

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-24160

(P2019-24160A)

(43) 公開日 平成31年2月14日(2019.2.14)

(51) Int.Cl. F I テーマコード (参考)  
 HO4R 3/00 (2006.01) HO4R 3/00 320 5D220  
 HO4R 3/02 (2006.01) HO4R 3/02

審査請求 未請求 請求項の数 7 O L (全 20 頁)

(21) 出願番号 特願2017-142480 (P2017-142480)  
 (22) 出願日 平成29年7月24日 (2017.7.24)  
 特許法第30条第2項適用申請有り (1) 発行日  
 2017年2月15日 刊行物 「平成28年度 九州  
 大学 芸術工学部 音響設計学科 卒業研究概要集 1  
 DS13157R」 発行者名 九州大学

(71) 出願人 00004226  
 日本電信電話株式会社  
 東京都千代田区大手町一丁目5番1号  
 (71) 出願人 504145342  
 国立大学法人九州大学  
 福岡県福岡市西区元岡744  
 (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  
 (72) 発明者 鎌本 優  
 東京都千代田区大手町一丁目5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】 抽出発生音補正装置、抽出発生音補正方法、プログラム

(57) 【要約】

【課題】聴感上の違和感を抑えるように、大きな会場で発生した音を抽出した抽出発生音を補正することができる抽出発生音補正技術を提供する。

【解決手段】 収音会場音 $y(k)$ を会場で再生された音及び会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を収音会場音 $y(k)$ から会場で発生した音を抽出したサンプル番号 $k$ の音とし、収音会場音 $y(k)$ と抽出発生音 $e(k)$ から、収音会場音パワー $Y_k$ と抽出発生音パワー $E_k$ を算出するパワー算出部110と、収音会場音パワー $Y_k$ と抽出発生音パワー $E_k$ を用いて、抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正部120とを含み、抽出発生音補正部120は、補正済抽出発生音 $e'(k)$ を、比 $E_k/Y_k$ が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として生成する。

【選択図】 図4

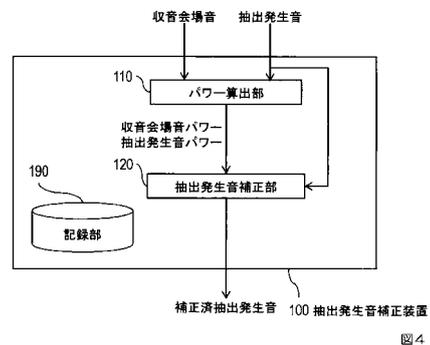


図4

## 【特許請求の範囲】

## 【請求項 1】

サンプル番号 $k$ を $k=1, 2, \dots$ 、収音会場音 $y(k)$ を会場で再生された音及び当該会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を前記収音会場音 $y(k)$ から当該会場で発生した音を抽出したサンプル番号 $k$ の音とし、

前記収音会場音 $y(k)$ と前記抽出発生音 $e(k)$ から、前記収音会場音 $y(k)$ のパワーである収音会場音パワー $Y_k$ と前記抽出発生音 $e(k)$ のパワーである抽出発生音パワー $E_k$ を算出するパワー算出部と、

前記収音会場音パワー $Y_k$ と前記抽出発生音パワー $E_k$ を用いて、前記抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正部と

を含む抽出発生音補正装置であり、

前記抽出発生音補正部は、前記補正済抽出発生音 $e'(k)$ を、前記収音会場音パワー $Y_k$ に対する前記抽出発生音パワー $E_k$ の比 $E_k/Y_k$ が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として生成する

抽出発生音補正装置。

10

## 【請求項 2】

サンプル番号 $k$ を $k=1, 2, \dots$ 、収音会場音 $y(k)$ を会場で再生された音及び当該会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を前記収音会場音 $y(k)$ から当該会場で発生した音を抽出したサンプル番号 $k$ の音とし、

前記抽出発生音 $e(k)$ から、前記抽出発生音 $e(k)$ の時間方向におけるスパースさの程度を表すスパース尺度 $S_k$ を算出するスパース尺度算出部と、

前記スパース尺度 $S_k$ を用いて、前記抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正部と

を含む抽出発生音補正装置であり、

前記抽出発生音補正部は、前記補正済抽出発生音 $e'(k)$ を、前記スパース尺度 $S_k$ が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として生成する

抽出発生音補正装置。

20

## 【請求項 3】

請求項 1 または 2 に記載の抽出発生音補正装置であって、

さらに、

前記補正済抽出発生音 $e'(k)$ から、前記補正済抽出発生音 $e'(k)$ を生成するために用いる補正済抽出発生音パラメータ $p(k)$ を生成する補正済抽出発生音パラメータ生成部と

を含むことを特徴とする抽出発生音補正装置。

30

## 【請求項 4】

請求項 1 ないし 3 のいずれか 1 項に記載の抽出発生音補正装置であって、

前記小さいことを示す所定の範囲にあることを示す値は、サンプルごとに更新されることを特徴とする抽出発生音補正装置。

## 【請求項 5】

サンプル番号 $k$ を $k=1, 2, \dots$ 、収音会場音 $y(k)$ を会場で再生された音及び当該会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を前記収音会場音 $y(k)$ から当該会場で発生した音を抽出したサンプル番号 $k$ の音とし、

抽出発生音補正装置が、前記収音会場音 $y(k)$ と前記抽出発生音 $e(k)$ から、前記収音会場音 $y(k)$ のパワーである収音会場音パワー $Y_k$ と前記抽出発生音 $e(k)$ のパワーである抽出発生音パワー $E_k$ を算出するパワー算出ステップと、

前記抽出発生音補正装置が、前記収音会場音パワー $Y_k$ と前記抽出発生音パワー $E_k$ を用いて、前記抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正ステップと

を含む抽出発生音補正方法であり、

前記抽出発生音補正ステップは、前記補正済抽出発生音 $e'(k)$ を、前記収音会場音パワ

40

50

$-Y_k$  に対する前記抽出発生音パワー  $E_k$  の比  $E_k/Y_k$  が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$  とし、それ以外の場合は、 $e'(k)=e(k)$  として生成する

抽出発生音補正方法。

【請求項 6】

サンプル番号  $k$  を  $k=1, 2, \dots$ 、収音会場音  $y(k)$  を会場で再生された音及び当該会場で発生した音を収音したサンプル番号  $k$  の音、抽出発生音  $e(k)$  を前記収音会場音  $y(k)$  から当該会場で発生した音を抽出したサンプル番号  $k$  の音とし、

抽出発生音補正装置が、前記抽出発生音  $e(k)$  から、前記抽出発生音  $e(k)$  の時間方向におけるスパースの程度を表すスパース尺度  $S_k$  を算出するスパース尺度算出ステップと、

前記抽出発生音補正装置が、前記スパース尺度  $S_k$  を用いて、前記抽出発生音  $e(k)$  から補正済抽出発生音  $e'(k)$  を生成する抽出発生音補正ステップと

を含む抽出発生音補正方法であり、

前記抽出発生音補正ステップは、前記補正済抽出発生音  $e'(k)$  を、前記スパース尺度  $S_k$  が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$  とし、それ以外の場合は、 $e'(k)=e(k)$  として生成する

抽出発生音補正方法。

【請求項 7】

請求項 1 ないし 4 のいずれか 1 項に記載の抽出発生音補正装置としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音場空間で発生した音を抽出する技術に関する。

【背景技術】

【0002】

会場で発生した音を抽出するために、エコーキャンセラ技術が用いられる（非特許文献 1）。以下、図 1 を参照して、エコーキャンセラ技術について説明する。

【0003】

まず、会場 A で発生した音  $x(k)$  ( $k=1, 2, \dots$ 、以下、 $k$  をサンプル番号という) が会場 B に伝送される。会場 B に伝送された入力音  $x(k)$  を、例えば、 $K$  個のサンプルで構成されるフレームごとに処理する場合、 $K$  はフレーム長となる。また、入力音  $x(k)$  は、例えば、電話の音声、会場 A に設置されたコンテンツ配信システムの音である。

【0004】

次に、入力音  $x(k)$  が会場 B に設置されたスピーカ 910 から再生され、（スピーカ 910 からマイク 920 までの）会場 B の音響エコー  $h_0(k)$  が畳み込まれた畳み込み入力音  $x(k) * h_0(k)$  として会場 B に設置されたマイク 920 により収音される。ここで、 $*$  は畳み込み演算子を表す。また、音響エコーのことを伝達関数やインパルス応答ということもあるが、以下では、音響エコーまたは伝達関数ということにする。

