

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7145446号  
(P7145446)

(45)発行日 令和4年10月3日(2022.10.3)

(24)登録日 令和4年9月22日(2022.9.22)

(51)国際特許分類

F I

G 0 1 S 3/808(2006.01)

G 0 1 S 3/808

G 1 0 L 21/10 (2013.01)

G 1 0 L 21/10

請求項の数 15 (全20頁)

(21)出願番号	特願2018-134381(P2018-134381)	(73)特許権者	314012076
(22)出願日	平成30年7月17日(2018.7.17)		パナソニックIPマネジメント株式会社
(65)公開番号	特開2020-12704(P2020-12704A)		大阪府大阪市中央区城見2丁目1番61号
(43)公開日	令和2年1月23日(2020.1.23)	(74)代理人	110002000
審査請求日	令和3年6月25日(2021.6.25)		特許業務法人栄光特許事務所
		(72)発明者	林 裕昭
			大阪府門真市大字門真1006番地 パナソニック株式会社内
		(72)発明者	溝淵 翔平
			大阪府門真市大字門真1006番地 パナソニック株式会社内
		審査官	高 場 正光

最終頁に続く

(54)【発明の名称】 音処理装置、音処理方法およびプログラム

(57)【特許請求の範囲】

【請求項1】

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信する通信部と、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定する推定部と、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析する分析部と、

前記推定された音源の位置から生じた音の種類の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示する制御部と、を備え、

前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、

前記制御部は、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、

音処理装置。

【請求項2】

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するヒートマップ処理部、をさらに備え、

前記制御部は、前記ヒートマップ上の前記推定された音源の位置に対応する音パラメー

タを、前記分析結果の内容に応じた色を付与した視覚画像に変換して前記画像に重畳する、  
請求項 1 に記載の音処理装置。

【請求項 3】

前記推定部は、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、

請求項 2 に記載の音処理装置。

【請求項 4】

前記分析部は、前記音の種類の分析結果として、人の声または物の音を特定する、

請求項 1 ～ 3 のうちいずれか一項に記載の音処理装置。

10

【請求項 5】

前記音源は、前記監視対象エリアにおいて複数存在する、

請求項 1 ～ 4 のうちいずれか一項に記載の音処理装置。

【請求項 6】

前記制御部は、前記音の種類の分析結果が予め指定された音の種類と一致した場合に、前記音の種類の分析結果を前記モニタに表示する、

請求項 1 ～ 5 のうちいずれか一項に記載の音処理装置。

【請求項 7】

前記マイクアレイにより集音された音データを用いて、前記マイクアレイから前記推定された音源の位置に向かう方向の音データを強調処理する強調音生成部、をさらに備える、  
請求項 1 に記載の音処理装置。

20

【請求項 8】

音処理装置における音処理方法であって、

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、

前記推定された音源の位置から生じた音の種類の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、を有し、

30

前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、  
前記重畳のステップは、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、

音処理方法。

【請求項 9】

コンピュータである音処理装置に、

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、

40

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、

前記推定された音源の位置から生じた音の種類の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、を実行させるためのプログラムであって、更に、

前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、  
前記重畳のステップは、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、

プログラム。

50

## 【請求項 10】

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信する通信部と、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定する推定部と、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析する分析部と、

前記推定された音源の位置から生じた音の種類の結果を、前記カメラにより撮像された画像に重畳してモニタに表示する制御部と、

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するヒートマップ処理部と、を備え、

前記制御部は、前記ヒートマップ上の前記推定された音源の位置に対応する音パラメータを、前記分析結果の内容に応じた色を付与した視覚画像に変換して前記画像に重畳し、

前記推定部は、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、

音処理装置。

## 【請求項 11】

音処理装置における音処理方法であって、

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、

前記推定された音源の位置から生じた音の種類の結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するステップと、を有し、

前記重畳のステップは、前記ヒートマップ上の前記推定された音源の位置に対応する音パラメータを、前記分析結果の内容に応じた色を付与した視覚画像に変換して前記画像に重畳し、

前記推定のステップは、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、

音処理方法。

## 【請求項 12】

コンピュータである音処理装置に、

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、

前記推定された音源の位置から生じた音の種類の結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、

10

20

30

40

50

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するステップと、を実行させるためのプログラムであって、更に、

前記重畳のステップは、前記ヒートマップ上の前記推定された音源の位置に対応する音パラメータを、前記分析結果の内容に応じた色を付与した視覚画像に変換して前記画像に重畳し、

前記推定のステップは、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、  
プログラム。

10

【請求項 1 3】

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信する通信部と、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定する推定部と、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析する分析部と、

前記推定された音源の位置から生じた音の種類の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示する制御部と、を備え、

20

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するヒートマップ処理部、をさらに備え、

前記推定部は、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、

音処理装置。

【請求項 1 4】

音処理装置における音処理方法であって、

30

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、

前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、

前記推定された音源の位置から生じた音の種類の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、

前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するステップと、を有し、

40

前記推定のステップは、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、

音処理方法。

【請求項 1 5】

コンピュータである音処理装置に、

監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、

50

前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、  
前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、  
前記推定された音源の位置から生じた音の種類を分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、  
前記カメラにより撮像された画像を構成する画素または所定数の画素からなるブロックごとに、前記監視対象エリアの音の大きさを示す音パラメータを導出し、その導出された音パラメータを対応する画素またはブロックに割り当てたヒートマップを生成するステップと、を実行させるためのプログラムであって、更に、  
前記推定のステップは、前記生成されたヒートマップを構成する音パラメータを正規化し、前記正規化された画素またはブロックごとの音パラメータと既定の閾値との比較に基づいて、前記監視対象エリア内に存在する音源の位置を推定する、  
プログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、集音された音を処理する音処理装置、音処理方法およびプログラムに関する。

【背景技術】

【0002】

20

特許文献1には、集音情報（音信号）を得るためのマイクロホンを透過型ヘッドマウントディスプレイの前面に配設し、マイクロホンの集音情報に基づき、位置情報に対応した視覚映像内の位置に、音源から出る音のレベル情報を円の大きさで表示する音情報表示装置が開示されている。この音情報表示装置は、音源位置が視覚映像内にある場合にはその位置あるいは近傍に音情報を表示し、一方、音源位置が表示面内から外れる場合にはその音源位置に近い視覚映像の端部に音情報を表示する。

【先行技術文献】

【特許文献】

【0003】

【文献】特開2012-133250号公報

30

【発明の概要】

【発明が解決しようとする課題】

【0004】

特許文献1によれば、音源の位置が透過型ヘッドマウントディスプレイに表示される視覚映像内のどの位置であるかが概略的に判別可能である。しかし、特許文献1では、音源の種類が具体的にどのようなものであるかを判別することは想定されていない。このため、例えば監視目的で集音された音源がどのような種類の音であるかを判別できなければ、監視員等がその音源位置まで出向いて音源を逐一確認する必要があり、監視精度が劣化するだけでなく監視員の利便性が低下するという課題があった。

【0005】

40

本開示は、上述した従来の状況に鑑みて案出され、監視対象エリアにおいて集音された音源からの音の種類を判別し、監視精度の劣化ならびに監視員の利便性の低下を抑制する音処理装置、音処理方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0006】

本開示は、監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信する通信部と、前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定する推定部と、前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析する分析部と、前記推定された音源の位置から生じた音の種類を分析結果を、前記カメラにより撮像さ

50

れた画像に重畳してモニタに表示する制御部と、を備え、前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、前記制御部は、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、音処理装置を提供する。

【0007】

また、本開示は、音処理装置における音処理方法であって、監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、前記推定された音源の位置から生じた音の種類

10

の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、を有し、前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、前記重畳のステップは、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、音処理方法を提供する。

