}_t^{(1)(i+1)}(\omega) & & \\ & \ddots & \\ & & {}_s \hat{\lambda}_t^{(M)(i+1)}(\omega) \end{bmatrix} \quad (24)$$

$${}_s \hat{\lambda}_t^{(m)(i+1)}(\omega) = \frac{{}_s \hat{v}_t^{(m)(i+1)}}{\hat{A}_t^{(m)(i+1)}(e^{j\omega})} \quad (25)$$

10

$$\hat{A}_t^{(m)(i+1)}(z) = 1 - \hat{a}_{t,1}^{(m)(i+1)} z^{-1} - \dots - \hat{a}_{t,Q}^{(m)(i+1)} z^{-Q} \quad (26)$$

更新された音源信号のパワースペクトル密度の推定値 ${}_s \hat{\lambda}_t^{(i+1)}(\omega)$ は、回帰行列推定値更新部 305 に入力される。

【0039】

回帰行列推定値更新部 305 は、音源信号のパワースペクトル密度の推定値 ${}_s \hat{\lambda}_t^{(i+1)}(\omega)$ と、分離行列の初期値 ${}_w \hat{\Lambda}^{(0)}$ 又は更新された分離行列の推定値 ${}_w \hat{\Lambda}^{(i)}$ と、観測信号とを入力として回帰行列の推定値 ${}_g \hat{\Theta}^{(i+1)}$ を式(27)に従って更新する(ステップ S305)。

20

【数 19】

$${}_g \hat{\Theta}^{(i+1)} = \arg \max_{\Theta} L(\Theta; \mathbf{y}) \Big|_{{}_s \Theta = {}_s \hat{\Theta}^{(i+1)}, {}_w \Theta = {}_w \hat{\Theta}^{(i)}} \quad (27)$$

【0040】

式(27)に示す回帰行列の更新則について説明する。まず p 番目の周波数帯域における回帰行列の各成分を式(28)に示すように単一のベクトルにまとめる。

【数 20】

$$\mathbf{g}_p = \left[\mathbf{g}_{1,p}, \dots, \mathbf{g}_{K_p,p} \right]_{1 \times M^2 K_p} \quad (28)$$

30

$$\mathbf{g}_{k,p} = \left[\mathbf{g}_{k,p}^{(1)\tau}, \dots, \mathbf{g}_{k,p}^{(M)\tau} \right]_{1 \times M^2} \quad (29)$$

右下の添え字は行列の大きさを表す。ここで、 $\mathbf{g}_{k,p}^{(m)}$ は $\mathbf{G}_{k,p}$ の m 番目の列を表すものとする。以降、 \mathbf{g}_p を回帰行列の成分ベクトルと称する。成分ベクトル \mathbf{g}_p の全周波数帯域に渡る集合 $\{\mathbf{g}_p\}_{p=0}^{p=P-1}$ は回帰行列 ${}_g \hat{\Theta}$ に一致する。

【0041】

次に p 番目の周波数帯域における 1 フレーム前の観測信号行列 $Y_{t-1,p}$ を式(30)に示すように定義する。

40

【数 21】

$$Y_{t-1,p} = \left[\tilde{\mathbf{y}}_{t-1,p}, \dots, \tilde{\mathbf{y}}_{t-K_p,p} \right]_{M \times M^2 K_p} \quad (30)$$

$$\tilde{\mathbf{y}}_{t-k,p} = \begin{bmatrix} \mathbf{y}_{t-k,p}^\tau & & 0 \\ & \ddots & \\ 0 & & \mathbf{y}_{t-k,p}^\tau \end{bmatrix}_{M \times M^2} \quad (31)$$

50

【 0 0 4 2 】

これらを用いて、回帰行列の成分ベクトルの更新後の推定値 $g_p^{(i+1)}$ は式 (32) で計算出来る。

【 数 2 2 】

$$\begin{aligned} \left(\hat{g}_p^{(i+1)}\right)^H &= \left(\sum_{t=0}^{T-1} Y_{t-1,p}^H \hat{W}_p^{(i)} \hat{\Lambda}_t^{(i+1)} \left(\frac{2\pi p}{P}\right)^{-1} \left(\hat{W}_p^{(i)}\right)^H Y_{t-1,p} \right)^{-1} \\ &\times \left(\sum_{t=0}^{T-1} Y_{t-1,p}^H \hat{W}_p^{(i)} \hat{\Lambda}_t^{(i+1)} \left(\frac{2\pi p}{P}\right)^{-1} \left(\hat{W}_p^{(i)}\right)^H y_{t,p} \right) \end{aligned} \quad (32)$$

