(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5582915号
(P5582915)

(45)発行日　平成26年9月3日(2014.9.3)　(24)登録日　平成26年7月25日(2014.7.25)

(51)Int.Cl.　　F I
G10G 3/04 (2006.01)　　G10G 3/04
G10L 25/51 (2013.01)　　G10L 25/51 300

請求項の数 11　(全 30 頁)

(21)出願番号　特願2010-177968(P2010-177968)
(22)出願日　平成22年8月6日(2010.8.6)
(65)公開番号　特開2011-39511(P2011-39511A)
(43)公開日　平成23年2月24日(2011.2.24)
審査請求日　平成24年11月27日(2012.11.27)
(31)優先権主張番号　61/234,076
(32)優先日　平成21年8月14日(2009.8.14)
(33)優先権主張国　米国(US)

(73)特許権者　000005326
本田技研工業株式会社
東京都港区南青山二丁目１番１号
(74)代理人　100064908
弁理士　志賀　正武
(74)代理人　100108578
弁理士　高橋　詔男
(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】楽譜位置推定装置、楽譜位置推定方法および楽譜位置推定ロボット

(57)【特許請求の範囲】
【請求項１】
音響信号取得部と、
前記音響信号取得部が取得する音響信号に対応する楽譜情報を取得する楽譜情報取得部と、
前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出部と、
前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出部と、
前記音響信号のビート位置を推定するビート位置推定部と、
前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング部と、
を備えることを特徴とする楽譜位置推定装置。
【請求項２】
前記音響信号の特徴量抽出部は、
フレーム時刻ｔ毎にバンドパスフィルタによって前記１つの楽音ｃ（ｉ，ｔ）（ｉは１～１２の整数）を抽出し、前記抽出した前記１つの楽音ｃ（ｉ，ｔ）に対して、次式

【数１】

$$\begin{aligned} c'(i,t) = &-c(i+1,t-1)-2c(i+1,t)-c(i+1,t+1) \\ &-c(i,t-1)+6c(i,t)+3c(i,t+1) \\ &-c(i-1,t-1)-2c(i-1,t)-c(i-1,t+1) \end{aligned} \quad \cdots(1)$$

の畳み込みを周期的に行い、前記畳み込みが行われたｃ’（ｉ，ｔ）に基づいて音響クロマベクトルを算出する請求項１に記載の楽譜位置推定装置。

【請求項３】

前記楽譜情報の特徴量抽出部は、前記楽譜情報から音符の出現頻度であるレアネスを算出し、

前記マッチング部は、前記レアネスを用いてマッチングを行う

ことを特徴とする請求項１または請求項２に記載の楽譜位置推定装置。

【請求項４】

前記マッチング部は、前記算出されたレアネスと前記抽出された音響信号の特徴量と楽譜情報の特徴量との積に基づきマッチングを行う

ことを特徴とする請求項３に記載の楽譜位置推定装置。

【請求項５】

前記レアネスは、フレーム内の所定の区間における前記音符の出現頻度である

ことを特徴とする請求項３または請求項４に記載の楽譜位置推定装置。

【請求項６】

前記マッチング部は、

前記楽譜情報において進むべきフレーム数Ｆに対して次式のように重み付けを行い（ただしf_mは前記楽譜情報のｍ番目のオンセットのフレーム、f_{m+k}は前記楽譜情報のｍ＋ｋ番目のオンセット時刻、ｋは進むべき前記楽譜情報のオンセット時刻、σは重み付けの分散値）、

【数２】

$$W(k) = \exp\left(-\frac{(f_{m+k}-f_m-F)^2}{2\sigma^2}\right) \quad \cdots(2)$$

前記楽譜情報におけるｍ番目のオンセットのフレームと、前記音響信号におけるｎ番目のオンセット時刻との類似性Ｓ（ｎ、ｍ）を算出し、

前記重み付けした値Ｗ（ｋ）と前記算出された類似性Ｓ（ｎ、ｍ）を用いて、次式の範囲内で探索を行うことで前記進むべき楽譜情報のオンセット時刻ｋを算出する

【数３】

$$k = \arg\max_{l} W(l)S(n+1,m+l) \quad \cdots(3)$$

ことを特徴とする請求項１から請求項５のいずれか１項に記載の楽譜位置推定装置。

【請求項７】

前記音響信号の特徴量抽出部は、前記音響信号の特徴量を、クロマベクトルを用いて抽出し、

前記楽譜情報の特徴量抽出部は、前記楽譜情報の特徴量を、クロマベクトルを用いて抽出する

ことを特徴とする請求項１から請求項６のいずれか１項に記載の楽譜位置推定装置。
【請求項８】
前記音響信号の特徴量抽出部は、抽出した音響信号の特徴量において高周波成分に重み付けを行い、重み付けした特徴量に基づき音符の出だしのタイミング時刻を算出し、
前記マッチング部は、算出された音符の出だしのタイミング時刻を用いて、マッチングを行う
ことを特徴とする請求項１から請求項７のいずれか１項に記載の楽譜位置推定装置。
【請求項９】
前記ビート位置推定部は、異なる複数の観測誤差モデルを、スイッチング・カルマン・フィルタにより切り替えることでビート位置の推定を行う
ことを特徴とする請求項１から請求項８のいずれか１項に記載の楽譜位置推定装置。
【請求項１０】
楽譜位置推定装置の楽譜位置推定方法において、
音響信号取得部が、音響信号を取得する音響信号取得工程と、
楽譜情報取得部が、前記音響信号に対応する楽譜情報を取得する楽譜情報取得工程と、
音響信号の特徴量抽出部が、前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出工程と、
楽譜情報の特徴量抽出部が、前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出工程と、
ビート位置推定部が、前記音響信号のビート位置を推定するビート位置推定工程と、
マッチング部が、前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング工程と、
を含むことを特徴とする楽譜位置推定方法。
【請求項１１】
音響信号取得部と、
前記音響信号取得部が取得した音響信号に対して抑圧処理を行うことで、演奏に対応する音響信号を抽出する音響信号分離部と、
前記音響信号分離部が抽出した音響信号に対応する楽譜情報を取得する楽譜情報取得部と、
前記音響信号分離部が抽出した前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出部と、
前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出部と、
前記音響信号分離部が抽出した音響信号のビート位置を推定するビート位置推定部と、
前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング部と、
を備えることを特徴とする楽譜位置推定ロボット。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、楽譜位置推定装置、楽譜位置推定方法および楽譜位置推定ロボットに関する。
【背景技術】
【０００２】
近年、ロボットの身体的な機能の顕著な向上により、人間社会に関わる家事や看護を行

う人の援助などが試みられている。このように、日々の場面で人間とロボットとが共存していく上で、ロボットには人間と自然な相互作用ができるようにしていくことが求められている。

【０００３】

人間とロボットとの相互作用におけるコミュニケーションとして、音楽を通したコミュニケーションがある。音楽は、人間同士のコミュニケーションにおいても重要な役割を果たし、例えば言葉が通じ合わない人間同士でも音楽を通じて友好的で楽しい時間を共有することができる場合もある。このため、ロボットには、音楽を通して人間と相互作用を行うことができることが、人間と調和して共存していく上で重要になってくる。

【０００４】

ロボットが音楽を通して人間とコミュニケーションを行う場面として、例えば、ロボットが、伴奏または歌声に合わせて歌ったり、音楽に合わせて自身の胴体を動かしたりすることが考えられる。

【０００５】

このようなロボットにおいて、楽譜情報を解析し、解析結果に基づいて動作することが知られている。

楽譜に記載されている音符が、どの音符かを認識する技術として、楽譜の画像データを音符データに変換して、楽譜を自動認識する技術が提案されている（例えば、特許文献１参照）。また、楽譜データと予めグルーピングされた構造分析データにもとづき、楽曲データの拍節構造を分析し、演奏されている音響信号からテンポを推定する技術として、ビートトラッキング法が提案されている（例えば、特許文献２参照）。

【先行技術文献】

【特許文献】

【０００６】

【特許文献１】特許第３１４７８４６号公報

【特許文献２】特開２００６－２０１２７８号公報

【発明の概要】

【発明が解決しようとする課題】

【０００７】

特許文献２の拍節構造を分析する技術では、楽譜に基づく構造のみを分析していた。このため、ロボット自身が集音した音響信号に合わせてロボットに歌唱を行わせようとしても、音楽が途中から開始されると、楽譜のどの部分が不明であるので、演奏されている曲のビート時間やテンポの抽出に失敗する場合もあるという問題点があった。また、人間が行う演奏の場合、演奏のテンポが変動することもあり、この結果、ロボットが、演奏されている曲のビート時間やテンポの抽出に失敗する場合もあるという問題点があった。

以上のように、従来技術では、楽譜データに基づいて曲の拍節構造やビート時間やテンポを抽出していたので、実際の演奏が行われている場合、楽譜のどの位置が演奏されているかを精度良く検出することができなかった。

【０００８】

本発明は、上記の問題点に鑑みてなされたものであって、演奏に対する楽譜位置の推定を行う楽譜位置推定装置、楽譜位置推定方法および楽譜位置推定ロボットを提供することを課題としている。

【課題を解決するための手段】

【０００９】

上記目的を達成するため、本発明に係る楽譜位置推定装置は、音響信号取得部と、前記音響信号取得部が取得する音響信号に対応する楽譜情報を取得する楽譜情報取得部と、前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出部と、前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出部と、前記音響信号のビート位置を推定す

るビート位置推定部と、前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング部と、を備えることを特徴としている。

また、本発明に係る楽譜位置推定装置において、前記音響信号の特徴量抽出部は、フレーム時刻ｔ毎にバンドパスフィルタによって前記１つの楽音ｃ（ｉ，ｔ）（ｉは１～１２の整数）を抽出し、前記抽出した前記１つの楽音ｃ（ｉ，ｔ）に対して、次式

【数１】

$$\begin{aligned} c'(i,t) = &-c(i+1,t-1) - 2c(i+1,t) - c(i+1,t+1) \\ &- c(i,t-1) + 6c(i,t) + 3c(i,t+1) \\ &- c(i-1,t-1) - 2c(i-1,t) - c(i-1,t+1) \end{aligned} \quad \cdots(1)$$

の畳み込みを周期的に行い、前記畳み込みが行われたｃ’（ｉ，ｔ）に基づいて音響クロマベクトルを算出するようにしてもよい。

【００１０】

また、本発明に係る楽譜位置推定装置において、前記楽譜情報の特徴量抽出部は、前記楽譜情報から音符の出現頻度であるレアネスを算出し、前記マッチング部は、前記レアネスを用いてマッチングを行うようにしてもよい。

【００１１】

また、本発明に係る楽譜位置推定装置において、前記マッチング部は、前記算出されたレアネスと前記抽出された音響信号の特徴量と楽譜情報の特徴量との積に基づきマッチングを行うようにしてもよい。

【００１２】

また、本発明に係る楽譜位置推定装置において、前記レアネスは、フレーム内の所定の区間における前記音符の出現頻度であるようにしてもよい。

また、本発明に係る楽譜位置推定装置において、前記マッチング部は、前記楽譜情報において進むべきフレーム数Ｆに対して次式のように重み付けを行い（ただしf_mは前記楽譜情報のｍ番目のオンセットのフレーム、f_{m+k}は前記楽譜情報のｍ＋ｋ番目のオンセット時刻、ｋは進むべき前記楽譜情報のオンセット時刻、σは重み付けの分散値）、

【数２】

$$W(k) = \exp\left(-\frac{(f_{m+k} - f_m - F)^2}{2\sigma^2} \right) \quad \cdots(2)$$

前記楽譜情報におけるｍ番目のオンセットのフレームと、前記音響信号におけるｎ番目のオンセット時刻との類似性Ｓ（ｎ、ｍ）を算出し、

前記重み付けした値Ｗ（ｋ）と前記算出された類似性Ｓ（ｎ、ｍ）を用いて、次式の範囲内で探索を行うことで前記進むべき楽譜情報のオンセット時刻ｋを算出するようにしてもよい

