

(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2008-35259

(P2008-35259A)

(43) 公開日 平成20年2月14日(2008.2.14)

| | | |
|-----------------------------|-----------------|-------------|
| (51) Int.Cl. | F I | テーマコード (参考) |
| H04R 3/00 (2006.01) | H04R 3/00 320 | 5D020 |
| G10L 21/02 (2006.01) | G10L 21/02 201D | |
| H04R 1/40 (2006.01) | G10L 21/02 202B | |
| | G10L 21/02 203Z | |
| | H04R 1/40 320Z | |

審査請求 未請求 請求項の数 10 OL (全 22 頁)

(21) 出願番号 特願2006-207006 (P2006-207006)
 (22) 出願日 平成18年7月28日 (2006.7.28)

(71) 出願人 000001199
 株式会社神戸製鋼所
 兵庫県神戸市中央区脇浜町二丁目10番26号
 (74) 代理人 100084135
 弁理士 本庄 武男
 (72) 発明者 稗方 孝之
 兵庫県神戸市西区高塚台1丁目5番5号
 株式会社神戸製鋼所神戸総合技術研究所内
 (72) 発明者 池田 陽平
 兵庫県神戸市西区高塚台1丁目5番5号
 株式会社神戸製鋼所神戸総合技術研究所内
 Fターム(参考) 5D020 BB04

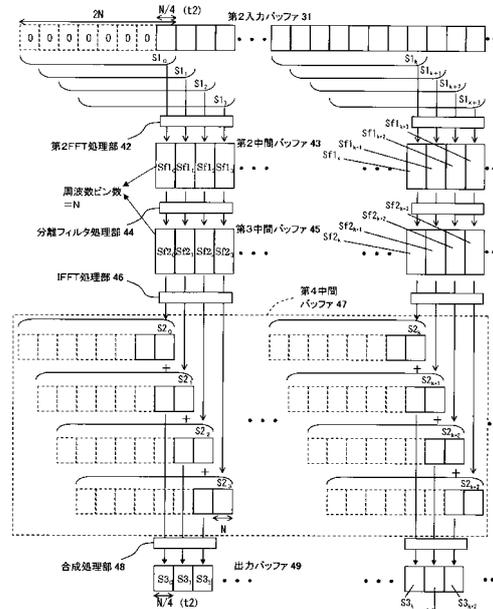
(54) 【発明の名称】 音源分離装置、音源分離方法及び音源分離プログラム

(57) 【要約】

【課題】ICA法による音源分離処理を行うに当たり、高い音源分離性能を確保しつつ、出力遅延（混合音声信号が発生した時点から、その混合音声信号から分離生成された分離信号が出力されるまでの遅延）を短くすることができること。

【解決手段】フィルタ処理の入力信号として用いる第2周波数領域信号S1を得るための第2フーリエ変換処理の実行周期t2の方が、分離行列の学習計算に用いる第1周波数領域信号を得るための第1フーリエ変換処理の実行周期t1よりも短くする。ここで、第2時間領域信号S1の時間長が、第1時間領域信号S0の時間長よりも短く設定されている場合、学習計算により得られる第1分離行列の行列要素を、複数のグループごとに集約することにより、フィルタ処理で用いる第2分離行列を設定する。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

複数の音源が存在する音響空間に複数のマイクロホンが存在する状態で、該マイクロホン各々を通じて入力される音響信号を一定のサンプリング周期で逐次デジタル化して得られる複数の混合音響信号に基づいて、1以上の前記音源に対応する音響信号である分離信号を分離生成する音源分離装置であって、

所定の第1の時間長分の新たな前記混合音響信号が得られるごとに、該第1の時間長以上の時間長分の最新の前記混合音声信号である第1の時間領域信号にフーリエ変換処理を施し、該フーリエ変換処理により得られる第1の周波数領域信号を所定の記憶手段に一時記憶させる第1のフーリエ変換手段と、

前記第1のフーリエ変換手段により得られた1又は複数の前記第1の周波数領域信号に基づいて周波数領域での独立成分分析法による学習計算を行うことにより、所定の第1の分離行列を算出する分離行列学習計算手段と、

前記分離行列学習計算手段により算出された前記第1の分離行列に基づいて前記分離信号の分離生成に用いる第2の分離行列を設定して更新する分離行列設定手段と、

前記第1の時間長よりも短い予め定められた第2の時間長分の新たな前記混合音響信号が得られるごとに、該第2の時間長の2倍の時間長分の最新の前記混合音声信号を含む第2の時間領域信号にフーリエ変換処理を施し、該フーリエ変換処理により得られる第2の周波数領域信号を所定の記憶手段に一時記憶させる第2のフーリエ変換手段と、

前記第2のフーリエ変換手段により新たな前記第2の周波数領域信号が得られるごとに、該第2の周波数領域信号に対して前記分離行列設定手段により更新される前記第2の分離行列に基づくフィルタ処理を施し、これにより得られる第3の周波数領域信号を所定の記憶手段に一時記憶させる分離フィルタ処理手段と、

前記分離フィルタ処理手段により新たな前記第3の周波数領域信号が得られるごとに、該第3の周波数領域信号に逆フーリエ変換処理を施し、該逆フーリエ変換処理により得られる第3の時間領域信号を所定の記憶手段に一時記憶させる逆フーリエ変換手段と、

前記逆フーリエ変換手段により新たな前記第3の時間領域信号が得られるごとに、該第3の時間領域信号とその1回前に得られた前記第3の時間領域信号とにおける時間帯が重複する部分の両信号を合成することによって新たな前記分離信号を生成する信号合成手段と、

を具備してなることを特徴とする音源分離装置。

【請求項 2】

前記第1の時間領域信号の時間長と前記第2の時間領域信号の時間長とが等しく設定されており、

前記分離行列設定手段が、前記第1の分離行列をそのまま前記第2の分離行列として設定してなる請求項1に記載の音源分離装置。

【請求項 3】

前記第2の時間領域信号の時間長が、前記第1の時間領域信号の時間長よりも短く設定されており、

前記分離行列設定手段が、前記第1の分離行列を構成する行列要素を複数のグループごとに集約することにより得られる行列を前記第2の分離行列として設定してなる請求項1に記載の音源分離装置。

【請求項 4】

前記第2の時間領域信号の時間長が、その2倍以上の整数倍が前記第1の時間領域信号の時間長となるように設定されてなる請求項3に記載の音源分離装置。

【請求項 5】

前記分離行列設定手段における前記集約が、前記第1の分離行列を構成する行列要素について、複数のグループごとに1つの行列要素を選択すること、又は複数のグループごとに行列要素の平均若しくは加重平均を算出することである請求項3又は4のいずれかに記載の音源分離装置。

10

20

30

40

50

【請求項 6】

前記第 2 の時間領域信号が、前記第 2 の時間長の 2 倍以上の予め定められた時間長分の最新の前記混合音声信号である請求項 1 ~ 5 のいずれかに記載の音源分離装置。

【請求項 7】

前記第 2 の時間領域信号が、前記第 2 の時間長の 2 倍の時間長分の最新の前記混合音声信号に所定の数の定数信号が付加された信号である請求項 1 ~ 5 のいずれかに記載の音源分離装置。

【請求項 8】

前記第 2 の時間領域信号が、前記第 2 の時間長の 2 倍の時間長分の最新の前記混合音声信号に所定の数の 0 値信号が付加された信号である請求項 7 に記載の音源分離装置。

10

【請求項 9】

複数の音源が存在する音響空間に複数のマイクロホンが存在する状態で、該マイクロホン各々を通じて入力される音響信号を一定のサンプリング周期で逐次デジタル化して得られる複数の混合音響信号に基づいて、1 以上の前記音源に対応する音響信号である分離信号を分離生成する処理を所定のプロセッサにより実行する音源分離方法であって、

所定の第 1 の時間長分の新たな前記混合音響信号が得られるごとに、該第 1 の時間長以上の時間長分の最新の前記混合音声信号である第 1 の時間領域信号にフーリエ変換処理を施し、該フーリエ変換処理により得られる第 1 の周波数領域信号を所定の記憶手段に一時記憶させる第 1 のフーリエ変換手順と、

前記第 1 のフーリエ変換手順により得られた 1 又は複数の前記第 1 の周波数領域信号に基づいて周波数領域での独立成分分析法による学習計算を行うことにより、所定の第 1 の分離行列を算出する分離行列学習計算手順と、

20

前記分離行列学習計算手順により算出された前記第 1 の分離行列に基づいて前記分離信号の分離生成に用いる第 2 の分離行列を設定して更新する分離行列設定手順と、

前記第 1 の時間長よりも短い予め定められた第 2 の時間長分の新たな前記混合音響信号が得られるごとに、該第 2 の時間長の 2 倍の時間長分の最新の前記混合音声信号を含む信号である第 2 の時間領域信号を設定する時間領域信号設定手順と、

前記時間領域信号設定手順により設定された前記第 2 の時間領域信号各々にフーリエ変換処理を施し、該フーリエ変換処理により得られる第 2 の周波数領域信号を所定の記憶手段に一時記憶させる第 2 のフーリエ変換手順と、

30

前記第 2 のフーリエ変換手順により新たな前記第 2 の周波数領域信号が得られるごとに、該第 2 の周波数領域信号に対して前記分離行列設定手順により更新される前記第 2 の分離行列に基づくフィルタ処理を施し、これにより得られる第 3 の周波数領域信号を所定の記憶手段に一時記憶させる分離フィルタ処理手順と、

前記分離フィルタ処理手順により新たな前記第 3 の周波数領域信号が得られるごとに、該第 3 の周波数領域信号に逆フーリエ変換処理を施し、該逆フーリエ変換処理により得られる第 3 の時間領域信号を所定の記憶手段に一時記憶させる逆フーリエ変換手順と、

前記逆フーリエ変換手順により新たな前記第 3 の時間領域信号が得られるごとに、該第 3 の時間領域信号とその 1 回前に得られた前記第 3 の時間領域信号とにおける時間帯が重複する部分の両信号を合成することによって新たな前記分離信号を生成する信号合成手順と、

40

を有してなることを特徴とする音源分離方法。

【請求項 10】

所定のプロセッサを、複数の音源が存在する音響空間に複数のマイクロホンが存在する状態で、該マイクロホン各々を通じて入力される音響信号を一定のサンプリング周期で逐次デジタル化して得られる複数の混合音響信号に基づいて、1 以上の前記音源に対応する音響信号である分離信号を分離生成する音源分離装置として機能させるための音源分離プログラムであって、

