4 音楽ロボットのための実時間音楽情報処理

奥乃 博[1]　中臺 一博[2]　大塚 琢馬[1]

1 京都大学大学院情報学研究科知能情報学専攻
2 (株)ホンダ・リサーチ・インスティチュート・ジャパン

音楽のリズムに合わせて振る舞う音楽ロボットを目標に据えると，音楽情報処理の課題が見えてくる．

音楽ロボットが要求する機能

音楽は，人と人とをつなぎ，ともに楽しむための重要なメディアである．最近では，受動的に音楽鑑賞をするだけでなく，音楽ソースは所与であっても，それを加工して，自分なりの楽しみ方をするという能動的音楽鑑賞も注目を浴びている．たとえば，バーチャルシンガー初音ミクの歌声に合わせて動画を作成したり，逆に動画に合わせてを初音ミクを歌わせるといった作品が，多数ニコニコ動画サイトなどにアップロードされており，その品質はプロのコンテンツを凌駕するものも少なくない．楽器の合奏を行うのは，より積極的な音楽鑑賞である．

合奏，セッション，ダンスの相手やボーカル役にロボットを使うのは，ロボットと人との共生を追求する上で重要な課題である．このようなパートナーとしての音楽ロボットには，次のような機能が要求される．

(1) 実時間で音楽認識機能，
(2) 音楽表現生成機能，
(3) ロボット自身の耳で演奏される音を聞く機能，
(4) ロボットの演奏音，発声音をモニタ・抑制する機能．

水本らは，これらの要求条件に基づいて，図-1 に示す音楽ロボットの一般的なアーキテクチャを設計している[1]．本アーキテクチャは，音楽表現モジュールと音楽認識モジュールから構成される．音楽表現モジュールは，音楽表現のプラニングを担当する Conceptualizer と，音楽表現の生成を行い，具体的な動作命令生成を担当する Formulator から構成される．一方，音楽認識モジュールは，音源分離と音楽認識器から構成される．

ロボットに聴覚機能を装備する上で，音の取得法は重要である．歌唱ロボットやダンスロボットでは，自分の耳で聞くのではなく，他者の演奏をファイル経由やネットワーク経由で「聞く」という設定，さらには，音楽の開始時間とロボットの演奏や挙動の開始時刻だけを同期させ，音楽そのものは聞かないという設定も，さらには，MIDI 信号を使う設定も，よくとられる．これを見て，ロボットも音楽を聞いて，挙動ができるのだと誤解されることが少なからずある．エンタテインメントとしては大成功であるが，人と共生する音楽ロボットの研究

