

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5150573号  
(P5150573)

(45) 発行日 平成25年2月20日 (2013. 2. 20)

(24) 登録日 平成24年12月7日 (2012.12.7)

(51) Int. Cl.	F I	
<b>G 1 O L 13/02 (2013.01)</b>	G 1 O L 13/02	1 3 O C
<b>G 1 O G 3/04 (2006.01)</b>	G 1 O G 3/04	
<b>G 1 O H 1/40 (2006.01)</b>	G 1 O H 1/40	
<b>G 1 O H 1/00 (2006.01)</b>	G 1 O H 1/00	B
<b>G 1 O L 25/51 (2013.01)</b>	G 1 O L 11/00	4 O 2 K
請求項の数 3 (全 22 頁) 最終頁に続く		

(21) 出願番号	特願2009-166049 (P2009-166049)	(73) 特許権者	000005326
(22) 出願日	平成21年7月14日 (2009. 7. 14)		本田技研工業株式会社
(65) 公開番号	特開2010-26513 (P2010-26513A)		東京都港区南青山二丁目1番1号
(43) 公開日	平成22年2月4日 (2010. 2. 4)	(74) 代理人	100064908
審査請求日	平成23年11月24日 (2011. 11. 24)		弁理士 志賀 正武
(31) 優先権主張番号	61/081, 057	(74) 代理人	100108578
(32) 優先日	平成20年7月16日 (2008. 7. 16)		弁理士 高橋 詔男
(33) 優先権主張国	米国 (US)	(74) 代理人	100146835
			弁理士 佐伯 義文
		(74) 代理人	100094400
			弁理士 鈴木 三義
		(74) 代理人	100107836
			弁理士 西 和哉
		(74) 代理人	100108453
			弁理士 村山 靖彦
最終頁に続く			

(54) 【発明の名称】 ロボット

(57) 【特許請求の範囲】

【請求項1】

音楽音響を集音して音楽音響信号に変換する集音手段と、  
音声合成処理によって歌唱又はスキャットに係る自己音声信号を生成する音声信号生成手段と、

前記自己音声信号を音に変換して出力する音出力手段と、

前記音楽音響信号と前記自己音声信号とを入力し、前記音楽音響信号から前記自己音声信号の音声成分を抑制した音響信号を生成する自己音声抑制手段と、

前記音響信号にフィルタ処理を行ってオンセットを強調するフィルタ手段と、

前記オンセットの強調された音響信号に、相互相関関数を適用した時間周波数パターンマッチングを行ってビート間隔信頼度を計算するビート間隔信頼度計算手段と、

前記計算されたビート間隔信頼度に基づいてビート間隔を推定するビート間隔推定手段と、

前記フィルタ手段においてオンセットの強調された音響信号と前記ビート間隔推定手段において推定されたビート間隔とに基づいて、ビート時刻信頼度を計算するビート時刻信頼度計算手段と、

前記計算されたビート時刻信頼度に基づいてビート時刻を推定するビート時刻推定手段と、

前記それぞれ推定されたビート間隔及びビート時刻に基づいて、現在時刻よりも先のビート時刻を予測するビート時刻予測手段と、

10

20

前記ビート間隔と前記予測されたビート時刻とに基づいて、前記音声信号生成手段で生成される自己音声信号を同期化する同期化手段と、  
を備えたことを特徴とするロボット。

【請求項 2】

前記ビート時刻予測手段は、

現在時刻から、少なくとも前記音声信号生成手段における処理遅延時間分の時間を経過した時刻以降のビート時刻を予測することを特徴とする請求項 1 記載のロボット。

【請求項 3】

前記ビート間隔推定手段で推定されたビート間隔に基づいて、ビート間隔の変動が所定の許容値より小さい期間を音楽区間として検出する音楽区間検出手段をさらに備え、

前記音声信号生成手段は、前記音楽区間として検出された場合に前記自己音声信号を生成することを特徴とする請求項 1 又は 2 記載のロボット。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音楽やスカット等のビートを有する音響情報からテンポ及びビート時刻を推定するビートトラッキング技術を適用して音楽インタラクションを行うロボットの技術に関する。

【背景技術】

【0002】

近年、ヒューマノイドやホームロボット等、人間とソーシャル・インタラクションを行うロボットの研究が盛んに行われている。その中でも、ロボットに自らの耳で音楽を聴かせ、その音楽に合わせて歌唱させたり体を動かしたりさせる音楽インタラクションに関する研究は、ロボットに自然で豊かな表現をさせるために重要である。この技術分野においては、例えば、マイクロホンで集音したライブ音楽からリアルタイムにビートを抽出し、そのビートに同期させてロボットを踊らせる技術が知られている（例えば、特許文献 1 参照）。

【0003】

このようなロボットに音楽を聴かせて、その音楽のリズムに合わせてロボットを動作させる場合、音楽の音響情報からテンポを推定する必要がある。従来では、音響情報に基づいた自己相関関数を計算することによりテンポを推定していた（例えば、特許文献 1, 2 参照）。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2007 - 33851 号公報

【特許文献 2】特開 2002 - 116754 号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

ところで、音楽を聴かせたロボットに、その音楽の音響情報からビートを抽出しテンポを推定させる場合、主に二つの対応すべき技術的課題がある。第一の課題は、雑音に対するロバスト性の確保である。ロボットに音楽を聴かせるためには、集音装置、例えばマイクロホンを搭載することが必要である。ロボットの外見の見栄えを考慮すると、集音装置はロボット体内に組み込まれることが望ましい。

【0006】

そうすると、その集音装置が集音する音には様々な雑音が含まれることになる。すなわち、集音装置が集音する音には、ロボットの周囲で発生する環境音はもちろんのこと、ロボット自身から発生する様々な音が雑音として含まれる。例えば、ロボット自身から発生する音として、ロボットの足音、体内で駆動するモータの動作音、自発音声等が挙げられ

10

20

30

40

50

る。特に、自発音声は、音声発生源であるスピーカが集音装置の比較的近くに組み込まれるため、周囲の環境音よりも入力レベルの大きな雑音となる。このように、集音された音楽の音響信号のS/N比が悪くなると、音響信号からビートを抽出する精度が低くなり、その結果テンポ推定の精度も低くなる。

【0007】

特に、ロボットの音楽インタラクションに求められる、集音した音楽に合わせて歌唱させたり発声させたりする動作においては、雑音となる自発音声のビートに周期性があるため、ロボットによるテンポ推定動作に不利な影響を与えることになる。

【0008】

第二の課題は、テンポ変動への追従性（適応性）及びテンポ推定の安定性の確保である。例えば、人間の演奏や歌唱による音楽のテンポは、演奏者や歌唱者の技量によって、又は楽曲の曲調等によって常に一定ではなく、曲の途中で変動することが通常である。ロボットに、そのようなテンポの一定しない音楽を聴かせて、その音楽のビートに同期させて動作させる場合、テンポ変動への高い追従性が求められる。その一方で、テンポが比較的一定しているときは、安定してテンポを推定できることが望ましい。一般的に、自己相関の計算を行って安定してテンポを推定するためには、テンポ推定処理で用いられる時間窓を長く設定するのがよいが、そのかわりテンポ変動への追従性は悪くなる。すなわち、テンポ変動への追従性の確保と、テンポ推定の安定性の確保とはトレードオフの関係にある。しかしながら、ロボットの音楽インタラクションにおいては、その両方の性能を良好に保つ必要がある。

【0009】

ここで、第一及び第二の課題の関係をみると、第一の課題である雑音に対するロバスト性を確保するためには、第二の課題の一方であるテンポ推定の安定性を確保する必要があると考えられるが、その場合、第二の課題の他方であるテンポ変動への追従性の確保が困難になるという問題がある。

【0010】

上記特許文献1, 2には、上記第一の課題についての明示的記載及び示唆は一切ない。また、特許文献1, 2を含む従来技術では、テンポ推定処理において時間方向の自己相関を求めており、テンポ推定の安定性を確保するために時間窓を長く設定するとテンポ変動への追従性が悪くなり上記第二の課題に対応できない。

【0011】

そこで、本発明は上記問題に鑑みてなされたものであり、雑音に対するロバスト性を確保するとともに、テンポ変動への追従性及びテンポ推定の安定性を確保して精度の高い音楽インタラクションを行うことのできるロボット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】