【0008】

また、本開示は、コンピュータである音処理装置に、監視対象エリアに配置されたマイクアレイおよびカメラとの間でそれぞれ通信するステップと、前記マイクアレイにより集音された音データに基づいて、前記監視対象エリア内に存在する音源の位置を推定するステップと、前記マイクアレイにより集音された音データに基づいて、前記推定された音源の位置から生じた音の種類を分析するステップと、前記推定された音源の位置から生じた音の種類

20

の分析結果を、前記カメラにより撮像された画像に重畳してモニタに表示するステップと、を実行させるためのプログラムであって、更に、前記音の種類の分析結果は、前記モニタに表示された設定画面により選択可能であり、前記重畳のステップは、前記設定画面により選択された音の種類の分析結果を前記撮像された画像に重畳する、プログラムを提供する。

【発明の効果】

【0009】

本開示によれば、監視対象エリアにおいて集音された音源からの音の種類を判別し、監視精度の劣化ならびに監視員の利便性の低下を抑制することができる。

【図面の簡単な説明】

30

【0010】

【図1】実施の形態1に係る音源可視化システムの概要例を示す図

【図2】音源可視化システムのシステム構成ならびに処理閲覧装置のハードウェア構成の一例を示すブロック図

【図3】処理閲覧装置における音源可視化処理の動作手順例を示すフローチャート

【図4】ステップS2の音源位置推定処理の動作手順例を示すフローチャート

【図5】撮像画像上の推定された音源位置および音を強調するための指向方向を示す図

【図6】音圧を正規化しないで音源位置を推定するための音圧分布例を示す図

【図7】音圧を正規化して音源位置を推定するための音圧分布例を示す図

【図8】発話および物音の音圧ヒートマップが撮像画像に重畳して表示されたモニタの表示画面例を示す図

40

【図9】発話の音圧ヒートマップが撮像画像に重畳して表示されたモニタの表示画面例を示す図

【図10】物音の音圧ヒートマップが撮像画像に重畳して表示されたモニタの表示画面例を示す図

【発明を実施するための形態】

【0011】

以下、適宜図面を参照しながら、本開示に係る音処理装置、音処理方法およびプログラムを具体的に開示した実施の形態を詳細に説明する。但し、必要以上に詳細な説明は省略する場合がある。例えば、既によく知られた事項の詳細説明や実質的に同一の構成に対す

50

る重複説明を省略する場合がある。これは、以下の説明が不必要に冗長になるのを避け、当業者の理解を容易にするためである。なお、添付図面及び以下の説明は、当業者が本開示を十分に理解するために提供されるのであって、これらにより特許請求の範囲に記載の主題を限定することは意図されていない。

【0012】

図1は、実施の形態1に係る音源可視化システム5の概要例を示す図である。音源可視化システム5は、監視カメラシステムの一例として、監視対象エリアSAを撮像するとともに、監視対象エリアSA内で生じる音あるいは音声を集音する。図1では、監視対象エリアSAとして、オフィス100の室内が図示されている。オフィス100の室内では、デスク101を前に座っている2人の人物hm1, hm2が会話し、さらに、コピー機120が作動している。また、オフィス100の監視室(図示略)に控える監視員(音源可視化システム5のユーザの一例)は、音源可視化システム5を用いて、オフィス100の室内の状況を、撮像される画像と集音される音とによって監視している。

10

【0013】

なお、以下の実施の形態1では、監視対象エリアSAとして、オフィス100の室内を例示しているが、オフィス100の室内に限定されず、人の発する音声と物の発する音とが混在して生じる施設(例えば、工場、店舗、公共施設等)でよい。

【0014】

図1に示されるように、音源可視化システム5は、カメラ付マイクアレイ10と、処理閲覧装置50と、レコーダ70とを含む構成である。カメラ付マイクアレイ10と、処理閲覧装置50と、レコーダ70とは、ネットワークNWを介して互いにデータあるいは情報の送受信が可能に接続される。

20

【0015】

カメラ付マイクアレイ10は、オフィス100の天井に固定して設置され、監視対象エリアSAを上方から監視する。カメラ付マイクアレイ10は、例えば所定の厚みを有する円柱状の筐体を有し、この筐体内に、鉛直方向の同軸上に配置されたマイクアレイ20とカメラ30とを有する。つまり、カメラ30のレンズ(図示略)の光軸とマイクアレイ20の筐体の軸線とは共通する。

【0016】

マイクアレイ20は、例えば複数の無指向性のマイクロホンを含む。マイクロホンは、筐体の前面(つまり、カメラ付マイクアレイ10の天井との設置面から鉛直下方を向く方向の面)において同心円状に等間隔で配置される。マイクロホンの数は、例えば16個、32個、64個、128個でよい。マイクロホンには、例えばエレクトレットコンデンサーマイクロホン(ECM: Electret Condenser Microphone)が用いられる。マイクアレイ20は、複数のマイクロホンと、複数のマイクロホンのそれぞれからの出力信号をそれぞれ増幅するための複数の増幅器と、複数のA/D変換器とを有する。マイクロホンと増幅器とA/D変換器とはそれぞれ同数設けられる。各増幅器から出力されるアナログ信号は、A/D変換器でそれぞれデジタル信号に変換される。

30

【0017】

マイクアレイ20により集音される音は、例えば、人物hm1, hm2のように人間が発する音声に限らず、コピー機120、掃除機(図示略)、冷蔵庫(図示略)、空調機(図示略)等の機械的な動作音、その他の音が含まれる。つまり、マイクアレイ20により集音される音は、可聴周波数(つまり、20Hz~20kHz)域の音に限らず、可聴周波数より低い低周波音、可聴周波数を超える超音波音を含んでもよい。

40

【0018】

カメラ30は、監視対象エリアSAを画角内に収めるように設置されており、監視対象エリアSAを撮像する。カメラ30は、画角が固定された固定カメラでよいし、監視対象エリアSAを含む全域を撮像可能な魚眼レンズを搭載したカメラ(つまり、全方位カメラ)でもよい。

【0019】

50

処理閲覧装置 50 は、マイクアレイ 20 により集音された音の音データを基に、監視対象エリア SA において生じた音声あるいは音の音源位置を推定するとともに、その推定された音源から生じた音声あるいは音の種類を判別（特定）する。処理閲覧装置 50 は、音源から生じた音声あるいは音の大きさを示す音パラメータ（後述参照）を、カメラ 30 で撮像された監視対象エリア SA の撮像画像に重畳して表示する。

【0020】

レコーダ 70 は、記録データ md（図 2 参照）を蓄積する。記録データ md は、マイクアレイ 20 により集音された音の音データ sd と、カメラ 30 により撮像された画像の画像データ gd とを対応付けた構成である。なお、レコーダ 70 は、ネットワーク NW に接続されることなく、USB（Universal Serial Bus）等のインターフェースを介して、処理閲覧装置 50 に直接接続されてもよい。

10

【0021】

図 2 は、音源可視化システム 5 のシステム構成ならびに処理閲覧装置 50 のハードウェア構成の一例を示すブロック図である。処理閲覧装置 50 は、例えば PC（Personal Computer）55 と、モニタ 61 と、操作デバイス 62 と、スピーカ 63 とを有する。PC 55 は、プロセッサ 56 と、メモリ 57 と、外部インターフェース 60 とを有する。

【0022】

メモリ 57 は、ROM（Read Only Memory）58 および RAM（Random Access Memory）59 を含む。RAM 59 は、プロセッサ 56 のワーキングメモリとして使用され、プロセッサ 56 が ROM 58 から読み出した各種の実行ファイル在处理中に生成したデータ等を一時的に保持する。ROM 58 は、プロセッサ 56 により読み出されて実行される各種の実行ファイル（例えば、後述する音圧ヒートマップ処理部 561、音源位置推定部 562、強調音生成部 563、音分析部 564、ヒートマップ色付け部 565 のそれぞれを機能的に構成するためのプログラムおよびデータ）を格納する。