所定のプロセッサを、

所定の第 1 の時間長分の新たな前記混合音響信号が得られるごとに、該第 1 の時間長以

50

上の時間長分の最新の前記混合音声信号である第1の時間領域信号にフーリエ変換処理を施し、該フーリエ変換処理により得られる第1の周波数領域信号を所定の記憶手段に一時記憶させる第1のフーリエ変換手段と、

前記第1のフーリエ変換手段により得られた1又は複数の前記第1の周波数領域信号に基づいて周波数領域での独立成分分析法による学習計算を行うことにより、所定の第1の分離行列を算出する分離行列学習計算手段と、

前記分離行列学習計算手段により算出された前記第1の分離行列に基づいて前記分離信号の分離生成に用いる第2の分離行列を設定して更新する分離行列設定手段と、

前記第1の時間長よりも短い予め定められた第2の時間長分の新たな前記混合音響信号が得られるごとに、該第2の時間長の2倍の時間長分の最新の前記混合音声信号を含む第2の時間領域信号にフーリエ変換処理を施し、該フーリエ変換処理により得られる第2の周波数領域信号を所定の記憶手段に一時記憶させる第2のフーリエ変換手段と、

前記第2のフーリエ変換手段により新たな前記第2の周波数領域信号が得られるごとに、該第2の周波数領域信号に対して前記分離行列設定手段により更新される前記第2の分離行列に基づくフィルタ処理を施し、これにより得られる第3の周波数領域信号を所定の記憶手段に一時記憶させる分離フィルタ処理手段と、

前記分離フィルタ処理手段により新たな前記第3の周波数領域信号が得られるごとに、該第3の周波数領域信号に逆フーリエ変換処理を施し、該逆フーリエ変換処理により得られる第3の時間領域信号を所定の記憶手段に一時記憶させる逆フーリエ変換手段と、

前記逆フーリエ変換手段により新たな前記第3の時間領域信号が得られるごとに、該第3の時間領域信号とその1回前に得られた前記第3の時間領域信号とにおける時間帯が重複する部分の両信号を合成することによって新たな前記分離信号を生成する信号合成手段と、

の各手段として機能させるための音源分離プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、所定の音響空間に複数の音源と複数の音声入力手段とが存在する状態で、その音声入力手段各々を通じて逐次入力される複数の混合音声信号（音源各々からの音源信号が重畳された信号）に対し、所定の分離行列を用いた行列演算を施すことにより前記音源信号に対応する複数の分離信号を逐次生成する機能を備えた音源分離装置、音源分離方法及び音源分離プログラムに関するものである。

【背景技術】

【0002】

所定の音響空間に複数の音源と複数のマイクロホン（音声入力手段に相当）とが存在する場合、その複数のマイクロホンごとに、複数の音源各々からの個別の音声信号（以下、音源信号という）が重畳された音声信号（以下、混合音声信号という）が取得される。このようにして取得（入力）された複数の前記混合音声信号のみに基づいて、前記音源信号各々を同定（分離）する音源分離処理の方式は、ブラインド音源分離方式（Blind Source Separation方式、以下、BSS方式という）と呼ばれる。なお、本明細書において、「音声」は、人が発する声に限らず各種の音響を含む概念を表す用語として用いている。従って、例えば、音響入力手段と音声入力手段とは同義であり、また、混合音響信号と混合音声信号とは同義である。また、本明細書において、「演算」、「計算」及び「算出」は同義である。

さらに、BSS方式の音源分離処理の1つに、独立成分分析法（Independent Component Analysis、以下、ICA法という）による音源分離処理がある。

複数のマイクロホンを通じて入力される複数の前記混合音声信号（時系列の（時間領域の）音声信号）に含まれる音源信号それぞれは、統計的に独立である。ICA法による音源分離処理は、各音源信号が統計的に独立であることを前提に、入力された複数の前記混合音声信号に基づいて、所定の分離行列（逆混合行列）を学習計算により最適化する処理

10

20

30

40

50

を有する。さらに、ICA法による音源分離処理は、学習計算により最適化された分離行列を用いて、入力された複数の混合音声信号に対してフィルタ処理（行列演算）を行うことを含み、これによって前記音源信号が同定（音源分離）される。

ここで、ICA法における分離行列の最適化は、所定の時間長分の混合音声信号に対し、分離行列を用いたフィルタ処理（行列演算）を行うことによる分離信号（同定された信号）の算出と、その分離信号を用いた逆行列演算等による分離行列の更新と、を逐次繰り返す学習計算により行われる。

このようなICA法による音源分離処理は、例えば、非特許文献1や非特許文献2等に詳説されている。

【0003】

BSS方式の音源分離処理を行うためのICA法は、時間領域(Time-Domain)でのICA法（以下、TDICA法という）と、周波数領域(Frequency-Domain)でのICA法（以下、FDICA法という）とに大別される。

TDICA法は、一般に、広い周波数帯域において音源信号それぞれの独立性を評価できる手法であり、分離行列の学習計算において、最適点近傍における収束性が高い。このため、TDICA法によれば、最適化レベルの高い分離行列を得ることができ、高精度で音源信号を分離できる（分離性能が高い）。しかしながら、TDICA法は、分離行列の学習計算に非常に複雑な（演算負荷の高い）処理（畳み込み混合に対する処理）を必要とするため、リアルタイム処理には適さない。

一方、例えば特許文献1等に示されるFDICA法は、フーリエ変換処理によって混合音声信号を時間領域の信号から周波数領域の信号に変換することにより、畳み込み混合の問題を、複数の分割された周波数帯域である周波数ビン（特許文献1ではサブバンド）ごとの瞬時混合の問題に変換した上で、分離行列の学習計算を行う手法である。このFDICA法によれば、分離行列（分離フィルタ処理に用いる行列）の最適化（学習計算）を、安定かつ高速に行うことができる。従って、FDICA法は、リアルタイム音源分離処理に適している。

【0004】

以下、図7を参照しつつ、FDICA法による分離行列の学習計算について説明する。図7はFDICA法による分離行列の学習計算を行う学習計算ユニットZ1の概略構成を表すブロック図である。

図7には、2つの音源1、2からの音源信号 $S_1(t)$ 、 $S_2(t)$ を2つのマイクロホン111、112を通じて入力した2チャンネル（各チャンネルは、マイクロホンそれぞれに対応）の混合音声信号 $x_1(t)$ 、 $x_2(t)$ に基づいて、分離行列 $W(f)$ の学習計算を行う例について示しているが、2チャンネル以上であっても同様である。なお、混合音声信号 $x_1(t)$ 、 $x_2(t)$ は、A/D変換器によって一定のサンプリング周期（一定のサンプリング周波数といってもよい）でデジタル化された信号であるが、図7において、A/D変換器の記載を省略している。

FDICA法では、まず、FFT処理部13が、入力された混合音声信号 $x(t)$ が所定の周期（所定のサンプル数）ごとに区分された信号であるフレームそれぞれについて、フーリエ変換処理を行う。これにより、混合音声信号（入力信号）が、時間領域の信号から周波数領域の信号へ変換される。フーリエ変換後の信号は、周波数ビンと呼ばれる所定範囲の周波数帯域ごとに区分された信号となる。そして、そのフーリエ変換処理後の各チャンネルの信号について、分離フィルタ処理部11fが、分離行列 $W(f)$ に基づくフィルタ処理（行列演算処理）を施すことによって音源分離（音源信号の同定）を行う。ここで f を周波数ビン、 m を分析フレーム番号とすると、分離信号（同定信号） $y(f, m)$ は、次の(1)式のように表すことができる。

【数1】

$$Y(f, m) = W(f) \cdot X(f, m) \quad \dots(1)$$

そして、(1)式における分離フィルタ（分離行列） $W(f)$ は、不図示のプロセッサ（

10

20

30

40

50

例えば、コンピュータが備えるCPUが、次の(2)式により表される処理(以下、単位処理という)を繰り返す逐次計算(学習計算)を実行することによって求められる。ここで、前記単位処理を実行する際、前記プロセッサは、まず、前回(i)の出力 $y(f)$ を(2)式に適用することによって今回(i+1)の $W(f)$ を求める。ここで、分離行列 $W(f)$ は、周波数ビンそれぞれに対応するフィルタ係数を行列要素とする行列であり、前記学習計算は、そのフィルタ係数各々の値を算出する計算である。

さらに、前記プロセッサは、今回求めた $W(f)$ を用いて所定時間長分の混合音声信号(周波数領域の信号)に対してフィルタ処理(行列演算)を施すことによって今回(i+1)の出力 $y(f)$ を求める。そして、前記プロセッサが、これら一連の処理(前記単位処理)を複数回繰り返すことにより、分離行列 $W(f)$ は、徐々に上記逐次計算(学習計算)で用いられる混合音声信号に適合した内容となる。

【数2】

$$\mathbf{W}_{(ICA)}^{[i+1]}(f) = \mathbf{W}_{(ICA)}^{[i]}(f) - \eta(f) \left[\text{off-diag} \left\{ \phi \left(\mathbf{Y}_{(ICA)}^{[i]}(f, m) \mathbf{Y}_{(ICA)}^{[i]}(f, m)^H \right) \right\}_m \right] \mathbf{W}_{(ICA)}^{[i]}(f) \quad \dots(2)$$

但し、 $\eta(f)$ は更新係数、 i は更新回数、 $\langle \dots \rangle$ は時間平均、 H はエルミート転置を表す。

off-diag X は行列Xの全ての対角要素をゼロに置き換える演算処理を表す。

$\phi(\dots)$ はシグモイド関数等を要素にもつ適当な非線形ベクトル関数を表す。

【0005】

ところで、FDICA法では、分離行列の学習計算に用いる周波数領域の混合音声信号(以下、学習用入力信号という)における周波数ビンの数(特許文献1に示されるサブバンドの数)が、その学習計算により得られる分離行列を用いてフィルタ処理を行った場合の分離性能に大きな影響を及ぼす。ここで、フーリエ変換処理において、出力信号(周波数領域の信号)の周波数ビン数は、入力信号(時間領域の信号)のサンプル数の1/2倍なるため、フーリエ変換処理の入力となる混合音声信号(デジタル信号)のサンプル数が、分離性能に大きな影響を及ぼすともいえる。また、混合音声信号をA/D変換する際のサンプリング周期は一定であるので、フーリエ変換処理の入力となる混合音声信号の時間長が、分離性能に大きな影響を及ぼすといってもよい。

特許文献1および非特許文献3には、例えば、混合音声信号のサンプリング周波数が8kHzである場合、フーリエ変換処理の入力信号(時間領域の信号)の長さ(フレーム長)を1024サンプル程度(時間に換算して128ms)とすれば、即ち、フーリエ変換処理の出力信号(周波数領域の信号)における周波数ビンの数(サブバンド数)が512程度となるようにすれば、高い分離性能が得られる(分離性能の高い分離行列が得られる)ことが示されている。