【0005】

マイク 920 により収音される音は、畳み込み入力音  $x(k) * h_0(k)$  だけではない。会場 B で発生する環境音（例えば、音声や拍手音） $c_1(k)$  もその音源からマイク 920 までの音響エコー  $h_1(k)$  が畳み込まれた畳み込み環境音  $c_1(k) * h_1(k)$  として収音される。また、マイク 920 にはノイズ  $n(k)$  も混入する。したがって、マイク 920 により収音される音  $y(k)$  （以下、収音会場音  $y(k)$  という。）は  $y(k) = x(k) * h_0(k) + c_1(k) * h_1(k) + n(k)$  と表すことができる。

【0006】

このまま収音会場音  $y(k)$  を会場 A に伝送し、会場 A で再生すると、ハウリングやダブルトークを生じてしまい、相手方にとって聞きづらい音になってしまう。そこで、エコーキャンセラ 930 は、畳み込み入力音  $x(k) * h_0(k)$  を近似する  $x'(k) = x(k) * h_0'(k)$  を推定し、会場 B で発生した音を抽出した抽出発生音  $e(k) = y(k) - x'(k) = c_1(k) * h_1(k) + n(k) + d(k)$  を

10

20

30

40

50

求め、この抽出発生音 $e(k)$ を会場 A に伝送する。

【0007】

しかし、推定された伝達関数 $h_0'(k)$ と真の伝達関数 $h_0(k)$ は異なるため、残留信号 $d(k)$ が生じてしまう。この問題を解決するためにNLMS(Normalized Least-Mean-Squares)法などの適応アルゴリズム(適応フィルタ)が使われる。NLMS法を用いると、伝達関数 $h_0'(k)$ を伝達関数 $h_0(k)$ に近づけることができる。特に、会場 B が小さい(つまり、スピーカとマイクの距離が比較的近くなる)場合は伝達関数 $h_0'(k)$ を精度よく推定できるため、残留信号 $d(k)$ の振幅値は十分小さくなり、聴感上の違和感を小さくすることができる。

【先行技術文献】

【非特許文献】

10

【0008】

【非特許文献1】北脇信彦, “デジタル音声・オーディオ技術”, 株式会社オーム社, pp.223-225.

【発明の概要】

【発明が解決しようとする課題】

【0009】

上記エコーキャンセラ技術を用いることによって、例えば、スピーカホンの用途のようにスピーカとマイクの位置が近い場合には、会場 B で発生した音を適切に抽出することができる。

【0010】

20

しかし、スピーカとマイクの位置が離れ、あえて残響を加えるように設計されている環境(例えば、コンサート会場や演劇会場のようなライブビューイングにおけるライブ会場(会場 A)とは別のビューイング会場(会場 B))では、伝達関数 $h_0'(k)$ を伝達関数 $h_0(k)$ に十分に近づけることができない。そのため、残留信号 $d(k)$ の振幅値が大きくなってしまい、会場 B で発生した音だけを抽出することができないことになる。

【0011】

スピーカホンによる音声通話などでエコーキャンセラを用いる場合には、会議室(会場 B)での音声を含む環境音 $c_1(k)$ が相手側に届かないと困るため、残留信号 $d(k)$ が多少残っていたとしても、抽出発生音 $e(k)$ をそのまま伝送する方が好ましい。

【0012】

30

しかし、ライブビューイングの場合には、ライブ会場(会場 A)での聴感を大きく損ねることになるため、抽出発生音 $e(k)$ をそのまま伝送するのは好ましくない。特に、ビューイング会場(会場 B)で発生した拍手音や手拍子音などの環境音 $c_1(k)$ そのものを伝送する代わりに、伝送先であるライブ会場(会場 A)で環境音 $c_1(k)$ を生成するために用いるパラメータを伝送する場合、ビューイング会場(会場 B)で発生した環境音 $c_1(k)$ を誤抽出してしまうと、伝送先で誤った形で拍手音や手拍子音(つまり、拍手音や手拍子音以外の音も含んだ音)を生成してしまい、もともとのコンテンツを楽しんでいるライブ会場(会場 A)の聴衆に迷惑がかかってしまうことになる。

【0013】

図2は、このような状況を示すものであり、残留信号 $d(k)$ が大きく残った(入力音 $x(k)$ と収音会場音 $y(k)$ の差異が大きい)場合における、入力音 $x(k)$ 、収音会場音 $y(k)$ 、抽出発生音 $e(k)$ の様子を示す図である。図2の網掛け部は、通常のエコーキャンセラを用いた場合に、残留信号 $d(k)$ が大きく、スピーカから再生された音 $x(k)$ が抽出発生音 $e(k)$ に残っており、会場 B の環境音 $c_1(k)$ が正しく抽出されていないこと示している。

40

【0014】

つまり、通常のエコーキャンセラでは、ビューイング会場(会場 B)のようにスピーカとマイクが近くにないような大きな会場で発生する音を、ライブ会場(会場 A)で再生する場合の聴感上の違和感を抑えた形で抽出することができないという問題がある。

【0015】

そこで本発明では、聴感上の違和感を抑えるように、大きな会場で発生した音を抽出し

50

た抽出発生音を補正することができる抽出発生音補正技術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】

【0016】

本発明の一態様は、サンプル番号 $k$ を $k=1,2,\dots$ 、収音会場音 $y(k)$ を会場で再生された音及び当該会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を前記収音会場音 $y(k)$ から当該会場で発生した音を抽出したサンプル番号 $k$ の音とし、前記収音会場音 $y(k)$ と前記抽出発生音 $e(k)$ から、前記収音会場音 $y(k)$ のパワーである収音会場音パワー $Y_k$ と前記抽出発生音 $e(k)$ のパワーである抽出発生音パワー $E_k$ を算出するパワー算出部と、前記収音会場音パワー $Y_k$ と前記抽出発生音パワー $E_k$ を用いて、前記抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正部とを含み、前記抽出発生音補正部は、前記補正済抽出発生音 $e'(k)$ を、前記収音会場音パワー $Y_k$ に対する前記抽出発生音パワー $E_k$ の比 $E_k/Y_k$ が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として生成する。

10

【0017】

本発明の一態様は、サンプル番号 $k$ を $k=1,2,\dots$ 、収音会場音 $y(k)$ を会場で再生された音及び当該会場で発生した音を収音したサンプル番号 $k$ の音、抽出発生音 $e(k)$ を前記収音会場音 $y(k)$ から当該会場で発生した音を抽出したサンプル番号 $k$ の音とし、前記抽出発生音 $e(k)$ から、前記抽出発生音 $e(k)$ の時間方向におけるスパースさの程度を表すスパース尺度 $S_k$ を算出するスパース尺度算出部と、前記スパース尺度 $S_k$ を用いて、前記抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する抽出発生音補正部とを含み、前記抽出発生音補正部は、前記補正済抽出発生音 $e'(k)$ を、前記スパース尺度 $S_k$ が小さいことを示す所定の範囲にある場合は、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として生成する。

20

【発明の効果】

【0018】

本発明によれば、聴感上の違和感を抑えるように、大きな会場で発生した音を抽出した抽出発生音を補正することができる。

【図面の簡単な説明】

【0019】

【図1】エコーキャンセラ技術を説明するための図。

【図2】残留信号 $d(k)$ が大きく残った場合における、入力音 $x(k)$ 、収音会場音 $y(k)$ 、抽出発生音 $e(k)$ の様子の一例を示す図。

30

【図3】抽出発生音補正装置100を含む抽出発生音補正システム10の構成の一例を示すブロック図。

【図4】抽出発生音補正装置100の構成の一例を示すブロック図。

【図5】抽出発生音補正装置100の動作の一例を示すフローチャート。

【図6】抽出発生音補正装置100の入力である抽出発生音 $e(k)$ と出力である補正済抽出発生音 $e'(k)$ の様子の一例を示す図。

【図7】抽出発生音補正装置101の構成の一例を示すブロック図。

【図8】抽出発生音補正装置101の動作の一例を示すフローチャート。

【図9】環境音 $c1(k)$ が拍手音である場合の振幅の様子の一例を示す図。

40

【図10】抽出発生音補正装置200の構成の一例を示すブロック図。

【図11】抽出発生音補正装置200の動作の一例を示すフローチャート。

【図12】抽出発生音補正装置201の構成の一例を示すブロック図。

【図13】抽出発生音補正装置201の動作の一例を示すフローチャート。

【図14】抽出発生音補正装置300の構成の一例を示すブロック図。

【図15】抽出発生音補正装置300の動作の一例を示すフローチャート。

【図16】環境音 $c1(k)$ が拍手音である場合の補正済抽出発生音 $e'(k)$ の様子の一例を示す図。

【図17】抽出発生音補正装置301の構成の一例を示すブロック図。

【図18】抽出発生音補正装置301の動作の一例を示すフローチャート。

50

## 【発明を実施するための形態】

## 【0020】

以下、本発明の実施の形態について、詳細に説明する。なお、同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。