20

【0023】

プロセッサ 56 は、CPU（Central Processing Unit）、MPU（Micro Processing Unit）、DSP（Digital Signal Processor）、FPGA（Field Programmable Gate Array）を用いて構成される。プロセッサ 56 は、ROM 58 に格納された各種の実行ファイル（上述参照）を読み込んで協働することにより、各種の実行ファイルに対応した構成を機能的に実行する。プロセッサ 56 は、各種の実行ファイルに対応した構成として、音圧ヒートマップ処理部 561、音源位置推定部 562、強調音生成部 563、音分析部 564、およびヒートマップ色付け部 565 のそれぞれの機能を実行可能である。

30

【0024】

音圧ヒートマップ処理部 561（ヒートマップ処理部の一例）は、マイクアレイ 20 を構成するそれぞれのマイクロホンから出力される監視対象エリア SA における音データを基に、監視対象エリア SA が撮像された画像に対応して、音の大きさを示す音パラメータ（例えば音圧）を割り当てた音圧ヒートマップを生成する。具体的に、音圧ヒートマップ処理部 561 は、監視対象エリア SA で生じている音の大きさを示す音パラメータ（例えば音圧）を、監視対象エリア SA の撮像画像を構成する画素あるいは複数の画素からなるブロックごとに算出する。ブロックは、例えば  $2 \times 2$  個もしくは  $4 \times 4$  個の画素により構成される。 $*$ （アスタリスク）は、乗算の演算子である。音圧ヒートマップ処理部 561 は、画素あるいはブロックごとに算出された音パラメータの値を、視覚的な画像に変換して撮像画像に重畳することで、音圧ヒートマップを生成する。

40

【0025】

音源位置推定部 562 は、音圧ヒートマップ処理部 561 により生成された音圧ヒートマップを構成する音パラメータ（例えば音圧）のデータを正規化处理し、正規化处理された音パラメータ（例えば音圧）のデータを基にして、監視対象エリア SA における音源位置を推定する。音源位置推定部 562 は、例えば公知の白色化相互相関法（CSP（Cross-power Spectrum Phase analysis）法）に従って、マイクアレイ 20 により集音された音データを用いて、監視対象エリア SA での音源位置を推定する。CSP 法では、音源

50

位置推定部 562 は、監視対象エリア SA を複数のブロック（例えば、 $2 \times 2$  個もしくは  $4 \times 4$  個の画素により構成）に分割する。そして、音源位置推定部 562 は、マイクアレイ 20 で音が集音されると、それぞれのブロックごとに音の大きさを示す音パラメータ（例えば、マイクアレイ 20 を構成する複数のマイクロホンのうち、複数組からなる 2 つのマイクロホンによりそれぞれ集音された音データ間の相互相関値の正規化出力値）が閾値（既定値）を超えるか否かを判定することで、監視対象エリア SA 内での音源位置を概略的に推定できる。これにより、音源位置推定部 562 は、音源位置を推定する際、音圧データの生データ（Raw Data）を使用できる。従って、音源位置推定部 562 は、例えば従来技術のようにフィルタ等を使用して集音された生データ（つまり、音圧データ）を歪ませることなく、音源位置が音圧データに埋もれてしまうことを抑制できて的確に音源位置を推定できる。

10

#### 【0026】

強調音生成部 563 は、マイクアレイ 20（具体的には、マイクアレイ 20 の各マイクロホンの位置）から、音源位置推定部 562 により推定された音源位置に向かう指向方向に音データの指向性を形成することで、その指向方向における音データの強調処理を行う。また、強調音生成部 563 は、マイクアレイ 20 により集音された全方位の音データに、例えばユーザの操作に基づいて指定された任意の方向を主ビーム方向とする指向性をビームフォーミング（つまり、信号処理）によって形成でき、その指向方向の音データを強調できる。なお、マイクアレイ 20 によって集音された音をビームフォーミングして音データに指向性を形成する処理は、公知技術である。

20

#### 【0027】

音データの強調処理として、既知の遅延和方式を用いる場合、例えば、音源から発した音波は、各マイクロホンにそれぞれ到達して集音されるまでに到達時間差がある。強調音生成部 563 は、マイクアレイ 20 を構成する各マイクロホンが集音した音データに対し、各マイクロホンにおける音の到達時間差に対応する遅延時間を各遅延器で与えて全ての音波の位相を揃えた後、遅延処理後の音データを加算器で加算する。これにより、強調音生成部 563 は、所望の指向方向に音データの指向性を形成できる。

#### 【0028】

音分析部 564 は、強調音生成部 563 により指向方向に強調処理された音データを分析する。音データの分析では、ここでは、人の発話と物音とが判別される。発話の場合、人の声は、倍音を含む調波構造を有し、また、時間によって音圧が変化する特性を有する。一方、物音は、例えばホワイトノイズのように、時間的にあまり変化しない特性（白色性）を有する。音分析部 564 は、強調処理された音データを基に、上記特性の有無を判定することによって、発話か物音かを判別する。なお、音データの分析では、人の声と物音以外、例えば犬等の動物が吠える声を判別できるようにしてもよい。犬が吠える声は、人の声に含まれない、可聴音域を超えるような音成分を含む。音分析部は、強調処理した音が可聴音域を超えるような音成分を含む場合、犬の声と判断してもよい。

30

#### 【0029】

また、音分析部 564 は、人工知能（AI: Artificial Intelligent）を用いた機械学習によって音分析を行ってよい。例えば、プロセッサは、各種の音源から発せられる発話や物音等の音データを教師データとして、処理閲覧装置 50 に保持されるストレージ（図示略）に蓄積しておく。プロセッサ 56 は、上記教師データを用いてディープラーニング等の機械学習を行い、学習モデルを形成するニューラルネットワークの中間層のニューロンの数や重み付け係数を更新し、学習済みモデルを生成あるいは更新する。プロセッサ 56 は、この学習済みモデルに対し、音データを入力し、音源の種類を出力する。なお、上述した機械学習は、1 つ以上の統計的分類技術を用いて行っても良い。統計的分類技術としては、例えば、線形分類器（linear classifiers）、サポートベクターマシン（support vector machines）、二次分類器（quadratic classifiers）、カーネル密度推定（kernel estimation）、決定木（decision trees）、人工ニューラルネットワーク（artificial neural networks）、ベイジアン技術及び / 又はネットワーク（Bayesian techniques

40

50

and/or networks)、隠れマルコフモデル(hidden Markov models)、バイナリ分類子(binary classifiers)、マルチクラス分類器(multi-class classifiers)クラスタリング(a clustering technique)、ランダムフォレスト(a random forest technique)、ロジスティック回帰(a logistic regression technique)、線形回帰(a linear regression technique)、勾配ブースティング(a gradient boosting technique)などが挙げられる。ただし、使用される統計的分類技術はこれらに限定されない。

【0030】

ヒートマップ色付け部565は、音分析部564による音データの分析結果を基に、音圧ヒートマップ処理部561により生成された音圧ヒートマップの視覚的な画像に色付け処理を施す。ヒートマップ色付け部565は、音源の種類(ここでは、人の発話と物音)に応じて、画素あるいは複数の画素からなるブロックごとに、音圧ヒートマップの視覚的な画像に対し、赤色系の色(赤色、オレンジ色、黄色)、青色系の色(紺色、青色、緑色)で色付ける。また、ヒートマップ色付け部565は、音圧の大きさに応じて、色付けを変化させる。例えば、ヒートマップ色付け部565は、発話を赤色系の色で着色する場合、音圧が閾値TH1以上である場合、赤色に着色し、閾値TH1未満で閾値TH2以上である場合、オレンジ色で着色し、閾値TH2未満で閾値TH3以上である場合、黄色で着色する。同様に、ヒートマップ色付け部565は、物音を青色系の色で着色する場合、音圧が閾値TH1以上である場合、紺色に着色し、閾値TH1未満で閾値TH2以上である場合、青色で着色し、閾値TH2未満で閾値TH3以上である場合、緑色で着色する。

