

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6637014号
(P6637014)

(45) 発行日 令和2年1月29日(2020.1.29)

(24) 登録日 令和1年12月27日(2019.12.27)

(51) Int.Cl. F I
G 1 0 L 21/0272 (2013.01) G 1 0 L 21/0272 1 0 0 B
G 1 0 L 25/18 (2013.01) G 1 0 L 25/18

請求項の数 7 (全 29 頁)

<p>(21) 出願番号 特願2017-212311 (P2017-212311) (22) 出願日 平成29年11月2日(2017.11.2) (62) 分割の表示 特願2015-560567 (P2015-560567) の分割 原出願日 平成25年10月23日(2013.10.23) (65) 公開番号 特開2018-36666 (P2018-36666A) (43) 公開日 平成30年3月8日(2018.3.8) 審査請求日 平成29年11月2日(2017.11.2) (31) 優先権主張番号 61/772, 708 (32) 優先日 平成25年3月5日(2013.3.5) (33) 優先権主張国・地域又は機関 米国 (US) 前置審査</p>	<p>(73) 特許権者 597159765 フラウンホーファーゲゼルシャフト ツール フォルデルング デル アンゲヴァン テン フォルシユング エー. ファー. ドイツ国, デー-80686 ミュンヘン, ハンサシュトラーセ 27ツェー (74) 代理人 100112715 弁理士 松山 隆夫 (72) 発明者 ウーレ、 クリスティアン ドイツ国 90489 ニュルンベルク、 キューンホーファーシュトラーセ 7 (72) 発明者 ハーベッツ、 エマーヌエル ドイツ国 91080 シュパールドルフ、 シュヴェーデンシュトラーセ 13 最終頁に続く</p>
--	--

(54) 【発明の名称】 音声信号処理のためのマルチチャンネル直接・環境分解のための装置及び方法

(57) 【特許請求の範囲】

【請求項1】

2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための装置であって、前記2つ以上の音声入力チャンネル信号の各々は、直接信号部分及び環境信号部分を含み、前記装置は、

第1のパワースペクトル密度情報を推定し第2のパワースペクトル密度情報を推定することによりフィルタを決定するためのフィルタ決定部(110)と、

前記2つ以上の音声入力チャンネル信号に前記フィルタを適用することにより前記1つ以上の音声出力チャンネル信号を生成するための信号処理部(120)と、を備え、

前記フィルタは、前記第1のパワースペクトル密度情報および前記第2のパワースペクトル密度情報に依存し、

前記1つ以上の音声出力チャンネル信号は、前記フィルタに依存し、

前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力チャンネル信号について、前記音声入力チャンネル信号についてのパワースペクトル密度情報を推定することにより前記第1のパワースペクトル密度情報を推定するように構成され、かつ、前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力チャンネル信号について、前記音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を推定することにより前記第2のパワースペクトル密度情報を推定するように構成され、或いは、

前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力

チャンネル信号について、前記音声入力チャンネル信号についてのパワースペクトル密度情報を推定することにより前記第1のパワースペクトル密度情報を推定するように構成され、かつ、前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力チャンネル信号について、前記音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を推定することにより前記第2のパワースペクトル密度情報を推定するように構成され、或いは、

前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力チャンネル信号について、前記音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を推定することにより前記第1のパワースペクトル密度情報を推定するように構成され、かつ、前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号の各音声入力チャンネル信号について、前記音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を推定することにより前記第2のパワースペクトル密度情報を推定するように構成される、装置。

10

【請求項2】

請求項1に記載の装置であって、

前記装置は、更に、前記2つ以上の音声入力チャンネル信号を時間領域から時間周波数領域に変換するための分析フィルタバンク(605)を備え、

前記フィルタ決定部(110)は、前記時間周波数領域において表される前記音声入力チャンネル信号に応じて前記第1のパワースペクトル密度情報及び前記第2のパワースペクトル密度情報を推定することによって前記フィルタを決定するように構成され、

20

前記信号処理部(120)は、時間周波数領域において表される前記2つ以上の音声入力チャンネル信号に前記フィルタを適用することにより、前記時間周波数領域において表される前記1つ以上の音声出力チャンネル信号を生成するように構成され、

前記装置は、更に、時間周波数領域において表される前記1つ以上の音声出力チャンネル信号を、前記時間周波数領域から前記時間領域に変換するための合成フィルタバンク(625)を備える、装置。

【請求項3】

請求項1又は請求項2に記載の装置であって、フィルタ決定部(110)は、前記第1のパワースペクトル密度情報を推定するように構成され、

前記フィルタ決定部(110)は、前記第2のパワースペクトル密度情報を推定するように構成され、

30

前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号のうち少なくとも1つに応じて、数であるトレードオフ情報(i, j)を決定するように構成され、

前記フィルタ決定部(110)は、前記第1のパワースペクトル密度情報、前記第2のパワースペクトル密度情報および前記トレードオフ情報に応じて前記フィルタを決定するように構成される、装置。

【請求項4】

請求項3に記載の装置であって、前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号のうち少なくとも1つにトランジェントが存在するか否かに応じて前記トレードオフ情報(i, j)を決定するように構成される、装置。

40

【請求項5】

請求項3又は請求項4に記載の装置であって、前記フィルタ決定部(110)は、前記2つ以上の音声入力チャンネル信号のうち1つが送信される少なくとも1つの信号チャンネルにおける加算ノイズの存在に応じて前記トレードオフ情報(i, j)を決定するように構成される、装置。

【請求項6】

2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための方法であって、前記2つ以上の音声入力チャンネル信号の各々は、直接信号部分及び環境信号部分を含み、前記方法は、

50

第1のパワースペクトル密度情報を推定し第2のパワースペクトル密度情報を推定することによりフィルタを決定するステップと、

前記2つ以上の音声入力チャネル信号に前記フィルタを適用することにより前記1つ以上の音声出力チャネル信号を生成するステップと、を備え、

前記フィルタは、前記第1のパワースペクトル密度情報および前記第2のパワースペクトル密度情報に依存し、

前記1つ以上の音声出力チャネル信号は、前記フィルタに依存し、

前記第1のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号についてのパワースペクトル密度情報を推定することにより行われ、かつ、前記第2のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報を推定することにより行われ、或いは、

前記第1のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号についてのパワースペクトル密度情報を推定することにより行われ、かつ、前記第2のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号の直接信号部分についてのパワースペクトル密度情報を推定することにより行われ、或いは、

前記第1のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号の直接信号部分についてのパワースペクトル密度情報を推定することにより行われ、かつ、前記第2のパワースペクトル密度情報を推定することは、前記2つ以上の音声入力チャネル信号の各音声入力チャネル信号について、前記音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報を推定することにより行われる、方法。

【請求項7】

コンピュータ又はプロセッサにおいて実行されたときに請求項6に記載の方法を実現するためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声信号処理のためのマルチチャネル直接・環境分解のための装置及び方法に関する。

【背景技術】

【0002】

音声信号処理の重要性が増している。この分野では、音信号を直接音信号及び環境音信号に分離させることが重要な役割を演じている。

【0003】

一般的に、音響音は、直接音及び環境（又は拡散）音を混合したものからなる。直接音は、音源、例えば楽器、歌唱者又はスピーカーによって放射され、可能な限り最短の経路で受信機、例えば聴取者の耳の入り口又はマイクロフォンに到達する。

【0004】

直接音を聞く場合、これは音源の方向から来るものとして知覚される。位置及び他の空間的音特性に適切な聴覚の手がかりは、両耳間のレベル差、両耳間の時間差及び両耳間のコヒーレンスである。同一の両耳間のレベル差及び両耳間の時間差を引き起こす直接音波は、同じ方向から来るものとして知覚される。拡散音が存在しない場合、左耳及び右耳、又は他の多数のセンサに到達する信号はコヒーレントである。

【0005】

これに対し、環境音は、同じ環境音に寄与する互いに間隔を置いた多数の音源又は音反射境界によって放射される。音波が室内の壁に到達すると、その一部が反射され、室内の

10

20

30

40

50

全ての反射を重ね合わせたもの、即ち反響が環境音の主要な例である。他の例としては、聴衆の音（例えば拍手）、自然環境の音（例えば雨）及びその他の背景音（例えばがやがやとしたノイズ）が挙げられる。環境音は、拡散したもの、位置を判定できないものとして知覚され、聴取者には包み込まれる（「音の中に没入した」）ような印象を与える。互いに間隔を置いた多数のセンサを用いて環境音フィールドを捕捉した場合、記録された信号は少なくとも部分的に非コヒーレントである。

【0006】

音声信号を直接信号成分と環境信号成分とに分解することからは、音の後生成及び再生における様々な応用分野で利益が得られる。このような信号処理についての主な課題は、任意の数の入力チャンネル信号及び全ての可能な入力信号特性について高い音質を維持しながら、高精度の分離を達成することである。直接・環境分解（DAD）、即ち音声信号の直接信号成分及び環境信号成分への分解によって、信号成分を別個に再生又は変更することが可能となり、これは例えば音声信号のアップミックスに望ましいものである。

10

【0007】

アップミックスという用語は、入力信号がN個のチャンネルを有する場合にP個のチャンネルを有する信号を作成する（ただし $P > N$ ）プロセスを指す。その主な応用例として、入力信号で利用可能であるよりも多くのチャンネルを有するサラウンド音セットアップを用いた音声信号の再生がある。高度な信号処理アルゴリズムを用いてコンテンツを再生することにより、聴取者は、マルチチャンネル音再生セットアップの全ての利用可能なチャンネルを用いることが可能となる。このような処理により、入力信号を、意味のある信号成分（例えばステレオ画像における知覚位置、直接音対環境音、単一の楽器に基づくもの）、又はこれらの信号成分を減衰若しくは増強させた信号、へと分解することができる。

