

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5658588号  
(P5658588)

(45) 発行日 平成27年1月28日 (2015. 1. 28)

(24) 登録日 平成26年12月5日 (2014. 12. 5)

(51) Int. Cl. F I  
 H O 4 S 7/00 (2006. 01) H O 4 S 7/00 A  
 G 1 O L 21/10 (2013. 01) G 1 O L 21/10

請求項の数 9 (全 19 頁)

(21) 出願番号	特願2011-24296 (P2011-24296)	(73) 特許権者	000004352
(22) 出願日	平成23年2月7日 (2011. 2. 7)		日本放送協会
(65) 公開番号	特開2012-165195 (P2012-165195A)		東京都渋谷区神南2丁目2番1号
(43) 公開日	平成24年8月30日 (2012. 8. 30)	(74) 代理人	110001807
審査請求日	平成25年6月24日 (2013. 6. 24)		特許業務法人磯野国際特許商標事務所
特許法第30条第1項適用	(1) 社団法人日本音響学会から平成22年9月7日に発行された刊行物「日本音響学会2010年秋季研究発表会講演論文集」において発表 (2) 社団法人日本音響学会が平成22年9月14日~16日に開催した「日本音響学会2010年秋季研究発表会」において平成22年9月14日に文書をもって発表	(74) 代理人	100064414
			弁理士 磯野 道造
		(74) 代理人	100111545
			弁理士 多田 悦夫
		(72) 発明者	大出 訓史
			東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
		(72) 発明者	安藤 彰男
			東京都世田谷区砧一丁目10番11号 日本放送協会 放送技術研究所内
			最終頁に続く

(54) 【発明の名称】 聴覚臨場感評価装置及び聴覚臨場感評価プログラム

(57) 【特許請求の範囲】

【請求項1】

二つのマイクロホンによって計測された二つの音響信号に基づいて聴覚臨場感を評価する聴覚臨場感評価装置であって、

計測された前記二つの音響信号を周波数帯域ごとに分割する音響信号分割部と、

分割された前記二つの音響信号を用いて、前記二つの前記音響信号に関する相互相関関数を前記周波数帯域ごとに算出する相互相関関数算出部と、

算出された前記相互相関関数を、前記二つの音響信号のラグと、前記周波数帯域と、をフレームの座標軸とする二次元動画像の輝度、色相及び彩度のいずれかに変換することによって動画像を生成する動画像生成部と、

生成された前記動画像の複数のフレームを用いて前記動画像の移動ベクトルを算出することによって、音像の移動ベクトルを算出する音像移動ベクトル算出部と、

算出された前記音像の移動ベクトルに基づいて、前記音像の移動ベクトルが大きいほど聴覚臨場感評価値が大きくなるように当該聴覚臨場感評価値を算出する聴覚臨場感評価値算出部と、

を備えることを特徴とする聴覚臨場感評価装置。

【請求項2】

前記動画像生成部は、前記相互相関関数を前記輝度に変換する

ことを特徴とする請求項1に記載の聴覚臨場感評価装置。

【請求項3】

分割された前記二つの音響信号を用いて、前記二つの音響信号の音圧レベルを前記周波数帯域ごとに算出する音圧レベル算出部をさらに備え、

前記動画像生成部は、算出された前記音圧レベルを、前記輝度、前記色相及び前記彩度のうち、前記相互相関関数が変換されたもの以外のいずれかに変換することによって前記動画像を生成する

ことを特徴とする請求項 1 に記載の聴覚臨場感評価装置。

【請求項 4】

前記動画像生成部は、前記相互相関関数を前記輝度に変換するとともに、前記音圧レベルを前記彩度に変換する

ことを特徴とする請求項 3 に記載の聴覚臨場感評価装置。

10

【請求項 5】

前記音圧レベル算出部は、前記音圧レベルの差である音圧レベル差を前記周波数帯域ごとに算出し、

前記動画像生成部は、算出された前記音圧レベル差を、前記輝度、前記色相及び前記彩度のうち、前記相互相関関数及び前記音圧が変換されたもの以外に変換することによって前記動画像を生成する

ことを特徴とする請求項 3 に記載の聴覚臨場感評価装置。

【請求項 6】

前記動画像生成部は、前記相互相関関数を前記輝度に変換し、前記音圧レベルを前記彩度に変換するとともに、前記音圧レベル差を前記色相に変換する

ことを特徴とする請求項 5 に記載の聴覚臨場感評価装置。

20

【請求項 7】

前記相互相関関数算出部は、区間長の異なる複数の区間長別相互相関関数を算出し、算出された複数の区間長別相互相関関数に基づいて前記相互相関関数を算出する

ことを特徴とする請求項 1 から請求項 6 のいずれか一項に記載の聴覚臨場感評価装置。

【請求項 8】

前記映像移動ベクトル算出部は、前記動画像の三以上のフレームを用いて前記動画像の移動ベクトルを算出する

ことを特徴とする請求項 1 から請求項 7 のいずれか一項に記載の聴覚臨場感評価装置。

30

【請求項 9】

二つのマイクロホンによって計測された二つの音響信号に基づいて聴覚臨場感を評価する聴覚臨場感評価プログラムであって、

コンピュータを、

計測された前記二つの音響信号を周波数帯域ごとに分割する音響信号分割部、

分割された前記二つの音響信号を用いて、前記二つの前記音響信号に関する相互相関関数を前記周波数帯域ごとに算出する相互相関関数算出部、

算出された前記相互相関関数を、前記二つの音響信号のラグと、前記周波数帯域と、をフレームの座標軸とする二次元動画像の輝度、色相及び彩度のいずれかに変換することによって動画像を生成する動画像生成部、

生成された前記動画像の複数のフレームを用いて前記動画像の移動ベクトルを算出することによって、映像の移動ベクトルを算出する映像移動ベクトル算出部、及び、

算出された前記映像の移動ベクトルに基づいて、前記映像の移動ベクトルが大きいほど聴覚臨場感評価値が大きくなるように当該聴覚臨場感評価値を算出する聴覚臨場感評価値算出部、

40

として機能させることを特徴とする聴覚臨場感評価プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音響信号に基づいて聴覚における臨場感を評価する聴覚臨場感評価装置及び聴覚臨場感評価プログラムに関する。

50

## 【背景技術】

## 【0002】

複数話者が存在する場合の音声認識の前処理、ロボットの自律運動制御等に有用な技術として、音源方向の推定に関する研究が行われてきた（非特許文献1参照）。

かかる音源方向の推定技術は、多チャンネルマイクロホンによって構成されるマイクロホンアレー信号処理を行うものであり、複数のマイクロホンを配置するため、大規模な装置が必要であった。

## 【0003】

一方、人間は左右の耳で聞こえる2チャンネルの音響信号を用いて十分な精度の音源方向の推定を実現しており、これを模して、2チャンネルの音響信号を用いて音源方向の推定を行う研究も行われている（非特許文献2参照）。

かかる2チャンネルの音響信号を用いた音源方向の推定技術は、装置の小型化を図ることができるが、情報量が少ないため、推定の精度を確保するためには複雑な信号処理が必要であった（非特許文献3参照）。

## 【0004】

また、従来の2チャンネルの音響信号を用いた音源方向の推定技術は、一つの音源の方向しか推定することができず、また、推定対象が静止音源に限られていた（特許文献1参照）。

すなわち、従来の音源方向の推定技術は、高精度であることを目指すが故に装置として大規模になりがちで実時間処理が困難である上、一つの音源の方向しか推定することができないという問題を有していた。

## 【0005】

一方、符号化された音に対する客観的な評価法として、PEAQ（Perceptual Evaluation of Audio Quality）が開発されており、かかる評価法は、ITU（International Telecommunications Union：国際電気通信連合）によって規格化されている（ITU-R BS. 1387）。

かかる評価法は、ヒトの聴覚末梢系を模倣しているが、聴覚中枢系の部分に関しては、簡易なニューラルネットワークで代用している。

## 【0006】

また、ヒトが感じる音声品質の客観的な評価法として、特許文献2に記載された評価法が存在する。かかる評価法は、基準となる音に関する劣化度合いを客観的に評価する手法である。

## 【先行技術文献】

## 【特許文献】

## 【0007】

【特許文献1】特開平5 - 87903号公報

【特許文献2】特開2004 - 172753号公報

## 【非特許文献】

## 【0008】

【非特許文献1】Arthur N. Popper、Richard R. Fay、Eds., 「Sound Source Localization」、Springer Handbook of Auditory Research、2005年、Springer、New York