【0031】

なお、ヒートマップ色付け部565は、音圧の大きさによることなく、音源の種類ごとに単色で表示してもよい。例えば、発話である場合、閾値TH2以上である音に対し、赤色のみで着色し、物音である場合、閾値TH2以上である音に対し、青色のみで着色してもよい。また、ヒートマップ色付け部565は、音源の種類を区別することなく、音圧が閾値TH1以上である場合、赤色に着色し、閾値TH1未満で閾値TH2以上である場合、ピンク色で着色し、閾値TH2未満で閾値TH3以上である場合、青色で着色してもよい。

【0032】

外部インターフェース60は、処理閲覧装置50の通信部の一例としての役割を有し、ネットワークNWを介して、マイクアレイ20、カメラ30またはレコーダ70との間でデータあるいは情報の送受信(通信)を行う、通信回路により構成される。また、外部インターフェース60は、処理閲覧装置50を構成するモニタ61、操作デバイス62およびスピーカ63との間のデータあるいは情報の入出力を行う。

【0033】

モニタ61は、カメラ30により撮像された画像、この画像に対してPC50において色付けされて生成された音圧ヒートマップが重畳された画像を表示する。モニタ61は、例えば液晶ディスプレイ(Liquid Crystal Display)、有機EL(Electroluminescence)ディスプレイ、プラズマディスプレイ等を用いて構成される。なお、モニタ61は、PC55に外付けされる構成に限らず、PC55に内蔵されて構成されてもよい。

【0034】

操作デバイス62は、監視員(音源可視化システム5のユーザの一例)により使用され、PC55に対し、監視対象エリアSAにおける音源を可視化するための各種の入力操作を受け付けてPC55に出力する。操作デバイス62は、例えばマウス、キーボード、タッチパッド、トラックボール等の入力機器を用いて構成される。

【0035】

スピーカ63は、音声を出力する。なお、スピーカ63は、PC55に外付けされる構成に限らず、PC55に内蔵されて構成されてもよい。

【0036】

次に、実施の形態1に係る音源可視化システム5の動作を示す。

【0037】

先ず、監視対象エリア S A で発せられる音を可視化する動作を示す。図 1 では、音源として、会話する 2 人の発話と、コピー機 1 2 0 の作動音とが例示されている。

【 0 0 3 8 】

図 3 は、処理閲覧装置 5 0 における音源可視化処理の動作手順例を示すフローチャートである。図 3 において、音圧ヒートマップ処理部 5 6 1 は、マイクアレイ 2 0 を構成するそれぞれのマイクロホンから出力される監視対象エリア S A における音データを基に、監視対象エリア S A が撮像された画像に対応して、音の大きさを示す音パラメータ（例えば音圧）を割り当てた音圧ヒートマップを生成する（ S 1 ）。

【 0 0 3 9 】

音源位置推定部 5 6 2 は、ステップ S 1 において生成された音圧ヒートマップを基に、監視対象エリア S A における音源位置を推定する（ S 2 ）。ステップ S 2 における音源位置の推定処理の詳細については、後述する。なお、音源位置推定部 5 6 2 は、音圧ヒートマップを構成する画素あるいはブロック（上述参照）ごとの音パラメータ（例えば音圧）のピーク位置から音源位置を推定してもよいし、ユーザがモニタ 6 1 に表示される撮像画像を見ながら任意に指定した位置を音源位置として推定してもよい。

【 0 0 4 0 】

強調音生成部 5 6 3 は、マイクアレイ 2 0（具体的には、マイクアレイ 2 0 の各マイクロホンの位置）から、ステップ S 2 において推定された音源位置に向かう指向方向に音データの指向性を形成することで、その指向方向における音データの強調処理を行う（ S 3 ）。音分析部 5 6 4 は、ステップ S 3 において指向方向の音データが強調処理された後、その音データを分析する（ S 4 ）。図 1 では、会話する 2 人の人物 h m 1 , h m 2 の声（発話）と、コピー機 1 2 0 の作動音（物音）とが判別される。

【 0 0 4 1 】

ヒートマップ色付け部 5 6 5 は、ステップ S 4 における音データの分析結果を基に、ステップ S 1 において生成された音圧ヒートマップの視覚的な画像に色付け処理を施す（ S 5 ）。ヒートマップ色付け部 5 6 5 は、音源の種類と音圧の大きさに応じて、画素単位あるいは所定の画素ブロック単位に、音圧ヒートマップの視覚的な画像に色付けする（図 8 ~ 図 1 0 参照）。

【 0 0 4 2 】

具体的に、ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 3 以上である発話の領域 V 1 , V 2 に対し、音パラメータ（例えば音圧）が閾値 T H 1 以上である範囲を、赤色（図中右上がりの密なハッチ）で着色する。ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 1 未満で閾値 T H 2 以上である範囲を、オレンジ色（図中右上がりの中間ハッチ）で着色する。ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 2 未満で閾値 T H 3 以上である場合、黄色（図中右上がりの粗いハッチ）で着色する。

【 0 0 4 3 】

同様に、ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 3 以上である物音の領域 M 1 に対し、音パラメータ（例えば音圧）が閾値 T H 1 以上である範囲を、紺色（図中右下がりの密なハッチ）に着色する。ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 1 未満で閾値 T H 2 以上である範囲を、青色（図中右下がりの中間ハッチ）で着色する。ヒートマップ色付け部 5 6 5 は、音パラメータ（例えば音圧）が閾値 T H 2 未満で閾値 T H 3 以上である範囲を、緑色（図中右下がりの粗いハッチ）で着色する。なお、実施の形態 1 では、発話と物音とで閾値 T H 1 , T H 2 , T H 3 の値は同じであったが、異なる値であってもよい。また、発話と物音とで音圧と比較される閾値 T H 1 , T H 2 , T H 3 の数は、同じであったが、異なる数であってもよい。これにより、音圧ヒートマップの色付けを任意に設定できる。

【 0 0 4 4 】

ヒートマップ色付け部 5 6 5 は、色付けされた音圧ヒートマップを撮像画像の画素あるいは画素ブロック単位に展開して R A M 5 9 に記憶する。プロセッサ 5 6 は、全ての音源

10

20

30

40

50

位置に対し、S 3 ~ S 5 の処理を行ったか否かを判別する ( S 6 )。全ての音源位置に対して上記処理を行っていない場合、プロセッサ 5 6 の処理はステップ S 3 に戻り、プロセッサ 5 6 は、同様の処理を繰り返す。一方、ステップ S 6 で全ての音源位置に対して上記処理を行った場合、プロセッサ 5 6 は、R A M 5 9 に展開された音圧ヒートマップが重畳された撮像画像を、ユーザの設定に従ってモニタ 6 1 に表示する ( S 7、図 8 ~ 図 1 0 参照)。この後、プロセッサ 5 6 は、図 3 に示す処理を終了する。