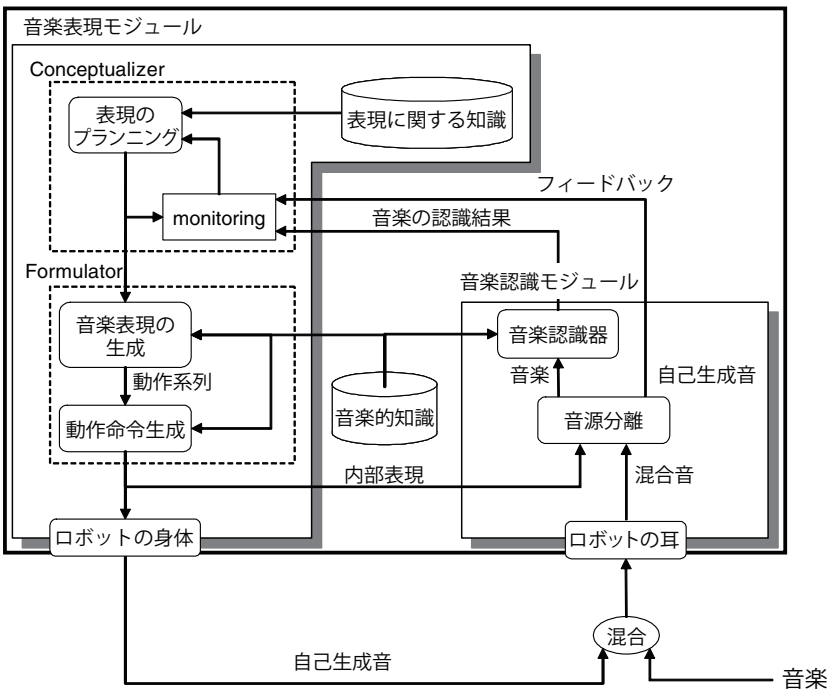


図-1　音楽ロボットの一般的なアーキテクチャ[1]

開発を進める上では，この誤解が大きな支障となる:「そんな機能は，もうXXXのロボットでできているじゃないの」と．

ロボットが自分自身の耳で音楽を聞くと，対象とする音楽だけでなく，反射や残響といった部屋の伝達関数の影響に加えて，自分自身の発声や演奏あるいは雑音が混入することになる．残響処理は，音響処理での重要な研究課題として認識されている．一方，自己生成音は，音声認識ではダブルトークあるいはバージイン発話と呼ばれており，その処理の中心は相手の発話を検知したら，自発話を止めるというレベルでしか扱われていない．混合音中から対象となる音楽を抽出するという音楽ロボットでは不可欠な機能の研究は，緒についたばかりである．

自己発声音を抑制し，相手の音を聞く

人は，発話をするときに，自分の発話をモニタしているとされるが，このモニタは高次の言語理解を担う聴覚連合野で行われることを意味するわけではない．最近の脳科学の知見によると，内耳で電気信号に変換された音響信号が後頭部の一次聴覚野に届いた後，そこで処理が終わり，聴覚連合野に届けられない場合がある．つまり，自分の声を，(1) 低次の信号処理と，(2) 高次の意味の理解，の2段階で処理していると考えられる．図-1のアーキテクチャでは，前者は音源分離で，後者はConceptualizerのモニタ機能で処理するように設計されている．

前者の自己発声音抑制機能として，武田らは，自分の声が分かっているという条件下で，バージイン発話中から相手の発話だけを分離するセミブラインドセパレーション (SB-ICA)[2] に取り組んでいる（図-2参照）．ロボット自身の歌声と，相手の出す歌声，話声，楽器音が，それぞれ部屋の残響特性を表現する伝達関数がかかって，ロボットの耳に混入される．SB-ICAでは，残響音を仮想的に別音源と見なした多入力に対して，独立成分分析 (ICA) を適応し，ほぼ実時間で，音源分離と残響除去とを実現している．得られる出力は，残響を削除した自己生成音と相手の生成音である．本セミブラインドセパレーション機能は，2種類の音楽ロボット[1), 6)]に応用され，まず自己発声音が抑制され，相手の生成した音楽音だけが，音楽認識機能であるビートトラッキングに渡される．

実時間でビートを認識し，追跡する

音楽ロボットだけでなく，自動楽器伴奏では，演奏された音楽のビートを実時間で追跡する実時間ビートトラッキング機能が不可欠である．ビートトラッキングとは，

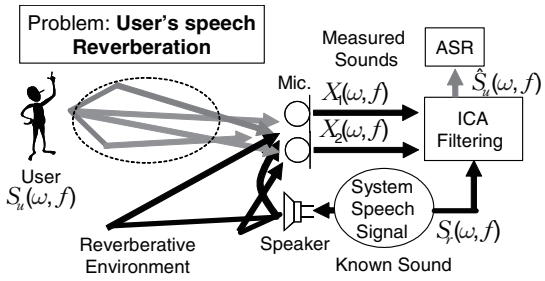


図-2　残響下での音楽ロボットの自己発声音抑制の摸式図[2]

人間でたとえると音楽に合わせて手を叩くときに行うように，音楽から“拍”を抽出する処理である．実時間ビートトラッキングには，楽譜情報を使用する場合と，使用しない場合がある．前者は，楽譜追跡，楽譜照合，楽譜アラインメントなどとも呼ばれており，実時間処理にはオフラインアルゴリズムが使われる[5)]．