【数３】

$$k = \arg\max_{l} W(l) S(n+1, m+l) \quad \cdots(3)$$

【００１３】

また、本発明に係る楽譜位置推定装置において、前記音響信号の特徴量抽出部は、前記

音響信号の特徴量を、クロマベクトルを用いて抽出し、前記楽譜情報の特徴量抽出部は、前記楽譜情報の特徴量を、クロマベクトルを用いて抽出するようにしてもよい。

【００１４】

また、本発明に係る楽譜位置推定装置において、前記音響信号の特徴量抽出部は、抽出した音響信号の特徴量において高周波成分に重み付けを行い、重み付けした特徴量に基づき音符の出だしのタイミング時刻を算出し、前記マッチング部は、算出された音符の出だしのタイミング時刻を用いて、マッチングを行うようにしてもよい。

【００１５】

また、本発明に係る楽譜位置推定装置において、前記ビート位置推定部は、異なる複数の観測誤差モデルを、スイッチング・カルマン・フィルタにより切り替えることでビート位置の推定を行うようにしてもよい。

【００１６】

上記目的を達成するため、本発明に係る楽譜位置推定装置における楽譜位置推定方法は、音響信号取得部が、音響信号を取得する音響信号取得工程と、楽譜情報取得部が、前記音響信号に対応する楽譜情報を取得する楽譜情報取得工程と、音響信号の特徴量抽出部が、前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出工程と、楽譜情報の特徴量抽出部が、前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出工程と、ビート位置推定部が、前記音響信号のビート位置を推定するビート位置推定工程と、マッチング部が、前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング工程と、を含むことを特徴としている。

【００１７】

上記目的を達成するため、本発明に係る楽譜位置推定ロボットは、音響信号取得部と、前記音響信号取得部が取得した音響信号に対して抑圧処理を行うことで、演奏に対応する音響信号を抽出する音響信号分離部と、前記音響信号分離部が抽出した音響信号に対応する楽譜情報を取得する楽譜情報取得部と、前記音響信号分離部が抽出した前記音響信号に含まれる音階を構成する楽音のうちの１つの楽音に隣接する他の楽音のパワーを減じ、さらに前のフレーム時のパワーを減じて当該１つの楽音を強調し、前記強調された楽音を用いて当該音響信号の特徴量を抽出する音響信号の特徴量抽出部と、前記楽譜情報の特徴量を抽出する楽譜情報の特徴量抽出部と、前記音響信号分離部が抽出した音響信号のビート位置を推定するビート位置推定部と、前記推定されたビート位置を用いて、前記音響信号の特徴量と前記楽譜情報の特徴量とのマッチングを行うことで、前記音響信号が対応する前記楽譜情報における位置を推定するマッチング部と、を備えることを特徴としている。

【発明の効果】

【００１８】

本発明によれば、取得した音響信号から特徴量とビート位置とを抽出し、取得した楽譜情報から特徴量を抽出する。そして、抽出したビート位置を用いて、音響情報の特徴量と楽譜情報の特徴量とをマッチングすることで、音響信号が対応する前記楽譜情報における位置を推定するようにした。この結果、音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、楽譜情報から音符の出現頻度であるレアネスを算出し、算出したレアネスを用いてマッチングを行うようにしたので、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、レアネスと音響信号の特徴量と楽譜情報の特徴量との積に基づいてマッチングを行うようにしたので、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、音符の出現頻度の低さをレアネスとして用いるようにしたので、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、音響信号の特徴量と楽譜情報の特徴量とをクロマベクトルを用いて抽出するようにしたので、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、音響信号の特徴量において高周波成分に重み付けを行い、重み付けした特徴量に基づき音符の出だしのタイミング時刻を用いて、マッチングを行うようにしたので、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

本発明によれば、異なる複数の観測誤差モデルを、スイッチング・カルマン・フィルタにより切り替えることでビート位置の推定を行うようにしたので、演奏が楽譜通りのテンポから外れた場合においても、精度良く音響信号に基づいて、楽譜位置を正確に推定することが可能になる。

【図面の簡単な説明】

【００１９】

【図１】本実施形態に係る楽譜位置推定装置１００を組み込んだロボット１の一例を説明する図である。

【図２】同実施形態に係る楽譜位置推定装置１００のブロック図の一例を示す図である。

【図３】楽器演奏時の音響信号のスペクトラムの一例を示す図である。

【図４】楽器演奏時の音響信号の残響波形（パワーエンベロープ）の一例を示す図である。

【図５】実際の演奏に基づく音響信号と楽譜のクロマベクトルの一例を示す図である。

【図６】音楽演奏におけるスピード、又はテンポの変化を示したものである。

【図７】同実施形態に係る楽譜位置推定部１２０の構成を説明するブロック図である。

【図８】同実施形態に係る音響信号からの特徴量抽出部４１０がクロマベクトルとオンセット時刻と抽出する際に用いる式における記号を説明するリストである。

【図９】同実施形態に係る音響信号と楽譜からクロマベクトルを算出する過程を説明する図である。

【図１０】同実施形態に係るオンセット時刻抽出手順の概略を説明する図である。

【図１１】同実施形態に係るレアネスを説明する図である。

【図１２】同実施形態に係るカルマン・フィルタを適用したビートトラッキングを説明する図である。

【図１３】同実施形態に係る楽譜位置推定処理のフローチャートである。

【図１４】楽譜位置推定装置を備えるロボット１と音源の設置関係を説明する図である。

【図１５】２種類の音楽信号（（ｖ）と（ｖｉ））と４つの手法（（ｉ）～（ｉｖ））の結果を示している。

【図１６】クリーン信号時の各手法の累積絶対値誤差平均値で分類された楽曲数を示している。

【図１７】残響あり信号時の各手法の累積絶対値誤差平均値で分類された楽曲数を示している。

【発明を実施するための形態】

【００２０】

以下、図面を用いて本発明の実施形態について詳細に説明する。なお、本発明は斯かる実施形態に限定されず、その技術思想の範囲内で種々の変更が可能である。

【００２１】

図１は、本実施形態における楽譜位置推定装置１００を組み込んだロボット１の一例を説明する図である。ロボット１は、図１に示すように、基体部１１と、基体部１１にそれぞれ可動連結される頭部１２（可動部）と、脚部１３（可動部）と、腕部１４（可動部）とを備えている。また、ロボット１は、背負う格好で基体部１１に収納部１５を装着している。なお、基体部１１には、スピーカ２０が収納され、頭部１２にはマイクロホン３０が収納されている。なお、図１は、ロボット１を側面から見た図であり、マイクロホン３０およびスピーカ２０は、例えば正面から左右対称にそれぞれ複数収納されている。

【００２２】

図２は、本実施形態における楽譜位置推定装置１００のブロック図の一例を示す図である。図２のように、楽譜位置推定装置１００にはマイクロホン３０、スピーカ２０が接続されている。また、楽譜位置推定装置１００は、音響信号分離部１１０と、楽譜位置推定部１２０と、歌声生成部１３０とを備えている。また、音響信号分離部１１０は、自己生成音抑制フィルタ部１１１を備え、楽譜位置推定部１２０は、楽譜データベース１２１と楽曲位置推定部１２２を備え、歌声生成部１３０は、歌詞とメロディーのデータベース１３１と音声生成部１３２を備えている。

【００２３】

マイクロホン３０は、演奏（伴奏）の音と、ロボット１自身のスピーカ２０を介して出力した音声信号（歌声）とが混合された音を集音し、集音した音を音響信号に変換して音響信号分離部１１０に出力する。

音響信号分離部１１０には、マイクロホン３０から集音された音響信号と、歌声生成部１３０から生成された音声信号とが入力される。音響信号分離部１１０の自己生成音抑制フィルタ部１１１は、入力された音響信号に対して、独立成分分析（ＩＣＡ；Ｉｎｄｅｐｅｎｄｅｎｔ　Ｃｏｍｐｏｎｅｎｔ　Ａｎａｌｙｓｉｓ；）を行って、生成された音声信号と音響信号に含まれる残響音を抑圧する。これにより、音響信号分離部１１０は、演奏に関わる音響信号を分離して抽出する。音響信号分離部１１０は、抽出した音響信号を楽譜位置推定部１２０に出力する。

【００２４】

楽譜位置推定部１２０（楽譜情報取得部、音響信号の特徴量抽出部、楽譜情報の特徴量抽出部、ビート位置推定部、マッチング部）には、音響信号分離部１１０から分離された音響信号が入力される。楽譜位置推定部１２０の楽曲位置推定部１２２は、入力された音響信号から特徴量である音響クロマベクトルとオンセット時刻を算出する。楽曲位置推定部１２２は、楽譜データベース１２１から演奏されている曲の楽譜データを読み出し、楽譜データから特徴量である楽譜クロマベクトルと音符の出現頻度であるレアネスを算出する。楽曲位置推定部１２２は、入力された音響信号からビートトラッキングを行い、リズム間隔（テンポ）を検出する。楽曲位置推定部１２２は、抽出したリズム間隔（テンポ）に基づき、スイッチング・カルマン・フィルタ（ＳＫＦ；Ｓｗｉｔｃｈｉｎｇ　Ｋａｌｍａｎ　Ｆｉｌｔｅｒ）を用いて、テンポの外れ値やノイズ分を推定し、安定したリズム間隔（テンポ）を抽出する。楽曲位置推定部１２２（音響信号の特徴量抽出部、楽譜情報の特徴量抽出部、ビート位置推定部、マッチング部）は、抽出したリズム間隔（テンポ）と、算出した音響クロマベクトル、オンセット時刻情報、楽譜クロマベクトルおよびレアネスとを用いて、演奏による音響信号と楽譜とのマッチングを行う。つまり、楽曲位置推定部１２２は、演奏されている曲が、楽譜のどの位置であるかを推定する。楽譜位置推定部１２０は、推定した楽譜位置を示す楽譜位置情報を歌声生成部１３０に出力する。

なお、楽譜データベース１２１に予め楽譜データが記憶されている例を説明したが、楽楽譜位置推定部１２０は、入力された楽譜データを楽譜データベース１２１に書き込んで記憶させるようにしてもよい。

【００２５】

歌声生成部１３０には、推定された楽譜位置情報が入力される。歌声生成部１３０の音声生成部１３２は、入力された楽譜位置情報に基づき、歌詞とメロディーのデータベース１３１に記憶されている情報を用いて、公知の手法により演奏に合わせた歌声の音声信号を生成する。歌声生成部１３０は、生成した歌声の音声信号を、スピーカ２０を介して出力する。

【００２６】

次に、音響信号分離部１１０が、独立成分分析を用いて、生成された音声信号と音響信号に含まれる残響音を抑圧する概要について説明する。独立成分分析は、音源同士の独立性（確率密度）を仮定して分離を行う。ロボット１がマイクロホン３０を介して取得した音響信号は、演奏されている音とロボット１がスピーカ２０から出力した音声信号とが混合された信号である。この混合された信号のうち、ロボット１がスピーカ２０から出力し

た音声信号は、音声生成部１３２で生成した信号であるため既知である。このため、音響信号分離部１１０は、周波数領域で独立成分分析を行い、混合された信号に含まれるロボット１の音声信号を抑圧することで、演奏されている音を分離する。

【００２７】

次に、本実施形態における、楽譜位置推定装置１００に採用した技術の概略を説明する。演奏されている音楽（伴奏）からビートやテンポを抽出して、楽譜のどの位置が演奏されているのかを推定する場合、大きく分けると３つの技術がある。

