

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5294603号
(P5294603)

(45) 発行日 平成25年9月18日(2013.9.18)

(24) 登録日 平成25年6月21日(2013.6.21)

(51) Int.Cl.

F I

HO4R	3/00	(2006.01)	HO4R	3/00	320
G1OL	21/0272	(2013.01)	G1OL	21/02	202A
G1OL	21/0308	(2013.01)	G1OL	21/02	203A
HO4R	1/40	(2006.01)	G1OL	21/02	203Z
G1OL	21/028	(2013.01)	HO4R	1/40	320A

請求項の数 8 (全 13 頁) 最終頁に続く

(21) 出願番号 特願2007-259797 (P2007-259797)
 (22) 出願日 平成19年10月3日 (2007.10.3)
 (65) 公開番号 特開2009-89315 (P2009-89315A)
 (43) 公開日 平成21年4月23日 (2009.4.23)
 審査請求日 平成22年8月18日 (2010.8.18)

(73) 特許権者 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
 (74) 代理人 100121706
 弁理士 中尾 直樹
 (74) 代理人 100128705
 弁理士 中村 幸雄
 (74) 代理人 100147773
 弁理士 義村 宗洋
 (74) 代理人 100066153
 弁理士 草野 卓
 (72) 発明者 守谷 健弘
 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】 音響信号推定装置、音響信号合成装置、音響信号推定合成装置、音響信号推定方法、音響信号合成方法、音響信号推定合成方法、これらの方法を用いたプログラム、及び記録媒体

(57) 【特許請求の範囲】

【請求項1】

K個(Kは2以上の整数)のマイクで収録したKチャンネルの音響信号を、チャンネルk(k=1, 2, ..., K)ごとに所定の周波数帯域ごとに分割して帯域信号 X_k ()を生成する帯域分割部と、

周波数帯域ごとに、各マイクkの位置での各音源m(m=1, 2, ..., M、Mは周波数帯域における音源の数)からの信号 $U_{k,m}$ を推定し、前記帯域信号 X_k ()と前記信号 $U_{k,m}$ とから、

【数5】

$$N_k(\omega) = X_k(\omega) - \sum_{m=1}^{M\omega} U_{k,\omega,m}$$

10

である残差帯域信号 N_k ()を求める音源推定部と、
 を備える音響信号推定装置。

【請求項2】

周波数帯域ごとの各音源m(m=1, 2, ..., M、Mは周波数帯域における音源の数)の位置または方向 D_{m} と当該音源の位置または方向に対応づけられた周波数帯域ごとの強度と位相 S_{m} と、前記音源の位置または方向に対応づけられていない信号で各チャンネルk(k=1, 2, ..., K、Kはチャンネルの数)と周波数帯域に対応づ

20

けられた信号である残差帯域信号 $N_k(\omega)$ 、音を合成する位置を入力とする音響信号合成装置であって、

周波数帯域ごとに、各音源の位置または方向 $D_{k,m}$ と周波数帯域ごとの強度と位相 $S_{k,m}$ から、指定された位置での前記各音源からの信号を合成した帯域信号 $Z(\omega)$ を推定する帯域信号成分推定部と、

各マイク k と前記指定された位置との距離に応じた重みを α_k としたとき、周波数帯域ごとに、前記帯域信号 $Z(\omega)$ と前記残差帯域信号 $N_k(\omega)$ とから、

【数 6】

$$Y(\omega) = Z(\omega) + \sum_{k=1}^K \alpha_k \cdot N_k(\omega) \quad 10$$

である前記指定された位置での帯域信号 $Y(\omega)$ を求める帯域信号成分加算部と、

前記指定された位置での帯域信号 $Y(\omega)$ を、時間領域の信号に変換する帯域統合部とを備える音響信号合成装置。

【請求項 3】

K 個 (K は 2 以上の整数) のマイクで収録した K チャンネルの音響信号を、チャンネル k ($k = 1, 2, \dots, K$) ごとに所定の周波数帯域ごとに分割して帯域信号 $X_k(\omega)$ を生成する帯域分割部と、

周波数帯域ごとに、各音源 m ($m = 1, 2, \dots, M$ 、 M は周波数帯域における音源の数) の位置または方向 $D_{k,m}$ と強度と位相 $S_{k,m}$ と各マイク k の位置での各音源 m からの信号 $U_{k,m}(\omega)$ を推定し、前記帯域信号 $X_k(\omega)$ と前記信号 $U_{k,m}(\omega)$ とから、

【数 7】

$$N_k(\omega) = X_k(\omega) - \sum_{m=1}^M U_{k,m}(\omega)$$

である残差帯域信号 $N_k(\omega)$ を求める音源推定部と、

各音源の位置または方向 $D_{k,m}$ と周波数帯域ごとの強度と位相 $S_{k,m}$ 、および各チャンネルの残差帯域信号 $N_k(\omega)$ を記録する記録部と、

周波数帯域ごとに、各音源の位置または方向 $D_{k,m}$ と周波数帯域ごとの強度と位相 $S_{k,m}$ から、指定された位置での各音源からの信号を合成した帯域信号 $Z(\omega)$ を推定する帯域信号成分推定部と、

各マイク k と前記指定された位置との距離に応じた重みを α_k としたとき、周波数帯域ごとに、前記帯域信号 $Z(\omega)$ と前記残差帯域信号 $N_k(\omega)$ とから、