1984年頃からDannenbergは，実時間伴奏のためのオンラインアルゴリズムの研究を行っており，独奏者の生演奏から得られた音響信号をクロマベクトルという12音高のクラスで表現し，その表現と楽譜との照合を動的計画法で行い，時間情報を生成し，あらかじめ録音された伴奏をその時間情報に従って実時間で演奏している．彼らの一連の研究開発から，楽譜追跡応用として，The PinanoTutor，Smart Music，Music Plus Oneなどの教育用ソフトウェアが開発されている．

多重奏演奏から，楽譜を使用せずに，ビートを抽出し，追跡する実時間ビートトラッキングの研究は，後藤の研究を嚆矢とする．1993年から始まった研究では，4拍子を前提として，バスドラムやスネアドラムの音の周波数帯域をカバーするバンドパスフィルタから得られたパワー変化や，コード変化度などのボトムアップ情報の自己相関関数から楽曲のテンポを求める．さらに，あらかじめ学習したドラムパターンなどのトップダウン情報とのマッチングを通じてビート位置，あるいは小節構造を分析する．本手法は，スペクトルピークを追跡するエージェント群から構成されるマルチエージェントシステムで実現され，富士通製並列コンピュータAP-1000上に実装されて，実時間処理を達成している[3)]．具体的な応用として開発されたビートに合わせて踊るCGダンサー“Cindy”の応答のよさは，事前に作成されたのかと思わせるほどである．

Scheirerらは，4拍子のポピュラー音楽だけでなく，さまざまなジャンルに対応可能なビートトラッキング手法を開発した．本手法は，音楽音響信号を複数のバンドパスフィルタに通過させ，それぞれから出力されたパワーエンベロープをさまざまなテンポに対応する櫛型共振フィルタにかけることでテンポ推定を行う．テンポ推定

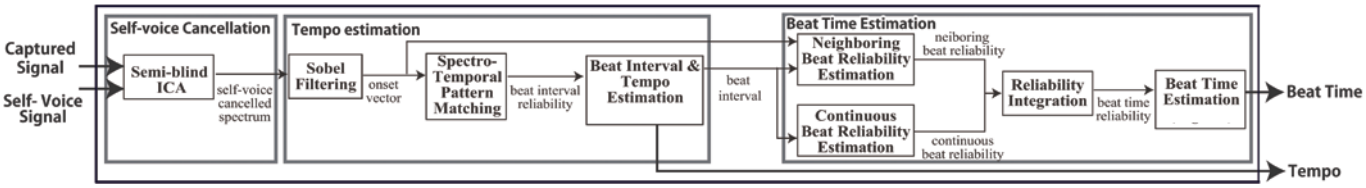


図 -3　時間周波数テンプレートマッチング（STPM）ビートトラッキングアルゴリズム概要 [6)]

には，櫛型フィルタを用いることで，本来の倍や半分のテンポに推定する間違いを抑制し，櫛型フィルタから出力エネルギーをもとに，ビート位置を推定している．評価実験により，ポピュラー音楽だけでなく，クラシック，非西洋音楽などのいずれのジャンルでも正しくビートトラッキング可能であることを示した．ただし，変化に富むドラムパターンや，複雑なリズムの曲に対しては，必ずしも人間の感じるリズムが出力されるわけではない．