【００２８】

第１の技術は、演奏されている音響信号に含まれている様々な楽器音の差異をどのように判別するかである。図３は、楽器演奏時の音響信号のスペクトラムの一例を示す図である。図３（ａ）は、ピアノでＡ４の音（４４０［Ｈｚ］）を鳴らしたときの音響信号のスペクトラムを表し、図３（ｂ）は、フルートでＡ４の音を鳴らしたときの音響信号のスペクトラムを表している。縦軸は信号の大きさを表し、横軸は周波数を表している。図３（ａ）と図３（ｂ）のように、同じ周波数範囲で分析した各スペクトラムは、同じ基本周波数４４０［Ｈｚ］のＡ４の音でも、楽器によってスペクトラムの形状や成分が異なっている。

【００２９】

図４は、楽器演奏時の音響信号の残響波形（パワーエンベロープ）の一例を示す図である。図４（ａ）は、ピアノにおける音響信号の残響波形を表し、図４（ｂ）は、フルートにおける音響信号のスペクトラムを表す。縦軸は信号の大きさを表し、横軸は時間を表している。通常、楽器の残響波形は、アタック(出だし)部分（２０１，２１１）、減衰部分（２０２，２１２）、持続部分（２０３，２１３）、リリース(消滅)部分（２０４，２１４）から構成されている。図４（ａ）のように、ピアノやギターのような楽器の残響波形は、下降的な持続部分２０３を有し、図４（ｂ）のように、フルートやヴァイオリン、サキソフォンなどの楽器の残響波形は、永続性の持続部分２１３を有している。

複合的な音符が様々な楽器から同時に演奏された場合、言い換えれば、和音音響を扱う場合、各音符の基本周波数を検出すること、又は持続音を認識することは、更に難しくなる。

【００３０】

このため、本実施形態では、演奏における波形の出だしであるオンセット時刻（２０５，２１５）に着目する。

楽譜位置推定部１２０は、１２段階のクロマベクトル（音響特徴量）を用いて周波数領域の特徴量を抽出する。そして、楽譜位置推定部１２０は、抽出した周波数領域の特徴量に基づき、時間領域の特徴量であるオンセット時刻を算出する。クロマベクトルの利点としては、様々な楽器のスペクトル形状の変化への頑健性と和音音響信号に対する有効性が挙げられる。クロマベクトルは、基本周波数の代わりに各１２音名、つまり、Ｃ、Ｃ＃、…、Ｂなどのパワーを抽出する。本実施形態では、図４（ａ）の出だし部分２０５、図４（ｂ）の出だし部分２１５のように、パワーの急激な上昇周辺の頂点を「オンセット時刻」と定義する。オンセット時刻の抽出は、楽譜同期を行う上で、各音符のスタート時間を得るために必要である。さらに和音音響信号では、オンセット時刻の抽出は、時間領域におけるパワーの上昇部分として、持続部分やリリース部分より簡単に抽出できる。

【００３１】

次に、第２の技術は、演奏されている音響信号と楽譜の相違を推定するかである。図５は、実際の演奏に基づく音響信号と楽譜のクロマベクトルの一例を示す図である。図５（ａ）は、楽譜のクロマベクトルを表し、図５（ｂ）は、実際の演奏に基づく音響信号のクロマベクトルを表している。図５（ａ）と図５（ｂ）における縦軸は、１２段階の音の種類を表し、図５（ａ）における横軸は楽譜におけるビートを表し、図５（ｂ）における横軸は時間を表している。また、図５（ａ）と図５（ｂ）において、縦実線３１１は、各音（音符）のオンセット時刻を表している。楽譜中のオンセット時刻は、各音符フレームの開始部分として定義される。

図５（ａ）と図５（ｂ）のように、実際の演奏による音響信号に基づくクロマベクトルと楽譜に基づくクロマベクトルには差異が見られる。実線で囲まれている符号３０１の領域では、図５（ａ）ではクロマベクトルが存在せず、図５（ｂ）ではクロマベクトルが存在している。すなわち、楽譜中には音符が無い部分にもかかわらず、実際の演奏においては、前の音のパワーが持続している。点線で囲まれている符号３０２の領域では、逆に、図５（ａ）ではクロマベクトルが存在しているのに、図５（ｂ）ではクロマベクトルがほとんど検出できない。

さらに、楽譜において、各音符の音量は明示されていない。

【００３２】

以上により、本実施形態において、ほとんど使用されない音名の音符は、音響信号において、時として顕著に表されるという考えに基づき、音響信号と楽譜との相違を軽減する。まず、演奏される曲の楽譜を予め取得して、楽譜データベース１２１に登録しておく。そして、楽曲位置推定部１２２は、演奏される曲の楽譜を解析し、各音符の使用頻度を各々算出する。この楽譜中の各音名の出現頻度をレアネス（ｒａｒｅｎｅｓｓ）と定義する。レアネスの定義は情報エントロピーに類似している。図５（ａ）において、音名Ｂの数は他の音名の数より少ないため、音名Ｂのレアネスは高い。対照的に、音名Ｃや音名Ｅは楽譜中で頻繁に使用されているためレアネスは低い。

さらに、楽曲位置推定部１２２は、このように算出した各音名に対して、算出したレアネスに基づき重み付けを行う。

このように重み付けを行うことで、低頻出音符は高頻出音符に比べて和音音響信号からより簡単に抽出される可能性がある。

【００３３】

次に、第３の技術は、演奏されている音響信号のテンポの変動を推定するかである。安定したテンポ推定は、ロボット１が正確に楽譜に同期して歌唱を実行するだけでなく、演奏されている曲に合わせてロボット１が滑らかで心地よい歌声を出力することにとっても不可欠である。人間が行う演奏においては、楽譜で指示されているテンポから外れる場合もある。さらに、公知のビートトラッキングを用いたテンポ推定時にも発生する。

図６は、音楽演奏におけるスピード、又はテンポの変化を示したものである。図６（ａ）は、人間の演奏に厳密に一致させたＭＩＤＩ（登録商標（Ｍｕｓｉｃａｌ　Ｉｎｓｔｒｕｍｅｎｔ　Ｄｉｇｉｔａｌ　Ｉｎｔｅｒｆａｃｅ；電子楽器デジタルインタフェース））データから算出したビートの時間変動を示す図である。各テンポは楽譜中の音符の長さをその時間の長さで分割して得られる。図６（ｂ）は、ビートトラッキングにおけるビートの時間変動を示す図である。テンポ列は相当数の外れ値を含む。外れ値は一般的にドラムのパターンの変化によって引き起こされる。図６において、縦軸は、時間あたりのビート数を表し、横軸は時間を表している。

このため、本実施形態では、楽曲位置推定部１２２は、テンポ推定にスイッチング・カルマン・フィルタ（ＳＫＦ）を用いる。ＳＫＦは、誤りを含む一連のテンポから、次のテンポ推定を可能にする。

【００３４】

次に、楽譜位置推定部１２０が行う処理について、図７～図１２を用いて、詳細に説明する。図７は、楽譜位置推定部１２０の構成を説明するブロック図である。図７のように、楽譜位置推定部１２０は、楽譜データベース１２１と楽曲位置推定部１２２とを備えている。また、楽曲位置推定部１２２は、音響信号からの特徴量抽出部４１０（音響信号の特徴量抽出部）と、楽譜からの特徴量抽出４２０（楽譜情報の特徴量抽出部）と、ビート間隔(テンポ)算出４３０と、マッチング部４４０と、テンポ推定部４５０（ビート位置推定部）を備えている。また、マッチング部４４０は、類似度計算部４４１と重み付け計算部４４２を備えている。また、テンポ推定部４５０は、小さな観測誤差モデル４５１と外れ値となる大きな観測誤差モデル４５２を備えている。

【００３５】

［音響信号からの特徴量抽出］

音響信号からの特徴量抽出部４１０には、音響信号分離部１１０により分離された音響信号が入力される。音響信号からの特徴量抽出部４１０は、入力された音響信号から、音響クロマベクトルとオンセット時刻とを抽出し、抽出したクロマベクトルとオンセット時刻情報をビート間隔(テンポ)算出４３０に出力する。

図８は、音響信号からの特徴量抽出部４１０がクロマベクトルとオンセット時刻情報と抽出する際に用いる式における記号を説明するリストである。図８において、ｉは、西洋音階における１２音（Ｃ、Ｃ＃、Ｄ、Ｄ＃、Ｅ、Ｆ、Ｆ＃、Ｇ、Ｇ＃、Ａ、Ａ＃、Ｂ）の名前のインデックスである。ｔは、音響信号のフレーム時間である。ｎは、音響信号におけるオンセット時刻のためのインデックスである。t_nは、音響信号におけるｎ番目のオンセット時刻である。ｆは、楽譜のフレーム・インデックスである。ｍは、楽譜におけるオンセット時刻のためのインデックスである。f_mは、楽譜におけるｍ番目のオンセット時刻である。

【００３６】

音響信号からの特徴量抽出部４１０は、短時間フーリエ変換（ＳＴＦＴ；ｓｈｏｒｔ－ｔｉｍｅ　Ｆｏｕｒｉｅｒ　ｔｒａｎｓｆｏｒｍａｔｉｏｎ）を用いて、入力された音響信号からスペクトラムを算出する。短時間フーリエ変換は、入力された音響信号にハニング等の窓関数を音声信号に乗じて有限期間内で、解析位置をシフトしながらスペクトラムを算出する技術である。本実施形態では、ハニング窓が４０９６［ポイント]、変位間隔が５１２［ポイント]、サンプリングレートが４４．１［kHz］の設定を用いた。ここでフレーム時間ｔ、周波数ωの時のｐ（ｔ，ω）をパワーとする。

クロマベクトルｃ（ｔ）＝［ｃ（１，ｔ），ｃ（２，ｔ），…，ｃ（１２，ｔ）］^{T}（^{T}はベクトルの転置を意味する）はフレーム時間ｔ毎に生成される。図９のように、音響信号からの特徴量抽出部４１０が各音名のバンドパス・フィルタにより１２音名のうちの1つに対応した各成分を抽出し、抽出した１２音名のうちの1つに対応した各成分は数式（１）のように表される。図９は、音響信号と楽譜からクロマベクトルを算出する過程を説明する図であり、図９（ａ）は、音響信号からクロマベクトルを算出する過程を説明する図である。

【００３７】

【数１】

$$c(i,t)=\sum_{h=Oct_L}^{Oct_H}\int_{-\infty}^{\infty}BPF_{i,h}(\omega)p(t,\omega)d\omega \qquad \cdots(1)$$

【００３８】

式（１）において、$BPF_{i,h}$は、ｈ番目のオクターブにおける音名ｉのバンドパス・フィルタである。また、Oct_LとOct_Hは、それぞれ考慮される下限オクターブ及び上限オクターブである。周波帯のピークは、音の基本周波数である。周波帯の端は、隣接する音の周波数である。例えば、基本周波数４４０［Ｈｚ］である音“Ａ４”（第４オクターブの音“Ａ”)のＢＰＦは、４４０［Ｈｚ］にその周波帯のピークがある。その周波帯の一端は、“Ｇ＃”（第４オクターブの音“Ｇ＃”)の４１５［Ｈｚ］であり、“Ａ＃”の４６６［Ｈｚ］である。本実施形態では、Oct_L＝３及びOct_H＝７とする。言い換えれば、下限音は“Ｃ３”の１３１［Ｈｚ］とし、上限音は“Ｂ７”、３９５１［Ｈｚ］とした。

次に、音響信号からの特徴量抽出部４１０は、音名を強調するために、式（１）に対して、次式（２）の畳み込みを行う。

【００３９】

【数２】

$$c'(i,t) = -c(i+1,t-1) - 2c(i+1,t) - c(i+1,t+1) - c(i,t-1) + 6c(i,t) + 3c(i,t+1) - c(i-1,t-1) - 2c(i-1,t) - c(i-1,t+1) \quad \cdots(2)$$

【００４０】

音響信号からの特徴量抽出部４１０は、式（２）の畳み込みを、ｉに対して周期的に処理を行う。例えば、ｉ＝１（音名は“Ｃ”）のとき、ｃ（ｉ－１，ｔ）は、ｃ（１２，ｔ）（音名は“Ｂ”）に置き換えられる。

式（２）の畳み込みにより、隣接する音名のパワーを減じ、他よりパワーを持つ成分が強調され、画像処理におけるエッジ抽出に類似する可能性がある。前のフレーム時間のパワーが減じられることで、パワーの増加量は強調される。

次に、音響信号からの特徴量抽出部４１０は、次式（３）により、音響信号から音響クロマベクトル$c_{sig}(i,t)$（ｉ，ｔ）を算出することで特徴量を抽出する。

【００４１】

【数３】

$$c_{sig}(i,t) = \begin{cases} c'(i,t) & (c'(i,t) > 0) \\ 0 & otherwise \end{cases} \quad \cdots(3)$$

【００４２】

次に、音響信号からの特徴量抽出部４１０は、入力された音響信号からＲｏｄｅｔ他により提案されたオンセット抽出手法（手法１）を使用して、オンセット時刻を抽出する。

【００４３】

文献１（手法１）　X. Rodet and F. Jaillet. Detection and modeling of fast attack transients. In International Computer Music Conference, pages 30-33, 2001.

