

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6887620号
(P6887620)

(45) 発行日 令和3年6月16日(2021.6.16)

(24) 登録日 令和3年5月21日(2021.5.21)

(51) Int. Cl.		F I			
H04R	3/00	(2006.01)	H04R	3/00	310
G10K	15/04	(2006.01)	G10K	15/04	302F
G10L	25/72	(2013.01)	G10L	25/72	Z

請求項の数 7 (全 16 頁)

(21) 出願番号	特願2017-86836 (P2017-86836)	(73) 特許権者	000004226
(22) 出願日	平成29年4月26日(2017.4.26)		日本電信電話株式会社
(65) 公開番号	特開2018-186386 (P2018-186386A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成30年11月22日(2018.11.22)	(73) 特許権者	504145342
審査請求日	令和1年7月29日(2019.7.29)		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	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

最終頁に続く

(54) 【発明の名称】 環境音合成システム、その方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

環境音分析装置から伝送元の音響信号の音量に関する環境音量パラメタを取得して環境音を生成する環境音合成システムであって、

Nを1以上の整数の何れかとし、前記環境音はN個の再生部で再生されるものとし、

前記環境音量パラメタによって表される音量を、N個の再生部で再生させて実現するためのパラメタである駆動パラメタに変換する駆動パラメタ算出部と、

環境音のテンプレートである環境音素片テンプレートを記憶するテンプレート記憶部と

環境音素片テンプレートの環境音の音量に対応する情報とその環境音素片テンプレートの長さに対応する情報とを用いて、前記駆動パラメタの示す音量が平均的に実現されるように、その環境音素片テンプレートを発生させるタイミングを決定し、再生音を合成する再生音合成部とを含む、

環境音合成システム。

【請求項2】

請求項1の環境音合成システムであって、

前記再生音合成部は、前記駆動パラメタの示す音量が実現されるために必要な人数に対応する人がそれぞれ前記再生部の存在する位置で環境音を発しているかのような再生音を合成する、

環境音合成システム。

10

20

【請求項 3】

請求項 2 の環境音合成システムであって、

前記駆動パラメタ算出部は、前記環境音量パラメタに基づき、1つの再生部で実現する環境音の人数を駆動パラメタとして求め、

前記再生音合成部は、前記駆動パラメタの整数部に対応する人がそれぞれ前記再生部の存在する位置で環境音を発しているかのような再生音を合成し、前記駆動パラメタの少数部に対応する人がN個の再生部全体で実現されるように確率的に人数調整を行う、

環境音合成システム。

【請求項 4】

環境音分析装置から伝送元の音響信号の音量に関する環境音量パラメタを取得して環境音を生成する環境音合成方法であって、 10

Nを1以上の整数の何れかとし、前記環境音はN個の再生部で再生されるものとし、テンプレート記憶部には、環境音のテンプレートである環境音素片テンプレートが記憶されているものとし、

前記環境音量パラメタによって表される音量を、N個の再生部で再生させて実現するためのパラメタである駆動パラメタに変換する駆動パラメタ算出ステップと、

環境音素片テンプレートの環境音の音量に対応する情報とその環境音素片テンプレートの長さに対応する情報とを用いて、前記駆動パラメタの示す音量が平均的に実現されるように、その環境音素片テンプレートを発生させるタイミングを決定し、再生音を合成する再生音合成ステップとを含む、 20

環境音合成方法。

【請求項 5】

請求項 4 の環境音合成方法であって、

前記再生音合成ステップは、前記駆動パラメタの示す音量が実現されるために必要な人数に対応する人がそれぞれ前記再生部の存在する位置で環境音を発しているかのような再生音を合成する、

環境音合成方法。

【請求項 6】

請求項 5 の環境音合成方法であって、

前記駆動パラメタ算出ステップは、前記環境音量パラメタに基づき、1つの再生部で実現する環境音の人数を駆動パラメタとして求め、 30

前記再生音合成ステップは、前記駆動パラメタの整数部に対応する人がそれぞれ前記再生部の存在する位置で環境音を発しているかのような再生音を合成し、前記駆動パラメタの少数部に対応する人がN個の再生部全体で実現されるように確率的に人数調整を行う、

環境音合成方法。

【請求項 7】

請求項 1 から請求項 3 の何れかの環境音合成システムとして、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】 40

【0001】

本発明は、伝送元で収録された環境音を、伝送先で再生する環境音合成システム、その方法及びプログラムに関する。

【背景技術】

【0002】

実測データをもとに算出した個人差や、速度・大きさの揺らぎの程度を利用して、単独のユーザと同期するように複数の拍手音を合成し出力する技術が提案されている（非特許文献1）。また、ある地点の音を別の場所に伝送し再生する技術として、音響符号化技術が知られている。例えば、非特許文献2では、聴覚マスキングを巧みに利用し、また楽器の特性を利用して低域の成分を高域にコピーして使うという楽音の特性に合わせたモデル 50

により、低ビットレートで品質の高い音響符号化技術が提案されている。

【0003】

非特許文献1は、ユーザと同調する複数の人がその場にいるような環境を仮想的に実現することを目的としたものであり、ユーザの拍手のピッチに合わせて仮想的な拍手音を合成する技術であり、実在する遠隔地の場の状況（拍手音や手拍子）を、別の場所に伝送し再現することはできなかった。また、声援・掛け声などの拍手音以外の環境音を伝送し再現することは対象としていない。また、拍手音や声援・掛け声などの環境音は純粋な音声や楽器音とは異なり白色雑音に近いため、非特許文献2のような従来の音響符号化技術ではうまく表現できず、音質が劣化していた。

【0004】

伝送元において収録された拍手や手拍子音、声援・掛け声などの環境音を効率よく伝送し、伝送先で伝送元の場の雰囲気再現することができる環境音合成装置として、特許文献1が知られている。

【0005】

特許文献1の環境音合成装置では、テンプレート記憶部に1フレーム分（一定時間分）の環境音のテンプレートと当該テンプレートの環境音の音量に対応する情報とを対応付けて記憶しておき、音源合成部が受信した環境音量パラメタと同じ音量大きさのテンプレートをテンプレート記憶部から選択し、選択したテンプレートを合成して環境音を生成する。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2014-63145号公報

【非特許文献】

【0007】

【非特許文献1】西村竜一、宮里勉、「仮想的集団による拍手音の合成」、電子情報通信学会技術研究報告、電子情報通信学会、1999年3月、MVE,マルチメディア・仮想環境基礎、98(684), p.17 24,

【非特許文献2】Stefan Meltzer and Gerald Moser, "MPEG 4 HE AAC v2 audio coding for today's digital media world," EBU technical review, Jan., 2006.

