

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6095223号
(P6095223)

(45) 発行日 平成29年3月15日 (2017. 3. 15)

(24) 登録日 平成29年2月24日 (2017. 2. 24)

(51) Int. Cl. F I
G 1 O L 19/00 (2013. 01) G 1 O L 19/00 2 5 0
G 1 O L 13/02 (2013. 01) G 1 O L 13/02 1 1 0 Z

請求項の数 15 (全 26 頁)

(21) 出願番号	特願2013-169037 (P2013-169037)	(73) 特許権者	000004226
(22) 出願日	平成25年8月16日 (2013. 8. 16)		日本電信電話株式会社
(65) 公開番号	特開2014-63145 (P2014-63145A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成26年4月10日 (2014. 4. 10)	(73) 特許権者	504145342
審査請求日	平成27年7月24日 (2015. 7. 24)		国立大学法人九州大学
(31) 優先権主張番号	特願2012-186420 (P2012-186420)		福岡県福岡市東区箱崎六丁目10番1号
(32) 優先日	平成24年8月27日 (2012. 8. 27)	(74) 代理人	100121706
(33) 優先権主張国	日本国 (JP)		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二丁目3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】 環境音合成装置、環境音伝送システム、環境音合成方法、環境音伝送方法、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装置であって、

前記環境音分析装置から前記環境音量パラメタを受信するデータ受信部と、

一人の人間によって発生される1フレーム分の環境音のテンプレート(以下、環境音素片テンプレートという)を記憶するテンプレート記憶部と、

前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情報を記憶するテンプレート音量記憶部と、

前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記環境音量パラメタと前記環境音素片テンプレートの音量に基づいて定まる時間間隔ごとに配置して合成し、環境音を生成する音源合成部と、を備えることを特徴とする環境音合成装置。

【請求項2】

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装置であって、

前記環境音分析装置から前記環境音量パラメタを受信するデータ受信部と、

一人の人間による一拍分の環境音のテンプレート(以下、環境音素片テンプレートという)を記憶するテンプレート記憶部と、

前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情

10

20

報を記憶するテンプレート音量記憶部と、

前記環境音量パラメタの音量に対応する情報を前記環境音素片テンプレートの音量に対応する情報で除算した値を人数として出力する人数推定部と、

前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記人数が増えるに従って小さくなるように定められた時間間隔ごとに配置して合成し、環境音を生成する音源合成部と、
を備えることを特徴とする環境音合成装置。

【請求項 3】

請求項 2 に記載の環境音合成装置であって、

前記テンプレート記憶部に、複数種類の環境音素片テンプレートと各環境音素片テンプレートの出力確率が対応付けて記憶され、

前記音源合成部は、

前記テンプレート記憶部から環境音素片テンプレートを前記出力確率に応じて 1 以上選択して合成すること

を特徴とする環境音合成装置。

【請求項 4】

環境音分析装置と、環境音合成装置とを備える環境音伝送システムであって、

前記環境音分析装置は、

音響信号を取得して、当該音響信号の音量に対応する値を計算する音量計算部と、

前記音響信号の音量に対応する値を取得して、当該音響信号の音量に対応する値を量子化して、そのインデックスを環境音量パラメタとして出力するパラメタ変換部と、

前記環境音量パラメタを前記環境音合成装置に送信するデータ送信部とを備え、

前記環境音合成装置は、

前記環境音分析装置から前記環境音量パラメタを受信するデータ受信部と、

一人の人間によって発生される 1 フレーム分の環境音のテンプレート（以下、環境音素片テンプレートという）を記憶するテンプレート記憶部と、

前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情報を記憶するテンプレート音量記憶部と、

前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記環境音量パラメタと前記環境音素片テンプレートの音量に基づいて定まる時間間隔ごとに配置して合成し、環境音を生成する音源合成部と、
を備えることを特徴とする環境音伝送システム。

【請求項 5】

環境音分析装置と、環境音合成装置とを備える環境音伝送システムであって、

前記環境音分析装置は、

音響信号を取得して、当該音響信号の音量に対応する値を計算する音量計算部と、

前記音響信号の音量に対応する値を取得して、当該音響信号の音量に対応する値を量子化して、そのインデックスを環境音量パラメタとして出力するパラメタ変換部と、

前記環境音量パラメタを前記環境音合成装置に送信するデータ送信部とを備え、

前記環境音合成装置は、

前記環境音分析装置から前記環境音量パラメタを受信するデータ受信部と、

一人の人間による一拍分の環境音のテンプレート（以下、環境音素片テンプレートという）を記憶するテンプレート記憶部と、

前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情報を記憶するテンプレート音量記憶部と、

前記環境音量パラメタの音量に対応する情報を前記環境音素片テンプレートの音量に対応する情報で除算した値を人数として出力する人数推定部と、

前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記人数が増えるに従って小さくなるように定められた時間間隔ごとに配置して合成し、環境音を生成する音源合成部と、

10

20

30

40

50

を備えることを特徴とする環境音伝送システム。

【請求項 6】

請求項 4 または 5 に記載の環境音伝送システムであって、
前記環境音分析装置のパラメタ変換部が、
前記音響信号の音量に対応する値として平均エネルギー E_j (j はフレームのインデックスを表す 1 以上の整数) を計算し、
前記平均エネルギー E_j から

【数 1 2】

$$F_j = \begin{cases} -\sqrt{E_j} - 1 & \left(\sqrt{E_j} \text{ が奇数の場合} \right) \\ \sqrt{E_j} & \left(\sqrt{E_j} \text{ が偶数の場合} \right) \end{cases}$$

10

で表される列 F_j を計算し、前記列 F_j を 2 回、2 で割って値 G_j を求め、前記値 G_j を符号に変換して前記値 G_j の各々にシンボルを割り当て、前記値 G_j に割り当てられたシンボルを j 番目のフレームのパラメタ P_j として出力することを特徴とする環境音伝送システム。

【請求項 7】

請求項 5 に記載の環境音伝送システムであって、
前記環境音分析装置が、
前記音響信号が拍手音であるか否かを判別し、判別結果が拍手音でない場合には、環境音量パラメタ = 0、および周期なしを示す周期情報を出力し、前記判別結果が拍手音である場合には、拍手音を示すフラグを出力する拍手区間検出部と、

20

前記拍手音を示すフラグを取得して、自己相関関数を用いて前記音響信号の周期を検出して、前記周期が検出された場合には自己相関関数のピーク間隔を周期情報として出力し、前記周期が検出されない場合には周期なしを示す周期情報を出力する周期性判定部とをさらに備え、

前記データ送信部が、前記環境音量パラメタとともに前記周期情報を前記環境音合成装置に送信し、

前記データ受信部が、前記環境音量パラメタとともに前記周期情報を前記環境音分析装置から受信し、

前記音源合成部が、前記環境音量パラメタが 0 でなく前記周期情報が周期なしを示す場合に、前記音響信号が拍手音であると判定し、前記環境音量パラメタが 0 でなく前記周期情報が周期ありを示す場合に、前記音響信号が手拍子音であると判定して、前記拍手音と判定された場合と比較して前記手拍子音と判定された場合の前記時間間隔が長くなるように前記環境音素片テンプレートを配置して合成すること

30

を特徴とする環境音伝送システム。

【請求項 8】

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、

前記環境音量パラメタを受信するデータ受信ステップと、

一人の人間によって発生される 1 フレーム分の環境音のテンプレート (以下、環境音素片テンプレートという) を記憶するテンプレート記憶部と、前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情報を記憶するテンプレート音量記憶部とを参照して、前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記環境音量パラメタと前記環境音素片テンプレートの音量に基づいて定まる時間間隔ごとに配置して合成し、環境音を生成する音源合成ステップと、

40

を備えることを特徴とする環境音合成方法。

【請求項 9】

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、

50

前記環境音量パラメタを受信するデータ受信ステップと、
 一人の人間による一拍分の環境音のテンプレート（以下、環境音素片テンプレートという）の音量に対応する情報を記憶するテンプレート音量記憶部を参照して、前記環境音量パラメタの音量に対応する情報を前記環境音素片テンプレートの音量に対応する情報で除算した値を人数として出力する人数推定ステップと、
 前記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を参照して、前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記人数が増えるに従って小さくなるように定められた時間間隔ごとに配置して合成し、環境音を生成する音源合成ステップと、
 を備えることを特徴とする環境音合成方法。

10

【請求項10】

環境音分析装置と、環境音合成装置とが実行する環境音伝送方法であって、
 前記環境音分析装置は、
 音響信号を取得して、当該音響信号の音量に対応する値を計算する音量計算ステップと、
 、
 前記音響信号の音量に対応する値を取得して、当該音響信号の音量に対応する値を量子化して、そのインデックスを環境音量パラメタとして出力するパラメタ変換ステップと、
 前記環境音量パラメタを前記環境音合成装置に送信するデータ送信ステップとを実行し、

20

前記環境音合成装置は、
 前記環境音分析装置から前記環境音量パラメタを受信するデータ受信ステップと、
一人の人間によって発生される1フレーム分の環境音のテンプレート（以下、環境音素片テンプレートという）を記憶するテンプレート記憶部と、前記テンプレート記憶部に記憶されている環境音素片テンプレートの音量に対応する情報を記憶するテンプレート音量記憶部とを参照して、前記テンプレート記憶部から前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記環境音量パラメタと前記環境音素片テンプレートの音量に基づいて定まる時間間隔ごとに配置して合成し、環境音を生成する音源合成ステップと、
 を実行することを特徴とする環境音伝送方法。

30

【請求項11】

環境音分析装置と、環境音合成装置とが実行する環境音伝送方法であって、
 前記環境音分析装置は、
 音響信号を取得して、当該音響信号の音量に対応する値を計算する音量計算ステップと、
 、
 前記音響信号の音量に対応する値を取得して、当該音響信号の音量に対応する値を量子化して、そのインデックスを環境音量パラメタとして出力するパラメタ変換ステップと、
 前記環境音量パラメタを前記環境音合成装置に送信するデータ送信ステップとを実行し、