【非特許文献2】永田仁史、「2チャンネル信号に基づく到来音方向推定の計算モデル」、日本比較生理生化学会、2010年、第27巻、第1号、p.10-18

【非特許文献3】Nicoleta Roman、DeLiang Wang、「Binaural Tracking of Multiple Moving Sources」、IEEE Transactions on Audio, Speech, and Language Processing、2008年、Vol. 16、No. 4、p. 728-739

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0009】

収音再生技術の発展に伴い、2.2.2マルチチャンネル音響システム、Wave Field Synth

10

20

30

40

50

esis等といった高い臨場感を実現可能な音響システムが多数開発されており、音の品質の劣化度合いではなく、品質の向上度合いを客観的に評価することが求められている。

【0010】

ここで、K. Ozawa、Y. Chujo、Y. Suzuki、T. Sone、「Contents which yield high auditory-presence in sound reproduction」、Kansei Engineering International、2002年、Vol. 3、No. 4、p. 25-30、及び、K. Ozawa、Y. Chujo、「Content Presence vs. System Presence in Audio Reproduction Systems」、Proc. of the Second International Symposium on Universal Communication (ISUC2008)、2008年、p. 50-55に記載されているように、音響機器によって再生された音響信号による音像が、聴取者に対して相対的に移動している程、臨場感があると聴取者が感じる事が明らかになっている。

10

【0011】

本願発明者は、鋭意研究の結果、音の品質の向上度合いを客観的に評価するためには、聴覚における臨場感、すなわち聴覚臨場感を定量化して客観的に評価すればよいのではないかという考えに至った。

【0012】

本発明は、前記した事情に鑑みて創案されたものであり、聴覚臨場感を客観的に評価することが可能な聴覚臨場感評価装置及び聴覚臨場感評価プログラムを提供することを課題とする。

【課題を解決するための手段】

【0013】

前記課題を解決するため、本発明の聴覚臨場感評価装置は、二つのマイクによって計測された二つの音響信号に基づいて聴覚臨場感を評価する聴覚臨場感評価装置であって、音響信号分割部と、相互相関関数算出部と、動画像生成部と、音像移動ベクトル算出部と、聴覚臨場感評価値算出部と、を備えることを特徴とする。

20

【0014】

かかる構成によると、音響信号分割部によって、計測された前記二つの音響信号を例えばある一定時間ごとにある区間長だけ切り出し、周波数帯域ごとに分割し、相互相関関数算出部によって、分割された前記二つの音響信号を用いて、前記二つの前記音響信号に関する相互相関関数を前記周波数帯域ごとに算出する。そして、動画像生成部によって、算出された前記相互相関関数を、前記二つの音響信号のラグと、前記周波数帯域と、をフレームの座標軸とする二次元動画像の輝度、色相及び彩度のいずれかに変換することによって動画像を生成し、音像移動ベクトル算出部によって、生成された前記動画像の複数のフレームを用いて前記動画像の移動ベクトルを算出することによって、音像の移動ベクトルを算出し、聴覚臨場感評価値算出部によって、算出された前記音像の移動ベクトルに基づいて、前記音像の移動ベクトルが大きいほど聴覚臨場感評価値が大きくなるように当該聴覚臨場感評価値を算出することができる。

30

【0015】

前記動画像生成部は、前記相互相関関数を前記輝度に変換する構成であってもよい。かかる構成によると、相互相関関数と相関する輝度成分を有するモノクロ動画像を生成し、生成されたモノクロ動画像の移動ベクトルから聴覚臨場感評価値を算出することができる。

40

【0016】

聴覚臨場感評価装置は、音圧レベル算出部をさらに備える構成であってもよい。

【0017】

かかる構成によると、音圧レベル算出部によって、分割された前記二つの音響信号を用いて、前記二つの音響信号の音圧レベルを前記周波数帯域ごとに算出し、動画像生成部によって、算出された前記音圧レベルを、前記輝度、前記色相及び前記彩度のうち、前記相互相関関数が変換されたもの以外のいずれかに変換することによって前記動画像を生成することができる。すなわち、聴覚臨場感評価装置は、より多くのパラメータを用いて聴覚臨場感評価値を算出するので、聴覚臨場感の算出精度が向上する。

50

## 【0018】

前記動画像生成部は、前記相互相関関数を前記輝度に変換するとともに、前記音圧レベルを前記彩度に変換する構成であってもよい。かかる構成によると、相互相関関数と相関する輝度成分及び音圧レベルと相関する彩度成分を有するカラー動画像を生成し、生成されたカラー動画像の移動ベクトルから聴覚臨場感評価値を算出することができる。

## 【0019】

また、前記音圧レベル算出部は、前記音圧レベルの差である音圧レベル差を前記周波数帯域ごとに算出し、前記動画像生成部は、算出された前記音圧レベル差を、前記輝度、前記色相及び前記彩度のうち、前記相互相関関数及び前記音圧が変換されたもの以外に変換することによって前記動画像を生成する構成であってもよい。すなわち、聴覚臨場感評価装置は、より多くのパラメータを用いて聴覚臨場感評価値を算出するので、聴覚臨場感の算出精度が向上する。

10

## 【0020】

さらに、前記動画像生成部は、前記相互相関関数を前記輝度に変換し、前記音圧レベルを前記彩度に変換するとともに、前記音圧レベル差を前記色相に変換する構成であってもよい。かかる構成によると、相互相関関数と相関する輝度成分、音圧レベルと相関する彩度成分、及び音圧レベル差と相関する色相成分を有するカラー動画像を生成し、生成されたカラー動画像の移動ベクトルから聴覚臨場感評価値を算出することができる。

## 【0021】

前記相互相関関数算出部は、区間長の異なる複数の区間長別相互相関関数を算出し、算出された複数の区間長別相互相関関数に基づいて前記相互相関関数を算出する構成であってもよい。単一の区間長の相互相関関数のみを算出した場合には、偶々相関性が高く算出される等相互相関関数の値と実際の音像位置とが正確に一致しない場合があるが、かかる構成によると、複数の区間長を用いることで、音像位置の誤検出を防ぐことができる。

20

## 【0022】

前記音像移動ベクトル算出部は、前記動画像の三以上のフレームを用いて前記動画像の移動ベクトルを算出する構成であってもよい。かかる構成によると、移動ベクトルの算出精度が向上する。

## 【0023】

また、本発明は、コンピュータを前記した聴覚臨場感評価装置として機能させる聴覚臨場感評価プログラムとしても具現化可能である。

30

## 【発明の効果】

## 【0024】

本発明によれば、聴覚臨場感を客観的に評価することができる。

## 【図面の簡単な説明】

## 【0025】

【図1】本発明の実施形態に係る聴覚臨場感評価システムを示すブロック図である。

【図2】一対のマイクロホンによって計測された音響信号を示すグラフである。

【図3】(a)は相互相関関数を示すグラフであり、(b)は、相互相関関数の値と輝度との関係を示す図であり、(c)は、相互相関関数から生成された画素列を示す図である

40

【図4】(a)は音圧レベル差を示すグラフであり、(b)は音圧レベル差による重みを示すグラフであり、(c)は画素列に対する重み付けを示す図である。

【図5】(a)は動画像を示す図であり、(b)は移動ベクトルを示す図である。

【図6】(a)~(d)は、音像移動量の算出手法を説明するための図である。

【図7】(a)(b)は、単一音像の移動ベクトルの算出例を説明するための図である。

【図8】(a)(b)は、複数音像の移動ベクトルの算出例を説明するための図である。

## 【発明を実施するための形態】

## 【0026】

以下、本発明の実施形態について、本発明を再生音場の評価に適用した場合を例にとり

50

、適宜図面を参照しながら説明する。同様の部分には同一符号を付し、重複する説明を省略する。本発明において、「音像」は、聴取者が音源の存在を知覚する位置を指し、ステレオスピーカ等によって具現化された仮想的な音源の位置であったり、実際の音源の位置であったりする。また、本発明において、「聴覚臨場感」は、聴取者が聴覚において感じる臨場感を指す。従来、臨場感は、AV機器の性能を語る際の重要なキーワードであるにも関わらず、曖昧な意味で用いられていたのに対し、本発明の聴覚臨場感推定装置は、聴覚臨場感評価値を定量的かつ客観的に推定することができる。