【発明の概要】

【発明が解決しようとする課題】

【0008】

しかしながら、特許文献1では、伝送先における残響を考慮していない。そのため、生成された環境音を再生すると、一点から環境音が発せられるように聞こえてしまい、実際には一点ではなく所定の空間から発せられる環境音を適切に再現することが難しい。特に、伝送先の空間が広い場合にその傾向が強くなる。

【0009】

そこで本発明では、伝送先において、多数の再生部が存在する場合にそれらを効果的に活用することにより、伝送先であたかも伝送元の音源がその場(伝送先)に存在するかのように環境音を再現することができる環境音合成システム、その方法及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】

【0010】

上記の課題を解決するために、本発明の一態様によれば、環境音合成システムは、環境音分析装置から伝送元の音響信号の音量に関する環境音量パラメタを取得して環境音を生成する。Nを1以上の整数の何れかとし、環境音はN個の再生部で再生されるものとし、環境音合成システムは、環境音量パラメタによって表される音量を、N個の再生部で再生させて実現するためのパラメタである駆動パラメタに変換する駆動パラメタ算出部と、環境音のテンプレートである環境音素片テンプレートを記憶するテンプレート記憶部と、環

10

20

30

40

50

環境音素片テンプレートの環境音の音量に対応する情報とその環境音素片テンプレートの長さに対応する情報とを用いて、駆動パラメタの示す音量が平均的に実現されるように、その環境音素片テンプレートを発生させるタイミングを決定し、再生音を合成する再生音合成部とを含む。

【0011】

上記の課題を解決するために、本発明の他の態様によれば、環境音合成方法は、環境音分析装置から伝送元の音響信号の音量に関する環境音量パラメタを取得して環境音を生成する。Nを1以上の整数の何れかとし、環境音はN個の再生部で再生されるものとし、テンプレート記憶部には、環境音のテンプレートである環境音素片テンプレートが記憶されているものとし、環境音合成方法は、環境音量パラメタによって表される音量を、N個の再生部で再生させて実現するためのパラメタである駆動パラメタに変換する駆動パラメタ算出ステップと、環境音素片テンプレートの環境音の音量に対応する情報とその環境音素片テンプレートの長さに対応する情報とを用いて、駆動パラメタの示す音量が平均的に実現されるように、その環境音素片テンプレートを発生させるタイミングを決定し、再生音を合成する再生音合成ステップとを含む。

【発明の効果】

【0012】

本発明によれば、伝送先であたかも伝送元の音源がその場(伝送先)に存在するかのよう

に環境音を再現することができるという効果を奏する。

【図面の簡単な説明】

【0013】

【図1】環境音伝送システムの機能ブロック図。

【図2】端末指示装置の機能ブロック図。

【図3】端末指示装置の処理フローの例を示す図。

【図4】環境音再生端末の機能ブロック図。

【図5】環境音再生端末の処理フローの例を示す図。

【図6】合成タスクを説明するための図。

【発明を実施するための形態】

【0014】

以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、ベクトルや行列の各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

【0015】

<第一実施形態のポイント>

伝送元で発生した環境音の音量に対応する情報を伝送先に送り、その音量を伝送先に分散して配置された複数の再生部(スピーカ)を駆動することによって実現する。

本実施形態では環境音再生端末が伝送先の会場内に分散して複数配置されており、この環境音再生端末が全体として伝送元で生じた環境音の音量を模擬できるような動作を行う。

【0016】

<第一実施形態に係る環境音伝送システム>

図1は第一実施形態に係る環境音伝送システム1000の構成を示すブロック図である。

本実施形態の環境音伝送システム1000は、伝送元の環境音分析装置1と、伝送先の端末指示装置21、N個の環境音再生端末22-n(あわせて環境音合成サブシステム2)から構成される。Nは1以上の整数の何れかであり、n=1,2,...,Nである。

【0017】

<環境音分析装置1>

拍手や手拍子音、声援・掛け声などの環境音の総音量は、観客の人数が多いほど大きく

なる。環境音分析装置 1 では伝送元の環境音そのものを伝送するのではなく、環境音の音量を表す情報だけを抽出することにより、收音された環境音を効率よく低遅延に伝送することが可能となっている。図 1 に示すように、環境音分析装置 1 は、入力された音響信号（環境音）の音量に対応する情報（環境音量パラメタ P_i ）を抽出し、出力する。ここで j はフレームのインデックスを示し、環境音分析装置 1 はフレーム毎に環境音量パラメタ P_i を出力する。環境音量パラメタ P_i の抽出方法としては様々な方法を用いることができる。例えば、特許文献 1 の方法により環境音量パラメタ P_i を抽出してもよく、伝送元で收音した音響信号の平均エネルギー E_i に対応する情報等を環境音量パラメタ P_i として用いることができる。

【 0 0 1 8 】

< 環境音合成サブシステム 2 >

環境音合成サブシステム 2 は、端末指示装置 2 1 と N 台の環境音再生端末 2 2 - n によって構成される。環境音分析装置 1 より入力された環境音量パラメタ P_i に基づき端末指示装置 2 1 は駆動パラメタ D_i を求め、 N 台の環境音再生端末 2 2 - n に送信し、 N 台の環境音再生端末 2 2 - n から予め記憶された環境音のテンプレート(以下、環境音素片テンプレートともいう)が再生される頻度をコントロールすることによって環境音を合成する。