【 0 0 4 3 】

式 (30) を全ての周波数帯域番号 p ($0 \leq p \leq P-1$) に渡って計算することで回帰行列の推定値 $g_p^{(i+1)}$ を更新することが出来る。こうして求めた $g_p^{(i+1)}$ が式 (25) の解である。更新した回帰行列の推定値 $g_p^{(i+1)}$ は、分離行列推定値更新部 306 に入力される。

分離行列推定値更新部 306 は、回帰行列の推定値 $g_p^{(i+1)}$ と、音源信号のパワースペクトル密度の推定値と、観測信号とを入力として分離行列の推定値 $w_p^{(i+1)}$ を式 (4) に従って更新する (ステップ S306)。

【 0 0 4 4 】

【 数 2 3 】

$$w_p \hat{\Theta}^{(i+1)} = \arg \max_{w_p \Theta} L(\Theta; y) \Big|_{s \Theta = s \hat{\Theta}^{(i+1)}, g \Theta = g \hat{\Theta}^{(i+1)}} \quad (33)$$

式 (33) に示す分離行列の更新則について説明する。式 (34) は例えば準ニュートン法を用いて解くことが出来る。マイクロホンの数が 2 本 ($M=2$ 、ステレオ録音) の場合、より効率的な閉形式のアルゴリズムが存在する。ここでは、そのアルゴリズムを用いて分離行列の推定値 $w_p^{(i+1)}$ を更新する例を説明する。

【 0 0 4 5 】

まず、式 (7) で求められる残響除去信号 $x_{t,p}^{(i+1)}$ の共分散行列 R_p を、マイクロホン毎に式 (34) と式 (35) のように定義する。

【 数 2 4 】

$$R_p^{(1)} = \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{s \hat{\lambda}_t^{(1)(i+1)} (2\pi p / P)} \hat{x}_{t,p}^{(i+1)} \hat{x}_{t,p}^{(i+1)H} \quad (34)$$

$$R_p^{(2)} = \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{s \hat{\lambda}_t^{(2)(i+1)} (2\pi p / P)} \hat{x}_{t,p}^{(i+1)} \hat{x}_{t,p}^{(i+1)H} \quad (35)$$

【 0 0 4 6 】

この場合、分離行列の推定値 $w_p^{(i+1)}$ は次に示すアルゴリズムで求めることが出来る。

【 0 0 4 7 】

まず、式 (36) を満たす行列 $R_p^{(1)}$ の逆平方根 V_p を計算する。

【 数 2 5 】

$$V_p^H R_p^{(1)} V_p = I_2 \quad (36)$$

【 0 0 4 8 】

ここで、 I_2 は二次元単位行列を表す。次に式 (37) を満たすユニタリ行列 E_p と、

10

20

30

40

50

それに付随して定まる $d_p^{(1)}$ と $d_p^{(2)}$ を計算する。

【数 2 6】

$$E_p^H V_p^H R_p^{(2)} V_p E_p = \begin{bmatrix} d_p^{(1)} & 0 \\ 0 & d_p^{(2)} \end{bmatrix} \quad (37)$$

【0049】

E_p は、具体的には $V_p^H R_p^{(2)} V_p$ を固有値分解することで求められる。そして、 V_p 、 E_p 、 $d_p^{(2)}$ を用いて $\hat{W}_p^{(i+1)}$ を式 (38) で計算する。

【数 2 7】

$$\hat{W}_p^{(i+1)} = V_p E_p \begin{bmatrix} 1 & 0 \\ 0 & d_p^{(2)-\frac{1}{2}} \end{bmatrix} \quad (38)$$

式 (36) を全ての周波数帯域番号 p ($0 \leq p \leq P - 1$) に渡って計算することで、分離行列の推定値 $\hat{w}^{(i+1)}$ を得る。こうして求めた $\hat{w}^{(i+1)}$ が式 (31) の解である。