【0027】

再生側の機器としては、ステレオスピーカ、5.1チャンネルシステム、ヘッドホン等があるが、聴取者は、最終的に両耳に到達した音響信号によって聴覚臨場感を知覚する。したがって、本発明の聴覚臨場感評価装置は、聴取者の両耳と同様に、2チャンネルの音響信号に基づいて音像の移動量を推定することとする。

10

【0028】

図1に示すように、本発明の実施形態に係る聴覚臨場感評価システム1は、再生側の機器として、記憶媒体再生装置10と、スピーカ群20と、を備えている。記憶媒体再生装置10は、図示しない記憶媒体に記憶されたデータを読み取り、ステレオスピーカ、5.1チャンネルシステム、ヘッドホン等からなるスピーカ群20を介して音響信号を再生する。

【0029】

本発明の実施形態に係る聴覚臨場感評価システム1は、評価側の機器として、一对のマイクロホン30L、30Rと、聴覚臨場感評価装置40と、通知装置50と、を備える。

20

【0030】

<マイクロホン>

一对のマイクロホン30L、30Rは、左右に配置された2チャンネル構造を呈し、スピーカ群20で再生された音響信号を計測し、計測された2チャンネルの音響信号を聴覚臨場感評価装置40へ出力する。本明細書では、スピーカ群20で再生された音波だけでなく、当該音波のマイクロホン30L、30Rによる計測結果も音響信号と記載する。

【0031】

<聴覚臨場感評価装置>

聴覚臨場感評価装置40は、CPU(Central Processing Unit)、RAM(Random Access Memory)、ROM(Read-Only Memory)、入出力回路等から構成されており、機能部として、音響信号分割部41と、相互相関数算出部42と、音圧レベル算出部43と、動画像生成部44と、音像移動ベクトル算出部45と、聴覚臨場感評価値算出部46と、を備える。

30

【0032】

音響信号分割部

音響信号分割部41は、一对のマイクロホン30L、30Rから出力された2チャンネルの音響信号をある一定時間ごとにある区間長だけ取得し、取得された2チャンネルの音響信号をM個のバンドパスフィルタによって周波数帯域ごとにM個に分割し、分割された音響信号を相互相関数算出部42及び音圧レベル算出部43へ出力する。

40

【0033】

本実施形態において、音響信号分割部41は、5個のオクターブバンドパスフィルタを備えており、マイクロホン30Lから出力された音響信号Lを、中心周波数が125[Hz]である周波数帯域 $f_1$ の第一の音響信号 $L_1$ 、中心周波数が250[Hz]である周波数帯域 $f_2$ の第二の音響信号 $L_2$ 、中心周波数が500[Hz]である周波数帯域 $f_3$ の第三の音響信号 $L_3$ 、中心周波数が1000[Hz]である周波数帯域 $f_4$ の第四の音響信号 $L_4$ 、中心周波数が2000[Hz]である周波数帯域 $f_5$ の第五の音響信号 $L_5$ に分割する。

【0034】

同様に、音響信号分割部41は、マイクロホン30Rから出力された音響信号を、中心

50

周波数が125 [Hz]である周波数帯域 $f_1$ の第一の音響信号 $R_1$ 、中心周波数が250 [Hz]である周波数帯域 $f_2$ の第二の音響信号 $R_2$ 、中心周波数が500 [Hz]である周波数帯域 $f_3$ の第三の音響信号 $R_3$ 、中心周波数が1000 [Hz]である周波数帯域 $f_4$ の第四の音響信号 $R_4$ 、中心周波数が2000 [Hz]である周波数帯域 $f_5$ の第五の音響信号 $R_5$ に分割する。

【0035】

なお、音響信号の分割数は、5個に限定されず、対象とする音像の種類に応じて適宜変更可能である。例えば、自動車騒音のように広帯域な雑音成分を含む音を対象とする場合には、可聴周波数帯域(20 [Hz] ~ 20 [kHz])を1/1オクターブバンド幅で区切った $M=10$ 程度とすることができる。また、楽器音のように高調波構造を有する音を対象とする場合には、1/10オクターブバンド幅で区切った $M=40$ 程度とすることができる。また、汎用的なモデルのバンドパスフィルタとして、50チャンネル程度の聴覚フィルタ(聴覚末梢系の特性を考慮したバンドパスフィルタであり、代表的なものにROEXフィルタ、Gamma-toneフィルタ等がある)を採用することによって、聴覚特性を反映した分割を行うことができる。

10

【0036】

相互相関関数算出部

相互相関関数算出部42は、音響信号分割部41から出力された各音響信号 $L_1 \sim L_5$ 、 $R_1 \sim R_5$ を取得し、取得された各音響信号 $L_1 \sim L_5$ 、 $R_1 \sim R_5$ に基づいて、周波数帯域ごとの相互相関関数を算出し、算出された相互相関関数を動画像生成部44へ出力する。

20

【0037】

本実施形態において、相互相関関数算出部42は、算出周期ごと(時刻 $t_1, t_2, \dots$ )に、音響信号 $L_1, R_1$ の相互相関(CC: Cross Correlation)関数を算出する。

相互相関関数算出部42は、図2(a)に示すように、時刻 $t_1$ を先頭として区切った区間長(時間窓幅ともいう) $T_1$ の音響信号 $L_1, R_1$ を抽出して抽出された音響信号 $L_1, R_1$ の相関値を算出するとともに、音響信号 $R_1$ のみをラグ $m$ ずつずらして区切って抽出し直して相関値を算出することを繰り返すことによって、図3(a)の下段に示すように、ラグ $m$ を変数とする相互相関関数を算出する。相互相関関数は、-1から+1までの値をとり、音響信号 $L_1, R_1$ が一致した場合に+1となり、音響信号 $L_1, R_1$ の波形が異なるほど値が小さくなる。本発明において、ラグ $m$ は、音響信号 $L_1, R_1$ における相互相関関数を算出するための区間のずれ(時間差)である。

30

【0038】

前記した時刻 $t_1, t_2, \dots$ を先頭として区切った区間長 $T_1$ 、 $p$ 番目の算出周期、 $q$ 番目の周波数帯域における時刻 $n$ の関数である音響信号 $L_{p,q}(n), R_{p,q}(n)$ に対する相互相関関数 $r_{p,q}(m)$ は、下記式によって算出される。

【0039】

【数1】

$$r_{p,q}(m) = \frac{1}{T_1 \cdot \sigma_{p,q}^L \cdot \sigma_{p,q}^R} \sum_{n=0}^{N-1} L_{p,q}(n) \cdot R_{p,q}(m+n)$$

40

【0040】

ここで、 $L_{p,q}, R_{p,q}$ は、それぞれ $L_{p,q}(n), R_{p,q}(n)$ の標準偏差である。また、 $N$ は、相互相関関数を算出する期間の長さ $T$ に対応する離散時間である。

【0041】

なお、区間長は、対象とする音像の種類に応じて適宜変更可能である。例えば、定常的な騒音を対象とする場合には、100 ~ 200 [ms]程度とし、時間的な変化が大きい楽器音を対象とする場合には、10 ~ 20 [ms]程度とすることができる。かかる区間

50

長は、相互相関関数算出部 4 2 に予め記憶されていてもよく、聴覚臨場感評価装置 4 0 に接続されたキーボード、マウス等からなる入力装置（図示せず）をユーザが操作することによって設定される構成であってもよい。

【 0 0 4 2 】

また、算出周期（時刻  $t_1$  ,  $t_2$  , ... の間隔）と、区間長とは、一致しても一致しなくてもよい。

【 0 0 4 3 】

また、相互相関関数算出部 4 2 は、聴覚臨場感評価値算出部 4 6 によって算出された音像の移動量（移動ベクトルの大きさ）に基づいて、算出周期を変更する構成であってもよい。例えば、音像の移動量が小さい場合には、算出周期を長く設定することによって演算量を削減することができる。この場合において、二次元動画像の一般的なフレームレートである  $1/30$  [s] を算出周期として相互相関関数を算出する場合であっても、区間長を  $1/30$  [s] よりも長く設定し、音響信号  $R_1$  の抽出範囲をオーバーラップさせながら相互相関関数を算出することができる。相互相関関数算出部 4 2 には、音像の移動量と、相互相関関数の算出周期と、の関係がデータベースとして予め記憶されており、相互相関関数算出部 4 2 は、聴覚臨場感評価値算出部 4 6 によって算出された音像の移動量を用いてデータベースを検索することによって、音像の移動量の対応する算出周期を読み出し、読み出された算出周期に基づいて相互相関関数を算出する。