【数 8】

$$Y(\omega) = Z(\omega) + \sum_{k=1}^K \alpha_k \cdot N_k(\omega) \quad 40$$

である前記指定された位置での帯域信号 $Y(\omega)$ を求める帯域信号成分加算部と、

前記指定された位置での帯域信号 $Y(\omega)$ を、時間領域の信号に変換する帯域統合部とを備える音響信号推定合成装置。

【請求項 4】

帯域分割部で、 K 個 (K は 2 以上の整数) のマイクで収録した K チャンネルの音響信号を、チャンネル k ($k = 1, 2, \dots, K$) ごとに所定の周波数帯域ごとに分割して帯域信号 $X_k(\omega)$ を生成する帯域分割ステップと、

音源推定部で、周波数帯域ごとに、各マイク k の位置での各音源 m ($m = 1, 2, \dots, M$ 、 M は周波数帯域における音源の数) からの信号 $U_{k,m}(\omega)$ を推定し、前記

帯域信号 $X_k(\omega)$ と前記信号 $U_{k, \omega, m}$ とから、

【数 9】

$$N_k(\omega) = X_k(\omega) - \sum_{m=1}^{M\omega} U_{k, \omega, m}$$

である残差帯域信号 $N_k(\omega)$ を求める音源推定ステップと、
を有する音響信号推定方法。

【請求項 5】

周波数帯域ごとの各音源 m ($m = 1, 2, \dots, M$ 、 M は周波数帯域における音源の数) の位置または方向 $D_{\omega, m}$ と当該音源の位置または方向に対応づけられた周波数帯域ごとの強度と位相 $S_{\omega, m}$ と、前記音源の位置または方向に対応づけられていない信号で各チャンネル k ($k = 1, 2, \dots, K$ 、 K はチャンネルの数) と周波数帯域に対応づけられた信号である残差帯域信号 $N_k(\omega)$ 、音を合成する位置を入力とする音響信号合成方法であって、

10

帯域信号成分推定部で、周波数帯域ごとに、各音源の位置または方向 $D_{\omega, m}$ と周波数帯域ごとの強度と位相 $S_{\omega, m}$ から、指定された位置での前記各音源からの信号を合成した帯域信号 $Z(\omega)$ を推定する帯域信号成分推定ステップと、

帯域信号成分加算部で、各マイク k と前記指定された位置との距離に応じた重みを α_k としたとき、周波数帯域ごとに、前記帯域信号 $Z(\omega)$ と前記残差帯域信号 $N_k(\omega)$ とから、

20

【数 10】

$$Y(\omega) = Z(\omega) + \sum_{k=1}^K \alpha_k \cdot N_k(\omega)$$

である前記指定された位置での帯域信号 $Y(\omega)$ を求める帯域信号成分加算ステップと、
帯域統合部で、前記指定された位置での帯域信号 $Y(\omega)$ を、時間領域の信号に変換する帯域統合ステップと
を有する音響信号合成方法。

30

【請求項 6】

帯域分割部で、 K 個 (K は 2 以上の整数) のマイクで收音した K チャンネルの音響信号を、チャンネル k ($k = 1, 2, \dots, K$) ごとに所定の周波数帯域ごとに分割して帯域信号 $X_k(\omega)$ を生成する帯域分割ステップと、

音源推定部で、周波数帯域ごとに、各音源 m ($m = 1, 2, \dots, M$ 、 M は周波数帯域における音源の数) の位置または方向 $D_{\omega, m}$ と強度と位相 $S_{\omega, m}$ と各マイク k の位置での各音源 m からの信号 $U_{k, \omega, m}$ を推定し、前記帯域信号 $X_k(\omega)$ と前記信号 $U_{k, \omega, m}$ とから、

【数 11】

$$N_k(\omega) = X_k(\omega) - \sum_{m=1}^{M\omega} U_{k, \omega, m}$$

40

である残差帯域信号 $N_k(\omega)$ を求める音源推定ステップと、

帯域信号成分推定部で、周波数帯域ごとに、各音源の位置または方向 $D_{\omega, m}$ と周波数帯域ごとの強度と位相 $S_{\omega, m}$ から、指定された位置での各音源からの信号を合成した帯域信号 $Z(\omega)$ を推定する帯域信号成分推定ステップと、

帯域信号成分加算部で、各マイク k と前記指定された位置との距離に応じた重みを α_k としたとき、周波数帯域ごとに、前記帯域信号 $Z(\omega)$ と前記残差帯域信号 $N_k(\omega)$ とから、

50

【数 1 2】

$$Y(\omega) = Z(\omega) + \sum_{k=1}^K \alpha_k \cdot N_k(\omega)$$

である前記指定された位置での帯域信号 Y () を求める帯域信号成分加算ステップと、帯域統合部で、前記指定された位置での帯域信号 Y () を、時間領域の信号に変換する帯域統合ステップとを有する音響信号推定合成方法。

10

【請求項 7】

請求項 4 から 6 のいずれかに記載の方法をコンピュータに実行させるプログラム。

【請求項 8】

請求項 7 記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数チャンネルの音響信号から音源の位置または方向と強度と位相を推定し、任意の位置の音響信号を合成する音響信号推定装置、音響信号合成装置、音響信号推定合成装置、音響信号推定方法、音響信号合成方法、音響信号推定合成方法、これらの方法を用いたプログラム、及び記録媒体に関する。

20

【背景技術】

【0002】

立体的音響信号を複数のマイクで收音し、音源を分離したり、雑音を抑圧したりする手法は良く知られている。音源の位置はセンサーで収集できる。また、アレーマイクで個別の音に分離して収集することもできる。その手段として、S A F I A 法（非特許文献 1）や C S C C 法（非特許文献 2）が知られている。