40

前記環境音合成装置は、
 前記環境音分析装置から前記環境音量パラメタを受信するデータ受信ステップと、
 一人の人間による一拍分の環境音のテンプレート（以下、環境音素片テンプレートという）の音量に対応する情報を記憶するテンプレート音量記憶部を参照し、前記環境音量パラメタの音量に対応する情報を前記環境音素片テンプレートの音量に対応する情報で除算した値を人数として出力する人数推定ステップと、
 前記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を参照して、前記環境音素片テンプレートを選択し、当該選択された環境音素片テンプレートを、前記人数が増えるに従って小さくなるように定められた時間間隔ごとに配置して合成し、環境音を生成する音源合成ステップと、
 を実行することを特徴とする環境音伝送方法。

50

【請求項12】

請求項 10 または 11 に記載の環境音伝送方法であって、
 前記環境音分析装置が実行するパラメタ変換ステップにおいて、
 前記音響信号の音量に対応する値として平均エネルギー E_j (j はフレームのインデックスを表す 1 以上の整数) を計算し、
 前記平均エネルギー E_j から

【数 13】

$$F_j = \begin{cases} -\sqrt{E_j} - 1 & \left(\sqrt{E_j} \text{が奇数の場合} \right) \\ \sqrt{E_j} & \left(\sqrt{E_j} \text{が偶数の場合} \right) \end{cases}$$

で表される列 F_j を計算し、前記列 F_j を 2 回、2 で割って値 G_j を求め、前記値 G_j を符号に変換して前記値 G_j の各々にシンボルを割り当て、前記値 G_j に割り当てられたシンボルを j 番目のフレームのパラメタ P_j として出力すること
 を特徴とする環境音伝送方法。 10

【請求項 13】

請求項 11 に記載の環境音伝送方法であって、
 前記環境音分析装置が、
 前記音響信号が拍手音であるか否かを判別し、判別結果が拍手音でない場合には、環境音量パラメタ = 0、および周期なしを示す周期情報を出力し、前記判別結果が拍手音である場合には、拍手音を示すフラグを出力する拍手区間検出ステップと、
 前記拍手音を示すフラグを取得して、自己相関関数を用いて前記音響信号の周期を検出して、前記周期が検出された場合には自己相関関数のピーク間隔を周期情報として出力し、前記周期が検出されない場合には周期なしを示す周期情報を出力する周期性判定ステップとをさらに実行し、 20

前記データ送信ステップにおいて、前記環境音量パラメタとともに前記周期情報を前記環境音合成装置に送信し、

前記データ受信ステップにおいて、前記環境音量パラメタとともに前記周期情報を前記環境音分析装置から受信し、

前記音源合成ステップにおいて、前記環境音量パラメタが 0 でなく前記周期情報が周期なしを示す場合に、前記音響信号が拍手音であると判定し、前記環境音量パラメタが 0 でなく前記周期情報が周期ありを示す場合に、前記音響信号が手拍子音であると判定して、前記拍手音と判定された場合と比較して前記手拍子音と判定された場合の前記時間間隔が長くなるように前記環境音素片テンプレートを配置して合成すること
 を特徴とする環境音伝送方法。 30

【請求項 14】

コンピュータを、請求項 1 から 3 のいずれかに記載の環境音合成装置として機能させるためのプログラム。

【請求項 15】

コンピュータを、請求項 4 から 7 の何れかに記載の環境音伝送システムとして機能させるためのプログラム。

【発明の詳細な説明】 40

【技術分野】

【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装置、環境音伝送システム、環境音合成方法、環境音伝送方法、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 50

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。

【先行技術文献】

【非特許文献】

【0003】

【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.

10

【発明の概要】

【発明が解決しようとする課題】

【0004】

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。そこで本発明では、伝送元において收音され

20

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

を再現することができる環境音合成装置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】

【0005】

本発明の環境音合成装置は、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ことを特徴とし、データ受信部と、テンプレート記憶部と、音源合成部とを備える。

【0006】

データ受信部は、環境音分析装置から環境音量パラメタを受信する。テンプレート記憶部は、1フレーム分の環境音のテンプレート（以下、テンプレートという）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する。音源合成部は、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。

30

【発明の効果】

【0007】

本発明の環境音合成装置によれば、伝送元において收音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を再現することができる。

【図面の簡単な説明】

【0008】

40

【図1】本発明の環境音伝送システムの構成例を示すブロック図。

【図2】実施例1の環境音分析装置の構成を示すブロック図。

【図3】実施例1の環境音分析装置の動作を示すフローチャート。

【図4】実施例2の環境音分析装置の構成を示すブロック図。

【図5】実施例2の環境音分析装置の動作を示すフローチャート。

【図6】実施例2のパラメタ変換部のパラメタ生成手順を例示する図。

【図7】実施例2の変形例1の環境音分析装置の構成を示すブロック図。

【図8】実施例2の変形例1の環境音分析装置の動作を示すフローチャート。

【図9】実施例3の環境音合成装置の構成を示すブロック図。

【図10】実施例3の環境音合成装置の動作を示すフローチャート。

50

- 【図 1 1】実施例 4 の環境音合成装置の構成を示すブロック図。
 【図 1 2】実施例 4 の環境音合成装置の動作を示すフローチャート。
 【図 1 3】実施例 4 の音源合成部の環境音素片テンプレート合成手順を例示する図。
 【図 1 4】実施例 5 及びその変形例の環境音分析装置の構成を示すブロック図。
 【図 1 5】実施例 5 及びその変形例の環境音分析装置の動作を示すフローチャート。
 【図 1 6】拍手音や手拍子音の周波数成分の時間変化を例示する図。
 【図 1 7】実施例 6 の環境音合成装置の構成を示すブロック図。
 【図 1 8】実施例 6 の環境音合成装置の動作を示すフローチャート。
 【図 1 9】実施例 7、8 の環境音合成装置の構成を示すブロック図。
 【図 2 0】実施例 7、8 の環境音合成装置の動作を示すフローチャート。
 【図 2 1】テンプレート記憶部に環境音素片テンプレートと出力確率とを対応付けて記憶する例を例示する図。

10

【発明を実施するための形態】

【0009】

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。

【0010】

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。本発明では、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伝送する。そして、伝送先では予め記憶された環境音のテンプレートを、音量を表す情報に応じて変換することにより、伝送元の環境音（に類似した音）を再生する。

20

【0011】

また、拍手や手拍子音の一拍（一度両手を合わせて打つこと）は、音響パワーの個人差が小さい。また、一拍と一拍の時間間隔（以下、拍手間隔ともいう）の個人差も小さく、200ms～300ms程度である。したがって、ある人の拍手音（一拍分）を環境音素片テンプレートとして用意しておき、それを個人差に応じたゆらぎ（200ms～300ms）を持たせた間隔で繰り返し再生することにより、別の人の拍手音に類似した音を構成することができる。

【0012】

<環境音伝送システム>

30

以下、図 1 を参照して本発明の環境音伝送システムについて説明する。図 1 は本発明の環境音伝送システムの構成例を示すブロック図である。図 1 に示すように、本発明の環境音伝送システムは、伝送元の環境音分析装置と、伝送先の環境音合成装置から構成される。環境音分析装置は、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_j 、以下単にパラメタともいう）を抽出し、出力する。環境音合成装置は、予め記憶された環境音のテンプレートを用いて、入力された環境音量パラメタ P_j に合わせてテンプレートを変換することにより環境音を合成し、出力する。以下、実施例 1 において環境音分析装置 1、実施例 2 において環境音分析装置 2、実施例 2 の変形例 1 において環境音分析装置 2'、実施例 3 において環境音合成装置 3、実施例 4 において環境音合成装置 4、実施例 5 において環境音分析装置 5、実施例 6 において環境音合成装置 6、実施例 7 において環境音合成装置 7、実施例 8 において環境音合成装置 8、実施例 9 において環境音合成装置 9、実施例 10 において環境音合成装置 10 をそれぞれ説明する。また、環境音分析装置 1 と環境音合成装置 3 との組み合わせを環境音伝送システム 1000、環境音分析装置 2 と環境音合成装置 3 との組み合わせを環境音伝送システム 2000、環境音分析装置 2' と環境音合成装置 3 との組み合わせを環境音伝送システム 2000'、環境音分析装置 1 と環境音合成装置 4 との組み合わせを環境音伝送システム 3000、環境音分析装置 2 と環境音合成装置 4 との組み合わせを環境音伝送システム 4000、環境音分析装置 2' と環境音合成装置 4 との組み合わせを環境音伝送システム 4000'、環境音分析装置 5 と環境音合成装置 6 との組み合わせを環境音伝送システム 5000 と呼ぶ。

40

【実施例 1】

50

【 0 0 1 3 】

以下、図 2、図 3 を参照して本発明の実施例 1 の環境音分析装置について説明する。図 2 は本実施例の環境音分析装置 1 の構成を示すブロック図である。図 3 は本実施例の環境音分析装置 1 の動作を示すフローチャートである。図 2 に示すように、本実施例の環境音分析装置 1 は、收音部 1 1 と、音量計算部 1 2 と、パラメタ変換部 1 3 と、データ送信部 1 4 とを備える。

【 0 0 1 4 】

< 收音部 1 1 >

收音部 1 1 は伝送元の音を收音する (S 1 1)。ここでは、收音部 1 1 には伝送元の拍手音が入力されるものとする。

10

【 0 0 1 5 】

< 音量計算部 1 2 >

音量計算部 1 2 は、拍手音の音響信号を取得する。音量計算部 1 2 が取得する拍手音の音響信号は、所定のサンプリング周波数でサンプリングされた信号列とする。ここで、 X_j を第 j フレームの音響信号とし、 $X_j = (x_j(1), x_j(2), \dots, x_j(N))$ (N はフレームあたりのサンプル数) とする。例えば 8 k H z サンプリングのときに 1 フレーム 2 0 m s とすると、 $N = 1 6 0$ である。なお、遅延が短い方が良ければフレームの長さを短くし、遅延が長くなっても良ければ、フレームの長さを長くすれば良い。音量計算部 1 2 は、フレーム毎に、入力された拍手音の音響信号の音量に対応する値 (以下、「拍手音量に対応する値」ともいう) を求めて出力する。具体的には、音量計算部 1 2 は、フレーム毎に、入力された拍手音の音響信号 $X_j = (x_j(1), x_j(2), \dots, x_j(N))$ の平均エネルギー

