

## 音楽共演ロボット：開始・終了キューの画像認識による 人間のフルート奏者との実時間同期

リム アンジェリカ<sup>†1</sup> 水 本 武 志<sup>†1</sup> 大 塚 琢 馬<sup>†1</sup>  
古谷 ルイ 賢造 カイエ<sup>†1</sup> 尾 形 哲 也<sup>†1</sup> 奥 乃 博<sup>†1</sup>

聞く、見るは、共演者が仲間の演奏者とテンポを合わせて演奏するために重要なスキルである。画像キュー(cue, 合図)を検知し、他の演奏者に耳を傾けることによって、演奏者はいっせいに演奏を始め、テンポの緩急の合図に合わせ、さらに、いっせいに演奏を終えることができる。本稿では、人間のフルート奏者がアンサンブルリーダーを担い、ロボットは伴奏者として人間の演奏に追従する問題を扱う。まず、フルート奏者の3種類のジェスチャを提案し、画像キューによる認識、音響ビートと画像キューとの統合によるテンポ推定について述べ、テルミン演奏共演ロボットのジェスチャ認識について報告する。初期実験で3タイプの画像キューが83%以上で検出できること、また、画像キューと音響ビート検出とを組み合わせることにより、テンポ検出が0.5秒以内に行えればビート追跡が安定することが分かった。この結果、フルート奏者の指示に合わせて共演者音楽ロボットがテルミンを演奏し、歌を歌うことが可能となった。

**Musical Robot Co-Player:  
Real-time Synchronization with a Human Flutist  
Recognizing Visual Start and End Cues**

ANGELICA LIM,<sup>†1</sup> TAKESHI MIZUMOTO,<sup>†1</sup>  
TAKUMA OTSUKA,<sup>†1</sup> LOUIS-KENZO FURUYA CAHIER,<sup>†1</sup>  
TETSUYA OGATA<sup>†1</sup> and HIROSHI G. OKUNO<sup>†1</sup>

Listening and watching are important skills for co-players to play in time with fellow musicians. By detecting visual cues and listening to other players, musicians can start together, stop together, and follow a leader's visual cues of changes in tempo. In this paper, we formalize three visual cues for the case of flutists, and describe how our thereminist robot co-player system detects them. Initial experiments show over 83% detection rates for our 3 types of

visual cues. Additionally, by coupling visual cues and acoustic beat detection, the robot can extract a tempo in half a second. The resulting robot co-player can play theremin and sing a song with the lead of a human flutist.

### 1. はじめに

音楽を使ったロボットと人の共生は、音声対話と比べると言語特有の制約が少なく、世代、性別、地域、人種を超えた普遍的なものとなる可能性がある。これまでに開発されてきた音楽ロボットは、演奏を人に聞かせたり、見せたりするいわゆるエンターテインメントロボットであった。ロボットが人と共演する場合も、事前に作りこまれたロボットの演奏に人の奏者が合わせることで実現されていた。人同士が楽器演奏を楽しむように、ロボットが人と共演する機能が提供できれば、共演相手がない、共演相手と時間が合わない、という理由で合奏ができない場合、合奏の練習回数を増やしたいという場合にも、音楽ロボットが対応できるようになる。また、人がロボットとの共演を通じて、複数人とロボットとの共演へと発展させることもできよう。このように音楽ロボットが直接的、間接的に人同士の共生に貢献をすれば、音楽を通じた人とロボットとのインタラクションの有用性・重要性が高まると期待される。

音楽を通じた人とロボットとの合奏や合唱におけるインタラクションを実現するためには、実時間音楽認識機能、音楽表現生成機能、ロボットが自分の耳で聞くときに入って来る自己生成音の抑制機能、自分自身の演奏音・発聲音のモニタ機能、演奏全体のモニタ機能などが必要である<sup>1)</sup>。これらの機能の上に、リズムレベル、メロディーレベル、ハーモニーレベルでのインタラクション、さらには、音響レベルだけでなく、身体レベルでのインタラクションが実現されることになる。本稿では、リズムレベル、および、身体レベルでのインタラクションのうち、最も原初的なリズム同期に焦点を絞って、合奏で自然に行われるジェスチャの認識、および音情報との統合の重要性について議論する。

アンサンブル演奏では、身体レベルでの様々なインタラクションを通じて演奏者間でコ

†1 京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

本稿は、International Conference on Intelligent Robots and Systems, “Robot Musical Accompaniment: Integrating Audio and Visual Cues for Real-time Synchronization with a Human Flutist”, pp.1964–1969, IEEE, RSJ, Taipei, 18–22 Oct. 2010 の拡張版である。

ミュニケーションが行われている。たとえば、アマチュア演奏家同士の演奏であっても、お互いに演奏音を聞き、共演者を見て、お互いの演奏を合わせようとする。Fredrickson<sup>2)</sup>は、バンド奏者は指揮者を見るだけでなく、共演者の演奏も聞いて同期すると実証している。たとえ指揮者がいなくても、奏者は視覚を使ってコミュニケーションできる。ピアノ二重奏の研究<sup>3)</sup>からは、頭の動き、強調された指の持ち上げ、アイコンタクトが演奏者間の同期イベントをやりとりするのに使用されていることが分かっている。

本稿では、楽譜に従ったアンサンブル演奏で、人の奏者がリードする場合に、ロボットが人の指示を読み取って、それに従って演奏するという課題を取り上げる。また、音楽演奏ロボットには水本らが開発した電子楽器テルミンを演奏する HRP-2 上に実装されたシステム<sup>4),5)</sup>を使用する。ソフトウェアによる音楽伴奏システムの歴史は古く、様々な手法が開発されてきている<sup>6)-10)</sup>。楽譜に従ったアンサンブル演奏では、楽譜追跡が重要な機能となるが、人との合奏に必要なオンライン処理を行った場合には追跡エラーが累積し、累積エラーのために演奏が進むにつれて合奏の同期がずれることが避けられないので、何らかのエラー修復機能の必要性が指摘されている<sup>11)</sup>。