#### 【 0 0 4 5 】

図 4 は、ステップ S 2 の音源位置推定処理の動作手順例を示すフローチャートである。音源位置推定部 5 6 2 は、音圧ヒートマップ処理部 5 6 1 により生成された音圧ヒートマップを構成する音パラメータ (例えば音圧) のデータを正規化する ( S 1 1 )。この正規化方法として、例えば、一般によく知られている Z - s c o r e ( Z 値) が用いられる。音圧データの Z 値は、音圧データ群の数値から平均値を引いて、標準偏差で割ることで求められる。音圧データの Z 値群では、平均値が値 0、標準偏差が値 1 になる。従って、音圧が大きく異なる音圧データ群において、数値の比較が可能となり、閾値 T H を共通に設定できる。

#### 【 0 0 4 6 】

音源位置推定部 5 6 2 は、音パラメータ (例えば音圧) が閾値 T H m 以上の範囲の音データを抽出する ( S 1 2 )。閾値 T H m は、音データを抽出するために、予め設定された固定値である。閾値 T H m は、音圧ヒートマップを生成するために設定される、最小の閾値 T H 3 と同じ値であってもよいし、異なる値であってもよい。

#### 【 0 0 4 7 】

音源位置推定部 5 6 2 は、抽出された音データの範囲ごとにピーク位置を抽出する ( S 1 3 )。ピーク位置の抽出は、公知の技術を用いて行われる。例えば、抽出した範囲の音データに含まれるノイズの影響を除去した後、ノイズ除去後の音データを微分し、値 0 を横切る位置をピーク位置として抽出する方法が挙げられる。

#### 【 0 0 4 8 】

音源位置推定部 5 6 2 は、抽出した音データの範囲を全て抽出したか否かを判別する ( S 1 4 )。全て抽出していない場合、音源位置推定部 5 6 2 は、ステップ S 1 3 の処理に戻る。一方、全て抽出した場合、音源位置推定部 5 6 2 は、図 4 に示す処理を終了する。

#### 【 0 0 4 9 】

図 5 は、撮像画像 G Z 1 上の推定された音源位置および音を強調するための指向方向を示す図である。なお、図 5 では、撮像画像 G Z 1 には、音ヒートマップを色付けする単位として、複数の画素あるいは画素ブロックが格子状に便宜的に描画されているが、モニタ 6 1 に表示される画像には含まれない。

#### 【 0 0 5 0 】

撮像画像 G Z 1 上には、2 人の人物 h m 1 , h m 2 が会話することで抽出される発話の領域 V 1 , V 2 と、コピー機 1 2 0 が作動することで抽出される物音の領域 M 1 とが設定される。この場合、人物 h m 1 , h m 2 の口元付近が音源位置 s p 1 , s p 2 となる。また、コピー機 1 2 0 の作動音を発する駆動機構が音源位置 s p 3 となる。

#### 【 0 0 5 1 】

マイクアレイ 2 0 は、撮像画像 G Z 1 の中心 o に位置する。プロセッサ 5 6 は、中心位置 o から方向 d 1 に音データの指向性を形成すると、音源位置 s p 1 である人物 h m 1 が発話する音声を強調処理できる。同様に、プロセッサ 5 6 は、中心位置 o から方向 d 2 に指向性を形成すると、音源位置 s p 2 である人物 h m 2 が発話する音声を強調処理できる。プロセッサ 5 6 は、中心位置 o から方向 d 3 に指向性を形成すると、コピー機 1 2 0 の作動音 (物音) を強調処理できる。

#### 【 0 0 5 2 】

図 6 は、音圧を正規化しないで音源位置を推定するための音圧分布例を示す図である。図 6 ( A ) ~ 図 6 ( D ) の各グラフにおいて、縦軸は音圧を示し、横軸は中心位置 o の鉛直方向から方位 d a に沿った音源方向 (角度) を示す。

10

20

30

40

50

## 【 0 0 5 3 】

図 6 ( A ) の音圧分布では、マイクアレイ 2 0 で集音された音の音データ ( 生データ ) の音圧が閾値  $THa$  以上である音のピーク波形  $p_1$  ,  $p_2$  が 2 つ現れる。小さな音のピーク波形  $p_1$  は、コピー機 1 2 0 の作動音による。大きな音のピーク波形  $p_2$  は、人物  $hm_1$  による発話による。閾値  $THa$  は、2 つのピーク波形  $p_1$  ,  $p_2$  のレベルより小さい。従って、図 6 ( B ) の音圧分布では、閾値  $THa$  以上の 2 つのピーク波形  $p_1$  ,  $p_2$  の部分が、音源のある範囲  $ra_1$  ,  $ra_2$  として、音源位置推定部 5 6 2 によって抽出される。

## 【 0 0 5 4 】

一方、図 6 ( C ) の音圧分布では、図 6 ( A ) の場合と比べて、全体の音圧レベルが下がっている。この場合、音圧が閾値  $THa$  以上である音のピーク波形  $p_4$  が 1 つ現れる。小さな音のピーク波形  $p_3$  は、コピー機 1 2 0 の作動音による。コピー機 1 2 0 が待機状態にあり、ピーク波形  $p_3$  の音圧は小さい。ピーク波形  $p_3$  と比べて大きな音のピーク波形  $p_4$  は、人物  $hm_1$  による発話による。閾値  $THa$  は、図 6 ( A ) の場合と同じ値である。従って、図 6 ( D ) の音圧分布では、閾値  $THa$  以上の 1 つのピーク波形  $p_4$  の部分が、音源のある範囲  $ra_4$  として、音源位置推定部 5 6 2 によって抽出される。

## 【 0 0 5 5 】

図 7 は、音圧を正規化して音源位置を推定するための音圧分布例を示す図である。図 7 ( A ) ~ 図 7 ( D ) の各グラフにおいて、縦軸は音圧を示し、横軸は中心位置  $o$  の鉛直方向から方位  $da$  に沿った音源方向 ( 角度 ) を示す。マイクアレイ 2 0 で集音された音の音データ ( 生データ ) は、音源位置推定部 5 6 2 によって正規化される。

## 【 0 0 5 6 】

図 7 ( A ) の音圧分布では、正規化された音データの音圧が閾値  $THb$  以上である音のピーク波形  $p_5$  ,  $p_6$  が 2 つ現れる。小さな音のピーク波形  $p_5$  は、コピー機 1 2 0 の作動音による。大きな音のピーク波形  $p_6$  は、人物  $hm_1$  による発話による。閾値  $THb$  は、2 つのピーク波形  $p_5$  ,  $p_6$  のレベルより小さい。従って、図 7 ( B ) の音圧分布では、閾値  $THb$  以上の 2 つのピーク波形  $p_5$  ,  $p_6$  の部分が、音源のある範囲  $ra_5$  ,  $ra_6$  として、音源位置推定部 5 6 2 によって抽出される。

## 【 0 0 5 7 】

一方、図 7 ( C ) の音圧分布では、図 6 ( D ) と同様、全体の音圧レベルが下がっている。しかし、正規化された音データは、図 7 ( A ) と同様の大きさを有する。従って、この場合も、音圧が閾値  $THb$  以上である音のピーク波形  $p_7$  ,  $p_8$  が 2 つ現れる。小さな音のピーク波形  $p_7$  は、コピー機 1 2 0 の作動音による。コピー機 1 2 0 が待機状態にあり、音データの音圧は小さいが、正規化された音圧の値は変わらない。大きな音のピーク波形  $p_8$  は、人物  $hm_1$  による発話による。閾値  $THb$  は、図 7 ( A ) の場合と同じ値である。従って、図 7 ( D ) の音圧分布では、閾値  $THb$  以上の 2 つのピーク波形  $p_7$  ,  $p_8$  の部分が、音源のある範囲  $ra_7$  ,  $ra_8$  として、音源位置推定部 5 6 2 によって抽出される。

## 【 0 0 5 8 】

このように、正規化された音データ場合、音圧が相対値で表されるので、音圧が小さいピーク波形であっても、閾値を可変させることなく、音源位置が音圧データに埋もれることなく推定できる。