【非特許文献 1】青木真理子，山口義和，古家賢一，片岡章俊，“音源分離方式 S A F I A を用いた高騒音下における近接音源の分離抽出”，電子情報通信学会誌 A，Vol. J88-A，No.4，pp.468-479，2005。

30

【非特許文献 2】松本恭輔，小野順貴，嵯峨山茂樹，“位相差拘束付複素スペクトル円心（C S C C）法による雑音抑圧の検討”，日本音響学会講演論文集，3-1-11，pp.499-500，2006。

【発明の開示】

【発明が解決しようとする課題】

【0003】

一般的に、複数のマイクは音源から離れた位置に設置され、常時、音を收音している。しかし、音源の位置や数は明確ではなく、時間とともに変動することも想定される。このような場合に、任意の位置で收音される音を求めるには、いくつかのパラメータを仮定して音源を分離する手法では対応できない。

40

本発明はこのような問題点を解決し、複数のマイクで收音された音から、音源を推定し、任意の位置での音を合成する方法を提供することにある。

【課題を解決するための手段】

【0004】

本発明の音響信号推定装置は、帯域分割部と音源推定部から構成される。帯域分割部は、複数のマイクで收音した複数チャンネルの音響信号を、チャンネルごとに所定の周波数帯域ごとに分割して帯域信号を生成する。音源推定部は、周波数帯域ごとに音源の位置または方向と強度と位相を推定する。そして、チャンネルごとに音源からの信号を帯域信号から除いて残差帯域信号を求める。つまり、1 以上の音源が推定できた周波数帯域は、チャンネル

50

ごとに音源からの信号を帯域信号から除いて残差帯域信号を求め、音源が推定できなかった周波数帯域は、各チャンネルの帯域信号を残差帯域信号とする。

【0005】

本発明の音響信号合成装置は、帯域信号成分推定部と帯域信号成分加算部と帯域統合部から構成され、各音源の位置または方向と周波数帯域ごとの強度と位相、各チャンネルの残差帯域信号、音を合成する位置を入力とする。帯域信号成分推定部は、各音源の位置または方向と周波数帯域ごとの強度と位相から、指定された位置での各音源からの帯域信号を推定する。帯域信号成分加算部は、推定された各音源からの帯域信号と各チャンネルの残差帯域信号とを重み付き加算することで、指定された位置での帯域信号を求める。帯域統合部は、指定された位置での帯域信号を、時間領域の信号に変換する。

10

【0006】

本発明の音響信号推定合成装置は、上述の音響信号推定装置と記録部と音響信号合成装置から構成される。記録部は、音響信号推定装置から出力される各音源の位置または方向と周波数帯域ごとの強度と位相、および各チャンネルの残差帯域信号を記録する。音響信号合成装置は、記録部に記録された推定された各音源の位置または方向と周波数帯域ごとの強度と位相、各チャンネルの残差帯域信号、收音される音を合成する位置を入力とする。

なお、音響信号推定装置や音響信号合成装置は、上述の記録部を内部に備えていてもよい。

【発明の効果】

【0007】

本発明の音響信号推定装置によれば、複数のマイクで收音した複数チャンネルの音響信号から、1以上の音源の位置または方向と周波数帯域ごとの強度と位相を推定し、各チャンネルの残差帯域信号を求める。したがって、音源が推定できた音と雑音などの音源が推定できない音に分けることができる。

20

本発明の音響信号合成装置によれば、音源が推定できた音については、音源の位置または方向から指定された位置で收音される音を計算できる。また、音源が推定できない音については、各チャンネルの残差帯域信号（帯域信号に含まれる音源が特定できない信号）から指定された位置で收音される音を計算できる。そして、これらを重み付け加算するので、指定された位置での音を合成できる。

【0008】

本発明の音響信号推定合成装置によれば、上述の音響信号推定装置と音響信号合成装置の効果の有するので、複数のマイクで收音した複数チャンネルの音響信号から、指定された位置での合成できる。

30

このような効果があるので、例えば、複数の場所のカメラから任意の視点の画像・映像を合成する自由視点映像システムに対応した音響信号の合成も可能となる。

【発明を実施するための最良の形態】

【0009】

以下に、図を示しながら本発明の原理と実施形態を説明する。

原理

図1は、4つのマイクと伝播した音が平面波と近似できるほど遠方の音源からの音の様子を示す例である。一般的には、最も離れたマイク同士の間隔より、10倍以上音源が離れた場合、平面波と近似できる。図1では、4つのマイク501～504は直線状に配置されている。音源Aからの音は、マイクの配置と垂直な方向から来るとする。この場合には、到達する音波の波面がそろうので、音源Aからの各マイクへの入力信号は同一となる。音源Bからの音は、マイクの配置に対して垂直ではない方向から来るとする。この場合、各マイクへの音源Bからの音の到達時間が異なる。また、帯域ごとの帯域信号成分で見ると位相が異なる。図2に、マイク501～504が設置されている場所501～504での、音源Aから伝播された音のスペクトルの例を示す。図3に場所501～504での音源Bから伝播された音のスペクトルの例を示す。図3(A)は場所501での音源Bからの音のスペクトル、(B)は場所502での音源Bからの音のスペクトル、(C)は場