20

【0008】

アップミックスの2つの概念が広く知られている。

【0009】

1. ガイド型アップミックス：アップミックスプロセスをガイドする追加の情報を用いてアップミックスする。この追加の情報は、入力信号において特定の方法で「符号化」されるか、又は追加的に記憶され得る。

【0010】

2. 非ガイド型アップミックス：出力信号は、追加の情報なしに排他的に音声入力信号から得られる。

30

【0011】

高度なアップミックス方法は、更に、直接信号及び環境信号の位置付けに関して分類することができる。即ち、「直接・環境方式」と「帯内」方式とに区別される。直接・環境ベースの技術の中核的な要素は、環境信号を抽出し、これを例えばマルチチャンネルサラウンド音セットアップの後方チャンネル又は高さチャンネルに入力することである。後方チャンネルまたは高さチャンネルを用いて環境音を再生することによって、聴取者には包み込まれる（「音の中に没入した」）ような印象が与えられる。更に、直接音源を、ステレオパノラマ中の知覚位置に従って前方チャンネルの間で配置することができる。これに対し、「帯内」方式は、全ての利用可能なラウドスピーカーを用いて聴取者の周囲の全ての音（直接音及び環境音）を位置付けることをめざすものである。

40

【0012】

音声信号を直接信号及び環境信号に分解することによって、例えばこれをスケージング又はフィルタリングすることによって環境音又は直接音に別個に変更を加えることも可能となる。一使用例として、過剰な量の環境音を伴って捕捉された音楽演奏の録音処理が挙げられる。別の使用例として、音声生成（例えば映画のサウンド又は音楽におけるもの）であって、異なる位置で捕捉されたため異なる環境音特性を有する音声信号を組み合わせる場合が挙げられる。

【0013】

いずれの場合でも、このような信号処理のための要件は、任意の数の入力チャンネル信号

50

及び全ての可能な入力信号特性について高い音質を維持しながら、高程度の分離を達成することである。

【 0 0 1 4 】

D A D、又は直接信号成分若しくは環境信号成分の減衰若しくは増強についての先行技術における様々な方策が提案されており、以下に簡単に説明する。

【 0 0 1 5 】

公知の概念は、マイクロフォン録音から望ましくない背景ノイズを除去することを目的としたスピーチ信号の処理に関する。

【 0 0 1 6 】

[1]においては、2つの入力チャンネルを有するスピーチ録音からの反響を減衰させる方法が記載されている。入力信号における無相関（又は拡散）信号成分を減衰させることによって反響信号成分を低減させる。この処理は、時間周波数領域で実現されるため、サブバンド信号は、スペクトル重み付け方法を用いて処理される。実数値重み付け因子は、パワースペクトル密度（P S D）を用いて

【 0 0 1 7 】

【数 1】

$$\phi_{xx}(m, k) = E \{ X(m, k) X^*(m, k) \} \quad (1)$$

$$\phi_{yy}(m, k) = E \{ Y(m, k) Y^*(m, k) \} \quad (2)$$

$$\phi_{xy}(m, k) = E \{ X(m, k) Y^*(m, k) \} \quad (3)$$

により計算され、ここで $X(m, k)$ 及び $Y(m, k)$ は、時間領域入力信号 $x_t[n]$ 及び $y_t[n]$ の時間周波数領域表現を示し、 $E\{\cdot\}$ は、期待演算であり、 X^* は、 X の複素共役である。

【 0 0 1 8 】

この文献の著者等は、 $\phi_{xy}(m, k)$ に比例する場合、例えば正規化された相互相関関数（又はコヒーレンス関数）に等しい重みを用いる場合に、異なるスペクトル重み付け関数が有効であると指摘している。

【 0 0 1 9 】

【数 2】

$$\rho(m, k) = \frac{|\Phi_{xy}(m, k)|}{\sqrt{\Phi_{xx}(m, k) \Phi_{yy}(m, k)}} \quad (4)$$

【 0 0 2 0 】

これと同様の理論で、[2]に記載の方法では、周波数帯域で計算された正規化された相互相関関数から導出した重みによるスペクトル重み付けを用いて環境信号を抽出する。式(4)（この文献の著者等は、「チャンネル間短時間コヒーレント関数」という用語を用いている）を参照。[1]と比較すると、拡散信号成分を減衰させる代わりに、 $(1 - \rho(m, k))$ の単調で一様な関数であるスペクトル重みを用いて直接信号成分を減衰させている点で異なっている。

【 0 0 2 1 】

[3]においては、マルチチャンネル・ウィーナフィルタリングを用いて2つのチャンネルを有する入力信号をアップミックスする応用例における分解が記載されている。この処理は、時間周波数領域で行われる。入力信号は、環境信号及び（1周波数帯当り）1つのアクティブな直接源を混合したものとしてモデル化され、ここで、1つのチャンネルにおける直接信号は、第2のチャンネル、即ち振幅パンニングにおける直接信号成分のスケーリングされたコピーへと制限される。正規化された相互相関と、両方のチャンネルにおける入力信号パワーとを用いて、パンニング係数と直接信号及び環境信号のパワーとを推定する。直接出力信号及び環境出力信号は、実数値重み付け係数によって入力信号の線形組み合わせから導出される。追加の後スケーリングを適用することにより、出力信号のパワーが推定

10

20

30

40

50

量に等しくなるようにする。

【 0 0 2 2 】

[4]に記載の方法では、環境パワーの推定値に基づいてスペクトル重み付けを用いて環境信号を抽出する。環境パワーの推定は、両方のチャンネルにおける直接信号成分が完全に相関しており、環境チャンネル信号が互いに及び直接信号と相関しておらず、且つ両方のチャンネルにおける環境パワーが等しいという想定に基づいている。

【 0 0 2 3 】

[5]においては、方向性音声符号化 (D i r A C) に基づくステレオ信号のアップミックス方法が記載されている。D i r A C は、到来方向、拡散性及び音の場のスペクトルを分析及び再生することをめざすものである。ステレオ入力信号をアップミックスするために、入力信号の無エコー性 B フォーマット録音をシミュレートする。

10

【 0 0 2 4 】

[6]においては、適応フィルタアルゴリズムを用いたステレオ音声信号から無相関の反響を抽出する方法であって、1つのチャンネル信号における直接信号成分を、最小平均二乗 (L M S) アルゴリズムによって他のチャンネル信号を用いて予測することをめざすものが記載されている。次に、入力信号から推定直接信号を減算することにより環境信号を導出する。この方策の理論は、予測は相関の信号についてのみ有効であり、予測エラーは無相関の信号に似るといえるものである。L M S 原理に基づく様々な適応フィルタアルゴリズム、例えば L M S 又は正規化 L M S (N L M S) アルゴリズムが存在し、有効である。

【 0 0 2 5 】

20

[7]においては、2つのチャンネルよりも多くのチャンネルを有する入力信号を分解するために、まずマルチチャンネル信号をダウンミックスして2チャンネルステレオ信号を得てから、[3]で示されたステレオ入力信号処理方法を適用する方法が記載されている。

【 0 0 2 6 】

[8]に記載の方法では、モノ信号を処理するために、スペクトル重み付けを用いて環境信号を抽出し、スペクトル重みは、特徴抽出及び教師有り学習を用いて計算される。

【 0 0 2 7 】

アップミックスの応用例におけるモノ録音から環境信号を抽出するもう1つの方法では、入力信号の時間周波数領域表現と、これを圧縮したものの、好ましくは負でない行列の因数分解を用いて計算されたものとの差から時間周波数領域表現を得る [9]。

30

【 0 0 2 8 】

[1 0]には、音声信号における反響信号成分を、反響信号を生成した反響システムの大きさ伝達関数の推定値に基づいて抽出し変化させる方法が記載されている。信号成分の周波数領域表現の大きさの推定値は、再帰的フィルタリングによって導出され、変更を加えることができる。

【 0 0 2 9 】

本発明の目的は、音声信号処理のためのマルチチャンネル直接・環境分解のための改善された概念を提供することである。本発明の目的は、請求項 1 に記載の装置、請求項 1 4 に記載の方法、及び請求項 1 5 に記載のコンピュータプログラムによって解決される。

【 0 0 3 0 】

40

2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための装置が提供される。2つ以上の音声入力チャンネル信号の各々は、直接信号部分及び環境信号部分を含む。装置は、第1のパワースペクトル密度情報を推定し第2のパワースペクトル密度情報を推定することによりフィルタを決定するためのフィルタ決定部を備える。更に、装置は、2つ以上の音声入力チャンネル信号にフィルタを適用することにより1つ以上の音声出力チャンネル信号を生成するための信号処理部を備える。第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示

50

し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示す。或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。

【0031】

実施例は、音声入力信号を直接信号成分及び環境信号成分に分解し、これらを音の後生成及び再生に適用するための概念を提供する。このような信号処理における主な課題は、任意の数の入力チャンネル信号及び全ての可能な入力信号特性について高い音質を維持しながら、高精度の分離を達成することである。本願により提供される概念は、時間周波数領域におけるマルチチャンネル信号処理であって、平均平方誤差の意味での条件付き最適解につながるものであり、例えば推定された所望の信号の歪み又は残差干渉の低減に対する条件を受けるものに基づく。

10

【0032】

音声入力信号を直接信号成分及び環境信号成分に分解するための実施例が提供される。更に、環境信号成分を計算するためのフィルタの導出が提供され、更に、フィルタの応用例における実施例が記載される。

【0033】

いくつかの実施例は、1つのチャンネルよりも多くのチャンネルを有する入力信号を伴う直接・環境方式に従う非ガイド型アップミックスに関する。

20

【0034】

本願に記載の分解の想定される応用例として、同じ数のチャンネルを有する出力信号を入力信号として計算することへの関心が集まっている。この応用例においては、実施例は、分離及び音質の観点で極めて良好な結果を提供するが、それは、直接信号が入力チャンネル間で時間遅延される入力信号に対処できるからである。他の概念、例えば[3]で提案された概念とは対照的に、実施例は、入力信号における直接音がスケーリングのみによってパンニングされる(振幅パンニング)のではなく、各々のチャンネルにおける直接信号間の時間差をも導入することによってパンニングされることを想定している。