20

【 0 0 1 6 】

【数 1】

$$E_j = \frac{x_j(1)^2 + x_j(2)^2 + \dots + x_j(N)^2}{N}$$

【 0 0 1 7 】

を計算する (S 1 2)。

【 0 0 1 8 】

< パラメタ変換部 1 3 >

パラメタ変換部 1 3 は、音量計算部 1 2 から出力された拍手音量に対応する値を取得する。パラメタ変換部 1 3 は、取得した拍手音量に対応する値を量子化し、環境音量パラメタを出力する。具体的には、パラメタ変換部 1 3 は、平均エネルギー E_j の取りうる範囲 (例えば $x_j(i)$ ($i = 1, 2, \dots, N$) が符号付き 1 6 b i t の場合は最小値が 0 で最大値が 2^30 となる) をあらかじめ定められた場合の数 (例えば 1 6 b i t) に量子化し、そのインデックスを環境音量パラメタ P_j として出力する (S 1 3)。

30

【 0 0 1 9 】

< データ送信部 1 4 >

データ送信部 1 4 は、パラメタ変換部 1 3 が出力した環境音量パラメタ P_j を伝送先の環境音合成装置 3 (または 4) に送信する (S 1 4)。環境音合成装置 3 については実施例 3 に、環境音合成装置 4 については実施例 4 に記載する。

40

【 0 0 2 0 】

このように、本実施例の環境音分析装置 1 によれば、伝送元において收音された拍手音を効率よく低遅延に伝送することができる。

【 0 0 2 1 】

[実施例 1 の動作例 2]

上述の実施例 1 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、伝送元で收音され

50

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。

【0022】

実施例1の動作例2における環境音分析装置1は、環境音分析装置1の收音部11、音量計算部12、パラメタ変換部13、データ送信部14の各部で取り扱われる拍手音および拍手音量が、環境音及び環境音の音量に置き換わる点を除いては、上述の動作例と同じである。

【0023】

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を決定的な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、伝送元において收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。

【実施例2】

【0024】

以下、図4、図5、図6を参照して本発明の実施例2の環境音分析装置について説明する。図4は本実施例の環境音分析装置2の構成を示すブロック図である。図5は本実施例の環境音分析装置2の動作を示すフローチャートである。図6は本実施例のパラメタ変換部23のパラメタ生成手順を例示する図である。図4に示すように、本実施例の環境音分析装置2は、收音部11と、音量計算部12と、パラメタ変換部23と、データ送信部14とを備える。收音部11、音量計算部12、データ送信部14は実施例1の環境音分析装置1における同一番号の各構成部と同じであるから説明を適宜略する。

【0025】

<音量計算部12>

音量計算部12は、48kHzサンプリングでサンプリングされた信号列であり、1フレーム6サンプル(N=6)で構成される信号列 $X_j = (x_j(1), x_j(2), \dots, x_j(6))$ を取得する。音量計算部12は、フレーム毎に、入力された拍手音響信号 $X_j = (x_j(1), x_j(2), \dots, x_j(6))$ から、平均エネルギー

【0026】

【数2】

$$E_j = \frac{x_j(1)^2 + x_j(2)^2 + \dots + x_j(6)^2}{6}$$

【0027】

を計算する(S12)。

【0028】

<パラメタ変換部23>

パラメタ変換部23は、求めた平均エネルギー E_j を以下の式により変形した列 F_j を求める。

【0029】

【数3】

$$F_j = \begin{cases} -\sqrt{E_j} - 1 & \left(\sqrt{E_j} \text{が奇数の場合} \right) \\ \sqrt{E_j} & \left(\sqrt{E_j} \text{が偶数の場合} \right) \end{cases} \dots (1)$$

【0030】

つまり、図6に示すように、ガウス関数や床関数により整数値化された F_j の取りうる値(0~32768)のうち、奇数の値に負の符号を与え、さらに1を減じる。これにより、 F_j はすべて偶数の値を取るようになる。次に、全部偶数になった F_j の各々を2で

10

20

30

40

50

割る（右に1ビットシフトでも構わない）。この値をG.711準拠の範囲に収めるために、μ-lawを使うのであればさらに2で割り（右に1ビットシフトでもよい）値G_jを求める。そして、Gに対して、ITU-T G.711の符号化処理を行い、G_jをG.711の符号（番号）に変換する。48kHzサンプリング6サンプル分を一塊（1フレーム）にすると8kHzの1サンプル分に相当するので、上記のG_j毎にG.711のシンボル1つを割り当てることができる。割り当てられたシンボル列をパラメタP_jとして出力する（S23）。パラメタP_jは通常の音声と同様に固定電話回線を用いて伝送すると遅延が短くすむ。式（1）のかわりに式（2）のように対数を用いてもよい。

【0031】

【数4】

$$F_j = \begin{cases} -\log E_j - 1 & (\log E_j \text{が奇数の場合}) \\ \log E_j & (\log E_j \text{が偶数の場合}) \end{cases} \dots (2)$$

【0032】

また、平方根演算や対数演算は多項式近似（テイラー展開など）で演算量を削減してもよい。

【0033】

[実施例2の変形例1]

以下、図7、図8を参照して実施例2のパラメタ変換部23に変更を加えた変形例1の環境音分析装置について説明する。図7は本変形例の環境音分析装置2'の構成を示すブロック図である。図8は本変形例の環境音分析装置2'の動作を示すフローチャートである。図7に示すように、本変形例の環境音分析装置2'は、收音部11と、音量計算部12と、パラメタ変換部23'と、データ送信部14とを備える。收音部11、音量計算部12、データ送信部14は実施例2の環境音分析装置2における同一番号の各構成部と同じであるから説明を適宜略する。

【0034】

<パラメタ変換部23'>

パラメタ変換部23'は、図6のようなマッピング演算の代わりに、F_jの取りうる0~32768の値を直接8ビットのシンボルにマッピングするマッピングテーブル23Aを予め備えており、マッピングテーブル23Aを参照してパラメタP_jを求める（S23'）。または、パラメタ変換部23'は、F_jの取りうる0~32768の値をあらかじめビットシフト等により場合の数を減らしてから、マッピングテーブル23Aを用いてパラメタP_jを求めてもよい。この場合はマッピングテーブル23Aの大きさを削減できる。F_jはデシベル単位に変換したものをを用いてもよい。

【0035】

実施例2及び変形例1の環境音分析装置は以下の効果を有する。收音された拍手音の音響信号は正の値となるため、E_jの平方根の値の取りうる範囲は正の整数値、例えばx_j(n)(n=1, 2, ..., N)が符号付き16ビットの場合は最小値が0で最大値が32768となる。このまま、パラメタ変換部でITU-T G.711の符号化を行うと、符号化効率が悪くなるという問題がある。上記式（1）の変形を行うと、例えばx_j(n)(n=1, 2, ..., N)が符号付き16ビットの場合は、F_jの取りうる範囲は-16384から16384になる。そこで、パラメタ変換部においてE_jの取りうる範囲が負の整数値から正の整数値の範囲となるように変換した値F_jを用いることにより、符号化効率を向上させることができ、パラメタP_jの情報量を削減することができる。つまり、伝送遅延をより少なくすることが可能となる。

【0036】

[実施例2の動作例2]

上述の実施例2および実施例2の変形例1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2(2')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

10

20

30

40

50

としても良いし、伝送元で収録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。

【0037】

実施例2の動作例2においては、環境音分析装置2（2'）の收音部11、音量計算部12、パラメタ変換部23または23'、データ送信部14の各部で取り扱われる拍手音および拍手音量が、環境音及び環境音の音量に置き換わる点を除いては、上述の動作例と同じである。

【実施例3】

【0038】

以下、図9、図10を参照して本発明の実施例3の環境音合成装置について説明する。図9は本実施例の環境音合成装置3の構成を示すブロック図である。図10は本実施例の環境音合成装置3の動作を示すフローチャートである。図9に示すように、本実施例の環境音合成装置3は、データ受信部31と、音源合成部32と、テンプレート記憶部33と、再生部34とを備える。環境音合成装置3は環境音分析装置1（2、2'）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装置である。以下、実施例1、2で詳述した動作例に従い、環境音の例として拍手音を用いて説明を進める。

10

【0039】

<データ受信部31>

データ受信部31は、環境音分析装置から環境音量パラメタ P_j を受信する（S31）

20

【0040】

<テンプレート記憶部33>

テンプレート記憶部33には、拍手音の各音量バリエーションに対して複数の拍手音（1フレーム分）のテンプレートが記憶されている。つまり、テンプレート記憶部33には、 i をフレームのインデックスとした場合に、1フレーム分の拍手音を含む環境音のテンプレート T_i と当該テンプレートの環境音の音量に対応する情報 E'_i とが対応付けて記憶されているものとする。なお、テンプレートの環境音の音量に対応する値は、各テンプレート T_i を入力として、上記実施例1または2の音量計算部12及びパラメタ変換部13（23）と同じ方法により求めることができる。なお、実施例1または2のどの方法を用いるかは、環境音分析装置と環境音合成装置との間で統一しておくものとする。

30

【0041】

<音源合成部32>

音源合成部32は、入力された環境音量パラメタ P_j と同じ音量大きさのテンプレートのうちいずれか1つをテンプレート記憶部33からランダムに選択する。つまり、 $P_j = E'_i$ を満たす E'_i に対応づけられているテンプレート T_i のうち、いずれか1つをランダムに選択する。音源合成部32は、選択したテンプレートを、必要に応じて前のフレームと補間をして、1フレーム分の音響信号を合成して環境音（この動作例では拍手音）を生成する（S32）。例えば、20msのフレームあたり環境音量パラメタに8bitのバリエーションがあったとすると、400bit/secで拍手音を伝送できる。