【0050】

また、以上説明したアルゴリズムの代わりに公知の音源分離技術を用いて分離行列の推定値 $\hat{w}^{(i+1)}$ を求めても良い。公知の音源分離技術としては、FastICA、ベル＝セイノフスキーのアルゴリズム、チコツキ＝ウンベハウエンのアルゴリズム、JADE 等が利用出来る。ここに上げた音源分離技術は、例えば参考文献「A.Hyvarinen, J.Karhunen, E. Oja, 根元幾 (訳)、川勝真喜 (訳) “ 詳解独立成分分析信号解析の新しい世界 ” 東京電機大学出版局、2005」に記載されている。

【0051】

以上のようにして更新された回帰行列の推定値 $\hat{g}^{(i+1)}$ と、分離行列の推定値 $\hat{w}^{(i+1)}$ とが、収束したか否かを収束判定部 307 が判定する (ステップ S307)。収束しているか否かを判定する方法としては、例えば繰り返し回数 i が所定数に到達していれば収束していると判定しても良いし、繰り返しによって得られる対数尤度関数 (式 (9)) の値の増分が、所定の閾値よりも小さければ収束していると判定しても良い。収束するまでステップ S303 ~ ステップ S306 の動作が繰り返され、収束した回帰行列の推定値 $\hat{g}^{(i+1)}$ と、分離行列の推定値 $\hat{w}^{(i+1)}$ とが、強調信号計算部 14 に出力される。それと同時に、パラメータ推定値記録部 308 に両パラメータを記録しても良い。

【0052】

〔実験結果〕

この発明の音響信号強調方法の効果を確認する目的で実験を行った。実験条件を説明する。音源信号は、ASJ-JNAS データベースから男女各 1 話者の音声抽出した。この音源信号を残響時間が約 0.6 秒の部屋で互いに異なる位置から再生し、2 個のマイクロホンで録音した。録音時の標本化周波数は 8 kHz とした。音響信号強調装置は、実施例 2 で説明した構成とした。帯域分割部には、ポリフェーズフィルタバンク分析を用いた。帯域分割数は 256、間引き率は 128 とした。

【0053】

音響信号の極の個数は $Q = 12$ とした。残響除去フレーム数 K_p は、1500 Hz 未満の周波数帯域では $K_p = 40$ 、それ以上の周波数帯域では周波数が高くなる程小さな値に設定した。収束判定は、繰り返し回数が 3 回で収束しているものとした。その結果を示す音響信号の波形を図 5 に示す。図 5 の縦軸は振幅、横軸は時間 (秒) である。図 5 (a) が一方のマイクロホンで録音した観測音響信号、図 5 (b) が従来の音源分離技術で強調された女性話者の音声、図 5 (c) が従来の音源分離技術で強調された男性話者の音声、図 5 (d) がこの発明で強調した女性話者の音声、図 5 (e) がこの発明で強調した男性

10

20

30

40

50

話者の音声、図5(f)が女性話者の音源信号、図5(g)が男性話者の音源信号である。

【0054】

従来の音源分離技術による強調音声と、この発明による強調音声とを比較すると、男女共にこの発明の音響信号強調方法の方が残響成分を抑圧して、より良好に音源分離を達成していることが見て取れる。例えば、男性話者は冒頭約1秒の間発話していない。しかし、従来技術は図5(c)の冒頭1秒間に現れているように女性話者の干渉音声を除去出来ていない。一方、この発明による男性話者の強調音声を示す図5(e)の同一の区間の女性の音声は、十分抑圧されている。このようにこの発明の音響信号強調方法は、従来技術よりも干渉音声を除去することが出来た。なお、図5に示した実験のSN比の平均改善率を比較すると従来技術の2.1dBに対してこの発明は4.9dBであった。このようにこの発明の音響信号強調装置によれば、高品質な強調音響信号を得ることが出来る。

10

【0055】

なお、この発明の技術思想に基づく音響信号強調装置とその方法は、上述の実施形態に限定されるものではなく、この発明の趣旨を逸脱しない範囲で適宜変更が可能である。上記した装置及び方法において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されるとしてもよい。