【0035】

更に、実施例は、1つ又は2つのチャンネルを有する入力信号しか処理できない先行技術の全ての他の概念(上記を参照)とは対照的に、任意の数のチャンネルを有する入力信号に対する演算を行うことができる。

30

【0036】

実施例の他の利点は、制御パラメータの利用、環境PSD行列の推定、及びフィルタの更なる変更が挙げられ、これについては後述する。

【0037】

いくつかの実施例は、全ての入力音オブジェクトについて一貫性のある環境音を提供する。入力信号を直接音及び環境音に分解したとき、いくつかの実施例では、適切な音声信号処理を用いて環境音特性を適合し、他の実施例では、環境信号成分の代わりに人工的な反響及び他の人工的な環境音を用いる。

40

【0038】

実施例によると、装置は、更に、2つ以上の音声入力チャンネル信号を時間領域から時間周波数領域に変換するように構成された分析フィルタバンクを備えることができる。フィルタ決定部は、時間周波数領域で表される音声入力チャンネル信号に応じて第1のパワースペクトル密度情報及び第2のパワースペクトル密度情報を推定することによってフィルタを決定するように構成することができる。信号処理部は、時間周波数領域で表される2つ以上の音声入力チャンネル信号にフィルタを適用することにより、時間周波数領域で表される1つ以上の音声出力チャンネル信号を生成するように構成することができる。また、装置は、更に、時間周波数領域で表される1つ以上の音声出力チャンネル信号を、時間周波数領域から時間領域に変換するように構成された合成フィルタバンクを備えることができる。

50

【 0 0 3 9 】

更に、2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための方法が提供される。2つ以上の音声入力チャンネル信号の各々は、直接信号部分及び環境信号部分を含む。方法は、

- 第1のパワースペクトル密度情報を推定し第2のパワースペクトル密度情報を推定することによりフィルタを決定するステップと、

- 2つ以上の音声入力チャンネル信号にフィルタを適用することにより前記1つ以上の音声出力チャンネル信号を生成するステップと、を備える。

【 0 0 4 0 】

第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示す。或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。

10

【 0 0 4 1 】

更に、コンピュータ又は信号プロセッサにおいて実行されたときに上述の方法を実現するためのコンピュータプログラムが提供される。

20

【 0 0 4 2 】

以下、本発明の実施例について、図面を参照してより詳細に説明する。

【 図面の簡単な説明 】

【 0 0 4 3 】

【 図 1 】 図 1 は、実施例による2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための装置を示す図である。

【 図 2 】 図 2 は、実施例によるクラシック音楽の5チャンネル録音の分解の入力信号及び出力信号であって、入力信号（左列）、環境出力信号（中列）及び直接出力信号（右列）を示す図である。

30

【 図 3 】 図 3 は、実施例による環境信号推定及び直接信号推定を用いた分解の基本的な概観を示す図である。

【 図 4 】 図 4 は、実施例による直接信号推定を用いた分解の基本的な概観を示す図である。

【 図 5 】 図 5 は、実施例による環境信号推定を用いた分解の基本的な概観を示す図である。

【 図 6 a 】 図 6 a は、別の実施例による装置であって、分析フィルタバンク及び合成フィルタバンクを更に備えた装置を示す図である。

【 図 6 b 】 図 6 b は、更なる実施例による装置であって、直接信号成分の抽出を示し、ブロック A F B は、N個の分析フィルタバンク（各々のチャンネルにつき1つ）の組であり、S F B は、1組の合成フィルタバンクであるものを示す図である。

40

【 0 0 4 4 】

図 1 は、実施例による2つ以上の音声入力チャンネル信号に応じて1つ以上の音声出力チャンネル信号を生成するための装置を示す。2つ以上の音声入力チャンネル信号の各々は、直接信号部分及び環境信号部分を含む。

【 0 0 4 5 】

装置は、第1のパワースペクトル密度情報を推定し第2のパワースペクトル密度情報を推定することによりフィルタを決定するためのフィルタ決定部 1 1 0 を備える。

【 0 0 4 6 】

更に、装置は、2つ以上の音声入力チャンネル信号にフィルタを適用することにより1つ

50

以上の音声出力チャンネル信号を生成するための信号処理部 120 を備える。

【0047】

第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。

【0048】

或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示す。

【0049】

或いは、第1のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の直接信号部分についてのパワースペクトル密度情報を示し、第2のパワースペクトル密度情報は、2つ以上の音声入力チャンネル信号の環境信号部分についてのパワースペクトル密度情報を示す。

【0050】

実施例によっては、音声入力信号を直接信号成分及び環境信号成分に分解するための概念が提供され、これらを音の後生成及び再生に適用することができる。このような信号処理における主な課題は、任意の数の入力チャンネル信号及び全ての可能な入力信号特性について高い音質を維持しながら、高程度の分離を達成することである。本願により提供される実施例は、時間周波数領域におけるマルチチャンネル信号処理に基づくものであり、平均平方誤差の意味での最適解であって、推定された所望の信号の歪み又は残余干渉の低減に対する条件を受けるものが提供される。

【0051】

まず、本発明の実施例が基づく発明概念について説明する。

【0052】

N 個の入力チャンネル信号 $\mathbf{y}_t[n]$ が受信されると想定する。

【0053】

【数3】

$$\mathbf{y}_t[n] = [y_1[n] \cdots y_N[n]]^T \quad (5)$$

【0054】

例えば、 $N \geq 2$ である。本願によって提供される概念の目的は、入力チャンネル信号 $y_1[n] \cdots y_N[n]$ ($= [\mathbf{y}_t[n]]^T$) を、 $\mathbf{d}_t[n] = [d_1[n] \cdots d_N[n]]^T$ で示される N 個の直接信号成分、及び/又は、 $\mathbf{a}_t[n] = [a_1[n] \cdots a_N[n]]^T$ で示される N 個の環境信号成分、に分解することである。この処理は、全ての入力チャンネルに適用することができ、又は、入力信号チャンネルは、別個に処理されるチャンネルのサブセットに分割される。

【0055】

実施例によると、直接信号成分 $d_1[n] \cdots d_N[n]$ のうちの1つ以上、及び/又は、環境信号成分 $a_1[n] \cdots a_N[n]$ のうちの1つ以上を、2つ以上の入力チャンネル信号 $y_1[n] \cdots y_N[n]$ から推定して1つ以上の出力チャンネル信号として、直接信号成分 $d_1[n] \cdots d_N[n]$ 及び/又は環境信号成分 $a_1[n] \cdots a_N[n]$ についての1つ以上の推定値 ($\hat{d}_1[n] \cdots \hat{d}_N[n]$, $\hat{a}_1[n] \cdots \hat{a}_N[n]$) を得る。

【0056】

10

20

30

40

いくつかの実施例において提供される出力の一例を図2に示し、ここで $N=5$ である。1つ以上の音声出力チャンネル信号 $\hat{d}_1[n] \cdots \hat{d}_N[n]$ ($= [\hat{\mathbf{d}}_t[n]]^T$), $\hat{a}_1[n] \cdots \hat{a}_N[n]$ ($= [\hat{\mathbf{a}}_t[n]]^T$) は、図3に示すように、直接信号成分及び環境信号成分を独立に推定することによって得られる。これに代えて、2つの信号のうち的一方 ($\mathbf{d}_t[n]$ 又は $\mathbf{a}_t[n]$) についての推定値 ($\hat{\mathbf{d}}_t[n]$ 又は $\hat{\mathbf{a}}_t[n]$) を計算し、入力信号から第1の結果を減じることによって他方の信号を得る。図4は、まず直接信号成分 $\mathbf{d}_t[n]$ を推定し、それから、入力信号から直接信号の推定値を減じることによって環境信号成分 $\mathbf{a}_t[n]$ を導出する処理を示す。これと同様の理論で、図5のブロック図に示すように、まず、環境信号成分の推定値を導出することができる。

10

【0057】

実施例によると、この処理は、例えば、時間周波数領域で実行することができる。入力音声信号の時間周波数領域表現は、例えば、フィルタバンク（分析フィルタバンク）、例えば短時間フーリエ変換（STFT）を用いて得ることができる。

【0058】

図6aに示す実施例によると、分析フィルタバンク605は、音声入力チャンネル信号 $y_t[n]$ を時間領域から時間周波数領域に変換する。更に、図6aでは、合成フィルタバンク625は、直接信号成分 $\hat{d}[m, 1] \cdots \hat{d}[m, k]$ の推定値を時間周波数領域から時間領域に変換して音声出力チャンネル信号 $\hat{d}_1[n] \cdots \hat{d}_N[n]$ ($= [\hat{\mathbf{d}}_t[n]]^T$) を得る。

20

【0059】

図6aの実施例においては、分析フィルタバンク605は、2つ以上の音声入力チャンネル信号を時間領域から時間周波数領域に変換するように構成される。フィルタ決定部110は、時間周波数領域で表される音声入力チャンネル信号に応じて第1のプワースペクトル密度情報及び第2のプワースペクトル密度情報を推定することによってフィルタを決定するように構成される。信号処理部120は、時間周波数領域で表される2つ以上の音声入力チャンネル信号にフィルタを適用することにより、時間周波数領域で表される1つ以上の音声出力チャンネル信号を生成するように構成される。合成フィルタバンク625は、時間周波数領域で表される1つ以上の音声出力チャンネル信号を、時間周波数領域から時間領域に変換するように構成される。

30

【0060】

時間周波数領域表現は、時間の経過に伴って発展する或る数のサブバンド信号を含む。任意には、隣接するサブバンドを線形に組み合わせてより広いサブバンド信号とすることで計算上の複雑度を低減させることができる。入力信号における各々のサブバンドは、以下に詳細に説明するように個別に処理される。時間領域出力信号は、それぞれフィルタバンク、即ち合成フィルタバンクの逆処理を適用することによって得られる。全ての信号がゼロの平均値を有するものと想定され、時間周波数領域信号は、複雑なランダム変数としてモデル化することができる。