人の音楽演奏をビートレベルで追跡するロボットは複数開発されている。たとえば、ジョージア工科大学のドラム演奏ロボット HAILE<sup>12)</sup>は、スペクトルパワーに基づくビート追跡法で人のドラムビートを検出し、検出したビートから、テンポを計算し、それに従って即興演奏を行う。村田ら<sup>13)</sup>は、ビートを追跡し、ポップミュージックを聞き、ステップを踏みながら歌うロボットを開発している。これらのシステムが追跡対象としているのは打楽器音であり、パワーの急激な立ち上がりにより音符オンセットの認識ができるので、ビート追跡が容易である。一方、後述するように、このような方法では、バイオリンやフルートといったパワー変換が必ずしも急激でないような楽器音に対するビート追跡には向いていない。また、ギターでは裏拍が多用されるので、単純なビート追跡や楽譜追跡では、倍速のテンポになったり、半拍ずれることになったりするので、何らかの工夫が必要である<sup>14)</sup>。

音楽演奏での画像キュー（cue, 合図）についても、複数の先行研究が行われている。インタラクティブなマリンバ演奏ロボット Shimon<sup>15)</sup>は、人の共演者を見て、ソロチェンジを指示し、アンサンブルリーダ（以下、単にリーダと呼ぶ）を担う。しかし、逆に入人がリーダを担っても、ロボットは同じ動きを検出することはできない。早稲田大学のフルートロボット<sup>17)</sup>、サキソフォンロボット<sup>18)</sup>は、人のサキソフォン奏者の線形の動きを検出し、音量やビブラートを変えることができる。同様に、マルチモーダル演奏システム<sup>19)</sup>は、フルート奏者を音響と画像を通じて追跡し、フルート奏者がフルートを下方に向け、低音を演奏す

るときには、録音済みの音楽を低音で再生する。後者 2 つのアプローチの問題点は、検出しようとするキューを示す奏者の動きが、フルートが参加する合奏でよく目にするような「フルートの口から遠い端を上下に動かす」という自然なジェスチャとは違っていることである。

画像情報以外の情報を使用する研究も行われている。白鳥ら<sup>20)</sup>は、モーションキャプチャのデータから舞蹈動作のリズムをバッチで取り出し、人の舞蹈動作での基本動作を抽出し、ヒューマノイドロボット HRP-2<sup>4)</sup>での舞蹈動作生成に使用し、留め動作という動きの微小な静止姿勢の検出に応用している。特に、視聴覚情報統合により、リズムに合っていないような位置での留め候補は排除し、留め動作認識の高性能化が図られている。堀内ら<sup>21)</sup>は、フルート奏者のプレス（息継ぎ）による合図（キュー）を検出し、演奏を開始する伴奏システムを開発している。ただし、本システムではフルート用集音マイクをフルートに巧妙に設置しているので、演奏者の胸元に付けたピンマイク、あるいは、ロボットの耳で演奏者の息継ぎの音を安定して集音するような状況への適用は難しい。また、伴奏の演奏開始の精度を向上させるために、システムの内部処理にかかる時間を推定し、その時間分だけを予測されたテンポから引く方法を提案している。さらに、実験では、様々な条件で被験者実験を行っており、有益な知見が得られているものの、被験者はヘッドフォンで伴奏を聞いているので、自分の演奏音と伴奏を同時に聞いた場合の影響などは排除されており、実際の演奏時にどの程度有効であるかは未知である。また、本手法は、あくまで演奏開始時だけが考慮されており、曲の途中でのテンポ変化やフェルマータの終了などへの対応は検討されていない。

本稿では、アンサンブルリーダが音楽演奏において通常使用すると考えられる次に示す 3 つのジェスチャ認識の有効性について検討する：

- (1) 演奏の開始,
- (2) フェルマータの終了,
- (3) テンポの変更.

なお、本稿で想定する「自然な」ジェスチャは、「フルートの口から遠い端を上下に動かす」ものに限定する。

具体的には、人間のフルート奏者をアンサンブルリーダとするときに、ロボット共演者が合奏するために最低限必要な機能として、以下の 3 点に絞り込んだ（図 1）：

- 3 つのジェスチャによる画像キュー検出 ,
- オンセットという音響キューの検出 ,
- テンポ変更のジェスチャと音響キューを統合して人の演奏速度を抽出 .



図 1 テルミン演奏ロボット HRP-2 がアンサンブルリーダ（フルート演奏者）の視聴覚キューに合わせながら電子楽器テルミンを演奏

Fig. 1 Theremin-playing robot, HRP-2, plays theremin by synchronizing with audio and visual cues of a human ensemble leader (flutist).

本稿では、合奏に次の仮定をおく、共演者ロボットのための視聴覚キュー認識手法について述べる。

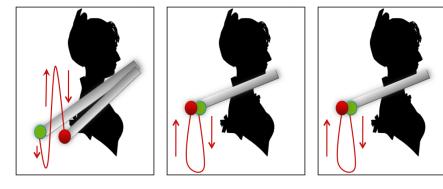
- (A) ロボットは、人間の演奏者の指示によってテンポを変化する。
- (B) ロボットは、人間の指示以外では一定のテンポで演奏する。
- (C) フルート奏者は、フルート遠端の上下動をもってロボットへ演奏の開始、フェルマータの終了、テンポの変更を合図する。

なお、合奏をするための準備情報を極力減らすために、ロボットは人間の演奏するパート譜は持たず、自分が演奏するパート譜しか持っていないものとする。

本稿で採用した視聴覚情報統合は、ギター演奏とロボットとの合奏システムにおけるビート追跡でも採用されており<sup>14),15)</sup>、合奏のための有効なアプローチであるといえる。

## 2. 口ボット共演者

本口ボット共演者は、視聴覚、つまり、画像と音響信号とを相補的に使用して、アンサンブルリーダの指示を認識し、それに合わせて演奏を行う。画像キューを使って演奏を開始し、音響と画像を組み合わせてテンポ変化に適応し、再び画像キューを使ってリーダに合わせて演奏を終了する。以下、まず画像キュー認識アルゴリズムについて説明をし、次に、音響処理でのオンセッタ検出について概略を述べる。最後に、実時間視聴覚情報統合について述べる。以下、音響信号から得たビートキューを音響ビートキュー、画像列から得たキューを画像ビートキューと呼ぶ。



(a) 開始キュー (b) 終了キュー (c) ビートキュー

図 2 フルート用画像キューの軌跡による分類  
Fig. 2 Trajectories of flute visual cues: Start, End and Beat Cues.