【0012】

上記の課題を解決するため、請求項1記載のロボット（例えば、実施例における脚式移動音楽ロボット4）は、音楽音響を集音して音楽音響信号（例えば、実施例における音楽音響信号MA）に変換する集音手段（例えば、実施例における耳機能部310）と、音声合成処理によって歌唱又はスキヤットに係る自己音声信号（例えば、実施例における自己音声信号SV）を生成する音声信号生成手段（例えば、実施例における歌唱制御部220, スキヤット制御部230）と、前記自己音声信号を音に変換して出力する音出力手段（例えば、実施例における発声機能部320）と、前記音楽音響信号と前記自己音声信号とを入力し、前記音楽音響信号から前記自己音声信号の音声成分を抑制した音響信号を生成する自己音声抑制手段（例えば、実施例における自己発声音抑制部10）と、前記音響信号にフィルタ処理を行ってオンセットを強調するフィルタ手段（例えば、実施例におけるSobelフィルタ部21）と、前記オンセットの強調された音響信号に、相互相関関数を適用した時間周波数パターンマッチングを行ってビート間隔信頼度を計算するビート間隔信頼度計算手段（例えば、実施例における時間周波数パターンマッチング部22）と、前記計算されたビート間隔信頼度に基づいてビート間隔を推定する（例えば、実施例にお

10

20

30

40

50

けるテンポTP)ビート間隔推定手段(例えば、実施例におけるビート間隔推定部23)と、前記フィルタ手段においてオンセットの強調された音響信号と前記ビート間隔推定手段において推定されたビート間隔とに基づいて、ビート時刻信頼度を計算するビート時刻信頼度計算手段(例えば、実施例における近接ビート信頼度計算部31,連続ビート信頼度計算部32,ビート時刻信頼度計算部33)と、前記計算されたビート時刻信頼度に基づいてビート時刻を推定する(例えば、実施例におけるビート時刻BT)ビート時刻推定手段(例えば、実施例におけるビート時刻推定部34)と、前記それぞれ推定されたビート間隔及びビート時刻に基づいて、現在時刻よりも先のビート時刻を予測するビート時刻予測手段(例えば、実施例におけるビート時刻予測部210)と、前記ビート間隔と前記予測されたビート時刻とに基づいて、前記音声信号生成手段で生成される自己音声信号を同期化する同期化手段(例えば、実施例における歌唱制御部220,スキャット制御部230)と、を備えたことを特徴とする。

10

請求項2記載のロボットは、前記ビート時刻予測手段は、現在時刻から、少なくとも前記音声信号生成手段における処理遅延時間分の時間を経過した時刻以降のビート時刻を予測することを特徴とする。

請求項3記載のロボットは、前記ビート間隔推定手段で推定されたビート間隔に基づいて、ビート間隔の変動が所定の許容値より小さい期間を音楽区間として検出する音楽区間検出手段(例えば、実施例における音楽区間検出部110)をさらに備え、前記音声信号生成手段は、前記音楽区間として検出された場合に前記自己音声信号を生成することを特徴とする。

20

#### 【発明の効果】

##### 【0013】

請求項1記載の発明によれば、雑音に対するロバスト性を確保するとともに、テンポ変動への追従性及びテンポ推定の安定性を確保して音楽インタラクションを行うことができる。

請求項2記載の発明によれば、処理遅延時間を考慮して推定ビート時刻から未来のビート時刻を予測するため、リアルタイムな音楽インタラクションを行うことができる。

請求項3記載の発明によれば、音楽区間を検出することによって、ビートが抽出できない区間を非音楽区間であると判定するため、不安定期間の影響を受けることを少なくして音楽インタラクションを行うことができる。

30

#### 【図面の簡単な説明】

##### 【0014】

【図1】本発明の実施形態であるロボットに組み込まれるビートトラッキング装置のブロック構成図である。

【図2】本実施形態における、推定ビート間隔を決定するビート間隔推定アルゴリズムを説明するための図である。

【図3】本実施形態における、ビート時刻を推定するビート時刻推定アルゴリズムを説明するための図である。

【図4】本発明の実施例である脚式移動音楽ロボットの概略の正面図である。

【図5】本実施例である脚式移動音楽ロボットの概略の側面図である。

40

【図6】本実施例である脚式移動音楽ロボットの、主に音楽インタラクションに関する部分のブロック構成図である。

【図7】本実施例における楽曲IDテーブルの例である。

【図8】推定されたテンポに係るビート間隔時間に基づいてビート時刻を予測して外挿する様子(第2の例)を模式的に表した図である。

【図9】本実施例における、ビートトラッキング性能(ビートトラッキング成功率)についての実験結果を示したチャートである。

【図10】従来技術を使用した場合の、ビートトラッキング性能(ビートトラッキング成功率)についての実験結果を示したチャートである。

【図11】本実施例における、ビートトラッキング性能(テンポ変化時からの平均遅延時

50

間) についての実験結果を示した図である。

【図12】本実施例におけるテンポ推定の実験結果のグラフである。

【図13】本実施例における、ビートトラッキング性能(ビート予測成功率) についての実験結果を示した図である。

【図14】推定されたテンポに係るビート間隔時間に基づいてビート時刻を予測して外挿する様子(第3の例)を模式的に表した図である。

【図15】推定されたテンポに係るビート間隔時間に基づいてビート時刻を予測して外挿する様子(第4の例)を模式的に表した図である。

【発明を実施するための形態】

【0015】

以下、本発明を実施するための形態について、図面を参照して詳細に説明する。ここでは、本発明の実施形態であるロボットに組み込まれるリアルタイム・ビートトラッキング装置(以下、ビートトラッキング装置という。)について説明する。このロボットは、後述する実施例でその詳細を説明するが、マイクロホンで集音した音楽からビートを抽出し、そのビートに合わせて足踏みをしたり、歌唱やスキヤットによる自己発声音をスピーカから出力したりして音楽インタラクションを行うものである。

【0016】

図1に、ビートトラッキング装置のブロック構成図を示す。同図において、ビートトラッキング装置1は、自己発声音抑制部10と、テンポ推定部20と、ビート時刻推定部30とを備えている。

【0017】

自己発声音抑制部10は、セミブラインド独立成分分析部(以下、SB-ICA部という。)11を備えている(SB-ICA: Semi-Blind Independent Component Analysis)。SB-ICA部11には、2チャンネルの音声信号が入力される。そのうち、第1のチャンネルは音楽音響信号MAであり、第2のチャンネルは自己音声信号SVである。音楽音響信号MAは、ロボットに備えられたマイクロホンで集音された音楽から得られた音響信号である。ここでいう音楽とは、一般的にいう歌唱や演奏による音楽やスキヤット等、拍のある音響信号のことをいう。自己音声信号SVは、ロボットの音声信号生成部(例えば、後述する実施例における歌唱制御部及びスキヤット制御部)で生成されてスピーカの入力部に入力される音声合成音に係る音声信号である。

【0018】

自己音声信号SVは、ロボットの音声信号生成部が生成した音声信号であるため、雑音を無視できるクリーンな信号である。一方、音楽音響信号MAは、マイクロホンで集音した音響信号であるため雑音が含まれている。特に、ロボットに音楽を聴かせながら、足踏み、歌唱、スキヤット等を行わせた場合には、これらの動作に伴って発生する音が、ロボットに聴かせている音楽と同様の周期性を有する雑音となって音楽音響信号MAに含まれることになる。

【0019】

そこで、SB-ICA部11は、音楽音響信号MA及び自己音声信号SVを入力して周波数解析処理を行ったのち、音楽音響情報から自己音声成分のエコーキャンセルを行って、自己発声音の抑制されたスペクトラムである自己発声音抑制スペクトラムを出力するものである。

【0020】

具体的には、SB-ICA部11は、音楽音響信号MA及び自己音声信号SVを、例えば、44.1kHz、16ビットで同期させてサンプリングしたのち、窓長を4096ポイント、シフト長を512ポイントに設定した短時間フーリエ変換を適用した周波数解析処理を行う。この周波数解析処理によって第1及び第2のチャンネルそれぞれについて得られたスペクトルを、スペクトル $Y(t, )$ 及び $S(t, )$ とする。なお、 $t,$  は、それぞれ時間フレーム及び周波数を表すインデックスである。

10

20

30

40

50

## 【0021】

次に、SB-ICA部11は、スペクトル $Y(t, \quad)$ 及び $S(t, \quad)$ をもとに、SB-ICA処理を行って自己発声音抑制スペクトラム $p(t, \quad)$ を求める。SB-ICA処理の計算方法を数式1に示す。なお、数式1においては、式を簡略化するためにの記載を省略してある。

## 【0022】

## 【数1】

$$\begin{pmatrix} P(t) \\ S(t) \\ \vdots \\ S(t-M) \end{pmatrix} = \begin{pmatrix} A & W(0) & \cdots & W(M) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} Y(t) \\ S(t) \\ \vdots \\ S(t-M) \end{pmatrix} \quad (\text{数式1})$$