【 0 0 4 4 】

また、相互相関関数算出部 4 2 は、区間長の異なる複数の区間長別相互相関関数を算出し、算出された複数の区間長別相互相関関数に基づいて相互相関関数を算出する構成であってもよい。例えば、相互相関関数算出部 4 2 には、区間長  $T_1$  ,  $T_2$  ,  $T_3$  が予め記憶されており、相互相関関数算出部 4 2 は、図 2 ( b ) に示すように、区間長  $T_1$  における区間長別相互相関関数を算出するとともに、区間長  $T_1$  よりも長い区間長  $T_2$  ,  $T_3$  における区間長別相互相関関数を算出し、算出された 3 個の区間長別相互相関関数の相加平均を算出することによって、相互相関関数を算出することができる。

【 0 0 4 5 】

相互相関関数算出部 4 2 は、他の周波数帯域  $f_2 \sim f_5$  に関しても同様に相互相関関数を算出する。

【 0 0 4 6 】

音圧レベル算出部

音圧レベル算出部 4 3 は、音響信号分割部 4 1 から出力された各音響信号  $L_1 \sim L_5$  ,  $R_1 \sim R_5$  を取得し、取得された各音響信号  $L_1 \sim L_5$  ,  $R_1 \sim R_5$  に基づいて、周波数帯域ごとの音圧レベル、及び、左右の音響信号の音圧の差である音圧レベル差を算出し、算出された音圧レベル及び音圧レベル差を動画像生成部 4 4 へ出力するとともに、算出された音圧レベルを聴覚臨場感評価値算出部 4 6 へ出力する。

【 0 0 4 7 】

動画像生成部

動画像生成部 4 4 は、相互相関関数算出部 4 2 から出力された周波数帯域ごとの相互相関関数と、音圧レベル算出部 4 3 から出力された音圧レベル及び音圧レベル差に基づいて、動画像（より詳細には、二次元動画像）を生成し、生成された動画像を音像移動ベクトル算出部 4 5 へ出力する。

【 0 0 4 8 】

図 5 ( a ) に示すように、動画像は、時刻  $t_1$  におけるフレーム 1 0 0、時刻  $t_2$  におけるフレーム 2 0 0、時刻  $t_3$  におけるフレーム 3 0 0、時刻  $t_4$  におけるフレーム 4 0 0、... から構成されており、動画像の一のフレームは、縦軸を周波数（周波数帯域  $f_1 \sim f_5$ ）、横軸をラグ  $m$  とする平面である。本実施形態では、一のフレームは、縦  $5 \times$  横  $8$  の画素から構成されており、動画像生成部 4 4 は、相互相関関数を対応する画素の輝度、音圧レベルを対応する画素の彩度、音圧レベル差を対応する画素の色相に変換することによって、動画像を生成する。

10

20

30

40

50



## 【 0 0 4 9 】

例えば、相互相関関数のみを用いて動画像を生成する場合には、動画像生成部 4 4 は、相互相関関数が - 1 の場合に最低輝度となり、相互相関関数が + 1 の場合に最高輝度となるように各画素の画素データを生成し（図 3（b）参照）、かかる画素データによって構成された一連の画像からなる動画像を生成する。ここで、動画像生成部 4 4 は、図 3（a）の下段に示す周波数帯域  $f_1$  の相互相関関数から、図 3（c）の下段に示す輝度を有する画素列 1 1 0 を生成し、図 3（a）の上段に示す周波数帯域  $f_2$  の相互相関関数から、図 3（c）の上段に示す輝度を有する画素列 1 2 0 を生成する。

## 【 0 0 5 0 】

また、動画像生成部 4 4 は、図 4（a）に示す周波数帯域ごとの音響信号  $L_1$ （又は  $L_2 \sim L_5$ ）、 $R_1$ （又は  $R_2 \sim R_5$ ）の音圧レベル 2 1 0、2 2 0 の音圧レベル差 2 3 0 に基づいて、図 4（b）に示すラグ  $m$  ごとの重みを算出し、算出された重みによって重み付けされた輝度を有する画素列 1 1 0' を生成する構成であってもよい。

10

## 【 0 0 5 1 】

相関の高い音響信号が左右の耳にある音圧レベル差で到来した場合、左右のレベル差によって、正面を 0 度として  $\pm 90$  度のいずれかの方向に音像が知覚される。動画像生成部 4 4 は、かかる音圧レベル差と音像の知覚方向との関係を予め記憶しておき、音圧レベル差 2 3 0 を用いてかかる関係を参照することによって、何度の方向に音像があるかを推測することができる。動画像生成部 4 4 は、音像があると推定される方向に重みとして 1 を設定し、そこから遠ざかる方向につれて、重みとして小さい値を設定する。

20

## 【 0 0 5 2 】

一方、相互相関関数の値が最大になるラグ  $m$  が音像のある方向であるから、ラグ  $m$  と音像の方向とは一対一に対応する。したがって、動画像生成部 4 4 は、重み付けによって、音圧レベル差 2 3 0 に基づいて音像があると推定された付近の相互相関関数の値はそのままにし、音圧レベル差 2 3 0 に基づいて音像がないと推定された付近の相互相関関数の値は小さくなるようにする。こうすることで、正弦波のように相互相関関数が 1 となるラグが数か所でてくるような場合でも、動画像生成部 4 4 は、音圧レベル差 2 3 0 に基づいて音像の方向を判定することができる。

## 【 0 0 5 3 】

また、動画像生成部 4 4 は、相互相関関数を輝度又は色相に変換して動画像を生成する場合には、相互相関関数がマイナスの部分をも 0 とみなして動画像を生成する構成であってもよい。これは、音像が 1 個の場合には、相互相関関数が最大となるラグに対応する位置が音像の位置と一致するためである。

30

## 【 0 0 5 4 】

## 音像移動ベクトル算出部

音像移動ベクトル算出部 4 5 は、動画像生成部 4 4 から出力された動画像の複数のフレームを用いて動画像の移動ベクトルを算出することによって、音像の移動ベクトルを算出し、算出された音像の移動ベクトルを聴覚臨場感評価値算出部 4 6 へ出力する。動画像の移動ベクトルの算出手法としては、オプティカルフロー、パターンマッチング、特徴量追従アルゴリズム等といった、画像情報処理のための公知のアルゴリズムを適宜採用することができる。

40

## 【 0 0 5 5 】

音像移動ベクトル算出部 4 5 は、3 個以上（本実施形態では、時刻  $t_1 \sim t_4$  に対応する 4 個）のフレームを蓄積し、蓄積されたフレームの輝度、色相及び彩度に基づいて、周波数帯域ごとの移動ベクトルを算出する。詳細には、音像移動ベクトル算出部 4 5 は、図 5（b）に示すように、時刻  $t_1$  における画素列 1 1 1（ここでは、説明のため、1 列の画素数を 1 2 個とする。）、時刻  $t_2$  における画素列 2 1 1、時刻  $t_3$  における画素列 3 1 1 を用いて、周波数帯域  $f_1$  における移動ベクトルを算出するとともに、他の周波数帯域  $f_2 \sim f_5$  についても同様の処理を行う。ここで、時刻  $t_1$ 、 $t_2$  の 2 個のフレームを用いて移動ベクトルを算出する場合には、音像移動ベクトル算出部 4 5 は、時刻  $t_1$  にお

50

ける画素列 1 1 1 の左から 4 個目の画素が、時刻  $t_2$  においても画素列 2 1 1 の左から 4 個目に留まっていると誤算出するおそれがある（実線の矢印）。そのため、音像移動ベクトル算出部 4 5 は、時刻  $t_3$  における画素列 3 1 1 をさらに用いることによって、時刻  $t_1$  における画素列 1 1 1 の左から 1 ~ 3 個目の画素が、時刻  $t_2$  における画素列 2 1 1 では左から 5 ~ 7 個目、時刻  $t_3$  における画素列では左から 1 0 ~ 1 2 個目へと移動していると推定するため、時刻  $t_1$  における画素列 1 1 1 の左から 4 個目の画素が、時刻  $t_2$  においては画素列 2 1 1 の左から 8 個目へ移動していると算出することができる（点線の矢印）。このように、3 個以上のフレームを用いて移動ベクトルを算出することによって、移動ベクトルの誤算出を防ぐことができる。