## 【0021】

ここまでの説明では、会場Bに伝送されてくる入力音 $x(k)$ をモノラル信号、会場Bで発生する環境音は $c1(k)$ の1つであるとして説明してきたが、入力音 $x(k)$ はステレオ信号であってもよいし、会場Bで発生する環境音は複数あってもよい。

## 【0022】

例えば、入力音がモノラル信号 $x(k)$ であり、会場Bで発生する環境音が $c1(k), \dots, cM(k)$ のM個(Mは1以上の整数)であった場合、入力音 $x(k)$ 、環境音 $c1(k), \dots, cM(k)$ の音響エコーをそれぞれ $h0(k), h1(k), \dots, hM(k)$ 、ノイズを $n(k)$ とすると、マイクにより收音される收音会場音 $y(k)$ は $y(k)=x(k)*h0(k)+c1(k)*h1(k)+\dots+cM(k)*hM(k)+n(k)$ と表すことができる。さらに、伝達関数 $h0(k)$ を推定した伝達関数を $h0'(k)$ 、残留信号を $d(k)$ とすると、抽出発生音 $e(k)$ は $e(k)=y(k)-x(k)*h0'(k)=c1(k)*h1(k)+\dots+cM(k)*hM(k)+n(k)+d(k)$ と表すことができる。

## 【0023】

また、入力音がステレオ信号 $xR(k), xL(k)$ であり、会場Bで発生する環境音が $c1(k), \dots, cM(k)$ のM個(Mは1以上の整数)であった場合、入力音 $xR(k), xL(k)$ 、環境音 $c1(k), \dots, cM(k)$ の音響エコーをそれぞれ $hR0(k), hL0(k), h1(k), \dots, hM(k)$ 、ノイズを $n(k)$ とすると、マイクにより收音される收音会場音 $y(k)$ は $y(k)=xR(k)*hR0(k)+xL(k)*hL0(k)+c1(k)*h1(k)+\dots+cM(k)*hM(k)+n(k)$ と表すことができる。さらに、伝達関数 $hR0(k), hL0(k)$ を推定した伝達関数をそれぞれ $hR0'(k), hL0'(k)$ 、残留信号を $d(k)$ とすると、抽出発生音 $e(k)$ は $e(k)=y(k)-\{xR(k)*hR0'(k)+xL(k)*hL0'(k)\}=c1(k)*h1(k)+\dots+cM(k)*hM(k)+n(k)+d(k)$ と表すことができる。

## 【0024】

3つの例からわかるように、收音会場音 $y(k)$ は、畳み込み入力音と畳み込み環境音とノイズの和として表現することができる。また、抽出発生音 $e(k)$ は、畳み込み環境音とノイズと残響信号の和として表現することができる。

## 【0025】

音響環境としては、モノラルやステレオ以外に、サラウンド環境、22.2ch環境など様々なものが存在するが、收音会場音 $y(k)$ 、抽出発生音 $e(k)$ は上記3つの例と同様に扱うことができる。

## 【0026】

そこで、以下説明する各実施形態では、入力音がモノラル信号であり、環境音が1つである場合を用いて説明することにする。

## 【0027】

まず、各実施形態で用いる用語について説明する。

## 【0028】

サンプル番号 $k$ を $k=1, 2, \dots$ とする。入力音 $x(k)$ をK個のサンプルで構成されるフレームごとに処理する場合、i番目のフレーム $Xi$ は、 $Xi=[x(1) \ x(2) \ \dots \ x(K)]$ というベクトルで表現することができる。ここで、 $i(i=1, 2, \dots)$ はフレーム番号を表す。

## 【0029】

以下、説明を簡素化するために、 $Xi=x(k)$ 、つまり、各フレームは1個のサンプルで構成されるものとして、各実施形態を説明するが、K個のサンプルで構成されとした場合と内容は変わらない。

## 【0030】

收音会場音 $y(k)$ を会場で再生された音及び会場で発生した音を收音したサンプル番号 $k$ の音とする。会場で再生された音とは、別の会場から伝送されてきて再生された音のことであり、先ほどの例で言えば、ライブ会場の音である。また、会場で発生した音とは、先

10

20

30

40

50

ほどの例で言えば、ビューイング会場で発生した拍手音などの環境音やノイズのことである。

【 0 0 3 1 】

抽出発生音 $e(k)$ を収音会場音 $y(k)$ から会場で発生した音を抽出したサンプル番号 $k$ の音とする。

【 0 0 3 2 】

< 第一実施形態 >

以下、図3を参照して抽出発生音補正システム10について説明する。図3は、抽出発生音補正システム10の構成を示すブロック図である。図3に示すように抽出発生音補正システム10は、スピーカ910、マイク920、エコーキャンセラ930、抽出発生音補正装置100を含む。

10

【 0 0 3 3 】

スピーカ910、マイク920、エコーキャンセラ930は、図1のそれと同一の機能を有するものである。具体的には、スピーカ910は、伝送されてきた会場Aの音 $x(k)$ を会場Bにて再生する装置である。マイク920は、スピーカ910から再生された入力音 $x(k)$ を畳み込んだ畳み込み入力音 $x(k)*h_0(k)$ と会場Bで発生した環境音 $c_1(k)$ を畳み込んだ畳み込み環境音 $c_1(k)*h_1(k)$ とノイズ $n(k)$ を収音する装置である。エコーキャンセラ930は、入力音 $x(k)$ とマイク920により収音された収音会場音 $y(k) (=x(k)*h_0(k)+c_1(k)*h_1(k)+n(k))$ から抽出発生音 $e(k)$ を生成する。抽出発生音 $e(k)=c_1(k)*h_1(k)+n(k)+d(k)$ は、会場Bで発生した音をエコーキャンセラ930が抽出した音である。ここで、 $d(k)$ は

20

【 0 0 3 4 】

なお、スピーカ910からマイク920への伝達遅延を考慮し、入力音 $x(k)$ の代わりに、入力音 $x(k)$ に所定の遅延を加えてからエコーキャンセラ930に入力するようにしてもよい。また、スピーカ910からマイク920への伝達遅延を考慮し、入力音 $x(k)$ の代わりに、環境音やノイズが十分小さいときに測定した伝達関数を入力音 $x(k)$ に畳み込んだ畳み込み入力音をエコーキャンセラ930に入力するようにしてもよい。

【 0 0 3 5 】

抽出発生音補正装置100は、収音会場音 $y(k)$ と抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する。ここで、補正済抽出発生音 $e'(k)$ は、収音会場音 $y(k)$ と抽出発生音 $e(k)$ を用いて抽出発生音 $e(k)$ を補正したサンプル番号 $k$ の音である。一般に、補正済抽出発生音 $e'(k)$ は、残留信号 $d(k)$ の振幅値を聴感上の違和感がなくなるほど十分小さくしたもののとなる。

30

【 0 0 3 6 】

以下、図4～図5を参照して抽出発生音補正装置100について説明する。図4は、抽出発生音補正装置100の構成を示すブロック図である。図5は、抽出発生音補正装置100の動作を示すフローチャートである。図4に示すように抽出発生音補正装置100は、パワー算出部110、抽出発生音補正部120、記録部190を含む。記録部190は、抽出発生音補正装置100の処理に必要な情報を適宜記録する構成部である。

【 0 0 3 7 】

図5に従い抽出発生音補正装置100の動作について説明する。パワー算出部110は、収音会場音 $y(k)$ と抽出発生音 $e(k)$ から、収音会場音 $y(k)$ のパワーである収音会場音パワー $Y_k$ と抽出発生音 $e(k)$ のパワーである抽出発生音パワー $E_k$ を算出する(S110)。例えば、収音会場音パワー $Y_k$ を、収音会場音 $y(k)$ を含む一定時間(Kサンプル分(ただし、Kは1以上の整数))の収音会場音のパワーとして算出する。サンプル $k$ 以前のK個のサンプルを用いる場合は次式のようなになる。

40

【 0 0 3 8 】

【数 1】

$$Y_k = \sum_{i=k-K+1}^k y(i) \times y(i) \quad \dots(1)$$

【0039】

また、抽出発生音パワー $E_k$ を、抽出発生音 $e(k)$ を含む一定時間(Kサンプル分)の抽出発生音のパワーとして算出する。サンプル $k$ 以前の $K$ 個のサンプルを用いる場合は次式のようになる。

【0040】

【数 2】

$$E_k = \sum_{i=k-K+1}^k e(i) \times e(i) \quad \dots(2)$$

【0041】

また、以下のように、上記の収音会場音パワー $Y_k$ 、抽出発生音パワー $E_k$ をそれぞれサンプル数 $K$ で割った値を収音会場音パワー $Y_k$ 、抽出発生音パワー $E_k$ としてもよい。