Klapuri らは，小節，拍，8 分音符などのより細かいリズム（tatum）間の関係を隠れマルコフモデルを用いて記述し，楽曲の小節構造を含めたビートトラッキング手法を開発した．このような確率モデルの導入は，音楽に関する事前知識の利用に通じ，特に小節構造の推定に有効であることが示されている．Davies らは，音の時間変化の自己相関関数を求めることで，高精度なテンポ，ビート位置推定手法を開発した．音の変化の指標として，高調波成分と，短時間フーリエ変換結果の時間変化を利用している．前者はドラム音やピアノの打鍵音などの打撃音には高調波成分に大きなエネルギーが観察できる事実に基づき，後者はバイオリンの音など，エネルギー変化の少ない音がもたらすリズムの抽出を狙っている．評価実験により，ジャズやロック，クラシック音楽などさまざまなジャンルにおいてもテンポ推定性能が高いことが示された．なお，テンポ推定に自己相関関数を用いるこれらの手法では，テンポ変化の追従に 2 秒程度の時間を要し，実際のテンポの倍や半分になるという推定テンポ誤りも生じやすい．

村田らは，時間周波数テンプレートマッチング（STPM）による実時間ビートトラッキング [6)] に取り組んでいる．STPM リアルタイムビートトラッキングアルゴリズムは，図 -3 に示したように，前章の述べた自己発声音抑制に加えて，テンポ推定，ビート時刻推定の計 3 つのモジュールから構成されており，入力信号に対してビート時刻とテンポを出力する．

テンポ推定には，一般的には，時間領域，もしくは各周波数ラインの 1 次元的な自己相関からビート間隔（テンポ）を推定することが多い．この際に安定性を重視して，自己相関関数の窓長を長くとることが多いが，その分，追従性が損なわれる．STPM は，画像処理で用いられる手法を利用して，この制約を緩和している．具体的には，自己発声音抑制後の信号のパワースペクトログラム上で，まず，音声認識や音楽認識で用いられるメルフィルタバンク適用し，周波数の次元数を 64 次元に圧縮する．次に，パワーが急激に上昇している時刻はオンセットである可能性が高く，オンセットとビート時刻やテンポは密接な関係があるという仮定の下，時間方向のエッジ強調と周波数方向の平滑化を同時に行う Sobel フィルタを適用する．さらに，正規化相互相関によるパターンマッチングを行い「ビート間隔」を検出する．ここで，パターンマッチング前に正規化を施すことにより，雑音が平均化され定常雑音がさらに抑制される．また，マッチングは時間周波数領域で 2 次元的に行う．これらにより，雑音に対するテンポ推定のロバスト性を保ったまま，テンポ変化への高速な追従性を確保している．

ビート時刻推定では，あるフレームとそのビート間隔前のフレームがともにビート時刻となる信頼度である「近接ビート信頼度」と各時刻において推定されたビート間隔で，ビートが連続的に存在しているかを示す信頼度である「連続ビート信頼度」の 2 つの尺度を導入し，これらを統合してビート時刻信頼度を算出する．最終的に，ビート時刻信頼度のピーク時刻を利用してビート時刻を推定する．後藤らの実時間ビートトラッカーでは，テンポ推定に 6 ～ 10 秒程度の窓長を持った自己相関関数を用いているのに対し，STPM では，1 秒程度の窓長を実現し，テンポ推定の安定性と追従性のトレードオフを緩和している．

さらに，大塚は，STPM システムを拡張し，ビートトラッキング誤りに対処するために楽譜の利用を検討している．

音高（ピッチ）推定を実時間で行う

実時間ピッチ推定研究のベースラインとなるシステムは，後藤が開発したメロディーとベースの音高推定を実時間で行う PreFEst[4)] である．PreFEst では，まず PreFEst-core で潜在的な基本周波数 F0 の相対的な有意度を MAP 推定で求め，次に，PreFEst-back-end で，時間的連続性をマルチエージェントシステムで追跡することにより，メロディーラインとベースラインで有意な F0 をそれぞれ求めている．

Dannenberg らの実時間伴奏のための実時間楽譜追跡では，音高はクロマベクトルで表現されている．一般的に，音高表現には，信号処理が軽く，12 次元ベクトルとコンパクトな表現であり，意外と高次の音楽表現に適しているという観点から，クロマベクトルが使用されることが多い．