40

【0042】

<再生部34>

再生部34は、音源合成部32が合成した拍手音を再生する（S34）。

【0043】

このように、本実施例の環境音合成装置3によれば、テンプレート記憶部33に拍手音の各音量バリエーションに対して複数のテンプレートを保持しておき、音源合成部32が音量の条件を充たす複数のテンプレートから1つのテンプレートをランダムに選択するため、合成された拍手音が定常的なパターンとして聞こえないようにすることができる。

【0044】

[実施例3の動作例2]

50

実施例 3 では、伝送元の環境音の例として拍手音を対象とし、伝送元の拍手音の音量に関するパラメタを取得して、伝送先で拍手音を生成する環境音合成装置 3 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、声援や掛け声や、伝送元で収録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、伝送元の環境音量パラメタが入力され、伝送先で環境音を合成してもよい。

【 0 0 4 5 】

実施例 3 の動作例 2 では、実施例 3 の環境音合成装置 3 のデータ受信部 3 1 と、音源合成部 3 2 と、テンプレート記憶部 3 3 と、再生部 3 4 において、拍手音が環境音に置き換わる点を除いては、上述の動作例と同じである。

【実施例 4】

【 0 0 4 6 】

以下、図 1 1、図 1 2、図 1 3 を参照して本発明の実施例 4 の環境音合成装置について説明する。図 1 1 は本実施例の環境音合成装置 4 の構成を示すブロック図である。図 1 2 は本実施例の環境音合成装置 4 の動作を示すフローチャートである。図 1 3 は本実施例の音源合成部 4 2 の環境音素片テンプレート合成手順を例示する図である。図 1 1 に示すように、本実施例の環境音合成装置 4 は、データ受信部 3 1 と、音源合成部 4 2 と、テンプレート記憶部 4 3 と、再生部 3 4 と、人数推定部 4 5 と、テンプレート音量記憶部 4 6 とを備える。データ受信部 3 1、再生部 3 4 は実施例 3 の環境音合成装置 3 における同一番号の各構成部と同じであるから説明を省略する。

【 0 0 4 7 】

<テンプレート記憶部 4 3 >

テンプレート記憶部 4 3 には、一人の人間による一拍分の拍手音（300ms 程度）のテンプレートの複数のバリエーションが記憶されている。本実施例では環境音の例として拍手音を扱うため、拍手音のテンプレートを環境音素片テンプレートのバリエーションのひとつとする。従って、以下では拍手音のテンプレートを環境音素片テンプレートともいう。例えば、異なる人の一拍分の拍手音をそれぞれ異なる環境音素片テンプレートとして記憶しておく。以下、単にテンプレートという場合には、所定フレーム長の複数人による拍手音（環境音）全体を収録したテンプレートを指すものとし、環境音素片テンプレートという場合には、一人の人間による一拍分の拍手音（環境音）のテンプレートを指すものとする。

【 0 0 4 8 】

<テンプレート音量記憶部 4 6 >

テンプレート音量記憶部 4 6 には、テンプレート記憶部 4 3 に記憶されている環境音素片テンプレートの音量に対応する情報（具体的には、実施例 1 または 2 の音量計算部 1 2 により計算される、平均エネルギー）が記憶されている。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 4 3 に記憶されている環境音素片テンプレートのいずれか一つについて計算された平均エネルギーを環境音素片テンプレートの音量に対応する情報として記憶しておいてもよい。また、テンプレート記憶部 4 3 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を、環境音素片テンプレートの音量に対応する情報としてテンプレート音量記憶部 4 6 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としてテンプレート音量記憶部 4 6 に記憶しておいてもよい。

【 0 0 4 9 】

なお、テンプレート音量記憶部 4 6 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 4 3 からランダムに選択した環境音素片テンプレートについて計算した平均エネルギーを環境音素片テンプレートの音量に対応する情報として用いてもよい。

【 0 0 5 0 】

<人数推定部 4 5 >

人数推定部 4 5 は、環境音量パラメタ P_j に応じて音量のゲイン調整を行うための構成である。人数推定部 4 5 は、伝送元から出力された環境音量パラメタ P_j を取得し、当該環境音量パラメタ P_j から音量に対応する情報 E'_j を求める。具体的には、実施例 1 または 2 のパラメタ変換部 1 3 (2 3) と逆の処理を行うことにより、音量に対応する情報 E'_j を得る。人数推定部 4 5 は、音量に対応する情報 E'_j を環境音素片テンプレートの音量に対応する情報で除算した値の整数値 (小数点以下を四捨五入、または切り捨てた値) を拍手の人数 M として出力する (S 4 5)。

【 0 0 5 1 】

< 音源合成部 4 2 >

音源合成部 4 2 は、テンプレート記憶部 4 3 から環境音素片テンプレートをランダムに選択して合成する (S 4 2)。環境音量パラメタにより一人分の拍手を合成する場合 (人数推定部 4 5 において $M = 1$ となった場合) は、図 1 3 A のように、約 3 0 0 m s ごとにランダムに選択された環境音素片テンプレート T_i を用いて合成した波形を拍手音として出力する。前述のように合成の時間間隔は約 3 0 0 m s でよいが、より好ましくは 3 0 0 m s を中心として時間間隔に揺らぎを持たせてもよい。時間間隔に揺らぎを持たせることによってさらに自然な拍手音を合成することができる。たとえば 3 0 0 m s を中心としてガウス分布にしたがう乱数により、 \pm 数 1 0 m s の揺らぎを持たせればよい。例えば音源合成部 4 2 は

【 0 0 5 2 】

【 数 5 】

$$Y_i = (i \cdot \tau + \sigma_i) \cdot T_i \quad (i = 0, 1, 2, \dots) \quad \dots(3)$$

【 0 0 5 3 】

によりテンプレートを変換した拍手音 Y_i ($i = 0, 1, 2, \dots$) を出力する (S 4 2)。別の表現方法で書くと、時系列テンプレート信号 $T_i = (t_i [1] \quad t_i [2] \quad \dots \quad t_i [P])$ と拍手タイミングを表すインパルス $(i \cdot \tau + \sigma_i)$ の畳み込み演算で Y_i を出力とする。

【 0 0 5 4 】

【 数 6 】

$$Y_i = \delta(i \cdot \tau + \sigma_i) * T_i \quad (i = 0, 1, 2, \dots) \quad \dots(3')$$

【 0 0 5 5 】

ここで * は畳み込み演算を表す。ここで、 $\tau = 3 0 0$ m s であり、 σ_i は $- 1 0$ m s $\sigma_i + 1 0$ m s の範囲で生成した乱数である。環境音量パラメタにより M 人分の拍手を合成する場合は、図 1 3 B のように、時間間隔を約 3 0 0 / M (m s) ごとにランダムに選択された環境音素片テンプレートを用いて合成された波形を拍手音として出力する。人数 M の逆数を使って、時間間隔を約 3 0 0 / M (m s) と設定することで、拍手の人数 M が増えるに従って時間間隔が小さくなるように設定することができる。この場合もガウス分布やラプラス分布に従う乱数によって、揺らぎを持たせることができる。例えば音源合成部 4 2 は、

【 0 0 5 6 】

【 数 7 】

$$Y_i = \sum_{m=1}^M \left(\frac{1}{M} m \cdot \tau + \sigma_m \right) \cdot T_i \quad (i = 0, 1, 2, \dots) \quad \dots(4)$$

【 0 0 5 7 】

によりテンプレートを変換した環境音 Y_i ($i = 0, 1, 2, \dots$) を出力する (S 4 2)。

【 0 0 5 8 】

このように、本実施例の環境音合成装置 4 によれば、実施例 3 のように音量ごとにテン

10

20

30

40

50

プレートを用意しておく必要がなく、テンプレート記憶部 4 3 に記憶しておく環境音素片テンプレートの数も少なくてもよい。環境音合成装置 4 のメモリ量を削減することができる。

【 0 0 5 9 】

[実施例 4 の動作例 2]

実施例 4 は、伝送元の伝送元の環境音の例として拍手音を対象とし、伝送元の拍手音の音量に関するパラメタを取得して、伝送先で拍手音を生成する環境音合成装置 4 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。上述では、一人の人間による一拍分の拍手音 (3 0 0 m s 程度) のテンプレートを環境音素片テンプレートの例として示したが、これに限らず、たとえば、一人の人間による一拍分の声援、掛け声のテンプレート

10

【 0 0 6 0 】

実施例 4 の動作例 2 では、実施例 4 の環境音合成装置 4 のデータ受信部 3 1 と、音源合成部 4 2 と、テンプレート記憶部 4 3 と、再生部 3 4 と、人数推定部 4 5 と、テンプレート音量記憶部 4 6 において取り扱われるデータが拍手音から環境音に置き換わる点を除いては、上述の動作例と同じである。

【 0 0 6 1 】

なお、音源合成部 4 2 において、式 (3) の代わりに、時系列テンプレート信号 $T_i = (t_i [1] \quad t_i [2] \quad \dots \quad t_i [P])$ と環境音タイミングを表すインパルス $(m \cdot \tau + \sigma_m)$ の畳み込み演算で Y_i を出力としても良い。

20

【 0 0 6 2 】

【数 8】

$$Y_i = \delta(m \cdot \tau + \sigma_m) * T_i \quad (i = 0, 1, 2, \dots) \quad \dots (3')$$

【 0 0 6 3 】

ここで * は畳み込み演算を表す。

【 0 0 6 4 】

また、テンプレート記憶部 4 3 に記憶しておく環境音素片テンプレートの波形のエネルギーをあらかじめ正規化してあってもよい。その場合は、人数推定部 4 5 のパラメタに応じて、音量 (ゲイン) を調整すればよい。この場合もメモリ量を少なくしながらバリエーションを増やすことができる。