【 0 0 1 9 】

< 端末指示装置 2 1 >

本実施形態の端末指示装置 2 1 は、環境音分析装置 1 から伝送元の音響信号の音量に関する環境音量パラメタ P_i を取得し、環境音量パラメタ P_i を駆動パラメタ D_i に変換した上で各環境音再生端末 2 2 - n に駆動パラメタ D_i を分配する。

本実施形態の端末指示装置 2 1 は、データ受信部 2 1 1 と、駆動パラメタ算出部 2 1 2 と、駆動パラメタ D_i を送信する端末指示部 2 1 3 とを備える。図 2 は端末指示装置 2 1 の機能ブロック図を、図 3 はその処理フローの例を示す。

【 0 0 2 0 】

< データ受信部 2 1 1 >

データ受信部 2 1 1 は、環境音分析装置 1 から環境音量パラメタ P_i を受信し（ $S 2 1 1$ ）、駆動パラメタ算出部 2 1 2 に出力する。

【 0 0 2 1 】

< 駆動パラメタ算出部 2 1 2 >

駆動パラメタ算出部 2 1 2 は、環境音量パラメタ P_i を入力とし、環境音量パラメタ P_i から音量を抽出し、さらに駆動パラメタ D_i に変換し（ $S 2 1 2$ ）、出力する。本実施形態では環境音量パラメタ P_i は伝送元の環境音の音量に対応する値であり、特許文献 1 に示されているように環境音量パラメタ P_i から音量に対応する情報(平均エネルギー E_i)を抽出できる。駆動パラメタ算出部 2 1 2 は情報(平均エネルギー E_i)によって表される音量を、 N 個の環境音再生端末 2 2 - n （再生部 2 2 3 - n ）全体を駆動させて実現するためのパラメタである駆動パラメタ D_i に変換する。

【 0 0 2 2 】

本実施形態では音量に対応する情報(平均エネルギー E_i)を環境音再生端末 2 2 - n の総数 N を用いて、

$$D_i = E_i / N$$

により駆動パラメタ D_i に変換する。また、本実施形態から明らかであるが、駆動パラメタ算出部 2 1 2 は環境音再生端末 2 2 - n の総数 N の値を保持しているものとする。

【 0 0 2 3 】

< 端末指示部 2 1 3 >

端末指示部 2 1 3 は、駆動パラメタ D_i を入力とし、 N 個の環境音再生端末 2 2 - n 全てに駆動パラメタ D_i を分配する（ $S 2 1 3$ ）。本実施形態では端末指示部 2 1 3 は N 個の環境音再生端末 2 2 - n に同じ情報を伝える。環境音再生端末の個別識別の必要がないため、単純な通信方式で伝えることができる。例えばbluetooth(登録商標)、wifi(登録商標)などで実現できる。

10

20

30

40

50

【 0 0 2 4 】

< 環境音再生端末 2 2 - n >

本実施形態の環境音再生端末 2 2 - n は、N個存在し、それぞれ端末指示装置 2 1 から駆動パラメタ D_i を取得し、その駆動パラメタ D_i を用いて再生部から音を発生させることを特徴とする。

図 4 は環境音再生端末 2 2 - n の機能ブロック図を、図 5 はその処理フローの例を示す。環境音再生端末 2 2 - n は、再生情報受信部 2 2 1 - n、再生音合成部 2 2 2 - n、再生部 2 2 3 - n 及びテンプレート記憶部 2 2 4 - n を備える。N個の環境音再生端末 2 2 - n 全体の発する音であたかも伝送元の間人がその場で環境音を発しているかのように環境音を合成するように動作する。

10

【 0 0 2 5 】

< 再生情報受信部 2 2 1 - n >

再生情報受信部 2 2 1 - n は、端末指示装置 2 1 から駆動パラメタ D_i を受信(取得)し (S 2 2 1)、再生音合成部 2 2 2 - n に出力する。

【 0 0 2 6 】

< テンプレート記憶部 2 2 4 - n >

テンプレート記憶部 2 2 4 - n には、L個の環境音素片テンプレート T_i と、そのL個の環境音素片テンプレート T_i の音量にそれぞれ対応するL個の情報 G_i と、L個の環境音素片テンプレートの長さに対応するL個の情報 L_i とが組みになって保存されている。ただし、 i は環境音素片テンプレートのインデックスを示し、 $i=1,2, \dots, L$ であり、 L は1以上の自然数の何れかである。ここで環境音素片テンプレートとは一人の人間の行動によって生じる環境音(300ms程度)をさし、環境音素片テンプレートの音量に対応する情報とは例えば特許文献 1 に記載された手法によって計算される環境音素片テンプレートの持つ音量である。例えば、環境音素片テンプレート T_i がHサンプルで構成される信号列 $X_i=(x_i(1), x_i(2), \dots, x_i(H))$ からなる場合、信号列 $X_i=(x_i(1), x_i(2), \dots, x_i(H))$ から、環境音素片テンプレート T_i の平均エネルギー

20

【数 1】

$$E_i = \frac{x_i(1)^2 + x_i(2)^2 + \dots + x_i(H)^2}{H}$$

を計算し、この平均エネルギー E_i を環境音素片テンプレート T_i の音量に対応する情報 G_i として利用する。

【 0 0 2 7 】

本実施形態では拍手音を例に用いるため、異なる人の一拍分の拍手音の環境音素片テンプレート T_i とその長さに対応する情報 L_i と、その音量(平均エネルギー)に対応する情報 G_i が組になって記憶されている。例えば、環境音素片テンプレート T_i の長さに対応する情報 L_i としては長さそのもの[ms]を用いることができる。