【0006】

次に、図8を参照しつつ、FDICA法による音源分離処理をリアルタイムで実行する場合の従来の処理手順について説明する。図8は従来のFDICA法による音源分離処理の流れを表すブロック図である。

図8に示す例では、FDICA法による音源分離処理は、学習演算部34、第2FFT処理部42'、分離フィルタ処理部44'、IFFT処理部46'及び合成処理部48'により実行される。これら学習演算部34、第2FFT処理部42'、分離フィルタ処理部44'、IFFT処理部46'、合成処理部48'は、例えば、DSP(Digital Signal Processor)等の演算用のプロセッサ及びそのプロセッサにより実行されるプログラムが記憶されたROM等の記憶手段、並びにRAM等のその他の周辺装置により構成される。

また、図8に示す各バッファ(第1入力バッファ31、第1中間バッファ33、第2入力バッファ41'、第2中間バッファ43'、第3中間バッファ45'、第4中間バッファ47'、出力バッファ49')は、説明の便宜上、非常に多くのデータを蓄積可能であるかのように記載されている。しかしながら、実際には、各バッファは、記憶するデータ

10

20

30

40

50

のうち不要になったものが順次消去され、これにより生じる空き領域が再利用されるので、その記憶容量は必要十分な量に設定されている。

【0007】

一定のサンプリング周期でデジタル化された各チャンネルの混合音声信号（音響信号）は、Nサンプル分ずつ第1入力バッファ31と第2入力バッファ41'とに入力（伝送）される。例えば、混合音声信号のサンプリング周波数が8kHzである場合、 $N = 512$ 程度とする。この場合、Nサンプル分の混合音声信号の時間長は64msである。

そして、第1FFT処理部32は、新たなNサンプル分の混合音声信号が第1入力バッファ31に入力されるごとに、それを含む最新の2Nサンプル分の混合音声信号（以下、第1時間領域信号S0という）について、フーリエ変換処理を実行し、その処理結果である周波数領域の信号（以下、第1周波数領域信号Sf0という）を、第1中間バッファ33に一時記憶させる。ここで、第1入力バッファ31に蓄積された信号のサンプル数が、2N個に満たない場合（処理開始後の初期の段階）には、不足する数だけ0値が充当された信号について、フーリエ変換処理が実行される。第1FFT処理部32の1回のフーリエ変換処理により得られる第1周波数領域信号Sf0の周波数ビンの数は、第1周波数領域信号Sf0のサンプル数の1/2倍（=N個）である。

また、学習演算部34は、第1中間バッファ33に、所定時間長T[sec]分の第1周波数領域信号Sf0が記録されるごとに、そのT[sec]分の信号Sf0に基づいて、分離行列 $W(f)$ の学習計算、即ち、分離行列 $W(f)$ を構成するフィルタ係数（行列要素）の学習計算を行う。さらに、学習演算部34は、所定のタイミングで、分離フィルタ処理部44'で用いられる分離行列を、学習後の分離行列に更新する（即ち、分離行列のフィルタ係数の値を学習後の値に更新する）。通常、学習演算部34は、学習計算の終了後、分離フィルタ処理部44'のフィルタ処理が最初に終了した直後に、分離行列を更新する。

【0008】

一方、第2FFT処理部42'も、新たなNサンプル分の混合音声信号が第2入力バッファ41'に入力されるごとに、それを含む最新の2Nサンプル分の混合音声信号（以下、第2時間領域信号S1という）について、フーリエ変換処理を実行し、その処理結果である周波数領域の信号（以下、第2周波数領域信号Sf1という）を、第2中間バッファ43'に一時記憶させる。このように、第2FFT処理部42'は、順次Nサンプル分ずつ時間帯が重複する第2時間領域信号S1（混合音声信号）について、フーリエ変換処理を実行する。ここで、第2入力バッファ41'に蓄積された信号のサンプル数が、2N個に満たない場合（処理開始後の初期の段階）には、不足する数だけ0値が充当された信号について、フーリエ変換処理が実行される。なお、この第2周波数領域信号Sf1の周波数ビンの数も、第2周波数領域信号Sf1のサンプル数の1/2倍（=N個）である。

また、分離フィルタ処理部44'は、第2中間バッファ43'に、新たな第2周波数領域信号Sf1が記録されるごとに、その新たな第2周波数領域信号Sf1について、分離行列を用いたフィルタ処理（行列演算）を行い、その処理により得られる信号（以下、第3周波数領域信号Sf2という）を、第3中間バッファ45'に一時記憶させる。このフィルタ処理に用いられる分離行列は、前述した学習演算部34によって更新されるものである。なお、学習演算部34により最初に分離行列が更新されるまでは、分離フィルタ処理部44'は、予め定められた初期値が設定された分離行列（初期行列）を用いてフィルタ処理を行う。ここで、第2周波数領域信号Sf1と第3周波数領域信号Sf2とは、周波数ビンの数が等しいことはいうまでもない。

【0009】

また、IFFT処理部46'は、第3中間バッファ45'に新たな第3周波数領域信号Sf2が記録されるごとに、その新たな第3周波数領域信号Sf2について、逆フーリエ変換処理を実行し、その処理結果である時間領域の信号（以下、第3時間領域信号S2という）を、第4中間バッファ47'に一時記憶させる。この第3時間領域信号S2のサンプル数は、第3周波数領域信号Sf2の周波数ビンの数（=N）の2倍（=2N）である。前述したように、第2FFT処理部42'が、Nサンプル分ずつ時間帯が重複する第2

10

20

30

40

50

時間領域信号 S 1 (混合音声信号) についてフーリエ変換処理を実行するので、第 4 中間バッファ 4 7' に記録される連続する 2 つの第 3 時間領域信号 S 2 も、相互に N サンプル分だけ時間帯が重複 (オーバーラップ) している。

また、合成処理部 4 8' は、第 4 中間バッファ 4 7' に新たな第 3 時間領域信号 S 2 が記録されるごとに、以下に示す合成処理を実行することによって新たな分離信号 S 3 を生成して出力バッファ 4 9' に一時記憶させる。

ここで、前記合成処理は、I F F T 処理部 4 6' によって得られた新たな第 3 時間領域信号 S 2 とその 1 回前に得られた第 3 時間領域信号 S 2 とについて、それらにおける時間帯が重複する部分の両信号 (それぞれ N サンプル分の信号) を、例えばクロスフェードの重み付けをして加算すること等によって合成する処理である。これにより、平滑化された分離信号 S 3 が得られる。

以上の処理により、混合音声信号に対していくらかの遅延 (時間遅れ) が生じるものの、音源に対応する分離信号 S 3 が、リアルタイムで出力バッファ 4 9' に記録される。

また、フィルタ処理に用いられる分離行列が、学習演算部 3 4 によって音響環境の変化に適合したものに適宜更新される。

【 0 0 1 0 】

次に、図 9 を参照しつつ、図 8 に示した従来の音源分離処理によって生じる出力遅延について説明する。図 9 は従来の F D I C A 法による音源分離処理における信号入出力の状態遷移を表すブロック図である。

ここで、出力遅延とは、混合音声信号が発生した時点から、その混合音声信号から分離生成された分離信号が出力されるまでの遅延を指す。

以下、A / D 変換処理によって得られる混合音声信号 (デジタル信号) を一時記憶するバッファを入力バッファ 2 3 と称する。この入力バッファ 2 3 から、N サンプル分の混合音声信号が、前記第 1 入力バッファ 3 1 及び第 2 入力バッファ 4 1' に転送される。また、図 9 において、入力ポイント P t 1 は、入力バッファ 2 3 に対する信号の書き込み位置 (書き込みポインタの指示位置) を表し、出力ポイント P t 2 は、出力バッファ 4 9' からの信号の読み出し位置 (読み出しポインタの指示位置) を表す。これら入力ポイント P t 1 及び出力ポイント P t 2 は、混合音声信号のサンプリング周期と同じ周期で同期して順次移動する。また、これら入力ポイント P t 1 及び出力ポイント P t 2 は、2 N サンプル分の記憶容量を有する入力バッファ 2 3 及び出力バッファ 4 9' 各々において巡回移動する。

【 0 0 1 1 】

図 9 (a) は、処理開始時の状態を表す。入力バッファ 2 3 及び出力バッファ 4 9' のいずれにも信号は蓄積されていない (例えば、0 値が埋められた状態) である。

図 9 (b) は、図 9 (a) の状態の後、入力バッファ 2 3 に、入力ポイント P t 1 の移動に従って順次新たな信号が書き込まれ、N サンプル分の信号が蓄積した時点の状態を表す。この時点で、N サンプル分の信号 (図中、入力 (1) と記した信号) が、音源分離処理を行う部分 (以下、音源分離処理部 A という) に転送され、音源分離処理が実行される。具体的には、N サンプル分の信号が、前記第 1 入力バッファ 3 1 及び第 2 入力バッファ 4 1' に転送 (記録) され、図 8 に基づき説明した音源分離処理が実行される。また、入力バッファ 2 3 において、音源分離処理部 A への転送が終了した信号は消去される。

図 9 (c) は、図 9 (b) の状態の後、音源分離処理部 A によって N サンプル分の分離信号 (図中、出力 (1) と記した信号) が生成され、その分離信号が出力バッファ 4 9' に書き込まれた時点の状態を表す。この分離信号 (出力 (1)) は、図 8 における分離信号 S 3 に相当するものである。

この図 9 (c) の状態では、出力ポイント P t 2 は、分離信号が書き込まれていない位置にあるので、分離信号 (出力 (1)) は未だ出力されない。

図 9 (d) は、図 9 (c) の状態の後、入力バッファ 2 3 に、さらに新たな信号が書き込まれ、次の N サンプル分の信号 (図中、入力 (2) と記した信号) が蓄積した時点の状態を表す。この時点で、新たな N サンプル分の信号 (入力 (2)) が、前記音源分離処理

10

20

30

40

50

部 A に転送され、音源分離処理が実行される。

この図 9 (d) の状態において、出力ポイント P t 2 が、前回の分離信号 (出力 (1)) の書き込み位置の先頭にあるので、分離信号 (出力 (1)) の出力が始まる。

図 9 (e) は、図 9 (d) の状態の後、音源分離処理部 A によって新たな N サンプル分の分離信号 (図中、出力 (2) と記した信号) が生成され、その分離信号が出力バッファ 4 9 ' に書き込まれた時点の状態を表す。図 9 (d) の時点から図 9 (e) の時点までの間、出力ポイント P t 2 の移動に従って、前回の分離信号 (出力 (1)) が 1 サンプルずつ順次出力される。また、出力バッファ 4 9 ' において、出力が終了した信号は消去される。

【特許文献 1】特開 2 0 0 3 - 2 7 1 1 6 8 号公報

【非特許文献 1】猿渡洋、「アレー信号処理を用いたブラインド音源分離の基礎」電子情報通信学会技術報告、vol.EA2001-7、pp.49-56、April 2001.