30

【実施例 5】

【 0 0 6 5 】

以下、図 1 4、図 1 5、図 1 6 を参照して本発明の実施例 5 及びその変形例の環境音分析装置について説明する。図 1 4 は本実施例及びその変形例の環境音分析装置 5 の構成を示すブロック図である。図 1 5 は本実施例及びその変形例の環境音分析装置 5 の動作を示すフローチャートである。図 1 6 は拍手音や手拍子音の周波数成分の時間変化を例示する図である。図 1 6 A は、拍手音が鳴りやんで無音状態に移行した場合の周波数成分の時間変化の例を示す図である。図 1 6 B は、手拍子が行われている場合の周波数成分の時間変化の例を示す図である。図 1 6 C は、拍手が行われている場合の周波数成分の時間変化の例を示す図である。図 1 4 に示すように、本実施例の環境音分析装置 5 は、收音部 1 1 と、音量計算部 1 2 と、パラメタ変換部 1 3 と、データ送受信部 5 4 と、拍手区間検出部 5 5 と、周期性判定部 5 6 とを備える。本実施例の変形例としてパラメタ変換部 1 3 をパラメタ変換部 2 3、2 3' に適宜変更可能である。また、收音部 1 1、音量計算部 1 2、パラメタ変換部 1 3、2 3、2 3' は実施例 1、実施例 2、変形例 1 の環境音分析装置 1、2、2' における同一番号の各構成部と同じであるから適宜説明を略する。

40

【 0 0 6 6 】

< 拍手区間検出部 5 5 >

拍手区間検出部 5 5 は、音響信号が拍手音であるか否かを判別する機能を有する。具体的には、拍手区間検出部 5 5 は、VAD (Voice __ Activity __ Detect

50

ion)やSAD(Sound_Activity_Detection)を用いて入力された音響信号をフレーム毎に分析し、「無音区間」「音声(音楽)区間」「その他」のいずれに該当するかを判別し、音響信号が無音区間または音声区間と判別された場合には、環境音量パラメタ $P_j = 0$ と設定する。例えば拍手区間検出部55が、ITU-T_G.720.1を用いてVAD_GSAD分析する場合、拍手区間検出部55は判別結果として0:無音、1:ノイズ、2:音楽、3:音声の何れかのフラグを生成する。この場合、拍手区間検出部55は、判別結果が無音区間(フラグ0)、音声区間(フラグ2または3)である場合(つまり、拍手音でない場合)には、環境音量パラメタ $P_j = 0$ と設定する。一方、拍手区間検出部55は、判別結果がノイズ(フラグ1)である場合には、これを拍手音であるものとして、フラグ=1と設定する。

10

【0067】

環境音量パラメタ $P_j = 0$ の場合(S5Y)、拍手区間検出部55はデータ送信部54に環境音量パラメタ $P_j = 0$ 、および周期情報 $T = 0$ 、または周期なしを示す周期情報を出し、ステップS54に移行する。一方、環境音量パラメタ P_j が0でない(フラグ=1)場合(S5N)、フラグ=1を出力してステップS56に移行する。

【0068】

<周期性判定部56>

周期性判定部56は、環境音量パラメタ P_j が0でない場合(G.720.1を用いてVAD_GSAD分析する場合、フラグ1:ノイズと判別された場合)当該フレームについて、1秒から数秒程度(例えば300フレーム:3秒)の窓幅で分析を行う。分析したフレームが図16Bに示すように一定の周期性がある場合には、周期性判定部56は、当該フレームの音響信号を手拍子の音響信号と判断し、周期が検出された場合には、自己相関関数のピーク間隔を周期情報として出力する(S56)。演算量削減のために、G.720.1の内部変数であるRMS値(フレームのパワーの平方根)を用いて、300フレーム分の自己相関関数を求め、周期Tを求めてもよい(フレームの自己相関ではなく信号の自己相関を用いても良い)。

20

【0069】

周期情報の検出には、例えば次のような方法を用いる。G.720.1の内部変数である10ms毎のRMS値(フレームのパワーの平方根)の自己相関関数を $R(i)$ ($i = 0, 1, 2, \dots, 300$)とする。この例では3秒の分析フレームで分析したことになる。この例では、100ms以上の周期を持つ場合には手拍子、そうでない場合には拍手と判定されることになる。手拍子の間隔が100ms以上であるとすると、 $R(i)$ の i の値を大きくしていき、相関の値 $R(i)$ が増加から減少に転じたときの i を T_1 とする。さらに i の値を増加させていき、相関の値 $R(i)$ が増加から減少に転じたときの i を T_2 とする。 $T_2 - T_1$ の値が閾値の範囲内(例えば11~265)であれば、 $T_2 - T_1$ を周期として $T = T_2 - T_1$ を出力する。なお、ピークの検出は、心電図のR-R間隔の検出など様々な方法があるので、既存の技術を適宜用いれば良い。

30

【0070】

周期が検出されない場合には、周期性判定部56は、 $T = 0$ または周期なしを示す周期情報を出し、周期を表すインデックスが上記のように例えば11~265であれば255通りの周期を表すことができ、周期が無いというインデックスを加えて256通りの条件を表す8ビットを伝送すればよい。

40

【0071】

音量計算部12とパラメタ変換部13(23、23')は、環境音量パラメタ P_j が0でない(G.720.1を用いてVAD_GSAD分析する場合、フラグ1:ノイズと判別された)フレームについて、実施例1または実施例2、あるいは変形例2の音量計算部12とパラメタ変換部13(23、23')と同様の処理を行うことにより、環境音量パラメタ P_j を計算し出力する(S12、S13、S23、S23')。なお、音量計算部12が出力する E_j として、G.720.1の内部変数であるRMS値(フレームのパワーの平方根)を出力してもよい。

50

【0072】

<データ送信部54>

データ送信部54は、環境音量パラメタ P_j とともに周期情報を後述する環境音合成装置6に送信する(S54)。

【実施例6】

【0073】

以下、図17、図18を参照して本発明の実施例6の環境音合成装置について説明する。実施例6の環境音合成装置6は、実施例5の環境音分析装置5に対応する装置である。図17は本実施例の環境音合成装置6の構成を示すブロック図である。図18は本実施例の環境音合成装置6の動作を示すフローチャートである。図17に示すように、本実施例の環境音合成装置6は、データ受信部61と、音源合成部62と、テンプレート記憶部43と、再生部34と、人数推定部45と、テンプレート音量記憶部46とを備える。データ受信部61と音源合成部62以外の各構成部は実施例3、または実施例4の環境音合成装置3、4における同一番号の各構成部と同じであるから説明を省略する。

10

【0074】

<データ受信部61>

データ受信部61は、環境音量パラメタ P_j とともに周期情報を環境音分析装置5から受信する(S61)。

【0075】

<音源合成部62>

音源合成部62は、あるフレームの音響信号の環境音量パラメタ P_j が0でなく(S6AN)、周期情報が $T=0$ 、または周期が存在しないことを示している場合に(S6BN)、当該フレームの音響信号が拍手音であると判定して、実施例4の環境音合成装置4の音源合成部42と同じ処理を行い、拍手音を出力する(S62A)。一方、音源合成部62は、あるフレームの音響信号の環境音量パラメタ P_j が0でなく(S6AN)、周期情報が $T \neq 0$ である(周期情報ありを示す)場合に(S6BY)、当該フレームの音響信号が手拍子であると判定して、その周期(たとえば500ms)が中心となるような実施例4の時よりも分散の小さいガウス分布やラプラス分布に従う揺らぎを持たせた波形を合成し、拍手音(実際は手拍子音)として出力する(S62B)。音源合成部62は、例えば上記式(3)または(4)において、 σ の値を周期情報(例えば $T=500\text{ms}$)としてテンプレートを変換した拍手音 Y_i ($i=0, 1, 2, \dots$)を出力する(S62B)。この例では、手拍子音における $\sigma=500(\text{ms})$ として前述の拍手音の合成の際に設定された $\sigma=300(\text{ms})$ より長い時間間隔に設定しているため、拍手音と判定された場合と比較して手拍子音と判定された場合の時間間隔が長くなるように環境音素片テンプレートが配置され合成される。

20

30

【0076】

このように、実施例5の環境音分析装置5と本実施例の環境音合成装置6により構成される環境音伝送システム5000によれば、周期性判定部56がフレームを分析して周期情報を生成し、音源合成部62において、周期情報の有無を考慮して拍手、手拍子の何れかの音響信号を生成するため、伝送元において収音された拍手や手拍子音を効率よく伝送する効果に加えて、拍手、手拍子の双方をより正確に合成することができる。

40

【実施例7】

【0077】

以下、図19、図20を参照して実施例4の変形実施例である本発明の実施例7の環境音合成装置について説明する。図19は本実施例の環境音合成装置7の構成を示すブロック図である。図20は本実施例の環境音合成装置7の動作を示すフローチャートである。図19に示すように、本実施例の環境音合成装置7は、データ受信部31と、音源合成部72と、テンプレート記憶部73と、再生部34と、人数推定部45と、テンプレート音量記憶部46とを備える。実施例4との違いは、実施例4における音源合成部42と、テンプレート記憶部43が、本実施例においてそれぞれ音源合成部72と、テンプレート記

50

憶部 73 に変更されている点のみである。よって、音源合成部 72、テンプレート記憶部 73 以外の各構成については説明を省略する。

【0078】

実施例 4 では音源合成部 42 においてテンプレート記憶部 43 から環境音素片テンプレートをランダムに 1 つ選択して、伝送先での環境音を合成していたが、本実施例では、音源合成部 42 に代わって、音源合成部 72 が、テンプレート記憶部 73 から環境音素片テンプレートを出力確率に応じて 1 以上選択して合成する (S72)。

【0079】

特に、声援・掛け声は拍手音と比較すると個人差が大きく、その内容にいくつかバリエーションがある。例えば、出演者が複数人いる場合などは、出演者の個人名やニックネームが声援や掛け声に含まれることがあり、出演者毎に異なる声援・掛け声のパターンが存在する。