【 0 0 2 8 】

なお、1人分の拍手音の音量の差は小さいので、テンプレート記憶部 2 2 4 - n に記憶されている環境音素片テンプレート T_i のいずれか一つについて計算された平均エネルギー E_i を、L個全ての環境音素片テンプレート T_i の音量に対応する情報 G_i として1つ記憶しておいてもよい。また、テンプレート記憶部 2 2 4 - n に記憶されている全環境音素片テンプレートの平均エネルギーの平均値 E を、L個全ての環境音素片テンプレート T_i の音量に対応する情報 G_i としてテンプレート記憶部 2 2 4 - n に記憶しておいてもよい。あるいは、予め定めた定数を音量に対応する情報 G_i としてテンプレート記憶部 2 2 4 - n に記憶しておいても良い。また、テンプレート記憶部 2 2 4 - n に予め環境音素片テンプレートの音量に対応する情報を記憶せず、その都度テンプレート記憶部 2 2 4 - n からランダムに選択した環境音素片テンプレート T_i について計算した平均エネルギー E_i を環境音素片テンプレート T_i の音量に対応する情報 G_i として用いても良い。 l は1,2, ..., Lの何れかである。

40

【 0 0 2 9 】

50

また、テンプレート記憶部 224 - n に予め環境音素片テンプレート T_i の長さに対応する情報 L_i を記憶せず、その都度テンプレート記憶部 224 - n からランダムに選択した環境音素片テンプレート T_i から取得(計算)しても良い。

よって、テンプレート記憶部 224 - n には、少なくとも L 個の環境音素片テンプレート T_i が記憶されていればよい。また、 L 個の環境音素片テンプレート T_i は、テンプレート記憶部 224 - n 毎に異なるものを用いてもよいし、同じものを用いてもよい。

【0030】

<再生音合成部 222 - n >

再生音合成部 222 - n は、駆動パラメタ D_i を入力とし、テンプレート記憶部 224 - n から (L 個の環境音素片テンプレート T_i の中から)ランダムに1つの環境音素片テンプレート T_i を選択し、環境音素片テンプレート T_i の環境音の音量に対応する情報 G_i と環境音素片テンプレート T_i の環境音の長さに対応する情報 L_i とを用いて、駆動パラメタ D_i の示す音量が平均的に実現されるように、再生音を合成し ($S222$)、出力する。その繰り返し頻度(環境音素片テンプレート T_i を発生させるタイミング)は駆動パラメタ D_i をもとに決定され、駆動パラメタ D_i の示す音量が平均的に実現されるように決定される。

【0031】

例えば、再生音合成部 222 - n は、駆動パラメタ D_i と、選択した環境音素片テンプレート T_i の音量、長さに対応する情報 G_i, L_i を用いて、時間 L_i/G_i を計算する。この時間 L_i/G_i 中にランダムに環境音素片テンプレート T_i を1回発生させて再生部 223 - n へ出力し、時間 L_i/G_i が経過した後再び駆動パラメタ D_i を取得し、同じ動作を繰り返せばよい。なお、駆動パラメタ $D_i=0$ のとき、または、時間 L_i/G_i がある時間 (例えば1[s]) を超える場合には、前述のある時間 (例えば1[s]) 経過後に強制的に駆動パラメタ D_i を取得し再び動作させてもよい。

【0032】

<再生部 223 - n >

再生部 223 - n は、再生音合成部 222 - n が合成した拍手音(環境音素片テンプレート T_i)を再生する ($S223$)。

【0033】

<効果 >

以上の構成により、複数のスピーカ(再生部)が配置された場合でも各々が自然な音量の拍手音を合成することが出来る。なお、単に同じ音を分けて再生すると小さな拍手音になってしまうが、本実施形態では上述の構成により、このような問題を解決している。

観客の中に分散配置されたスピーカから音を出すことにより、特別な残響処理を施さずとも観客の位置から観客の出すドライな環境音が再生されるため、自然な環境音を伝送先に再現することが出来る。

【0034】

<変形例 >

第一実施形態では、伝送元の環境音の例として拍手音を対象とし、伝送元の拍手音を分析する環境音分析装置 1 の動作例を説明したが、これに限らず拍手音以外の環境音を対象としても良い。例えば、声援や掛け声などを環境音としても良いし、伝送元で収音される音の中から伝送元会場のメインコンテンツの音を除いた音響信号(雑音を含む)を環境音としても良い。

【0035】

拍手音や声援・掛け声、雑音などは、いずれも伝送元の会場の雰囲気決定づける重要な要素である一方で、いろいろな音響信号が混合された白色雑音に近い信号である。前述したようにこれらの音を環境音と呼ぶ。伝送元で環境音が発せられたタイミング及び音量が保たれていれば、信号そのものは伝送元の環境音と全く同じ信号でなくとも、場の雰囲気を再現することができる。そこで、環境音分析装置 1 において、伝送元の環境音の音量に関するパラメタを抽出することで、伝送元において収音された環境音を効率よく低遅延に伝送することができる。

10

20

30

40

50

【 0 0 3 6 】

本実施形態では、端末指示部 2 1 3 は、駆動パラメタ D_i を入力とし、N個の環境音再生端末 2 2 - n 全てに駆動パラメタ D_i を分配しているが、環境音再生端末 2 2 - n 毎に駆動パラメタ D_i に対応する情報を分配する構成としてもよい。例えば、端末指示部 2 1 3 において、環境音再生端末 2 2 - n 毎の時間 G/D_i を計算し、インデックス i とともに出力する構成としてもよい。再生音合成部 2 2 2 - n では、インデックス i に対応する環境音素片テンプレート T_i をテンプレート記憶部 2 2 4 - n から取り出し、時間 G/D_i 中にランダムに環境音素片テンプレート T_i を 1 回発生させて再生部 2 2 3 - n に出力する。この場合、環境音再生端末 2 2 - n の個別識別が必要となる。