【００４４】

オンセット抽出において、特に高周波領域に位置するオンセット時刻のパワー増加量を利用する。音階のある楽器の音のオンセット時刻は、ドラムのような打楽器のオンセット時刻に比べ、より高い周波領域に重心がある。このように、この手法は、音階のある楽器のオンセット時刻検出に特に効果的である。

まず、音響信号からの特徴量抽出部４１０は、高周波成分と呼ばれるパワーを次式（４）により算出する。

【００４５】

【数４】

$$h(t) = \sum_{\omega} \omega p(t,\omega) \quad \cdots(4)$$

【００４６】

高周波成分は重み付けされたパワーであり、その重みは周波数に対して直線的に増加する。音響信号からの特徴量抽出部４１０は、図１０のように、オンセット時刻t_nを、中央値フィルタを用いてｈ（ｔ）のピークを選択することにより判断する。図１０は、オンセット時刻抽出手順の概略を説明する図である。図１０のように、入力された音響信号の

スペクトラムを算出した後（図１０（ａ））、音響信号からの特徴量抽出部４１０は、高周波成分に重み付けしたパワーを算出する（図１０（ｂ））。そして、音響信号からの特徴量抽出部４１０は、重み付けしたパワーに対して中央値フィルタを適用し、パワーのピーク部分の時刻をオンセット時刻として算出する（図１０（ｃ））。

【００４７】

音響信号からの特徴量抽出部４１０は、抽出した音響クロマベクトルとオンセット時刻情報とをマッチング部４４０に出力する。

【００４８】

［楽譜からの特徴量抽出］

楽譜からの特徴量抽出４２０は、楽譜データベース１２１に記憶されている楽譜のなかから必要な楽譜データを読み出す。なお、本実施形態では、予め演奏される曲名がロボット１に入力されているとし、楽譜からの特徴量抽出４２０は、指定されている曲の楽譜データを選択して読み出す。

次に、楽譜からの特徴量抽出４２０は、読み出した楽譜データを、図９（ｂ）のように、１小節の４８分の１と同等の長さのフレームに分割する。このフレーム解決法では、６分音符や３連音符の処理が可能である。本実施形態では、楽譜のクロマベクトルを、次式（５）を用いて算出することで特徴量を抽出する。図９（ｂ）は、楽譜からクロマベクトルを算出する過程を説明する図である。

【００４９】

【数５】

$$c_{sco}(i,m)=\begin{cases}1 & pitch\ name\ i\ starts\ at\ frame\ f_m\\ 0 & otherwise\end{cases} \quad \cdots(5)$$

【００５０】

式（５）において、f_mは、楽譜中のm番目のオンセット時刻を表している。

次に、楽譜からの特徴量抽出４２０は、抽出したクロマベクトルから、フレームf_mにおける各音名（ｉ）のレアネスｒ（ｉ，ｍ）を、次式（７）により算出する。

【００５１】

【数６】

$$n(i,m)=\frac{\sum_{p\in M} c_{sco}(i,p)}{\sum_{i=1}^{12}\sum_{p\in M} c_{sco}(i,p)} \quad \cdots(6)$$

【００５２】

【数７】

$$r(i,m)=\begin{cases}-\log_2 n(i,m) & (n(i,m)>0)\\ \max_i(-\log_2 n(i,m)) & (n(i,m)=0)\end{cases} \quad \cdots(7)$$

【００５３】

Ｍはフレームf_mに中心があり、長さが２小節分のフレーム範囲を意味している。したがって、ｎ（ｉ，ｍ）は、フレームf_m周辺の各音名の分布を表している。

楽譜からの特徴量抽出４２０は、抽出した楽譜のクロマベクトルとレアネスとをマッチ

ング部４４０に出力する。

【００５４】

図１１は、レアネスを説明する図である。図１１（ａ）～図１１（ｃ）において、縦軸は音名を表し、横軸は時間を表している。図１１（ａ）は楽譜のクロマベクトルを表す図であり、図１１（ｂ）は演奏された音響信号のクロマベクトルを表す図である。図１１（ｃ）～図１１（ｅ）は、レアネスの算出方法を説明する図である。

図１１（ｃ）のように、楽譜からの特徴量抽出４２０は、図１１（ａ）の楽譜クロマベクトルについて、フレーム毎に前後２小節区間で、各音の出現頻度（使用頻度）を計算する。そして、図１１（ｄ）のように、楽譜からの特徴量抽出４２０は、前後２小節区間における各音名ｉの使用頻度p_iを算出する。次に、図１１（ｅ）のように、楽譜からの特徴量抽出４２０は、式（７）を用いて算出した各音名ｉの使用頻度ｐｉの対数を取ってレアネスr_iを算出する。式（７）および図１１（ｅ）のように、$-\log p_i$は、使用頻度が低い音名ｉを抽出することを意味している。

【００５５】

楽譜からの特徴量抽出４２０は、抽出した楽譜クロマベクトルとレアネスとをマッチング部４４０に出力する。

【００５６】

［ビートトラッキング］

ビート間隔(テンポ)算出４３０は、村田らにより開発されたビートトラッキング手法（手法２）を用いて、入力された音響信号からビート間隔(テンポ)を算出する。

【００５７】

文献２（手法２） K. Murata, K. Nakadai, K. Yoshii, R. Takeda, T. Torii, H. G. Okuno, Y. Hasegawa, and H. Tsujino. A robot uses its own microphone to synchronize its steps to musical beats while scatting and singing. In IROS, pages 2459-2464, 2008.

【００５８】

まず、ビート間隔(テンポ)算出４３０は、周波数が直線的音階にあるスペクトログラムｐ（ｔ，ω）は、６４段階のメル尺度に周波数があるｐｍｅｌ（ｔ，φ）に次式（９）を用いて変換する。ビート間隔(テンポ)算出４３０は、オンセットのベクトルｄ（ｔ，φ）を、次式（８）を用いて算出する。なお、式（８）で式（９）用いているファイ、すなわち、ｄ（ｔ、ファイ）のファイと、ｐｍｅｌ（ｔ，φ）およびｄ（ｔ，φ）で用いているφ（ファイ）は同じである。

【００５９】

【数８】

$$d(t,\varphi)=\begin{cases} p_{mel}^{sobel}(t,\varphi) & \left(p_{mel}^{sobel}(t,\varphi)>0\right) \\ 0 & otherwise \end{cases} \quad \cdots(8)$$

【００６０】

【数９】

$$\begin{aligned} p_{mel}^{sobel}(t,\varphi) = & -p_{mel}(t-1,\varphi+1)+p_{mel}(t+1,\varphi+1) \\ & -2p_{mel}(t-1,\varphi)+2p_{mel}(t+1,\varphi) \\ & -p_{mel}(t-1,\varphi-1)+p_{mel}(t+1,\varphi-1) \end{aligned} \quad \cdots(9)$$

【００６１】

式（９）は、ゾーベル・フィルタによるオンセット強調を意味する。

次に、ビート間隔(テンポ)算出４３０は、ビート間隔(テンポ)推定を行う。ビート間隔(テンポ)算出４３０は、ビート間隔の信頼性Ｒ（ｔ，ｋ）を、正規化相互相関を用いて次式（１０）により算出する。

【００６２】

【数１０】

$$R(t,k)=\frac{\sum_{j}\sum_{l=0}^{P_w-1} d(t-l,j)\,d(t-k-l,j)}{\sqrt{\sum_{j}\sum_{k=l}^{P_w-1} d(t-l,j)^2 \sum_{j}\sum_{l=0}^{P_w-1} d(t-k-l,j)^2}} \quad \cdots(10)$$

【００６３】

式（１０）において、P_wは、信頼性算出におけるウィンドウの長さであり、ｋは時間シフトパラメータである。ビート間隔(テンポ)算出４３０は、ビート間隔Ｉ（ｔ）を時間シフト値ｋに基づいて判断する。また、ビート間隔の信頼性Ｒ（ｔ，ｋ）は、局所的なピークの値をとる。

【００６４】

ビート間隔(テンポ)算出４３０は、このように算出したビート間隔(テンポ)情報をテンポ推定部４５０に出力する。

【００６５】

［音響信号と楽譜のマッチング］

マッチング部４４０には、音響信号からの特徴量抽出部４１０が抽出した音響クロマベクトルとオンセット時刻情報と、楽譜からの特徴量抽出４２０が抽出した楽譜クロマベクトルとレアネスと、テンポ推定部４５０が推定した安定化したテンポ情報とが入力される。マッチング部４４０は、（t_n，f_m）を最終マッチング対とする。t_nは音響信号における時間、f_mは楽譜のフレーム・インデックスである。t_{n+1}で検出された音響信号の新しいオンセット時刻及びその時間のテンポを考える場合、楽譜中の進むべきフレームの数Ｆは、マッチング部４４０により次式（１１）のように推定される。

【００６６】

【数１１】

$$F=A\left(t_{n+1}-t_n\right) \quad \cdots(11)$$

【００６７】

式（１１）において、係数Ａは、テンポに対応し、音楽が速く進むと、係数Ａは大きくなる。また、楽譜フレームf_{m+k}の重み付けを次式（１２）のように定義する。

【００６８】

【数１２】

$$W(k)=\exp\left(-\frac{\left(f_{m+k}-f_m-F\right)^2}{2\sigma^2}\right) \quad \cdots(12)$$

【００６９】

式（１２）、ｋは進むべき楽譜中のオンセット時刻数であり、σは重み付けの分散値である。本実施形態では、σ=24として実行したが、これは音符の長さの半分に相当する。ここでｋは、負数となる可能性もあることに留意する。負数ｋの場合、（t_{n+1}，f_{m-1}）のようなマッチングを考慮することになるが、それはマッチングが楽譜内で逆行することを意味する。

【００７０】

マッチング部４４０は、対（t_n，f_m）の類似性を、次式（１３）を用いて算出する。

【００７１】

【数１３】

$$s(n,m)=\sum_{i=1}^{12}\sum_{\tau=t_n}^{t_{n+1}} r(i,m)c_{sco}(i,m)c_{sig}(i,\tau) \quad \cdots(13)$$

【００７２】

式（１３）において、ｉは音名、ｒ（ｉ，ｍ）はレアネス、c_{sco}及びc_{sig}はそれぞれ楽譜及び音響信号から生成されたクロマベクトルである。すなわち、マッチング部４４０は、対（t_n，f_m）の類似性を、レアネスと音響クロマベクトルと楽譜クロマベクトルの積に基づいて算出する。