【非特許文献 2】高谷智哉他、「SIMOモデルに基づくICAを用いた高忠実度なブラインド音源分離」電子情報通信学会技術報告、vol.US2002-87、EA2002-108、January 2003.

【非特許文献 3】猿渡洋、「音声・音響信号を対象としたブラインド音源分離」電子情報通信学会DSP研究会、DSP2001-194、pp.59-66、March 2002.

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 2 】

図 9 からわかるように、従来の音源分離処理では、音源分離処理部 A の前段及び後段における信号の受け渡しに関して、図 9 (a) の時点から図 9 (d) の時点までの間において、 $2N$ サンプル分の信号の時間長に相当する出力遅延が生じる。さらに、音源分離処理部 A 内においても、前記合成処理部 4 8 ' による合成処理により、 N サンプル分の信号の時間長に相当する出力遅延が生じる。従って、従来の音源分離処理では、全体として $3N$ サンプル分の信号の時間長に相当する出力遅延が生じるという問題点があった。

例えば、信号のサンプリング周波数が 8 kHz である場合、FDICA 法により分離性能の高い分離行列が得られるように 1 フレームを 1 0 2 4 サンプルの信号 (即ち、 $N = 5 1 2$) とすると、1 9 2 [msec] の出力遅延が生じる。

この 1 9 2 [msec] という出力遅延は、リアルタイムで動作する装置において許容し難い遅延である。例えば、デジタル携帯電話における通信の遅延時間は、一般に 5 0 [msec] 以下である。このデジタル携帯電話に従来の FDICA 法による音源分離を適用した場合、トータルの遅延時間が 2 4 2 [msec] となり、実用に耐えない。同様に、補聴器に従来の FDICA 法による音源分離を適用した場合も、利用者の目に映る映像と、補聴器を通じて聞こえる音声との時間のズレが大き過ぎて実用に耐えない。

ここで、入力ポイント P t 1 と出力ポイント P t 2 との位置関係を、予め図 9 に示した位置関係とは異なる位置関係に設定することにより、出力遅延は、 $3N$ サンプル分の信号の時間長以下にすることは可能である。しかしながら、その場合でも、出力遅延を、 $2N$ サンプル分の信号の時間長に、音源分離処理に要する時間を加算した時間にまで短縮できるに過ぎない。即ち、FDICA 法による音源分離処理では、出力遅延の時間が、フィルタ処理の入力信号として用いる周波数領域の信号 S f 1 を得るためのフーリエ変換処理 (前記第 2 FFT 処理部 4 2 ' の処理) の実行周期 (N サンプルの信号の時間長 t_N) の 2 倍強から 3 倍程度の時間となる。

一方、出力遅延の時間は、1 フレームの長さを短く (サンプル数を小さく) すれば短縮できる。しかしながら、1 フレームの長さを短くすることは、音源分離性能の悪化につながるという問題が生じる。

本発明は上記事情に鑑みてなされたものである。本発明の目的とするところは、ICA 法による音源分離処理を行うに当たり、高い音源分離性能を確保しつつ、出力遅延 (混合音声信号が発生した時点から、その混合音声信号から分離生成された分離信号が出力されるまでの遅延) を短くすることができる音源分離装置、音源分離方法及び音源分離プログラムを提供することにある。

10

20

30

40

50

【課題を解決するための手段】

【0013】

本発明は、それぞれ複数の音源からの信号が重畳された複数の（複数チャンネルの）混合音響信号に基づいて、1以上の前記音源に対応する音響信号である分離信号を分離生成する音源分離装置に適用されるものである。ここで、各混合音声信号は、複数の音源が存在する音響空間に複数のマイクロホンが存在する状態で、そのマイクロホン各々を通じて入力される音響信号を一定のサンプリング周期で逐次デジタル化することにより得られる信号（デジタル信号）である。

そして、上記目的を達成するために本発明は、次の(1)～(7)に示す構成要素を備える。

(1) 所定の第1の時間 t_1 の長さ分の新たな前記混合音響信号が得られるごとに、その第1の時間 t_1 以上の長さ分の最新の前記混合音声信号（以下、第1の時間領域信号という）にフーリエ変換処理を施し、そのフーリエ変換処理により得られる信号（以下、第1の周波数領域信号という）を所定の記憶手段に一時記憶させる手段（以下、第1のフーリエ変換手段という）。なお、前記第1の時間領域信号及び前記第1のフーリエ変換手段は、それぞれ図8における信号 S_0 及び第1FFT処理部32に相当するものである。

(2) 前記第1のフーリエ変換手段により得られた1又は複数の前記第1の周波数領域信号に基づいて、周波数領域での独立成分分析法（FDICA法）による学習計算を行うことにより、所定の分離行列（以下、第1の分離行列という）を算出する手段（以下、分離行列学習計算手段という）。なお、この分離行列学習計算手段は、図8における学習演算部34に相当するものである。

(3) 前記分離行列学習計算手段により算出された前記第1の分離行列に基づいて、前記分離信号の分離生成（即ち、フィルタ処理）に用いる行列（以下、第2の分離行列という）を設定して更新する手段（以下、分離行列設定手段という）。

(4) 前記第1の時間 t_1 の長さよりも短い予め定められた第2の時間 t_2 の長さ分の新たな前記混合音響信号が得られるごとに、その第2の時間 t_2 の2倍の長さ分の最新の前記混合音声信号を含む信号（以下、第2の時間領域信号という）にフーリエ変換処理を施し、そのフーリエ変換処理により得られる信号（以下、第2の周波数領域信号という）を、所定の記憶手段に一時記憶させる手段（以下、第2のフーリエ変換手段という）。なお、前記第2の時間領域信号及び前記第2のフーリエ変換手段は、それぞれ図8における信号 S_1 及び第2FFT処理部42'に相当するものである。

(5) 前記第2のフーリエ変換手段により新たな前記第2の周波数領域信号が得られるごとに、その新たな第2の周波数領域信号に対し、前記分離行列設定手段により更新される前記第2の分離行列に基づくフィルタ処理を施し、これにより得られる信号（以下、第3の周波数領域信号という）を所定の記憶手段に一時記憶させる手段（以下、分離フィルタ処理手段という）。

(6) 前記分離フィルタ処理手段により新たな前記第3の周波数領域信号が得られるごとに、その新たな第3の周波数領域信号に逆フーリエ変換処理を施し、その逆フーリエ変換処理により得られる信号（以下、第3の時間領域信号という）を所定の記憶手段に一時記憶させる手段（以下、逆フーリエ変換手段という）。なお、この逆フーリエ変換手段は、図8におけるIFFT処理部46'に相当するものである。

(7) 前記逆フーリエ変換手段により新たな前記第3の時間領域信号が得られるごとに、その新たな第3の時間領域信号と、その1回前に得られた前記第3の時間領域信号とにおける時間帯が重複する部分の両信号を合成することにより、新たな前記分離信号を生成する手段（以下、信号合成手段とという）。なお、この信号合成手段は、図8における合成処理部48'に相当するものである。

ここで、上記(1)～(7)において、信号の「時間の長さ」とその長短とにより特定する記載を、信号の「サンプル数」とその大小により特定する記載に置き換えた場合、その置き換えを行う前後の記載内容は、同じ意味を表す。

【0014】

10

20

30

40

50

前述したように、FDICA法による音源分離処理では、出力遅延の時間は、フィルタ処理の入力信号として用いる周波数領域の信号（前述した信号 S_f1 ）を得るためのフーリエ変換処理の実行周期の2倍強から3倍程度の時間となる。

これに対し、本発明に係る音源分離装置では、フィルタ処理の入力信号として用いる前記第2の周波数領域信号を得るためのフーリエ変換（前記第2のフーリエ変換手段の処理）の実行周期（前記第2の時間 t_2 ）の方が、分離行列の学習計算に用いる周波数領域の信号を得るためのフーリエ変換（前記第1のフーリエ変換手段の処理）の実行周期（前記第1の時間 t_1 ）よりも短い。従って、前記第2の時間 t_2 を従来よりも十分に短く設定すること（図9におけるサンプル数 N を小さく設定することと同じ）により、出力遅延の時間を従来よりも大幅に短縮できる。

一方、分離行列の学習計算に対応するフーリエ変換処理（前記第1のフーリエ変換手段の処理）の実行周期（前記第1の時間 t_1 ）は、前記第2の時間 t_2 に関わらず、十分長い時間（例えば、8kHzのサンプリング周期 $\times 1024$ サンプルの信号の長さ相当）に設定できる。これにより、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

【0015】

ところで、前述したように、フーリエ変換処理では、出力信号（周波数領域信号）の周波数ビンの数は、入力信号（時間領域信号）のサンプル数の $1/2$ 倍となる。また、FDICA法による学習計算により得られる分離行列の行列要素（即ち、フィルタ係数）の数は、その学習計算で用いる前記第1周波数領域信号における周波数ビンの数と同じである。さらに、フィルタ処理の入力信号（前記第1周波数領域信号）における周波数ビンの数と、フィルタ処理に用いる分離行列の行列要素の数（フィルタ係数の数）とは、一致していなければならない。

ここで、前記第1の時間領域信号の時間長と前記第2の時間領域信号の時間長とが等しく設定されていれば（即ち、両信号のサンプル数が同じであれば）、前記第1フーリエ変換手段の処理により得られる信号の周波数ビンの数と、前記第2フーリエ変換手段の処理により得られる信号の周波数ビンの数とは一致する。この場合、前記分離行列設定手段は、前記第1の分離行列をそのまま前記第2の分離行列として設定することができる。

【0016】

一方、前記第2の時間領域信号の時間長が、前記第1の時間領域信号の時間長よりも短く設定されている場合、学習計算により得られる前記第1の分離行列の行列要素の数は、フィルタ処理で用いる分離行列において必要十分な行列要素の数よりも多くなる。従って、前記分離行列設定手段は、前記第1の分離行列をそのまま前記第2の分離行列として設定することができない。

この場合、前記分離行列設定手段は、前記第1の分離行列を構成する行列要素を複数のグループごとに集約することにより得られる行列を前記第2の分離行列として設定する。これにより、必要十分な数の行列要素（フィルタ係数）が設定されたフィルタ処理用の分離行列（前記第2の分離行列）を設定することができる。

