

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4449987号
(P4449987)

(45) 発行日 平成22年4月14日(2010.4.14)

(24) 登録日 平成22年2月5日(2010.2.5)

(51) Int.Cl.

F I

G 1 O L 21/02 (2006.01)

G 1 O L 21/02 2 O 1 D

G 1 O L 11/00 (2006.01)

G 1 O L 11/00 4 O 2 Z

G 1 O L 11/02 (2006.01)

G 1 O L 11/02

請求項の数 9 (全 23 頁)

(21) 出願番号 特願2007-35410 (P2007-35410)
 (22) 出願日 平成19年2月15日(2007.2.15)
 (65) 公開番号 特開2008-197577 (P2008-197577A)
 (43) 公開日 平成20年8月28日(2008.8.28)
 審査請求日 平成20年9月22日(2008.9.22)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100095957
 弁理士 亀谷 美明
 (74) 代理人 100096389
 弁理士 金本 哲男
 (74) 代理人 100101557
 弁理士 萩原 康司
 (72) 発明者 難波 隆一
 東京都港区港南1丁目7番1号 ソニー株
 式会社内
 (72) 発明者 安部 素嗣
 東京都港区港南1丁目7番1号 ソニー株
 式会社内

最終頁に続く

(54) 【発明の名称】 音声処理装置、音声処理方法およびプログラム

(57) 【特許請求の範囲】

【請求項 1】

入力音声に特定音源から発せられた第一の音声が含まれているか否かを判定する音声判定部と；

前記音声判定部により前記入力音声に前記第一の音声が含まれていると判定された場合、前記入力音声を前記第一の音声と前記特定音源以外の音源から発せられた第二の音声とに分離する音声分離部と；

前記音声分離部により分離された前記第一の音声と前記第二の音声を、任意の音量比率で混合する音声混合部と；

を備え、

前記音声判定部は、前記入力音声の音源の位置が收音位置から設定距離の範囲内であり、前記入力音声にインパルス音が含まれ、前記入力音声が過去の平均音量と比較して大きい場合、前記入力音声に特定音源から発せられた前記第一の音声が含まれていると判定することを特徴とする、音声処理装置。

【請求項 2】

前記第一の音声は、前記入力音声を收音する際に用いられた装置の操作者に起因する音声を含み、

前記第二の音声は、收音対象から発せられた音声を含むことを特徴とする、請求項 1 に記載の音声処理装置。

【請求項 3】

映像を撮像する撮像部をさらに備え、

前記音声判定部は、請求項 1 に記載の場合に加えて、入力音声に含まれる 1 または 2 以上の音源から発せられた音声の音量または位相の少なくともいずれかに基づいて前記音源の位置情報を算出する位置情報算出部を備え、前記入力音声の音源の位置が前記撮像部の撮像方向の後方であると前記位置情報算出部により算出され、前記入力音声が入間の音声と一致または近似する音質である場合にも、前記入力音声に特定音源から発せられた前記第一の音声が含まれていると判定することを特徴とする、請求項 1 に記載の音声処理装置。

【請求項 4】

前記入力音声を收音する複数の收音部と；

10

前記音声混合部により混合された混合音声を記憶媒体に記録する記録部と；

を備えることを特徴とする、請求項 1 に記載の音声処理装置。

【請求項 5】

前記入力音声を記憶している記憶媒体と；

前記記憶媒体に記憶されている前記入力音声を再生し、前記音声判定部および前記音声分離部へ出力する再生部と；

を備えることを特徴とする、請求項 1 に記載の音声処理装置。

【請求項 6】

前記入力音声の音量が補正されている場合、前記音声分離部により分離された前記第二の音声の音量に、前記補正の程度に応じた逆補正を行なう音量補正部、を備えることを特徴とする、請求項 1 に記載の音声処理装置。

20

【請求項 7】

コンピュータを、

入力音声に特定音源から発せられた第一の音声が含まれているか否かを前記音源の位置情報に基づいて判定する音声判定部と；

前記音声判定部により前記入力音声に前記第一の音声が含まれていると判定された場合、前記入力音声を前記第一の音声と前記特定音源以外の音源から発せられた第二の音声とに分離する音声分離部と；

前記音声分離部により分離された前記第一の音声と前記第二の音声とを、任意の音量比率で混合する音声混合部と；

30

を備え、

前記音声判定部は、前記入力音声の音源の位置が收音位置から設定距離の範囲内であり、前記入力音声にインパルス音が含まれ、前記入力音声は過去の平均音量と比較して大きい場合、前記入力音声に特定音源から発せられた前記第一の音声が含まれていると判定することを特徴とする音声処理装置として機能させるための、プログラム。

【請求項 8】

コンピュータを、

映像を撮像する撮像部をさらに備え、

前記音声判定部は、入力音声に含まれる 1 または 2 以上の音源から発せられた音声の音量または位相の少なくともいずれかに基づいて前記音源の位置情報を算出する位置情報算出部を備え、請求項 7 に記載の場合に加えて、前記入力音声の音源の位置が前記撮像部の撮像方向の後方であると前記位置情報算出部により算出され、前記入力音声が入間の音声と一致または近似する音質である場合にも、前記入力音声に特定音源から発せられた前記第一の音声が含まれていると判定することを特徴とする音声処理装置として機能させるための、請求項 7 に記載のプログラム。

40

【請求項 9】

入力音声に特定音源から発せられた第一の音声が含まれているか否かを前記音源の位置情報に基づいて判定する判定ステップと；

前記入力音声に前記第一の音声が含まれていると判定された場合、前記入力音声を前記第一の音声と前記特定音源以外の音源から発せられた第二の音声とに分離する分離ステッ

50

プと；

前記分離された前記第一の音声と前記第二の音声を、任意の音量比率で混合する混合ステップと；

を含み、

前記判定ステップにおいて、前記入力音声の音源の位置が收音位置から設定距離の範囲内であり、前記入力音声にインパルス音が含まれ、前記入力音声が過去の平均音量と比較して大きい場合、前記入力音声に特定音源から発せられた前記第一の音声が含まれていると判定することを特徴とする、音声処理方法。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、音声処理装置、音声処理方法およびプログラムに関する。

【背景技術】

【0002】

近日、被写体の映像および被写体から発せられた音声を記録可能な映像音声記録装置が広く普及している。映像音声記録装置の操作者は、映像音声記録装置の撮像方向を調整したり、映像音声記録装置に設けられた操作手段を操作して被写体の映像を拡大または縮小することができる。

【0003】

20

ここで、音声の音量は音源から離れるにつれて減少する。したがって、上記のような映像音声記録装置には、映像音声記録装置の操作者の声や操作手段の操作音などの操作者に起因する音声、被写体が発する音声より大きな音量で記録される場合があった。

【0004】

特許文献1には、このような操作者に起因する音声の音量が抑制された音声を記録するための音声処理装置が開示されている。具体的には、当該音声処理装置は、前左用、前右用、後左用、後右用、および着脱可能なマイクロホンの計5本の指向性マイクロホンを備える。したがって、後ろ中央に位置する操作者の声は前左用、前右用、後左用および後右用のいずれのマイクロホンにもほとんど收音されず、必要や目的に応じて着脱可能なマイクロホンに收音させることができる。

30

【0005】

また、特許文献2には、複数の音源からの音声が含まれる混合音声のうちの1つ以上の音源からの信号をICA(Independent Component Analysis)法に基づくBSS(Blind Source Separation)方式を用いて分離する技術が開示されている。

【0006】

【特許文献1】特開2005-341073号公報

【特許文献2】特開2006-154314号公報

【発明の開示】

【発明が解決しようとする課題】

40

【0007】

しかし、従来の音声処理装置では、多数のマイクロホンを設ける必要があったため、音声処理装置のハードウェア規模が大きくなってしまふ。また、従来の音声処理装置はマイクロホンの指向性を利用して操作者の音声を選別するため、操作者の位置に対して制約が課されるという問題があった。

【0008】

そこで、本発明は、上記問題に鑑みてなされたものであり、本発明の目的とするところは、特定音源から発せられた音声の全体に占める音量比率を調整して記録することが可能な、新規かつ改良された音声処理装置、音声処理方法およびプログラムを提供することにある。

50

【課題を解決するための手段】

【0009】

上記課題を解決するために、本発明のある観点によれば、入力音声に特定音源から発せられた第一の音声が含まれているか否かを音源の位置情報に基づいて判定する音声判定部と、音声判定部により入力音声に第一の音声が含まれていると判定された場合、入力音声を第一の音声と特定音源以外の音源から発せられた第二の音声とに分離する音声分離部と、音声分離部により分離された第一の音声と第二の音声を、任意の音量比率で混合する音声混合部と、を備えることを特徴とする、音声処理装置が提供される。