10

## 【0023】

数式1において、残響を考慮するためのフレーム数を $M$ とした。すなわち、スピーカからマイクロホンまでの伝達系により、 $M$ フレームにわたって残響が生じることを仮定して、 $S(t, \quad)$ 、 $S(t-1, \quad)$ 、 $S(t-2, \quad)$ 、 $\cdots$ 、 $S(t-M, \quad)$ の反射モデルを採用している。例えば、実験においては $M=8$ フレームを設定することができる。また、数式1における $A$ 、 $W$ は分離フィルタを示し、SB-ICA部11では適応的にこれらの推定を行う。そして、数式1によれば、 $p(t, \quad) = Y(t, \quad) - S(t, \quad)$ となるようなスペクトラムが計算される。

20

## 【0024】

したがって、SB-ICA部11は、SB-ICA処理の入力及び出力に既知の信号である $S(t, \quad)$ を用いるとともに、伝達系による残響を考慮したことにより、雑音除去の効果を有しながら自己発声音の抑制を精度よく行うことができる。

## 【0025】

テンポ推定部20は、Sobelフィルタ部21と、時間周波数パターンマッチング部(以下、STPM部という。)22と、ビート間隔推定部23とを備えている(STPM: Spectro-Temporal Pattern Matching)。

30

## 【0026】

Sobelフィルタ部21は、テンポ推定部20のビート間隔推定処理の前処理として位置づけられるものであり、自己発声音抑制部10から供給された自己発声音抑制スペクトラム $p(t, \quad)$ について、楽音のオンセット(音響信号のレベルが急激に大きくなる部分)を強調するためのフィルタである。結果として、ビート成分の雑音に対するロバスト性が向上する。

## 【0027】

具体的には、Sobelフィルタ部21は、自己発声音抑制スペクトル $p(t, \quad)$ に対して、音声認識処理や音楽認識処理で用いられるメルフィルタバンクを適用し、周波数の次元数を64次元に圧縮する。そして、得られたメルスケールでのパワースペクトルを $P_{mel}(t, f)$ とする。なお、メル周波数軸での周波数インデックスを $f$ とする。ここで、スペクトログラムにおいてパワーが急激に上昇する時刻は楽音のオンセットである可能性が高く、そのオンセットとビート時刻やテンポとは密接な関係がある。そこで、時間方向のエッジ強調と周波数方向の平滑化とを同時に行うことのできるSobelフィルタを用いてスペクトラムを整形する。パワースペクトル $P_{mel}(t, f)$ にフィルタ処理を行って出力 $P_{sobel}(t, f)$ を出力するSobelフィルタの計算式を数式2に示す。

40

## 【0028】

50

【数 2】

$$\begin{aligned}
 P_{sobel}(t, f) = & -P_{mel}(t-1, f+1) + P_{mel}(t+1, f+1) \\
 & -P_{mel}(t-1, f-1) + P_{mel}(t+1, f-1) \\
 & -2P_{mel}(t-1, f) + 2P_{mel}(t+1, f)
 \end{aligned}
 \tag{数式 2}$$

【0029】

さらに、ビート時刻に対応するパワーの立ち上がり部を抽出するため、数式 3 の処理を行ってフレームごとに 62 次元 ( $f = 1, 2, \dots, 62$ ) のオンセットベクトル  $d(t, f)$  を求める。

【0030】

【数 3】

$$d(t, f) = \begin{cases} P_{sobel}(t, f) & \text{if } P_{sobel}(t, f) > 0 \\ 0 & \text{otherwise} \end{cases}
 \tag{数式 3}$$

【0031】

テンポ推定部 20 のビート間隔推定処理は、STPM 部 22 及びビート間隔推定部 23 によって行われる。ここで、隣り合う二つのビートの時間間隔を、「ビート間隔」と定義する。STPM 部 22 は、Sobel フィルタ部 21 で求められたオンセットベクトル  $d(t, f)$  を用いて正規化相互相関関数による時間周波数パターンマッチング処理を行ってビート間隔信頼度  $R(t, i)$  を計算する。この正規化相互相関関数の計算式を数式 4 に示す。なお、数式 4 において、オンセットベクトルのマッチングに用いる次元数を  $F_w$  とする。例えば、 $F_w$  には 62 次元全てとなる 62 を適用することができる。また、マッチングの窓長を  $P_w$ 、シフトパラメータを  $i$  とする。

【0032】

【数 4】

$$R(t, i) = \frac{\sum_{j=1}^{F_w} \sum_{k=0}^{P_w-1} d(t-k, j) d(t-i-k, j)}{\sqrt{\sum_{j=1}^{F_w} \sum_{k=0}^{P_w-1} d(t-k, j)^2 \sum_{j=1}^{F_w} \sum_{k=0}^{P_w-1} d(t-i-k, j)^2}}
 \tag{数式 4}$$

【0033】

数式 4 に示した正規化相互相関関数は、時間方向と周波数方向との二次元による相互相関をとるものであるため、周波数方向に深くする一方時間方向の窓長を減らすことができる。すなわち、STPM 部 22 は、雑音に対する処理の安定性を確保したまま、処理遅延時間を少なくすることができる。また、数式 4 の分母に示された正規化項は、信号処理における白色化に相当する部分である。よって、STPM 部 22 は、Sobel フィルタ部 21 における雑音抑制効果に加えて、さらに定常雑音の抑制効果を有するものである。

【0034】

10

20

30

40

50

ビート間隔推定部 23 は、STPM部 22 で計算されたビート間隔信頼度  $R(t, i)$  からビート間隔を推定する。具体的には、次のようにしてビート間隔を推定する。ビート間隔推定部 23 は、前処理として、数式 5 によりローカルピーク  $R_{peak}(t, i)$  を計算する。

【0035】

【数5】

$$R_{peak}(t, i) = \begin{cases} R(t, i) & \text{if } R(t, i-1) < R(t, i) < R(t, i+1) \\ 0 & \text{otherwise} \end{cases} \quad (\text{数式 5})$$

10

【0036】

ビート間隔推定部 23 は、数式 5 により求めたローカルピーク  $R_{peak}(t, i)$  のうち最上位から二つのローカルピークを抽出する。そして、これらのローカルピークに対応するビート間隔  $i$  を、ローカルピーク  $R_{peak}(t, i)$  の値の大きな方からビート間隔  $I_1(t)$  及び  $I_2(t)$  として選択する。そして、ビート間隔推定部 23 は、これらビート間隔  $I_1(t)$  及び  $I_2(t)$  を用いてビート間隔候補  $I_c(t)$  を求め、さらに推定ビート間隔  $I(t)$  を推定する。

【0037】

20

図 2 に、推定ビート間隔  $I(t)$  を決定するためのビート間隔推定アルゴリズムを示して具体的に説明する。同図において、抽出された二つのローカルピーク  $R_{peak}(t, i)$  の信頼度の差が大きい場合は、ビート間隔  $I_1(t)$  をビート間隔候補  $I_c(t)$  とする。なお、差の尺度は常数  $\alpha$  によって定められ、例えば、常数  $\alpha$  を 0.7 とすることができる。

【0038】

一方、その差が小さい場合には、裏拍が抽出されているおそれがあり、よってビート間隔  $I_1(t)$  が得るべきビート間隔ではないことがある。特に、正の整数分の整数倍（例えば、 $1/2$ 、 $2/1$ 、 $5/4$ 、 $3/4$ 、 $2/3$ 、 $4/3$  等）が誤検出されやすい。よって、それを考慮して、ビート間隔  $I_1(t)$  及び  $I_2(t)$  の差分を用いたビート間隔候補  $I_c(t)$  の推定を行う。より具体的には、ビート間隔  $I_1(t)$  及び  $I_2(t)$  の差分を差分  $I_d(t)$  とし、 $I_1(t) - n \times I_d(t)$  の絶対値、又は  $I_2(t) - n \times I_d(t)$  の絶対値が閾値  $\beta$  よりも小さい場合に、 $n \times I_d(t)$  をビート間隔候補  $I_c(t)$  にする。これにおいて、2 から  $N_{max}$  までの整数である変数  $n$  の範囲で探索を行う。なお、 $N_{max}$  は、四分音符の長さまでを考慮して 4 に設定することができる。

30

【0039】

次に、得られたビート間隔候補  $I_c(t)$  と一つ前のフレームのビート間隔  $I(t-1)$  とを用いて上記と同様の処理を行って、最終的な推定ビート間隔  $I(t)$  を推定する。

【0040】

次に、ビート間隔推定部 23 は、ビート間隔推定処理によって推定した  $T_I$  フレーム分のビート間隔群に対する中央値として、数式 6 の計算によってテンポ  $TP = I_m(t)$  を求める。なお、 $T_I$  は、例えば 13 フレーム（約 150ms）とすることができる。

40

【0041】

【数6】

$$I_m(t) = \text{median}(I(t_i)) \quad (t_i = t, t-1, \dots, t-T_I) \quad (\text{数式 6})$$