40

50

所 5 0 3 での音源 B からの音のスペクトル、(D) は場所 5 0 4 での音源 B からの音のスペクトルである。また、図 4 から図 6 に場所 5 0 1、5 0 2、5 0 3 での、音源 A と音源 B からの音のスペクトルを示す。図 4 が場所 5 0 1 での音源 A と音源 B からの音のスペクトル、図 5 が場所 5 0 2 での音源 A と音源 B からの音のスペクトル、図 6 が場所 5 0 3 での音源 A と音源 B からの音のスペクトルである。

【 0 0 1 0 】

本発明では、このように複数のマイクで収録した音のスペクトルから、音源の方向、音源のスペクトルを推定する。なお、図 7 のように球面波を前提とした場合には、音源の方向を推定するのではなく、音源の位置を推定することになる。そして、各マイクでの推定した音源からの音のスペクトルを計算し、残差として残る信号(残差信号)を音源が特定できない雑音として扱う。そして、推定された音源の位置とスペクトルから、音響波形が欲しい位置(指定された位置)での各音源からの音のスペクトルを求める。また、指定された位置の残差信号のスペクトルは、指定された位置の近くのマイクの残差信号を、指定された位置とマイクとの距離を考慮した重み付け加算して求める。これらを加算することで、指定された位置での音響波形を合成する。

10

【 0 0 1 1 】

音源の方向やスペクトルを推定する方法は、従来から存在する方法を用いればよい。例えば、各マイクで収録した音の位相差を用いる方法などがある。音の位相差は、例えば、マイクが 2 つの場合には、相互相関関数の計算である時間差でピークがはっきりと出れば、1 つの音源があると判断できる。また、マイクが 2 つ以上の場合、例えば、1 つの音源を仮定して連立方程式を解くか、位相差を周波数領域で評価すると、得られた結果が 1 つの音源とみなせるか否か判断できる。つまり、一般的に、2 つ以上のマイクがあれば、収録された音の個々の周波数帯域での位相の違いから、音源方向を推定できる。

20

【 0 0 1 2 】

S A F I A 法では、個々の帯域では主要な音源の成分は 1 つであると仮定し、音源の位置と、その音源からの音を求める。音源のスペクトルには、強い部分と弱い部分があり、ある帯域に注目すると主要な成分が複数の音源から来ることは比較的少ない。例えば、図 4 から図 6 に示したように、音源 A からの音のスペクトルと音源 B からの音のスペクトルでは、スペクトルが存在する周波数のほとんどが異なる(例えば、図 4 から図 6 の注目帯域 a と注目帯域 c)。したがって、帯域分割した場合、ある帯域では、音源 A または音源 B の一方の音が主となり、他方はほとんどない。S A F I A 法は、このような特性を利用している。

30

【 0 0 1 3 】

C S C C 法では、他の音源からの入力スペクトルが一定となる場合、あるいはそのように換算した場合、複数のマイクに対する単一音源からのスペクトルの複素平面上での配置から音源方向とその信号成分を分離して推定する。注目帯域 a の例の場合のように、音源 A からの成分はほとんどない場合や、各信号に遅延を与えるなどして音源 A からの信号成分がすべてのマイクに共通となるように換算できる場合には、場所 5 0 1 ~ 5 0 4 の音源 B の成分から、音源 B の方向が精度よく推定できる。なお、音源の位置の推定精度は、他の音がどの程度あるかに依存する。注目帯域 c の場合には、音源 B からの成分がほとんどないので、場所 5 0 1 ~ 5 0 4 の音源 A の成分から、音源 A の方向が精度よく推定できる。この場合は、どの場所のスペクトルも同じなので、マイクの設置方向と垂直な方向に音源 A が存在することが分かる。注目帯域 b の場合には、音源 A の成分も音源 B の成分も強いため、単純な分離は難しい。この場合、音源方向の推定の信頼度が高い帯域(例えば、注目帯域 a、注目帯域 c)で推定した音源の位置を用いて、音源 A からの成分と音源 B からの成分とを推定する。この例では、音源 A からの成分は、マイクの場所によらないので、定数とみなすことができる。

40

【 0 0 1 4 】

その他にも複数の音源を音源数以上の数のマイクの信号から分離する技術がある(特開 2 0 0 6 - 2 4 3 6 6 4 号公報)。また、帯域を分割すれば、音源が発生する周波数成分

50

が偏るので、マイクの数が少なくても分離可能となる（特開2007-198977号公報）。

【0015】

本発明でも、複数の音源があることを前提に複数のマイクで収録した信号を、音源ごとに分離することで、音源の方向（または位置）、音源のスペクトルを推定する。したがって、上述の信号の分離方法や類似の方法を用いる点では共通するし、どの方法を用いるかは適宜選択すればよい。しかし、本発明の目的は、任意の位置での音を合成することであり、音源ごとに音を分離することではない。つまり、本発明では、音を正確に分離できることよりも、結果的に指定された位置での音のように合成できることが重要である。そこで、本発明では、上述のいずれかの方法で可能な範囲まで、音源の位置または方向と音源帯域信号（周波数帯域ごとの強度と位相）とを推定し、残る信号を音源の位置が特定できない残差信号として扱う。残差信号は、マイクごとに求められる。そして、音源ごとの方向と周波数帯域ごとの音源帯域信号（複素スペクトル）、マイク（チャンネル）ごとの周波数帯域ごとの残差信号（残差帯域信号）が記録される。指定された位置での音の合成では、各音源の位置または方向と音源帯域信号（周波数帯域ごとの強度と位相）から、指定された位置での各音源からの帯域信号を推定する。そして、推定された各音源からの帯域信号と各チャンネルの残差帯域信号とを重み付き加算することで、指定された位置での帯域信号を求める。最後に、指定された位置での帯域信号を、時間領域の信号に変換する。