【0010】

かかる構成においては、音声分離部が、入力音声に含まれる特定音源から発せられた第一の音声を分離し、音声混合部が、例えば第一の音声と入力音声に含まれる他の音声である第二の音声とを、第一の音声が占める音量比率が、入力音声に占める第一の音声の音量比率より低減されるように混合する。したがって、入力音声のうち特定音源から発せられた第一の音声の音量が不要に大きい場合、音声混合部は、第二の音声の音量比率が入力音声に占める第二の音声の音量比率より増大させた混合音声を得ることができる。その結果、当該音声処理装置によれば、第二の音声の音量が第一の音声に埋もれてしまうことを防止できる。

【0011】

また、音声混合部は、例えば近傍から発せられた第一の音声と入力音声に含まれる他の音声である第二の音声とを、第一の音声の音量比率が、入力音声に占める第一の音声の音量比率より増加されるように混合してもよい。かかる構成によれば、音声収録者自身の音声の収録を所望する場合、音声収録者が発した第一の音声の音量比率を強調することができる。なお、音声判定部により入力音声に第一の音声が含まれていないと判定された場合、音声分離部は入力音声の分離を行なわなくてもよい。

【0012】

特定音源は入力音声の収録位置から設定距離範囲内に位置してもよい。すなわち、第一の音声が入力音声の収録位置から設定距離範囲内から発せられた音声であってもよい。ここで、音声の音量は距離が離れるにつれて減少するため、収録位置に近い音源から発せられた音声ほど入力音声に大きな音量で収録される場合が多い。したがって、音声混合部は、入力音声の収録位置から近い第一の音声の音量比率を抑制し、収録位置からの音源距離の差に起因するアンバランスな音量関係を是正することができる。

【0013】

第一の音声は、入力音声を收音する際に用いられた装置の操作者に起因する音声を含み、第二の音声は、收音対象から発せられた音声を含んでもよい。かかる構成によれば、入力音声を收音する際に用いられた装置の近傍で該装置を操作している操作者により発せられた第一の音声の音量比率を抑制し、收音対象から発せられた第二の音声の音量が第一の音声により埋もれてしまうことを防止可能である。

【0014】

音声判定部は、入力音声の音量または音質の少なくともいずれかに基づいて入力音声に第一の音声が含まれているか否かを判定してもよい。ここで、音声判定部は、入力音声の音量または位相に基づいて入力音声の音源の位置情報、または入力音声に含まれる1または2以上の音源から発せられた音声ごとの音源の位置情報を推定してもよい。

【0015】

当該音声処理装置は、映像を撮像する撮像部をさらに備え、音声判定部は、入力音声に含まれる1または2以上の音源から発せられた音声の音量または位相の少なくともいずれかに基づいて音源の位置情報を算出する位置情報算出部を備え、入力音声の音源の位置が撮像部の撮像方向の後方であると位置情報算出部により算出され、入力音声が入力の音声と一致または近似する音質である場合、入力音声に特定音源から発せられた第一の音声が含まれていると判定してもよい。ここで、操作者は撮像部の撮像方向の後方から音声処理装置を操作する場合が多い。したがって、音声判定部は、入力音声の音源の位置が撮像部

の撮像方向の後方であり、入力音声が入間の音声と一致または近似する音質である場合、入力音声に第一の音声として操作者の音声が含まれていると判定することができる。その結果、音声混合部により操作者の音声の音量比率が低減された混合音声を得ることができる。

【0016】

入力音声の音源の位置が收音位置から設定距離の範囲内であり、入力音声にインパルス音が含まれ、入力音声が過去の平均音量と比較して大きい場合、音声判定部は、入力音声に特定音源から発せられた第一の音声が含まれていると判定してもよい。ここで、入力音声を収録する装置の操作者が該装置のボタンを操作したり該装置を持ち替えると「パチン」、

「パン」などのインパルス音が発生する場合が多い。また、該インパルス音は該装置において発生するため、比較的大きな音量で收音される可能性が高い。したがって、音声判定部は、入力音声の音源の位置が收音位置から設定距離の範囲内であり、入力音声にインパルス音が含まれ、入力音声

が過去の平均音量と比較して大きい場合、入力音声に第一の音声として操作者の動作に起因するノイズが支配的に含まれていると判定することができる。その結果、音声混合部により操作者の動作に起因するノイズの音量比率が低減された混合音声を得ることができる。

10

【0017】

当該音声処理装置は、入力音声を收音する複数の收音部と、音声混合部により混合された混合音声を記憶媒体に記録する記録部と、を備えてもよい。かかる構成においては、記録部は記憶媒体に、第一の音声

が占める音量比率が入力音声に占める第一の音声の音量比率より低減された混合音声を記録する。したがって、該混合音声を再生する再生装置に特殊な音量補正機能を実装することなく、該再生装置において第一の音声の占める音量比率が調整された混合音声を再生することが可能となる。

20

【0018】

当該音声処理装置は、入力音声を記憶している記憶媒体と、記憶媒体に記憶されている入力音声を再生し、位置情報算出部、音声判定部および音声分離部の少なくともいずれかに出力する再生部と、を備えてもよい。かかる構成においては、位置情報算出部、音声判定部および音声分離部は再生部から入力される入力音声に基づいて混合音声を生成し、混合音声を再生音声として出力することができる。したがって、記憶媒体に入力音声を記録する記録装置に特殊な音量補正機能を実装することなく、第一の音声の占める音量比率が調整された混合音声を再生することが可能となる。

30

【0019】

当該音声処理装置は、入力音声の音量が補正されている場合、音声分離部により分離された第二の音声の音量に、補正の程度に応じた逆補正を行なう音量補正部を備えてもよい。例えば、第一の音声の音量が過大であったために入力音声の音量が全体として抑制された場合、第二の音声の音量も抑制されてしまっている。音量補正部は、このような場合、入力音声の音量が抑制された程度に応じて第二の音声の音量を増大させ、第二の音声

が過小となることを防止できる。

【0020】

また、上記課題を解決するために、本発明の別の観点によれば、入力音声の分離を行う音声分離部と、音声分離部により分離された音声に特定音源から発せられた第一の音声が含まれているか否かを判定する音声判定部と、音声分離部により分離された第一の音声と特定音源以外の音源から発せられた第二の音声を、任意の音量比率で混合する音声混合部と、を備えることを特徴とする、音声処理装置が提供される。

40

【0021】

また、上記課題を解決するために、本発明の別の観点によれば、コンピュータを、入力音声に特定音源から発せられた第一の音声が含まれているか否かを音源の位置情報に基づいて判定する音声判定部と、音声判定部により入力音声に第一の音声が含まれていると判定された場合、入力音声を第一の音声と特定音源以外の音源から発せられた第二の音声とに分離する音声分離部と、音声分離部により分離された第一の音声と第二の音声を任意の

50

音量比率で混合する音声混合部とを備えることを特徴とする音声処理装置として機能させるための、プログラムが提供される。

【 0 0 2 2 】

かかるプログラムは、例えばCPU、ROMまたはRAMなどを含むコンピュータのハードウェア資源に、上記のような位置情報算出部、音声判定部および音声分離部の機能を実行させることができる。すなわち、当該プログラムを用いるコンピュータを、上述の音声処理装置として機能させることが可能である。

【 0 0 2 3 】

音声判定部は、音源の位置情報、入力音声の音量または音質の少なくともいずれかに基づいて入力音声に第一の音声が含まれているか否かを判定してもよい。

10

【 0 0 2 4 】

映像を撮像する撮像部をさらに備え、音声判定部は、入力音声に含まれる1または2以上の音源から発せられた音声の音量または位相の少なくともいずれかに基づいて音源の位置情報を算出する位置情報算出部を備え、入力音声の音源の位置が撮像部の撮像方向の後方であると位置情報算出部により算出され、入力音声が人間の音声と一致または近似する音質である場合、入力音声に特定音源から発せられた第一の音声が含まれていると判定してもよい。

【 0 0 2 5 】

入力音声の音源の位置が收音位置から設定距離の範囲内であり、入力音声にインパルス音が含まれ、入力音声が過去の平均音量と比較して大きい場合、音声判定部は、入力音声に特定音源から発せられた第一の音声が含まれていると判定してもよい。