また、上記装置における処理手段をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、各装置における処理手段がコンピュータ上で実現される。

20

【0056】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD(Digital Versatile Disc)、DVD-RAM(Random Access Memory)、CD-ROM(Compact Disc Read Only Memory)、CD-R(Recordable)/RW(ReWritable)等を、光磁気記録媒体として、MO(Magneto Optical disc)等を、半導体メモリとしてフラッシュメモリ等を用いることができる。

30

【0057】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記録装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

また、各手段は、コンピュータ上で所定のプログラムを実行させることにより構成することにしてもよいし、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

40

【図面の簡単な説明】

【0058】

【図1】この発明の音響信号強調装置100の機能構成例を示す図。

【図2】音響信号強調装置100の動作フローを示す図。

【図3】パラメータ推定部30の機能構成例を示す図。

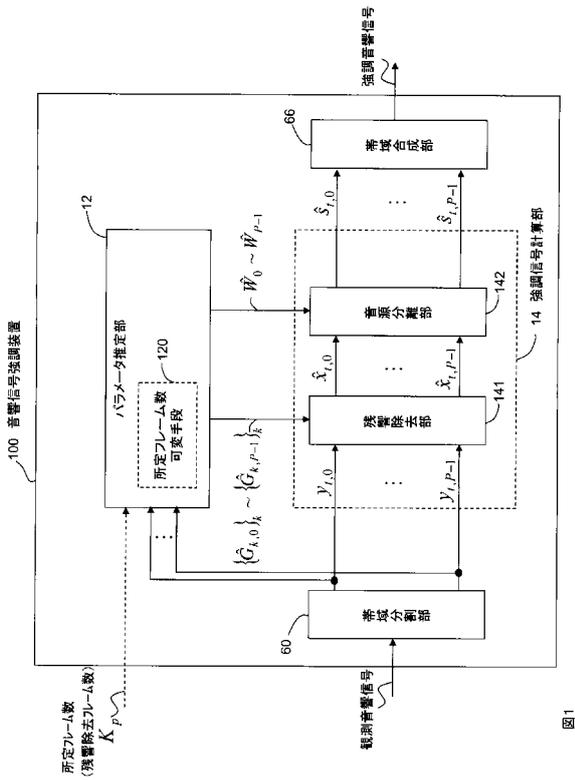
【図4】パラメータ推定部30の動作フローを示す図。

【図5】音響信号強調装置100を用いた音源分離実験の結果を示す図であり、(a)は観測音響信号を示す図、(b)は従来の音源分離技術で強調した女性話者の音声を示す図、(c)は従来の音源分離技術で強調した男性話者の音声を示す図、(d)はこの発明で強調した女性話者の音声を示す図、(e)はこの発明で強調した男性話者の音声を示す図、(f)は女性話者の音源信号を示す図、(g)は男性話者の音源信号を示す図である。