【 0 0 3 7 】

さらに、端末指示装置 2 1 側で各再生部 2 2 3 - n で再生する環境音素片テンプレート T_i を発生させ、各再生部 2 2 3 - n に出力する構成としてもよい。ただし、音声データ(環境音素片テンプレート T_i)をそのまま伝送するため、駆動パラメタ D_i のみを伝送する場合に比べ、端末指示装置 2 1 と環境音再生端末 2 2 - n との間の伝送量が増える。この場合、環境音再生端末 2 2 - n は再生部 2 2 3 - n のみを含めばよい。

【 0 0 3 8 】

なお、駆動パラメタ算出部、再生音合成部、テンプレート記憶部からなる構成を環境音合成システムともいう。また、環境音合成システムは、再生部を含まなくともよく、再生部を別装置としてもよい。

【 0 0 3 9 】

< 第二実施形態のポイント >

第一実施形態では、N個の環境音再生端末 2 2 - n 全体から再生される音量が伝送元の環境音の音量と近い状態が実現できる。しかし、例えば、ある環境音再生端末 2 2 - n が一人分の拍手音しか出さない場合、再生音合成部 2 2 2 - n では、様々な環境音素片テンプレート T_i がランダムに用いられるために、同一人物が出しているかのような環境音を再現しづらいという問題があった。

【 0 0 4 0 】

そこで、環境音を合成する際に、環境音再生端末 2 2 - n 上で一人の人が出す環境音をひとつのタスク(以下、合成タスクともいう)として管理し、その合成タスクの増減によって音量を調節することにより、あたかも環境音再生端末 2 2 - n の場所に人がいるかのような自然な環境音を合成することが可能となる。

【 0 0 4 1 】

以下にその具体的な構成を記す。

< 第二実施形態に係る環境音伝送システム >

第一実施形態と異なる部分を中心に説明する。

【 0 0 4 2 】

< 環境音再生端末 3 2 - n >

本実施形態では、一人の人物が出す環境音をひとつの合成タスクとして管理することにより、伝送先の会場内に会場元で計測された環境音の音量を自然に再現できる数の人間が伝送先に存在するかのような自然な環境音を再現する。環境音再生端末 3 2 - n 上で一人の人物が出す環境音をひとつの合成タスクとして管理し、この合成タスクの増減によって環境音の音量をコントロールする。

【 0 0 4 3 】

< 環境音再生端末 3 2 - n の再生音合成部 3 2 2 - n >

本実施形態では駆動パラメタ D_i によって表される音量の拍手を行うのに必要な人数があたかも環境音再生端末 3 2 - n (再生部 2 2 3)の存在する位置で拍手をしているかのような再生音を発生する。そのために、特定の環境音素片テンプレート T_i をある一定の周期で繰り返すことにより一人の人物の環境音を模擬し、この模擬した人物の環境音の発生と終了をコントロールすることにより駆動パラメタ D_i の表す音量を実現する。

【 0 0 4 4 】

10

20

30

40

50

まず、拍手音が収録された環境音素片テンプレート T_i を用いて一人の人物の環境音を模擬する方法を示す。本実施形態では環境音素片テンプレート T_i を(情報 G_i に対応する)その長さごとに再生することによって実現する。このように同じ環境音素片テンプレート T_i を繰り返すことによって、時間 G_i/D_i ごとにランダムに選択した環境音素片テンプレート T_i を再生するより同一性が高まり、繰返し周期も一定になるためあたかも一人の人間が拍手しているかのような環境音を作成することが出来る。

【 0 0 4 5 】

さらに発生間隔 G_i に $\pm 10\text{ms}$ 前後の揺らぎを持たせることで、より自然な一人の人物の拍手音を再現することが出来る。以降説明の簡単のため、前述の方法で環境音素片テンプレート T_i によって一人の人物の環境音を再現している状況を合成タスク L_k とし(k は合成タスクのインデックスであり、 $k=1,2, \dots, J$ とする)、その合成タスク集合を $[L]=[L_1, L_2, \dots, L_J]$ とする。ただし、 G_i は伝送元の環境音の音量に合わせて増減する値である。図 6 は合成タスクを説明するための図である。

【 0 0 4 6 】

以下、具体的な再生音合成部 3 2 2 - n の動作を示す。再生音合成部 3 2 2 - n は、駆動パラメタ D_i を入力とし、駆動パラメタ D_i の示す音量が実現されるために必要な人数に対応する人がそれぞれ再生部 2 2 3 - n の存在する位置で環境音を発しているかのような再生音を合成し(S 3 2 2)、出力する。例えば、再生音合成部 3 2 2 - n は、現在合成中の音量 O_i 、直前に取得した駆動パラメタ D_i を保持しており、合成タスク集合 $[L]$ によって再生音を合成している。新たに駆動パラメタ D_i を取得し、駆動パラメタ D_i の値により以下の動作を行う。

【 0 0 4 7 】

(A) $D_i = D_{i-1}$ の場合(環境音の音量に変化がない場合)

$O_i = O_{i-1}$ とする。現在の合成タスク集合 $[L]$ による合成音再生を継続する。つまり、合成タスク集合 $[L]$ に含まれる J 個の合成タスク $L_k(k=1,2, \dots, J)$ にそれぞれ対応する環境音素片テンプレート T_i を(情報 G_i に対応する)その長さ毎に発生させて再生部 2 2 3 - n に出力する。

【 0 0 4 8 】

(B) $D_i > D_{i-1}$ の場合(環境音の音量が大きくなる場合)

(i) $O_i = O_{i-1}$ とする。

(ii) L 個の環境音素片テンプレート T_i の中からランダムに1個の環境音素片テンプレート T_i を選択する。 $(D_i - O_i) > G_i$ であれば(iii)へ、それ以外 $((D_i - O_i) \leq G_i)$ は(iv)へ移行する。

(iii)環境音素片テンプレート T_i をもとに合成タスク L_k を新たに合成し、合成タスク集合 $[L]$ に加え、 $J = J + 1$ 、 $O_i = O_{i-1} + G_i$ とする。(ii)へ戻る。