ここで、前記第2の時間領域信号の時間長は、前記第1の時間領域信号の時間長よりも短く設定される場合、その2倍以上の整数倍が前記第1の時間領域信号の時間長となるように設定されることが望ましい。

これにより、前記第1の分離行列における行列要素のグループと、前記第2の分離行列における行列要素との対応関係が明確になる。

また、前記分離行列設定手段における前記集約とは、例えば、前記第1の分離行列を構成する行列要素について、複数のグループごとに1つの行列要素を選択することや、或いは複数のグループごとに行列要素の平均値若しくは加重平均値を算出すること等である。

ここで、学習計算に対応するフーリエ変換処理と、フィルタ処理に対応するフーリエ変換処理とで、入力信号の時間長（サンプル数）が異なることは、音源分離性能に影響するとも考えられる。しかしながら、後述する実験結果によれば、その影響は比較的小さい。

【0017】

10

20

30

40

50

また、前記第 2 の時間領域信号としては、以下のようなものが考えられる。

例えば、前記第 2 の時間領域信号が、前記第 2 の時間長の 2 倍以上の予め定められた時間長分の最新の前記混合音声信号であることが考えられる。

或いは、前記第 2 の時間領域信号が、前記第 2 の時間長の 2 倍の時間長分の最新の前記混合音声信号に所定の数の定数信号（例えば、0 値信号）が付加された信号であることも考えられる。なお、0 値信号とは、値が 0 の信号である。

また、本発明は、以上に示した音源分離装置が備える各手段が実行する処理を、所定のプロセッサにより実行する音源分離方法として捉えることもできる。

同様に、本発明は、所定のプロセッサを、以上に示した音源分離装置が備える各手段として機能させるための音源分離プログラムとして捉えることもできる。

【発明の効果】

【0018】

本発明によれば、フィルタ処理の入力信号として用いる前記第 2 の周波数領域信号を得るためのフーリエ変換（前記第 2 のフーリエ変換手段の処理）の実行周期（前記第 2 の時間 t_2 ）を十分に短く設定することにより、出力遅延の時間を従来よりも大幅に短縮できる。

また、分離行列の学習計算に対応するフーリエ変換（前記第 1 のフーリエ変換手段の処理）の実行周期（前記第 1 の時間 t_1 ）は、前記第 2 の時間 t_2 に関わらず、十分長い時間（例えば、8 kHz のサンプリング周期 \times 1024 サンプルの信号の長さ相当）に設定できる。これにより、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

【発明を実施するための最良の形態】

【0019】

以下添付図面を参照しながら、本発明の実施の形態について説明し、本発明の理解に供する。尚、以下の実施の形態は、本発明を具体化した一例であって、本発明の技術的範囲を限定する性格のものではない。

ここに、図 1 は本発明の実施形態に係る音源分離装置 X の概略構成を表すブロック図、図 2 は音源分離装置 X によるフィルタ処理（第 1 実施例）の流れを表すブロック図、図 3 は音源分離装置 X によるフィルタ処理（第 2 実施例）の流れを表すブロック図、図 4 は音源分離装置 X による時間領域信号の設定処理の様子を表す図、図 5 は音源分離装置 X による第 1 の実施例の処理と従来の音源分離処理との性能比較実験の結果を表すグラフ、図 6 は音源分離装置 X による第 2 の実施例の処理と従来の音源分離処理との性能比較実験の結果を表すグラフである。

【0020】

以下、図 1 に示すブロック図を参照しつつ、本発明の実施形態に係る音源分離装置 X について説明する。

音源分離装置 X は、複数の音源 1、2 が存在する音響空間に配置される複数のマイクロホン 111、112（音響入力手段）に接続される。

そして、音源分離装置 X は、そのマイクロホン 111、112 各々を通じて逐次入力される複数の混合音声信号 $x_i(t)$ から、各音源 1、2 のうちの 1 つ以上に対応する音源信号を分離（同定）した分離信号（即ち、音源信号を同定した信号） $y_i(t)$ を逐次生成してスピーカ（音声出力手段）に対してリアルタイム出力するものである。ここで、混合音声信号は、音源 1、2 各々からの音源信号（個別の音声信号）が重畳された信号であり、一定のサンプリング周期で逐次デジタル化されて入力されるデジタル信号である。

【0021】

図 1 に示すように、音源分離装置 X は、A/D コンバータ 21（図中、ADC と表記）、D/A コンバータ 22（図中、DAC と表記）、入力バッファ 23、デジタル処理部 Y を備えている。

また、デジタル処理部 Y は、第 1 入力バッファ 31、第 1 FFT 処理部 32、第 1 中間バッファ 33、学習演算部 34、第 2 入力バッファ 41、第 2 FFT 処理部 42、第 2 中

10

20

30

40

50

間バッファ 4 3、分離フィルタ処理部 4 4、第 3 中間バッファ 4 5、I F F T 処理部 4 6、第 4 中間バッファ 4 7、合成処理部 4 8 及び出力バッファ 4 9 を備えている。

ここで、デジタル処理部 Y は、例えば、D S P (Digital Signal Processor) 等の演算用のプロセッサ及びそのプロセッサにより実行されるプログラムが記憶された R O M 等の記憶手段、並びに R A M 等のその他の周辺装置により構成される。また、デジタル処理部 Y は、1 つの C P U 及びその周辺装置を有するコンピュータと、そのコンピュータにより実行されるプログラムとにより構成される場合もある。また、デジタル処理部 Y が有する機能は、所定のコンピュータ (音源分離装置が備えるプロセッサを含む) に実行させる音源分離プログラムとしても提供可能である。

なお、図 1 には、入力される混合音声信号 $x_i(t)$ のチャンネル数 (即ち、マイクロホンの数) が 2 つである例について示しているが、チャンネル数 n は、分離対象とする音源信号の数以上であれば、3 チャンネル以上であっても同様の構成により実現できる。

【 0 0 2 2 】

A / D コンバータ 2 1 は、複数のマイクロホン 1 1 1、1 1 2 各々から入力されるアナログの混合音声信号各々を、一定のサンプリング周期 (即ち、一定のサンプリング周波数) でサンプリングすることにより、デジタルの混合音声信号 $X_i(t)$ に変換し、変換後の信号を入力バッファ 2 3 に出力する (書き込む) ものである。例えば、各音源信号 $S_i(t)$ が人の声の音声信号である場合、8 k H z 程度のサンプリング周期でデジタル化すればよい。

入力バッファ 2 3 は、A / D コンバータ 2 1 によりデジタル化された混合音声信号を一時記憶するメモリである。新たな混合音声信号 $S_i(t)$ が入力バッファ 2 3 に $N / 4$ サンプル分だけ蓄積されるごとに、その $N / 4$ サンプル分の混合音声信号 $S_i(t)$ は、入力バッファ 2 3 から第 1 入力バッファ 3 1 及び第 2 入力バッファ 4 1 のそれぞれに伝送される。従って、入力バッファ 2 3 の記憶容量は、 $N / 2$ サンプル分 (= $N / 4 \times 2$) 以上あれば足りる。

【 0 0 2 3 】

音源分離装置 X において、第 1 入力バッファ 3 1、第 1 F F T 処理部 3 2、第 1 中間バッファ 3 3 及び学習演算部 3 4 は、それぞれ図 8 に示した従来の第 1 入力バッファ 3 1、第 1 F F T 処理部 3 2、第 1 中間バッファ 3 3 及び学習演算部 3 4 と同じ処理を実行するものである。

即ち、第 1 F F T 処理部 3 2 は、第 1 入力バッファ 3 1 に N サンプル分の新たな混合音声信号 $S_i(t)$ が記録されるごとに、フーリエ変換処理を実行する。なお、第 1 F F T 処理部 3 2 の処理の実行周期 (ここでは、 N サンプル分の信号の時間長) を、以下、第 1 の時間 t_1 という。

より具体的には、第 1 F F T 処理部 3 2 は、 N サンプル以上、即ち、第 1 の時間 t_1 の長さ分以上 (ここでは、 $2 N$ サンプル分) の最新の混合音声信号である第 1 時間領域信号 S_0 にフーリエ変換処理を施し、これにより得られる第 1 周波数領域信号 $S_f 0$ を、第 1 中間バッファ 3 3 に一時記憶させる (第 1 のフーリエ変換手段の一例)。

また、学習演算部 3 4 (分離行列学習計算手段の一例) は、所定の時間 T_{sec} ごとに、第 1 中間バッファ 3 3 に一時記憶された最新の時間 T_{sec} 分の第 1 周波数領域信号 $S_f 0$ を読み出し、読み出した信号に基づいて、前述した F D I C A 法 (周波数領域での独立成分分析法) による学習計算を行う。

さらに、学習演算部 3 4 は、その学習計算により算出される分離行列 (以下、第 1 分離行列という) に基づいて、分離信号の分離生成 (フィルタ処理) に用いる分離行列 (以下、第 2 分離行列という) を設定及び更新する (分離行列設定手段の一例)。なお、第 2 分離行列の設定方法については後述する。

【 0 0 2 4 】

[第 1 実施例]

次に、図 2 を参照しつつ、音源分離装置 X によるフィルタ処理の第 1 実施例について説明する。図 2 は、音源分離装置 X によるフィルタ処理 (第 1 実施例) の流れを表すプロッ

10

20

30

40

50

ク図である。

ここで、図 2 に示す各バッファ（第 2 入力バッファ 4 1、第 2 中間バッファ 4 3、第 3 中間バッファ 4 5、第 4 中間バッファ 4 7、出力バッファ 4 9）は、説明の便宜上、非常に多くのデータを蓄積可能であるかのように記載されている。しかしながら、実際には、各バッファは、記憶するデータのうち不要になったものが順次消去され、これにより生じる空き領域が再利用されるので、その記憶容量は必要十分な量に設定されている。

【 0 0 2 5 】

第 2 F F T 処理部 4 2（第 2 のフーリエ変換手段の一例）は、新たな $N / 4$ サンプル分の混合音声信号（第 2 の時間長分の新たな前記混合音響信号の一例）が第 2 入力バッファ 4 1 に入力（記録）されるごとに、その 2 倍の時間長分（ $N / 2$ サンプル分）の最新の混合音声信号を含む第 2 時間領域信号 $S 1$ について、フーリエ変換処理を実行し、その処理結果である第 2 周波数領域信号 $S f 1$ を、第 2 中間バッファ 4 3 に一時記憶させる。なお、第 2 F F T 処理部 4 2 の処理の実行周期（ここでは、 $N / 4$ サンプル分の信号の時間長）を、以下、第 2 の時間 $t 2$ という。