### 2.1 画像キューの設計

ジェスチャ認識を設計するにあたっては、どのジェスチャがどのような意図で実行されるのかを調査し、要求される機能を洗い出すことが必要である。フルートを演奏するときの動きを京都大学オーケストラのフルートパートの部員の調査に加えて、我々がこれまで行ってきた Web 上でのアンサンブルでのフルート演奏のビデオを複数調査した結果「フルートの遠端の上下がキューに用いられている」という知見を得た。この調査結果と似たような知見が、フルートと同じ吹奏楽器であるクラリネット奏者のクラリネットの動きについて得られている<sup>23)</sup>。すなわち、クラリネット奏者はクラリネットのベル（開口部）を上下に動かしてリズムをとっているとの知見である。

このようにして、フルート奏者がリーダを担っている場合にとる顕著なジェスチャとして、以下に述べる 3 つのものを同定した（図 2）。

- (1) 開始キュー (*start cue*) は、楽曲や新たな楽章の最初の音符を同期するのに使用される。我々は開始キューを図 2(a) に示したようにフルートの遠端の Down-Up-Down の動きと定義する。他のアンサンブル奏者とのアイコンタクトはあってもよいが、開始キューの前には楽器の動きはないものとする。
- (2) 終了キュー (*end cue*) は、リーダがフェルマータの「打ち切り」に使用する。たとえば、楽譜でのフェルマータでは、リーダは、全演奏者が演奏を終えるべきときを指示しなければならない。フルート演奏者間では、終了キューは、フルートの遠端の円周状の動きが使われる。本稿では、図 2(b) に示したように単純に正面から見たときの下-上運動で定義する。このジェスチャは、フェルマータ演奏中の動きのない状態に引き続いてとられるものとする。
- (3) ビートキュー (*beat cue*) は、音楽でのビートを指示するのに使用され、これからテ

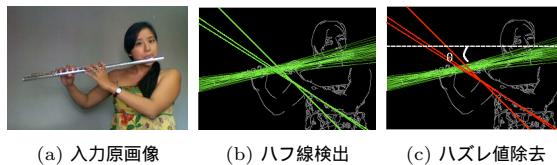


図 3 フルートの直線抽出による検出。ハズレ値は赤で、追跡されたフルートの角度は白で表示。簡単のため無背景を使用

Fig. 3 Detection of a flute by extracting lines. (a) Original input image, (b) processed image with detected Hough lines and (c) outliers marked in red, with the flute angle to track in white.

ンポ推定を行う。リーダがビートジェスチャをはやく行うと、奏者はみんなスピードアップする必要がある。我々は図 2(c)に示したように、ビートキューをフルートの下-上動作として定義する。終了キューと異なり、ビートキューの前は静止してはいけない。

これら 3 つのジェスチャはあくまで調査した範囲内でのフルート奏者から観測された結果であり、すべてのフルート奏者が演奏時にそのようなジェスチャをしているわけではない。多くのフルート奏者に共通する音楽演奏シンボルとしてのジェスチャを同定することは、音楽ロボットにおけるインタラクションの高度化を実現するうえでの画像処理の可能性を検討する出発点になると考えられる。

## 2.2 画像キュー検出法

これら 3 つのジェスチャを検出するためには、フルートそのものを追跡するのが、演奏者の個人性（服装、体型など）に依存する度合いが低いので、自然と考えられる。本稿では、図 3(a)に示したように、フルート奏者はロボット内蔵カメラの正面を向いているものとする。演奏者自身だけで共演を楽しむという観点からは演奏者同士が正対するのは妥当な仮定であると考える。

フルートの位置検出には、まずハフ（Hough）変換による直線検出法を使用し、ビデオ画像系列全体から直線を検出する（図 3(b)）。クラシック用フルートは、複数の運指用キー（ボタン）がついた棒として特徴づけられるので、ハフ線検出結果として得られる複数の直線群は、フルートの傾きをほぼ近似していると考えられる。さらに、背景や衣類から検出される余分な線はフルートに対するハズレ値であるので、RAMSAC ハズレ値検出アルゴリズム<sup>24)</sup>を使用して、これらの不必要的線を除去する。その結果を図 3(c)に示す。残った線の集合に対して平均角度  $\theta$  を求め、これをフルートの傾きとする。この角度検出は各時間で行

い、フルートがどちらの方向へ動いているかを決定する。形式的には、時刻  $t - 1$  時間でのビデオフレーム  $F_{t-1}$  と、時刻  $t$  時点でのフレーム  $F_t$  の間に生じたフルートの角度  $\theta$  の瞬時変化を求ることである。

$$\Delta\theta = \theta(F_t) - \theta(F_{t-1}) \quad (1)$$

この  $\Delta\theta$  より、フルートが現在、Down, Up, Still のいずれの状態にあるのかを決定する。

$$STATE(\Delta\theta) = \begin{cases} \text{Down} & \text{if } \Delta\theta < -\text{threshold} \\ \text{Up} & \text{if } \Delta\theta > \text{threshold} \\ \text{Still} & \text{otherwise} \end{cases} \quad (2)$$

ここで閾値 threshold は、小さな揺らぎを無視し、フルート角度の変化が十分に大きいことを保証するために使用する。3 つのキュー各々に対応する FSM ( Finite State Machine ) (図 4) を並列に使用し、フルートの現在の状態を追跡する。前述のように、ビートキューと終了キューは、その前の状態が Still かどうかの違いがあるので、FSM で Still 状態で一定時間とどまるように制約を追加し、ビートキューの可能性があるかどうかを判定している。

ロボットがキューを検出するごとに、システムは文脈に依存して得られた情報を使用するかどうかを決定する。特に、現在の楽譜位置に基づいてキューをフィルタしている。開始キューは、楽曲の最初にロボット共演者を制御するのに使用されるだけであり、終了キューは、音符が停止状態にあるときにだけ使用される。ビートキューは、音響キューが存在しない開始・終了時と異なり、音響キューが利用できるので、音響キューと統合することで時間解像度を向上させることが可能となる。

## 2.3 音響処理モジュール

前述のとおり、テンポ変化はビート間隔の伸縮に関連している。これらのビートを検出す方法の 1 つは、演奏された音符の開始点、すなわち、音符オンセットを検出することである。ビートは音符オンセットと一致することが多いので、まず、音符オンセット検出を実行し、ビート候補の集合を抽出し、次に画像キューとの統合を行う。