10

【0016】

[第1実施形態]

20

図8に、本発明の音響信号推定合成装置の機能構成例を示す。また、図9に、音響信号推定合成装置の処理フローの例を示す。本発明の音響信号推定合成装置100は、帯域分割部110、音源推定部120、記録部130、帯域信号成分推定部140、帯域信号成分加算部150、帯域統合部160から構成される。帯域分割部110は、K個（Kは2以上の整数）のマイクで収録したKチャンネルの音響信号 $x_1(t)$, $x_2(t)$, ..., $x_K(t)$ を、チャンネルごとに所定の周波数帯域ごとに分割して帯域信号 $X_1(\omega)$, $X_2(\omega)$, ..., $X_K(\omega)$ を生成する（S110）。音響信号 $x_1(t)$ は、Tサンプルからなるフレーム中の1つのサンプル値（スカラー量）であり、tは0, ..., T-1の値を取る。このような音響信号 $x_1(t)$ から、所定の周波数帯域ごとの帯域信号 $X_1(\omega)$ を得る。帯域信号 $X_1(\omega)$ は、例えば複素スペクトルである。なお、帯域信号 $X_1(\omega)$ は帯域分割複素信号でもよいが、以下では複素スペクトルとして説明する。次式のように、時間領域のT点ごとのフレームを複素フーリエ変換し、T/2点の複素フーリエ係数を求めたものを帯域信号 $X_1(\omega)$ とする。

30

【数1】

$$X_1(\omega) = \sum_{t=0}^{T-1} x_1(t) \exp(t\omega 2\pi j / T)$$

ただし、 $\omega = 0, \dots, T/2$ 、jは虚数単位、 ω は円周率とする。

帯域信号 $X_1(\omega)$ は1番目のマイク（第1のチャンネル）の位置での信号の、周波数帯域ごとの振幅と位相を示している。サンプリング周波数をf [Hz]としたとき、 f/T [Hz]を中心周波数とする帯域信号とみなせる。なお、帯域分割部110への入力を、アナログの音響信号とし、帯域分割部110内でサンプリングした値を音響信号 $x_1(t)$ としてもよい。どの場合も、出力は同じである。

40

【0017】

音源推定部120は、従来から存在する方法で、周波数帯域ごとに音源の位置または方向 $D_{1,1}, D_{1,2}, \dots, D_{1,M}$ と音源帯域信号 $S_{1,1}, S_{1,2}, \dots, S_{1,M}$ を推定する（Mは周波数帯域での音源の数であり、0以上の整数である）。音源帯域信号 $S_{1,1}$ は、周波数帯域での第1の音源から伝搬した音によって、マイク近傍で生じる信号を計算するための強度と位相の情報（例えば、複素スペクトル）である。例えば、 $D_{1,1}$ が音源の位置を示しており、音を球面波とするのであれば、音源帯域信

50

号 $S_{k,1}$ は音源の位置での強度と位相を示す複素スペクトルとすればよい。また、 $D_{k,1}$ が音源の方向を示しており、音を平面波に近似とするのであれば、音源帯域信号 $S_{k,1}$ はある位置（音源の位置である必要はない）での強度と位相を示す複素スペクトルとすればよい。この推定の過程で、各マイクの位置での、それぞれの音源からの信号 $U_{k,m}$ も求めておく（ k はマイクの番号を示しており、 $1 \sim K$ の整数である）。信号 $U_{k,m}$ は、 k 番目のマイクの位置での周波数帯域 m 番目の音源からの信号を示している（ m は周波数帯域ごとに付された音源の番号であり、 $0 \sim M$ の整数である）。例えば、平面波で近似する場合であれば、音源帯域信号 $S_{k,1}$ の位置とマイク k の位置とを結ぶベクトルと音の伝搬方向の単位ベクトルとの内積（音の伝搬方向にどれだけ離れているかを示す値）から、音源帯域信号 $S_{k,1}$ の位置とマイク k の位置との位相差を求め、 $S_{k,1}$ の位相をその位相差だけシフトした信号をマイク k の位置での信号 $U_{k,1}$ とすればよい。

10

【0018】

1以上の音源が推定できた周波数帯域は、次式のようにチャンネルごとに音源からの信号を帯域信号から除いて残差帯域信号 $N_1(\omega)$ 、 $N_2(\omega)$ 、 \dots 、 $N_K(\omega)$ を求める。

【数2】

$$N_k(\omega) = X_k(\omega) - \sum_{m=1}^{M\omega} U_{k,\omega,m}$$

20

また、音源が推定できなかった周波数帯域は、 $N_k(\omega) = X_k(\omega)$ をすべての k （マイク）と m （周波数帯域）に対して計算することで、各チャンネルの帯域信号 $X_k(\omega)$ を残差帯域信号 $N_1(\omega)$ 、 $N_2(\omega)$ 、 \dots 、 $N_K(\omega)$ とする（S120）。つまり、チャンネルごとに、マイクの位置での推定できた音源からの信号を帯域信号から引くことで、残差帯域信号 $N_1(\omega)$ 、 $N_2(\omega)$ 、 \dots 、 $N_K(\omega)$ を求めている。このように、本発明では、音源の位置が推定できなかった信号を、残差帯域信号として扱うので、音源の位置が特定できなかった信号を無理やりいずれかの音源に割り振る必要がない。