実環境での応用が要求する音楽情報処理

人とロボットとがインタラクションを行うソフトウェアの実装上の最も重要な点は，次の 4 点である：

(1)実時間応答を保障するオンラインアルゴリズム，
(2)空間計算量を定数程度に抑えるアルゴリズム，
(3)使用される音響的環境に依存しにくいアルゴリズム，
(4) モジュール間での音響データを共有できるミドルウェア．

実時間応答を保障するためには，逐次的に処理結果を返すオンラインアルゴリズムを設計しなければならない．もちろん，処理結果を事前に準備しておき，そのデータの参照だけで実際の計算を行わずに済ませる時間計算量を空間計算量で置き換えることも可能である．ただし，そのようなデータを保持しておくとそれだけ空間計算量が増えるため，できるだけ空間計算量を定数程度に抑える工夫が必要となる．

また，ロボットはさまざまな音響的な環境で使用されることになるため，極力事前知識を最小にするアルゴリズムを使用する必要がある．一般に，音響処理等の要素技術では単体での性能を評価することが多いため，音楽ロボットに応用する場合には必ずしもうまく機能するわけではない．特に，応用される環境に対してどの程度の前提を置いているのかは，最終的なシステムの可搬性に大きく影響を及ぼす．

音楽ロボットでは複数の音響処理を行うことが多く，たとえば，入力音響信号を異なるモジュールで参照することが多い．たとえば，筆者らが開発した聖徳太子ロボットで使用しているロボット聴覚ソフトウェア HARK では，音源定位，音源分離，音声認識インタフェースなどの主たるモジュールは，FlowDesigner というミドルウェアを介して，音響データを共有する．ミドルウェアの使用以前には三話者同時発話認識処理に 7.9 秒かかっていたタスクに対して，ミドルウェアを使用することにより，話者の発話終了後 1.9 秒で応答が可能となっている．

音楽ロボットの紹介

音楽音響信号が入力され，それに合わせて動作する音楽ロボットを紹介する．琴坂らは，神経振動子を用いて，ドラム音やメトロノーム音など周期的な音響信号に同調してドラムを叩くロボットを開発した．生成される運動信号を，観測される音響信号に神経振動子モデルを通じて引き込みを起こさせることで，テンポ変化に追従させることに成功している．運動の生成を神経振動子を用いて修正することに主眼が置かれており，入力としてはメトロノーム音など単純な音に限られる．

Weinberg らによって開発された打楽器演奏ロボット Haile は，人間の打楽器演奏に合わせてロボット自身の腕で打楽器を演奏するものである．人間の演奏の知覚にはビートトラッキングモジュールが組み込まれており，Scheirer の手法が用いられている．さらに，演奏の知覚として，ビートトラッキング結果を元にリズム安定性や，演奏パターンの類似性などが評価され，Haile 自身が生成するフレーズの決定に用いられる．

ジョージア工科大学の Hoffmann は，マリンバロボット Shimon を開発し，人間のピアノ演奏に合わせてマリンバを演奏するデモを公開している．技術の詳細は不明である．デモを見る限り，電子ピアノが使用されているので，ピアノ演奏音をマイクロフォン経由で聞いていないのかもしれない．また，マリンバを演奏するためにバチを並行移動させるゴム車輪の音も，マイクロフォン経由の入力では，難しい問題となる．

Yoshii らは，後藤のビートトラッキングアルゴリズムを Honda ASIMO に実装し，ロボット自身に装着されたマイクロフォンから入力された音楽のテンポに合わせた足踏み動作を実現した．実時間ビートトラッキングにより，予測される次のビート時刻に合わせて足を動かすことで，音楽に合わせた動作を実現している．後藤らのビートトラッキングアルゴリズムに起因して，テンポ変化への追従に 10 秒程度の時間を要することや，ロボット自身の動作に伴って生じる動作音がテンポ推定に悪影響を与えるといった問題点が指摘されている．