40

【0061】

以下、定義及び想定について説明する。

【0062】

以下の定義は、考案された方法の記載全体を通して用いられる。N個のチャンネルを有するマルチチャンネル入力信号の時間周波数領域表現は、時間インデックス m 及びサブバンドインデックス k , $k = 1 \dots K$ を用いて

50

【 0 0 6 3 】

【 数 4 】

$$\mathbf{y}(m, k) = [Y_1(m, k) Y_2(m, k) \cdots Y_N(m, k)]^T \quad (6)$$

【 0 0 6 4 】

によって与えられ、直接信号成分 $\mathbf{d}(m, k)$ 及び環境信号成分 $\mathbf{a}(m, k)$ の加法的な混合であると想定され、即ち

【 0 0 6 5 】

【 数 5 】

$$\mathbf{y}(m, k) = \mathbf{d}(m, k) + \mathbf{a}(m, k) \quad (7)$$

であり、ここで

【 0 0 6 6 】

【 数 6 】

$$\mathbf{d}(m, k) = [D_1(m, k) D_2(m, k) \cdots D_N(m, k)]^T \quad (8)$$

$$\mathbf{a}(m, k) = [A_1(m, k) A_2(m, k) \cdots A_N(m, k)]^T \quad (9)$$

であり、ここで、 $D_i(m, k)$ は直接成分を示し、 $A_i(m, k)$ は、 i 番目のチャネルにおける環境成分を示す。

【 0 0 6 7 】

直接・環境分解の目的は、 $\mathbf{d}(m, k)$ 及び $\mathbf{a}(m, k)$ を推定することである。出力信号は、フィルタ行列 $\mathbf{H}_D(m, k)$ 若しくは $\mathbf{H}_A(m, k)$ 又はその両方を用いて計算される。フィルタ行列は、 $N \times N$ のサイズを有し、複素値であり、又は、いくつかの実施例では、例えば実数値であり得る。直接信号成分及び環境信号成分の N チャネル信号の推定値は、

【 0 0 6 8 】

【 数 7 】

$$\hat{\mathbf{d}}(m, k) = \mathbf{H}_D^H(m, k) \mathbf{y}(m, k) \quad (10)$$

$$\hat{\mathbf{a}}(m, k) = \mathbf{H}_A^H(m, k) \mathbf{y}(m, k) \quad (11)$$

から得ることができる。これに代えて、1つのフィルタ行列だけを用いても良く、図4に示す減算は、

【 0 0 6 9 】

【 数 8 】

$$\hat{\mathbf{d}}(m, k) = \mathbf{H}_D^H(m, k) \mathbf{y}(m, k) \quad (12)$$

$$\hat{\mathbf{a}}(m, k) = [\mathbf{I} - \mathbf{H}_D(m, k)]^H \mathbf{y}(m, k) \quad (13)$$

【 0 0 7 0 】

と表すことができ、ここで、 \mathbf{I} は、 $N \times N$ のサイズの単位行列であり、又は、図5に示すように、

【 0 0 7 1 】

10

20

30

40

【数 9】

$$\hat{\mathbf{a}}(m, k) = \mathbf{H}_A^H(m, k) \mathbf{y}(m, k) \quad (14)$$

$$\hat{\mathbf{d}}(m, k) = [\mathbf{I} - \mathbf{H}_A(m, k)]^H \mathbf{y}(m, k) \quad (15)$$

【0072】

と表すことができる。ここで、上付きの^Hは、行列又はベクトルの共役転置を示す。フィルタ行列 $\mathbf{H}_D(m, k)$ は、直接信号 $\hat{\mathbf{d}}(m, k)$ についての推定値を計算するために用いられる。フィルタ行列 $\mathbf{H}_A(m, k)$ は、環境信号 $\hat{\mathbf{a}}(m, k)$ についての推定値を計算するために用いられる。

10

【0073】

上記の式(10)～(15)においては、 $\mathbf{y}(m, k)$ は、2つ以上の音声入力チャンネル信号を示す。 $\hat{\mathbf{a}}(m, k)$ は、環境信号部分の推定値を示し、 $\hat{\mathbf{d}}(m, k)$ は、音声入力チャンネル信号の直接信号部分の推定値を示す。 $\hat{\mathbf{a}}(m, k)$ 及び/若しくは $\hat{\mathbf{d}}(m, k)$ 、又は $\hat{\mathbf{a}}(m, k)$ 及び/若しくは $\hat{\mathbf{d}}(m, k)$ における1つ以上のベクトル成分は、1つ以上の音声出力チャンネル信号であり得る。

20

【0074】

式(10)、(11)、(12)、(13)、(14)、(15)のうちの1つ、いくつか又は全てを図1、6aの信号処理部120によって用いて、図1、6aのフィルタを音声入力チャンネル信号に適用することができる。図1、6aのフィルタは、例えば、 $\mathbf{H}_D(m, k)$ 、 $\mathbf{H}_A(m, k)$ 、 $\mathbf{H}_D^H(m, k)$ 、 $\mathbf{H}_A^H(m, k)$ 、 $[\mathbf{I} - \mathbf{H}_D(m, k)]$ 又は $[\mathbf{I} - \mathbf{H}_A(m, k)]$ であり得る。しかし、他の実施例では、フィルタ決定部110によって決定され、信号処理部120によって用いられるフィルタは、行列でなく別の種類のフィルタであっても良い。例えば、他の実施例では、フィルタは、フィルタを定義する1つ以上のベクトルを含むことができる。更なる実施例では、フィルタは、フィルタを定義する複数の係数を含むことができる。

30

【0075】

フィルタ行列は、以下に説明するように信号統計の推定値から計算される。

【0076】

具体的には、フィルタ決定部110は、第1のパワースペクトル密度(PSD)情報及び第2のPSD情報を推定することによってフィルタを決定するように構成される。

【0077】

【数10】

$$\phi_{X_i X_j}(m, k) = E \{ X_i(m, k) X_j^*(m, k) \} \quad (16)$$

40

を定義し、ここで、 $E \{ \cdot \}$ は、期待値演算子であり、 X^* は、 X の複素共役を示す。 $i = j$ の場合にはPSDが、 $i \neq j$ の場合にはクロスPSDが得られる。

【0078】

$\mathbf{y}(m, k)$ 、 $\mathbf{d}(m, k)$ 、 $\mathbf{a}(m, k)$ についての共分散行列は、

【0079】

【数 1 1】

$$\Phi_y(m, k) = E \{ \mathbf{y}(m, k) \mathbf{y}^H(m, k) \} \quad (17)$$

$$\Phi_d(m, k) = E \{ \mathbf{d}(m, k) \mathbf{d}^H(m, k) \} \quad (18)$$

$$\Phi_a(m, k) = E \{ \mathbf{a}(m, k) \mathbf{a}^H(m, k) \} \quad (19)$$

【0080】

である。共分散行列 $\Phi_y(m, k)$, $\Phi_d(m, k)$, $\Phi_a(m, k)$ は、主対角線上の全てのチャンネルについての PSD の推定値を含み、対角線から外れた要素は、それぞれのチャンネル信号のクロス PSD の推定値である。従って、行列 $\Phi_y(m, k)$, $\Phi_d(m, k)$, $\Phi_a(m, k)$ の各々は、パワースペクトル密度情報の推定値を表す。

10

【0081】

式 (17) ~ (19) において、 $\Phi_y(m, k)$ は、2つ以上の音声入力チャンネル信号についてのパワースペクトル密度情報を示す。 $\Phi_d(m, k)$ は、2つ以上の音声入力チャンネル信号の直接信号成分についてのパワースペクトル密度情報を示す。 $\Phi_a(m, k)$ は、2つ以上の音声入力チャンネル信号の環境信号成分についてのパワースペクトル密度情報を示す。

20

【0082】

式 (17) , (18) , (19) の行列 $\Phi_y(m, k)$, $\Phi_d(m, k)$ 及び $\Phi_a(m, k)$ の各々は、パワースペクトル密度情報と見做すことができる。しかし、他の実施例では、第1及び第2のパワースペクトル密度情報は行列ではなく、任意の他の種類の好適な形式で表すことができる。例えば、実施例によると、第1及び/又は第2のパワースペクトル密度情報は、1つ以上のベクトルとして表すことができる。更なる実施例においては、第1及び/又は第2のパワースペクトル密度情報は、複数の係数として表すことができる。

【0083】

以下のように想定する。

【0084】

・ $D_i(m, k)$ 及び $A_i(m, k)$ は、相互に無相関である。

30

【0085】

【数 1 2】

$$E \{ D_i(m, k) A_j^*(m, k) \} = 0 \quad \forall i, j$$

【0086】

・ $A_i(m, k)$ 及び $A_j(m, k)$ は、相互に無相関である。

【0087】

【数 1 3】

$$E \{ A_i(m, k) A_j^*(m, k) \} = 0 \quad \forall i \neq j$$

40

【0088】

・ 環境パワーは、全てのチャンネルにおいて等しい。

【0089】

【数 1 4】

$$E \{ A_i(m, k) A_j^*(m, k) \} = \phi_A(m, k) \quad \forall i=j$$

【0090】

その結果、

【0091】

50

【数 15】

$$\Phi_y(m, k) = \Phi_d(m, k) + \Phi_a(m, k) \quad (20)$$

$$\Phi_a(m, k) = \phi_A(m, k) \mathbf{I}_{N \times N} \quad (21)$$

が成立する。

【0092】

式(20)の結果、行列 $\Phi_y(m, k)$ 、 $\Phi_d(m, k)$ 及び $\Phi_a(m, k)$ のうちの2つの行列が決定された時、行列のうちの3番目のものが直ちに利用可能となる。更なる結果として、