(iv)現在の合成タスク集合 $[L]$ による合成音再生を継続する。

上述の(i)~(iv)に係る処理により、環境音の音量が、環境音素片テンプレートの音量に対応する情報 G_i よりも大きくなる場合、新たな拍手音を合成し、伝送先で再生される環境音の音量を大きくする。

【 0 0 4 9 】

(C) $D_i < D_{i-1}$ の場合(環境音の音量が小さくなる場合)

(v) $O_i = O_{i-1}$ とする。

(vi)合成タスク集合 $[L]$ の中からランダムに1つ合成タスク $L_m(m$ は $1,2, \dots, J$ の何れか)を選択し、合成タスク L_m が用いる環境音素片テンプレート T_i の平均エネルギー E_m を音量に対応する情報 G_m として用いて、 $(D_i - O_i) < G_m$ であれば(vii)へ、それ以外 $((D_i - O_i) \geq G_m)$ は(viii)へ移行する。

(vii)合成タスク L_m の合成を終了し、合成タスク集合 $[L]$ から合成タスク L_m を除き、 $J = J - 1$ 、 $O_i = O_{i-1} - G_m$ とし、(vi)へ戻る。

(viii)現在の合成タスク集合 $[L]$ による合成音再生を継続する。

【 0 0 5 0 】

上述の(v)~(viii)に係る処理により、環境音の音量が、環境音素片テンプレートの音

10

20

30

40

50

量に対応する情報G_nよりも小さくなる場合、合成タスク集合[L]の中から合成タスクを削除し、伝送先で再生される環境音の音量を小さくする。

【0051】

<効果>

このような構成とすることで、第一実施形態と同様の効果を得ることができる。さらに、本実施形態によれば、環境音再生端末上で一人の人が出す環境音をひとつの合成タスクとして管理し、その合成タスクの増減によって音量を調節することにより、あたかも環境音再生端末の場所に人がいるかのような自然な環境音を合成することが可能となる。

【0052】

<変形例>

なお、合成タスク集合[L]はなるべく不均一なタイミングで拍手が行われるほうが自然である。ところが、環境音素片テンプレートT_iが全て同じ長さであったり、フレーム更新間隔が環境音素片テンプレートT_iの長さより長かったりすると、合成タスクL_iのオンセット(開始時刻)が揃ってしまう場合がある。その場合は前述の(iii)で、環境音素片テンプレートT_iを用いた合成タスクL_iを開始する時に0から まで間でランダムなタイムラグを入れることによって解決できる(図6のタイムラグ参照)。

【0053】

また、合成タスクL_iの合成終了を環境音素片テンプレートT_iの再生が終了した時点とし、環境音素片テンプレートT_iの再生途中で音が切れないように(プチッと切れないように)してもよい。

【0054】

さらに、合成タスクL_iが環境音素片テンプレートT_iで構成される場合に、その繰返し周期を 1 で設定したが、繰返し周期を 1 ~ 1+100ms程度のランダムな周期に設定しても良い。こうすることで、同じ環境音素片テンプレートT_iが合成タスクL_iに登録された場合の不自然さが軽減される。

【0055】

本実施形態では、端末指示部213は、駆動パラメタD_iを入力とし、N個の環境音再生端末22-n全てに駆動パラメタD_iを分配しているが、環境音再生端末22-n毎に駆動パラメタD_iに対応する情報を分配する構成としてもよい。例えば、端末指示部213において、環境音再生端末32-n毎の合成タスク集合[L]を求め、出力する構成としてもよい。再生音合成部322-nでは、合成タスク集合[L]による合成音再生を継続すればよい。

【0056】

さらに、端末指示装置21側で各再生部223-nに対応する合成タスク集合[L]に基づき合成音を生成し、各再生部223-nに出力する構成としてもよい。

【0057】

<第三実施形態のポイント>

第二実施形態において環境音再生端末22-n上で一人の人物が出す環境音をひとつの合成タスクL_iとして管理できるようになったが、その結果、1つの環境音再生端末22-nで1.5人分に相当する環境音を合成するということが出来なくなり、N個の環境音再生端末22-n全体で環境音の音量を再現する精度が悪化する(端数処理による精度の悪化)という問題が新たに生じる。

【0058】

そこで本実施形態では、それぞれの環境音再生端末42-nにおいては一人の人物が出す環境音をひとつの合成タスクL_iとして管理して音を合成しつつ、確率的に合成タスクの増減を管理することによりN個の環境音再生端末42-n全体で再現する環境音の音量の精度を向上させる。

【0059】

<第三実施形態に係る環境音伝送システム>

第二実施形態と異なる部分を中心に説明する。

10

20

30

40

50

【0060】

以下に具体的な構成を記述する。

< 端末指示装置 4 1 の駆動パラメタ算出部 4 1 2 >

駆動パラメタ算出部 4 1 2 は、環境音量パラメタ P_i を入力とし、環境音量パラメタ P_i に基づき、環境音再生端末 4 2 - n 一台あたりで実現するべき環境音の人数 M_i を求め (S 4 1 2)、駆動パラメタ D_i として端末指示部 4 1 3 に出力する。

【0061】

本実施形態では環境音再生端末 4 2 - n のテンプレート記憶部 4 2 4 - n に記憶されている全環境音素片テンプレート T_1, T_2, \dots, T_L の平均エネルギー E_1, E_2, \dots, E_L の平均値 E が環境音素片テンプレート T_1, T_2, \dots, T_L の音量に対応する情報 G として保持されているとする。駆動パラメタ算出部 4 1 2 は、伝送元から出力された環境音量パラメタ P_i を入力とし (取得し)、環境音量パラメタ P_i から音量に対応する情報 (平均エネルギー E') を求める。さらに、各環境音再生端末 4 2 - n で実現するべき環境音の人数 M_i を、例えば、 $M_i = E' / (GN)$ によって計算し、駆動パラメタ D_i として出力する。