【 0 0 5 6 】

10

また、音像移動ベクトル算出部 4 5 は、周波数特性が変動する音像に対して、周波数帯域ごとの音圧レベルの変化に応じて、周波数帯域を跨ぐように音像の移動ベクトルを算出してもよい。ここで、音像移動ベクトル算出部 4 5 は、音圧レベルの周波数成分の時間変化に基づいて、周波数、時間及び音圧レベルを軸とする三次元空間における音圧レベルのピーク（尾根部分）が周波数の軸上でシフトした場合に、音像の周波数特性が変動していると判定し、周波数帯域を跨ぐように音像の移動ベクトルを算出することができる。

【 0 0 5 7 】

聴覚臨場感評価値算出部

聴覚臨場感評価値算出部 4 6 は、音像移動ベクトル算出部 4 5 から出力された音像の移動ベクトルに基づいて、聴覚臨場感評価値を算出し、通知装置 5 0 へ出力する。本実施形態において、聴覚臨場感評価値算出部 4 6 は、閾値未満の音像の移動量を切り捨てるとともに、閾値以上の複数の音像の移動量を相加平均することによって、聴覚臨場感評価値を算出することができる。かかる聴覚臨場感評価値は、音像の移動ベクトルが大きいほど、当該聴覚臨場感評価値が大きくなるような値である。

20

【 0 0 5 8 】

聴覚臨場感評価値算出部 4 6 は、動画像の周波数帯域ごとの移動ベクトル、すなわち、音像の移動ベクトルに基づいて、周波数帯域ごとの音像移動量を推定し、推定された周波数帯域ごとの音像移動量の最大値を採用する、音像移動量の相加平均を算出する、又は、音像移動量の積算値を算出することによって、全体としての音像移動量を推定する。

【 0 0 5 9 】

30

本実施形態では、聴覚臨場感評価値算出部 4 6 は、音像移動量の積算値を採用する。すなわち、聴覚臨場感評価値算出部 4 6 は、図 6 ( a ) に示すように、前フレームと今フレームにおける各周波数帯域  $f_1 \sim f_5$  の移動ベクトル  $V_{1a}$  ,  $V_{1b}$  ,  $V_2 \sim V_5$  を算出する。

【 0 0 6 0 】

続いて、聴覚臨場感評価値算出部 4 6 は、図 6 ( b ) に示すように、各周波数帯域  $f_1 \sim f_5$  の移動ベクトルを統合する。ここでは、周波数帯域  $f_2$  の移動ベクトル  $V_2$  と周波数帯域  $f_5$  の移動ベクトル  $V_5$  が類似しているため統合されるとともに、周波数帯域  $f_1$  の左の移動ベクトル  $V_{1a}$  と周波数帯域  $f_1$  の左の移動ベクトル  $V_{1b}$  と周波数帯域  $f_4$  の移動ベクトル  $V_4$  とが類似しているため統合される。

40

【 0 0 6 1 】

ここで、聴覚臨場感評価値算出部 4 6 は、二つの移動ベクトルの向きが同じで、大きさがほぼ同じ（例えば、移動ベクトルの大きさの比率が一定範囲内）であれば、二つの移動ベクトルが類似していると判定する構成であってもよく、さらに、二つの移動ベクトルの音圧レベル及び音圧レベル差の少なくとも一方がほぼ同じ（比率が一定範囲内）である場合に、二つの移動ベクトルが類似していると判定する構成であってもよい。

【 0 0 6 2 】

続いて、聴覚臨場感評価値算出部 4 6 は、図 6 ( c ) に示すように、統合後の移動ベクトル  $V_{1b}$  ( =  $V_{1a} = V_4$  ) ,  $V_3$  ,  $V_5$  の大きさの和（積分値）を算出し、算出結果を音像移動量とする。

50

## 【 0 0 6 3 】

なお、聴覚臨場感評価値算出部 4 6 は、図 6 ( d ) に示すように、統合後の移動ベクトル  $V_{1b}$  ,  $V_3$  ,  $V_5$  を音圧によって補正し、補正後の移動ベクトル  $V_{1b}'$  ,  $V_3'$  ,  $V_5'$  の大きさの和 ( 積分値 ) を算出し、算出結果を音像移動量とする構成であってもよい。例えば、音圧レベルが大きいほど聴覚臨場感評価値に対する音像の移動ベクトルの寄与度が高くなるように重み付けを行うことによって、聴覚臨場感評価値の推定精度を向上させることができる。

## 【 0 0 6 4 】

続いて、聴覚臨場感評価値算出部 4 6 は、今フレームと次フレームにおける音像移動量を算出し、かかる算出を所定フレーム数に達するまで繰り返し、所定フレーム数分の音像移動量の和 ( 積算値 ) を算出し、算出結果を最終的な音像移動量、すなわち、聴覚臨場感評価値とする。

10

## 【 0 0 6 5 】

また、聴覚臨場感評価値算出部 4 6 は、フレーム数の異なる複数のフレーム数別聴覚臨場感評価値を算出し、算出された複数のフレーム数別聴覚臨場感評価値に基づいて聴覚臨場感評価値を算出する構成であってもよい。

例えば、聴覚臨場感評価値算出部 4 6 は、3 個のフレームにおけるフレーム数別聴覚臨場感評価値と、4 個のフレームにおけるフレーム数別聴覚臨場感評価値と、を算出し、算出された 2 個のフレーム数別聴覚臨場感評価値の相加平均を算出することによって、聴覚臨場感評価値を算出することができる。

20

## 【 0 0 6 6 】

また、音像が 1 個の場合には、聴覚臨場感評価値算出部 4 6 は、輝度 ( すなわち、相互相関関数 ) が最大となるラグ  $m$  に基づいて、ある時刻における音像の方向 ( 方位角 ) を推定することもできる。これは、音像が 1 個の場合には、相互相関関数が最大となるラグ  $m$  が、音像から各マイクロホン 3 0 L , 3 0 R までの距離の差と関連するためである。

## 【 0 0 6 7 】

また、聴覚臨場感評価値算出部 4 6 は、複数の周波数帯域の音圧レベルが同期して変動している場合には、当該複数の周波数帯域に同じ音像が存在するとみなして音像の移動ベクトルを統合することもできる。かかる構成によると、聴取者が知覚するのに似た音像の移動ベクトルを用いて聴覚臨場感評価値を算出することができる。

30

## 【 0 0 6 8 】

< 通知装置 >

通知装置 5 0 は、ディスプレイ、スピーカ等からなり、聴覚臨場感評価値算出部 4 6 から出力された聴覚臨場感評価値を表示したり音声出力したりすることによって、利用者へ通知する。

## 【 0 0 6 9 】

< 動作例 >

続いて、本発明の実施形態に係る聴覚臨場感評価装置 4 0 の動作例について説明する。まず、音響信号分割部 4 1 が、一对のマイクロホン 3 0 L , 3 0 R から出力された 2 チャネルの音響信号を取得し、取得された 2 チャネルの音響信号を 5 個のバンドパスフィルタによって周波数帯域が異なる 5 個の音響信号  $L_1 \sim L_5$  ,  $R_1 \sim R_5$  に分割する。

40

## 【 0 0 7 0 】

続いて、相互相関関数算出部 4 2 が、各音響信号  $L_1 \sim L_5$  ,  $R_1 \sim R_5$  に基づいて、周波数帯域ごとの相互相関関数を算出する。

## 【 0 0 7 1 】

続いて、音圧レベル算出部 4 3 が、各音響信号  $L_1 \sim L_5$  ,  $R_1 \sim R_5$  に基づいて、周波数帯域ごとの音圧レベル、及び、左右の音響信号の音圧の差である音圧レベル差を算出し、算出された音圧レベル及び音圧レベル差を動画像生成部 4 4 へ出力するとともに、算出された音圧レベルを聴覚臨場感評価値算出部 4 6 へ出力する。

## 【 0 0 7 2 】

50

続いて、動画像生成部 44 が、相互相関関数と、音圧レベル算出部 43 から出力された音圧レベル及び音圧レベル差に基づいて、相互相関関数を輝度、音圧レベルを彩度、音圧レベル差を色相とみなした各画素の画素データを生成し、かかる画素データによって構成された一連の画像からなる動画像を生成する。

【0073】

続いて、音像移動ベクトル算出部 45 が、動画像生成部 44 から出力された動画像の複数のフレームを用いて動画像の移動ベクトルを算出することによって、音像の移動ベクトルを算出する。

【0074】

続いて、聴覚臨場感評価値算出部 46 が、音像移動ベクトル算出部 45 から出力された音像の移動ベクトルに基づいて、聴覚臨場感評価値を算出し、通知装置 50 へ出力する。