最終マッチングが（t_n，f_m）の時、新しいマッチングは（t_{n+1}，f_{m+k}）となり、そのとき進むべき楽譜中のオンセット時刻数ｋは、次式（１４）である。

【００７３】

【数１４】

$$k=\underset{l}{\arg\max}\, W(l)S(n+1,m+l) \quad \cdots(14)$$

【００７４】

本実施形態では、マッチング部４４０が行う処理の実行中、各マッチングステップの進むべき楽譜中のオンセット時刻数ｋの探索範囲は、計算コスト削減のため２小節内に制限する。

【００７５】

マッチング部４４０は、式（１１）～（１４）を用いて、最終マッチング対（t_n，f_m）を算出し、算出した最終マッチング対（t_n，f_m）を歌声生成部１３０に出力する。

【００７６】

［スイッチング・カルマン・フィルタを用いたテンポ推定］

マッチング結果とビートトラッキング手法によるテンポ推定における２種類の誤差に対処するため、テンポ推定部４５０は、スイッチング・カルマン・フィルタ（手法３）を使用してテンポ推定を行う。

【００７７】
文献３（手法３） K. P. Murphy. Switching kalman filters. Technical report, 1998.
【００７８】
テンポ推定部４５０が対応すべき２つの誤差とは、「演奏スピードのわずかな変化による小さな誤差」と「ビートトラッキングによるテンポ推定の外れ値による誤差」である。テンポ推定部４５０は、スイッチング・カルマン・フィルタで構成され、小さな観測誤差モデル４５１と外れ値となる大きな観測誤差モデル４５２の２つのモデルを備える。
スイッチング・カルマン・フィルタとは、カルマン・フィルタ（ＫＦ）を拡張したものである。カルマン・フィルタは、状態遷移モデルと観測モデルを有する直線的予測フィルタであり、状態が観測不能の時、離散時間系列内の誤差を含む観測値からその状態を推定する。スイッチング・カルマン・フィルタは、複合的な状態遷移モデル及び観測モデルを有する。スイッチング・カルマン・フィルタが観測値を得る毎に、モデルを各モデルの可能性に基づき自動的に切り替える。
本実施形態において、スイッチング・カルマン・フィルタが備える小さな観測誤差モデル４５１と外れ値となる大きな観測誤差モデル４５２の２つのモデルにおいて、状態遷移などの他のモデリング成分は、２つのモデルに共通している。
【００７９】
ビート時間とビート間隔を推定するため、本実施形態では、Ｃｅｍｇｉｌ他により提案されたＳＫＦモデル（手法４）を使用する。
【００８０】
文献４（手法４） A. T. Cemgil, B. Kappen, P. Desain, and H. Honing. On tempo tracking: Tempogram representation and kalman filtering. Journal of New Music Research, 28:4:259-273, 2001.
ｋ番目のビート時間をb_k、その時間のビート間隔をΔ_k、そのテンポを一定とする。次のビート時間は$b_{k+1} = b_k + \Delta_k$、次のビート間隔は$\Delta_{k+1} = \Delta_k$として表される。ここで状態ベクトルを$x_k = [b_k \ \Delta_k]^T$とすると、状態遷移は次式（１５）のように表される。
【００８１】
【数１５】

$$x_{k+1} = F_k x_k + v_k = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_k + v_k \qquad \cdots(15)$$

【００８２】
式（１５）において、F_kは状態遷移行列であり、v_kは平均０の正規分散と共分散行列Ｑから導かれた遷移誤差ベクトルである。最新の状態をx_kと仮定すると、テンポ推定部４５０は、次のビート時間b_{k+1}をx_{k+1}の最初の成分として次式（１６）を用いて推定する。
【００８３】
【数１６】

$$x_{k+1} = F_k x_k \qquad \cdots(16)$$

【００８４】
ここで、観測ベクトルを$z_k = [b_k', \Delta_k']^T$とする。$b_k'$は、マッチング部

４４０がマッチング結果により算出したビート時間であり、Δ_k'は、ビート間隔(テンポ)算出４３０がビートトラッキングにより算出したビート間隔である。テンポ推定部４５０は、観測ベクトルを、次式（１７）を用いて算出する。

【００８５】

【数１７】

$$z_k = H_k x_k + w_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x_k + w_k \qquad \cdots(17)$$

【００８６】

式（１７）において、H_kは観測行列、w_kは平均０の正規分散と共分散行列Ｒから導かれた観測誤差ベクトルである。本実施形態において、テンポ推定部４５０により、ＳＫＦは、観測誤差共分散行列R^i（ｉ＝１，２）を切り換える。ここで、ｉはモデル数である。予備実験から、本実施形態では、R^iを以下の通りとした。小さな誤差モデルをR^1＝ｄｉａｇ（０．０２，０．００５）、外れ値モデルをR^2＝ｄｉａｇ（１，０．１２５）とし、ここでｄｉａｇ（a_1，・・・，a_n）は対角要素が左上から右下にa_1，・・・，a_nであるｎ×ｎの対角行列である。

【００８７】

図１２は、カルマン・フィルタを適用したビートトラッキングを説明する図である。縦軸はテンポ、横軸は時間を表している。図１２（ａ）は、ビートトラッキングにおける誤差を説明する図であり、図１２（ｂ）は、ビートトラッキング手法のみの解析結果とカルマン・フィルタ適用後の解析結果を示す図である。図１２（ａ）において、符号５０１の部分は小ノイズであり、符号５０２の部分がビートトラッキング手法で推定したテンポにおける外れ値の例である。

図１２（ｂ）において、実線５１１は、ビートトラッキング手法のみによるテンポの解析結果であり、点線５１２は、ビートトラッキング手法による解析結果に対し、さらに本実施形態の方法でカルマン・フィルタを適用した解析結果である。図１２（ｂ）のように、本実施形態の方法を適用した結果、ビートトラッキング手法のみと比較して、テンポの外れの影響を大幅に改善できる。

【００８８】

［ビート時間の観測］

図９（ｂ）で説明したように、楽譜が４８分音符に対応した長さのフレームに分割されているので、ビートは楽譜中の１２フレーム毎にある。テンポ推定部４５０は、算出したビート時間b_k'が、ｋ番目のビートフレームに音符が存在しないとき、マッチング部４４０が行ったマッチング結果により補間する。

テンポ推定部４５０は、算出したビート時間b_k'とビート間隔情報をマッチング部４４０に出力する。

【００８９】

［楽譜位置推定処理の手順］

次に、楽譜位置推定装置１００が行う楽譜位置推定処理の手順を、図１３を用いて説明する。図１３は、楽譜位置推定処理のフローチャートである。

まず、楽譜からの特徴量抽出４２０は、楽譜データベース１２１から楽譜データを読み出す。楽譜からの特徴量抽出４２０は、読み出した楽譜データから、式（５）～式（７）を用いて楽譜クロマベクトルとレアネスを算出し、算出した楽譜クロマベクトルとレアネスをマッチング部４４０に出力する（ステップＳ１）。

【００９０】

次に、楽曲位置推定部１２２は、マイクロホン３０が集音した音響信号に基づき、演奏

が継続しているか否かを判別する（ステップＳ２）。なお、この判別は、例えば、楽曲位置推定部１２２が、音響信号が継続している場合に曲が継続していると判別し、または、演奏されている曲の演奏位置が楽譜の終端ではない場合に演奏が継続していると判別する。

ステップＳ２の判別の結果、演奏が継続していないと判別された場合（ステップＳ２；Ｎｏ）、楽譜位置推定の処理を終了する。

【００９１】

ステップＳ２の判別の結果、演奏が継続していると判別された場合（ステップＳ２；Ｙｅｓ）、音響信号分離部１１０は、マイクロホン３０が集音した音響信号を、例えば１秒間分、音響信号分離部１１０が備えるバッファに記憶させる（ステップＳ３）。

次に、音響信号分離部１１０は、入力された音響信号と歌声生成部１３０が生成した音声信号を用いて、独立成分分析を行って残響音の抑圧と自身の歌声の抑圧を行うことで音響信号を抽出し、抽出した音響信号を楽譜位置推定部１２０に出力する。

次に、ビート間隔(テンポ)算出４３０は、入力された音楽信号に基づき、式（８）～式（１０）を用いてビートトラッキング手法によりビート間隔(テンポ)を推定し、推定したビート間隔(テンポ)をマッチング部４４０に出力する（ステップＳ４）。

【００９２】

次に、音響信号からの特徴量抽出部４１０は、入力された音響信号から式（４）を用いて、オンセット時刻情報を検出し、検出したオンセット時刻情報をマッチング部４４０に出力する（ステップＳ５）。

次に、音響信号からの特徴量抽出部４１０は、入力された音響信号に基づき、式（８）～式（３）を用いて音響クロマベクトルを抽出し、抽出した音響クロマベクトルをマッチング部４４０に出力する（ステップＳ６）。

【００９３】

次に、マッチング部４４０には、音響信号からの特徴量抽出部４１０から音響クロマベクトルとオンセット時刻情報と、楽譜からの特徴量抽出４２０から楽譜クロマベクトルとレアネスと、テンポ推定部４５０から推定された安定したテンポ情報とが入力される。マッチング部４４０は、式（１１）～式（１４）を用いて、入力された音響クロマベクトルと楽譜クロマベクトルとを、逐次、マッチング処理を行い、最終マッチング対（t_n，f_m）を推定する。マッチング部４４０は、推定した楽譜位置に対応する最終マッチング対（t_n，f_m）をテンポ推定部４５０と歌声生成部１３０に出力する（ステップＳ７）。

【００９４】

次に、テンポ推定部４５０には、ビート間隔(テンポ)算出４３０から入力されたビート間隔(テンポ)情報に基づき、式（１５）～式（３）を用いてビート時間b_k’とビート間隔情報を算出し、算出したビート時間b_k’とビート間隔情報をマッチング部４４０に出力する（ステップＳ８）。

また、テンポ推定部４５０には、マッチング部４４０から最終マッチング対（t_n，f_m）が入力される。テンポ推定部４５０は、算出したビート時間b_k’が、ｋ番目のビートフレームに音符が存在しないとき、マッチング部４４０が行ったマッチング結果により補間する。

なお、マッチング部４４０とテンポ推定部４５０とは、マッチング処理とテンポ推定を逐次的に行い、マッチング部４４０が、最終マッチング対（t_n，f_m）を推定する。

【００９５】

歌声生成部１３０の音声生成部１３２は、入力された最終マッチング対（t_n，f_m）に基づき、歌詞とメロディーのデータベース１３１を参照し、楽譜位置に合致する歌詞をメロディーに合わせて歌声を生成する。なお、ここで、「歌声」とは、楽譜位置推定装置１００からスピーカ２０を介して出力される音声データである。すなわち、楽譜位置推定装置１００を備えるロボット１のスピーカ２０を介して出力されるものであるので、便宜的に「歌声」という。また、本実施形態において、音声生成部１３２には、ＶＯＣＡＬＯＩＤ２（ＶＯＣＡＬＯＩＤ（登録商標））を用いて、歌声を生成した。ＶＯＣＡＬＯＩＤ

２（ＶＯＣＡＬＯＩＤ（登録商標））は、メロディーと歌詞を入力することでサンプリングされた人の声を元にした歌声を合成することができるエンジンのため、本実施形態では、さらに楽譜位置を情報として加え、実際の演奏から歌声が外れないようにしている。
音声生成部１３２は、生成した音声信号をスピーカ２０から出力する。

【００９６】
また、最終マッチング対（t_n，f_m）推定後、ステップＳ２～ステップＳ８を曲の演奏が終了するまで逐次的に行う。
以上の処理により、楽譜位置を推定し、推定した楽譜位置に合致する音声（歌声）を生成し、生成した音声をスピーカ２０から出力することで、ロボット１が演奏に合わせた歌唱を行うことが可能になる。また、本実施形態によれば、演奏されている音響信号に基づいて、楽譜の位置を推定するようにしたので、曲が途中から開始された場合においても、正確に楽譜の位置を推定することができる。