このように、音源分離処理装置 X では、第 2 F F T 処理部 4 2 によるフーリエ変換処理の実行周期（即ち、第 2 の時間 $t 2$ ）が、第 1 F F T 処理部 3 2 によるフーリエ変換処理の実行周期（即ち、第 1 の時間 $t 1$ ）よりも短い周期となるように予め設定されている。

また、第 2 F F T 処理部 4 2 は、少なくとも $N / 4$ サンプル分ずつ時間帯が順次重複する第 2 時間領域信号 $S 1$ （混合音声信号）について、フーリエ変換処理を実行する。ここで、第 2 入力バッファ 4 1 に蓄積された信号のサンプル数が、 $2 N$ 個に満たない場合（処理開始後の初期の段階）には、第 2 F F T 処理部 4 2 は、不足する数だけ 0 値を充当した信号について、フーリエ変換処理を実行する。

なお、この第 2 周波数領域信号 $S f 1$ の周波数ピンの数は、第 2 周波数領域信号 $S f 1$ のサンプル数の $1 / 2$ 倍（ $= N$ 個）である。

【 0 0 2 6 】

この第 1 実施例において、第 2 時間領域信号 $S 1$ は、例えば以下のようなものが考えられる。

まず、図 2 に示すように、第 2 時間領域信号 $S 1$ が、最新の $2 N$ サンプル分の混合音声信号であることが考えられる。

その他、第 2 時間領域信号 $S 1$ が、第 2 の時間 $t 2$ の 2 倍の時間長分の最新の混合音声信号（最新の $N / 2$ サンプル分の混合音声信号）に、 $(3 N / 4)$ 個の定数信号（例えば、0 値信号）が付加された信号であることも考えられる。そのような第 2 時間領域信号 $S 1$ は、例えば、第 2 F F T 処理部 4 2 がパディング処理を行うことによって設定される。

図 4 は、パディング処理によって第 2 時間領域信号 $S 1$ を設定する処理の様子を表すブロック図である。図 4 において、各升目は $N / 4$ サンプル分の混合音声信号のセットを表す。また、図 4 において、各升目に記す「0」は 0 値信号を表し、各升目に記す「1」～「3」は、 $N / 4$ サンプル分の混合音声信号の時系列の番号を表す。

図 4 (a) 「Case 1」は、最新の $(2 N / 4)$ サンプル分の混合音声信号を信号列の最後尾に配置し、残りの部分に $(6 N / 4)$ サンプル分の 0 値信号（定数信号の一例）を付加（充当）するパディング処理により、第 2 時間領域信号 $S 1$ （合計 $2 N$ サンプル分の信号）が設定される様子を表す。

図 4 (b) 「Case 2」は、最新の $(2 N / 4)$ サンプル分の混合音声信号を信号列の先頭に配置し、残りの部分に $(6 N / 4)$ サンプル分の 0 値信号（定数信号の一例）を付加（充当）するパディング処理により、第 2 時間領域信号 $S 1$ （合計 $2 N$ サンプル分の信号）が設定される様子を表す。

図 4 (c) 「Case 3」は、最新の $(2 N / 4)$ サンプル分の混合音声信号を信号列における中間の予め定められた位置に配置し、残りの部分に $(6 N / 4)$ サンプル分の 0 値信号（定数信号の一例）を付加（充当）するパディング処理により、第 2 時間領域信号 $S 1$ （合計 $2 N$ サンプル分の信号）が設定される様子を表す。

【 0 0 2 7 】

10

20

30

40

50

また、分離フィルタ処理部 4 4 (分離フィルタ処理手段)は、第 2 中間バッファ 4 3 に、新たな第 2 周波数領域信号 $S f 1$ が記録されるごとに、その信号 $S f 1$ について、分離行列を用いたフィルタ処理(行列演算)を行い、その処理により得られる第 3 周波数領域信号 $S f 2$ を、第 3 中間バッファ 4 5 に一時記憶させる。このフィルタ処理に用いられる分離行列は、前述した学習演算部 3 4 によって更新されるものである。なお、学習演算部 3 4 により最初に分離行列が更新されるまでは、分離フィルタ処理部 4 4 は、予め定められた初期値が設定された分離行列(初期行列)を用いてフィルタ処理を行う。ここで、第 2 周波数領域信号 $S f 1$ と第 3 周波数領域信号 $S f 2$ とは、周波数ピンの数が等しい($= N$) ことはいうまでもない。

【 0 0 2 8 】

また、IFFT 処理部 4 6 (逆フーリエ変換手段の一例)は、第 3 中間バッファ 4 5 に新たな第 3 周波数領域信号 $S f 2$ が記録されるごとに、その新たな第 3 周波数領域信号 $S f 2$ について、逆フーリエ変換処理を実行し、その処理結果である第 3 時間領域信号 $S 2$ を、第 4 中間バッファ 4 7 に一時記憶させる。この第 3 時間領域信号 $S 2$ のサンプル数は、第 3 周波数領域信号 $S f 2$ の周波数ピンの数($= N$)の 2 倍($= 2 N$)である。前述したように、第 2 F F T 処理部 4 2 が、($7 N / 4$) サンプル分ずつ時間帯が重複する第 2 時間領域信号 $S 1$ (混合音声信号)についてフーリエ変換処理を実行するので、第 4 中間バッファ 4 7 に記録される連続する 2 つの第 3 時間領域信号 $S 2$ も、相互に($7 N / 4$) サンプル分だけ時間帯が重複(オーバーラップ)している。

また、合成処理部 4 8 は、第 4 中間バッファ 4 7 に新たな第 3 時間領域信号 $S 2$ が記録されるごとに、以下に示す合成処理を実行することによって新たな分離信号 $S 3$ を生成し、その信号を出力バッファ 4 9 に一時記憶させる。

ここで、前記合成処理は、IFFT 処理部 4 6 によって得られた新たな第 3 時間領域信号 $S 2$ とその 1 回前に得られた第 3 時間領域信号 $S 2$ について、それらにおける時間帯が重複する一部分の両信号(ここでは、 $N / 4$ サンプル分の信号)を、例えばクロスフェードの重み付けをして加算すること等によって合成する処理である。これにより、平滑化された分離信号 $S 3$ が得られる。

以上の処理により、多少の出力遅延が生じるものの、音源に対応する分離信号 $S 3$ (前述した分離信号 $y_i(t)$ と同じもの)が、リアルタイムで出力バッファ 4 9 に記録される。

【 0 0 2 9 】

ところで、この第 1 実施例では、第 1 時間領域信号 $S 0$ の時間長 $t 1$ (サンプル数 $2 N$) と、第 2 時間領域信号 $S 1$ の時間長 $t 2$ (サンプル数 $2 N$) とが等しく設定されている。このため、第 1 F F T 処理部 3 2 の処理により得られる信号 $S f 0$ の周波数ピンの数($= N$) と、第 2 F F T 処理部 4 2 の処理により得られる信号 $S f 1$ の周波数ピンの数($= N$) とは一致する。

従って、学習演算部 3 4 (分離行列設定手段の一例)は、学習計算により得られる前記第 1 分離行列を、そのままフィルタ処理に用いる前記第 2 分離行列として設定する。

この学習演算部 3 4 の処理により、フィルタ処理に用いられる前記第 2 分離行列が、音響環境の変化に適合したものに適宜更新される。

【 0 0 3 0 】

第 1 実施例のフィルタ処理を実行する音源分離装置 X では、第 2 F F T 処理部 4 2 の処理の実行周期(時間 $t 2$)の方が、第 1 F F T 処理部 3 2 の処理の実行周期(時間 $t 1$)よりも短い。従って、前記第 2 の時間 $t 2$ を従来よりも十分に短く設定すること(ここでは、 $N / 4$ サンプルの信号の時間長)により、出力遅延の時間を従来よりも大幅に短縮できる。

一方、第 1 F F T 処理部 3 2 の処理の実行周期(時間 $t 1$)は、時間 $t 2$ に関わらず、十分長い時間(例えば、8 kHz のサンプリング周期 $\times 1 0 2 4$ サンプルの信号の長さ相当)に設定できる。これにより、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

10

20

30

40

50

【 0 0 3 1 】

以下、音源分離装置 X の効果について説明する。

前述したように F D I C A 法による音源分離処理では、出力遅延の時間が、フィルタ処理の入力信号として用いる第 2 周波数領域信号 $S_f 1$ を得るための処理（第 2 F F T 処理部 4 2 の処理）の実行周期 t_2 の 2 倍強から 3 倍程度の時間となる。

これに対し、音源分離装置 X では、第 2 F F T 処理部 4 2 の処理の実行周期 t_2 を、従来よりも十分に短く設定することができ、出力遅延の時間を従来よりも大幅に短縮できる。図 2 に示した実施例では、出力遅延の時間を、図 8 に示した従来の音源分離処理における出力遅延の時間に対して 4 分の 1 に短縮できる。

一方、分離行列の学習計算に対応するフーリエ変換処理（第 1 F F T 処理部 3 2 の処理）の実行周期（第 1 の時間 t_1 ）は、前記第 2 の時間 t_2 に関わらず、十分長い時間（例えば、8 kHz のサンプリング周期 $\times 1024$ サンプルの信号の長さ相当）に設定できる。これにより、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

【 0 0 3 2 】

図 5 は、音源分離装置 X による第 1 実施例の音源分離処理と従来の音源分離処理との性能比較実験の結果を表すグラフである。

実験条件は、以下の通りである。

まず、所定の空間において、ある基準位置から左右の等距離の位置それぞれに 2 つのマイクロホン 1 1 1、1 1 2 を所定方向（以下、正面方向という）に向けて配置した。ここで、前記基準位置を中心とした場合に、正面方向を 0° 方向とし、上方から見て時計回りの角度を θ とする。

そして、2 つの音源（第 1 音源及び第 2 音源）の種類及び配置方向は、次の 7 パターン（以下、音源パターン 1 ~ 音源パターン 7 という）とした。

音源パターン 1：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -30^\circ$ の方向。第 2 音源は発話する女性。第 2 音源の配置方向は $\theta = +30^\circ$ の方向。

音源パターン 2：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -60^\circ$ の方向。第 2 音源はエンジン音を発する自動車。第 2 音源の配置方向は $\theta = +60^\circ$ の方向。

音源パターン 3：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -60^\circ$ の方向。第 2 音源は所定のノイズ音を発する音源。第 2 音源の配置方向は $\theta = +60^\circ$ の方向。

音源パターン 4：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -60^\circ$ の方向。第 2 音源は所定のクラシック音楽を出力する音響機器。第 2 音源の配置方向は $\theta = +60^\circ$ の方向。

音源パターン 5：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = 0^\circ$ の方向。第 2 音源は発話する女性。第 2 音源の配置方向は $\theta = +60^\circ$ の方向。