50



## 【 0 0 4 2 】

図 1 の説明に戻り、ビート時刻推定部 3 0 は、近接ビート信頼度計算部 3 1 と、連続ビート信頼度計算部 3 2 と、ビート時刻信頼度計算部 3 3 と、ビート時刻推定部 3 4 とを備えている。

## 【 0 0 4 3 】

近接ビート信頼度計算部 3 1 は、あるフレームとビート間隔  $I(t)$  前のフレームとがともにビート時刻である信頼度を計算するものである。具体的には、処理フレーム  $t$  ごとに、フレーム  $t - i$  と 1 ビート間隔  $I(t)$  分前のフレーム  $t - i - I(t)$  がともにビート時刻である信頼度、すなわち近接ビート信頼度  $S_c(t, t - i)$  を、オンセットベクトル  $d(t, f)$  を用いて数式 7 により計算する。

## 【 0 0 4 4 】

## 【数 7】

$$S_c(t, t - i) = F_s(t - i) + F_s(t - i - I(t)) \quad (0 \leq i \leq I(t)) \quad (\text{数式 7})$$

$$F_s(t) = \sum_{f=1}^{F_W} d(t, f)$$

10

20

## 【 0 0 4 5 】

連続ビート信頼度計算部 3 2 は、各時刻において推定されたビート間隔  $I(t)$  でビートが連続的に存在することを示す信頼度を計算するものである。具体的には、処理フレーム  $t$  におけるフレーム  $t - i$  の連続ビート信頼度  $S_r(t, t - i)$  を、近接ビート信頼度  $S_c(t, t - i)$  を用いて数式 8 により計算する。なお、 $T_p(t, m)$  は、フレーム  $t$  を基準として  $m$  個前のビート時刻であり、 $N_{sr}$  は、連続ビート信頼度  $S_r(t, t - i)$  を評価する際に考慮すべきビート数である。

## 【 0 0 4 6 】

## 【数 8】

$$S_r(t, t - i) = \sum_m^{N_{sr}} S_c(T_p(t, m), i) \quad (0 \leq i \leq I(t)) \quad (\text{数式 8})$$

$$T_p(t, m) = \begin{cases} t & (m = 0) \\ T_p(t, m - 1) - I(T_p(t, m - 1)) & (m \geq 1) \end{cases}$$

30

## 【 0 0 4 7 】

連続ビート信頼度  $S_r(t, t - i)$  は、複数のビート列が見つかった場合に、どのビート列が最も信頼できるか判定する場合に有効である。

## 【 0 0 4 8 】

ビート時刻信頼度計算部 3 3 は、処理フレーム  $t$  におけるフレーム  $t - i$  のビート時刻信頼度  $S'(t, t - i)$  を、近接ビート信頼度  $S_c(t, t - i)$  と連続ビート信頼度  $S_r(t, t - i)$  とを用いて数式 9 により計算する。

## 【 0 0 4 9 】

40

【数 9】

$$S'(t, t-i) = S_c(t, t-i) S_r(t, t-i) \quad (\text{数式 9})$$

【0050】

そして、ビート時刻信頼度計算部 33 は、ビート時刻信頼度  $S'(t, t-i)$  間の時間的な重複を考慮し、数式 10 に示す加算平均を行って最終的なビート時刻信頼度  $S(t)$  を計算する。なお、 $S'(t)$ 、 $N_{S'(t)}$  は、フレーム  $t$  で値を有する  $S'(t, t-i)$  の集合、及びその集合の要素数を示す。 10

【0051】

【数 10】

$$S(t) = \frac{1}{N_{S'(t)}} \sum_{t_i \in S'_i(t)} S'(t_i, t) \quad (\text{数式 10})$$

【0052】

20

ビート時刻推定部 34 は、ビート時刻信頼度計算部 33 で計算されたビート時刻信頼度  $S(t)$  を用いてビート時刻  $BT$  を推定する。具体的に、図 3 に示すビート時刻  $T(n+1)$  を推定するためのビート時刻推定アルゴリズムを参照して説明する。同図のビート時刻推定アルゴリズムにおいて、 $n$  番目のビート時刻  $T(n)$  が得られており、 $n+1$  番目のビート時刻  $T(n+1)$  を推定するものとする。同図のビート時刻推定アルゴリズムでは、現在の処理フレーム  $t$  が、ビート時刻  $T(n)$  にビート間隔  $I(t)$  の  $3/4$  倍を加えた時刻を超えている場合に、ビート時刻信頼度  $S(t)$  から  $T(n) \pm 1/2 \cdot I(t)$  の範囲内で最大 3 個のピークを抽出する。その範囲内にピークが存在する場合 ( $N_p > 0$ ) は、 $T(n) + I(t)$  に最も近いピークをビート時刻  $T(n+1)$  にする。一方、ピークが存在しない場合は、 $T(n) + I(t)$  をビート時刻  $T(n+1)$  にする。そして、ビート時刻  $T(n+1)$  をビート時刻  $BT$  として出力する。 30

【0053】

以上により、本実施形態におけるビートトラッキング装置によれば、自己発声音抑制部によって、周波数解析処理ののち音楽音響情報から自己音声成分のエコーキャンセルを行うため、雑音除去の効果とともに自己発声音の抑制効果を発揮することができる。

【0054】

また、本実施形態におけるビートトラッキング装置によれば、自己発声音が抑制された音楽音響情報に *Soberl* フィルタ処理を行うため、楽音のオンセットが強調されてビート成分の雑音に対するロバスト性が向上する。

【0055】

40

また、本実施形態におけるビートトラッキング装置によれば、時間方向と周波数方向との二次元の正規化相互相関関数を計算してパターンマッチングをはかるものであるため、雑音に対する処理の安定性を確保したまま、処理遅延時間を少なくすることができる。

【0056】

また、本実施形態におけるビートトラッキング装置によれば、第一番目及び第二番目に高いローカルピークに対応する 2 つのビート間隔をビート間隔候補として選択し、これらのうちいずれがよりビート間隔として尤もらしいかを詳細に判定するため、裏拍を誤検出するおそれを抑制してビート間隔の推定を行うことができる。

【0057】

さらに、本実施形態におけるビートトラッキング装置によれば、近接ビート信頼度及び 50

連続ビート信頼度を計算してビート時刻信頼度を求めるため、ビートの集合から尤度の高いビート列についてのビート時刻を推定することができる。

【実施例】

【0058】

次に、本発明を実施するための実施例について図面を参照して説明する。図4に、本発明の実施例である脚式移動音楽ロボット（以下、音楽ロボットという。）の概略の正面図を示す。そして図5に、図4に示した音楽ロボットの概略の側面図を示す。図4において、音楽ロボット4は、基体部41と、これにそれぞれ可動連結される頭部42と、脚部43L, 43Rと、腕部44L, 44Rとを備えている。また、図5に示すように、音楽ロボット4は、背負う格好で収納部45を基体部41に装着している。

10

【0059】

図6に、主に音楽ロボット4の音楽インタラクションに関係する部分のブロック構成図を示す。同図において、音楽ロボット4は、ビートトラッキング装置1と、音楽認識装置100と、ロボット制御装置200とを備えている。なお、ここでのビートトラッキング装置1は、前述の実施形態におけるビートトラッキング装置を採用しているため、同一の符号を付している。そして、ビートトラッキング装置1と、音楽認識装置100と、ロボット制御装置200とは、収納部45に収納されている。

【0060】

音楽ロボット4の頭部42は、音楽ロボット4の周囲の音を集音するための耳機能部310を備えている。耳機能部310は、例えばマイクロホンを用いることができる。基体部41は、音楽ロボット4自身が発声すべき音を周囲に拡声するための発声機能部320を備えている。発声機能部320は、例えば音声信号を増幅するためのアンプ及びスピーカを用いることができる。脚部43L, 43Rは、脚機能部330を備えている。脚機能部330は、脚部43L, 43Rで上体を支えて自立させるだけでなく、二足歩行をさせたり足踏みさせたりする等、脚部43L, 43Rの動作を制御するものである。

20

【0061】

ビートトラッキング装置1は、前述の実施形態で説明したとおり、音楽ロボット4が音楽を聴いて得た音楽音響信号から、音楽ロボット4自身が発声する自己発声音による影響を抑制した音楽音響情報を抽出し、その音楽音響情報からテンポを推定するとともにビート時刻を推定するものである。ビートトラッキング装置1の自己発声音抑制部10は2チャンネル分の音声信号入力部を具備しており、その第1のチャンネルには、頭部42に備えられた耳機能部310から音楽音響信号MAが入力される。また、第2のチャンネルには、ロボット制御装置200から出力されて基体部41の発声機能部320に入力される自己音声信号SVの分岐された信号（これも自己音声信号SVという。）が入力される。