【0042】

【数 3】

$$Y_k = \left( \sum_{i=k-K+1}^k y(i) \times y(i) \right) / K \quad \dots(3)$$

$$E_k = \left( \sum_{i=k-K+1}^k e(i) \times e(i) \right) / K \quad \dots(4)$$

【0043】

抽出発生音補正部120は、収音会場音パワー $Y_k$ と抽出発生音パワー $E_k$ を用いて、抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する(S120)。例えば、 $E_k > Y_k$ の場合、 $e'(k) = e(k)$ とし、 $E_k < Y_k$ の場合、 $e'(k) = 0$ として補正済抽出発生音を生成する。 $E_k > Y_k$ の場合に $e'(k) = e(k)$ とする理由は、環境音 $c1(k)$ が入力音 $x(k)$ に比べて十分大きく、残留信号 $d(k)$ が大きいと考えられるためである。また、 $E_k < Y_k$ の場合、 $e'(k) = 0$ とする理由は、環境音 $c1(k)$ がスピーカ910から出力された入力音 $x(k)$ に比べて十分小さく、環境音 $c1(k)$ は埋もれていると考えられるためである。なお、後述する(変形例2)との関係でいえば、 $E_k < Y_k$ の場合、 $e(k)$ をパラメータ生成に用いても良好な結果を得られないと考えられるため、 $e'(k) = 0$ とする。

【0044】

図6は、抽出発生音補正装置100の入力である抽出発生音 $e(k)$ と出力である補正済抽出発生音 $e'(k)$ の様子を示す図である。図6を見ると、スピーカ910から再生された入力音 $x(k)$ の影響を減じ、環境音 $c1(k)$ が抽出できていることがわかる。

【0045】

なお、 $\alpha$ をあらかじめ定められた正の定数(以下、 $\alpha$ のことを倍率という)として、 $E_k > \alpha Y_k$ の場合、 $e'(k) = e(k)$ とし、 $E_k < \alpha Y_k$ の場合、 $e'(k) = 0$ として補正済抽出発生音を生成するようにしてもよい。

【0046】

もちろん、 $E_k > Y_k$ 、 $E_k < Y_k$ の代わりに、 $E_k > \alpha Y_k$ 、 $E_k < \alpha Y_k$ を、 $E_k > Y_k$ 、 $E_k < Y_k$ の代わりに、 $E_k > \alpha Y_k$ 、 $E_k < \alpha Y_k$ を用いてもよい。

【0047】

以上まとめると、 $\alpha = 1$ の場合も含めて、 $E_k / Y_k > \alpha$  または  $E_k / Y_k < \alpha$  となる場合、 $e'(k) = 0$ とし、それ以外の場合は、 $e'(k) = e(k)$ として補正済抽出発生音 $e'(k)$ を生成する。あらかじめ定められた正の定数 $\alpha$ に対して $E_k / Y_k > \alpha$  または  $E_k / Y_k < \alpha$  となることを、収音会場

10

20

30

40

50

音パワー $Y_k$ に対する抽出発生音パワー $E_k$ の比 $E_k/Y_k$ が小さいことを示す所定の範囲にあるという。このとき、 $E_k/Y_k$ が小さいことを示す所定の範囲にあることを示す値という。

【0048】

(変形例1)

抽出発生音補正部120における収音会場音パワー $Y_k$ と抽出発生音パワー $E_k$ の比較では、比較条件に用いる倍率 $\alpha_k$ を時間経過によらず不変であるものとして扱ったが、一定のタイミングで倍率 $\alpha_k$ を更新するようにしてもよい。一定のタイミングで更新するようにすると、会場での人数の変化や会場の温度の変化などに対して、倍率 $\alpha_k$ が追従することができるようになる。

【0049】

例えば、 $\alpha_0$  (ただし、 $\alpha_0$ はあらかじめ定められた正の定数)を倍率の初期値として、サンプルごとに倍率 $\alpha_k$ を更新して比較するようにしてもよい。具体的には、 $k=1, \dots$ として、 $E_k > \alpha_{k-1} Y_k$ の場合、 $e'(k)=e(k)$ とし、 $E_k \leq \alpha_{k-1} Y_k$ の場合、 $e'(k)=0$ として補正済抽出発生音 $e'(k)$ を生成する。次に、次のサンプルのための倍率として $\alpha_k = E_k/Y_k$ を求め、同様に $E_{k+1} > \alpha_k Y_{k+1}$ の場合、 $e'(k+1)=e(k+1)$ とし、 $E_{k+1} \leq \alpha_k Y_{k+1}$ の場合、 $e'(k+1)=0$ として補正済抽出発生音 $e'(k+1)$ を生成する。

【0050】

なお、倍率 $\alpha_k$ を $\alpha_k = (E_{k-L+1}/Y_{k-L+1} + \dots + E_k/Y_k)/L$ というようにLサンプル分の平均値としてもよい。さらに、単純な平均とする代わりに、倍率 $\alpha_k$ の計算に忘却係数を定めたフィルタを用いてもよいし、線形関数や非線形関数を用いてもかまわない。例えば、 $\alpha_k$ と $\alpha_{k-1}$ を実数の定数(ただし、 $\alpha_k > 0$ )として、ガウス窓を用いて定義される忘却係数 $W(i)$  ( $i=0, 1, \dots$ )により、倍率 $\alpha_k$ を以下のように求めてもよい。

【0051】

【数4】

$$W(i) = \omega \cdot \exp\left(-\frac{i^2}{\sigma^2}\right)$$

$$\alpha_k = \left( W(0) \cdot \frac{E_k}{Y_k} + W(1) \cdot \frac{E_{k-1}}{Y_{k-1}} + \dots + W(L-1) \cdot \frac{E_{k-L+1}}{Y_{k-L+1}} \right) / L$$

【0052】

(変形例2)

抽出発生音補正装置が、補正済抽出発生音 $e'(k)$ の代わりに、補正済抽出発生音 $e'(k)$ を生成するために用いるパラメータである補正済抽出発生音パラメータ $p(k)$ を生成するようにしてもよい。

【0053】

以下、図7～図8を参照して抽出発生音補正装置101について説明する。図7は、抽出発生音補正装置101の構成を示すブロック図である。図8は、抽出発生音補正装置101の動作を示すフローチャートである。図7からわかるように、抽出発生音補正装置101は、補正済抽出発生音パラメータ生成部130をさらに含む点においてのみ抽出発生音補正装置100と異なる。また、図8からわかるように、抽出発生音補正装置101の動作は、S130が追加されている点においてのみ抽出発生音補正装置100の動作と異なる。補正済抽出発生音パラメータ生成部130は、補正済抽出発生音 $e'(k)$ から補正済抽出発生音パラメータ $p(k)$ を生成する(S130)。補正済抽出発生音 $e'(k)$ を生成するために用いることのできるパラメータであれば、補正済抽出発生音パラメータはどのようなものであってもよい。例えば、補正済抽出発生音 $e'(k)$ を含む一定時間の補正済抽出発生音のパワー $E'_k$ を用いて、補正済抽出発生音パラメータ $p(k)$ を生成することができる(式(2)参照)。

【0054】

10

20

30

40

50

【数5】

$$E'_k = \sum_{i=k-K+1}^k e'(i) \times e'(i)$$

【0055】

具体的には、パワー $E'_k$ の取り得る値の範囲をあらかじめ定められた場合の数に量子化して得られるインデックスを補正済抽出発生音パラメータ $p(k)$ とする。

【0056】

なお、後述する式(8)や式(6)のように、一定時間での絶対値和や任意の累乗和などを上記パワー $E'_k$ の代わりに用いて、同様の手順により、補正済抽出発生音パラメータ $p(k)$ を生成するようにしてもよい。さらに、それらの対数を取ったもの( $\log_{i=k-K+1}^k e(i) \times e(i)$ ,  $\log_{i=k-K+1}^k |e(i)|$ ,  $\log_{i=k-K+1}^k e(i) \times e(i) \times e(i) \times e(i)$ )を用いて、同様の手順により、補正済抽出発生音パラメータ $p(k)$ を生成するようにしてもよい。

【0057】

また、 $p(k)$ のサンプリング周波数が伝送フォーマットのサンプリング周波数と異なることもある。この場合、伝送フォーマットのサンプリング周波数を考慮して、 $p(k)$ のサンプリング周波数を変更した新たなパラメータ $p'(j)$ ( $j=1,2,\dots$ 、ただし、 $j$ はサンプル番号)を生成するようにしてもよい。例えば、サンプリング周波数が48kHzである $p(k)$ を、伝送フォーマットのサンプリング周波数8kHzでサンプリングしたパラメータ $p'(k)$ に変更したいときは、次式のように平均値を求めて変更するようにしてもよいし、音響信号で使われるようなりサンブラを用いて変更するようにしてもよい。

【0058】

【数6】

$$p'(j) = (p(k) + p(k+1) + \dots + p(k+5)) / 6$$

$$p'(j+1) = (p(k+6) + p(k+7) + \dots + p(k+11)) / 6$$

...