【0019】

音源の位置を推定するか方向を推定するかは、音を球面波と仮定するか平面波と仮定するかで決まる。この仮定は、あらかじめ決めておく。また、どのような方法で音源の位置または方向と強度と位相を推定するかは、上述の方法などから適宜選択しておけばよい。なお、上述したように、本発明では正確に音源の位置（または方向）やスペクトルを推定することよりも、最終的に合成された音が、指定された位置での音らしくなることが重要である。ステップ S120 で推定された各音源の位置または方向と周波数帯域ごとの強度と位相、および各チャンネルの残差帯域信号は、記録部 130 に記録される。なお、記録される情報は、符号化された情報でもよい。

30

【0020】

帯域信号成分推定部 140 は、位置 P が指定されると、周波数帯域 m ごとの各音源の位置または方向 $D_{k,1}$ 、 $D_{k,2}$ 、 \dots 、 $D_{k,M}$ と音源帯域信号 $S_{k,1}$ 、 $S_{k,2}$ 、 \dots 、 $S_{k,M}$ から、指定された位置 P でのすべての音源からの音を合成した帯域信号 $Z(\omega)$ を推定する（S140）。例えば、周波数帯域 m ごとに、位置 P での各音源からの信号 $U_{P,m}$ を求める（ m は周波数帯域ごとに付された音源の番号であり、 $0 \sim M$ の整数である）。信号 $U_{P,m}$ の求め方は、音源推定部 120 の各マイクの位置での音源からの信号 $U_{k,m}$ の求め方と同じでよい。位置 P での各音源からの信号 $U_{P,m}$ を、次のように周波数帯域 m ごとに、加算すれば、帯域信号 $Z(\omega)$ を求めることができる。

40

【数3】

$$Z(\omega) = \sum_{m=1}^{M\omega} U_{P,\omega,m}$$

50

帯域信号成分加算部 150 は、推定されたすべての音源からの音を合成した帯域信号 $Z(\omega)$ () と各チャンネルの残差帯域信号 $N_1(\omega)$ () , $N_2(\omega)$ () , ... , $N_K(\omega)$ () とを重み付き加算することで、指定された位置 P での帯域信号 $Y(\omega)$ () を求める (S150)。例えば、次式のように、推定されたすべての音源からの音を合成した帯域信号 $Z(\omega)$ () には重み 1 を乗算し、各チャンネルの残差帯域信号には、すべてのチャンネルへの重みの合計が 1 となるように、各チャンネルのマイクと指定された位置 P との距離に応じた (例えば、反比例した) 重みを設定し、重みを乗算して加算すればよい。

【数 4】

$$Y(\omega) = Z(\omega) + \sum_{k=1}^K \alpha_k \cdot N_k(\omega)$$

10

$$\alpha_k = \frac{1/d_k}{\sum_k 1/d_k}$$

ただし、 d_k は、 k 番目のマイクと位置 P との距離とする。

【0021】

帯域統合部 160 は、指定された位置 P での帯域信号 $Y(\omega)$ () を、時間領域の信号 $y(t)$ () に変換する (S160)。例えば、信号 $y(t)$ () は、 T サンプルからなるフレーム内の 1 つのサンプル値であり、 t は $0, \dots, T-1$ の値を取る。

20

本発明の音響信号推定合成装置 100 はこのような構成なので、音源が推定できた音と雑音などの音源が推定できない音に分けることができる。そして、音源が推定できた音については、音源の位置または方向から指定された位置 P での音を計算できる。また、音源が推定できない音については、各チャンネルの残差帯域信号 (帯域信号に含まれる音源が特定できない信号) から指定された位置 P での音を計算できる。そして、これらを重み付け加算するので、指定された位置 P での音を合成できる。このような効果があるので、例えば、複数の場所のカメラから任意の視点の画像・映像を合成する自由視点映像システムに対応した音響信号の合成も可能となる。

【0022】

【変形例】

30

第 1 実施形態では、音響信号推定合成装置 100 を説明した。しかし、各音源の位置または方向と周波数帯域ごとの音源帯域信号、および各チャンネルの残差帯域信号を推定するまでを 1 つの装置 (音響信号推定装置) としても良い。また、各音源の位置または方向と周波数帯域ごとの音源帯域信号、および各チャンネルの残差帯域信号から、指定された位置 P での音を合成するまでを 1 つの装置 (音響信号合成装置) としても良い。

【0023】

音響信号推定装置 200 は、例えば、帯域分割部 110 と音源推定部 120 から構成される。記録部 130 は、音響信号推定装置 200 の内部に備えても良いし、外部でも良い。音響信号合成装置 300 は、例えば、帯域信号成分推定部 140、帯域信号成分加算部 150、帯域統合部 160 から構成される。

40

このように、いくつかの装置に分割して全体で音響信号推定合成装置を形成しても、第 1 実施形態と同じ効果を得ることができる。

【0024】

図 10 に、コンピュータの機能構成例を示す。なお、本発明の音響信号推定合成方法、音響信号推定方法、音響信号合成方法は、コンピュータ 2000 の記録部 2020 に、本発明の各構成部としてコンピュータ 2000 を動作させるプログラムを読み込ませ、制御部 2010、入力部 2030、出力部 2040 などを動作させることで、コンピュータに実行させることができる。また、コンピュータに読み込ませる方法としては、プログラムをコンピュータ読み取り可能な記録媒体に記録しておき、記録媒体からコンピュータに読み込ませる方法、サーバ等に記録されたプログラムを、電気通信回線等を通じてコンピュ