30

【0062】

音楽認識装置100は、ビートトラッキング装置1で推定されたテンポTPに基づいて、音楽ロボット4に歌唱させるための楽曲を決定し、その楽曲に関する楽曲情報をロボット制御装置200に出力するものである。音楽認識装置100は、音楽区間検出部110と、楽曲名同定部120と、楽曲情報検索部130と、楽曲データベース140とを備えている。

40

【0063】

音楽区間検出部110は、ビートトラッキング装置1から供給されたテンポTPに基づいて、安定したビート間隔が得られる時間を音楽区間として検出し、その音楽区間において音楽区間ステータス信号を出力するものである。具体的には、過去Aw個のフレームのうち、フレームxのビート間隔I(x)と現在の処理フレームtのビート間隔I(t)との差分が、ビート間隔の許容誤差よりも小さくなる関係を満たすフレームxの総数をNxとする。そのときのビート間隔安定度Sを数式11により求める。

【0064】

【数 1 1】

$$S = \frac{N_x}{A_w} \quad (\text{数式 1 1})$$

【 0 0 6 5 】

例えば、過去のフレーム数  $A_w = 300$  (約 3.5 秒に相当する。)、許容誤差  $\epsilon = 5$  (58 ミリ秒に相当する。)とした場合に、ビート間隔安定度  $S$  が 0.8 以上の場合に、音楽区間であると判定する。

10

【 0 0 6 6 】

楽曲名同定部 120 は、ビートトラッキング装置 1 から供給されたテンポ  $TP$  に最も近いテンポに対応する楽曲 ID を出力するものである。なお、本実施例では、各楽曲は各別のテンポを有するという前提とする。具体的には、楽曲名同定部 120 は、図 7 に示すような楽曲 ID テーブル 70 を予め有している。この楽曲 ID テーブル 70 は、60 M.M. から 120 M.M. までの複数のテンポにそれぞれ対応した楽曲 ID と、いずれのテンポにもマッチしない場合 (Unknown) に用いる楽曲 ID である「ID unknown」とを登録した表データである。同図の例によれば、楽曲 ID である ID001 - ID007 にそれぞれ対応する楽曲情報は、楽曲データベース 140 に格納されている。なお、テンポの単位である「M.M.」とは、1 分間あたりの四分音符の数を示すテンポ表記である。

20

【 0 0 6 7 】

楽曲名同定部 120 は、ビートトラッキング装置 1 から供給されたテンポ  $TP$  について、テンポ差が最も小さいテンポを楽曲 ID テーブル 70 から検索し、その検索されたテンポとテンポ  $TP$  との差分がテンポ差の許容値  $\epsilon$  以下である場合は、その検索されたテンポに関連づけられた楽曲 ID を出力する。一方、差分が許容値  $\epsilon$  よりも大きい場合は、「ID unknown」を楽曲 ID にして出力する。

【 0 0 6 8 】

楽曲情報検索部 130 は、楽曲名同定部 120 から供給された楽曲 ID が「ID unknown」でない場合は、その楽曲 ID をキーとして楽曲データベース 140 から楽曲情報を読み出し、音楽区間検出部 110 から供給される音楽区間ステータス信号のタイミングにしたがって出力する。楽曲情報は、例えば歌詞情報と、音の種類、長さ、音程等を含む譜面情報とを含むものである。楽曲情報は、楽曲 ID テーブル 70 の楽曲 ID (ID001 - ID007)、又はこれらの楽曲 ID と同一の ID に関連づけられて楽曲データベース 140 に格納されている。

30

【 0 0 6 9 】

一方、楽曲情報検索部 130 は、楽曲名同定部 120 から供給された楽曲 ID が「ID unknown」である場合は、楽曲データベース 140 には歌唱させるべき楽曲情報が格納されていないため、音楽ロボット 4 にスキヤットを発声させるためのスキヤット実行命令を、入来する音楽区間ステータス信号のタイミングにしたがって出力する。

40

【 0 0 7 0 】

ロボット制御装置 200 は、ビートトラッキング装置 1 で推定されたテンポ  $TP$  及びビート時刻  $BT$ 、並びに音楽認識装置 100 から供給された楽曲情報又はスキヤット実行命令に基づいて、ビート時刻に同期した歌唱若しくはスキヤット、若しくは足踏み、又はそれらを組み合わせた動作を行わせるものである。ロボット制御装置 200 は、ビート時刻予測部 210 と、歌唱制御部 220 と、スキヤット制御部 230 と、足踏み制御部 240 とを備えている。

【 0 0 7 1 】

ビート時刻予測部 210 は、ビートトラッキング装置 1 で推定されたテンポ  $TP$  及びビ

50

ート時刻  $B T$  に基づいて、音楽ロボット 4 における処理遅延時間を考慮した、現在時刻より未来のビート時刻の予測を行うものである。本実施例における処理遅延は、ビートトラック装置 1 における処理遅延、及びロボット制御装置 200 における処理遅延である。

【0072】

ビートトラック装置 1 における処理遅延は、主に数式 10 に示すビート時刻信頼度  $S(t)$  の計算処理と、ビート時刻推定アルゴリズムによるビート時刻  $T(n+1)$  の推定処理とに係るものである。すなわち、数式 10 において、フレーム  $t$  のビート時刻信頼度  $S(t)$  を算出するときには、全てのフレーム  $t_i$  が揃うまで待つ必要がある。フレーム  $t_i$  の最大値は、 $t + \max(I(t_i))$  で規定されるが、 $I(t_i)$  の最大値はビート時刻推定アルゴリズムの特性上、60M・M に対応するフレーム数であるため、正規化相互相関関数の窓長と同一の 1sec となる。ビート時刻推定処理では、 $t = T(n) + 3/4 \cdot I(t)$  の際のピーク抽出に  $T(n) + 3/2 \cdot I(t)$  までビート時刻信頼度が必要である。つまり、フレーム  $t$  のビート時刻信頼度が得られてから  $3/4 \cdot I(t)$  だけ待つ必要があり、この最大値は 0.75sec となる。

10

【0073】

また、ビートトラック装置 1 では、自己発音抑制部 10 における  $M$  フレーム分の遅延と、テンポ推定部 20 の Sobel フィルタ部 21 における 1 フレーム分の遅延とが生じるため、約 2sec の処理遅延時間が発生する。

【0074】

また、ロボット制御装置 200 における処理遅延は、主に歌唱制御部 220 における音声合成処理の処理による遅延である。

20

【0075】

そこで、ビート時刻予測部 210 は、ビート時刻推定部 30 で推定された最新のビート時刻  $B T$  にテンポ  $T P$  に係るビート間隔時間を外挿することにより、処理遅延時間よりも長い時間分先のビート時刻を予測する。

【0076】

具体的には、第 1 の例として数式 12 の計算によってビート時刻を予測することができる。数式 12 において、ビート時刻  $T(n)$  を、フレーム  $t$  までに推定されたビート時刻のうち最新のビート時刻とする。そして、数式 12 では、フレーム  $t$  より未来のビート時刻に対応するフレームのうち最もフレーム  $t$  に近いフレーム  $T'$  を計算する。

30

【0077】

【数 12】

$$T' = \begin{cases} T_{tmp} & \text{if } T_{tmp} \geq \frac{3}{2}I_m(t) + t \\ T_{tmp} + I_m(t) & \text{otherwise} \end{cases}$$

$$T_{tmp} = T(n) + I_m(t) + (t - T(n)) - \{(t - T(n)) \bmod I_m(t)\}$$

40

(数式 12)

【0078】

また、第 2 の例として、予め処理遅延時間が分かっている場合は、ビート時刻予測部 210 は、現在時刻から処理遅延時間分の時間を超えるまでテンポ  $T P$  をカウントし、超えたときのタイミングでビート時刻を外挿していく。図 8 に、第 2 の例によってビート時刻を外挿する様子を模式的に示す。同図の (a) (b) それぞれは、ビート時刻予測部 210 が、最新の推定ビート時刻である最新ビート時刻  $C B$  を取得したのちの現在時刻  $C T$  から処理遅延時間  $D T$  を超えた時刻に予測ビート時刻  $P B$  を外挿する様子を示している。同

50

図 ( a ) は、処理遅延時間  $D T$  に対して 1 ビート間隔が長いため 1 ビート間隔後に予測ビート時刻  $P B$  を外挿する様子を示している。そして、同図 ( b ) は、処理遅延時間  $D T$  に対して 1 ビート間隔が短く、3 ビート間隔後に予測ビート時刻  $P B$  を外挿する様子を示している。

【 0 0 7 9 】

また、第 3 の例として、ビート時刻予測部 2 1 0 は、予測ビート時刻が現在時刻からの処理遅延時間内に存在するときにその予測ビート時刻を確定予測ビートとして確定させる。但し、現在時刻より前に予測された最新の予測ビート時刻と、現在時刻からの処理遅延時間内に存在する最初の予測ビート時刻との時間間隔が所定の時間に満たないときは、処理遅延時間内に存在する予測ビート時刻の確定を行わない。