$$p'(j+i) = (p(k+i \times 6) + p(k+1+i \times 6) + \dots + p(k+5+i \times 6)) / 6$$

【0059】

さらに、 $p(k)$ や $p'(j)$ の量子化精度と伝送フォーマットの量子化精度が異なる場合には、量子化精度を変換したパラメータ $p^\wedge(k)$ や $p'^\wedge(j)$ を生成するようにしてもよい。例えば $p(k)$ の量子化精度が16ビット、 $p^\wedge(k)$ の量子化精度が8ビットである場合には、単純に $p(k)$ を8ビット右にシフト、つまり、 $p^\wedge(k) = p(k) \gg 8$ と演算して量子化精度を変換してもよい(8ビット右にシフトする代わりに、256で割るようにしてもよい)。もしくは、ITU-T G.711のA-law/ $\mu$ -law変換のように、16ビット精度を一度14ビット精度に落としてから、対数関数で変換して8ビット精度に変換するようにしてもよい。なお、以上の処理は、ルックアップテーブルを用いて、 $p^\wedge(k) = \text{LUT}(p(k))$ や $p'^\wedge(j) = \text{LUT}(p'(j))$ のように変換してもよい。

【0060】

本発明によれば、聴感上の違和感を抑えるように、大きな会場で発生した音を抽出した抽出発生音を補正することができる。これにより、スピーカとマイクが近しくなく、あえて残響を加えるように設計されている会場で発生した音も適切に抽出・補正することができるようになる。

【0061】

<第二実施形態>

環境音 $c1(k)$ が例えば拍手音である場合、図9の網掛け部に示すように、環境音 $c1(k)$ は時間方向にスパースな信号となる。そこで、抽出発生音 $e(k)$ の時間方向におけるスパースさの程度を調べ、スパースであると判定された場合、環境音 $c1(k)$ が入力音 $x(k)$ に比べて十分に大きいと判断して、 $e'(k) = e(k)$ とし、スパースでないと判定された場合、環境音 $c$

1(k) が入力音 $x(k)$ に比べて十分に小さく、環境音 $c1(k)$ が抽出できないとして、 $e'(k)=0$ として補正済抽出発生音を生成するようにする。

【0062】

以下、図10～図11を参照して抽出発生音補正装置200について説明する。図10は、抽出発生音補正装置200の構成を示すブロック図である。図11は、抽出発生音補正装置200の動作を示すフローチャートである。図10に示すように抽出発生音補正装置200は、スパース尺度算出部210、抽出発生音補正部220、記録部190を含む。

【0063】

図10に従い抽出発生音補正装置200の動作について説明する。スパース尺度算出部210は、抽出発生音 $e(k)$ から、抽出発生音 $e(k)$ のスパース尺度 $S_k$ を算出する( $S_{210}$ )。スパース尺度とは、時間方向におけるスパースさの程度を表す値である。例えば、サンプル $k$ 以前の $K$ 個のサンプルの抽出発生音を用いて、スパース尺度 $S_k$ を次式により算出してもよい。

【0064】

【数7】

$$E2_k = \sum_{i=k-K+1}^k e(i) \times e(i) \quad \dots(5)$$

$$E4_k = \sum_{i=k-K+1}^k e(i) \times e(i) \times e(i) \times e(i) \quad \dots(6)$$

$$S_k = E4_k / E2_k \quad \dots(7)$$

【0065】

なお、 $E2_k, E4_k$ のことを尖度といい、スパース尺度 $S_k$ のことを尖度係数ということもある。また、スパース尺度 $S_k$ を次式により算出してもよい。

【0066】

【数8】

$$E1_k = \sum_{i=k-K+1}^k |e(i)| \quad \dots(8)$$

$$E2_k = \sum_{i=k-K+1}^k e(i) \times e(i) \quad \dots(9)$$

$$S_k = E2_k / E1_k \quad \dots(10)$$

【0067】

さらに、別の例として、振幅絶対値の最大値と振幅絶対値の平均値から算出される比をスパース尺度 $S_k$ として用いてもよい。

【0068】

【数9】

$$E0_k = \max_{i=k-K+1, \dots, k} |e(i)| \quad \dots(11)$$

$$E1_k = \sum_{i=k-K+1}^k |e(i)| \quad \dots(12)$$

$$S_k = E0_k / (E1_k / K) \quad \dots(13)$$

【0069】

10

20

30

40

50

なお、振幅絶対値の最大値の代わりに、上位P個（ただし、 $P < K$ ）の平均値を $E0_k$ としてもよい。つまり、 $j=1, \dots, P$ に対して、 $f_j$ を $|e(i)|$ （ $i=k-K+1, \dots, k$ ）の中でj番目に大きい値として、 $E0_k=(f_1+f_2+\dots+f_p)/P$ としてもよい。

【0070】

抽出発生音補正部220は、スパース尺度 $S_k$ を用いて、抽出発生音 $e(k)$ から補正済抽出発生音 $e'(k)$ を生成する（S220）。例えば、 $\theta$ をあらかじめ定められた定数（以下、 $\theta$ のことを閾値ともいう）として、 $S_k > \theta$ の場合、 $e'(k)=e(k)$ とし、 $S_k \leq \theta$ の場合、 $e'(k)=0$ として補正済抽出発生音を生成する。

【0071】

もちろん、 $S_k > \theta$ 、 $S_k \leq \theta$ の代わりに、 $S_k > \theta$ 、 $S_k \leq \theta$ を用いてもよい。

10

【0072】

以上まとめると、 $S_k > \theta$  または $S_k \leq \theta$ となる場合、 $e'(k)=0$ とし、それ以外の場合は、 $e'(k)=e(k)$ として補正済抽出発生音 $e'(k)$ を生成する。あらかじめ定められた定数 $\theta$ に対して $S_k > \theta$  または $S_k \leq \theta$ となることを、スパース尺度 $S_k$ が小さいことを示す所定の範囲にあるという。このとき、 $\theta$ のことをスパース尺度 $S_k$ が小さいことを示す所定の範囲にあることを示す値という。

【0073】

（変形例1）

抽出発生音補正部220におけるスパース尺度 $S_k$ と閾値 $\theta$ の比較では、比較条件に用いる閾値 $\theta$ を時間経過によらず不変であるものとして扱ったが、一定のタイミングで閾値 $\theta$ を更新するようにしてもよい。一定のタイミングで更新するようにすると、会場での人数の変化や会場の温度の変化などに対して、閾値 $\theta$ が追従することができるようになる。

20

【0074】

例えば、 $\theta_0$ （ただし、 $\theta_0$ はあらかじめ定められた数）を閾値の初期値として、サンプルごとに閾値 $\theta$ を更新して比較するようにしてもよい。具体的には、 $k=1, \dots$ として、 $S_k > \theta_{k-1}$ の場合、 $e'(k)=e(k)$ とし、 $S_k \leq \theta_{k-1}$ の場合、 $e'(k)=0$ として補正済抽出発生音 $e'(k)$ を生成する。次に、次のサンプルのための閾値を $\theta_k=S_k$ とし、同様に $S_{k+1} > \theta_k$ の場合、 $e'(k+1)=e(k+1)$ とし、 $S_{k+1} \leq \theta_k$ の場合、 $e'(k+1)=0$ として補正済抽出発生音 $e'(k+1)$ を生成する。

【0075】

30

なお、閾値 $\theta_k$ を $\theta_k=(S_{k-L+1}+\dots+S_k)/L$ というようにLサンプル分の平均値としてもよい。さらに、単純な平均とする代わりに、閾値 $\theta_k$ の計算に忘却係数を定めたフィルタを用いてもよいし、線形関数や非線形関数を用いてもかまわない。例えば、 $\theta_k$ を実数の定数（ただし、 $\theta_k > 0$ ）として、ガウス窓を用いて定義される忘却係数 $W(i)$ （ $i=0, 1, \dots$ ）により、閾値 $\theta_k$ を以下のように求めてもよい。

【0076】

【数10】

$$W(i) = \omega \cdot \exp\left(-\frac{i^2}{\sigma^2}\right)$$

40

$$\beta_k = (W(0) \cdot S_k + W(1) \cdot S_{k-1} + \dots + W(L-1) \cdot S_{k-L+1}) / L$$

【0077】

（変形例2）

第一実施形態で説明したように、抽出発生音補正装置が、補正済抽出発生音 $e'(k)$ の代わりに、補正済抽出発生音 $e'(k)$ を生成するために用いるパラメータである補正済抽出発生音パラメータ $p(k)$ を生成するようにしてもよい。