音源パターン 6：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -60^\circ$ の方向。第 2 音源は所定のクラシック音楽を出力する音響機器。第 2 音源の配置方向は $\theta = 0^\circ$ の方向。

音源パターン 7：第 1 音源の種類は発話する男性。第 1 音源の配置方向は $\theta = -60^\circ$ の方向。第 2 音源はエンジン音を発する自動車。第 2 音源の配置方向は $\theta = 0^\circ$ の方向。

また、いずれの音源パターンにおいても、混合音声信号のサンプリング周波数は 8 kHz である。

また、評価値（グラフの縦軸）は、第 1 音源の信号を分離対象とする目的信号 (Signal) とした場合に、これに第 2 音源の信号成分 (Noise) がどの程度混入しているかを示す S N 比 (dB) である。S N 比の値が大きいほど、音源信号の分離性能が高いことを表す。

【 0 0 3 3 】

また、図 5 において、 g_1 は図 8 に示した従来の音源分離処理 ($N = 512$) の結果（従って出力遅延は 192 msec）を表す。また、 g_2 は図 8 に示した従来の音源分離処理において $N = 128$ とした場合の結果（従って、出力遅延は 48 msec）を表す。

10

20

30

40

50

一方、図5において、 $g \times 1$ は、音源分離装置Xによる第1実施例の音源分離処理において、 $N = 512$ であり、第2FFT処理部42への入力信号(第2時間領域信号S1)が最新の $2N$ サンプル分の混合音声信号である場合の結果(出力遅延は48msec)を表す。

また、 $g \times 2$ は、音源分離装置Xによる第1実施例の音源分離処理において、 $N = 512$ であり、第2FFT処理部42への入力信号(第2時間領域信号S1)が図4に示したパディング処理(0値充当)に基づく信号である場合の結果(出力遅延は48msec)を表す。

【0034】

図5(a)、(b)に示すグラフからわかるように、音源分離装置X1の処理結果 $g \times 1$ 、 $g \times 2$ は、従来処理結果 $g1$ に対し、出力遅延の時間が $1/4$ に短縮されているにもかかわらず、ほぼ同等の音源分離性能(同等のSN比)が得られることがわかる。

ちなみに、従来処理において、第1FFT処理部32及び第2FFT処理部42'の両方の処理周期を単に $1/4$ 倍($N = 128$)とした場合($g2$)、音源分離性能が大きく劣化することがわかる。

以上に示したように、音源分離処理装置Xによれば、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

【0035】

[第2実施例]

次に、図3を参照しつつ、音源分離装置Xによるフィルタ処理の第2実施例について説明する。図3は、音源分離装置Xによるフィルタ処理(第2実施例)の流れを表すブロック図である。

この第2実施例のフィルタ処理が、前記第1実施例のフィルタ処理と異なる点は、第2時間領域信号S1のサンプル数が少ない(信号の時間長が短い)点である。即ち、この第2実施例では、第2時間領域信号S1のサンプル数が、第1時間領域信号S0のサンプル数よりも短く設定されている。これは、第2時間領域信号S1の時間長が、第1時間領域信号S0の時間長よりも短く設定されているということと同じ意味である。

図3に示す例では、第2時間領域信号S1のサンプル数が($2N/4$)個に設定されている。これに対し、第1時間領域信号S0のサンプル数は、第1実施例の場合と同じ $2N$ 個である(図8参照)。即ち、第2時間領域信号S1の時間長の4倍(2倍以上の整数倍の一例)が、第1時間領域信号S0の時間長となるように設定されている。

これにより、第3時間領域信号S2のサンプル数も($2N/4$)個になる。しかしながら、第1実施例においても、合成処理部48は、時間帯が重複する $N/4$ サンプル分の信号についてのみ合成処理を行うものである。従って、第2実施例においても、合成処理部48の処理は、第1実施例の場合と特に変わりはない。第1実施例の場合と異なるのは、第3時間領域信号S2中に、合成処理に用いない信号が含まれないことだけである。

【0036】

一方、第2実施例では、第2時間領域信号S1の時間長が、第1時間領域信号S0の時間長よりも短く設定されている(サンプル数が少ない)ため、学習計算により得られる前記第1分離行列の行列要素(フィルタ係数)の数は、フィルタ処理で用いる前記第2分離行列において必要十分な行列要素の数よりも多くなる。従って、学習演算部34は、前記第1分離行列をそのまま前記第2分離行列として設定することができない。

図3に示す例では、第1時間領域信号S0のサンプル数($2N$)が、第2時間領域信号S1のサンプル数($=N/2$)の4倍となるため、前記第1分離行列の行列要素(フィルタ係数)4個と、前記第2分離行列の行列要素1個とが相互に対応する関係となる。

そこで、第2実施例では、学習演算部34(分離行列設定手段の一例)は、前記第1分離行列を構成する行列要素(フィルタ係数)を、前記第2分離行列の行列要素それぞれに対応する複数のグループに区分し、そのグループごとに行列要素(フィルタ係数)を集約することに、前記第2分離行列として設定する分離行列(行列要素)を算出する。

【0037】

10

20

30

40

50

ここで、前記第1分離行列の行列要素(フィルタ係数)を集約する方法としては、例えば、次の2つが考えられる。

その1つは、前記第1分離行列を構成する行列要素(フィルタ係数)について、複数のグループごとに1つの行列要素を代表値として選択するという集約処理が考えられる。以下、この集約を代表値集約という。

その他、前記第1分離行列を構成する行列要素(フィルタ係数)について、複数のグループごとに行列要素の平均値を算出、或いは予め定められた重み係数に基づく加重平均値を算出するという集約処理も考えられる。以下、この集約を平均値集約という。なお、この平均値集約には、各グループにおいて、その一部の行列要素についての平均値や加重平均値を算出することも含まれる。例えば、4個の行列要素(フィルタ係数)ごとにグループ化される場合、グループごとに所定の3つの行列要素について平均値を求めること等が考えられる。

10

これらいずれかの集約処理により、学習演算部34は、必要十分な数の行列要素(フィルタ係数)を有する前記第2分離行列を設定する。

このような第2実施例に係る音源分離処理によっても、前記第1実施例の場合と同様に、出力遅延の時間を短縮しつつ、高い音源分離性能を確保することができる。

ここで、学習計算に対応するフーリエ変換処理と、フィルタ処理に対応するフーリエ変換処理とで、入力信号の時間長(サンプル数)が異なることは、音源分離性能に影響するとも考えられる。しかしながら、以下に示す実験結果によれば、その影響は比較的小さい。

20

【0038】

図6は、音源分離装置Xによる第2実施例の音源分離処理と従来の音源分離処理との性能比較実験の結果を表すグラフである。

実験条件とした音源パターンは、前述した音源パターン1~音源パターン7と同じである。また、混合音声信号のサンプリング周波数は8kHzである。

さらに、評価値(グラフの縦軸)も、図5に示したものと同一SN比であり、その値が大きいほど、音源信号の分離性能が高いことを表す。

また、図6において、g1及びg2は、図5に示したg1及びg2と同じ実験の結果である。

一方、図6において、gx3は、音源分離装置Xによる第2実施例の処理において、 $N = 512$ であり、第2FFT処理部42への入力信号(第2時間領域信号S1)が最新の $N/2$ サンプル分の混合音声信号であり、前記平均値集約(通常の前記第2分離行列を設定した場合の結果(出力遅延は4.8msec))を表す。

30

また、gx4は、音源分離装置Xによる第2実施例の処理において、 $N = 512$ であり、第2FFT処理部42への入力信号(第2時間領域信号S1)が最新の $N/2$ サンプル分の混合音声信号であり、前記代表値集約によって前記第2分離行列を設定した場合の結果(出力遅延は4.8msec)を表す。

【0039】

図6(a)、(b)に示すグラフからわかるように、音源分離装置X1の処理結果gx3(平均値集約)は、従来の処理結果g1に対し、出力遅延の時間が $1/4$ に短縮されているにもかかわらず、それほど遜色のない音源分離性能(同等のSN比)が得られることがわかる。また、音源分離装置X1の処理結果gx3は、従来の音源分離処理において、第1FFT処理部32及び第2FFT処理部42'の両方の処理周期を単に $1/4$ 倍($N = 128$)とした場合(g2)に対し、高い音源分離性能(同等のSN比)が得られることがわかる。

40

一方、音源分離装置X1の処理結果gx4(代表値集約)は、前記平均値集約の場合の処理結果gx3ほどの分離性能は得られていない。しかしながら、処理結果gx4(代表値集約)は、音源パターン6や音源パターン7のように、音源の一方が正面に配置されているような音源パターンにおいて、処理結果g2よりも分離性能が改善されている。一般に、音源の1つが正面に配置された音源パターンは、ICA法による音源分離処理によっ

50

て高い分離性能が得られにくいパターンである。

従って、音源の存在方向を検知或いは推定することが可能である場合、音源の存在方向に応じて、前記第2分離行列を設定するための集約処理の方法を切り替えることが考えられる。同様に、音源の存在方向に応じて、音源分離処理の方法自体（本発明の音源分離処理か従来の音源分離処理か）を切り替えることも考えられる。