50

一々に読み込ませる方法などがある。

【図面の簡単な説明】

【0025】

【図1】 4つのマイクと遠くの音源から伝播した平面波の音の様子を示す図。

【図2】 場所501～504での音源Aから伝播された音のスペクトルの例を示す図。

【図3】 場所501～504での音源Bから伝播された音のスペクトルの例を示す図。

【図4】 場所501での音源Aと音源Bからの音のスペクトルを示す図。

【図5】 場所502での音源Aと音源Bからの音のスペクトルを示す図。

【図6】 場所503での音源Aと音源Bからの音のスペクトルを示す図。

【図7】 4つのマイクと音源から伝播した球面波の音の様子を示す図。

【図8】 音響信号推定合成装置の機能構成例を示す図。

【図9】 音響信号推定合成装置の処理フローの例を示す図。

【図10】 コンピュータの機能構成例を示す図。

【図1】

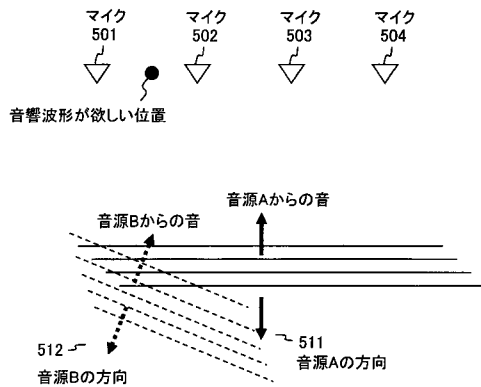


図1

【図2】

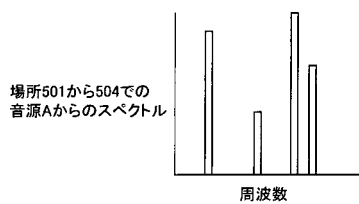


図2

【図3】

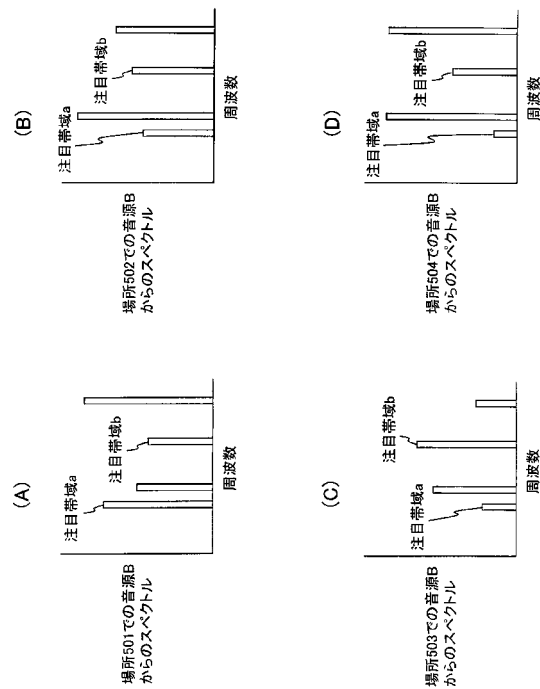


図3

【 図 4 】

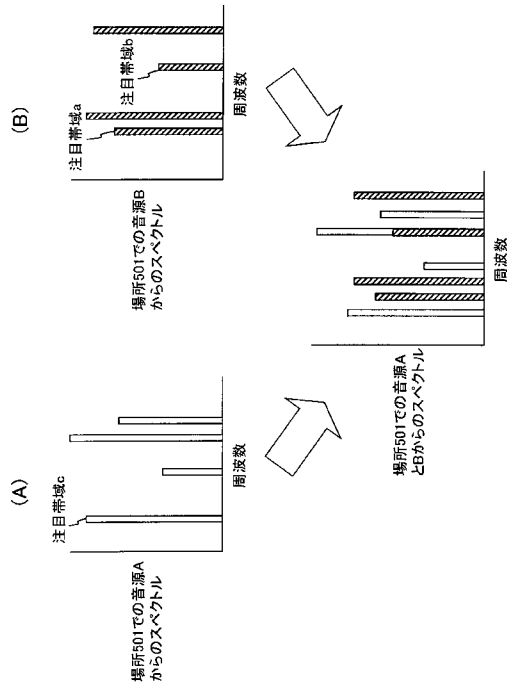


図4

【 図 5 】

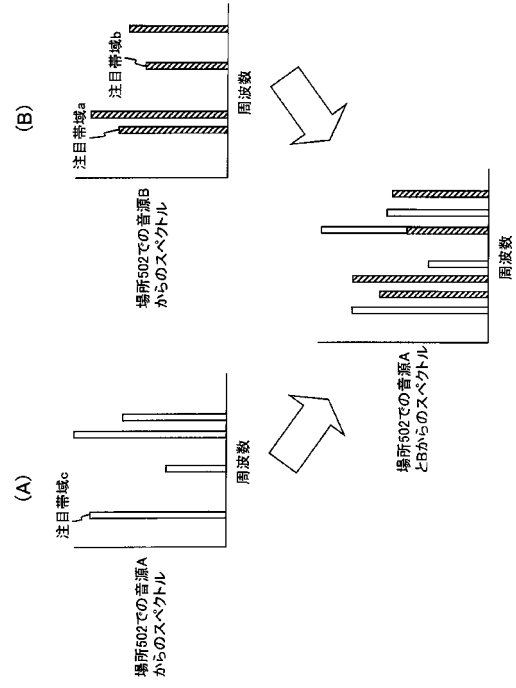


図5

【 図 6 】

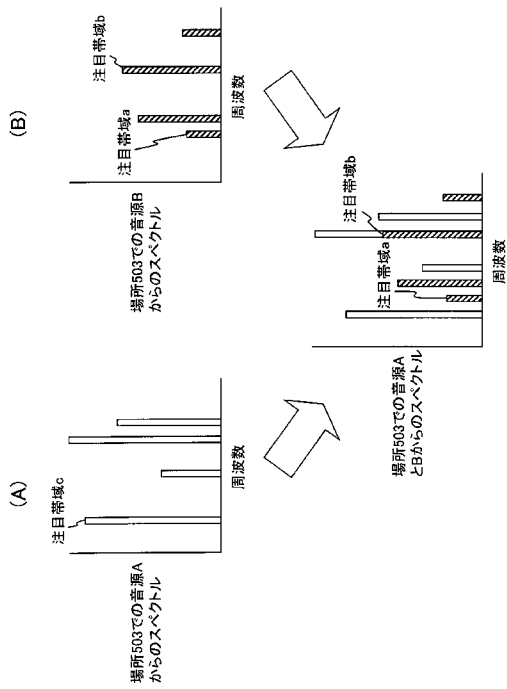


図6

【 図 7 】

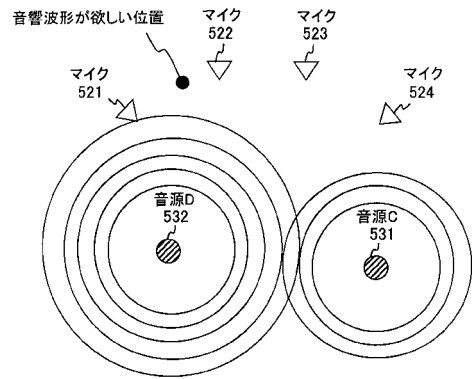