【0078】

以下、図12～図13を参照して抽出発生音補正装置201について説明する。図12は、抽出発生音補正装置201の構成を示すブロック図である。図13は、抽出発生音補

50

正装置 201 の動作を示すフローチャートである。図 12 からわかるように、抽出発生音補正装置 201 は、補正済抽出発生音パラメータ生成部 130 をさらに含む点においてのみ抽出発生音補正装置 200 と異なる。また、図 13 からわかるように、抽出発生音補正装置 201 の動作は、S130 が追加されている点においてのみ抽出発生音補正装置 200 の動作と異なる。補正済抽出発生音パラメータ生成部 130 は、補正済抽出発生音  $e'(k)$  から補正済抽出発生音パラメータ  $p(k)$  を生成する (S130)。補正済抽出発生音  $e'(k)$  を生成するために用いることのできるパラメータであれば、補正済抽出発生音パラメータはどのようなものであってもよい。

#### 【0079】

本発明によれば、聴感上の違和感を抑えるように、大きな会場で発生した音を抽出した抽出発生音を補正することができる。これにより、スピーカとマイクが近しくなく、あえて残響を加えるように設計されている会場で発生した音も適切に抽出・補正することができるようになる。

#### 【0080】

< 第三実施形態 >

第一実施形態ではパワー比  $E_k/Y_k$  を、第二実施形態ではスパース尺度  $S_k$  を用いて、抽出発生音を補正するための条件判定を行った。ここでは、パワー比  $E_k/Y_k$  とスパース尺度  $S_k$  の 2 つを用いた形で条件判定を行う。

#### 【0081】

以下、図 14 ~ 図 15 を参照して抽出発生音補正装置 300 について説明する。図 14 は、抽出発生音補正装置 300 の構成を示すブロック図である。図 15 は、抽出発生音補正装置 300 の動作を示すフローチャートである。図 14 に示すように抽出発生音補正装置 300 は、パワー算出部 110、スパース尺度算出部 210、抽出発生音補正部 320、記録部 190 を含む。

#### 【0082】

図 15 に従い抽出発生音補正装置 300 の動作について説明する。パワー算出部 110 は、收音会場音  $y(k)$  と抽出発生音  $e(k)$  から、收音会場音  $y(k)$  のパワーである收音会場音パワー  $Y_k$  と抽出発生音  $e(k)$  のパワーである抽出発生音パワー  $E_k$  を算出する (S110)。スパース尺度算出部 210 は、抽出発生音  $e(k)$  から、抽出発生音  $e(k)$  のスパース尺度  $S_k$  を算出する (S210)。発生音抽出部 320 は、收音会場音パワー  $Y_k$  と抽出発生音パワー  $E_k$  とスパース尺度  $S_k$  を用いて、抽出発生音  $e(k)$  から補正済抽出発生音  $e'(k)$  を生成する (S320)。例えば、倍率  $\alpha$ 、閾値  $\beta$  に対して、 $E_k > \alpha \cdot Y_k$  かつ  $S_k > \beta$  ( $E_k < \alpha \cdot Y_k$  かつ  $S_k < \beta$ ) となる場合、 $e'(k) = e(k)$  とし、それ以外の場合は  $e'(k) = 0$  として補正済抽出発生音を生成する。その際、例えば、 $Y_k, E_k$  については式(1)、式(2)を用いて、 $S_k$  については式(5) ~ 式(7)を用いて算出することができる。また、 $S_k$  については式(5) ~ 式(7)の代わりに、式(8) ~ 式(10)や式(11) ~ 式(13)を用いてもよい。

#### 【0083】

もちろん、例えば、倍率  $\alpha$ 、閾値  $\beta$  に対して、 $E_k < \alpha \cdot Y_k$  かつ  $S_k < \beta$  ( $E_k > \alpha \cdot Y_k$  かつ  $S_k > \beta$ ) となる場合、 $e'(k) = 0$  とし、それ以外の場合は  $e'(k) = e(k)$  として補正済抽出発生音を生成するようにしてもよい。

#### 【0084】

さらに、抽出発生音補正部 320 における抽出発生音を補正するための条件判定に用いる条件を複数にしてもよい。例えば、発生音抽出部 320 が、收音会場音パワー  $Y_k$  と抽出発生音パワー  $E_k$  とスパース尺度  $S_k$  を用いて、抽出発生音  $e(k)$  から補正済抽出発生音  $e'(k)$  を生成する際、倍率  $\alpha_1, \alpha_2$ 、閾値  $\beta_1, \beta_2$  (ただし、 $\alpha_1 < \alpha_2$ 、 $\beta_1 > \beta_2$ ) に対して、以下の第 1 条件、第 2 条件のいずれかが成立する場合、 $e'(k) = e(k)$  とし、それ以外の場合は  $e'(k) = 0$  として補正済抽出発生音を生成する。

第 1 条件： $E_k > \alpha_1 \cdot Y_k$  かつ  $S_k > \beta_1$

第 2 条件： $E_k > \alpha_2 \cdot Y_k$  かつ  $S_k > \beta_2$

#### 【0085】

10

20

30

40

50

その際、 $Y_k, E_k$ については式(3)、式(4)を用いて、 $S_k$ については式(11)～式(13)を用いて算出することができる。この場合、例えば、 $\alpha_1=0.5$ 、 $\alpha_2=1$ 、 $\beta_1=12$ 、 $\beta_2=9$ とすればよい。

【0086】

図16は、拍手音を環境音として、抽出発生音 $e(k)$ に対して第一実施形態のようにパワー比を用いて生成した補正済抽出発生音 $e'(k)$ (図16(b))と本実施形態のようにパワー比とスパース尺度を用いて生成した補正済抽出発生音 $e'(k)$ (図16(c))の様子を示す図である。図16(b)と図16(c)を比べると、本実施形態(図16(c))の方が、環境音である拍手音の抽出・補正の精度がよいことがわかる。

【0087】

もちろん、倍率 $\alpha_1$ 、 $\alpha_2$ 、閾値 $\beta_1$ 、 $\beta_2$ (ただし、 $\alpha_1 < \alpha_2$ 、 $\beta_1 > \beta_2$ )に対して、以下の第3条件、第4条件のいずれかが成立する場合、 $e'(k)=0$ とし、それ以外の場合は $e'(k)=e(k)$ として補正済抽出発生音を生成するようにしてもよい。

第3条件： $E_k \leq \alpha_1 \cdot Y_k$ かつ $S_k \leq \beta_1$

第4条件： $E_k \leq \alpha_2 \cdot Y_k$ かつ $S_k \leq \beta_2$

【0088】

(変形例1)

第一実施形態や第二実施形態と同様、倍率 $\alpha$ と閾値 $\beta$ をサンプルごとに変化させる構成としてもよい。

【0089】

(変形例2)

また、第一実施形態や第二実施形態で説明したように、抽出発生音補正装置が、補正済抽出発生音 $e'(k)$ の代わりに、補正済抽出発生音 $e'(k)$ を生成するために用いるパラメータである補正済抽出発生音パラメータ $p(k)$ を生成するようにしてもよい。

【0090】

以下、図17～図18を参照して抽出発生音補正装置301について説明する。図17は、抽出発生音補正装置301の構成を示すブロック図である。図18は、抽出発生音補正装置301の動作を示すフローチャートである。図17からわかるように、抽出発生音補正装置301は、補正済抽出発生音パラメータ生成部130をさらに含む点においてのみ抽出発生音補正装置300と異なる。また、図18からわかるように、抽出発生音補正装置301の動作は、S130が追加されている点においてのみ抽出発生音補正装置300の動作と異なる。補正済抽出発生音パラメータ生成部130は、補正済抽出発生音 $e'(k)$ から補正済抽出発生音パラメータ $p(k)$ を生成する(S130)。補正済抽出発生音 $e'(k)$ を生成するために用いることのできるパラメータであれば、補正済抽出発生音パラメータはどのようなものであってもよい。

【0091】

本発明によれば、聴感上の違和感を抑えるように、大きな会場で発生した音を抽出した抽出発生音を補正することができる。これにより、スピーカとマイクが近しくなく、あえて残響を加えるように設計されている会場で発生した音も適切に抽出・補正することができるようになる。

【0092】

<補記>

本発明の装置は、例えば単一のハードウェアエンティティとして、キーボードなどが接続可能な入力部、液晶ディスプレイなどが接続可能な出力部、ハードウェアエンティティの外部に通信可能な通信装置(例えば通信ケーブル)が接続可能な通信部、CPU(Central Processing Unit、キャッシュメモリやレジスタなどを備えていてもよい)、メモリであるRAMやROM、ハードディスクである外部記憶装置並びにこれらの入力部、出力部、通信部、CPU、RAM、ROM、外部記憶装置の間のデータのやり取りが可能なように接続するバスを有している。また必要に応じて、ハードウェアエンティティに、CD-ROMなどの記録媒体を読み書きできる装置(ドライブ)などを設けることとしてもよ