−2つ以上の音声入力チャネル信号についてのパワースペクトル密度情報、及び、2つ以上の音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報、又は−2つ以上の音声入力チャネル信号についてのパワースペクトル密度情報、及び、2つ以上の音声入力チャネル信号の直接信号部分についてのパワースペクトル密度情報、又は−2つ以上の音声入力チャネル信号の直接信号部分についてのパワースペクトル密度情報、及び、2つ以上の音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報を決定するだけで十分となるが、それは、上記3種類のPSD情報が行列として表されるのではなく、別の種類の好適な表現、例えば1つ以上のベクトル、又は例えば複数の係数等として利用可能な場合、3種類のパワースペクトル密度情報の関係から(例えば式(20))により、又は3種類のパワースペクトル密度情報(完全な入力信号のPSD、環境成分のPSD及び直接成分のPSD)の関係についての任意の他の再公式化により3番目の(推定されていない)パワースペクトル密度情報が直ちに明らかとなるからである。

10

20

【0093】

考案された方法の性能を評価するために、以下の信号を定義する。

【0094】

・直接信号歪み：

【0095】

【数16】

$$\mathbf{q}_d(m, k) = [\mathbf{I} - \mathbf{H}_D(m, k)]^H \mathbf{d}(m, k)$$

【0096】

・残差環境信号：

【0097】

【数17】

$$\mathbf{r}_a(m, k) = \mathbf{H}_D^H(m, k) \mathbf{a}(m, k)$$

【0098】

・環境信号歪み：

【0099】

【数18】

$$\mathbf{q}_a(m, k) = [\mathbf{I} - \mathbf{H}_A(m, k)]^H \mathbf{a}(m, k)$$

【0100】

・残差直接信号：

【0101】

【数19】

$$\mathbf{r}_d(m, k) = \mathbf{H}_A^H(m, k) \mathbf{d}(m, k)$$

【0102】

以下においては、フィルタ行列の導出について図4及び図5に従って説明する。読みやすくするために、サブバンドインデックス及び時間インデックスは省略する。

30

40

50

【 0 1 0 3 】

最初に、直接信号成分の推定についての実施例について説明する。

【 0 1 0 4 】

考案された方法の理論は、直接信号歪み \mathbf{q}_d を条件付きとしながら残差環境信号 \mathbf{r}_a が最小になるようにフィルタを計算することである。これは、条件付き最適化問題につながる。

【 0 1 0 5 】

【 数 2 0 】

$$\mathbf{H}_D(\beta_i) = \arg \min_{\mathbf{H}_D} E\{\|\mathbf{r}_a\|^2\} \quad (22)$$

$$\text{但し、} E\{\|\mathbf{q}_d\|^2\} \leq \sigma_{d, \max}^2$$

【 0 1 0 6 】

ここで、 $\sigma_{d, \max}^2$ は、最大の許容可能な直接信号歪みである。解は、

【 0 1 0 7 】

【 数 2 1 】

$$\mathbf{H}_D(\beta_i) = [\Phi_d + \beta_i \Phi_a]^{-1} \Phi_d \quad (23)$$

によって与えられる。i 番目のチャネルの直接出力信号を計算するためのフィルタは、

【 0 1 0 8 】

【 数 2 2 】

$$\mathbf{h}_{D,i}(\beta_i) = [\Phi_d + \beta_i \Phi_a]^{-1} \Phi_d \mathbf{u}_i \quad (24)$$

に等しく、ここで \mathbf{u}_i は、i 番目の位置における 1 を伴う長さ N の零ベクトルである。パラメータ β_i により、残差環境信号の低減と環境信号歪みとの間のトレードオフが可能となる。図 4 に示すシステムについては、直接出力信号におけるより低い残差環境レベルが、環境出力信号におけるより高い環境レベルにつながる。より少ない直接信号歪みは、環境出力信号における直接信号成分の良好な減衰につながる。時間及び周波数に依存するパラメータ β_i は、各々のチャネルについて別個に設定することができ、入力信号又は以下のように導出された信号によって制御することができる。

【 0 1 0 9 】

これと類似の解を得るには、条件付き最適化問題を

【 0 1 1 0 】

【 数 2 3 】

$$\mathbf{H}_D(\beta_i) = \arg \min_{\mathbf{H}_D} E\{\|\mathbf{q}_d\|^2\} \quad (25)$$

$$\text{但し、} E\{\|\mathbf{r}_a\|^2\} \leq \sigma_{a, \max}^2$$

【 0 1 1 1 】

として公式化することができる。 Φ_d が階数 1 のものである場合、i 番目のチャネル信号についての $\sigma_{d, \max}^2$ と β_i との関係は、

【 0 1 1 2 】

【数 2 4】

$$\sigma_{d, \max}^2 = \left(\frac{\beta_i}{\beta_i + \lambda} \right)^2 \phi_{D_i D_i} \quad (26)$$

として導出され、ここで $D_i D_i$ は、 i 番目のチャンネルにおける直接信号の PSD であり、 $\phi_{D_i D_i}$ はマルチチャンネル直接対環境比 (DAR)

【0 1 1 3】

【数 2 5】

$$\lambda = \text{tr} \{ \Phi_a^{-1} \Phi_d \} \quad (27)$$

$$= \text{tr} \{ \Phi_a^{-1} \Phi_y \} - N \quad (28)$$

10

であり、ここで、正方行列 A のトレースは、主対角線上の要素の和に等しい。

【0 1 1 4】

【数 2 6】

$$\text{tr} \{ \mathbf{K} \} = \sum_{i=1}^N k_{ii} \quad (m, k)$$

20

【0 1 1 5】

なお、 Φ_d が階数 1 のものであるという記載は単なる仮定である。この仮定が現実に真であるか否かに関わらず、本発明の実施例においては、 Φ_d の正確な結果が現実には Φ_d が階数 1 のものではない状況でも上述の式 (26), (27), (28) が用いられる。このような状況において、 Φ_d が階数 1 のものであるという仮定が現実には真でない場合でも本発明の実施例は良好な結果をもたらす。

【0 1 1 6】

以下、環境信号成分の推定について説明する。

【0 1 1 7】

考案された方法の理論は、環境信号歪み q_a を条件付きとしながら残差直接信号 r_d が 30
最小になるようにフィルタを計算することである。これは、条件付き最適化問題

【0 1 1 8】

【数 2 7】

$$\mathbf{H}_A(\beta_i) = \arg \min_{\mathbf{H}_A} E \{ \|\mathbf{r}_d\|^2 \} \quad (29)$$

$$\text{但し、} E \{ \|\mathbf{q}_a\|^2 \} \leq \sigma_{a, \max}^2$$

【0 1 1 9】

につながり、ここで、 $\sigma_{a, \max}^2$ は、最大の許容可能な直接信号歪みである。解は、 40

【0 1 2 0】

【数 2 8】

$$\mathbf{H}_A(\beta_i) = [\beta_i \Phi_d + \Phi_a]^{-1} \Phi_a \quad (30)$$

によって与えられる。 i 番目のチャンネルの環境出力信号を計算するためのフィルタは、

【0 1 2 1】

【数 2 9】

$$\mathbf{h}_{A,i}(\beta_i) = [\beta_i \Phi_d + \Phi_a]^{-1} \Phi_a \mathbf{u}_i \quad (31)$$

に等しい。

【0 1 2 2】

以下、本発明の概念を実現する実施例を詳細に記載する。

【0 1 2 3】

例えば、パワースペクトル密度情報を決定するために、音声入力チャネル信号 Φ_y のPSD行列を、短時間の移動平均又は再帰的平均を用いて直接推定することができる。環境PSD行列 Φ_a を、例えば以下に記載のように推定することができる。すると、直接PSD行列 Φ_d を、例えば式(20)を用いて得ることができる。

10

【0 1 2 4】

以下においても、各々のサブバンド(単一の直接源)において一時に1つ以下の直接音源がアクティブであり、従って Φ_d が階数1のものであると想定する。

【0 1 2 5】

なお、1つ以下の直接音源がアクティブであり、 Φ_d が階数1のものであるという記載は単なる仮定である。これらの仮定が現実には真であるか否かに関わらず、本発明の実施例においては、現実には1つよりも多くの直接音源がアクティブである状況でも、更に、 Φ_d の正確な結果が現実には Φ_d が階数1のものではない場合でも以下の式、特に式(32)、(33)が用いられる。このような状況において、1つ以下の直接音源がアクティブであり、 Φ_d が階数1のものであるという仮定が現実には真でない場合でも本発明の実施例は良好な結果をもたらす。

20

【0 1 2 6】

従って、1つ以下の直接音源がアクティブであり、 Φ_d が階数1のものであると仮定すると、式(23)を、

【0 1 2 7】

【数 3 0】

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1} \Phi_d}{\beta_i + \lambda} \quad (32)$$

$$\frac{\Phi_a^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33)$$

と書き替えることができる。式(33)は、式(22)についての条件付き最適化問題についての解をもたらす。

30

【0 1 2 8】

上述の式(32)、(33)において、 Φ_a^{-1} は、 Φ_a の逆行列である。 Φ_a^{-1} は、2つ以上の音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報をも示すことは明らかである。

40

【0 1 2 9】

$\mathbf{H}_D(\beta_i)$ を決定するために、 Φ_a^{-1} 及び Φ_d を決定する必要がある。 Φ_a が利用可能である場合、 Φ_a^{-1} を直ちに決定することができる。 λ は、式(27)、(28)に従って定義され、その値は、 Φ_a^{-1} 及び Φ_d が利用可能である場合に利用可能である。 Φ_a^{-1} 、 Φ_d 及び λ を決定する他に、 β_i についての好適な値を選択する必要がある。

【0 1 3 0】

更に、式(33)を再公式化する(式(20)を参照)ことによって、

50

【 0 1 3 1 】

【 数 3 1 】

$$\mathbf{H}_D(\beta_i) = \frac{(\Phi_y - \Phi_d)^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33a)$$