図7

【 図 8 】

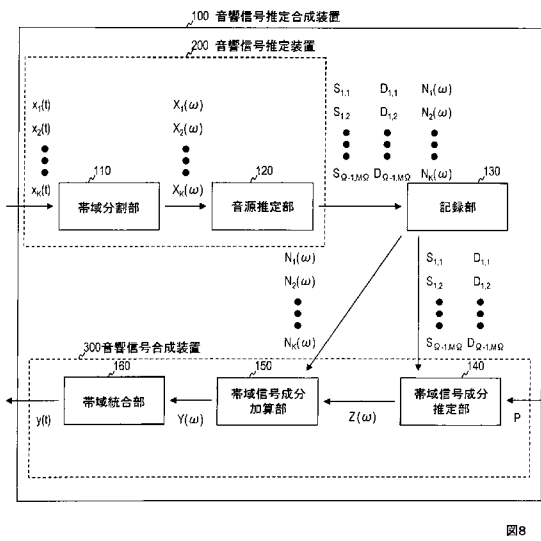


図8

【 図 9 】

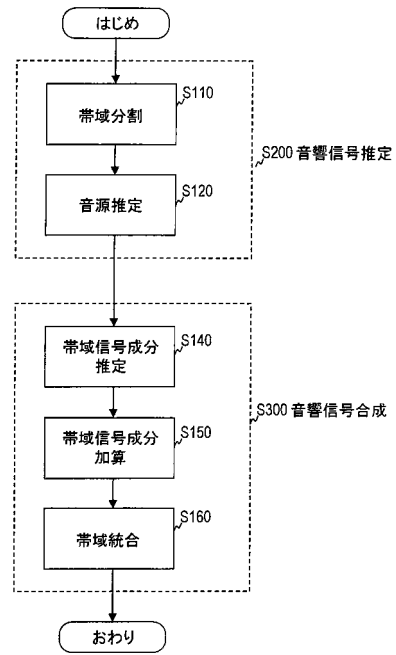


図9

【 図 10 】

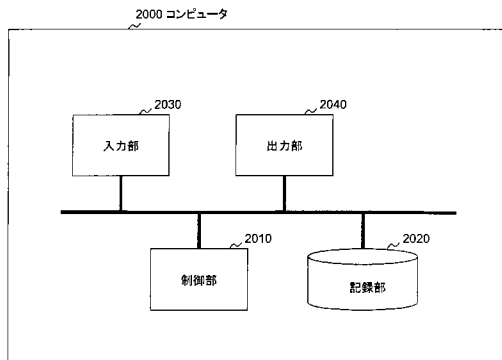


図10

フロントページの続き

(51)Int.Cl. F I
G 0 1 S 5/20 (2006.01) G 1 0 L 21/02 2 0 1 D
G 0 1 S 5/20

(72)発明者 原田 登
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 鎌本 優
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

審査官 渡邊 正宏

(56)参考文献 特開2007-074665(JP,A)

(58)調査した分野(Int.Cl., DB名)

G 0 1 S 1 / 7 2 - 1 / 8 2
G 0 1 S 3 / 8 0 - 3 / 8 6
G 0 1 S 5 / 1 8 - 5 / 3 0
G 0 1 S 7 / 5 2 - 7 / 6 4
G 0 1 S 1 5 / 0 0 - 1 5 / 9 6
G 1 0 K 1 5 / 0 0 - 1 5 / 0 6
G 1 0 L 1 1 / 0 0 - 1 3 / 0 8
G 1 0 L 1 9 / 0 0 - 2 1 / 0 6
H 0 4 R 1 / 2 0 - 1 / 4 0
H 0 4 R 3 / 0 0 - 3 / 1 4