水本らは，吉井の仕事を発展させ，4 拍子の音楽に対して，小節単位での認識を行い，その結果に基づき，次の拍を予想しながら，「1，2，3，4」という発声を行っている．このシステムでも，セミブラインド分離による自己発声音の抑制を使用している．

村田らが STPM リアルタイムビートトラッキングに基づいて構築した音楽ロボットのアーキテクチャを図 -4 に示す．システムは主に 3 つのサブシステム（リアルタイムビートトラッカー部，音楽認識部，ロボット制御部）とロボットに分けられる．リアルタイムビートト

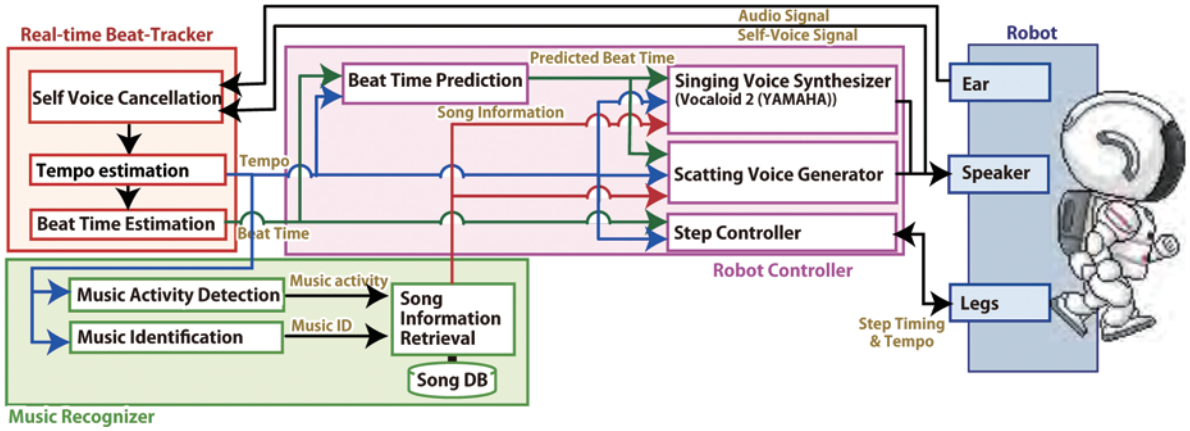
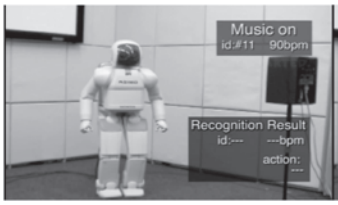
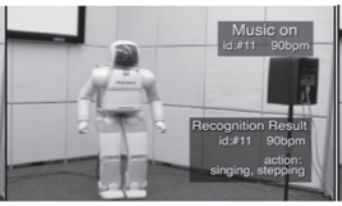


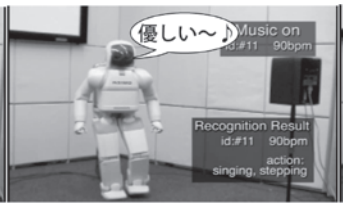
図 -4　音楽を聞き，歌いながらステップをする ASIMO の内部処理の概要 [6)]



(a) Music #11 is played.

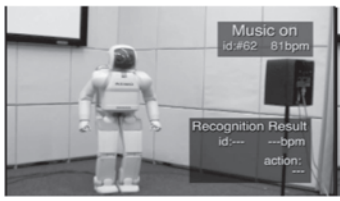


(b) Music is successfully identified.

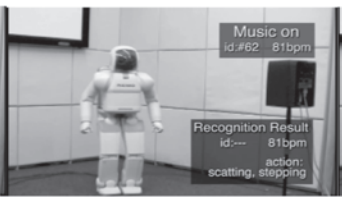


(c) Singing and Stepping started.