【0075】

< 音像の移動ベクトルの算出例 >

続いて、本発明の実施形態に係る聴覚臨場感評価装置 40 の音像移動ベクトル算出部 45 が、移動する音像の移動ベクトルを実際に算出した例について説明する。

【0076】

ここで、マイクロホン 30L, 30R として、ダミーヘッド（高研社製の SAMRAI）を使用するとともに、スピーカ群 20 として、単一のスピーカを使用し、このスピーカを、ダミーヘッドを中心とした半径 0.5 [m] の円上で、左（-90 [deg]）から右（90 [deg]）まで、又は右から左まで移動させたときの音響信号をダミーヘッドが検出した。また、音響信号分割部 41 のサンプリング周波数は 44100 [Hz] であり、中心周波数 125, 250, 500, 1000, 2000, 4000 [Hz] とするオクターブバンドパスフィルタを用いて音響信号分割部 41 を構成することにより、音響信号を 6 個の周波数帯域に分割した。

【0077】

相互相関関数算出部 42 は、相互相関関数の算出周期を 1/30 [s]、区間長を 1/5 [s] として相互相関関数を算出した。ここで、ダミーヘッドにおけるマイクロホン 30L, 30R 間の距離は 0.18 [m] であり、音速は 340 [m/s] であるため、音響信号がマイクロホン 30L, 30R に到達する時間差が最大となる（スピーカが左又は右にある）場合の時間差は、約 0.7 [ms] となる。そこで、ラグ m としては、0 を中心とした  $\pm 0.7$  [ms] に相当する部分にのみ着目した。

【0078】

また、音像移動ベクトル算出部 45 は、相互相関関数の最大値  $r_{p-1, q}(m)_{max}$ ,  $r_{p, q}(m)_{max}$  を有する画素同士を結ぶ手法を用いて動画像の移動ベクトル、すなわち、音像の移動ベクトルを算出した。

【0079】

単一音像の場合

単一のスピーカから音響信号としてホワイトノイズを出力した場合、すなわち、単一の音源を用いて一の音像を生成した場合における、計測開始時点において動画像生成部 44 が生成した画像（フレーム）を図 7（a）に示し、現在（途中経過時点）において動画生成部 44 が生成した画像を図 7（b）に示す。図 7（b）における矢印は、開始時点から現在までの、周波数帯域ごとの音像の移動ベクトルである。図 7（a）（b）を参照すると、全ての周波数帯域において音像が左から右へ移動している様子が分かる。

【0080】

複数音像の場合

左から右へ移動するスピーカからローパスノイズ（遮断周波数 700 [Hz]）を出力した音響信号と、右から左へ移動するスピーカからハイパスノイズ（遮断周波数 1000 [Hz]）を出力した音響信号と、を合成したものを音響信号分割部 41 に入力して分析を行った。この場合では、二つの音源を用いて二つの音像を生成している。この場合における、測開始時点において動画像生成部 44 が生成した画像（フレーム）を図 8（a）に

10

20

30

40

50

示し、現在（途中経過時点）において動画生成部 44 が生成した画像を図 8（b）に示す。図 8（b）における矢印は、開始時点から現在までの、周波数帯域ごとの音像の移動ベクトルである。図 8（a）（b）を参照すると、低周波数帯域においては音像が左から右へ移動し、高周波数帯域においては音像が右から左へ移動している様子が分かる。

【0081】

本発明の実施形態に係る聴覚臨場感評価装置 40 は、複数の音像に対して処理を行う場合であっても、音像の方向を用いずに音像の移動量に基づいて聴覚臨場感評価値を算出するので、簡易な処理で聴覚臨場感を客観的に評価することができる。

【0082】

以上、本発明の実施形態について実施形態を参照して説明したが、本発明は前記実施形態に限定されず、本発明の要旨を逸脱しない範囲で適宜設計変更可能である。例えば、音圧レベル算出部 43 を備えずに、動画像生成部 44 が相互相関関数を輝度と変換したモノクロの動画像を生成する構成であってもよい。また、動画像生成部 44 が相互相関関数を輝度、音圧レベルを彩度としたカラーの動画像を生成する構成であってもよい。この場合には、動画像生成部 44 は、各画素の色相を所定値（例えば、色相値 50%）に設定することができる。また、相互相関関数、音圧レベル及び音圧レベル差を、輝度、彩度及び色相のいずれかに変換して動画像を生成する構成であってもよい。

【0083】

また、聴覚臨場感評価装置 40 は、記憶媒体に記憶されていない音響信号を再生する装置、再生装置ではない実際の音源に対しても、聴覚臨場感評価値を算出することができる。また、音像移動ベクトル算出部 45 は、複数の周波数帯域の画素列を用いて一つの移動ベクトルを算出することによって、統合された移動ベクトルを算出する構成であってもよく、移動ベクトルごとの画像の特徴量をパターンとして予め記憶しており、パターンマッチングを用いて移動ベクトルを算出する構成であってもよい。また、本発明は、コンピュータを聴覚臨場感評価装置 40 として機能させる聴覚臨場感評価プログラムとしても具現化可能である。

【符号の説明】

【0084】

- 1 聴覚臨場感評価システム
- 30L, 30R マイクロホン
- 40 聴覚臨場感評価装置
- 41 音響信号分割部
- 42 相互相関関数算出部
- 43 音圧レベル算出部
- 44 動画像生成部
- 45 音像移動ベクトル算出部
- 46 聴覚臨場感評価値算出部