## 【 0 0 5 9 】

なお、図 6、図 7 では、マイクアレイ 2 0 の中心位置  $o$  の鉛直方向から方位  $da$  に沿った音源方向の音圧を 2 次元分布で示したが、中心位置  $o$  の鉛直方向から周囲の音源方向の音圧を 3 次元分布で表してもよい。

## 【 0 0 6 0 】

図 8 は、発話および物音の音圧ヒートマップ  $sh$  が撮像画像  $GZ_1$  に重畳して表示されたモニタ 6 1 の表示画面例を示す図である。モニタ 6 1 の画面には、音圧ヒートマップ  $sh$  が重畳された撮像画像  $GZ_1$  の他、ヒートマップ設定画面  $w_1$  が表示される。

## 【 0 0 6 1 】

10

20

30

40

50

ヒートマップ設定画面 w 1 には、発話のチェックボックス c b 1 および物音のチェックボックス c b 2 が表示される。ユーザは、発話のチェックボックス c b 1 および物音のチェックボックス c b 2 に対し、チェックマークを挿入可能である。発話のチェックボックス c b 1 および物音のチェックボックス c b 2 の少なくとも一方にチェックマークが入っている場合、チェックマークが入っている音源に対応する音圧ヒートマップ s h が撮像画像 G Z 1 に重畳して表示される。

【 0 0 6 2 】

図 8 では、発話のチェックボックス c b 1 および物音のチェックボックス c b 2 の双方に、チェックマークが入っている。従って、図 8 では、人物 h m 1 の発話による音圧ヒートマップ s h の着色画像 s h 1、人物 h m 2 の発話による音圧ヒートマップ s h の着色画像 s h 2、およびコピー機 1 2 0 の作動音による音圧ヒートマップ s h の着色画像 s h 3 が撮像画像 G Z 1 に重畳される。

10

【 0 0 6 3 】

図 9 は、発話の音圧ヒートマップ s h が撮像画像 G Z 1 に重畳して表示されたモニタ 6 1 の表示画面例を示す図である。図 9 では、発話のチェックボックス c b 1 にチェックマークが入っているが、物音のチェックボックス c b 2 には、チェックマークが入っていない。従って、図 9 では、人物 h m 1 の発話による音圧ヒートマップ s h の着色画像 s h 1、および人物 h m 2 の発話による音圧ヒートマップ s h の着色画像 s h 2 が撮像画像 G Z 1 に重畳される。

【 0 0 6 4 】

20

図 1 0 は、物音の音圧ヒートマップ s h が撮像画像 G Z 1 に重畳して表示されたモニタ 6 1 の表示画面例を示す図である。図 1 0 では、物音のチェックボックス c b 2 には、チェックマークが入っているが、発話のチェックボックス c b 1 にチェックマークが入っていない。従って、図 1 0 では、コピー機 1 2 0 の作動音による音圧ヒートマップ s h の着色画像 s h 3 が撮像画像 G Z 1 に重畳される。

【 0 0 6 5 】

このように、実施の形態 1 に係る音源可視化システム 5 によれば、音源位置を推定する場合、フィルタ処理されていないローデータ（生データ）を使用することで、音データが例えばフィルタ処理で歪んでしまうことにより音源が的確に推定できなくなることを防止できる。また、音源可視化システム 5 によれば、処理閲覧装置 5 0 は、音圧が閾値を超えることで音源を推定する場合、音圧の小さな音源の音であっても、音圧を正規化することで、正規化された音圧が閾値を超えるようになり、音圧の小さな音を発する音源でも、音源位置を推定可能である。一方、処理閲覧装置 5 0 は、推定された音源の種類を判別する場合、音源の方向に指向性を形成して音源が発する音を強調することで、音を精度良く分析できる。従って、音源の種類を正確に判別できるようになる。

30

【 0 0 6 6 】

以上により、実施の形態 1 に係る音源可視化システム 5 では、処理閲覧装置 5 0 は、外部インターフェース 6 0（通信部の一例）を介して、監視対象エリア S A に配置されたマイクアレイ 2 0 およびカメラ 3 0 との間でそれぞれ通信する。音源位置推定部 5 6 2（推定部の一例）は、マイクアレイ 2 0 により集音された音の音データに基づいて、監視対象エリア S A 内に存在する音源位置を推定する。音分析部 5 6 4（分析部の一例）は、マイクアレイ 2 0 により集音された音の音データに基づいて、推定された音源位置から生じた音の種類を分析する。プロセッサ 5 6（制御部の一例）は、推定された音源位置から生じた音の種類の分析結果を、カメラ 3 0 により撮像された画像に重畳してモニタ 6 1 に表示する。

40

【 0 0 6 7 】

これにより、音源可視化システム 5 によれば、処理閲覧装置 5 0 は、監視対象エリア S A において集音された音源からの音の種類を具体的に判別できるので、監視精度の劣化ならびに監視員の利便性の低下を抑制できる。

【 0 0 6 8 】

50

また、処理閲覧装置 50 は、カメラ 30 により撮像された画像を構成する画素または所定数の画素からなる画素ブロックごとに、監視対象エリア S A の音圧（音の大きさを示す音パラメータの一例）を導出し、その導出された音パラメータを対応する画素または画素ブロックに割り当てた音圧ヒートマップ h p を生成する音圧ヒートマップ処理部 561、をさらに備える。プロセッサ 56 は、音圧ヒートマップ h p 上の推定された音源の位置に対応する音圧レベル（音パラメータの一例）を、着色画像 s h 1, s h 2, s h 3（分析結果の内容に応じた色を付与した視覚画像の一例）に変換して撮像画像 G Z 1 に重畳する。これにより、音源可視化システム 5 によれば、監視員は、監視対象エリア S A において音源から発せられる音の大きさを視覚的に把握でき、監視員による音の監視精度（つまり、監視品質）を的確に向上できる。

10

**【0069】**

また、音源位置推定部 562 は、生成された音圧ヒートマップ h p を構成する音パラメータ（例えば音圧）のデータを正規化し、正規化された音圧ヒートマップ h p を構成する画素または画素ブロックごとの音圧レベルと既定の閾値 T H m との比較に基づいて、監視対象エリア S A 内に存在する音源の位置を推定する。これにより、音源可視化システム 5 によれば、処理閲覧装置 50 は、正規化された音データ（例えば音圧）が相対値で表されるので、音圧が小さいピーク波形であっても、閾値を変化することなく、音圧データに埋もれることなく音源位置を適正に推定できる。

**【0070】**

また、音分析部 564 は、音の種類の分析結果として、発話（人の声）または物音（物の音）を特定する。これにより、音源可視化システム 5 によれば、処理閲覧装置 50 は、監視すべき音の対象を特定でき、監視員に対して効率良く音の監視を行わせることができる。

20

**【0071】**

また、音源は、監視対象エリア S A において複数存在する。これにより、音源可視化システム 5 によれば、処理閲覧装置 50 は、複数の音を同時に監視でき、監視員による監視効率を高めることができる。

**【0072】**

また、プロセッサ 56 は、音の種類の分析結果が予め指定された音の種類と一致した場合に、音の種類の分析結果をモニタ 61 に表示する。これにより、監視員は、所望の音を選択して監視でき、音源が多く存在する場合に効果的にそれぞれの音源から生じた音の内容や特徴を監視できる。

30

**【0073】**

また、処理閲覧装置 50 は、マイクアレイ 20 により集音された音データを用いて、マイクアレイ 20 から推定された音源の位置に向かう方向の音データを強調処理する強調音生成部 563、をさらに備える。これにより、音源可視化システム 5 によれば、処理閲覧装置 50 は、音源から発生した音の種類の分析精度を向上できる。