20

【 0 0 2 6 】

また、上記課題を解決するために、本発明の別の観点によれば、入力音声に特定音源から発せられた第一の音声が含まれているか否かを音源の位置情報に基づいて判定するステップと、入力音声に第一の音声が含まれていると判定された場合、入力音声を第一の音声と特定音源以外の音源から発せられた第二の音声とに分離するステップと、分離された第一の音声と第二の音声を、任意の音量比率で混合するステップと、を含むことを特徴とする、音声処理方法が提供される。

【発明の効果】

【 0 0 2 7 】

以上説明したように本発明にかかる音声処理装置、音声処理方法およびプログラムよれば、特定音源から発せられた音声の全体に占める音量比率を任意に調整して出力、又は記録することができる。

30

【発明を実施するための最良の形態】

【 0 0 2 8 】

以下に添付図面を参照しながら、本発明の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

【 0 0 2 9 】

(第1の実施形態)

40

まず、本発明の第1の実施形態にかかる音声記録装置10について説明する。本実施形態の説明においては、図1および図2を参照して音声記録装置10が用いられる場面の一例を説明した後に、図3～図10を参照して音声記録装置10の構成および動作を説明する。

【 0 0 3 0 】

図1は、本実施形態にかかる音声記録装置10が用いられる場面の一例を示した説明図である。図1に示した例では、被写体である子供が品川区立一番小学校の校門の前に立っており、映像撮像機能が実装された音声記録装置10を手にした操作者が音声記録装置10を被写体に向けている。

【 0 0 3 1 】

50

また、被写体は、操作者の「おーい」という呼びかけに「はーい」という返事をしている。このとき、映像撮像機能が実装された音声記録装置 10 は、被写体の映像と共に操作者の「おーい」という呼びかけ、および被写体の「はーい」という返事を記録する。ここで、図 2 を参照して通常の音声記録方法によって記録される音声について説明する。

【0032】

図 2 は、通常の音声記録方法によって記録される音声の時間領域の振幅を示した説明図である。音声は、音源が点音源であると仮定すると、收音される音量は音源と收音位置との距離の二乗に反比例する。すなわち、收音位置が音源から離れるほど收音される音量は小さくなる。したがって、收音位置に近い操作者の「おーい」という呼びかけは、図 2 (a) に示すような振幅を有する音声として收音される。

10

【0033】

一方、收音位置から操作者より離れている被写体の「はーい」という返事は、図 2 (b) に示したように操作者の声に比べて小さな振幅の音声として收音される。この場合、通常の音声記録方法によれば、図 2 (c) に示したように、単純に操作者の「おーい」という呼びかけと被写体の「はーい」という返事が重畳された音声記録される。

【0034】

しかし、図 2 (c) に示した音声には操作者の「おーい」という呼びかけが支配的に含まれ、被写体の「はーい」という返事が埋もれてしまっている。同様に、操作者による操作ノイズが被写体の発する音声と比較して相対的に大きく記録されてしまう。このため、被写体の発する音声が操作者に起因する音声によってマスキングされ、操作者の意図した適切な音量バランスで被写体の発する音声を記録できない場合が多いという問題があった。

20

【0035】

そこで、上記の問題を一着眼点とし、本実施形態にかかる音声記録装置 10 が創作されるに至った。本実施形態にかかる音声記録装置 10 は、操作者に起因する音声の音量比率を抑制し、被写体の発する音声と操作者に起因する音声とを適切な音量バランスで記録することができる。以下、このような音声記録装置 10 の詳細な構成および動作について説明する。

【0036】

図 3 は、本実施形態にかかる音声処理装置の一例としての音声記録装置 10 の構成を示した機能ブロック図である。音声記録装置 10 は、音声收音部 110 と、音声判定部 120 と、音源分離部 140 と、音声混合部 150 と、記録部 160 と、記憶部 170 と、を備える。なお、図 1 においては音声記録装置 10 としてビデオカメラを示しているが、音声記録装置 10 はビデオカメラに限られず、PC (Personal Computer)、携帯電話、PHS (Personal Handyphone System)、携帯用音声処理装置、携帯用映像処理装置、PDA (Personal Digital Assistant)、家庭用ゲーム機器、携帯用ゲーム機器、などの情報処理装置であってもよい。

30

【0037】

音声收音部 110 は、音声を收音し、收音した音声を離散量子化する。また、音声收音部 110 は、物理的に分離された 2 以上の收音部 (例えば、マイクロホン) を含む。図 3 に示した例では、音声收音部 110 は、左音声 L を收音する收音部と右音声 R を收音する收音部の 2 つを含む。音声收音部 110 は、離散量子化した左音声 L および右音声 R を入力音声として音声判定部 120 および音源分離部 140 へ出力する。

40

【0038】

音声判定部 120 は、音声收音部 110 から入力された入力音声に、操作者の音声または操作者の動作に起因するノイズなど音声記録装置 10 の近傍から発せられた近傍音声 (第一の音声) が含まれているか否かを判定する。かかる音声判定部 120 の詳細な構成を図 4 を参照して説明する。

【0039】

50

図4は、音声判定部120の構成を示した機能ブロック図である。音声判定部120は、音量検出器124、平均音量検出器126および最大音量検出器128からなる音量検出部122と、スペクトル検出器132および音質検出器134からなる音質検出部130と、距離方向推定器136と、操作者音声推定器138と、を備える。なお、図4においては図面の明瞭性の観点から左音声Lおよび右音声Rを合わせて入力音声と示している。

【0040】

音量検出器124は、所定長さのフレーム単位（例えば、数10msec）で与えられる入力音声の音量値列（振幅）を検出し、検出した入力音声の音量値列を平均音量検出器126、最大音量検出器128、音質検出器134および距離方向推定器136に出力する。

10

【0041】

平均音量検出器126は、音量検出器124から入力されるフレーム単位の音量値列に基づいて、入力音声の音量平均値を例えばフレームごとに検出する。また、平均音量検出器126は、検出した音量平均値を音質検出器134および操作者音声推定器138に出力する。

【0042】

最大音量検出器128は、音量検出器124から入力されるフレーム単位の音量値列に基づいて、入力音声の音量最大値を例えばフレームごとに検出する。また、最大音量検出器128は、検出した入力音声の音量最大値を音質検出器134および操作者音声推定器138に出力する。

20

【0043】

スペクトル検出器132は、入力音声に例えばFFT（Fast Fourier Transform）処理を施し、入力音声の周波数領域における各スペクトルを検出する。スペクトル検出器132は、検出したスペクトルを音質検出器134および距離方向推定器136に出力する。

【0044】

音質検出器134は、入力音声、音量平均値、音量最大値およびスペクトルが入力され、かかる入力に基づいて入力音声の人間の音声らしさ、音楽らしさ、定常性、インパルス性などを検出し、操作者音声推定器138に出力する。人間の音声らしさは、入力音声の一部または全体が人間の音声と一致するか否か、あるいは人間の音声とどの程度近似するかなどを示す情報であってもよい。また、音楽らしさは、入力音声の一部または全体が音楽であるか否か、あるいは音楽とどの程度近似するかなどを示す情報であってもよい。

30

【0045】

定常性は、例えば空調音のように時間的にそれほど音声の統計的性質が変化しない性質を指す。インパルス性は、例えば打撃音、破裂音のように短時間にエネルギーが集中した雑音性の強い性質を指す。

【0046】

例えば、音質検出器134は、入力音声のスペクトル分布と人間の音声のスペクトル分布との一致度に基づいて人間の音声らしさを検出することができる。また、音質検出器134は、フレームごとの音量最大値を比較し、他のフレームと比較して音量最大値が大きいほどインパルス性が高いことを検出してもよい。

40

【0047】

なお、音質検出器134は、ゼロクロッシング法、LPC（Linear Predictive Coding）分析などの信号処理技術を用いて入力音声の音質を分析してもよい。ゼロクロッシング法によれば入力音声の基本周期が検出されるため、音質検出器134は該基本周期が人間の音声の基本周期（例えば100～200Hz）に含まれるか否かに基づいて人間の音声らしさを検出してもよい。

【0048】

距離方向推定器136は、入力音声、入力音声の音量値列、入力音声のスペクトルなど

50

が入力され、該入力に基づいて入力音声の音源または入力音声に含まれる支配的な音声が発せられた音源の方向情報および距離情報などの位置情報を推定する位置情報算出部としての機能を有する。かかる距離方向推定器 136 は、入力音声の位相、音量、音量値列、過去の平均音量値、最大音量値などによる音源の位置情報の推定方法を組み合わせることによって、残響や映像記録装置 10 本体による音声の反射の影響が大きい場合でも総合的に音源位置を推定することができる。距離方向推定器 136 による方向情報および距離情報の推定方法の一例を図 5 ～ 図 8 を参照して説明する。