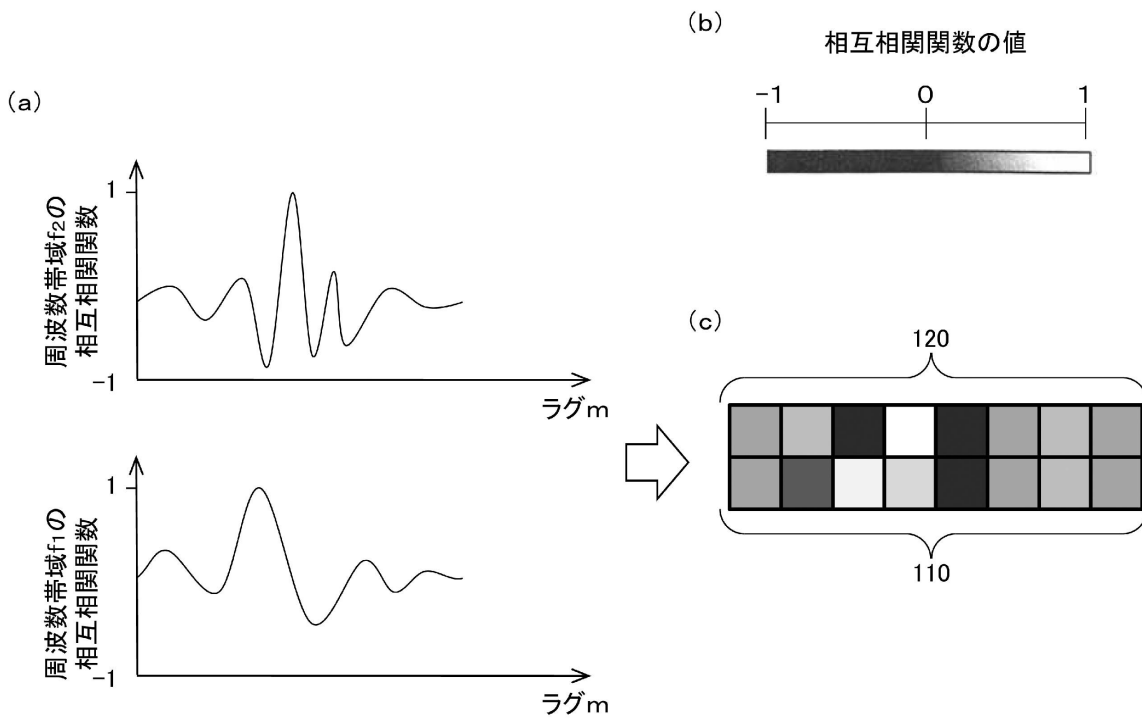
10

20

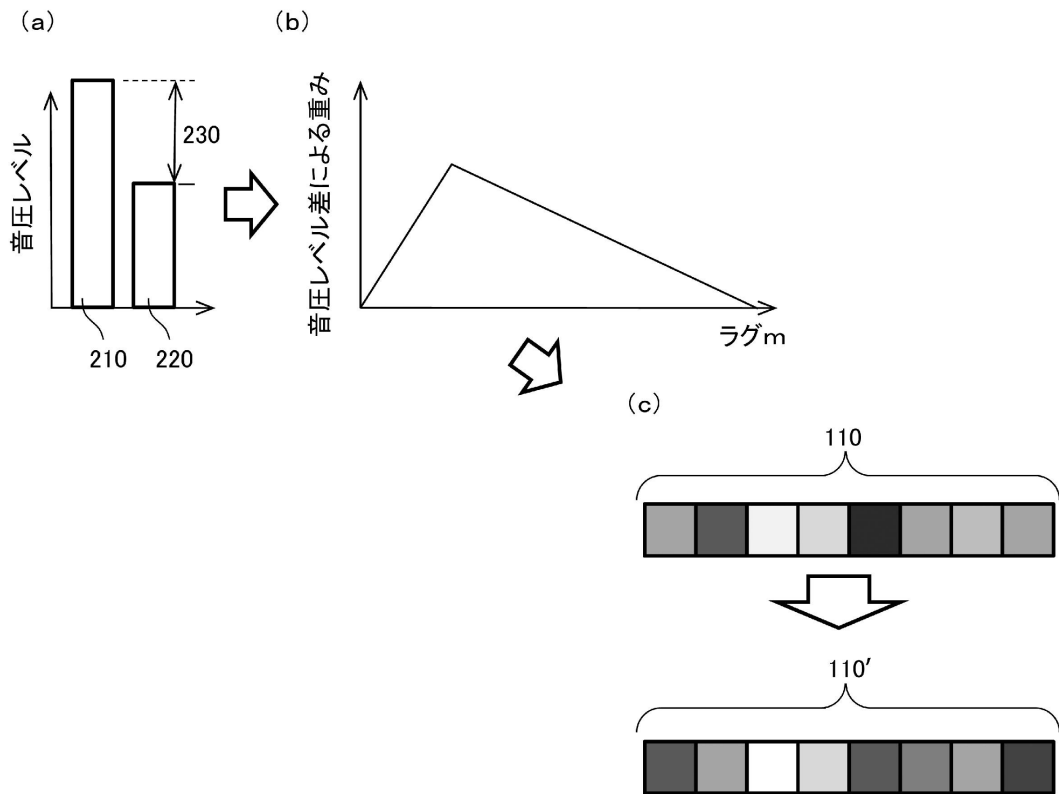
30



【 図 3 】

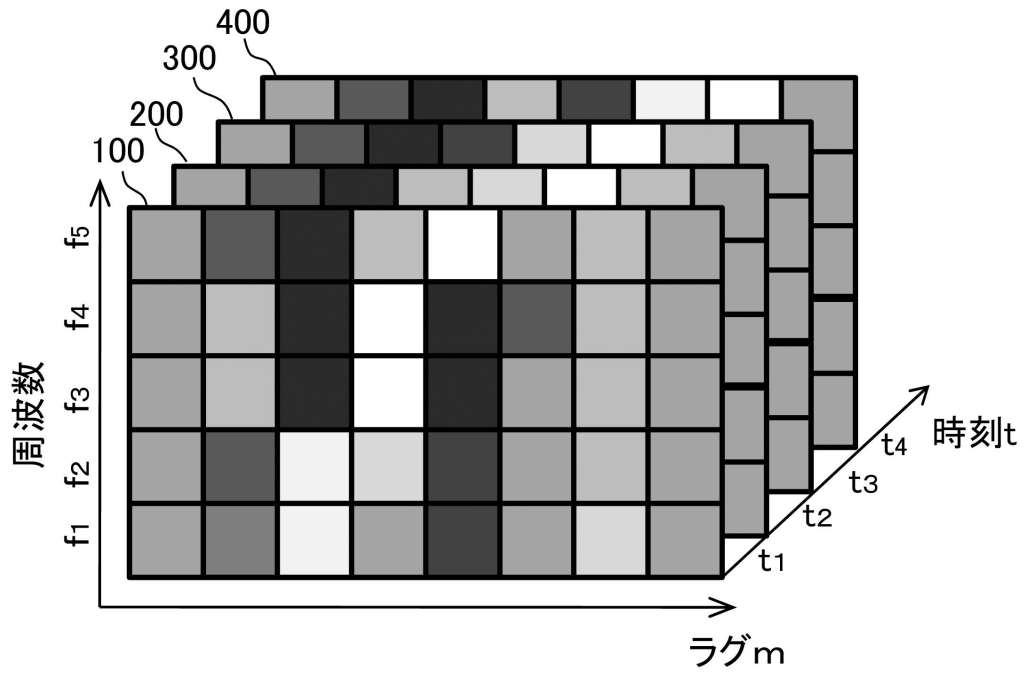


【 図 4 】

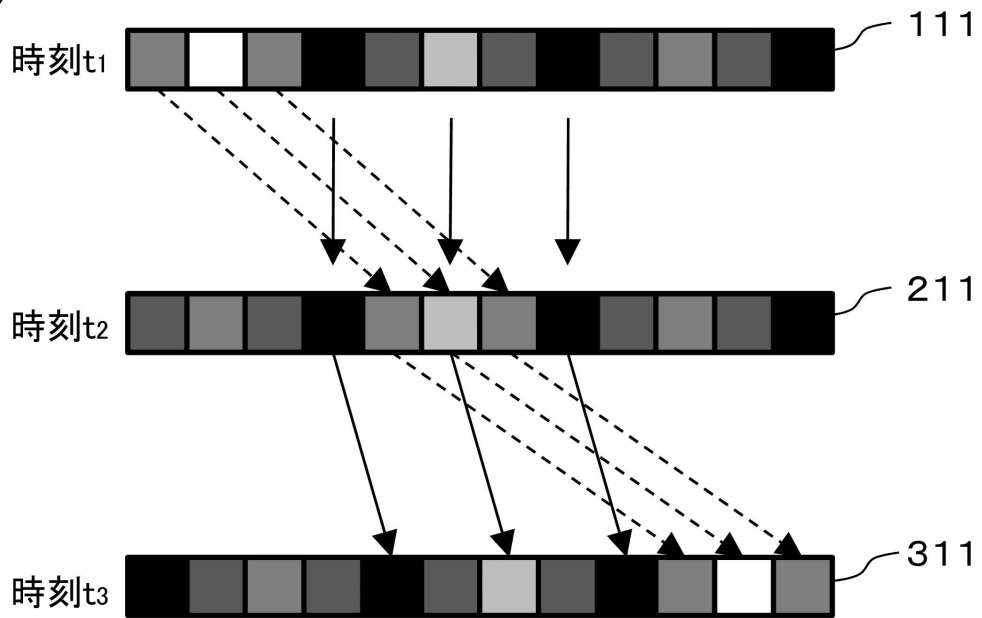


【図5】

(a)