10

【0080】

ただし、声援や掛け声の音のパターン(ミックスを打つとも表現される)はコンテンツによりきめられていることが多く、声援や掛け声のセリフのバリエーションは多くない。また、ある曲の決められたタイミングで発せられることが多く、音響パワーの時間差が小さい。したがって、複数種類の声援や掛け声の定型文を環境音素片テンプレートとして用意しておき、一つのテンプレートではなく、音量に応じたゆらぎを持たせた混合音にして再生することにより、毎回同じ声援や掛け声の音ではなく、発せられるタイミングにより別の環境音に類似した音を構成することができる。

20

【0081】

以下では、伝送元で複数の出演者を含むコンテンツが演じられている(もしくは再生されている)場合を例に、環境音を伝送元とは異なる伝送先の会場で合成する環境音合成装置 7 の動作例を説明する。

【0082】

前述したように、本実施例の環境音合成装置 7 は、実施例 4 における環境音合成装置 4 とほとんど同じ構成である。ただし、テンプレート記憶部 73 に、複数種類の環境音素片テンプレートと各環境音素片テンプレートの出力確率とが対応付けて記憶されている点が異なる。複数種類の環境音素片テンプレートとは、例えば、コンテンツの出演者の各々の個人名やニックネームに対応する声援や掛け声の音響信号であり、予め用意しておくものとする。

30

【0083】

以下、図 21 を参照して説明を続ける。図 21 は、テンプレート記憶部に環境音素片テンプレートと出力確率とを対応付けて記憶する例を例示する図である。

【0084】

図 21 の A では、環境音素片テンプレートの後の P_i ($i = 1, 2, \dots, n$)、例えば「アリス」の後の P_1 は、その対応する環境音素片テンプレートが出力される確率であり、 $P_1 + P_2 + \dots + P_n = 1$ である。実施例 4 の動作例 2 はこの確率が共通であって ($P_1 = P_2 = \dots = P_n = 1/n$) 環境音素片テンプレートをランダムに 1 つ選択して合成する場合に相当する。本実施例では、この確率がテンプレートごとに異なる。なお、図 21 の A は出演者ごとに 1 つの環境音テンプレートが記憶されている例であるが、図 21 の B のように出演者ごとに複数の環境音テンプレートが記憶されていて、各環境音テンプレートに出力確率が対応付けられていても良い。各テンプレートの確率は予め設定しておけば良い。例えば、伝送元や伝送先のコンサートホール等において各自が応援したい出演者を投票もしくは、事前に登録してもらうなどして、その会場の聴衆についての出演者の人気度のランキングを得ておき、その順位が高いほど出力確率が高くなるように、各テンプレートの出力確率を決定しておけば良い。あるいは、伝送先の聴衆に限らずに取得した人気度のランキングに基づいても良い。また、ある出演者の誕生日や引退など、特別なイベントを考慮して、該当出演者への環境音素片テンプレートの出力確率を高くするようにバイアスをかけてもよい。

40

50

【 0 0 8 5 】

これらの出力確率は、このシステムを動作させながらリアルタイムで変更してもよい。例えば、受聴者がボタンを押すなどのリアルタイムの投票数に応じて、出力確率値を時間的に変化させてもよい。

【 0 0 8 6 】

音源合成部 7 2 では、テンプレート記憶部 7 3 に記憶された出力確率に従って、テンプレートを 1 つ以上選択し、式 (3 ') または式 (4) に従って環境音の音響信号を合成する。人数推定部 4 5 で推定した人数 M が 2 名以上であって、テンプレートを 2 つ以上選択するとする。フレーム i について選択するテンプレートの数を S_i (S は 2 以上の整数) とすると、音源合成部 7 2 は、フレーム i ごとに

【 0 0 8 7 】

【数 9】

$$Y_i = Y_{i,1} * Y_{i,2} * \dots * Y_{i,S_i} \quad (i = 0, 1, 2, \dots)$$

【 0 0 8 8 】

により合成した音響信号 Y_i を出力する。ここで、 $T_{i,k}$ はフレーム i について選択した k 番目のテンプレートであり、

【 0 0 8 9 】

【数 1 0】

$$Y_{i,k} = \sum_{k=1}^{M_k} \left(\frac{1}{M_k} k \cdot \tau + \sigma_k \right) \cdot T_{i,k} \quad (k = 1, 2, \dots, S_i)$$

【 0 0 9 0 】

である。ここで、 M_k は、テンプレート $T_{i,k}$ に対応する人数であり、

【 0 0 9 1 】

【数 1 1】

$$\sum_{k=1}^{S_i} M_k = M$$

【 0 0 9 2 】

となるように、テンプレート $T_{i,k}$ の出力確率に応じて各 M_k の値を決定したものである。

【 0 0 9 3 】

以上の構成によれば、出演者の人気が反映された音量で、各出演者に対応する環境音が合成されるので、より場の盛り上がりや雰囲気を反映した環境音を合成することができる。

【実施例 8】

【 0 0 9 4 】

引き続き、図 1 9、図 2 0、図 2 1 を参照して実施例 8 の環境音合成装置 8 について説明する。図 1 9 に示すように、本実施例の環境音合成装置 8 は、データ受信部 3 1 と、音源合成部 8 2 と、テンプレート記憶部 8 3 と、再生部 3 4 と、人数推定部 4 5 と、テンプレート音量記憶部 4 6 とを備える。実施例 7 との違いは、実施例 7 における音源合成部 7 2 と、テンプレート記憶部 7 3 が、本実施例においてそれぞれ音源合成部 8 2 と、テンプレート記憶部 8 3 に変更されている点のみである。

【 0 0 9 5 】

実施例 7 ではテンプレートごとに確率値を決めていたが、本実施例では図 2 1 の C に示すテンプレート記憶部 8 3 のように、テンプレートを出演者毎にクラスタリングし、クラスタごとに出力確率を決定しても良い。この場合、音源合成部 8 2 では、テンプレート記

10

20

30

40

50

憶部 8 3 に記憶された出力確率に従って、出演者に対応するクラスタを 1 つ以上選択し、選択した各クラスタに対応する 1 以上の環境音素片テンプレートの中からランダムに 1 つまたは複数の環境音素片テンプレートを選択する。なお、クラスタ内で複数の環境音素片テンプレートを選択する場合には、そのテンプレートにより出力される音響信号の音量の総和が、当該クラスタに対して割り当てられる音量となるように、テンプレートごとの音量を決定する。

【実施例 9】

【0096】

以下、実施例 9 の環境音合成装置 9 について説明する。本実施例の環境音合成装置 9 は、実施例 7 または 8 の環境音合成装置で合成した音声をヘッドマウントディスプレイに出力する。

10

【0097】

ここで、テンプレート記憶部 7 3 (8 3) に記憶しておく各テンプレートもしくは各クラスタの出力確率は、ヘッドマウントディスプレイを装着する使用者の好みに応じて予め設定しておく。これにより、個人ごとに好みの環境音を提示することができる。ヘッドマウントディスプレイの装着者毎にあらかじめ定められた確率値に応じて、テンプレート記憶部 7 3 (8 3) から環境音素片テンプレートをあらかじめ定められた確率に応じて選択して合成すると、個人個人に応じた環境音を別々に提示することができる。ヘッドマウントディスプレイに限らず、ユーザが音響的に独立の場所でこのシステムを利用する場合には、場所ごとに異なる確率値でテンプレート記憶部 7 3 (8 3) から環境音素片テンプレートをあらかじめ定められた確率に応じて選択して合成することができる。

20

【実施例 10】

【0098】

以下、実施例 10 の環境音合成装置 10 について説明する。上述の実施例 7、8、9 では、演者ごとに異なる環境音素片テンプレートに対応させていたが、本実施例の環境音合成装置 10 では、演者全員に共通な環境音素片テンプレートを定義する。例えば確率値もしくはランダムに選択された環境音素片テンプレートがいわゆるミックス（声援や掛け声の音のパターン）でもよい。

【0099】

なお、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ことはいうまでもない。

30

【0100】

また、上述の構成をコンピュータによって実現する場合、各装置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、上記処理機能がコンピュータ上で実現される。

【0101】

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。

40

【0102】

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。

【0103】

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

50

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。

【0104】

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。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、本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。