【0049】

図 5 は、2 つの入力音声の位相差に基づいて入力音声の音源位置を推定する様子を示した説明図である。音源が点音源であると仮定すると、音声収音部 110 を構成するマイクロホン M1 およびマイクロホン M2 に到達する各入力音声の位相と各入力音声の位相差が測定できる。さらに、位相差と、入力音声の周波数 f および音速 c の値から、入力音声の音源位置までのマイクロホン M1 からの距離とマイクロホン M2 からの距離との差を算出できる。音源は、当該距離差が一定である点の集合上に存在する。このような距離差が一定である点の集合は、双曲線となることが知られている。

【0050】

例えば、マイクロホン M1 が $(x_1, 0)$ に位置し、マイクロホン M2 が $(x_2, 0)$ に位置すると仮定する（このように仮定しても一般性を失わない）。また、求める音源位置の集合上の点を (x, y) とおき、上記距離差を d とおくと、以下の数式 1 が成り立つ。

【数 1】

$$\sqrt{(x-x_1)^2+y^2}-\sqrt{(x-x_2)^2+y^2}=d$$

（数式 1）

【0051】

さらに、数式 1 は数式 2 のように展開でき、数式 2 を整理すると双曲線を表す数式 3 が導かれる。

【数 2】

$$\{(x-x_1)^2+2y^2+(x-x_2)^2-d^2\}^2=4\{(x-x_1)^2+y^2\}\{(x-x_2)^2+y^2\} \quad (\text{数式 2})$$

【数 3】

$$\frac{\left(x-\frac{x_1+x_2}{2}\right)^2}{\left(\frac{d}{2}\right)^2}-\frac{y^2}{\left(\frac{1}{2}\right)^2}=1 \quad (\text{数式 3})$$

【0052】

また、距離方向推定器 136 は、マイクロホン M1 およびマイクロホン M2 の各々が収音した入力音声の音量差に基づいて音源がマイクロホン M1 およびマイクロホン M2 のどちらの近傍であるかを判定できるため、例えば図 5 に示したようにマイクロホン M2 に近い双曲線 1 上に音源が存在すると判定することができる。

【0053】

なお、位相差算出に用いる入力音声の周波数 f は、マイクロホン M1 およびマイクロホン M2 間の距離に対して下記の数式 4 の条件を満たす必要がある。

10

20

30

40

【数 4】

$$f < \frac{c}{2d} \quad (\text{数式 4})$$

【0054】

図 6 は、3 つの入力音声の位相差に基づいて入力音声の音源位置を推定する様子を示した説明図である。図 6 に示したような音声収音部 110 を構成するマイクロホン M3、マイクロホン M4 およびマイクロホン M5 の配置を想定した場合、マイクロホン M3 およびマイクロホン M4 に到達する入力音声の位相に比較してマイクロホン M5 に到達する入力音声の位相が遅れていれば、距離方向推定器 136 は、音源が マイクロホン M3 およびマイクロホン M4 を結ぶ直線 1 に対してマイクロホン M5 の逆側に位置すると判定できる（前後判定）。

10

【0055】

さらに、距離方向推定器 136 は、マイクロホン M3 およびマイクロホン M4 の各々に到達する入力音声の位相差に基づいて音源が存在し得る双曲線 2 を算出し、マイクロホン M4 およびマイクロホン M5 の各々に到達する入力音声の位相差に基づいて音源が存在し得る双曲線 3 を算出することができる。その結果、距離方向推定器 136 は、双曲線 2 および双曲線 3 の交点 P1 を音源位置として推定することができる。

【0056】

20

図 7 は、2 つの入力音声の音量に基づいて入力音声の音源位置を推定する様子を示した説明図である。音源が点音源であると仮定すると、逆二乗則よりある点で観測される音量は距離の二乗に反比例する。図 7 に示したような音声収音部 110 を構成するマイクロホン M6 およびマイクロホン M7 を想定した場合、マイクロホン M6 およびマイクロホン M7 に到達する音量比が一定となる点の集合は円となる。距離方向推定器 136 は、音量検出器 124 から入力される音量の値から音量比を求め、音源の存在する円の半径及び中心位置を算出できる。

【0057】

図 7 に示したように、マイクロホン M6 が $(x_3, 0)$ に位置し、マイクロホン M7 が $(x_4, 0)$ に位置する場合（このように仮定しても一般性を失わない）、求める音源位置の集合上の点を (x, y) と置くと、各マイクロホンから音源までの距離 r_1 、 r_2 は以下の数式 5 のように表せる。

30

【数 5】

$$r_1 = \sqrt{(x - x_3)^2 + y^2} \quad r_2 = \sqrt{(x - x_4)^2 + y^2} \quad (\text{数式 5})$$

【0058】

ここで、逆二乗則より以下の数式 6 が成り立つ。

【数 6】

$$\frac{1}{r_1^2} : \frac{1}{r_2^2} = \text{一定} \quad (\text{数式 6})$$

40

【0059】

数式 6 は正の定数 d （例えば 4）を用いて数式 7 のように変形される。

【数 7】

$$\frac{r_2^2}{r_1^2} = d \quad (\text{数式 7})$$

【0060】

数式 7 を r_1 および r_2 に代入し、整理すると以下の数式 8 が導かれる。

50

【数 8】

$$\frac{(x-x_4)^2+y^2}{(x-x_3)^2+y^2}=d$$

$$\left(x-\frac{x_4-dx_3}{1-d}\right)^2+y^2=\frac{d(x_4-x_3)^2}{(1-d)^2} \quad (\text{数式 8})$$

【0061】

数式 8 より、距離方向推定器 136 は、図 7 に示したように、中心の座標が数式 9 で表され半径が数式 10 で表される円 1 上に音源が存在すると推定できる。

10

【数 9】

$$\left(\frac{x_4-dx_3}{1-d}, 0\right) \quad (\text{数式 9})$$

【数 10】

$$\left|\frac{x_4-x_3}{1-d}\right|\sqrt{d} \quad (\text{数式 10})$$

【0062】

図 8 は、3つの入力音声の音量に基づいて入力音声の音源位置を推定する様子を示した説明図である。図 8 に示したような音声収音部 110 を構成するマイクロホン M3、マイクロホン M4 およびマイクロホン M5 の配置を想定した場合、マイクロホン M3 およびマイクロホン M4 に到達する入力音声の位相に比較してマイクロホン M5 に到達する入力音声の位相が遅れていれば、距離方向推定器 136 は、音源がマイクロホン M3 およびマイクロホン M4 を結ぶ直線 2 に対してマイクロホン M5 の逆側に位置すると判定できる（前後判定）。

20

【0063】

さらに、距離方向推定器 136 は、マイクロホン M3 およびマイクロホン M4 の各々に到達する入力音声の音量比に基づいて音源が存在し得る円 2 を算出し、マイクロホン M4 およびマイクロホン M5 の各々に到達する入力音声の音量比に基づいて音源が存在し得る円 3 を算出することができる。その結果、距離方向推定器 136 は、円 2 および円 3 の交点 P2 を音源位置として推定することができる。なお、4つ以上のマイクロホンを使用した場合には、距離方向推定器 136 は、空間的な音源の配置を含め、より精度の高い推定が可能となる。

30

【0064】

距離方向推定器 136 は、上記のように各入力音声の位相差や音量比に基づいて入力音声の音源の位置を推定し、推定した音源の方向情報や距離情報を操作者音声推定器 138 に出力する。以下の表 1 に、上述した音量検出部 122、音質検出部 130 および距離方向推定器 136 の各構成の入出力をまとめた。

【表 1】

ブロック	入力	出力
音量検出器	入力音声	フレーム内の音量値列（振幅）
平均音量検出器	フレーム内の音量値列（振幅）	音量平均値
最大音量検出器	フレーム内の音量値列（振幅）	音量最大値
スペクトル検出器	入力音声	スペクトル
音質検出器	入力音声 音量平均値 音量最大値 スペクトル	人間の音声らしさ 楽音らしさ 定常 or 非定常 インパルス性
距離方向推定器	入力音声 フレーム内の音量値列（振幅） スペクトル	方向情報 距離情報

10

【 0 0 6 5 】

なお、入力音声に複数の音源から発せられた音声为重畳されている場合、距離方向推定器 1 3 6 は入力音声に支配的に含まれている音声の音源位置を正確に推定することは困難である。しかし、距離方向推定器 1 3 6 は入力音声に支配的に含まれている音声の音源位置に近い位置を推定することは可能である。また、当該推定された音源位置は音源分離部 1 4 0 において音声分離のための初期値として利用してもよい。また、距離方向推定器 1 3 6 が推定する音源位置に誤差があっても当該音声記録装置 1 0 は所望の動作をすることができる。