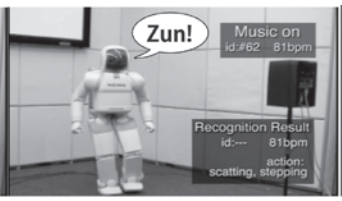
(d) It continues until music stopped.



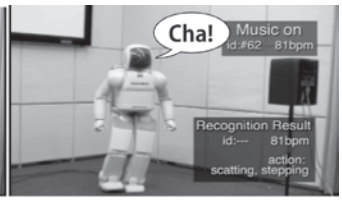
(e) Muic #62 is played.
(#62 is not in the music DB.)



(f) Tempo is successfully estimated, but music identification failed.



(g) Scatting and stepping started.



(h) It continues until music stopped.

図 -5　音楽ロボットの挙動のスナップショット [6)]

ラッカー部には，STPM リアルタイムビートトラッキングを用いている．音楽認識部では，歌の開始・終了のタイミングを検出するため，テンポ情報に基づく音楽区間検出機能，および歌詞，譜面情報を取得するため，テンポ情報に基づく曲名同定と曲名をキーとした曲情報検索機能が実現されている．ロボット制御部ではリアルタイムビートトラッカーによって検出されたビート時刻とビート間隔，音楽認識によって検索された曲情報を用いて，ビート時刻に同期した足踏み，歌唱機能の歌声合成エンジンに VOCALOID2 を用いた歌唱，ビート時刻に合わせた「ずん」「ちゃ」という口ずさみを行う．この際，ビートトラッキング処理の遅れ，ロボット動作の遅れを考慮し実時間動作を可能としている．ロボットは，Honda ASIMO を用いている．音楽信号収音用に，頭部にマイクを 1 本搭載しており，胸部には発声用にスピーカーを内蔵している．ASIMO の動作は，足踏みのみを扱っており，足踏み制御は足踏み間隔のみで行うインタフェースとなっている．また，ハードウェア的な制約から，足踏み間隔は 500 ～ 1,000 [ms] の範囲となっている．これに伴い，リアルタイムビートトラッカーのテンポ推定をこの間隔のテンポ換算である 60 ～ 120 M.M. に制限している．

実際に，構築した音楽ロボットの動作例を図 -5 に示す．正面のスピーカから，音楽(インストルメンタル)が流れ，ロボット搭載マイクの入力からその音楽のビートを検出する．検出したビートから曲名が判定できれば，ビートに合わせて足踏みを行いながら，歌唱を行う(図 -5 (a) ～ (d))．判定できない場合は，口ずさみ音声を出力する（図 -5 (e) ～ (h))．また，テンポ変化への高速追従が可能であることから，テンポが常に変化する人間の拍手のビートトラッキングも可能であり，人・ロボット音楽セッションの実現に向けた検討も行われている．

水本は，電子楽器テルミンの演奏法に取り組み，右手で音高を演奏し，左手で音量を調整するロボットに依存しないモデル化を行っている（図 -6)．このときに必要となる音高や音量調整のパラメータのチューニング作業では，音高を実時間で認識するアルゴリズムを使用している．テルミンは単音演奏楽器であるため，自己相関関数によるピッチ推定が遅延も少なく，精度が高い．