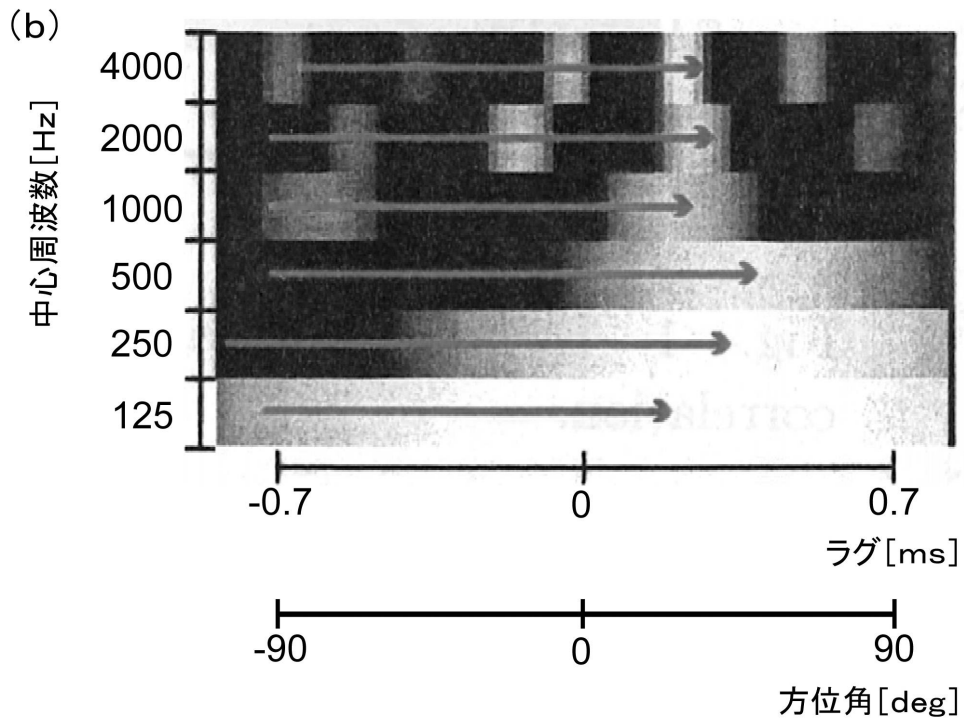
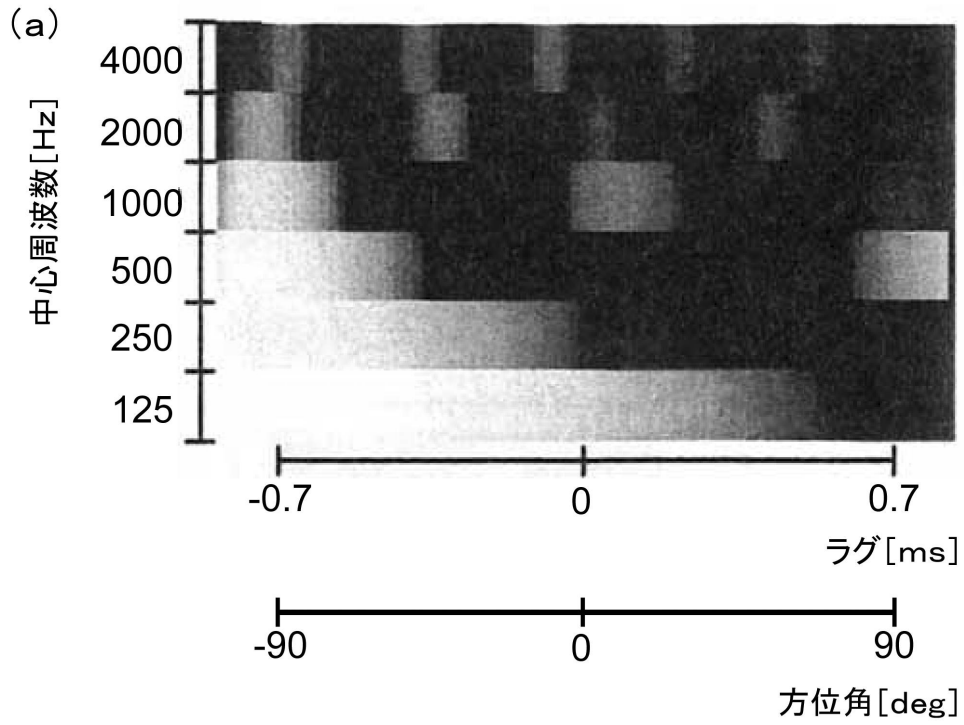


(b)

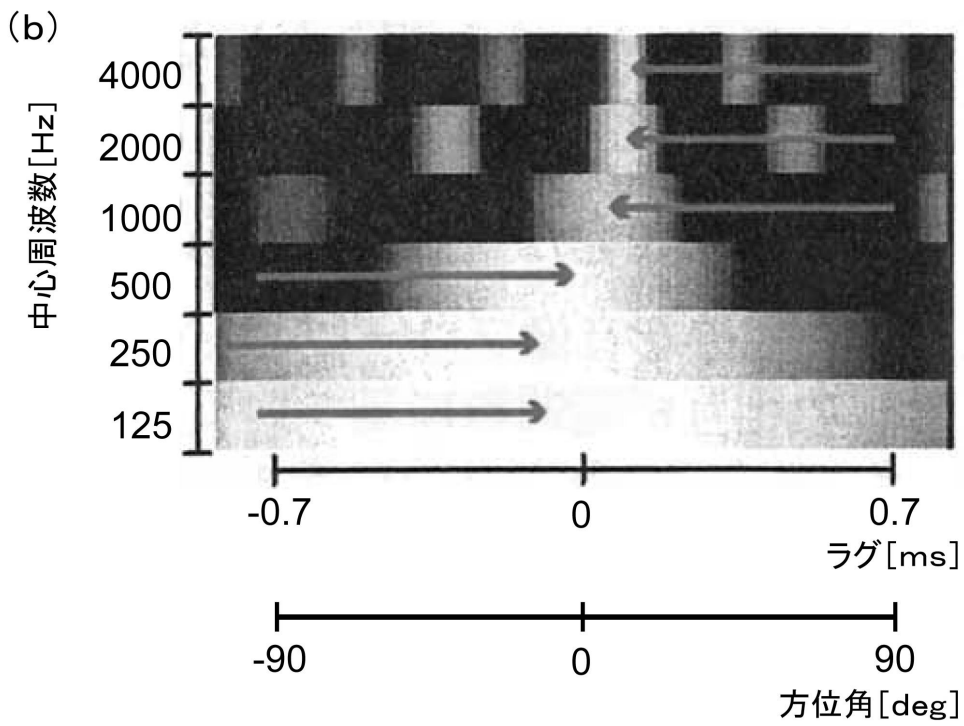
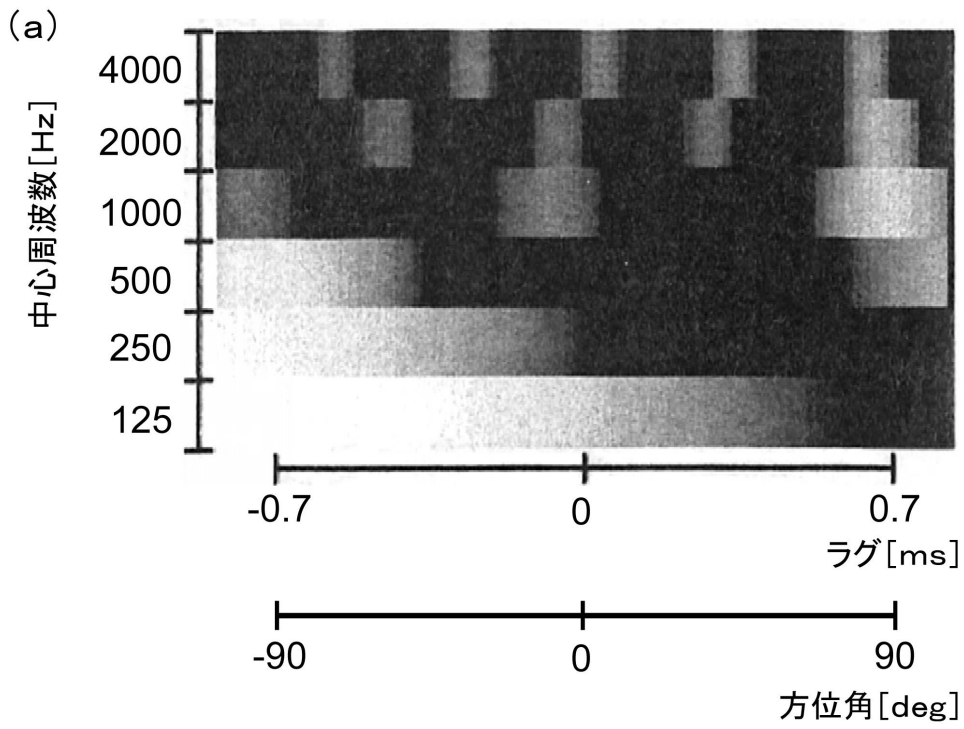




【図7】



【 図 8 】



---

フロントページの続き

(出願人による申告)平成21年度、独立行政法人 情報通信研究機構、委託研究「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発 課題工：感性情報認知・伝達技術」、産業技術力強化法第19条の適用を受ける特許出願

特許権者において、実施許諾の用意がある。

- (72)発明者 小澤 賢司  
山梨県甲府市武田四丁目4番37号 山梨大学内
- (72)発明者 木下 雄一朗  
山梨県甲府市武田四丁目4番37号 山梨大学内
- (72)発明者 齋藤 康大  
山梨県甲府市武田四丁目4番37号 山梨大学内

審査官 菊池 充

- (56)参考文献 特開2004-279525(JP,A)  
特開2005-338528(JP,A)  
特開2005-039311(JP,A)  
欧州特許出願公開第01458218(EP,A1)  
欧州特許出願公開第01499161(EP,A1)

(58)調査した分野(Int.Cl., DB名)

H04S 1/00 - 7/00  
G10L 19/00 - 21/18