20

【 0 0 6 6 】

図 4 を参照して音声判定部 1 2 0 の構成の説明に戻ると、操作者音声推定器 1 3 8 は、入力音声の音量、音質または位置情報の少なくともいずれかに基づき、入力音声に操作者の音声または操作者の動作に起因するノイズなど音声記録装置 1 0 の近傍である特定音源から発せられた近傍音声が含まれているか否かを総合的に判定する。また、操作者音声推定器 1 3 8 は、入力音声に近傍音声が含まれていると判定した場合、音源分離部 1 4 0 に入力音声に近傍音声が含まれる旨（操作者音声存在情報）や距離方向推定器 1 3 6 により推定された位置情報などを出力する音声判定部としての機能を有する。

30

【 0 0 6 7 】

具体的には、操作者音声推定器 1 3 8 は、入力音声の音源の位置が映像を撮像する撮像部（図示せず。）の撮像方向の後方であると距離方向推定器 1 3 6 に推定され、入力音声の音質が人間の音声と一致または近似する音質である場合、入力音声に近傍音声が含まれていると判定してもよい。ここで、図 9 に示すように、操作者は撮像部の撮像方向の後方、すなわちファインダーの左後方から音声記録装置 1 0 を操作する場合が多い（右利きで自分撮り以外の通常の撮影時）。

【 0 0 6 8 】

したがって、操作者音声推定器 1 3 8 は、入力音声の音源の位置が撮像部の撮像方向の後方であり、入力音声の音質が人間の音声と一致または近似する音質である場合、入力音声に近傍音声として操作者の音声が含まれていると判定することができる。その結果、後述の音声混合部 1 5 0 により操作者の音声の音量比率が低減された混合音声を得ることができる。

40

【 0 0 6 9 】

また、操作者音声推定器 1 3 8 は、入力音声の音源の位置が收音位置から設定距離（例えば、音声記録装置 1 0 の 1 m 以内など音声記録装置 1 0 の近傍）の範囲内であり、入力音声にインパルス音が含まれ、入力音声の音量が過去の平均音量と比較して大きい場合、入力音声に特定音源から発せられた近傍音声が含まれていると判定してもよい。ここで、音声記

50

録装置 10 の操作者が音声記録装置 10 のボタンを操作したり音声記録装置 10 を持ち替えると「パチン」、「パン」などのインパルス音が発生する場合が多い。また、該インパルス音は音声記録装置 10 において発生するため、比較的大きな音量で収音される可能性が高い。

【0070】

したがって、操作者音声推定器 138 は、入力音声の音源の位置が収音位置から設定距離の範囲内であり、入力音声にインパルス音が含まれ、入力音声が過去の平均音量と比較して大きい場合、入力音声に近傍音声として操作者の動作に起因するノイズが支配的に含まれていると判定することができる。その結果、後述の音声混合部 150 により操作者の動作に起因するノイズの音量比率が低減された混合音声を得ることができる。

10

【0071】

その他、操作者音声推定器 138 に入力される情報と、入力される情報に基づく操作者音声推定器 138 の判定結果の一例を以下の表 2 にまとめた。なお、近接センサー、温度センサーなどを組み合わせて用いて操作者音声推定器 138 における判定の精度をあげることも可能である。

【表 2】

操作者音声推定器入力									判定結果
音量			音質				方向及び距離		
音量	平均音量	最大音量	人間の音声らしさ	音楽らしさ	定常 or 非定常	インパルス性	方向	距離	
大	過去の平均音量と比較して大	大	高い	低い	非定常	普通	本体後方	近い	操作者音声
大	過去の平均音量と比較して大	大	低い	低い	非定常	高い	全方向	近い	操作ノイズ
小	過去の平均音量と比較して小	小	普通	普通	定常	低い	方向不明瞭	遠い	定常環境音
小	過去の平均音量より比較的小	中	低い	低い	非定常	高い	全方向	遠い	インパルス状環境音
中	過去の平均音量より比較的大	中～大	普通	普通	非定常	普通	本体前方	近い～遠い	被写体音声

20

30

【0072】

図 3 を参照して音声記録装置 10 の構成の説明に戻ると、音源分離部 140 は、音声判定部 120 から操作者音声存在情報が入力されると、音声判定部 120 から入力される音源の位置情報に基づき、音声収音部 110 から入力される入力音声进行操作者の音声などの近傍音声と、近傍音声以外の被写体の音声などの収音対象音声（第二の音声）とに分離する。その結果、音源分離部 140 は、入力される入力音声の数の倍の数の音声を出力する。図 3 においては、音源分離部 140 が左音声 L および右音声 R を入力音声として入力され、左近傍音声 L 1 および 右近傍音声 R 1 を近傍音声として出力し、左収音対象音声 L 2 および 右収音対象音声 R 2 を収音対象音声として出力する様子を示している。

40

【0073】

具体的には、音源分離部 140 は、独立成分解析を用いた手法（ICA）、音の時間周波数成分間の重なり度の少なさを利用する手法などを用いて音声を音源に応じて分離する音声分離部として機能する。

【0074】

音声混合部 150 は、音源分離部 140 から入力された近傍音声および収音対象音声を

50

、近傍音声が占める音量比率が、入力音声に占める近傍音声の音量比率より低減されるように混合する。かかる構成によれば、入力音声のうち特定音源から発せられた近傍音声の音量が不要に大きい場合、音声混合部150は、收音対象音声に占める音量比率が入力音声に占める收音対象音声の音量比率より増大させた混合音声を得ることができる。その結果、当該音声記録装置10によれば、收音対象音声に近傍音声に埋もれてしまうことを防止できる。

【0075】

なお、音声混合部150は、入力される左近傍音声L1および左收音対象音声L2を混合して混合左音声Lを生成し、入力される右近傍音声R1および右收音対象音声R2を混合して混合右音声Rを生成し、混合左音声Lおよび混合右音声Rを混合音声として記録部160に出力する。

10

【0076】

また、音声混合部150は、音源分離部140により分離された近傍音声および收音対象音声の平均音量比から適切な混合比率を算出し、算出した混合比率で近傍音声および收音対象音声を混合してもよい。また、音声混合部150は、前フレームに適用していた混合比率との差分が所定の上限值を超えない範囲で適用する混合比率を変化させてもよい。また、該混合比率はユーザ設定されるようにしてもよい。

【0077】

記録部160は、音声混合部150から入力された混合音声を記憶部170に記録する。記憶部170は、EEPROM(Electrically Erasable Programmable Read-Only Memory)、EPROM(Erasable Programmable Read Only Memory)などの不揮発性メモリや、ハードディスクおよび円盤型磁性体ディスクなどの磁気ディスクや、CD-R(Compact Disk Recordable)/RW(Rewritable)、DVD-R(Digital Versatile Disk Recordable)/RW/+R/+RW/RAM(Random Access Memory)およびBD(Blu-Ray Disc(登録商標))R/BD-REなどの光ディスクや、MO(Magneto Optical)ディスクなどの記憶媒体であってもよい。なお、記憶部170は、被写体の映像データも記憶することができる。

20

【0078】

このように本実施形態にかかる音声記録装置10によれば、記録部160が記憶部170に、近傍音声に占める音量比率が入力音声に占める近傍音声の音量比率より低減された混合音声を記録する。したがって、該混合音声を再生する再生装置に特殊な音量補正機能を実装することなく、該再生装置において近傍音声の占める音量比率が調整された混合音声を再生することが可能となる。

30

【0079】

以上、本実施形態にかかる音声記録装置10の構成を説明した。続いて、図10を参照して本実施形態にかかる音声記録装置10において実行される音声処理方法を説明する。

【0080】

図10は、本実施形態にかかる音声記録装置10において実行される音声処理方法の流れを示したフローチャートである。まず、音声記録装置10の音声收音部110は音声を收音する(S210)。入力音声が無かった場合には処理を終了し、入力音声があった場合には距離方向推定器136が入力音声の全体または一部が発せられた音源の距離や方向などの位置情報を推定する(S230)。

40

【0081】

その後、操作者音声推定器138は入力音声に操作者の発した音声、または操作者の動作に起因するノイズなどの近傍音声が含まれているか否かを判定する(S240)。操作者音声推定器138により入力音声に近傍音声が含まれていると判定された場合、音源分離部140は、入力音声を近傍音声とそれ以外の收音対象音声とに分離する(S250)。