【 0 1 3 2 】

となるようにすることができ、このため、音声入力チャネル信号についてのPSD情報 Φ_y 及び音声入力チャネル信号の直接信号部分についてのPSD情報 Φ_d のみを決定しなければならない。

10

【 0 1 3 3 】

更に、式(33)を再公式化する(式(20)を参照)ことによって、

【 0 1 3 4 】

【 数 3 2 】

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33b)$$

【 0 1 3 5 】

となるようにすることができ、このため、音声入力チャネル信号の環境信号部分についてのPSD情報 Φ_a^{-1} 及び音声入力チャネル信号の直接信号部分についてのPSD情報 Φ_d のみを決定しなければならない。

20

【 0 1 3 6 】

更に、式(33)を再公式化することによって、

【 0 1 3 7 】

【 数 3 3 】

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33c)$$

30

【 0 1 3 8 】

となるようにすることができ、こうして $\mathbf{H}_A(\beta_i)$ が決定される。

【 0 1 3 9 】

式(33c)によって、式(29)の条件付き最適化問題についての解が得られる。

【 0 1 4 0 】

これと同様に、式(33a)、(33b)を再公式化して、

【 0 1 4 1 】

【 数 3 4 】

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{(\Phi_y - \Phi_d)^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33d)$$

40

又は

【 0 1 4 2 】

【数 3 5】

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33e)$$

とすることができる。

【0 1 4 3】

なお、 $\mathbf{H}_D(\beta_i)$ を決定することによって、フィルタ $\mathbf{H}_A(\beta_i)$ は、 $\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \mathbf{H}_D(\beta_i)$ として直ちに利用可能である。

10

【0 1 4 4】

更に、 $\mathbf{H}_A(\beta_i)$ を決定することによって、フィルタ $\mathbf{H}_D(\beta_i)$ は、 $\mathbf{H}_D(\beta_i) = \mathbf{I}_{N \times N} - \mathbf{H}_A(\beta_i)$ として直ちに利用可能である。

【0 1 4 5】

上述のように、例えば式 (33) に従って $\mathbf{H}_D(\beta_i)$ を決定するために、 Φ_y 及び Φ_a を決定することができる。

【0 1 4 6】

音声信号 $\Phi_y(m, k)$ の PSD 行列を、例えば、再帰的平均

20

【0 1 4 7】

【数 3 6】

$$\Phi_y(m, k) = (1 - \alpha) \mathbf{y}(m, k) \mathbf{y}^H(m, k) + \alpha \Phi_y(m-1, k) \quad (34a)$$

を用いて、例えば直接推定することができ、ここで、 α は、積分時間を決定するフィルタ係数であり、又は、例えば、短時間移動重み付き平均

【0 1 4 8】

【数 3 7】

$$\Phi_y(m, k) = b_0 \cdot \mathbf{y}(m, k) \mathbf{y}^H(m, k) + b_1 \cdot \mathbf{y}(m-1, k) \mathbf{y}^H(m-1, k) + b_2 \cdot \mathbf{y}(m-2, k) \mathbf{y}^H(m-2, k) + \dots + b_L \cdot \mathbf{y}(m-L, k) \mathbf{y}^H(m-L, k) \quad (34b)$$

30

を用いて、例えば直接推定することができ、ここで、 L は、例えば PSD の計算に用いられる過去の値の数であり、 $b_0 \dots b_L$ は、例えば [0 1] の範囲内のフィルタ係数であり (例えば 0 フィルタ係数 1)、又は、例えば、式 (34b) に従い、ただし全ての $i = 0 \dots L$ について

【0 1 4 9】

【数 3 8】

$$b_i = \frac{1}{L+1}$$

40

による短時間移動平均を用いて、例えば直接推定することができる。

【0 1 5 0】

次に、実施例による環境 PSD 行列 Φ_a の推定について説明する。

【0 1 5 1】

環境 PSD 行列 Φ_a は、

50

【 0 1 5 2 】

【 数 3 9 】

$$\Phi_a = \hat{\phi}_A \mathbf{I}_{N \times N} \quad (35)$$

【 0 1 5 3 】

によって与えられ、ここで、 $\mathbf{I}_{N \times N}$ は、 $N \times N$ のサイズの単位行列である。 $\hat{\phi}_A$ は、例えば数である。

【 0 1 5 4 】

10

実施例による1つの解は、例えば、一定の値を用いて、式(21)を用いて $\hat{\phi}_A$ を実数である正の定数 ε に設定することによって得られる。この方策の利点は、計算上の複雑度を無視できることである。

【 0 1 5 5 】

実施例においては、フィルタ決定部110は、2つ以上の音声入力チャネル信号に応じて $\hat{\phi}_A$ を決定するように構成される。

【 0 1 5 6 】

極めて低い計算上の複雑度における1つの選択肢は、実施例によると、入力パワーの分数を用いて、 $\hat{\phi}_A$ を入力PSDの平均値若しくは最小値又はその分数に設定することであり、例えば

20

【 0 1 5 7 】

【 数 4 0 】

$$\hat{\phi}_A = \frac{g}{N} \text{tr}\{\Phi_y\} \quad (36)$$

であり、ここで、パラメータ g は、環境パワーの量を制御し、 $0 < g < 1$ である。

【 0 1 5 8 】

30

更なる実施例によると、推定は、算術平均に基づいて行われる。式(20)及び式(21)へつながる仮定の場合、

【 0 1 5 9 】

【 数 4 1 】

$$\hat{\phi}_A = \frac{1}{N} \text{tr}\{\Phi_y - \Phi_d\} \quad (37)$$

$$= \frac{1}{N} (\text{tr}\{\Phi_y\} - \text{tr}\{\Phi_d\}) \quad (38)$$

【 0 1 6 0 】

40

を用いてPSD $\hat{\phi}_A$ を計算できることを示すことができる。 $\text{tr}\{\Phi_y\}$ を、例えば式(34a)の再帰的積分、又は、例えば式(34b)の短時間移動重み付き平均を用いて直接計算できるのに対し、 $\text{tr}\{\Phi_d\}$ は、

【 0 1 6 1 】

【数 4 2】

$$\text{tr}\{\Phi_d\} = \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^N [(\phi_{Y_i Y_i} - \phi_{Y_j Y_j})^2] \quad (39)$$

$$+ 4\text{Re}\{\phi_{Y_i Y_j}\}^2 \quad (40)$$

として推定される。

【0 1 6 2】

これに代えて、 $N > 2$ の場合、2つの入力チャネル信号を選択して1組の信号チャネルのみについて $\hat{\phi}_A(m, k)$ を推定することによって PSD $\hat{\phi}_A(m, k)$ を計算することができる。より正確な結果を得るには、この手続きを1組よりも多くの入力チャネル信号に適用して、その結果を、例えば全体の推定値を平均することで組み合わせる。類似の環境パワーを有するチャネルについてのアプリアリを利用して、例えば5.1録音の後方チャネル全て及び前方チャネル全てにおいて別個に環境パワーを推定することによって、サブセットを選択することができる。

10

【0 1 6 3】

更に、式(20)、(35)から、

【0 1 6 4】

【数 4 3】

$$\Phi_d = \Phi_y - \hat{\phi}_A \mathbf{I}_{N \times N} \quad (35a)$$

20

が得られる。

【0 1 6 5】

いくつかの実施例によると、(例えば式(35)又は式(36)に従い、又は式(37)～(40)に従い) $\hat{\phi}_A$ を決定することにより、且つ、式(35a)を用いて音声入力チャネル信号の環境信号部分についてのパワースペクトル密度情報を得ることによって Φ_d が決定される。そして、例えば式(33a)を用いることで $\mathbf{H}_D(\beta_i)$ を決定することができる。

【0 1 6 6】

以下、パラメータ β_i についての選択について検討する。

【0 1 6 7】

β_i は、トレードオフパラメータである。トレードオフパラメータ β_i は、数である。

【0 1 6 8】

いくつかの実施例では、全ての音声入力チャネル信号について有効なただ1つのトレードオフパラメータ β_i を決定し、このトレードオフパラメータを音声入力チャネル信号のトレードオフ情報と見做す。

【0 1 6 9】

他の実施例では、2つ以上の音声入力チャネル信号の各々について1つのトレードオフパラメータ β_i を決定し、音声入力チャネル信号のこれら2つ以上のトレードオフパラメータが合わさってトレードオフ情報を構成する。

40

【0 1 7 0】

更なる実施例においては、トレードオフ情報は、パラメータとして表されるのではなく、異なる種類の好適なフォーマットとして表されることができる。

【0 1 7 1】

上述のように、パラメータ β_i によって、環境信号の低減と直接信号の歪みとの間のトレードオフが可能となる。これは一定のものとして選択されるか、又は図6bに示すように信号依存のものとして選択され得る。

【0 1 7 2】

図6 bは、更なる実施例による装置を示す。装置は、音声入力チャネル信号 $\mathbf{y}_t [n]$ を時間領域から時間周波数領域に変換するための分析フィルタバンク 605 を備える。更に、装置は、1つ以上の音声出力チャネル信号（例えば、音声入力チャネル信号の推定直接信号成分 $\hat{d}_1 [n] \cdots \hat{d}_N [n]$ ）を時間周波数領域から時間領域に変換するための合成フィルタバンク 625 を備える。

【0173】

複数K個のベータ決定部 1111 \cdots 11K1（「ベータを計算」）は、パラメータ β_i を決定する。更に、複数K個のサブフィルタ計算部 1112 \cdots 11K2 は、サブフィルタ $\mathbf{H}_D^H(m, 1) \cdots \mathbf{H}_D^H(m, K)$ を決定する。複数のベータ決定部 1111 \cdots 11K1 及び複数のサブフィルタ計算部 1112 \cdots 11K2 は、特定の実施例によると、合わさって図1及び図6 aのフィルタ決定部 110 を構成する。複数のサブフィルタ $\mathbf{H}_D^H(m, 1) \cdots \mathbf{H}_D^H(m, K)$ は、特定の実施例によると、合わさって図1及び図6 aのフィルタを構成する。