10

【 0 0 8 0 】

図 1 4 に、この第 3 の例によって予測ビート時刻を外挿する様子を模式的に示す。同図 ( a ) は、予測ビート時刻  $P B 2$  が現在時刻  $C T$  から処理遅延時間  $D T$  の時間内に存在する様子を示している。同図 ( a ) の例では、現在時刻  $C T$  より前に予測された最新の予測ビート時刻  $P B 1$  からテンポ  $T P$  のビート間隔の  $1 / 2$  の時間よりも先に予測ビート時刻  $P B 2$  が存在する。よって、この例の場合、ビート時刻予測部 2 1 0 は予測ビート時刻  $P B 2$  を確定予測ビートとして確定させる。

【 0 0 8 1 】

一方、図 1 4 ( b ) は、予測ビート時刻  $P B 2 a$  が現在時刻  $C T$  から処理遅延時間  $D T$  の時間内に存在する様子を示したものであるが、現在時刻  $C T$  より前に予測された最新の予測ビート時刻  $P B 1$  からテンポ  $T P$  のビート間隔の  $1 / 2$  の時間よりも前に予測ビート時刻  $P B 2$  が存在する様子を示している。よって、この例の場合、ビート時刻予測部 2 1 0 は予測ビート時刻  $P B 2$  を確定予測ビートとして確定させない。

20

【 0 0 8 2 】

図 1 4 のように、所定の時間をテンポ  $T P$  のビート間隔の  $1 / 2$  の時間とするのは好ましい例である。これは、音楽は、例えば四分音符と二分音符とが組み合わせられて表現される場合があり、ビート間隔が  $1 / 2$  倍や 2 倍に突然変化することがあり得るからである。また、第 3 の例を適用することによって、裏拍を強迫 ( ビート ) としてサンプリングしないようにする効果もある。

【 0 0 8 3 】

また、上述した第 1 - 第 3 の例による処理は、ビートトラッキング装置 1 でビートが推定されるたびに実行されるものであるが、音楽が無音になる等の理由によってビートが検出されない場合もある。その場合、ビートが検出されないまま、確定予測ビート時刻が現在時刻より過去になってしまう場合が起こり得る。第 4 の例として、ビート時刻予測部 2 1 0 は、最新の確定予測ビート時刻を起点にして予測処理を行うようにする。

30

【 0 0 8 4 】

図 1 5 に、第 4 の例によってビート時刻を外挿する様子を模式的に示す。同図は、ビート時刻予測部 2 1 0 が最新ビート時刻  $C B$  を取得したのを最後にビートが推定されず、予測ビート時刻  $O P B 1$  ,  $O P B 2$  を経て現在時刻  $C T$  になった例を示している。この場合、ビート時刻予測部 2 1 0 は、現在時刻  $C T$  より前に予測された最新の予測ビート時刻  $O P B 2$  を起点にして第 1 - 第 3 の例による予測処理を行うようにする。

40

【 0 0 8 5 】

歌唱制御部 2 2 0 は、音楽認識装置 1 0 0 の楽曲情報検索部 1 3 0 から供給される楽曲情報において、譜面情報における音符の時刻及び長さを、ビートトラッキング装置 1 で推定されたテンポ  $T P$  と、ビート時刻予測部 2 1 0 で予測された予測ビート時刻  $P B$  とに基づいて調整する。そして、歌唱制御部 2 2 0 は、楽曲情報の歌詞情報を用いて音声合成処理を行い、音声信号である歌声音声信号に変換して出力する。

【 0 0 8 6 】

スキャット制御部 2 3 0 は、音楽認識装置 1 0 0 の楽曲情報検索部 1 3 0 から供給されるスキャット実行命令を受信した場合に、「 D a b a D a b a D u b a 」や「 Z u n

50

Cha(ずん ちゃ)」等の予め記憶しているスキヤット用歌詞の発音タイミングを、ビートトラッキング装置1で推定されたテンポTPと、ビート時刻予測部210で予測された予測ビート時刻PBとに基づいて調整する。

【0087】

具体的には、スキヤット制御部230は、スキヤット用歌詞(例えば、「Daba」, 「Daba」, 「Duba」)から抽出したオンセットベクトル $d(t, f)$ の各ベクトル値の合計値のピークを「Daba」, 「Daba」, 「Duba」のスキヤットのビート時刻にする。そして、スキヤット制御部230は、各音でのビート時刻とスキヤットのビート時刻とを合わせて音声合成処理を行い、音声信号であるスキヤット音声信号に変換して出力する。

10

【0088】

歌唱制御部220から出力された歌声音声信号と、スキヤット制御部230から出力されたスキヤット音声信号とは、信号合成されて発声機能部320に供給されるとともに、ビートトラッキング装置1の自己発声音制御部10の第2チャンネルに供給される。なお、音楽区間検出部110から音楽区間ステータス信号が出力されている区間において、信号合成による自己音声信号を生成し出力するようにしてもよい。

【0089】

足踏み制御部240は、ビートトラッキング装置1で推定されたテンポTPと、ビート時刻予測部210で予測された予測ビート時刻PBと、音楽ロボット4の脚部43L, 43Rの端部である足部の地面への接地時刻とを用いたフィードバック則に基づいて足踏み動作のタイミングを生成する。

20

【0090】

次に、本実施例である音楽ロボット4を用いて音楽インタラクションの実験を行った結果について説明する。

【0091】

[実験1: ビートトラッキングの基本性能]

実験1用の評価データとして、RWC研究用音楽データベース(<http://staff.aist.go.jp/m.goto/RWC-MDB/>)におけるポピュラー音楽データベース(RWC-MDB-P-2001)のポピュラー音楽全100曲(日本語歌詞及び英語歌詞の楽曲)を使用した。各楽曲は、正しいビート時刻を容易に取得するため、MIDIデータを用いて生成した。ただし、MIDIデータは、得られたビート時刻の評価のためにのみ用いることとした。また、各楽曲の開始後30-90秒までの60秒間を評価データとして用いることとし、本実施例である音楽ロボット4による相互相関数ベースの方式と、自己相関数ベースの方式とでビートトラッキング成功率を比較した。なお、ビートトラッキング成功率の計算では、推定ビート時刻と正しいビート時刻との差が $\pm 100$ ms以内に収まっている場合に成功であるとした。具体的なビートトラッキング成功率 $r$ の計算例を数式13に示す。なお、 $N_{success}$ は推定成功ビート数であり、 $N_{total}$ は正しいビート総数である。

30

【0092】

【数13】

40

$$r = \frac{N_{success}}{N_{total}} \times 100 \quad (\text{数式13})$$

【0093】

[実験2: テンポ変化への追従速度]

実験2用の評価データとして、ポピュラー音楽データベース(RWC-MDB-P-2001)から生演奏録音の3曲を選択し、テンポ変化を含む音楽音響信号を作成した。具

50

体的には、楽曲番号 11, 18, 62 の楽曲を選択し (テンポはそれぞれ 90, 112, 81 M.M.)、これらの楽曲を No. 18 No. 11 No. 18 No. 62 の順に 60 秒ずつ区切ってつなげ合わせて 4 分間の音楽音響情報を作成した。この音楽音響情報を用いて、ビートトラッキングの遅延を、実験 1 と同様に本実施例と自己相関関数ベースの方式とで比較した。なお、ビートトラッキングの遅延時間は、実際にテンポが変化してからシステムがテンポ変化に追従するまでの時間とした。

【0094】

[実験 3 : ビート予測のノイズロバスト性能]

実験 3 用の評価データとして、ポピュラー音楽データベース (RWC-MDB-P-2001) の楽曲番号 62 の MIDI データを用いて生成したテンポが一定である楽曲を用いた。ただし、実験 1 と同様に、MIDI データはビート時刻の検証にのみ用いることとした。また、評価指標はビートトラッキング成功率を用いた。

【0095】

次に、上記実験 1 - 3 の実験結果について説明する。まず、実験 1 の結果を図 9, 10 のチャートに示す。図 9 は、本実施例についての、テンポに対するビートトラッキング成功率を示した実験結果である。また、図 10 は、自己相関関数ベースについての、同様の実験結果である。図 9, 10 において、ビートトラッキング成功率の平均値は、図 9 については約 79.5%、図 10 については約 72.8% であり、本実施例の方式の方が自己相関関数ベースよりも優れていることがわかる。

【0096】

また、図 9, 10 とともに、テンポが遅い場合のビートトラッキング成功率の低下がみられる。これは、テンポの遅い楽曲は、ドラム等のテンポ抽出のキーとなる楽器が少ない音楽であるためと推測される。しかしながら、テンポが 90 M.M. 付近を超える楽曲についての本実施例でのビートトラッキング成功率は 90% を超えており、本実施例のビートトラッキングの基本的性能が従来例よりも高いことがわかる。