50

【0082】

続いて、音声混合部150が音源分離部140により分離された近傍音声と収音対象音声とを任意の比率で混合し、混合音声を生成する(S260)。S260の後、またはS240において入力音声に操作者の発した音声、または操作者の動作に起因するノイズなどの近傍音声が含まれていないと判定された場合、記録部160は混合音声または入力音声を記憶部170に記録する(S270)。

【0083】

以上説明したように、本実施形態にかかる音声記録装置10は、音源分離部140が、入力音声に含まれる特定音源から発せられた近傍音声を距離方向推定器136により推定された入力音声の音源の位置情報に基づいて分離し、音声混合部150が、近傍音声と入力音声に含まれる他の音声である収音対象音声とを、近傍音声が占める音量比率が、入力音声に占める近傍音声の音量比率より低減されるように混合する。

【0084】

したがって、入力音声のうち特定音源から発せられた近傍音声の音量が不要に大きい場合、音声混合部150は、収音対象音声に占める音量比率が入力音声に占める収音対象音声の音量比率より増大させた混合音声を得ることができる。その結果、当該音声記録装置10によれば、近傍音声を相対的に抑制し、収音対象音声に近傍音声に埋もれてしまうことを防止できる。また、入力音声に含まれる操作者により発せられる音声、ノイズなどの近傍音声の影響を低減除去した高品質の混合音声を記録することができる。

【0085】

また、音声記録装置10は、記憶部170に近傍音声に占める音量比率が入力音声に占める近傍音声の音量比率より低減された混合音声を記録できる。したがって、該混合音声を再生する再生装置に特殊な音量補正機能を実装することなく、該再生装置において近傍音声の占める音量比率が調整された混合音声を再生することが可能となる。

【0086】

また、本実施形態にかかる音声記録装置10は、入力音声をソフトウェア的に処理し近傍音声および収音対象音声の音量比率を調整した混合音声を記録できるため、マイクロホンの数などのハードウェア規模を縮小することができる。

【0087】

(第2の実施形態)

次に、本発明の第2の実施形態にかかる音声再生装置11について説明する。本実施形態にかかる音声再生装置11は、既に記憶されている音声に含まれる近傍音声の占める音量比率が調整された混合音声を再生することができる。以下、図11を参照して当該音声再生装置11の構成を説明する。

【0088】

図11は、本実施形態にかかる音声再生装置11の構成を示した機能ブロック図である。本実施形態にかかる音声再生装置11は、音声判定部120と、音源分離部140と、音声混合部150と、記憶部172と、再生部174と、音声出力部180と、を備える。

なお、本実施形態の説明においては、第1の実施形態で説明した内容と実質的に同一である構成については説明を省略し、第1の実施形態と異なる構成に重きをおいて説明する。

【0089】

記憶部172は、音声の記録機能を有する任意の装置において記録された音声を記憶している。再生部174は、記憶部172が記憶している音声を読み出し、必要に応じてデコードを行なう。そして、再生部174は、記憶部172が記憶している音声を音声判定部120および音源分離部140に出力する。音声判定部120および音源分離部140は、再生部174からの出力を入力音声として扱い、第1の実施形態で説明した内容と実質的に同一な処理を行う。

【0090】

音声出力部180は、音声混合部150により混合された混合音声を出力する。音声出

10

20

30

40

50

力部 180 は、例えばスピーカであってもイヤホンであってもよい。なお、本実施形態にかかる記憶部 172 も、第 1 の実施形態における記憶部 170 と同様に E E P R O M、E P R P M などの不揮発性メモリや、ハードディスクおよび円盤型磁性体ディスクなどの磁気ディスクや、C D - R / R W、D V D - R / R W / + R / + R W / R A M および B D (B l u - R a y D i s c (登録商標)) R / B D - R E などの光ディスクや、M O ディスクなどの記憶媒体であってもよい。

【0091】

このように、本実施形態にかかる音声再生装置 11 は、音声判定部 120、音源分離部 140 および音声混合部 150 が再生部 174 から入力される入力音声に基づいて混合音声10を生成し、混合音声10を再生音声として出力することができる。したがって、記憶部 172 に入力音声を記録する記録装置に特殊な音量補正機能を実装することなく、近傍音声の占める音量比率が調整された混合音声10を再生することが可能となる。また、操作者により発せられる音声、ノイズなどの近傍音声の影響を低減除去した高品質の混合音声10を出力することができる。

【0092】

(第3の実施形態)

次に、本発明の第3の実施形態にかかる音声再生装置 12 について説明する。本実施形態にかかる音声再生装置 12 は、入力音声に A G C (A u t o G a i n C o n t r o l) が施されている場合、入力音声に含まれる収音対象音声の音量を逆補正し、収音対象20音声を強調(ブースト)することができる。以下、図 12 および図 13 を参照し、本実施形態にかかる音声再生装置 12 の構成および動作を説明する。

【0093】

図 12 は、本実施形態にかかる音声再生装置 12 の構成を示した機能ブロック図である。音声再生装置 12 は、音声判定部 120 と、音源分離部 140 と、音声混合部 150 と、記憶部 172 と、再生部 174 と、音声出力部 180 と、音量補正部 190 とを備える。

なお、本実施形態の説明においては、第2の実施形態で説明した内容と実質的に同一である構成については説明を省略し、第2の実施形態と異なる構成に重きをおいて説明する。

【0095】

本実施形態にかかる記憶部 172 は、一部または全体に A G C (音量補正) が施された30音声を記憶している。ここで、A G C は、音量の過大入力に対して自動的に音量レベルを下げ、音割れ防止を一つの目的とするコンプレッサーの機構である。かかる A G C が施された音声の音量について図 13 を参照して説明する。

【0096】

図 13 は、A G C の適用前の音声(原音)の音量と、A G C 適用後の音声の音量を対比的に表した説明図である。A G C は、A G C の適用前の音声の音量が閾値 t_h を越えると、アタックタイムとして設定された時間で所定の割合 ($r a t i o$) まで音量を圧縮する。図 13 に示した例では、アタックタイムとして設定された時間で、A G C の適用前の音声の音量がおよそ $1/2 \sim 2/3$ 程度に圧縮される場合を示している。その後、A G C の適用前の音声の音量が閾値 t_h を下回ると、リリースタイムとして設定された時間内に A40 G C を解除する。

【0097】

ここで、音声の音量が閾値 t_h を超え A G C が動作するのは、該音声の記録装置の近傍から過大レベルの近傍音声の入力があった場合が多い。すなわち、遠方音源の収音対象音声によって A G C が動作する場合は少ない。しかし、入力音声は全体として A G C により圧縮されるため、入力音声に含まれる近傍音声のみならず、もともと微弱な収音対象音声10がさらに圧縮されてしまう問題があった。

【0098】

そこで、上記の問題を一着眼点とし、本実施形態にかかる音声再生装置 12 が創作されるに至った。本実施形態にかかる音声再生装置 12 は、音量補正部 190 の機能に基づき50

、入力音声にA G Cが施されていた場合であっても収音対象音声をブーストすることができる。

【0099】

音量補正部190は、音源分離部140により分離された近傍音声の音量の変化からA G Cが適用されたであろうアタックタイムを検出し、音源分離部140により分離された収音対象音声において該アタックタイムに相当する区間を走査する。収音対象音声には、背景環境音や被写体が発した音声などが含まれ得るが、背景環境音のみ含まれる場合は音量レベルがほぼ一定であると近似することが可能である。したがって、音量補正部190は、収音対象音声の音量に所定レベル以上の変化が生じている区間にはA G Cが施されたと判定することができる。

10

【0100】

そこで、音量補正部190は、収音対象音声における当該区間の音量を該区間の前後区間の音量と同程度になるように調整する逆補正を行い、収音対象音声をブーストすることができる。

【0101】

なお、上記アタックタイムおよびリリースタイムの推定値、および音量補正部190が行なった逆補正の程度を保持しておくことにより、収音対象音声に被写体が発した音声が含まれる場合に有効活用することができる。すなわち、収音対象音声に被写体が発した音声が含まれる場合であっても、音量補正部190は近傍音声からアタックタイムを検出し、収音対象音声における該アタックタイムに相当する区間の前後にわたって音量値の走査を行う。音量補正部190は、走査の結果、アタックタイムまたはリリースタイムと一致する時間幅で音量値が変化している場合はA G Cが動作したと判定し、逆補正を行なうことができる。

20

【0102】

音声混合部150は、このように音量補正部190により音量が逆補正された収音対象音声と、音源分離部140により分離された近傍音声とを、全体に占める近傍音声の音量比率が抑制されるような音量比率で混合して混合音声を生成することができる。