音符オンセット検出法に対する要求条件は、

- (1) 実時間演奏を可能にする高速処理、
  - (2) パワー変化が緩やかなオンセット（ソフトオンセット、soft onset）の検出、
- である。前述のとおり、打楽器音が含まれる場合にはパワーの急激な立ち上がりを手がかりに音符オンセットを求めることができる。それに対して、バイオリンやフルートなどのではレガート音符が頻出するので、パワー変化の穏やかなソフトオンセットの検出機能は不可

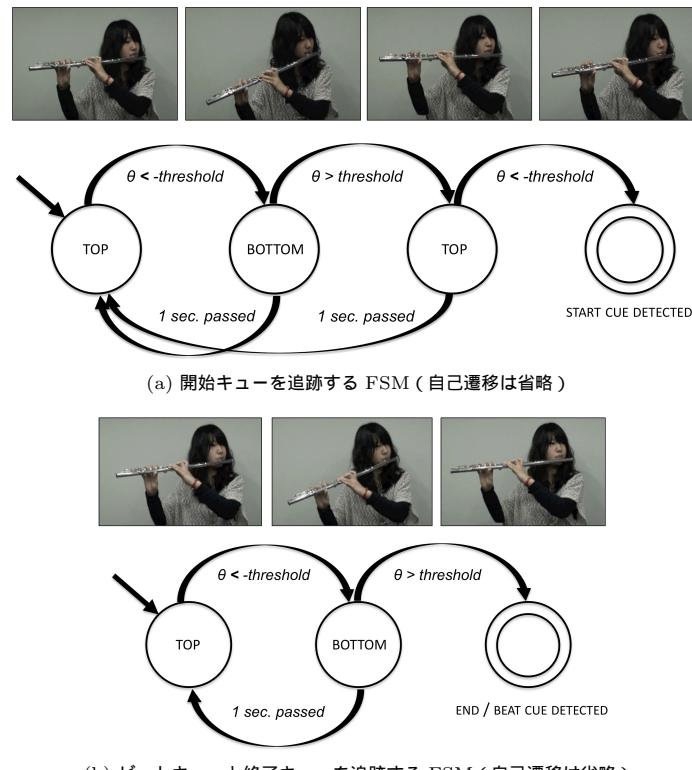


図 4 3つの画像キューを検出する FSM (Finite State Machine)  
Fig. 4 Finite state machines to detect three cues.

欠である。本稿では、Complex Domain Difference (CDD) 法<sup>25)</sup> という音符オンセット検出法を使用した。CDD 法は、スペクトルパワーと複素数領域での位相の双方の差を探し、パワー変化によりアタック音符を、位相摂動によりリガート音符を検出する。具体的には Aubio onset detection library<sup>26)</sup> を利用した。これは C 言語で実装されており、(1) の要求条件も満足する。

#### 2.4 情報統合モジュール

フルート奏者がテンポ変化を指示するには 2 つの方法が考えられる：(1) ジェスチャだけ

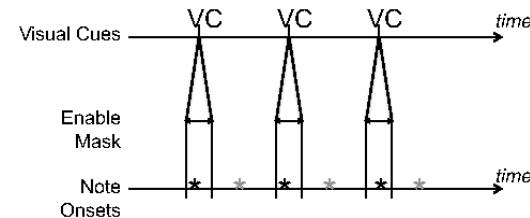


図 5 音響ビートキューと画像ビートキューとの照合の模式図。画像キューが enable mask として働き、そのキューの  $\pm \delta_1$  以内に発生した音符オンセットがマッチしたビートとして扱われる

Fig. 5 Our audio-visual matching scheme. Visual cues act as an enabler; detected note onsets which fall into a pre-specified range around visual cues are considered as matched beats.

を使用する場合、(2) フルート演奏とジェスチャの両方を使用する場合。後者に対しては、情報統合モジュールが、画像ビートキューと音響ビートキューとが同時にマッチするものを探して、テンポ変化を検出する。双方のモダリティから同時にビートキューが検出されると、情報統合モジュールは、アンサンブルリーダがビートを指示しようとしている（したがって、テンポを変えたいと思っている）可能性が非常に高いと判断する。

図 5 に示したように、画像キューは時間解像度が悪いので、情報統合のための enable mask として機能する。画像キューの  $\delta_1 = 150$  ms 前後、すなわち、300 ms を enable mask を用いて、音符オンセットの照合を試みる。この 150 ms は実験的に定めたものである。もし、300 ms の窓内に複数の音符オンセットが含まれる場合には、最初の音符オンセットをマッチしたキューとして選択する。いったん 2 つのビートがマッチすることが検出されると、直前に検出されていたビートとの差、すなわち、オンセット間隔 (Inter-Onset-Interval, IOI) を計算し、瞬時的なテンポが得られる。

ここで、リーダは、楽譜で指示されていない限り、突如大きなテンポ変化、たとえば、倍速に演奏を速めるといった指示はしないものと仮定する。したがって、テンポ変化が所与の閾値  $\delta_2$  以下であれば、このテンポ変化を受け入れ、新しいテンポがロボットの演奏モジュールへ送られる。なお、 $\delta_2$  は実験的に 100 ms と定めた。

テンポ (IOI) 検出のための情報統合アルゴリズムを図 6 に示す。まず、検出された画像キューの時間順整列リスト、音響キューの時間順整列リストをそれぞれ  $V$ ,  $A$  とする。今、時刻  $t_v$  の画像キューを  $V$  のリストに、時刻  $t_a$  の音響キューを  $A$  のリストに追加しようとしているものとする。マッチしたビート時刻の時間順整列リストを  $M$ 、その個数を  $|S|$  とする。また、 $V$  と  $A$  でのマッチしたイベント間の最大差を  $\delta_1$ 、現在のテンポ IOI を  $IOI_c$