【００９７】
［評価結果］
本実施形態における楽譜位置推定装置１００を用いて行った評価結果について説明する。まず、実験条件について説明する。評価に用いた楽曲は、後藤らにより作成されたＲＷＣ研究用音楽データベース（ＲＷＣ－ＭＤＢ－Ｐ－２００１；http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html）からのポピュラー音楽100曲を使用した。また、使用した楽曲については、歌唱部分や演奏部分を含むこれらの楽曲のフルバージョンを使用した。

【００９８】
楽譜同期の正解データは、評価者が各楽曲のＭＩＤＩファイルから生成した。これらＭＩＤＩファイルは、実際の演奏に厳密に同期される。誤差は、秒単位で、本実施形態により抽出されたビート時間と、正解データとの相違の絶対値として定義される。誤差は楽曲毎に平均化される。

【００９９】
評価は、以下の４種類について行い、評価結果を比較した。
（ｉ）本実施形態の方法；ＳＫＦおよびレアネス使用
（ｉｉ）ＳＫＦ無使用；テンポ推定への修正なし
（ｉｉｉ）レアネス無使用；全音符は同等のレアネスを有する状態
（ｉｖ）ビートトラッキング手法；この手法は、音楽の最初からビートを数えることにより楽譜位置を判断する。

【０１００】
さらに、楽譜位置推定装置１００のマイクロホン３０が集音する音は、室内環境における残響に影響を与えるのかをも評価するため、以下の２種類の音楽信号を使用して評価を行った。
（ｖ）クリーン音楽信号：残響なしの音楽信号
（ｖｉ）残響あり音楽信号：残響つきの音楽信号
残響は、インパルス応答畳み込みによりシミュレートしたものを使用した。図１４は、楽譜位置推定装置１００を備えるロボット１と音源の設置関係を説明する図である。図１４のように、評価用の音源は、ロボット１の正面から１００［ｃｍ］離した位置に設置したスピーカ６０１から出力した音源を用いた。この生成されたインパルス応答は、実験室において測定した。実験室での残響時間（ＲＴ２０）は，１５６［ｍｓｅｃ］である。講堂又は音楽ホールであれば、より長い残響時間となると考えられる。

【０１０１】
図１５は、２種類の音楽信号（（ｖ）と（ｖｉ））と４つの手法（（ｉ）～（ｉｖ））の結果を示している。各値は１００曲についての累積絶対値誤差の平均値と標準偏差である。クリーン信号及び残響あり信号の双方において、本実施形態による方法（ｉ）の誤差は、ビートトラッキング手法（ｉｖ）の誤差より少ない。本実施形態による方法（ｉ）は，誤差をクリーン信号で２９％、残響あり信号で１４％改善している。本実施形態による方法(ｉ)は、ＳＫＦを無使用の手法(ｉｉ)より誤差が少ないことから、ＳＫＦを用いるこ

とで誤差低減されていることがわかる。同様に、本実施形態の方法(ｉ)とレアネス無使用の手法(ｉｉｉ)の結果を比較すると、レアネスが誤差を減少させている。

さらに、ＳＫＦ無使用の手法(ｉｉ)は、レアネス無使用の手法(ｉｉｉ)より誤差が大きいので、ＳＫＦはレアネスより一層効果的であると言える。これは、しばしばレアネスが、楽譜中のフレームとドラム音のような誤ったオンセット時刻との間で高い類似性を誘引するからである。仮にドラム音が、高いレアネスを伴い、クロマベクトル成分中に大きなパワーを持つとすると、これが誤ったマッチングとなる。この問題を避けるため、楽譜位置推定装置１００では、単一音名でなく組み合わせた音名へのレアネスの考慮をすることが可能である。

【０１０２】

図１６は、クリーン信号時の各手法の累積絶対値誤差平均値で分類された楽曲数を示している。図１７は、残響あり信号時の各手法の累積絶対値誤差平均値で分類された楽曲数を示している。図１６と図１７において、より少ない平均誤差を有する楽曲数が多いほど、よりよい演奏を示している。クリーン信号では、本実施形態の方法（ｉ）では、２秒以下の誤差を有する楽曲が３１曲あるのに対し、ビートトラッキングのみの手法（ｉｖ）では９曲であった。

残響あり信号では、同様に、本実施形態の方法（ｉ）では、２秒以下の誤差を有する楽曲が３６曲あるのに対し、ビートトラッキングのみの手法（ｉｖ）では１２曲であった。このように、より少ない誤差で楽譜位置を推定できる点から、本実施形態の方法は、ビートトラッキング手法より優れている。これは音楽に合わせて自然な歌声を発生させることに関し不可欠なものである。

【０１０３】

本実施形態の方法による分類において、クリーン信号と残響あり信号との間に大差ないが、本実施形態の方法は、図１５のように、残響あり信号においてより多くの誤差を有する。したがって、実験室の残響が多くの誤差を含む曲に主に影響している。残響は少ない誤差を含む曲にはあまり影響しない。音楽ホール内のような、より長い残響を有する環境においては、楽譜同期の精度に悪影響を与えることも考えられる。

このため、本実施形態では、楽譜位置を推定するために、音響信号分離部１１０が独立成分分析を行って残響音の抑圧を行った後の音響信号を用いているので、この場合においても、残響の影響を軽減して、精度の高い楽譜同期を行うことができる。

【０１０４】

このため、ドラム音ありとドラム音なしの楽曲の誤差を比較することにより、本実施形態の方法の精度が、楽曲中でドラムが演奏されるか否かに依存することを評価した。ドラム音ありとドラム音なしの楽曲数は各８９と１１である。ドラムあり楽曲の累積絶対値誤差平均値は７．３７[秒]であり標準偏差は９．４[秒]である。一方、ドラムなし楽曲の平均累積誤差は２２．１[秒]であり標準偏差は１４．５[秒]である。ビートトラッキングによるテンポ推定は、ドラム音がない時、非常に大きな変動を生じやすい。これは、高い累積誤差を引き起こす、不正確なマッチングの原因となる。

本実施形態では、ドラムなどの低音領域の影響を軽減するため、図１０で説明したように、高周波成分に重み付けを行い、重み付けしたパワーからオンセット時刻を検出するようにしたので、より精度の高いマッチングを行うことができる。

【０１０５】

本実施形態では、楽譜位置推定装置１００をロボット１に適用し、ロボット１が演奏に合わせて歌う（歌声を、スピーカ２０を介して出力する）例について説明したが、推定した楽譜位置情報に基づき、さらにロボット１が演奏に合わせて自身の可動部を動かし、あたかもロボット１が演奏に合わせて、リズムに合わせて体を動かしているようにロボット１が備える制御部により制御するようにしてもよい。

【０１０６】

また、本実施形態では、楽譜位置推定装置１００をロボット１に適用する例を説明したが、他の装置に適用してもよく、例えば携帯電話等に適用するようにしてもよく、あるい

は演奏に合わせて歌う歌唱装置に適用してもよい。

【０１０７】

また、本実施形態では、マッチング部４４０において、レアネスを用いて重み付けを行う例を説明したが、重み付けは、他の要素により行うようにしてもよい。また、音符の出現頻度が低いと判定された場合においても、特定の前後するフレームにおいて、出現頻度の低い音符と判定された音符の出現頻度が高い場合などは、出現頻度が高い音符や、出現頻度が平均的なものを用いてもよい。

【０１０８】

また、本実施形態では、ビート間隔(テンポ)算出４３０で、楽譜が４８分音符に対応した長さのフレームに分割する例を説明したが、他の分割値でも良い。また、バッファを１秒間行う例を説明したが、バッファする時間は１秒でなくてもよく、処理に用いる時間以上分のデータを含むようにしてもよい。

【０１０９】

なお、実施形態の図２と図７の各部の機能を実現するための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、実行することにより各部の処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、ＯＳや周辺機器等のハードウェアを含むものとする。

また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

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、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＣＤ－ＲＯＭ等の可搬媒体、ＵＳＢ（Ｕｎｉｖｅｒｓａｌ　Ｓｅｒｉａｌ　Ｂｕｓ）　Ｉ／Ｆ（インタフェース）を介して接続されるＵＳＢメモリー、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ーのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、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で実現できるものであっても良い。

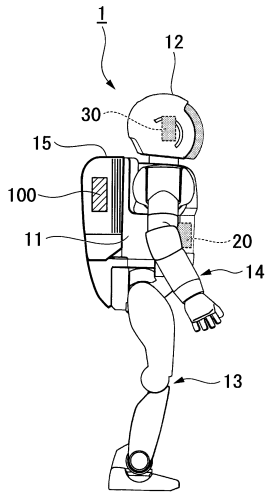
【符号の説明】

【０１１０】

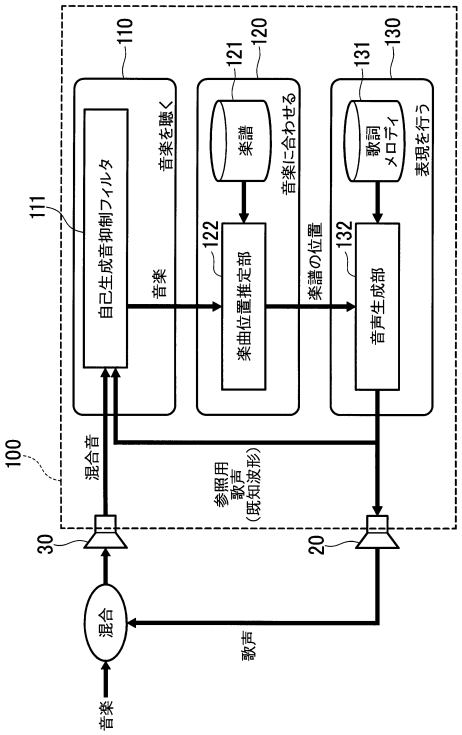
１・・・ロボット
１１・・・基体部
１２・・・頭部（可動部）
１３・・・脚部（可動部）
１４・・・腕部（可動部）
１５・・・収納部
２０・・・スピーカ
３０・・・マイクロホン
１００・・・楽譜位置推定装置
１１０・・・音響信号分離部
１１１・・・自己生成音抑制フィルタ部
１２０・・・楽譜位置推定部（楽譜情報取得部、音響信号の特徴量抽出部、楽譜情報の特徴量抽出部、ビート位置推定部、マッチング部）
１２１・・・楽譜データベース
１２２・・・楽曲位置推定部（音響信号の特徴量抽出部、楽譜情報の特徴量抽出部、ビート位置推定部、マッチング部）
１３０・・・歌声生成部

１３１・・・歌詞とメロディーのデータベース
１３２・・・音声生成部
４１０・・・音響信号からの特徴量抽出部（音響信号の特徴量抽出部）
４２０・・・楽譜からの特徴量抽出（楽譜情報の特徴量抽出部）
４３０・・・ビート間隔(テンポ)算出
４４０・・・マッチング部
４４１・・・類似度計算部
４４２・・・重み付け計算部
４５０・・・テンポ推定部（ビート位置推定部）
４５１・・・小さな観測誤差モデル
４５２・・・大きな観測誤差モデル