【0103】

以上説明したように、本発明の第3の実施形態にかかる音声再生装置12は、近傍音声の音量が過大であったために入力音声の音量が全体として抑制され、収音対象音声の音量も抑制されてしまっている場合、入力音声の音量が抑制された程度に応じて収音対象音声の音量を増大させ、収音対象音声が過小となることを防止できる。

30

【0104】

なお、本実施形態においては音量補正部190を音声再生装置12に設ける場合を説明したが、第1の実施形態で説明した音声記録装置10に設けることで、入力音声にA G Cが施されていてもA G Cの程度に応じてブーストされた収音対象音声を含む混合音声を記憶部170に記録することも可能である。

【0105】

以上、添付図面を参照しながら本発明の好適な実施形態について説明したが、本発明は係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、各種の変更例または修正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発明の技術的範囲に属するものと了解される。

40

【0106】

例えば、本明細書の音声記録装置10の処理における各ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むとしてもよい。

【0107】

また、図3には音声判定部120が音声収音部110により収音された入力音声に近傍音声が含まれるか否かを判定する例を示しているが、本発明はかかる例に限定されない。

50

例えば、音声判定部 120 は、音源分離部 140 により分離された音声が入力され、該分離された音声の音源位置を推定し、該分離された音声に近傍音声が含まれるか否かを判定し、該分離された音声を音声混合部 150 に出力してもよい。なお、この場合、音源分離部 140 は初期値無しでブラインドに音声を音源ごとに分離する。

【0108】

また、音声記録装置 10、音声再生装置 11、音声再生装置 12 に内蔵される CPU、ROM および RAM などのハードウェアを、上述した音声記録装置 10、音声再生装置 11、音声再生装置 12 の各構成と同等の機能を発揮させるためのコンピュータプログラムも作成可能である。また、該コンピュータプログラムを記憶させた記憶媒体も提供される。また、音声記録装置 10、音声再生装置 11、音声再生装置 12 の各機能ブロック図で示したそれぞれの機能ブロックをハードウェアで構成することで、一連の処理をハードウェアで実現することもできる。

【図面の簡単な説明】

【0109】

【図 1】本発明の第 1 の実施形態にかかる音声記録装置が用いられる場面の一例を示した説明図である。

【図 2】通常の音声記録方法によって記録される音声の時間領域の振幅を示した説明図である。

【図 3】同実施形態にかかる音声処理装置の一例としての音声記録装置の構成を示した機能ブロック図である。

【図 4】音声判定部の構成を示した機能ブロック図である。

【図 5】2 つの入力音声の位相差に基づいて入力音声の音源位置を推定する様子を示した説明図である。

【図 6】3 つの入力音声の位相差に基づいて入力音声の音源位置を推定する様子を示した説明図である。

【図 7】2 つの入力音声の音量に基づいて入力音声の音源位置を推定する様子を示した説明図である。

【図 8】3 つの入力音声の音量に基づいて入力音声の音源位置を推定する様子を示した説明図である。

【図 9】音声記録装置と操作者の位置関係を示した説明図である。

【図 10】同実施形態にかかる音声記録装置において実行される音声処理方法の流れを示したフローチャートである。

【図 11】本発明の第 2 の実施形態にかかる音声再生装置の構成を示した機能ブロック図である。

【図 12】本発明の第 3 の実施形態にかかる音声再生装置の構成を示した機能ブロック図である。

【図 13】A G C の適用前の音声の音量と、A G C 適用後の音声の音量を対比的に表した説明図である。

【符号の説明】

【0110】

10	音声記録装置
11、12	音声再生装置
110	音声収音部
120	音声判定部
124	音量検出器
134	音質検出器
136	距離方向推定器
138	操作者音声推定器
140	音源分離部
150	音声混合部