【0062】

< 端末指示装置 4 1 の端末指示部 4 1 3 >

端末指示部 4 1 3 は、駆動パラメタ $D_i (=M_i)$ を入力とし、N 個の環境音再生端末 4 2 - n 全てに駆動パラメタ D_i を分配する (S 4 1 3)。

【0063】

< 環境音再生端末 4 2 - n >

一人の人が出す環境音をひとつの合成タスク L_i として管理することにより、伝送先の会場内に会場元で計測された環境音の音量を自然に再現できる数の人間が伝送先に存在するかのよう自然な環境音を再現する。さらに、本実施形態では再現するべき音量に相当する人数が整数ではない場合も N 個の環境音再生端末 4 2 - n 全体で再現するべき音量に近づけるために、確率的に合成タスクの増減を行う。

【0064】

< 環境音再生端末 4 2 - n のテンプレート記憶部 4 2 4 - n >

本実施形態では、テンプレート記憶部 4 2 4 - n には、環境音素片テンプレート T_i とテンプレート記憶部 4 2 4 - n に記憶されている全環境音素片テンプレート T_1, T_2, \dots, T_L の平均エネルギー E_1, E_2, \dots, E_L の平均値 E が環境音素片テンプレート T_1, T_2, \dots, T_L の音量に対応する情報 G として保持されている。情報 G は、駆動パラメタ算出部 4 1 2 において、人数 $M_i (=E' / (GN))$ を求める際に利用される。

【0065】

< 環境音再生端末 4 2 - n の再生音合成部 4 2 2 - n >

再生音合成部 4 2 2 - n は、駆動パラメタ $D_i (=人数 M_i)$ を入力とし、駆動パラメタ D_i の整数部 $[D_i]$ によって表される音量の拍手を行うに必要な人数があたかもその装置の存在する位置で拍手をしているかのような再生音を発生し、駆動パラメタ D_i の少数部 d_i に相当する人数が N 個の環境音再生端末 4 2 - n 全体で実現されるように各環境音再生端末 4 2 - n で確率的に人数調整を行い、再生音を合成し (S 4 2 2)、出力する。

【0066】

第二実施形態に記載したように環境音素片テンプレート T_i によって一人の人物の環境音を再現する合成タスクを L_i とし、この L_i の数をコントロールすることで音量をコントロールする。現在再現中の人数を $|L|$ とする。 $|L|$ は、現在動作している合成タスク集合 $[L]$ に含まれる合成タスクの数と同じである。

【0067】

以下に $|L|$ を算出する具体的な動作例を示す。

(i) D_i の値を取得する。

(ii) $[D_i] = [D_{i-1}]$ で整数値が変わらない場合は (iv) へ、変わった場合は (iii) へ移行する。

(iii) 確率 $100 \cdot d_i / |L|$ で $|L| = [D_i] + 1$ とし、それ以外は $|L| = [D_i]$ とし、(iv) へ移行する。

(iv) 少数部分が増加 ($d_i - d_{i-1} > 0$) の場合は (v) へ、その他 ($d_i - d_{i-1} \leq 0$) の場合は (vi) へ移

10

20

30

40

50

行する。

(v) $j=[D_i]+1$ なら終了する。 $j=[D_i]$ の場合は

【数 2】

$$100 \frac{d_j - d_{j-1}}{1 - d_{j-1}} \%$$

の確率で $j=[D_i]+1$ として終了する。

【0068】

(vi) $j=[D_i]$ なら終了する。 $j=[D_i]+1$ の場合は

【数 3】

$$100 \frac{d_{j-1} - d_j}{d_j} \%$$

の確率で $j=[D_i]$ として終了する。

が算出された後に、合成タスク集合[L]を以下のように変える動作を行う。

から j が増加した場合は、ランダムに環境音素片テンプレートTを選択する。環境音素片テンプレートTをもとに合成タスクLを新たに合成し合成タスク集合[L]に加える。これを()回繰り返す。

から j が減少した場合は、現在の合成タスク集合[L]からランダムに合成タスクLを選択し、その動作を停止する。これを()回繰り返す。

の増減処理後も合成タスク集合[L]にもとづき引き続き再生音合成を行う。つまり、合成タスク集合[L]に含まれる k 個の合成タスクL($k=1,2, \dots, k$)にそれぞれ対応する環境音素片テンプレートTを(情報 k に対応する)その長さ毎に発生させて再生部 2 2 3 - nに出力する。

【0069】

<効果>

本実施形態によれば、環境音再生端末上で一人の人が出す環境音をひとつの合成タスクとして管理する場合でも、確率的に合成タスクの増減を管理することにより環境音再生端末全体で再現する環境音の音量の精度を向上させることができる。

【0070】

<変形例>

本実施形態では、端末指示部 4 1 3 は、駆動パラメタD_iを入力とし、N個の環境音再生端末 2 2 - n 全てに駆動パラメタD_iを分配しているが、環境音再生端末 2 2 - n 毎に駆動パラメタD_iに対応する情報を分配する構成としてもよい。例えば、端末指示部 4 1 3 において、環境音再生端末 4 2 - n 毎の j を求め、出力する構成としてもよい。再生音合成部 4 2 2 - n では、 j に基づき合成タスク集合[L]を更新し、合成音再生を継続すればよい。

【0071】

さらに、端末指示装置 4 1 側で各再生部 2 2 3 - n に対応する合成タスク集合[L]に基づき合成音を生成し、各再生部 2 2 3 - n に出力する構成としてもよい。

【0072】

<その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載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にあるいは個別に実行されてもよい。その他、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。

【0073】

<プログラム及び記録媒体>

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラム

10

20

30

40

50

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、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0074】

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。

【0075】

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てもよい。

【0076】

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、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部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施形態として、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い。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よって処理機能を実現する、いわゆるASP（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、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。

【0077】

また、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、各装置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 】