【0097】

次に、実験 2 の結果を図 11 の平均遅延時間測定結果に示す。また、図 12 に、音楽ロボット 4 の電源がオフである場合のテンポ推定の実験結果をグラフで示す。図 11, 12 から明らかとなり、本実施例は、従来の自己相関関数ベースに比べてテンポ変化への適応が速いことがわかる。図 11 によれば、本実施例 (STPM 処理) は、自己相関関数ベース (自己相関処理) に対して、スキヤットをしない場合で 1/10 程度、スキヤットをする場合で 1/20 程度の時間短縮の効果があることがわかる。

【0098】

また、図 12 によれば、実際のテンポ (Actual Tempo) に対する本実施例の遅延時間は、Delay = 2 sec であるのに対し、自己相関関数ベースの遅延時間は、Delay = 約 20 sec となっている。なお、同図の 100 sec 付近でビートトラッキングが乱れているのは、ビート時刻にオンセットのない部分が評価データに一時的に存在しているためである。よって、本実施例では、一時的 (短時間) にテンポが不安定になる場合があるが、従来の自己相関関数ベースに比べて不安定期間は特段に短い。また、本実施例では、音楽認識装置 100 の音楽区間検出部 110 が音楽区間を検出し、ビートが抽出できない区間を非音楽区間であると判定するため、本実施例である音楽ロボット 4 ではこのような不安定期間の影響は極めて少ない。

【0099】

実験 3 の結果を図 13 のビート予測成功率に示す。同図によれば、自己発声音がその周期性のためにビートトラッキングに影響を与えていることと、自己発声音抑制機能がこのような周期性ノイズに効果的に作用していることを示している。

【0100】

以上により、本実施例である音楽ロボットによれば、ビートトラッキング装置を備えたことによって雑音に対するロバスト性を確保するとともに、テンポ変動への追従性とテンポ推定の安定性とを兼ね備えることができる。

10

20

30

40

50



## 【 0 1 0 1 】

また、本実施例である音楽ロボットによれば、処理遅延時間を考慮して推定ビート時刻から未来のビート時刻を予測するため、リアルタイムな音楽インタラクションを行うことができる。

## 【 0 1 0 2 】

なお、上述した実施形態におけるビートトラッキング装置の一部又は全部の機能をコンピュータで実現するようにしてもよい。この場合、その機能を実現するためのビートトラッキング用プログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたビートトラッキング用プログラムをコンピュータシステムに読み込ませ、実行することによって実現してもよい。なお、ここでいう「コンピュータシステム」とは、OS (Operating System) や周辺機器のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、光ディスク、メモリカード等の可搬型記録媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持するものを含んでもよい。また上記のプログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせにより実現するものであってもよい。

## 【 0 1 0 3 】

以上、本発明の実施形態について図面を参照して詳述したが、具体的な構成はこの実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計等も含まれる。

## 【 符号の説明 】

## 【 0 1 0 4 】

1 リアルタイム・ビートトラッキング装置 (ビートトラッキング装置)

4 脚式移動音楽ロボット (ロボット)

1 0 自己発声音抑制部 (自己音声抑制手段)

1 1 セミブラインド独立成分分析部 (S B - I C A 部)

2 0 テンポ推定部

2 1 S o b e l フィルタ部 (フィルタ手段)

2 2 時間周波数パターンマッチング部 (ビート間隔信頼度計算手段)

2 3 ビート間隔推定部 (ビート間隔推定手段)

3 0 ビート時刻推定部

3 1 近接ビート信頼度計算部 (ビート時刻信頼度計算手段)

3 2 連続ビート信頼度計算部 (ビート時刻信頼度計算手段)

3 3 ビート時刻信頼度計算部 (ビート時刻信頼度計算手段)

3 4 ビート時刻推定部 (ビート時刻推定手段)

4 1 基体部

4 2 頭部

4 3 L , 4 3 R 脚部

4 4 L , 4 4 R 腕部

4 5 収納部

1 0 0 音楽認識装置

1 1 0 音楽区間検出部 (音楽区間検出手段)

1 2 0 楽曲名同定部

1 3 0 楽曲情報検索部

1 4 0 楽曲データベース

2 0 0 ロボット制御装置

10

20

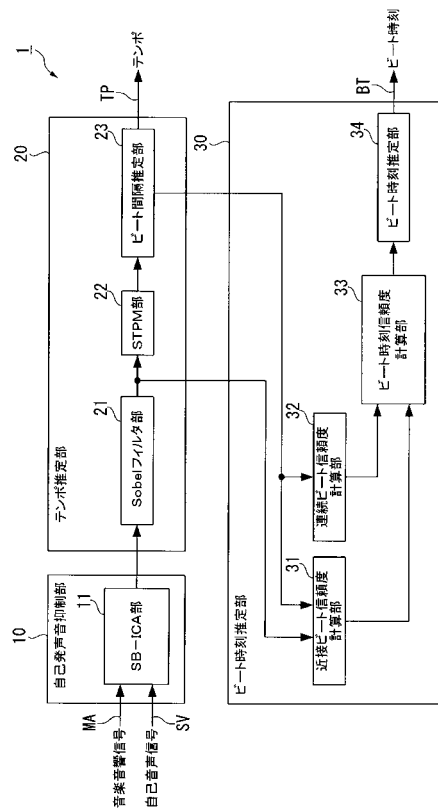
30

40

50

- 2 1 0 ビート時刻予測部
- 2 2 0 歌唱制御部
- 2 3 0 スキャット制御部
- 2 4 0 足踏み制御部
- 3 1 0 耳機能部
- 3 2 0 発声機能部
- 3 3 0 脚機能部
- MA 音楽音響信号
- SV 自己音声信号
- TP テンポ (ビート間隔)
- BT ビート時刻

【図 1】



SB-ICA部: セミブラインド独立成分分析部  
 STPM部: 時間周波数/クォンツメンツ部

【図 2】

```

EstimationBeatInterval(I1(t), I2(t))
Ic(t) ← I1(t)
if α Rpeak(t, I1) < Rpeak(t, I2)
  Id(t) ← | I1(t) - I2(t) |
  for n ← 2 to Nmax
    do {
      if | I1(t) - nId(t) | < δ or | I2(t) - nId(t) | < δ
        then {
          Ic(t) ← nId(t)
          break
        }
    }
I(t) ← EstimationBeatInterval(Ic(t), I(t-1))
return(I(t))
  
```

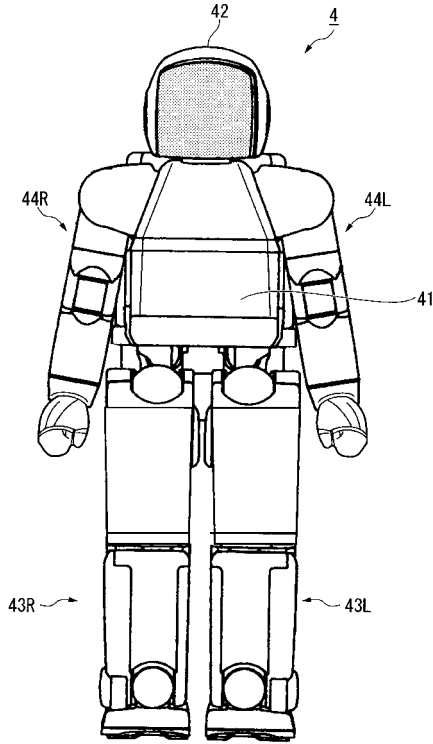
【図 3】

```

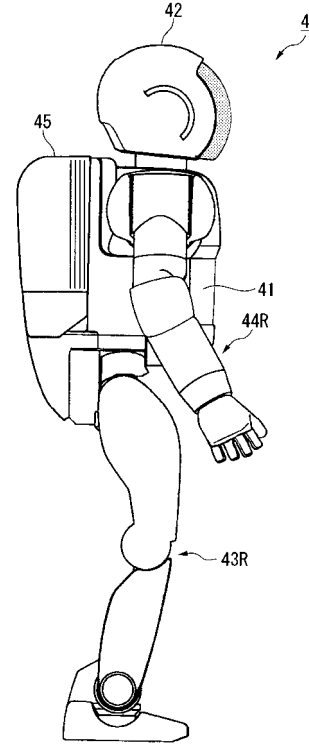
EstimationBeatTime(t, T(n), S(t), I(t))
procedure SearchPeaks(S(t), t, tr, Nmax)
  search Nmax peaks in S(t) within t ± 1/2 * tr
  comment: {t[i]} is an array of peak times.
           Np is the size of the array.
  return(t[i], Np)

main
T(n+1) ← nil
if t > T(n) + 3/4 * I(t)
  {
    (t[i], Np) ← SearchPeaks(S(t), T(n), I(t), 3)
    if Np > 0
      then {
        imin ← argmin( | T(n) + I(t) - t[i] | )
        T(n+1) ← t[imin]
      }
    else
      T(n+1) ← T(n) + I(t)
  }
return(T(n+1))
  