10

【0174】

更に、図6 bは、複数の信号サブ処理部 121 \cdots 12K を示し、各々の信号サブ処理部 121 \cdots 12K は、サブフィルタ $\mathbf{H}_D^H(m, 1) \cdots \mathbf{H}_D^H(m, K)$ のうちの1つを音声入力チャネル信号のうちの1つに適用して音声出力チャネル信号のうちの1つを得るように構成される。複数の信号サブ処理部 121 \cdots 12K は、特定の実施例によると、合わさって図1および図6 aの信号処理部を構成する。

【0175】

以下、信号分析を用いてパラメータ β_i を制御するための異なった使用例について説明する。

20

【0176】

最初に、トランジェント信号について検討する。

【0177】

実施例によると、フィルタ決定部 110 は、2つ以上の音声入力チャネル信号のうちの少なくとも1つにトランジェントが存在するか否かに応じてトレードオフ情報（ β_i, j ）を決定するように構成される。

【0178】

入力PSD行列の推定は、定常的な信号について最も有効である。一方で、トランジェント入力信号の分解の結果、トランジェント信号成分が環境出力信号へ漏れる恐れがある。非定常性の程度又はトランジェントの存在確率に関して信号分析を用いて β_i を制御して、信号がトランジェントを含むときに β_i が小さく、一様に保たれた部分で大きくなるようにすることで、フィルタ $\mathbf{H}_D(\beta_i)$ の適用の際の出力信号がより一貫性を持つようになる。非定常性の程度又はトランジェントの存在確率に関して信号分析を用いて β_i を制御して、信号がトランジェントを含むときに β_i が大きく、一様に保たれた部分で小さくなるようにすることで、フィルタ $\mathbf{H}_A(\beta_i)$ の適用の際の出力信号がより一貫性を持つようになる。

30

【0179】

次に、望ましくない環境信号について検討する。

40

【0180】

実施例においては、フィルタ決定部 110 は、2つ以上の音声入力チャネル信号のうちの1つが送信される少なくとも1つの信号チャネルにおける加算ノイズの存在に応じてトレードオフ情報（ β_i, j ）を決定するように構成される。

【0181】

提案される方法は、環境信号成分の性質に関わらず入力信号を分解する。ノイズの多い信号チャネルを介して入力信号が送信された場合、望ましくない加算ノイズ存在の確率を推定して β_i を制御することで出力DAR（直接対環境比）が増加するようにすることが有利である。

【0182】

50

次に、出力信号のレベルの制御について記載する。

【0183】

出力信号のレベルを制御するために、 β_i を i 番目のチャンネルについて別個に設定することができる。 i 番目のチャンネルの環境出力信号を計算するためのフィルタは、式(31)によって与えられる。

【0184】

任意の2つのチャンネルについて、所与の β_i につき、 i 番目及び j 番目の出力チャンネルにおける残差環境信号 $\mathbf{r}_{a,i}$ 及び $\mathbf{r}_{a,j}$ のPSDが等しくなる、即ち、

【0185】

【数44】

$$\mathbf{h}_{A,i}^H(\beta_i) \Phi_{a,A,i}(\beta_i) = \mathbf{h}_{A,j}^H(\beta_j) \Phi_{a,A,j}(\beta_j) \quad (41)$$

又は

【0186】

【数45】

$$\begin{aligned} (\mathbf{u}_i - \mathbf{h}_{D,i}(\beta_i))^H \Phi_a(\mathbf{u}_i - \mathbf{h}_{D,i}(\beta_i)) \\ = (\mathbf{u}_j - \mathbf{h}_{D,j}(\beta_j))^H \Phi_a(\mathbf{u}_j - \mathbf{h}_{D,j}(\beta_j)) \end{aligned} \quad (42)$$

となるように β_i を計算することができる。

【0187】

これに代えて、出力環境信号 $\hat{\mathbf{a}}_i$ 及び $\hat{\mathbf{a}}_j$ のPSDが全ての i 及び j の組について等しくなるように β_i を計算することができる。

【0188】

次に、パンニング情報の使用について検討する。

【0189】

入力チャンネルが2つある場合、パンニング情報は、サブバンドごとの両方のチャンネル間のレベル差を定量化する。パンニング情報を適用して β_i を制御することによって、出力信号の知覚幅を制御することができる。

【0190】

以下、出力環境チャンネル信号の等化について検討する。

【0191】

記載された処理は、全ての出力環境チャンネル信号のサブバンドパワーが等しくなることを確実にするものではない。全ての出力環境チャンネル信号のサブバンドパワーが等しくなることを確実にするために、上述のフィルタ \mathbf{H}_D を用いた実施例について以下に記載のようにフィルタに変更を加える。(主対角線上の各チャンネルの自動PSDを含む)環境出力信号の共分散行列は、

【0192】

【数46】

$$\Phi_{\hat{\mathbf{a}}} = (\mathbf{I} - \mathbf{H}_D)^H \Phi_y (\mathbf{I} - \mathbf{H}_D) \quad (43)$$

として得ることができる。

【0193】

全ての出力環境チャンネルのPSDが等しくなることを確実にするために、フィルタ \mathbf{H}_D を $\tilde{\mathbf{H}}_D$ に代える。

【0194】

10

20

30

40

【数 4 7】

$$\widetilde{\mathbf{H}}_D = \mathbf{I} - \mathbf{G}(\mathbf{I} - \mathbf{H}_D) = \mathbf{I} - \mathbf{G} + \mathbf{G}\mathbf{H}_D \quad (44)$$

【0 1 9 5】

ここで、 \mathbf{G} は、主対角線上の要素が

【0 1 9 6】

【数 4 8】

$$g_{ii} = \sqrt{\frac{\text{tr}\{\Phi_{\hat{a}_i}\}}{N \phi_{\hat{A}_i, \hat{A}_i}}}, \quad 1 \leq i \leq N \quad (45)$$

10

となる対角行列である。

【0 1 9 7】

上述のようにフィルタ \mathbf{H}_A を用いた実施例については、（主対角線上の各チャネルの自動PSDを含む）環境出力信号の共分散行列を、

【0 1 9 8】

【数 4 9】

$$\Phi_{\hat{a}} = \mathbf{H}_A^H \Phi_y \mathbf{H}_A \quad (46)$$

20

として得ることができる。

【0 1 9 9】

全ての出力環境チャネルのPSDが等しくなることを確実にするために、フィルタ \mathbf{H}_A をフィルタ $\widetilde{\mathbf{H}}_A$ に代える。

【0 2 0 0】

【数 5 0】

$$\widetilde{\mathbf{H}}_A = \mathbf{G}\mathbf{H}_A \quad (47)$$

30

【0 2 0 1】

装置の文脈でいくつかの局面を記載したが、これらの局面は対応の方法の記載をも表すものであり、ブロック又はデバイスは、方法ステップ又は方法ステップの特徴に対応することは明らかである。同様に、方法ステップの文脈で記載した局面は、対応する装置の対応するブロック若しくは項目又は特徴の記載をも表す。

【0 2 0 2】

本発明の分解された信号は、デジタル記憶媒体で記憶することができ、又は、無線伝送媒体又はインターネットのような有線伝送媒体のような伝送媒体、で送信することができる。

40

【0 2 0 3】

特定の実現要件に応じて、本発明の実施例は、ハードウェア又はソフトウェアによって実現され得る。その実現は、デジタル記憶媒体、例えばフロッピーディスク、DVD、CD、ROM、PROM、EPROM、EEPROM又はフラッシュメモリであって、電子的に読み出し可能な制御信号を格納しており、プログラム可能なコンピュータシステムと協働する（又は協働可能である）ことによりそれぞれの方法が実行されるようにするものを用いて実行され得る。

【0 2 0 4】

本発明のいくつかの実施例は、プログラム可能なコンピュータシステムと協働可能であ

50

ることによって本願明細書に記載の方法の1つが実行されるようにする、電子的に読み出し可能な制御信号を有する非一時的データキャリアを含む。

【0205】

一般的には、本発明の実施例は、プログラムコードを有するコンピュータプログラム製品であって、このコンピュータプログラム製品がコンピュータにおいて実行されるときに上記プログラムコードが上記方法の1つを実行するように動作するものとして実現される。プログラムコードは、例えば、機械読み取り可能キャリアに格納され得る。

【0206】

他の実施例は、機械読み取り可能キャリアに格納された、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを含む。

10

【0207】

従って、換言すると、本発明の方法の一実施例は、コンピュータプログラムであって、このコンピュータプログラムがコンピュータにおいて実行されるときに、本願明細書に記載の方法の1つを実行するためのプログラムコードを有するものである。

【0208】

従って、本発明の方法の更なる実施例は、データキャリア（又はデジタル記憶媒体若しくはコンピュータ読み取り可能媒体）であって、そこに記録された、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを含むものである。

【0209】

従って、本発明の方法の更なる実施例は、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを表すデータストリーム又は信号シーケンスである。データストリーム又は信号シーケンスは、例えば、インターネットを介したデータ通信接続を介して転送されるように構成され得る。

20

【0210】

更なる実施例は、本願明細書に記載の方法の1つを実行するように構成又は適合された処理手段、例えばコンピュータ又はプログラム可能論理装置を含む。

【0211】

更なる実施例は、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムをインストールしたコンピュータを含む。

【0212】

いくつかの実施例においては、プログラム可能論理装置（例えば、フィールドプログラマブルゲートアレイ）を用いて、本願明細書に記載の方法におけるいくつか又は全ての機能を実行しても良い。いくつかの実施例においては、フィールドプログラマブルゲートアレイは、マイクロプロセッサと協働して、本願明細書に記載の方法の1つを実行しても良い。一般的に、当該方法は、どのようなハードウェア装置によって実行されても良い。