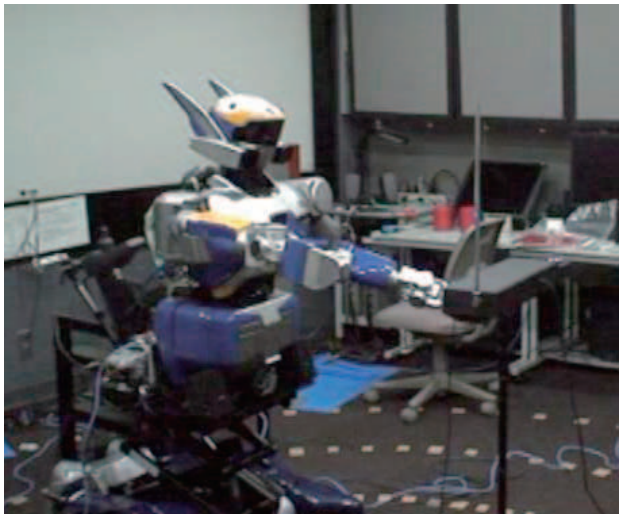


図 -6　テルミンを演奏する HRP-2

音楽ロボットの可能性を追求

計算機パワーの発展は目覚ましく，1980 年代には実時間処理には大型並列コンピュータが必要であったのが，21 世紀には Intel Core 2 Duo 搭載のノート PC でも十分であり，音楽の実時間処理の応用分野が広がっている．1995 年に後藤は「計算機は音楽に合わせて手拍子が打てるか？―リアルタイムビートトラッキングシステム―」を bit 誌に掲載し，並列コンピュータ上でのシステムを報告している．村田は，2008 年に人の手拍子に合わせて足踏みをし，歌を歌うロボットを ASIMO の上にノート PC で実現している．

コンピュータ音楽では，すでにアマチュアの作り込みによるコンテンツの質は，プロが提供するそれを大きく凌駕しているものも少なくない．小型ヒューマノイドロボットでも，アマチュア作品の方が研究開発者の作成した挙動よりも高度な挙動，たとえば，二足歩行や踊り，が可能になっている．実時間音楽情報処理の分野でも，論文にはなりにくい個別技術での作り込みと，小型ヒューマノイドロボットとをアマチュア的発想で組み合わせると，高機能な音楽ロボットの構築が期待できる．アマチュアが可能性を限界まで追求し，そのような経験や個別技術が蓄積されて，新たな実時間音楽情報処理技術が誕生し，それがコンテンツの品質の向上につながることを願っている．

本研究の一部は，科研費（基盤研究 (S)，特定領域研究「情報爆発」），グローバル COE，JST CrestMuse などの支援を受けた．

参考文献

1）Mizumoto, T., et al. : A Robot Listens to Music and Counts Its Beats Aloud by Separating Music from Counting Voice, Proc. of IEEE/RSJ IROS-2008, pp.1538-1548.
2）武田 他：残響下でのバージイン発話認識のための多入力独立成分分析を応用したロボット聴覚，日本ロボット学会誌，Vol.27 (2009) 印刷中．
3）後藤 他：ビートトラッキングシステムの並列計算機への実装―AP1000 によるリアルタイム音楽情報処理―，情報処理学会論文誌，Vol.37, No.7, pp.1460-1468 (1996) .
4）Goto, M. : A Real-time Music-scene-description System : Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, Speech Communication (ISCA Journal), 43:2, pp.311-329 (2004).
5）Dannenberg, R. and Raphael, C. : Music Score Alignment and Computer Accompaniment, CACM, 49:8, pp.38-43 (2006).
6）村田 他：ロボットを対象としたビートトラッキングロボットの提案とその音楽ロボットへの応用，日本ロボット学会誌，Vol.27 (2009) 印刷中．

（平成 21 年 7 月 6 日受付）

奥乃　博（正会員） okuno@kuis.kyoto-u.ac.jp

1972 年東京大学教養学部基礎科学科卒業．博士（工学）．NTT，JST，東京理科大学を経て，2001 年より京都大学大学院情報学研究科知能情報学専攻教授．

中臺 一博 nakadai@jp.honda-ri.com

1993 年東京大学工学部電子電気工学科，1995 年同工学研究系科情報工学専攻修了．博士（工学）．NTT，JST を経て，2003 年より（株）ホンダ・リサーチ・インスティチュート・ジャパン，シニア・リサーチャ．2006 年より東京工業大学情報理工学研究科連携准教授兼務．

大塚 琢馬（学生会員） ohtsuka@kuis.kyoto-u.ac.jp

2009 年京都大学工学部情報学科卒業．現在，同大情報学研究科知能情報学専攻修士課程 1 回生．