10

【図1】

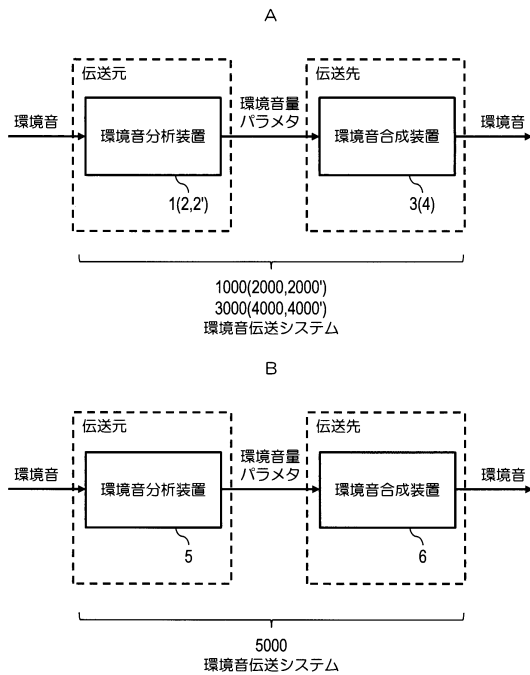


図1

【図2】

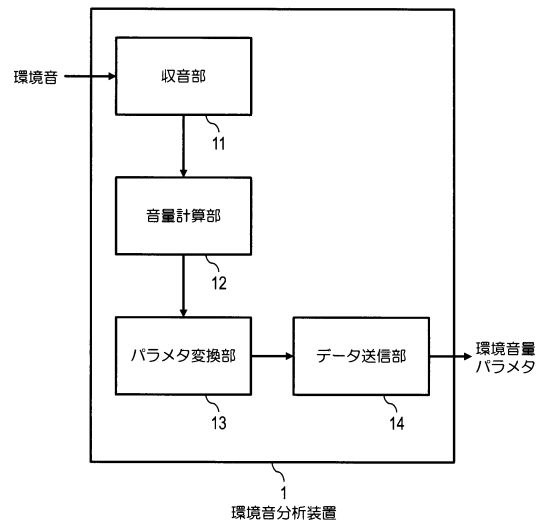


図2

【 図 3 】

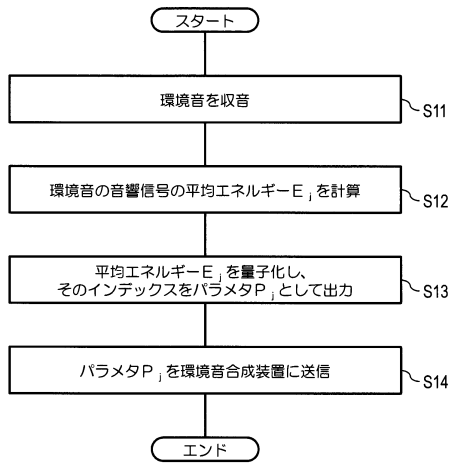


図3

【 図 4 】

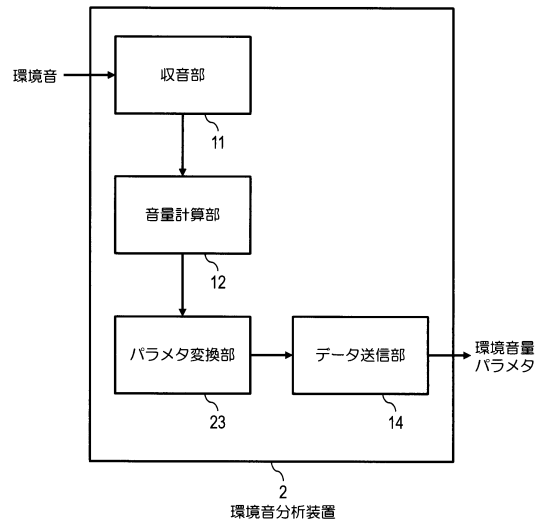


図4

【 図 5 】

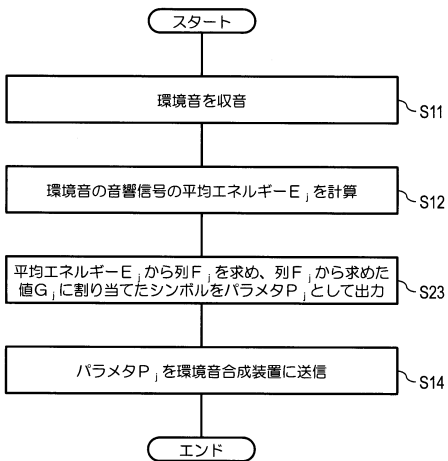


図5

【 図 6 】

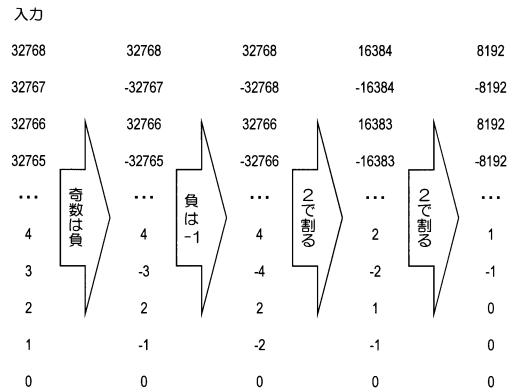


図6

【図7】

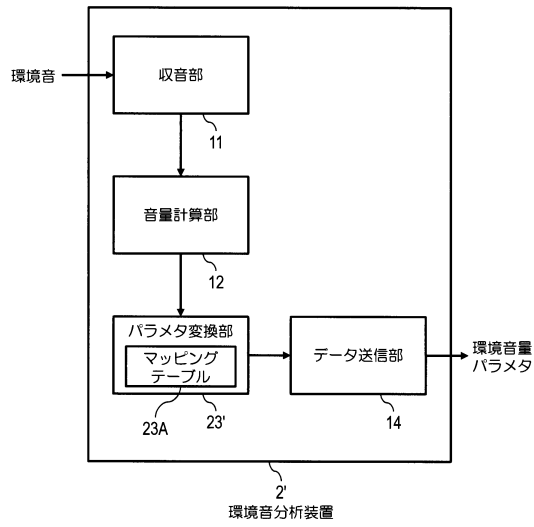


図7

【図8】

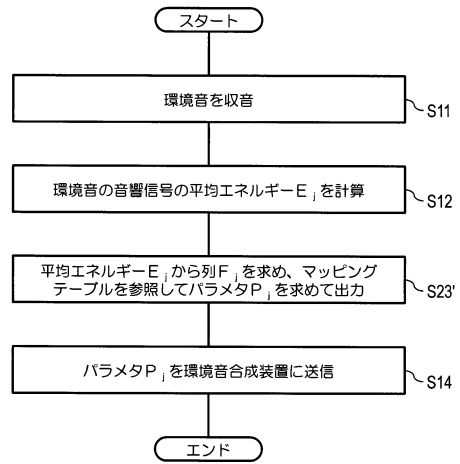


図8

【図9】

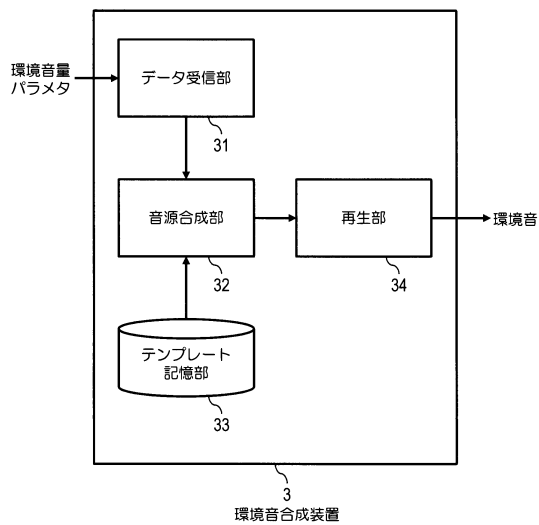


図9

【図10】

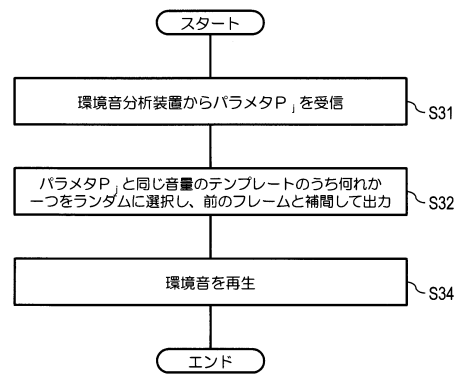


図10

【図11】

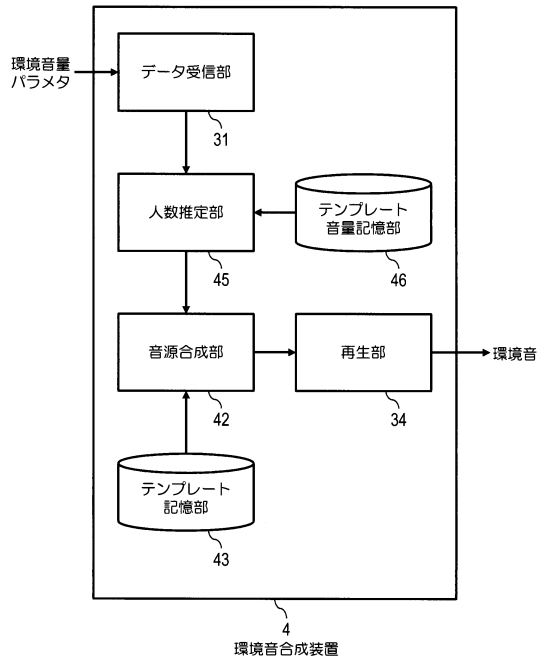


図11

【図12】

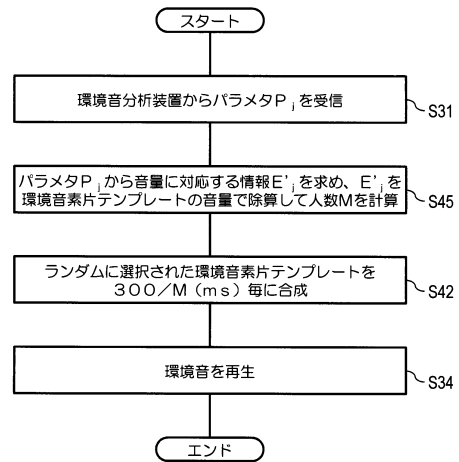


図12

【図13】

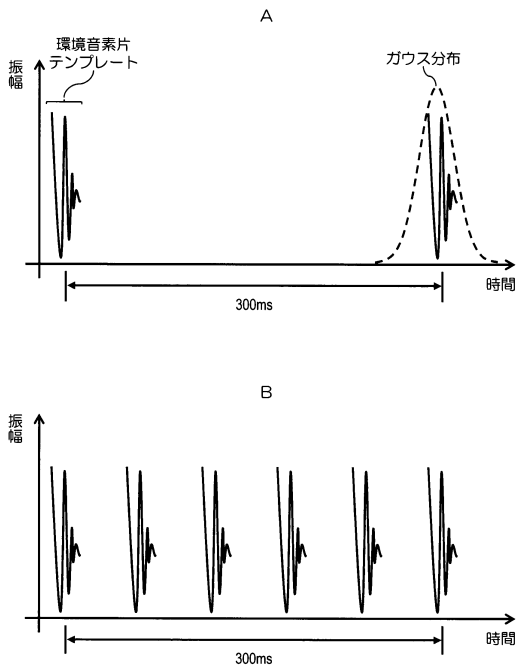
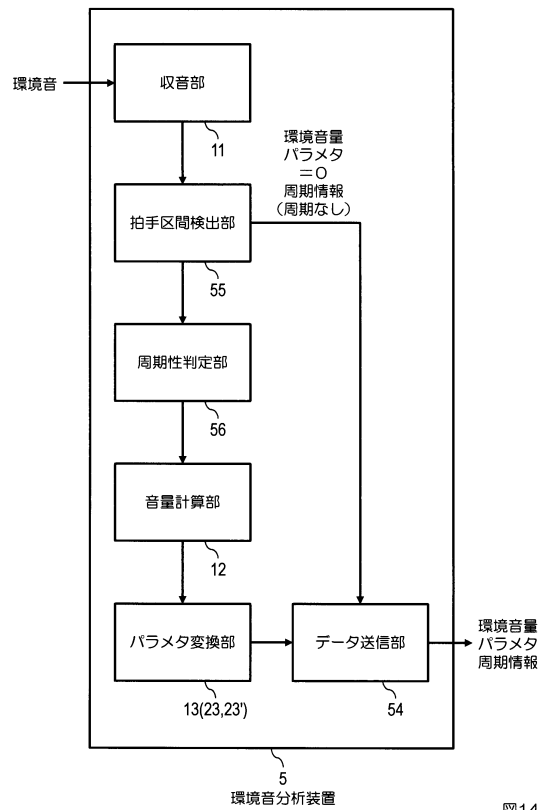