**【0074】**

以上、図面を参照しながら各種の実施の形態について説明したが、本開示はかかる例に限定されないことは言うまでもない。当業者であれば、特許請求の範囲に記載された範疇内において、各種の変更例、修正例、置換例、付加例、削除例、均等例に想到し得ることは明らかであり、それらについても当然に本開示の技術的範囲に属するものと了解される。また、発明の趣旨を逸脱しない範囲において、上述した各種の実施の形態における各構成要素を任意に組み合わせてもよい。

40

**【0075】**

例えば、上述した実施の形態 1 では、音圧ヒートマップを撮像画像に重畳して表示する際、音源の種類を判別可能に表示する他、音パラメータ（例えば音圧）を段階的に判別可能に表示したが、音パラメータ（例えば音圧）を段階的に表示することなく、音源の種類だけを判別可能に表示してもよい。具体的に、発話の音領域は、赤色（単色）で表示し、物音の音領域は青色（単色）で表示してもよい。

50

## 【 0 0 7 6 】

また、本開示は、上述した実施の形態 1 に係る処理閲覧装置 5 0 の機能を実現するプログラムを、ネットワークあるいは各種記憶媒体を介して処理閲覧装置 5 0 に供給し、コンピュータである処理閲覧装置 5 0 に内蔵されるプロセッサ 5 6 が読み出して実行するプログラム、及びこのプログラムが記憶された記録媒体も適用範囲である。

## 【産業上の利用可能性】

## 【 0 0 7 7 】

本開示は、監視対象エリアにおいて集音された音源からの音の種類を判別し、監視精度の劣化ならびに監視員の利便性の低下を抑制する音処理装置、音処理方法およびプログラムとして有用である。