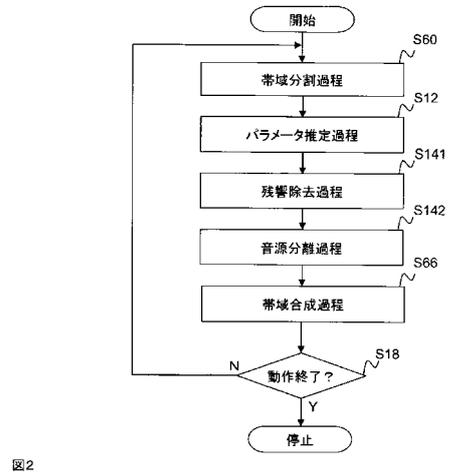
50

【図6】従来の音響信号強調装置600の機能構成例を示す図。

【図1】



【図2】



【 図 3 】

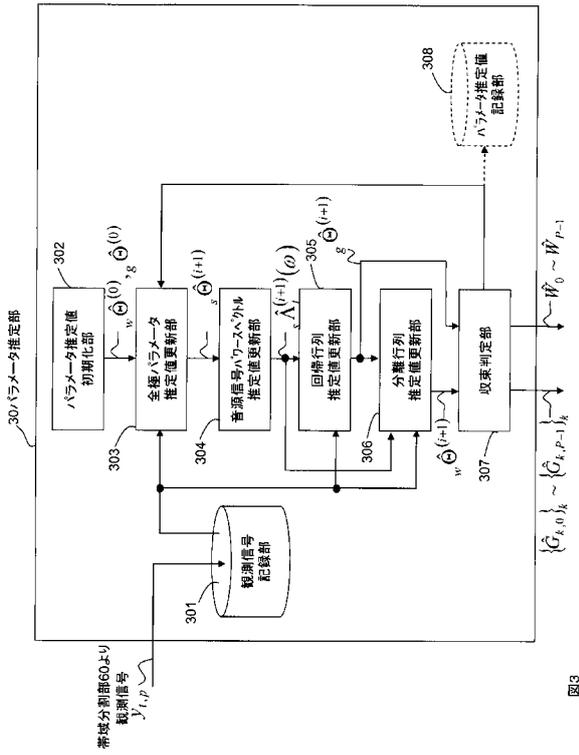


図3

【 図 4 】

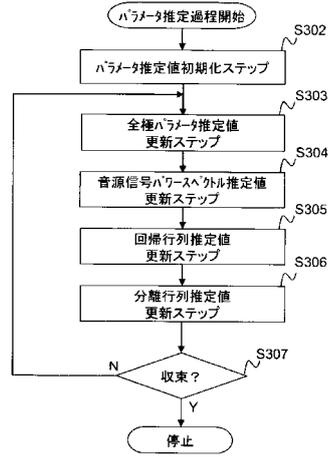
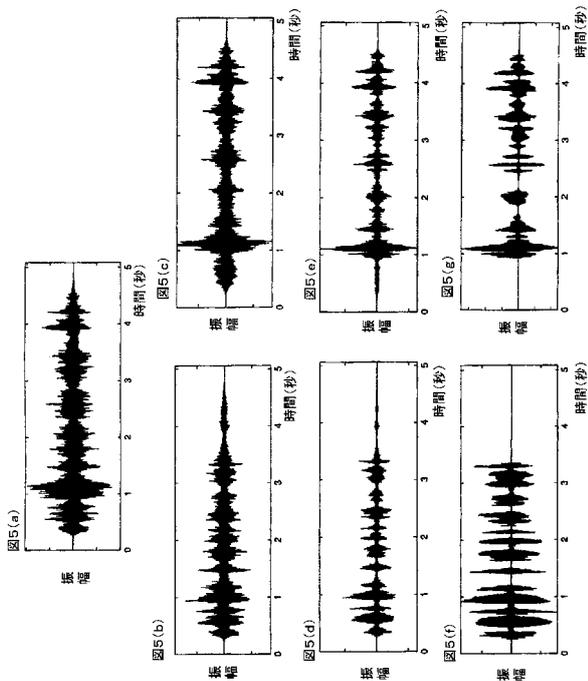


図4

【 図 5 】



【 図 6 】

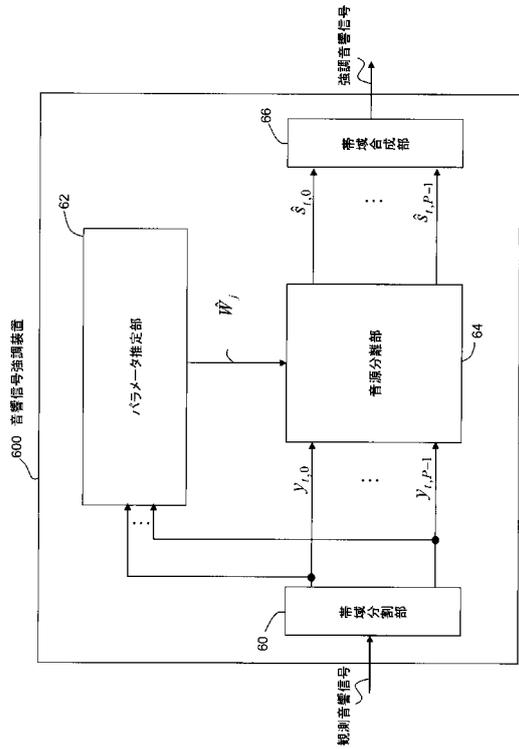


図6

フロントページの続き

(72)発明者 三好 正人

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

Fターム(参考) 5D020 CC05