とし，テンポ変化閾値を  $\delta_2$  とする。

与えられた画像キューと音響キューの間隔が  $\delta_1$  未満であれば，その音響キュー時刻を  $M$  に追加する。最近マッチした 2 つのキューの差が  $IOI_c$  から  $\delta_2$  以上であれば，その差を新しいテンポとして返し，それ以外のときには現在のテンポ  $IOI_c$  を返す。なお，新しいテンポを返すときには，音響イベントからのタイミングだけを使っている。この理由は音響キューの方がサンプリングレートが高いからである。音響信号は 44.1 KHz サンプリングであるのに対して，画像は毎秒たかだか 30 フレーム (fps) と低レートである。

音響キューがない場合は，3 個の連続した画像キューから 2 つの IOI を求め，その差が  $\delta_3$  未満であれば，隣接しているイベントと見なし，1 番目と 3 番目の画像キューの差の半分が  $IOI_c$  から  $\delta_2$  以上であれば，その差の半分を新しいテンポとして返す。音響キューが利用できない最後の場合には，使用するカメラのフレームレートに時間分解能が大きく影響を受ける。たとえば，それぞれ 20 fps, 25 fps のカメラを使う 2 つのシステムでは，5 fps の差があり，これは IOI では約 80 ms に相当する。

通常，複数のモダリティの処理は別々の計算機で行われることが多いので，それらの間の時間的なズレがあると，それがたとえ 100 ms 程度の小さな時間のズレであっても，テンポ推定に大きな影響を及ぼす。本情報統合スキームでは，Network Time Protocol (NTP)<sup>27)</sup> を使用して時間合わせを行っている。具体的には，1 Gbps イーサネットで接続されたすべ

```

if e is audio then
  A ← A + te
  if ∃v ∈ V, |te - tv| < δ1 then
    M ← M + te
    if |M| ≥ 2 and ||M[last] - M[last - 1]| - IOIc| < δ2 then
      return M[last] - M[last - 1]
  if e is video then
    V ← V + te
    if ∃a ∈ A, |te - ta| < δ1 then
      M ← M + min({ta | a ∈ A, |te - ta| < δ1})
      if |M| ≥ 2 and ||M[last] - M[last - 1]| - IOIc| < δ2 then
        return M[last] - M[last - 1]
    if |V| ≥ 3 and (V[last] - V[last - 1]) - (V[last - 1] - V[last - 2]) < δ3 then
      if (V[last] - V[last - 2])/2 - IOIc) < δ2 then
        return (V[last] - V[last - 2])/2

```

図 6 IOI (Inter-Onset-Interval) 検出アルゴリズム

Fig. 6 IOI (Inter-Onset-Interval) detection algorithm.

てのモジュールの時計を NTP によりミリセカンド単位で同期させている。

### 3. ロボット共演者システムの概要

図 7 にロボット共演者システムの概要を示す。人間のフルート奏者がアンサンブルリー<sup>ダ</sup>を担い，ロボットが電子楽器テルミンを演奏し，歌を歌う。本システムで使用したロボット共演者は，水本らが開発した汎用のテルミン演奏システム<sup>5)</sup>を川田工業製のヒューマノイドロボット HRP-2 上に移植したものである。汎用のテルミン演奏システムは，ロボットハードウェアに非依存部分と依存部分に分けて設計されている。電子楽器テルミンは，音高用のアンテナと音量用のアンテナを持ち，演奏者は手の位置とアンテナとの距離を変化させて，音高と音量を制御する。同演奏システムは，ロボットの腕の位置と音高と音量のモデルを有しており，キャリブレーションにより 2 つのモデルを獲得する。演奏は，まず楽譜が与えられると 2 つのモデルを使用して腕の運動計画を立案する。ここまで処理はロボットハードウェアに依存しない。次に，得られた腕の運動計画をロボットハードウェアに依存した処理を行い，ハードウェアコマンドに落とし込み，ロボットで実際に腕や体を動かし，テルミンを演奏する。今までに，HRP-2 だけでなく，川田工業製の上半身ヒューマノイド HIRO，ホンダのヒューマノイド，仏アルデバラン社の小型ヒューマノイド NAO に移植されており，任意の SMF (Standard Midi File) で与えられた楽譜を演奏することができる。

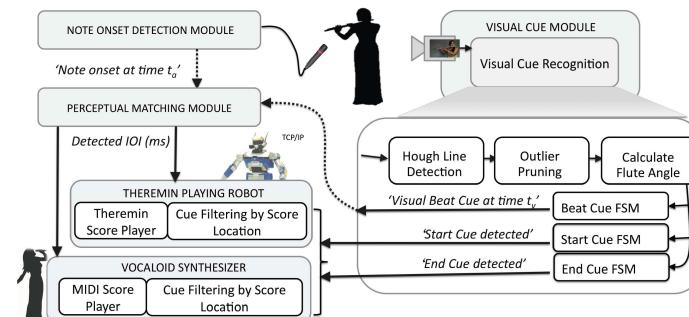


図 7 ロボット共演者システム：画像ジェスチャ認識と音響オンセット認識を統合したテンポ認識で，フルート奏者に合わせて電子楽器テルミンを演奏し，VOCALOID Prima で歌う

Fig. 7 Overview of our robot accompanist system: it detects tempo by integrating visual gesture recognition and audio onset detection, and plays theremin and sings a song by VOCALOID Prima with the lead of a flutist.

歌手部分は、ヤマハが開発した VOCALOID をベースにした商用の歌声作成システム PRIMA がベースとなっている。VOCALOID では初音ミクが有名であるが、PRIMA はその英語版である。歌声は事前に MIDI データとして作成してあるので、MIDI player がリーダのテンポ指示に応じて演奏すること、つまり、フルート演奏者の指示に合わせて歌うことができる。

テルミン演奏と歌手を制御するのがテンポ認識モジュールである。テンポ認識は前述の画像ビートキー検出、音響ビートキー検出、および、情報統合部から構成される。画像ビートキー検出では、HRP-2 の内蔵カメラ Point Grey 社製の Flea に TAMRON 製 219HB の 8 mm レンズを付けて、 $1,024 \times 728$  の解像度のグレースケール画像を最大 30 fps で読み込んでいる。なお、フルートの角度検出に、通常の CCD カメラ以外に距離情報の精度が高い TOF (Time-Of-Flight) カメラ、具体的には Swiss Ranger 社製の SR-4000 の使用も試みたが、フルートが金属材質であるためにうまく距離が取得できず、使用を断念した。