10

## 【符号の説明】

## 【 0 0 7 8 】

5 音源可視化システム

1 0 カメラ付マイクアレイ

2 0 マイクアレイ

3 0 カメラ

5 0 処理閲覧装置

5 6 プロセッサ

5 7 メモリ

5 8 R O M

20

5 9 R A M

6 0 外部インターフェース

6 1 モニタ

6 2 操作デバイス

6 3 スピーカ

7 0 レコーダ

5 6 1 音圧ヒートマップ処理部

5 6 2 音源位置推定部

5 6 3 強調音生成部

5 6 4 音分析部

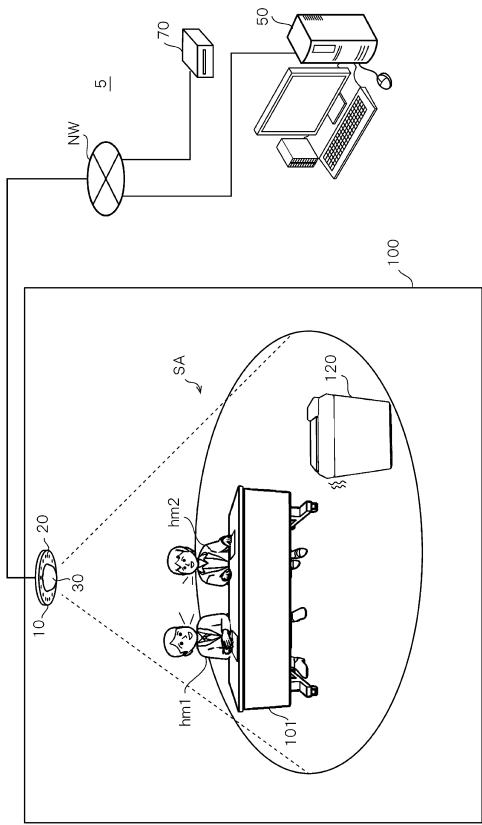
30

5 6 5 ヒートマップ色付け部

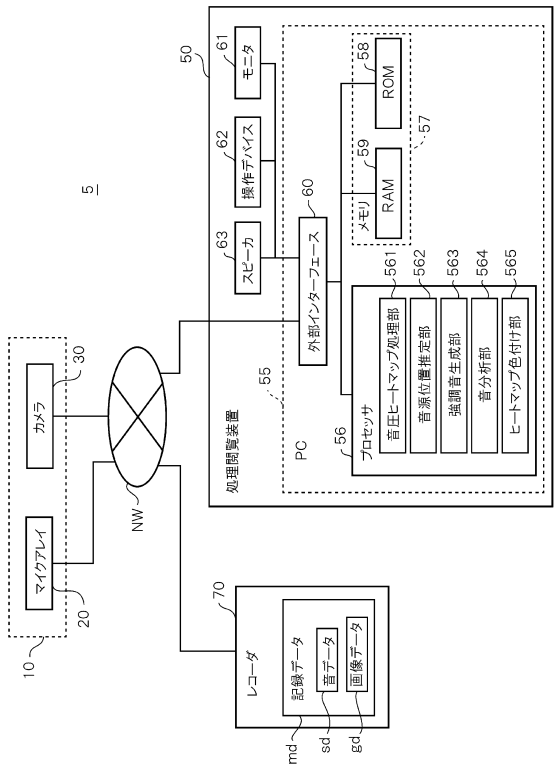
40

50

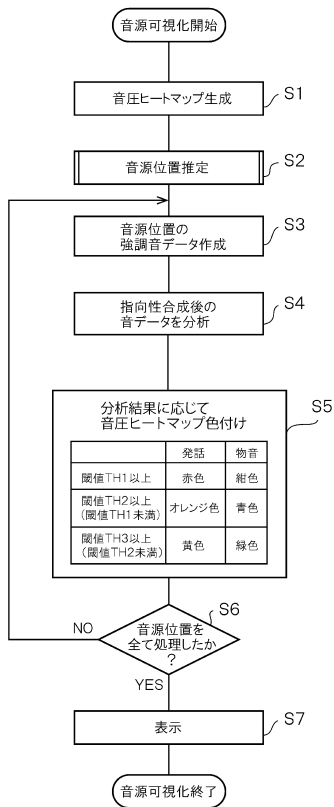
【図面】  
【図 1】



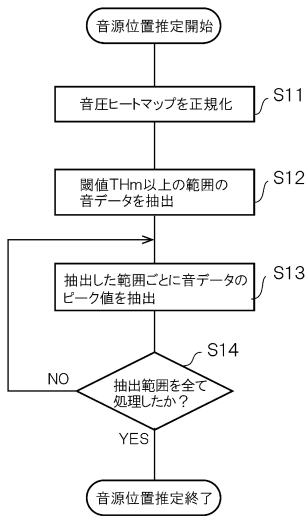
【図 2】



【図 3】



【図 4】



10

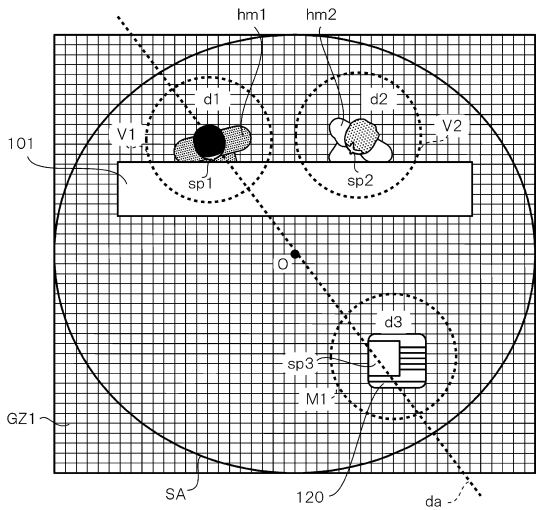
20

30

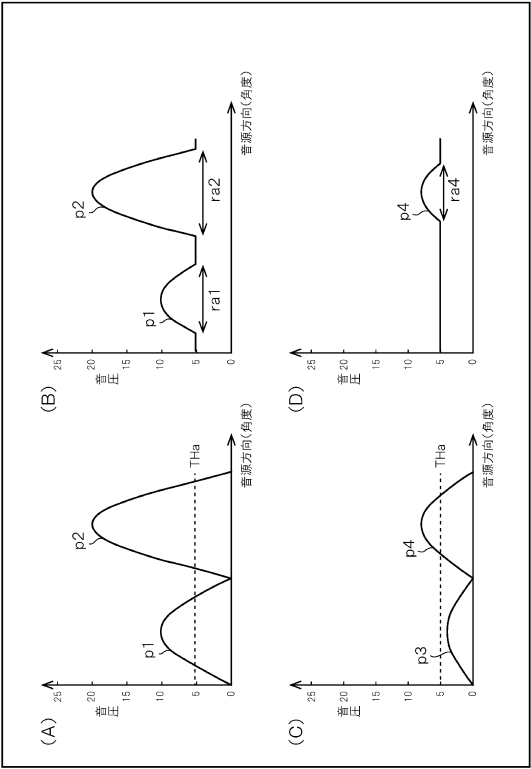
40

50

【図 5】



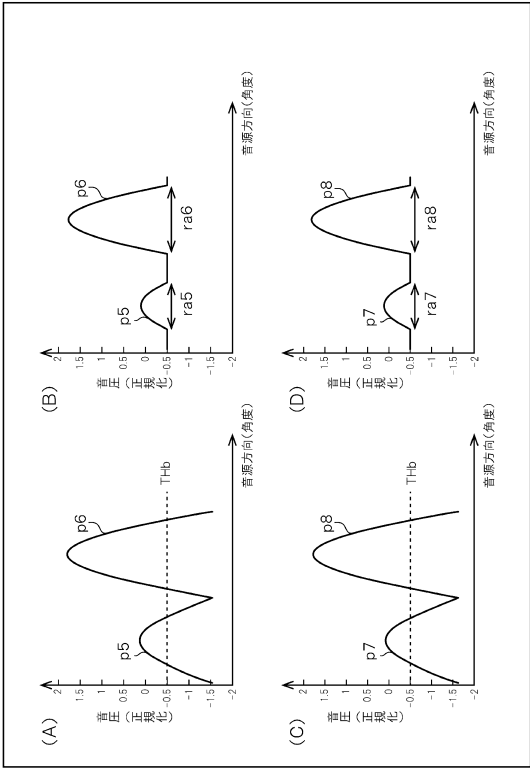
【図 6】



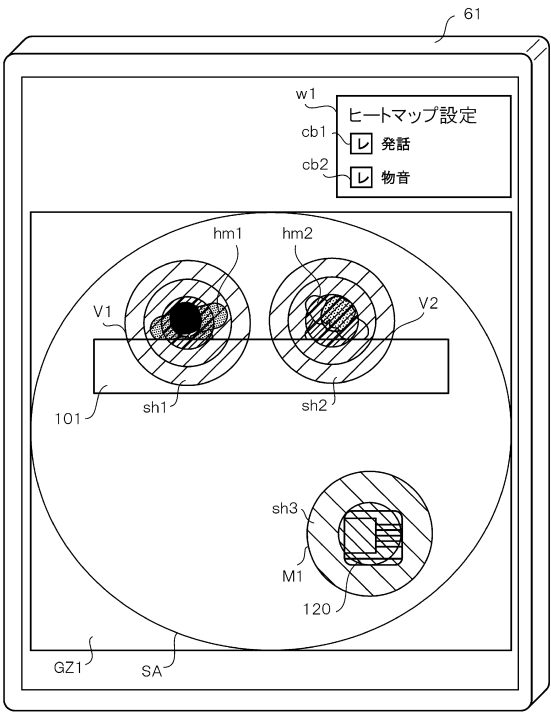
10

20

【図 7】



【図 8】

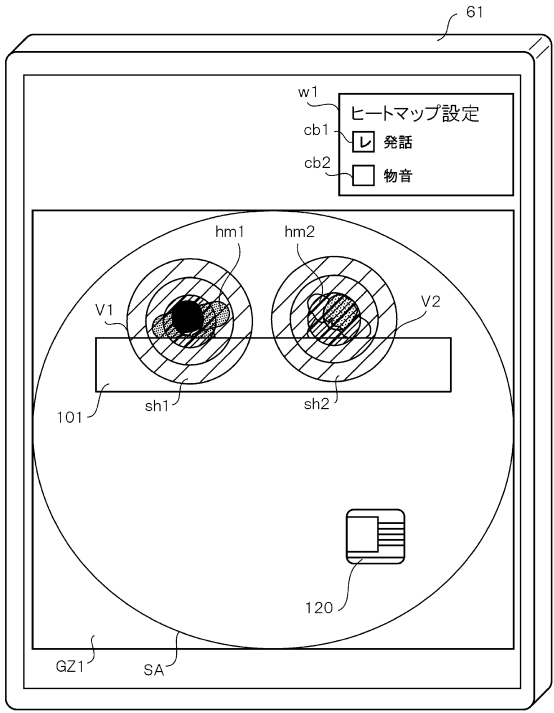


30

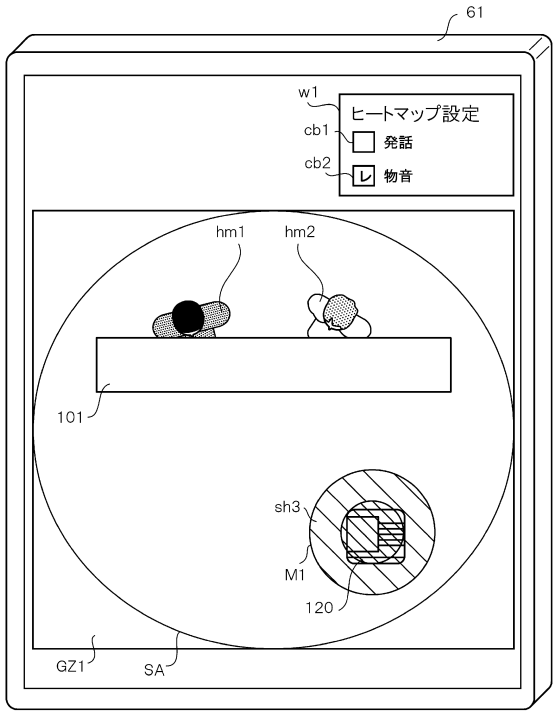
40

50

【図 9】



【図 10】



10

20

30

40

50

## フロントページの続き

- (56)参考文献 特開 2 0 1 7 - 0 2 2 6 6 4 ( J P , A )  
特開 2 0 1 3 - 1 0 6 2 9 8 ( J P , A )  
佐々木 洋子 Yoko Sasaki , “ 移動ロボットのための聴覚機能「モバイルオーディション」 ”  
 , 日本ロボット学会誌 , 日本 , 社団法人日本ロボット学会 The Robotics Society of Japan ,  
2010年01月15日 , Volume 28, Number 1 , Pages 14-17 , URL: <https://doi.org/10.7210/jrsj.28.14>
- (58)調査した分野 (Int.Cl. , D B 名)  
G 0 1 S 1 / 7 2 - G 0 1 S 1 / 8 2  
G 0 1 S 3 / 8 0 - G 0 1 S 3 / 8 6  
G 0 1 S 5 / 1 8 - G 0 1 S 5 / 3 0  
G 0 1 S 7 / 5 2 - G 0 1 S 7 / 6 4  
G 0 1 S 1 5 / 0 0 - G 0 1 S 1 5 / 9 6  
G 1 0 L 2 1 / 1 0 - G 1 0 L 2 1 / 1 4