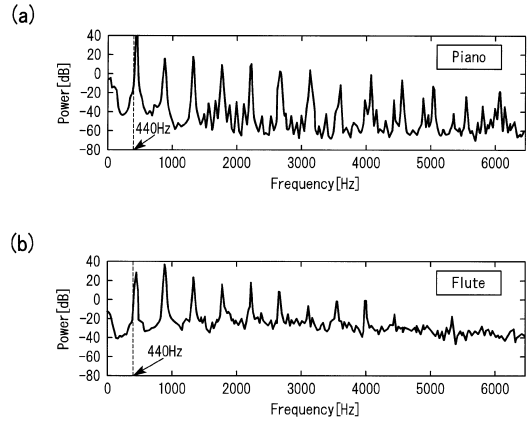
【図１】



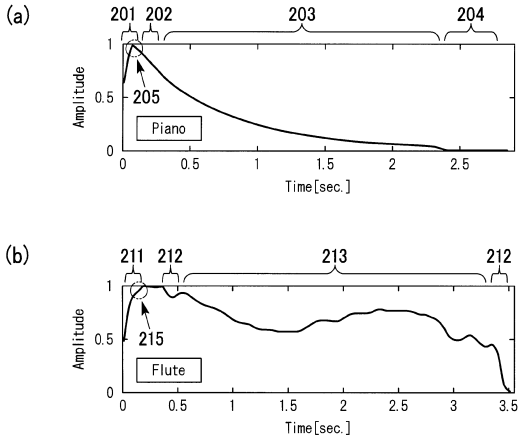
【図２】



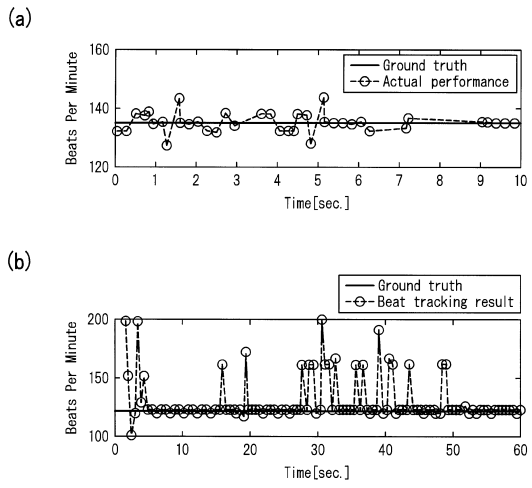
【図３】


(a)
Power[dB]
Piano
440Hz
Frequency[Hz]
(b)
Power[dB]
Flute
440Hz
Frequency[Hz]

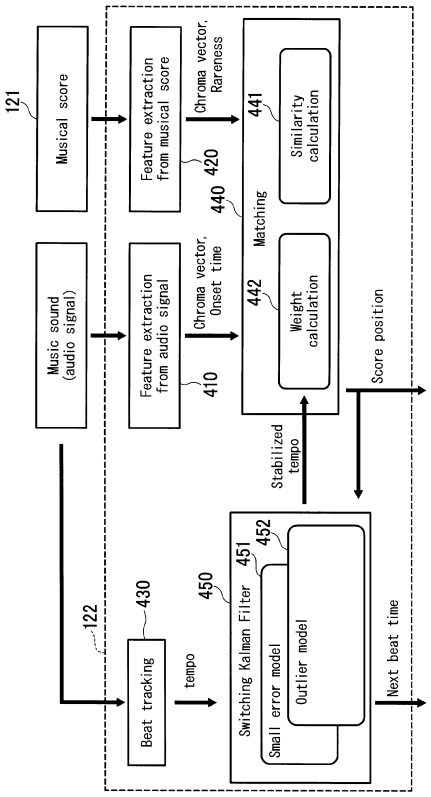
【図４】


(a)
201 202 203 204
Amplitude
205
Piano
Time[sec.]
(b)
211 212 213 212
Amplitude
215
Flute
Time[sec.]

【図６】


(a)
Beats Per Minute
Ground truth
Actual performance
Time[sec.]
(b)
Beats Per Minute
Ground truth
Beat tracking result
Time[sec.]

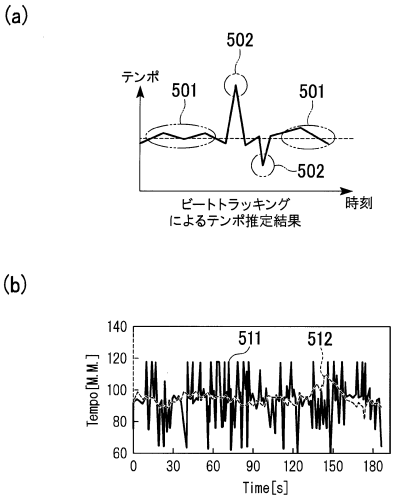
【図７】


121
Musical score
Feature extraction from musical score
Chroma vector, Rareness
420
441
Similarity calculation
440
Matching
Music sound (audio signal)
Feature extraction from audio signal
Chroma vector, Onset time
410
442
Weight calculation
Score position
Stabilized tempo
122
430
Beat tracking
tempo
450
Switching Kalman Filter
451
452
Small error model
Outlier model
Next beat time

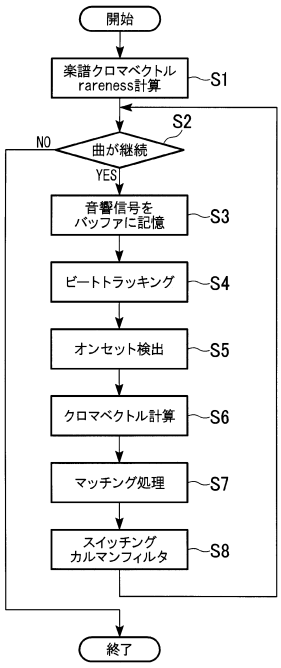
【図８】

Symbols	Definitions
i	index for 12 pitch names(C,C#,･･･,B)
t	time frame of audio signal
n	index for onsets in audio signal
t_n	n-th onset time in audio signal
f	frame index of musical score
m	index for onsets in musical score
f_m	m-th onset time frame in musical score

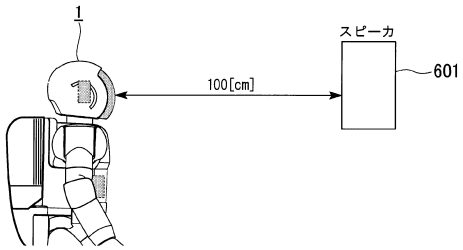
【図１２】



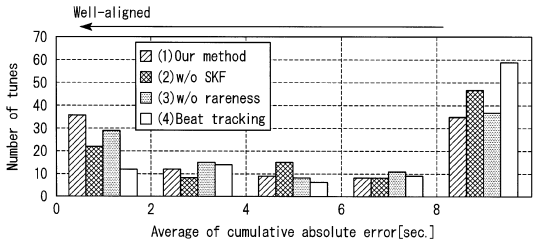
【図１３】



【図１４】



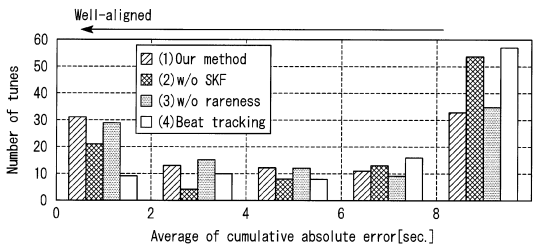
【図１７】



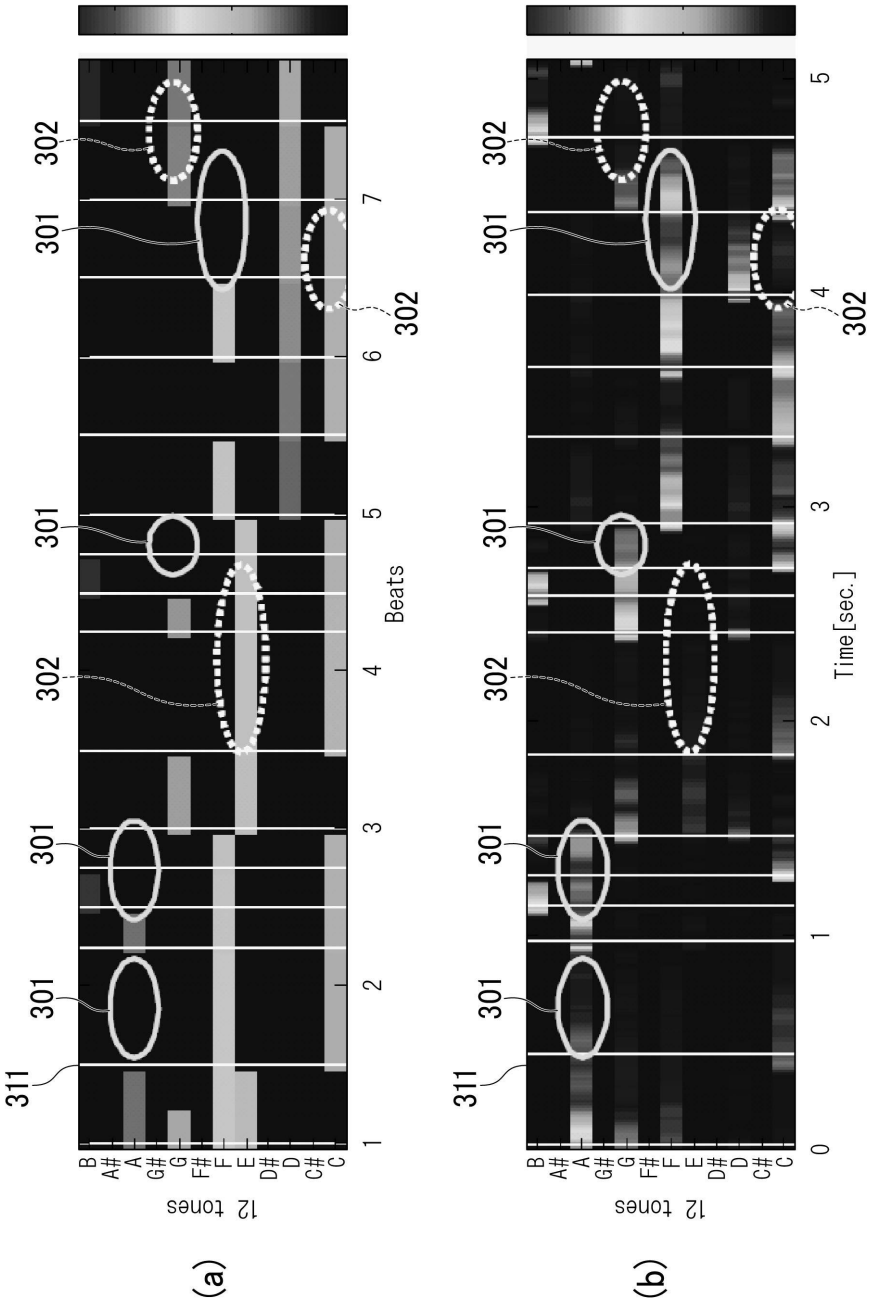
【図１５】

	Clean signal		Reverberated signal	
	Ave.	Std. dev.	Ave.	Std. dev.
(1)Our method	8.9	11.0	9.9	12.6
(2)w/o SKF	11.6	12.8	10.5	11.2
(3)w/o rareness	9.7	13.5	10.3	13.3
(4)Beat tracking	12.5	9.7	11.5	9.1