フルート音の取得には市販の廉価なエレクトレットコンデンサマイクロフォン Sony ECM-C10 を使用し、フルート奏者の下襟に装着した。音響処理には 2.13 GHz MacBook を使用した。

開始キーと終了キーが画像処理モジュールで検出されると、これらのコマンドはテルミン演奏ロボットへ送付され、ロボットが現在演奏中の楽譜の位置に応じて、楽曲の開始、あるいは、フェルマータの終了が行われる。現時点でのテンポがないとき（たとえば、曲の演奏前）には、テンポ検出モジュールは画像ビートキー検出だけを使用して初期テンポを決定する。それ以外は、前述したとおり情報統合モジュールから音響と画像の 2 つのモダリティからマッチするビートを抽出し、検出したテンポをテルミン演奏ロボットに送付する。

#### 4. 実験・評価およびデモ

本稿では 3 種類の実験を行い、ロボット共演者の実行可能性を評価する。最初の実験では、複数の照明条件下で画像キー検出アルゴリズムの評価を行う。2 番目の実験では、複数の被験者によるジェスチャ認識の性能評価を行う。3 番目の実験では、音響ビートキー検出と画像ビートキー検出とを情報統合によるテンポ検出の評価を行う。

##### 4.1 実験 1：画像開始キーと終了キーの検出

最初の実験では、開始キーと終了キーの検出モジュールの性能を評価する。中級のフルート奏者 1 名（被験者 A）が 3 つの異なる照度で各ジェスチャを 30 回ずつ実演し（図 8）、画像キーの検出を行う。なお、奏者の背景は図 8(a) から分かるように単純なものとなっ

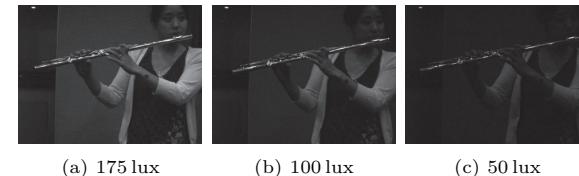


図 8 異なる照明条件でのロボットのカメラから取得した実画像

Fig. 8 Actual input images from robot's camera for our three experimental conditions.

表 1 各タイプのジェスチャの再現率

Table 1 Recognition rates of each type of gesture (Precision).

画像キー	175 lux	100 lux	50 lux
開始キー (%)	97	100	83
終了キー (%)	100	97	100

ている。

得られた各キーの再現率を結果を表 1 に示す。終了キーについては、ほぼ確実に得られ、開始キーも 50 lux 以外はほぼ確実にとれていることが分かる。

##### 4.2 実験 2：複数の被験者によるジェスチャ認識

前節での 1 名に加えて、京都大学オーケストラの 2 名のフルート演奏者を被験者とした。被験者 B は上級演奏者であり、被験者 C は中級演奏者である。実験は各人別々の日に行つた。使用した楽譜は次の 2 種類である。(1) 同時開始の楽譜、(2) フェルマータで同時終了の楽譜。各被験者はリーダ役で、2 回ずつ演奏をしてもらった。ただし、被験者にはリードの方法は知らせず、各自のやりかたでフルートを動かしてもらった。この様子をビデオ撮影をし、解析を行った。

解析結果から、以下の知見が得られた。

- (1) 実験を行った開始キーと終了キーはいずれの被験者でもうまく認識できた。
- (2) フルートの動きは、被験者 C は Down-Up-Down の単純形であったが、被験者 B は Up-Down-Up-Down と最初に軽く上にフルートが上がっていた。被験者 B の場合でも設計した FSM で受理できるので、いずれの場合にも図 4(a) で示した FSM でうまく認識できた。
- (3) 終了キーについては、いずれの奏者も Down-Up という動きをとっていたので、図 4(b) で示した FSM でうまく認識できた。

被験者  $\beta$  の動作は被験者  $\gamma$  よりも大きく、軽く息を吸う音が聞こえた。これは息を吸うために肺が膨らみ、その結果フルート本体が上昇する生理的な動きに対応していると考えられる。

#### 4.3 実験 3：音響オンセット検出 + 画像ビートキュー

3番目の実験では、175 lux に照明を固定して、情報統合モジュールのテンポ変化検出の性能を、被験者のそれと比較して評価する。被験者  $\alpha$  が 2つのレガート音符、A2 と B $\flat$ 2 をタンギングをせず滑らかに、交互に演奏する（図 9 参照）。音符の変化に対して、フルート

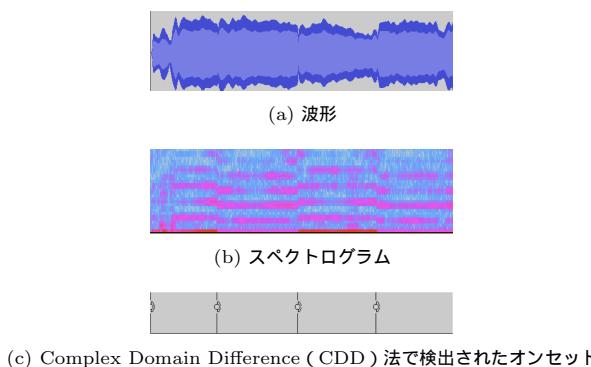


図 9 フルートでレガート音符として演奏された 4 音符。横軸は時間  
Fig. 9 Four notes played on flute with legato onsets.

奏者は画像ビートキューを実行する。被験者は、クラシック音楽の教育を受けた中級のクラリネット奏者であり、音符が変わるとコンピュータのキーをたたく。これにより、被験者が検出したテンポ (IOI で表現) が得られる。

図 10 に実験結果のタイムラインを示す。75 個のビート (音符) を演奏し、画像モジュールは 75 個すべての画像キューを正しく検出した。そのうち 72 回がマッチしたビートであり、間違った音符オンセットのうち、3 回が false positive, 3 回が false negative であった。本システムが検出した IOI と人間の被験者が検出した IOI との絶対誤差の平均は 46 ms であり、標準偏差は 32 ms である。

テンポ検出の相対誤差をヒストグラムで表すと、図 10 に示したように、白色雑音であるガウス分布に似た分布を示す。人がタップするタイミングパターンは白色/ピンク雑音に似ている<sup>28)</sup>と報告されているので、その影響が出ている可能性が否定できない。今後、本実験のように人間の実験結果を正解と見なすのではなく、メトロノームを用いた客観的な実験を行い、より正確な評価をする必要がある。