10

20

30

40

50

い。このようなハードウェア資源を備えた物理的実体としては、汎用コンピュータなどがある。

【0093】

ハードウェアエンティティの外部記憶装置には、上述の機能を実現するために必要となるプログラムおよびこのプログラムの処理において必要となるデータなどが記憶されている（外部記憶装置に限らず、例えばプログラムを読み出し専用記憶装置であるROMに記憶しておくこととしてもよい）。また、これらのプログラムの処理によって得られるデータなどは、RAMや外部記憶装置などに適宜に記憶される。

【0094】

ハードウェアエンティティでは、外部記憶装置（あるいはROMなど）に記憶された各プログラムとこの各プログラムの処理に必要なデータが必要に応じてメモリに読み込まれて、適宜にCPUで解釈実行・処理される。その結果、CPUが所定の機能（上記、...部、...手段などと表した各構成要件）を実現する。

【0095】

本発明は上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。また、上記実施形態において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されるとしてもよい。

【0096】

既述のように、上記実施形態において説明したハードウェアエンティティ（本発明の装置）における処理機能をコンピュータによって実現する場合、ハードウェアエンティティが有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記ハードウェアエンティティにおける処理機能がコンピュータ上で実現される。

【0097】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD（Digital Versatile Disc）、DVD-RAM（Random Access Memory）、CD-ROM（Compact Disc Read Only Memory）、CD-R（Recordable）/RW（ReWritable）等を、光磁気記録媒体として、MO（Magneto-Optical disc）等を、半導体メモリとしてEEP-ROM（Electrically Erasable and Programmable-Read Only Memory）等を用いることができる。

【0098】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0099】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみに

10

20

30

40

50

よって処理機能を実現する、いわゆる A S P (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

【0100】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、ハードウェアエンティティを構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【図1】