図13

【図14】



環境音分析装置

図14

【図15】

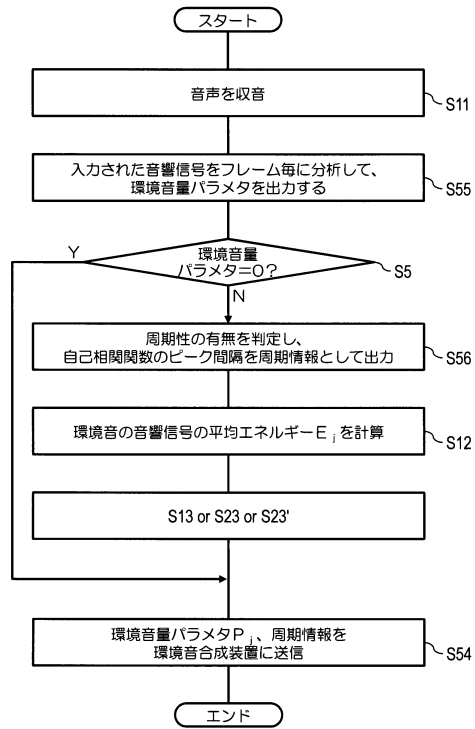


図15

【図17】

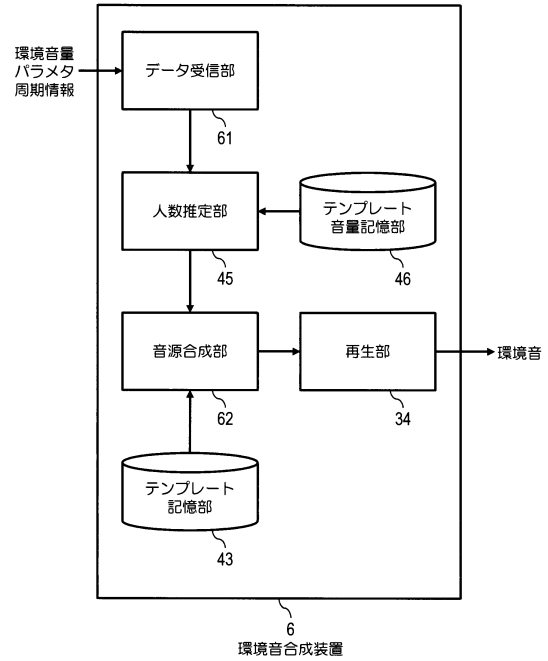


図17

【図18】

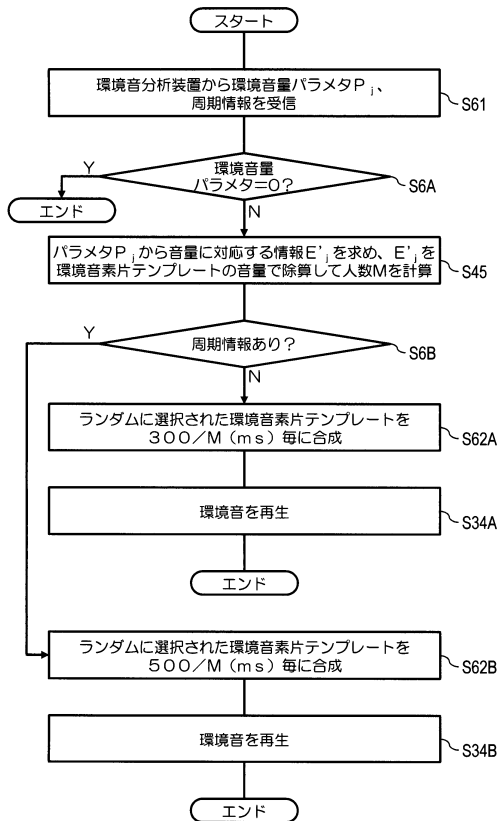


図18

【図19】

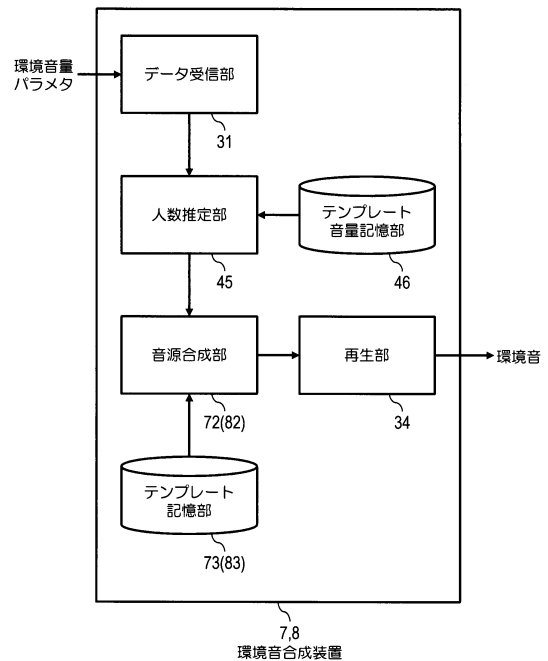


図19

【図20】

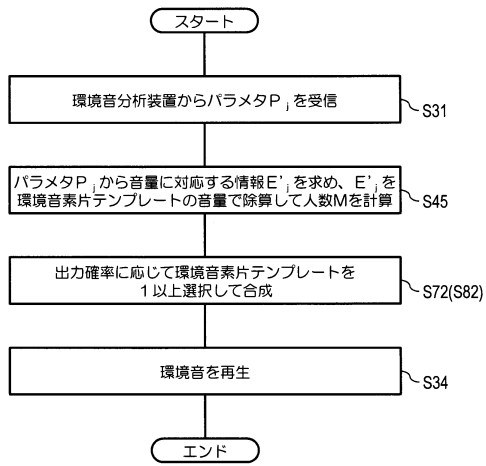


図20

【図21】

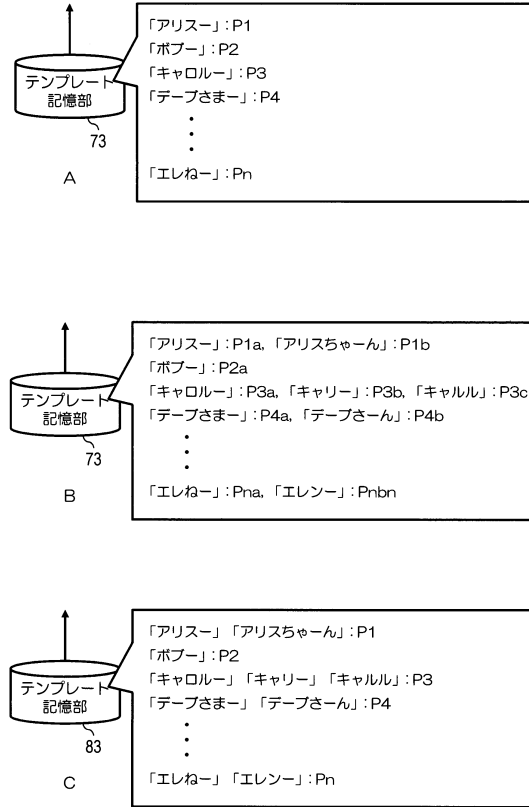


図21

【図16】

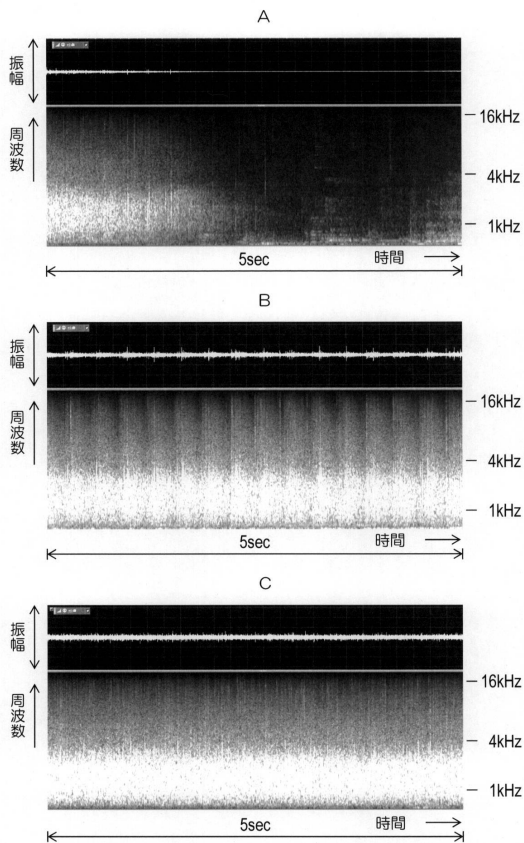


図16

フロントページの続き

(72)発明者 守谷 健弘

東京都千代田区大手町二丁目3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

審査官 安田 勇太

(56)参考文献 特開平06-203127(JP,A)

特開平05-037476(JP,A)

特開昭60-107933(JP,A)

特開平04-147198(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 19/26

G10L 13/00 - 13/10

G10L 25/00 - 25/94