```

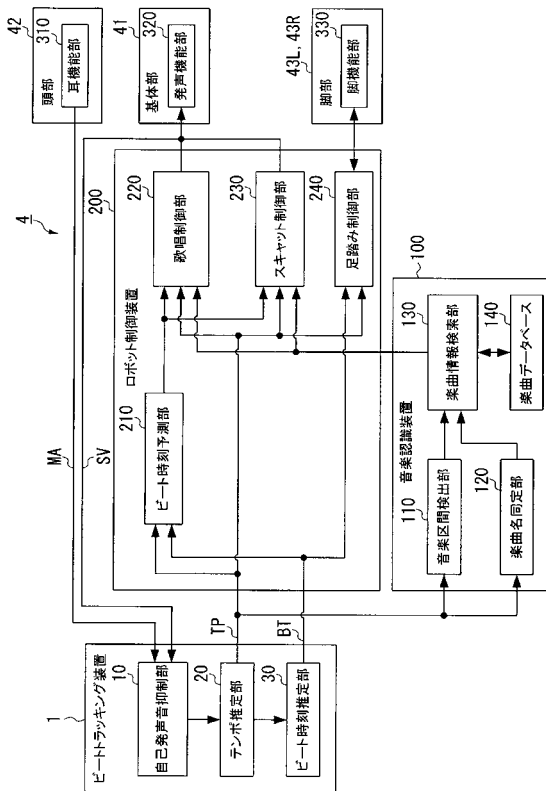
【図4】



【図5】



【図6】



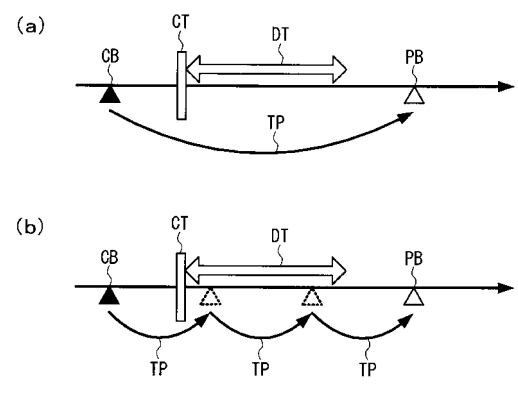
【図7】

70

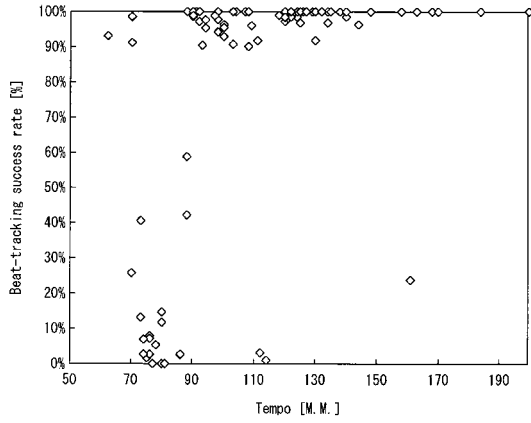
楽曲IDテーブル

テンポ	楽曲ID
60M. M.	ID001
70M. M.	ID002
80M. M.	ID003
...	...
120M. M.	ID007
Unknown	IDUnknown

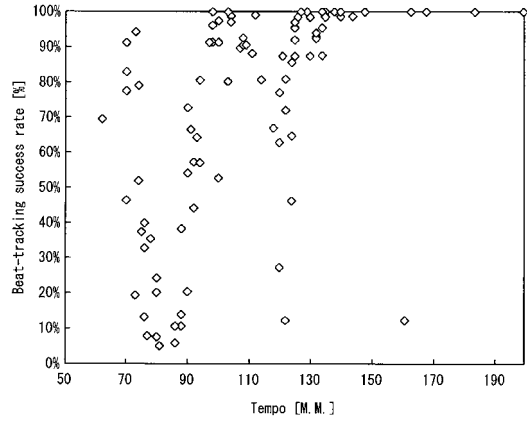
【図8】



【 図 9 】



【 図 10 】

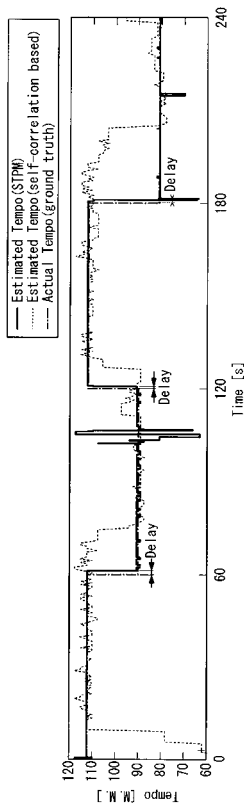


【 図 11 】

テンポ変化時からの平均遅延時間 単位:sec

ロボット電源	OFF		ON(足踏み動作あり)		
	なし	スキヤット	なし	スキヤット	歌唱
自己発声	なし	スキヤット	なし	スキヤット	歌唱
STPM処理	1.31	1.31	1.29	1.29	1.29
自己相関処理	11.24	29.91	14.66	20.43	N/A

【 図 12 】

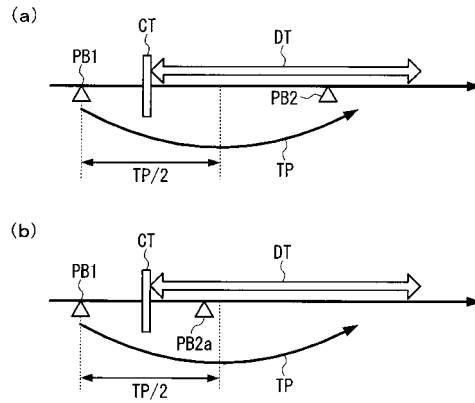


【 図 13 】

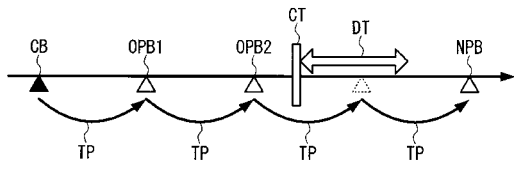
ビート予測の成功率 単位:%

ロボット電源	OFF		ON(足踏み動作あり)		
	なし	スキヤット	スキヤット	スキヤット	
自己発声	なし	スキヤット	スキヤット	スキヤット	
自己発声音抑制	—	あり	なし	あり	なし
成功率	73	76	54	77	54
裏拍の予測率	3	1	22	2	28

【 図 14 】



【 図 15 】



## フロントページの続き

(51)Int.Cl.		F I	
A 6 3 H 3/33 (2006.01)		A 6 3 H 3/33	C
A 6 3 H 11/00 (2006.01)		A 6 3 H 11/00	Z
G 1 0 L 21/0208 (2013.01)		G 1 0 L 21/02	1 0 2 B

- (72)発明者 中臺 一博  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 長谷川 雄二  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 辻野 広司  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 村田 和真  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 武田 龍  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 奥乃 博  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内

審査官 毛利 太郎

- (56)参考文献 特開 2 0 0 8 - 0 4 0 2 8 4 ( J P , A )  
特開 2 0 0 7 - 0 0 4 0 7 1 ( J P , A )  
特開 2 0 0 7 - 2 4 1 1 5 7 ( J P , A )  
特開 2 0 0 5 - 2 9 2 2 0 7 ( J P , A )  
特開 2 0 0 7 - 1 9 9 3 0 6 ( J P , A )  
特開 2 0 0 0 - 3 1 6 8 5 1 ( J P , A )  
特開平 1 1 - 2 8 2 4 8 3 ( J P , A )  
新田 恒雄 Tsuneo Nitta, 複合音響特徴平面 ( M A F P ) と L D A に基づく音響特徴抽出法  
Feature-Extraction Based on Multiple Acoustic-Feature Planes(MAFP) and LDA for Speech  
Recognition, 電子情報通信学会技術研究報告 V o l . 9 8 N o . 4 6 1 IEICE Technica  
I Report, 日本, 社団法人電子情報通信学会 The Institute of Electronics, Information and  
Communication Engineers, 1 9 9 8 年 1 2 月 1 1 日, p.63-69

## (58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 1 / 0 0 - 2 1 / 0 6  
G 1 0 G 1 / 0 0 - 7 / 0 2  
G 1 0 H 1 / 0 0 - 7 / 1 2  
A 6 3 H 3 / 3 3  
A 6 3 H 1 1 / 0 0