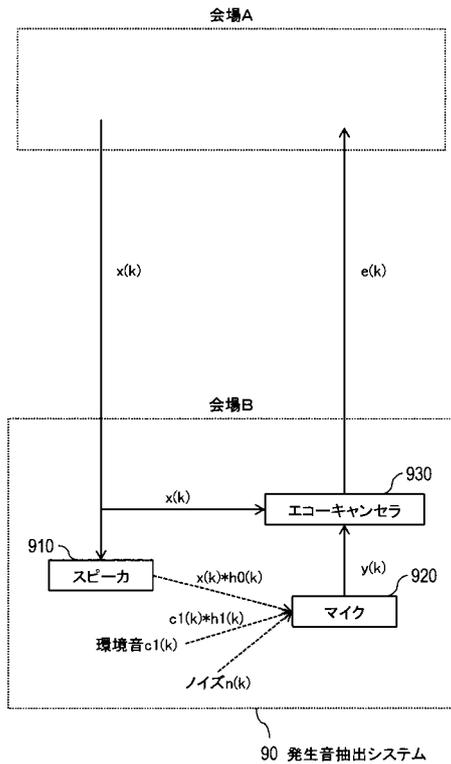


図1

【図2】

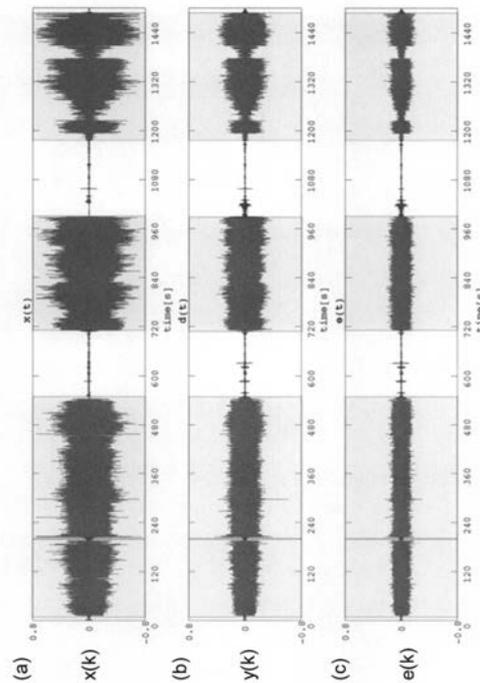


図2

【 図 3 】

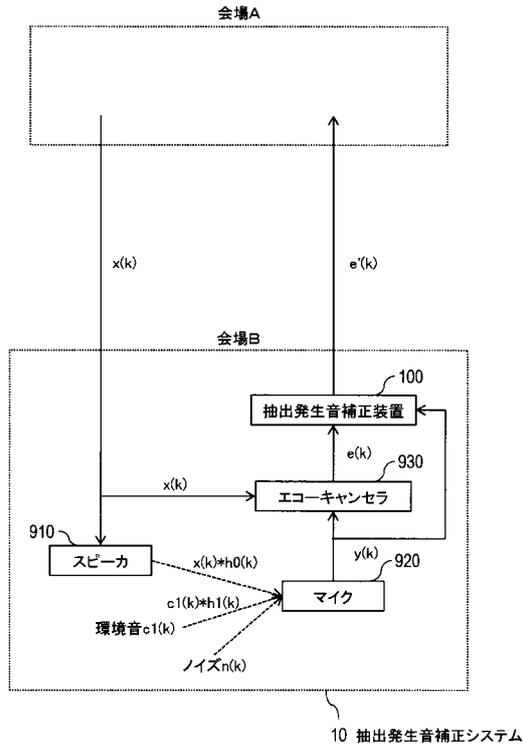


図3

【 図 4 】

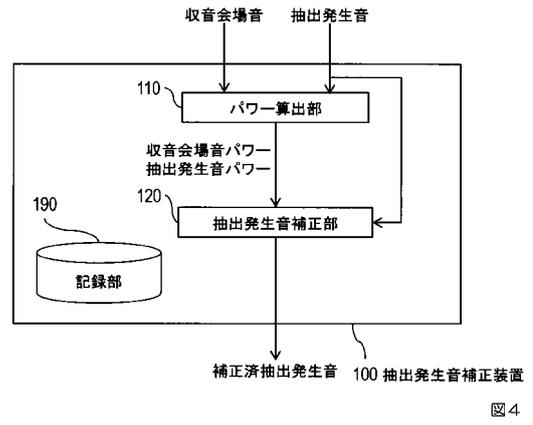


図4

【 図 5 】

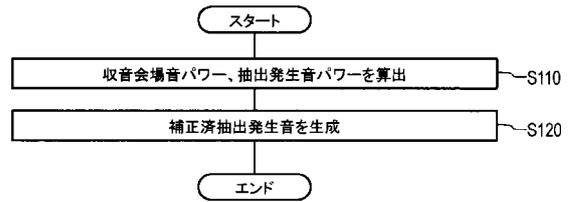


図5

【 図 6 】

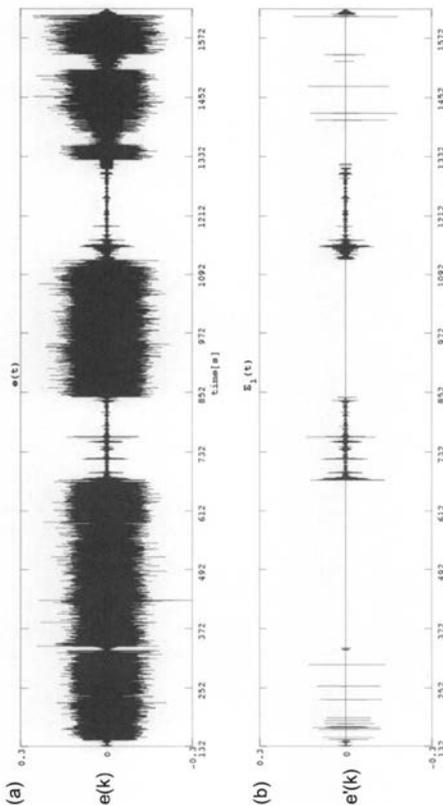


図6

【 図 7 】

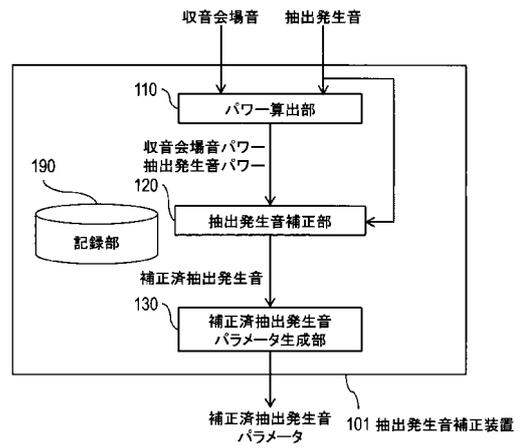


図7

【 図 8 】

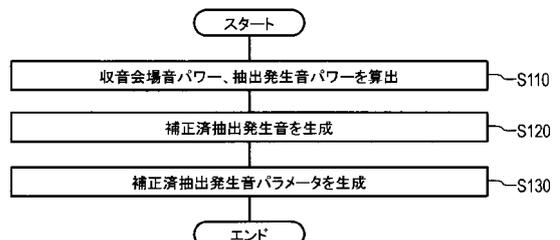


図8

【 図 9 】

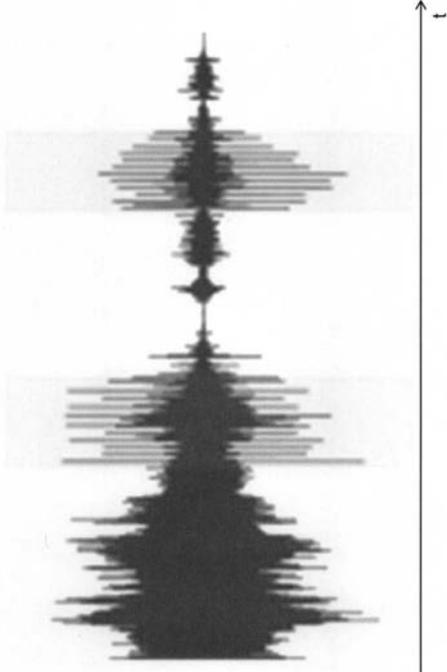


図9

【 図 1 0 】

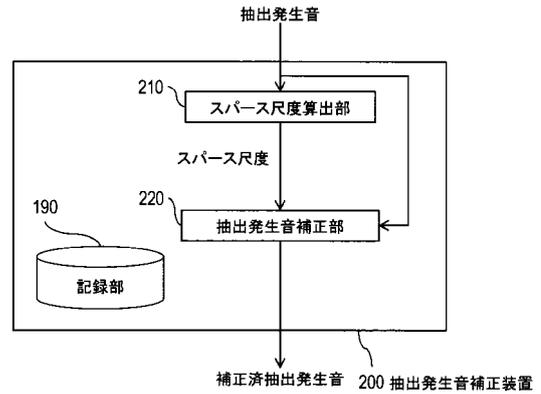


図10

【 図 1 1 】

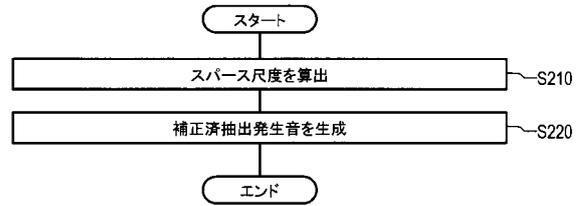


図11

【 図 1 2 】

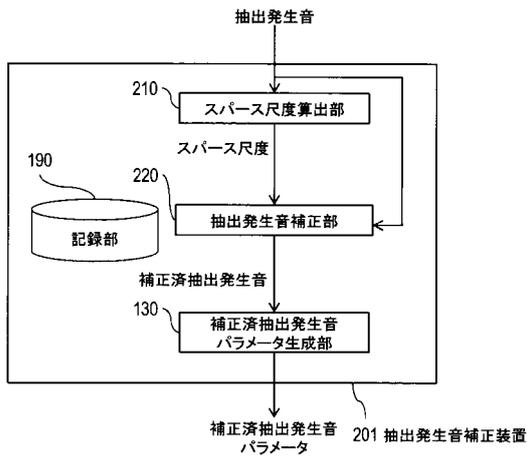


図12

【 図 1 4 】

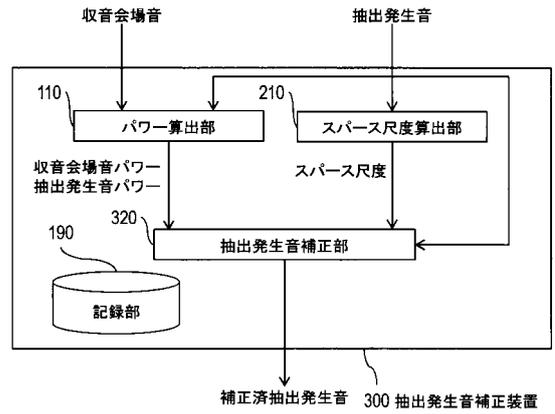


図14

【 図 1 3 】

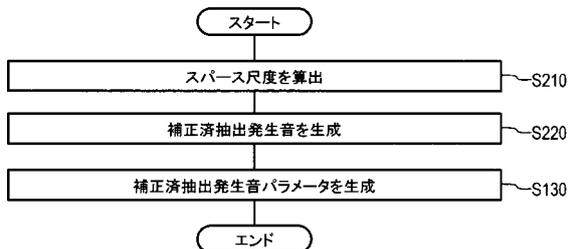


図13

【 図 1 5 】

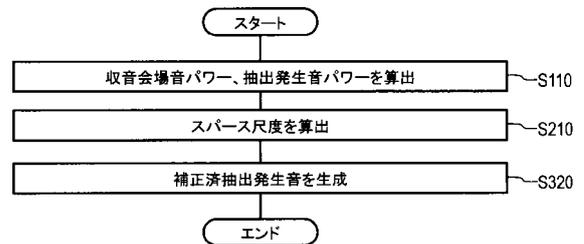


図15

【 図 1 6 】

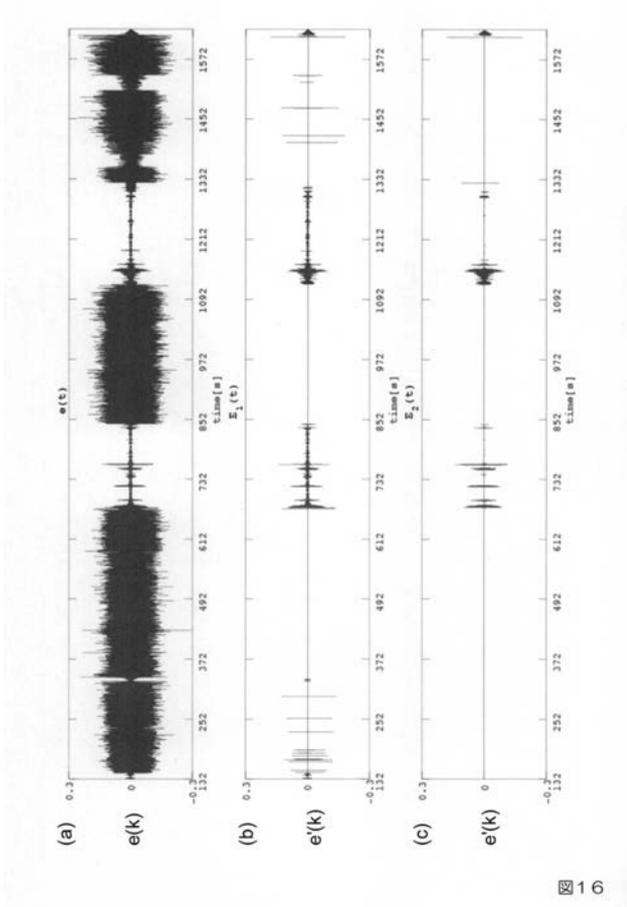


図 16

【 図 1 7 】

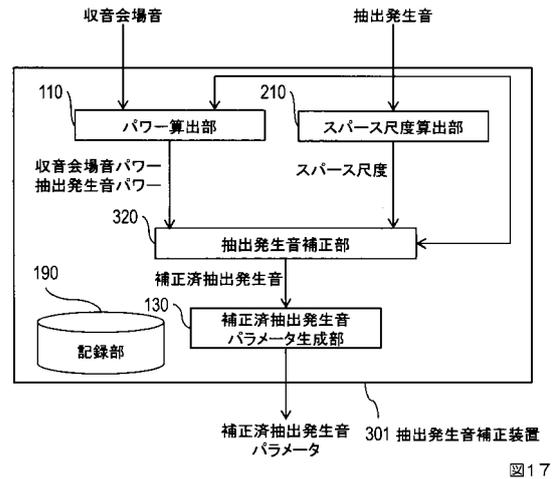


図 17

【 図 1 8 】

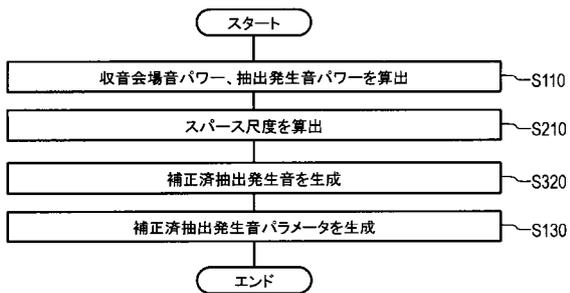


図 18

## フロントページの続き

- (72)発明者 佐藤 尚  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 白木 善史  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 杉浦 亮介  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 川西 隆仁  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 守谷 健弘  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 河原 一彦  
福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内
- (72)発明者 西川 萌恵  
福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内
- (72)発明者 尾本 章  
福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内
- (72)発明者 藤森 朗穂  
福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内
- Fターム(参考) 5D220 CC06