音符オンセット検出の絶対誤差の平均は 180 ms であり、標準偏差は 47 ms である。図 10 (a) からビート検出が 0.5 秒以内で行えれば、本手法はほぼ安定してテンポ推定を行っていることが分かる。マッチしたビートが 2 つ連続して検出されるごとに、テンポ IOI が瞬時的テンポとして計算される。2 番目の音響ビートがマイクロフォンに入力された時間と実際にロボット共演者の内部テンポが変更された時点との差から得られるテンポ検出の平均遅延は、上記の実験では 231 ms である。より詳しくいうと、2.13 GHz の Mac Book を音響キュー

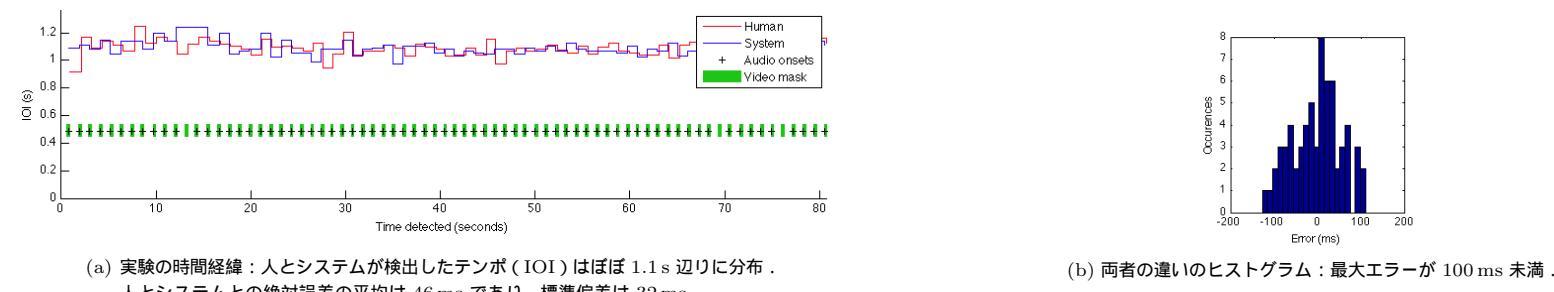


図 10 75 回ビートが演奏されたときの、システムと被験者のテンポ認識の比較  
Fig. 10 (Left) Experiment timeline: Over 75 notes played, both human and system detected tempos (IOI) remained around 1.1 s; average absolute error between human and system is 40 ms. (Right) A histogram of deviation between the human and system shows that most errors were less than 100 ms.

検出に，HRP-2 内部の画像処理ボードを画像キュー検出に，Lenovo T61p を情報統合に使用した場合，遅延 500 msec 以内でビートキューを検出することができた。

従来のビート追跡法では，過去の音符の履歴に基づいて，相互相關などの手法を用いてテンポを抽出していた。たとえば，村田らのビート追跡モジュール<sup>13)</sup>は，1 秒間の窓幅をパターン照合用に使用するので，テンポ変化の検出に 2 秒を要する。もし，本手法のような視聴覚情報統合によるテンポ推定により，直接瞬時的なテンポを 2 秒以内に検出できれば，村田の手法のようなテンポ推定に対して枝刈りのような処理を加えることにより，テンポ推定の精度を向上させ，テンポ追従の遅れを軽減できる。その結果，同期できずに演奏する時間をもっと縮めることも期待できる。

#### 4.4 フルート奏者がアンサンブルリーダの合奏

最終的な応用は，図 1 に示した実際のアンサンブル演奏である。フルート奏者がアンサンブルリーダを担い，リーダの指示に従って音楽共演ロボット HRP-2 がテルミンを演奏し，VOCALOID Prima を使って歌を歌う。取り上げた楽曲は，“Les Anges Dans Nos Campagnes”（荒野の果てに，讃美歌としても歌われる）というフランスの伝統的なクリスマスキャロルである。

- (1) 演奏開始前にリーダがフルート操作で 46 BPM で演奏することを指示，
- (2) リーダが開始キューを指示し，全員が同時に演奏を開始，
- (3) リーダが途中でビートキューで 66 BPM での演奏に変更を指示，
- (4) フェルマータで停止中にリーダが終了キューを指示し，全員が同時に終了。

本デモのように，共演者を求めていたフルート奏者が，自分がリーダとなって自分好みのアンサンブルを楽しむことができる。本演奏では，画像処理の精度を上げるために，背景が無地の部屋でアンサンブル演奏を行っている。なお，本デモのビデオは下記の URL から視聴可能である。<http://winnie.kuis.kyoto-u.ac.jp/members/angelica/>

#### 4.5 考察と今後の課題

本研究で得られた知見について，議論する。

##### 画像開始キューと終了キューの有効性：

演奏音がない状態でのジェスチャによる開始キューや終了キューが有効であることが分かり，それを活用したロボット共演者を開発することができた。ここでは，フルート奏者がフルートの口から遠い端をリズムをとるように振るというよく観察されるジェスチャを使用している。このジェスチャはフルート奏者にはそれほど違和感がないと考えられる。本研究で，画像キューの有効性が判明したので，この技術をギターのビート追跡にも応用してい

る<sup>14),15)</sup>。他の楽器，たとえば 2.1 節で述べたクラリネットなどに，本手法を応用することは可能である。

##### 画像ビートキューの有効性：

演奏の途中でテンポ変化を指示するのは，新たなメッセージを開始するような場合には開始キューと同様にある程度有効ではあるものの，演奏途中で演奏をしながらテンポ変化を指示するのは有効性が明らかにならない。今後，ビートトラッキング法を高度化するか，あるいは，楽譜追跡法を導入し，その有効性を検証していく必要がある。

##### フルート演奏に対する楽譜追跡：

本稿で提示したフルート演奏のビート追跡法の改良としては，パート譜を利用した楽譜追跡が有望である。特に，フルートはモノフォニックな楽器であるので，オンセットだけでなく音高も容易に抽出できる。したがって，古典的な手法である動的計画法による楽譜追跡<sup>22)</sup>が適用可能であると考えられる。ただし，フルート演奏用のパート譜を用意し，さらに，ロボット演奏用パート譜との時間的な対応付けをとっておくなどの前処理が必要となる。一方，本システムでは，ビート追跡が非力な分，ロボット演奏用パート譜（Standard MIDI file）があればすぐに合奏可能というメリットもある。