【産業上の利用可能性】

【0040】

本発明は、音源分離装置への利用が可能である。

【図面の簡単な説明】

【0041】

【図1】本発明の実施形態に係る音源分離装置Xの概略構成を表すブロック図。

【図2】音源分離装置Xによるフィルタ処理（第1実施例）の流れを表すブロック図。

【図3】音源分離装置Xによるフィルタ処理（第2実施例）の流れを表すブロック図。

【図4】音源分離装置Xによる時間領域信号の設定処理の様子を表す図。

【図5】音源分離装置Xによる第1の実施例の処理と従来の音源分離処理との性能比較実験の結果を表すグラフ。

【図6】音源分離装置Xによる第2の実施例の処理と従来の音源分離処理との性能比較実験の結果を表すグラフ。

【図7】FDICA法による分離行列の学習計算を行う学習計算ユニットZ1の概略構成を表すブロック図。

【図8】従来のFDICA法による音源分離処理の流れを表すブロック図。

【図9】従来のFDICA法による音源分離処理における信号入出力の状態遷移を表すブロック図。

【符号の説明】

【0042】

X ... 本発明の実施形態に係る音源分離装置

Y ... デジタル処理部

1、2 ... 音源

2 1 ... A / Dコンバータ

2 2 ... D / Aコンバータ

2 3 ... 入力バッファ

3 1 ... 第1入力バッファ

3 2 ... 第1FFT処理部

3 3 ... 第1中間バッファ

3 4 ... 学習演算部

4 1 ... 第2入力バッファ

4 2 ... 第2FFT処理部

4 3 ... 第2中間バッファ

4 4 ... 分離フィルタ処理部

4 5 ... 第3中間バッファ

4 6 ... IFFT処理部

4 7 ... 第4中間バッファ

4 8 ... 合成処理部

4 9 ... 出力バッファ

1 1 1、1 1 2 ... マイクロホン

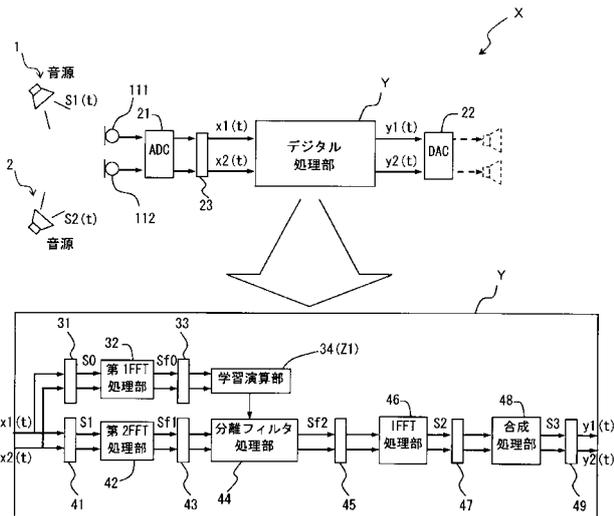
10

20

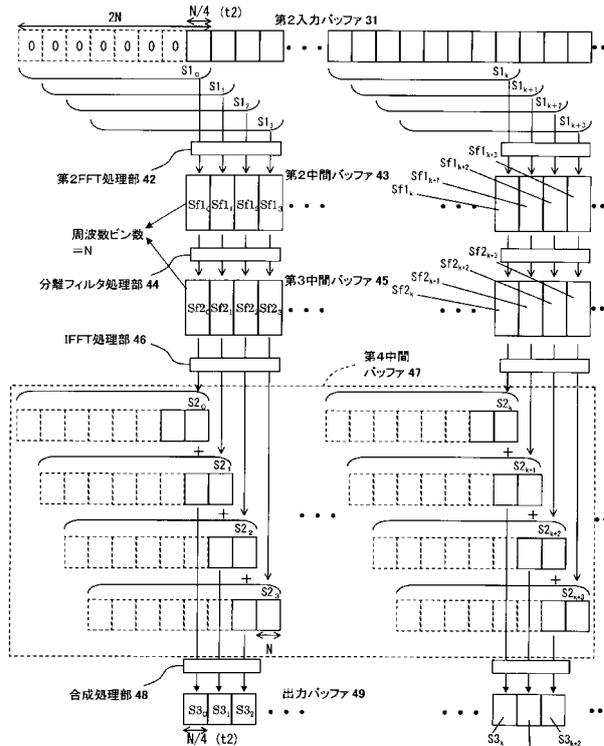
30

40

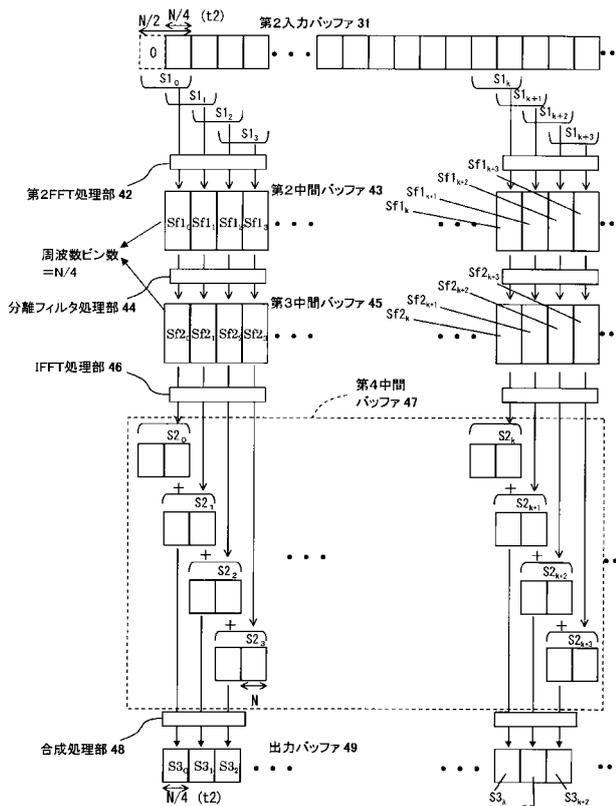
【図1】



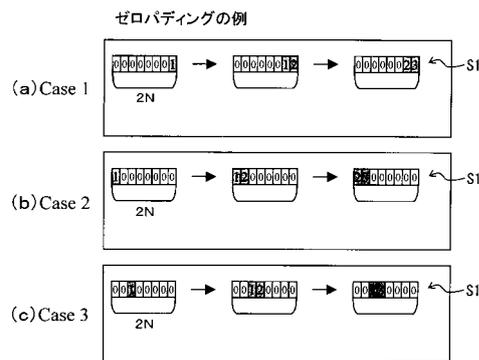
【図2】



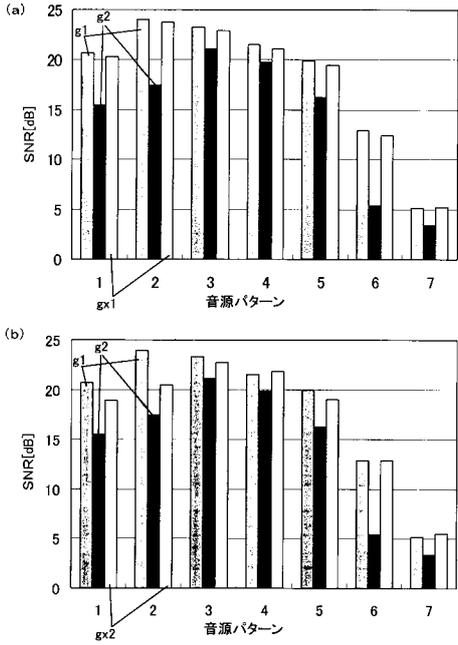
【図3】



【図4】

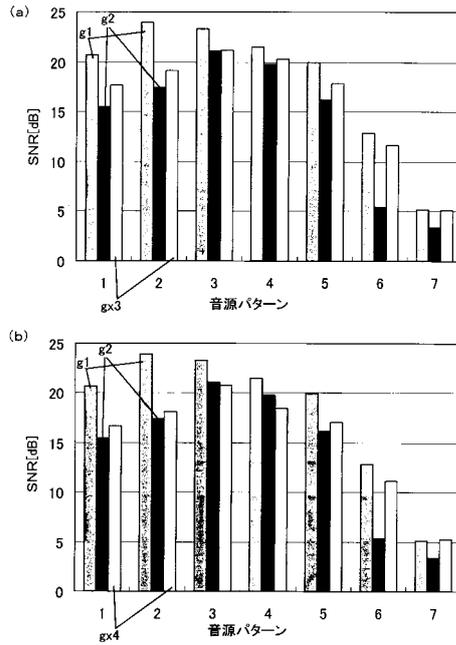


【図5】



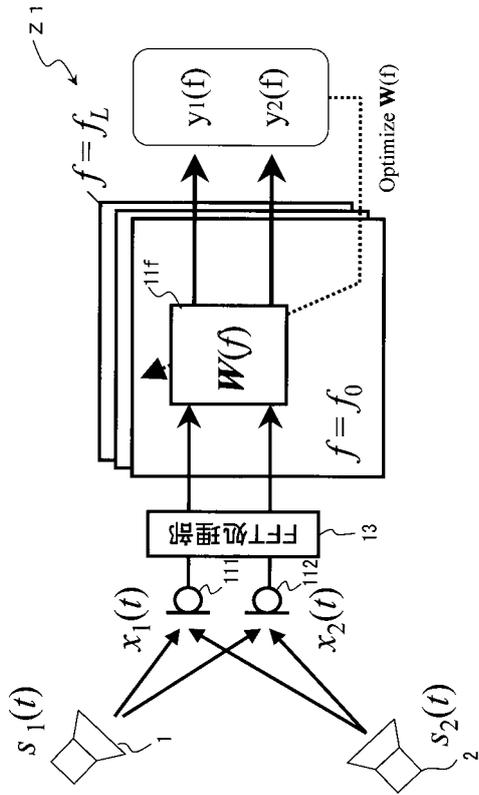
g1 : [遅延時間 192ms] 従来の結果
 g2 : [遅延時間 48ms] FFT入力信号サイズが1/4
 gx1 : [遅延時間 48ms] 本発明の結果(1)
 第2FFT入力が2N サンプル分の混合音声信号
 gx2 : [遅延時間 48ms] 本発明の結果(2)
 第2FFT入力がN/2 サンプル分の混合音声信号を含む
 ゼロパディング信号

【図6】

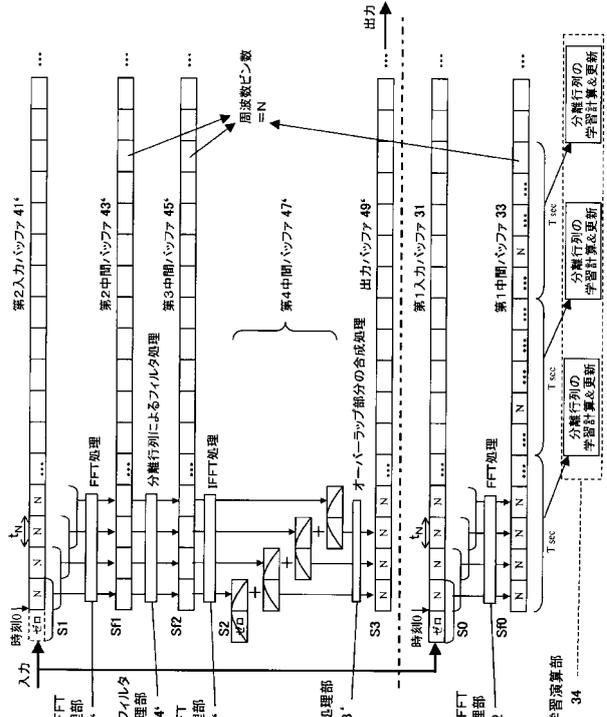


g1 : [遅延時間 192ms] 従来の結果
 g2 : [遅延時間 48ms] FFT入力信号サイズが1/4
 gx3 : [遅延時間 48ms] 本発明の結果(3)
 行列係数を平均値で集約
 gx4 : [遅延時間 48ms] 本発明の結果(2)
 行列係数を選択値で集約

【図7】



【図8】



【 図 9 】