10

20

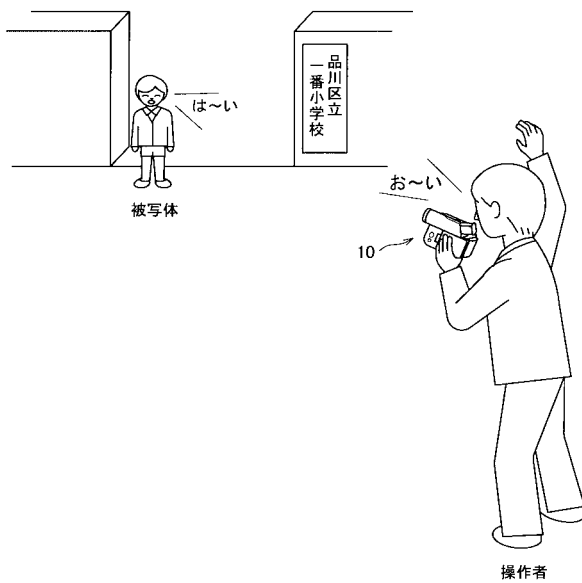
30

40

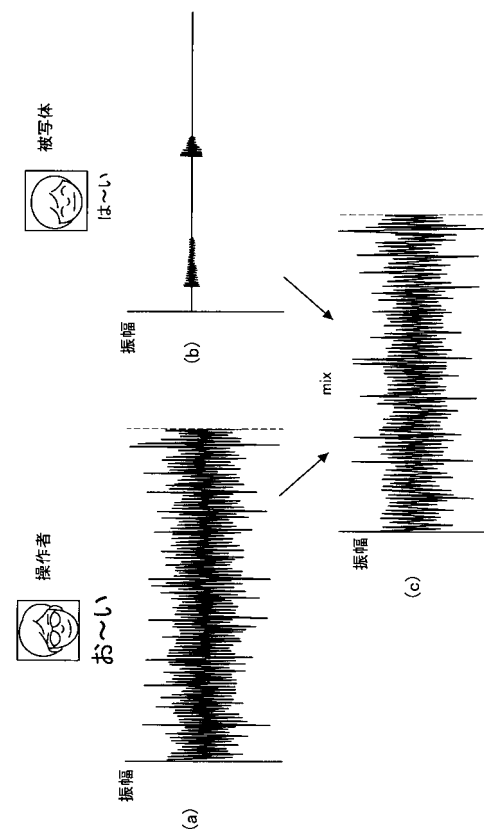
50

160 記録部
 170、172 記憶部
 174 再生部
 180 音声出力部
 190 音量補正部

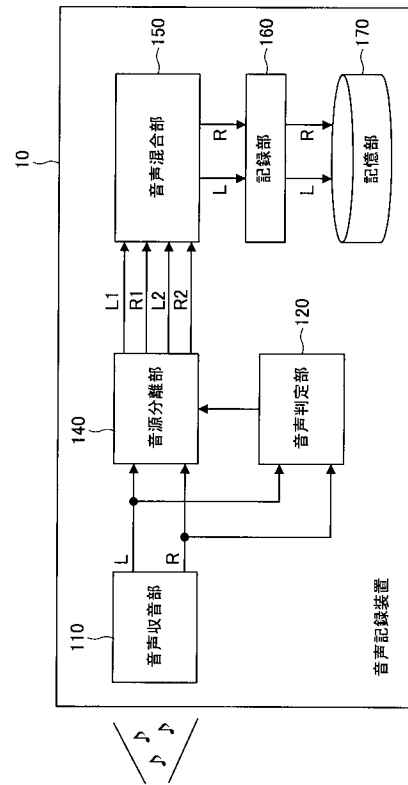
【図1】



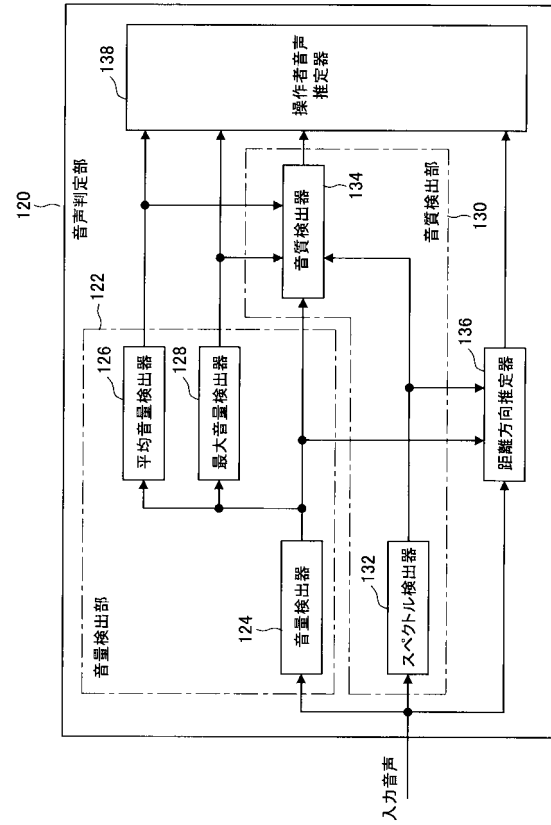
【図2】



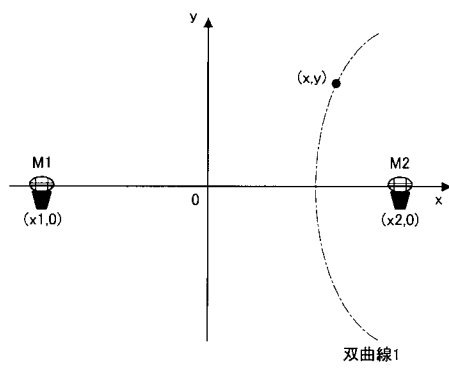
【図3】



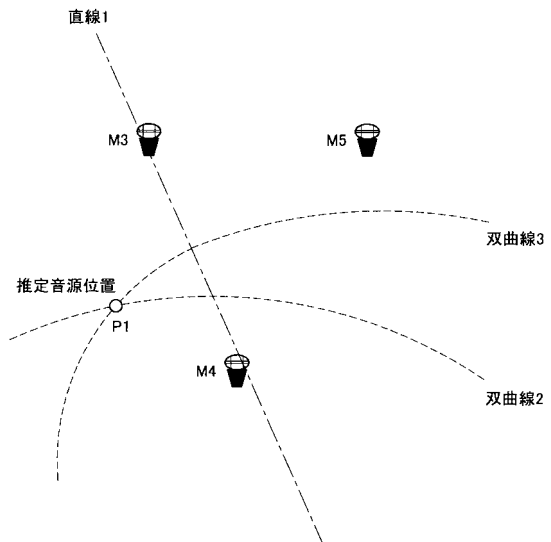
【図4】



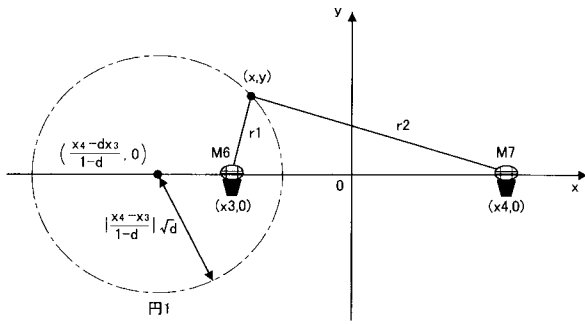
【図5】



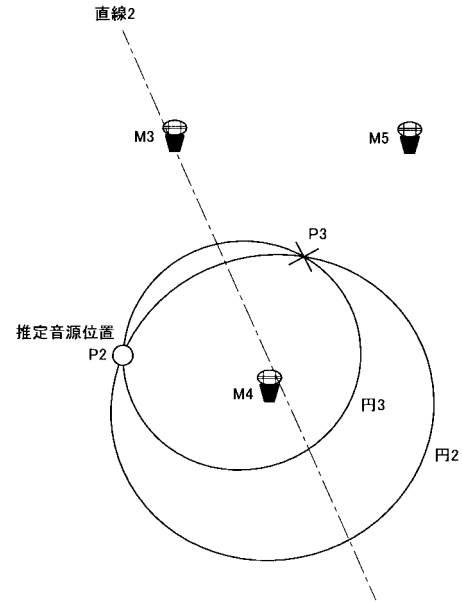
【図6】



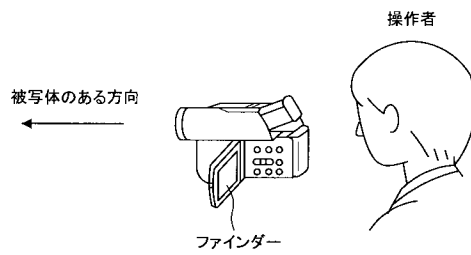
【図 7】



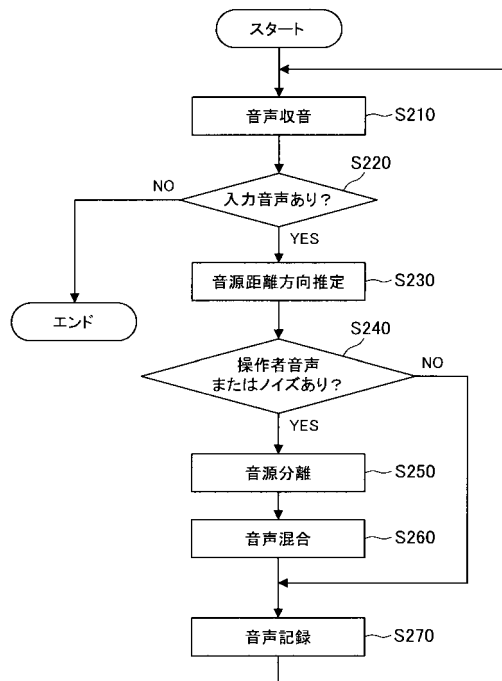
【図 8】



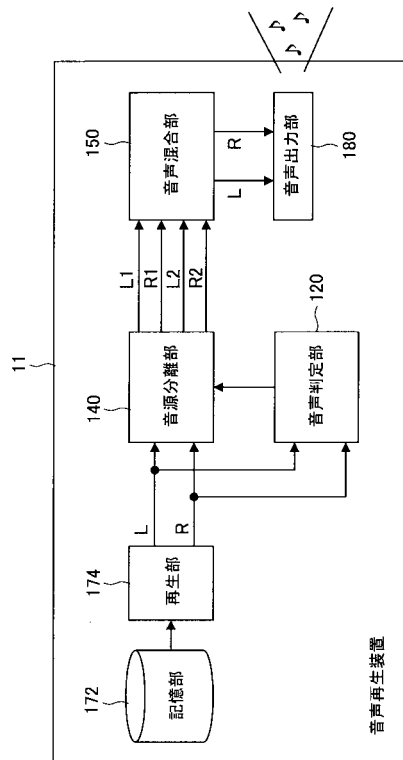
【図 9】



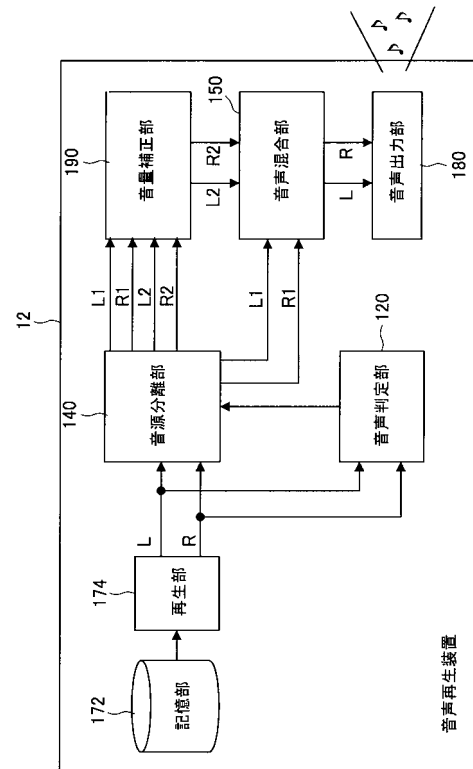
【図 10】



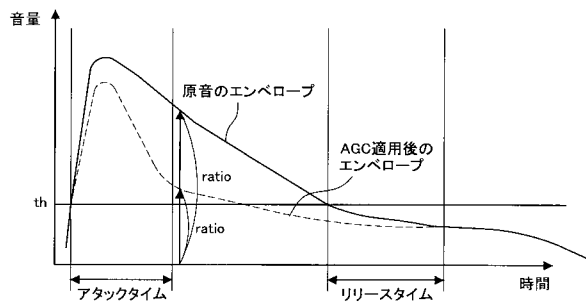
【図 1 1】



【図 1 2】



【図 1 3】



フロントページの続き

- (72)発明者 井上 晃
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 東山 恵祐
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 高橋 秀介
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 西口 正之
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 菊池 智紀

- (56)参考文献 特開平05-056007(JP,A)
特開2001-084713(JP,A)
特開2005-024736(JP,A)
特開2002-218583(JP,A)
特開2004-072306(JP,A)
実開平05-043200(JP,U)
特開平03-233439(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G10L 11/00-21/06
JSTPlus(JDreamII)