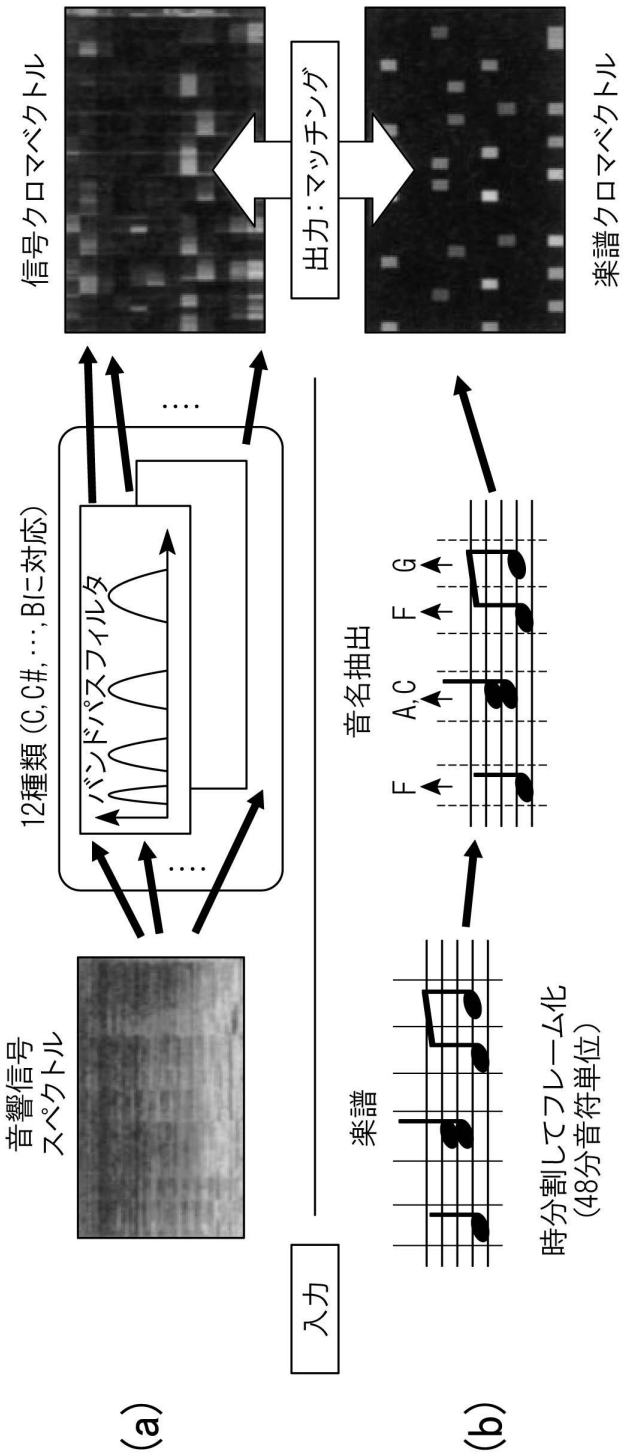
【図１６】



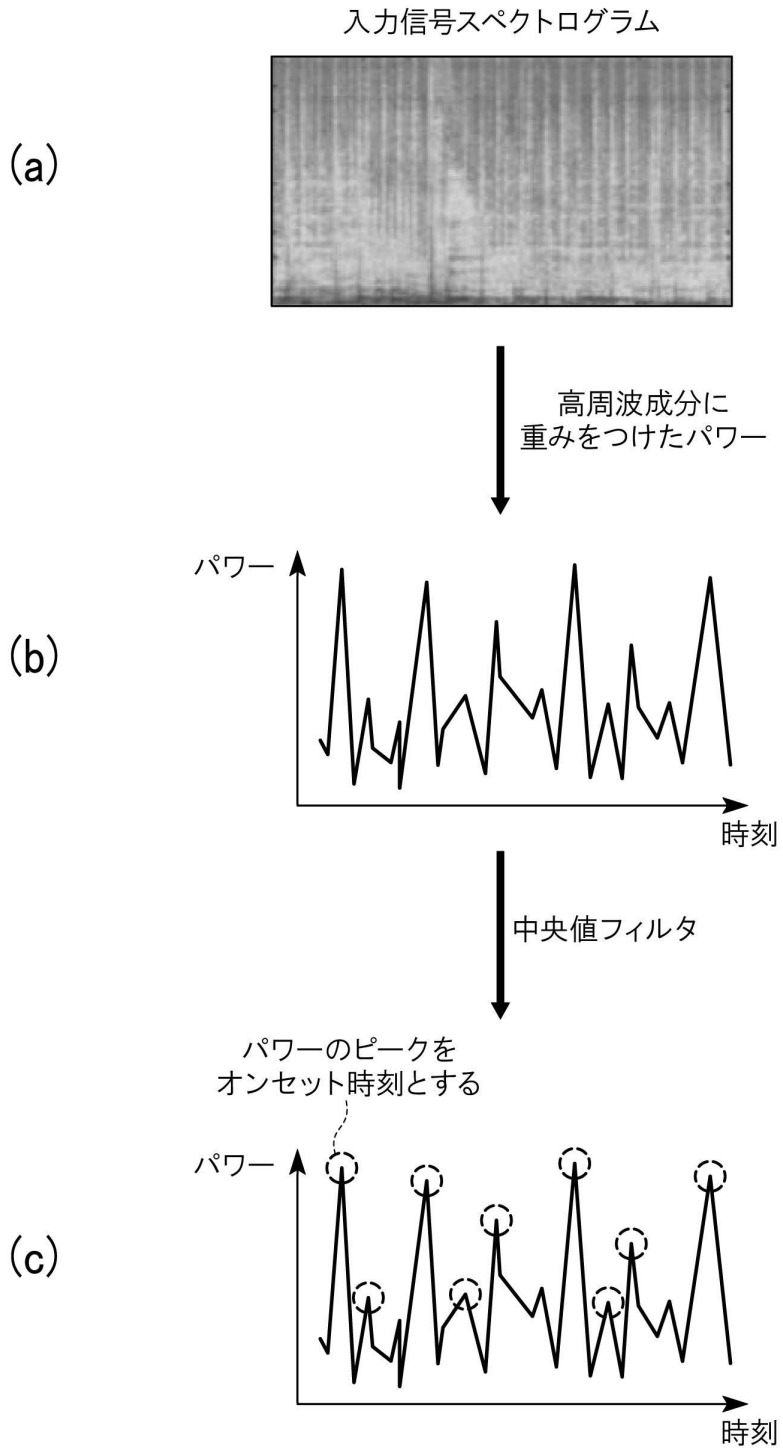
【図５】



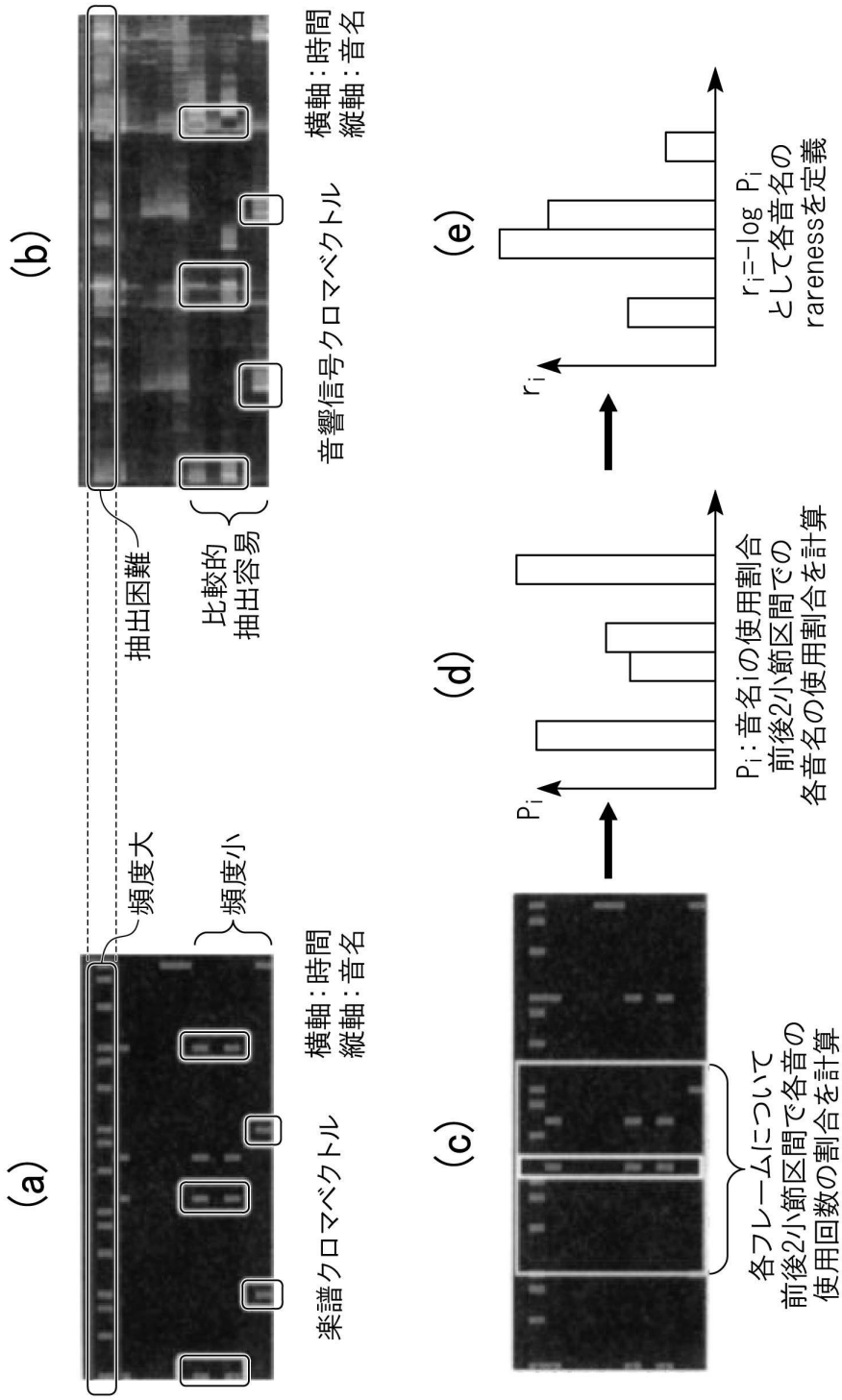
【図９】


信号クロマベクトル
出力：マッチング
楽譜クロマベクトル
12種類 (C, C#, …, Bに対応)
バンドパスフィルタ
音響信号
スペクトル
音名抽出
F A,C F G
楽譜
入力
時分割してフレーム化
(48分音符単位)
(a)
(b)

【図１０】


入力信号スペクトログラム
(a)
高周波成分に
重みをつけたパワー
パワー
(b)
時刻
中央値フィルタ
パワーのピークを
オンセット時刻とする
パワー
(c)
時刻

【図１１】


(a)
楽譜クロマベクトル
横軸：時間
縦軸：音名
頻度大
頻度小
(b)
音響信号クロマベクトル
横軸：時間
縦軸：音名
抽出困難
比較的
抽出容易
(c)
各フレームについて
前後2小節区間で各音の
使用回数の割合を計算
(d)
Pi
Pi：音名iの使用割合
前後2小節区間での
各音名の使用割合を計算
(e)
ri
ri=−log Pi
として各音名の
rarenessを定義

フロントページの続き

(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・インスティチュート・ジャパン内
(72)発明者　大塚　琢馬
京都府京都市左京区吉田本町　国立大学法人京都大学　大学院情報学研究科内
(72)発明者　奥乃　博
京都府京都市左京区吉田本町　国立大学法人京都大学　大学院情報学研究科内

審査官　上田　雄

(56)参考文献　米国特許第０６１０７５５９（ＵＳ，Ａ）
大塚琢馬、村田和馬、武田龍、中臺一博、高橋徹、尾形哲也、奥乃博，歌唱ロボットのためのビート情報とメロディ・ハーモニー情報の統合による音楽音響信号と楽譜の実時間同期手法の開発，情報処理学会第７１回全国大会講演論文集，日本，社団法人情報処理学会，２００９年　3月１０日，pp. 2-243 - 2-244
Juan Pablo Bello, Guiliano Monti and Mark Sandler, Techniques for automatic music transcription, Proc. ISMIR 2000, 米国，２０００年１０月２３日
A. T. Cemgil, B. Kappen, P. Desain and H. Honing, On tempo tracking: Tempogram Representation and Kalman filtering, Journal of New Music Research, ２００１年１２月, Volume 28, Issue 4 , pp. 259-273

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｇ　　１／００－３／０４
Ｇ１０Ｌ　　２５／５１