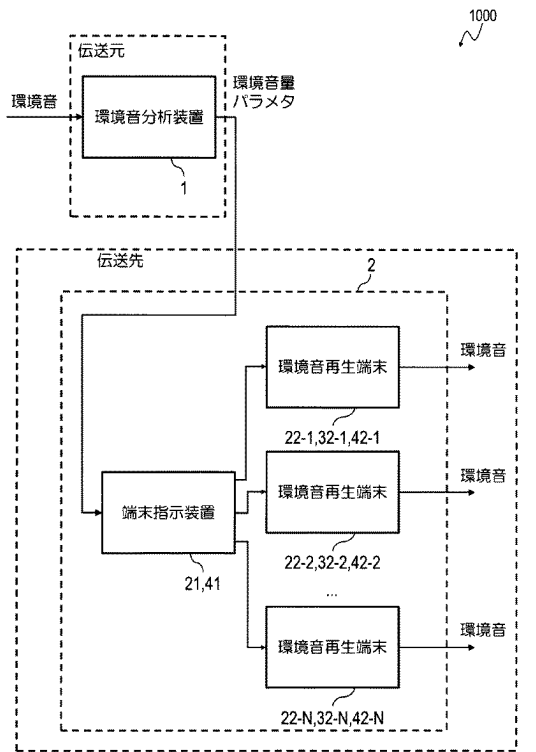


図1

【 図 2 】

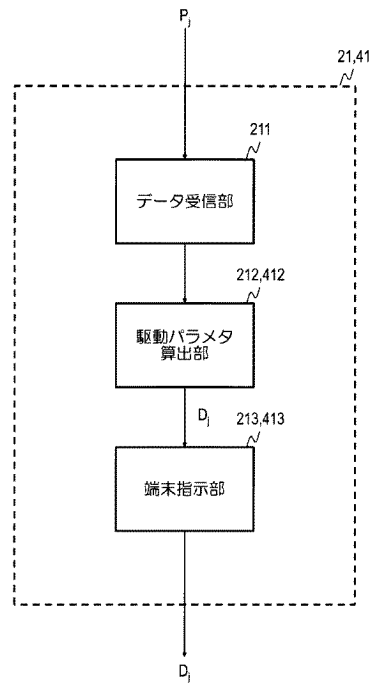


図2

【 図 3 】

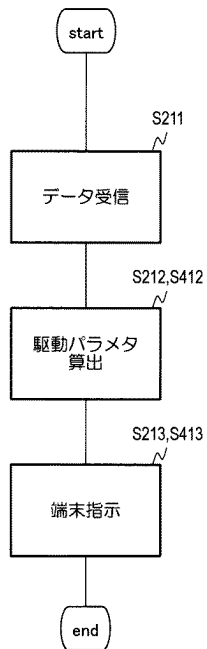


図3

【 図 4 】

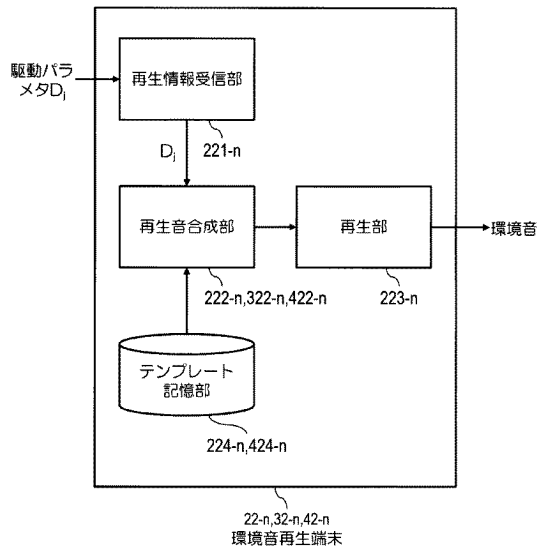


図4

【 図 5 】

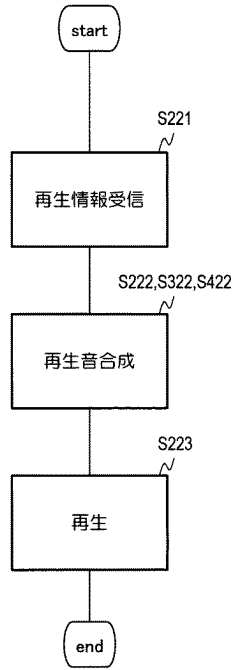


図5

【 図 6 】

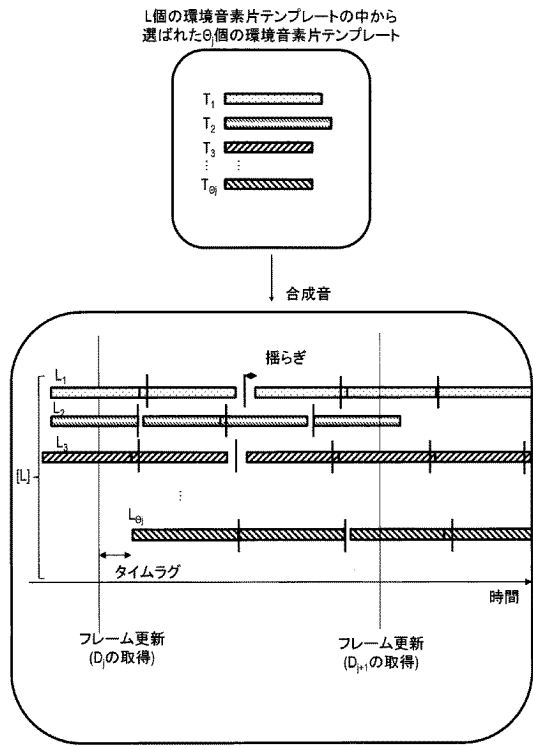


図6

フロントページの続き

- (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
- (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 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 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 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
- (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

審査官 渡邊 正宏

(56)参考文献 特開2012-093705(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10K 15/04
G10L 25/72
H04R 3/00 - 3/14