30

【0213】

上述の各実施例は、単に本発明の原理を例示するものである。本願明細書に記載の構成及び詳細を変更及び変形したものが当業者には明らかであることが理解される。従って、本願明細書における各実施例の記載及び説明として提示された特定の詳細によってではなく、添付の特許請求の範囲によってのみ限定されることが意図される。

40

【0214】

参考文献

[1] J. B. アレン (Allen)、D. A. バークリー (Berkeley)、J. ブラウアート (Blauert)、「スピーチ信号から室内反響を除去するためのマルチマイクロフォン信号処理技術 (Multimicrophone signal-processing technique to remove room reverberation from speech signals)」、米国音響協会報 (J. Acoust. Soc. Am)、62巻、1977年

[2] C. アヴェンダーノ (Avendano)、J. M. ジョット (Jot)、「マルチチャンネルアップミックスのための周波数領域方策 (A frequency-dom

50

ain approach to multi-channel upmix)」、音声工学協会報 (J. Audio Eng. Soc.)、52巻、2004年

[3] C. ファラー (Faller)、"ステレオ信号の多スピーカー式再生 (Multiple-loudspeaker playback of stereo signals)"、音声工学協会報 (J. Audio Eng. Soc.)、54巻、2006年

[4] J. メリマー (Merimaa)、M. グッドウィン (Goodwin)、J. M. ジョット (Jot)、"ステレオ録音からの相関ベースの環境抽出 (Correlation-based ambience extraction from stereo recordings)"、第123回AES会議集録 (Proc. of the AES 123rd Conv.)、2007年

10

[5] ヴィッレ・プルキ (Ville Pulkki)、"空間音再生及びステレオアップミックスにおける方向性音声符号化 (Directional audio coding in spatial sound reproduction and stereo upmixing)"、第28回国際AES会議集録 (Proc. of the AES 28th Int. Conf.)、2006年

[6] J. アッシャー (Usher)、J. ベネステイ (Benesty)、"空間音質の向上：新規の反響抽出音声アップミキサー (Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer)"、音声・スピーチ・言語処理に関するIEEE会報 (IEEE Trans. on Audio, Speech, and Language Processing)、15巻、2141~2150頁、2007年

20

[7] A. ヴァルター (Walther)、C. ファラー (Faller)、"サラウンド音信号の直接・環境分解及びアップミックス (Direct-ambient decomposition and upmix of surround sound signals)"、IEEE WASPAA集録 (Proc. of IEEE WASPAA)、2011年

[8] C. ウーレ (Uhle)、J. ヘレ (Herre)、S. ガイヤースベルガー (Geyersberger)、F. リッターブッシュ (Ridderbusch)、A. ヴァルター (Walter)、O. モーザー (Moser)、"環境信号を抽出するための重みづけ係数を得るための装置及び方法において環境信号を抽出するための装置及び方法、並びにコンピュータプログラム (Apparatus and method for extracting an ambient signal in an: apparatus and method for obtaining weighting coefficients for extracting an ambient signal and computer program)"、米国特許出願第2009/0080666号、2009年

30

[9] C. ウーレ (Uhle)、J. ヘレ (Herre)、A. ヴァルター (Walther)、O. ヘルムート (Hellmuth)、C. ヤンセン (Janssen)、"音声信号から環境信号を生成するための装置及び方法、音声信号からマルチチャネル音声信号を導出するための装置及び方法、並びにコンピュータプログラム (Apparatus and method for generating an ambient signal from an audio signal, apparatus and method for deriving a multi-channel audio signal from an audio signal and computer program)"、米国特許出願第2010/0030563号、2010年

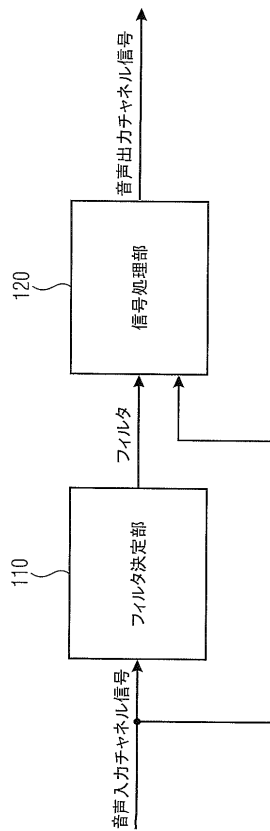
40

[10] G. ソウルオウダー (Soulodre)、"音声入力信号の反響コンテンツを抽出し変化させるためのシステム (System for extracting and changing the reverberant content of a

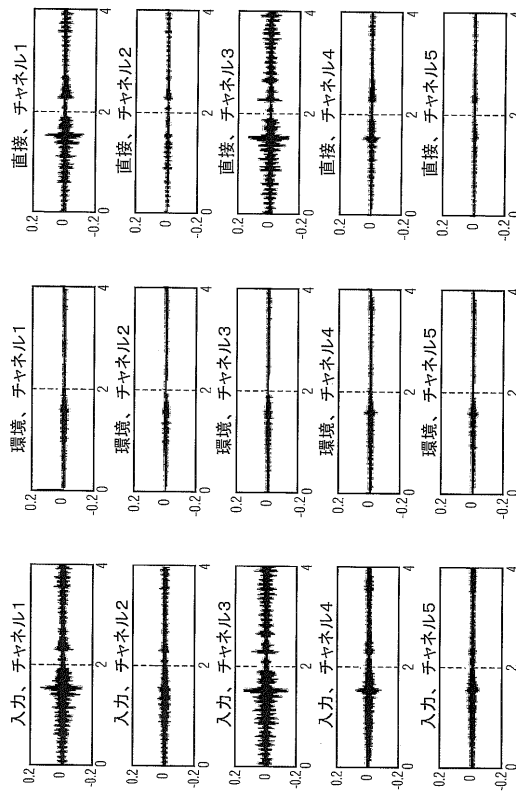
50

n audio input signal)」、米国特許第8,036,767号、特許日：2011年10月11日

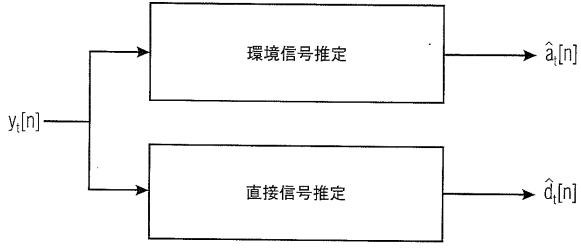
【図1】



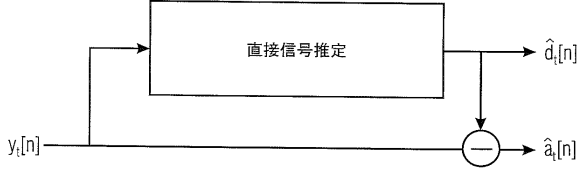
【図2】



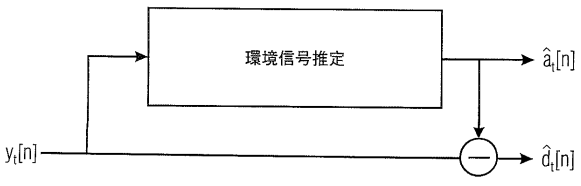
【図3】



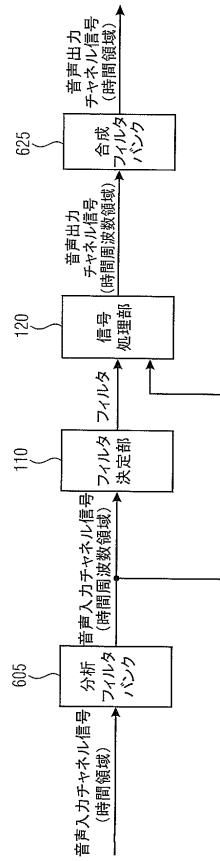
【図4】



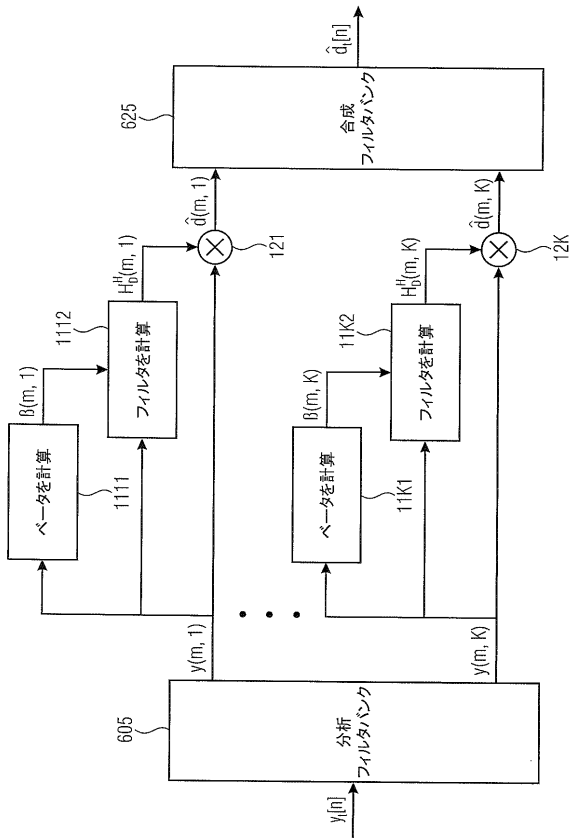
【図5】



【図6a】



【図6b】



フロントページの続き

- (72)発明者 ガンプ、 パトリック
ドイツ国 9 1 0 5 4 エルランゲン、 エスキルストゥーナシュトラッセ 2 9
- (72)発明者 クラッツ、 ミヒャエル
ドイツ国 9 1 0 5 4 エルランゲン、 フリードリヒシュトラッセ 9

審査官 岩田 淳

- (56)参考文献 特許第6 3 8 5 3 7 6 (J P , B 2)
特表2 0 0 9 - 5 2 2 9 4 2 (J P , A)

- (58)調査した分野(Int.Cl. , DB名)
G 1 0 L 1 9 / 0 0 - 9 9 / 0 0