##### 演奏のテンポ揺らぎへの対応：

演奏のテンポ揺らぎについては，合奏では避けることのできない課題であり，よりロバストなビート追跡法にはロボットのパート楽譜とタイミング付が行われている人間演奏者のパート楽譜を用いた楽譜追跡法が不可欠である。現在，パーティクルフィルタによる楽譜追跡法<sup>11),15)</sup>を開発中であり，本稿で使用しているビート追跡法よりも性能が向上している。今後，これらのシステムを組み込み，よりロバストな合奏システムの開発を行っていく。

ただし，このような楽譜追跡法を導入しても，演奏が進展するにつれてテンポ揺らぎによる楽譜追跡の累積エラーが大きくなり，うまく合奏ができない可能性が高い<sup>11)</sup>。このような累積エラーによる悪影響を避けるためには，演奏時刻の利用や適切なタイミングで強制的に同期させるようなエラーからの高次の復旧処理が不可欠であると考えられる<sup>10),22)</sup>。

本システムでの弱点の 1 つは，テンポ検出精度がフルート奏者の技量に大きく依存していることである。マッチング閾値を広げれば，検出誤りを小さくすることができる。今後さらなる評価を通じて適切な閾値を設定する必要がある。

##### 情報統合の可能性：

画像情報と音響情報とを統合して処理精度を上げる方法としては次の 3 つの方法が考えられる：

## (1) 画像情報を音響情報のフィルタとして使用：

本稿で使用した情報統合は、ビート検出のために得られた音響情報のうち、ビート検出に関与しないノイズを画像情報でフィルタリングし、ビート間隔から行うテンポ推定の精度を向上させている。

## (2) 音響情報を画像情報のフィルタとして使用：

白鳥らが使用した情報統合<sup>20)</sup>は、舞踏動作のセグメント化のために得られた画像情報のうち、セグメント化に関与しないノイズを音情報から得られたビートと合致するかどうかでフィルタリングし、セグメント化と留動作推定の精度を向上させている。

## (3) 画像情報と音響情報を同一レベルで使用：

糸原らが使用した情報統合<sup>14),15)</sup>は、ギター奏者の手の動きとギター演奏音から得たビートとをパーティクルフィルタで同時に扱うことによって、楽譜追跡とテンポ予測の精度を向上させている。

また、現行のシステムで視聴覚情報統合時の画像閾値を広げると、現行の視聴覚情報統合技法が有効なのは、演奏テンポが遅く、かつ、音符が少ない楽曲に限定されてしまうことになる。一般に、演奏者は楽譜中の技巧的な演奏フレーズでは、不必要的動きをすべて止める傾向にあり、易しい部分では動きが増す傾向にあると報告されている<sup>23)</sup>。今後、短時間に多くの音符が現れるリズミックな性質をうまく活用し、ジェスチャのような画像情報がなくてもテンポ検出ができるようにする必要がある。

## 今後の課題

ロボットと人との合奏の重要な研究の方向性は、ロボットに表現力を与えることである。現状では、ロボットはアンサンブルリーダの演奏にできるだけ忠実に従うように設計されている。もし、ロボットが自分自身のタイミング間隔を有しているとすれば、タイミング変化を独立に予期し、人との同期は軽微なもので済ませることも可能である。Mizumoto らは、カエルの合唱にヒントを得て、人とロボットが互いに独立な蔵本モデルによる非線形振動子の結合モデルとして合奏のテンポの引き込みを実現している<sup>29)</sup>。相手の演奏に対する信念モデルを有した高度な合奏機能も必要であろう。

現在、音響処理においてテンポ予測は実装中である。ただし、テンポの揺らぎが小さい場合には、そのようなテンポ予測は無視され、フルート奏者がジェスチャで指示したテンポを維持するようにしている。また、人の演奏パート譜を使用した楽譜追跡も実現し、よりロバストな合奏を目指してゆく。

今後、より多くの奏者と実験を行い、より広範囲な評価を行うこと、あるいは、接話型マ

イクロフォンではなく、ロボットのマイクロフォンの使用も重要である。この場合、ロボットの耳にはフルート演奏音だけでなく、ロボット自身の演奏音（テルミンと歌声）も混ざった混合音となるので、そこからフルート演奏音だけを抽出する機能<sup>13)</sup>も不可欠である。これについては、既開発のセミブラインド音源分離<sup>30)</sup>を使用する予定である。最後に、フルートやギターに加えて、クラリネット、バイオリン、尺八といった他の楽器についても検討をしてゆきたい。

## 5. 終わりに

本稿では、ロボットがフルート奏者をアンサンブルリーダとして演奏するときにリーダがジェスチャで開始を指示する開始キューやフェルマータでの停止状態をジェスチャで終了を指示する終了キューが合奏にとって有効であることを示した。また、演奏途中でのテンポ変化を指示する画像キューを使用することにより、演奏音から得られるビートの候補のうち、ノイズを画像キューでフィルタリングすることによりビート検出が0.5秒内で行えれば、ビート追跡の性能が安定することを示した。

我々の究極の目的は、人のような表現力を持って音楽を演奏することができるロボット共演者を構築することである。その第1歩として、テルミン演奏ロボットがフルート奏者の演奏を聞き、見て、テンポの指示に合わせて、テルミンを演奏し、歌を歌う三重奏を実現した。本稿で得られた知見はフルート演奏に限定的であるが、一部はギター演奏にも展開されており、より高性能化・ロバスト化を図り、他の楽器への展開も行う必要がある。

謝辞 最後に、本稿に対して詳細なコメントをいただいた査読者の方々に感謝する。本研究の一部は、科研費基盤研究(S)、科研費特定領域研究「情報爆発IT基盤」、および、Global COEの援助を受けた。

## 参考文献

- 1) 奥乃 博, 中臺一博, 大塚琢馬: 音楽ロボットのための実時間音楽情報処理, 情報処理, Vol.50, No.8, pp.729–734 (2009).
- 2) Fredrickson, W.E.: Band musicians' performance and eye contact as influenced by loss of a visual and/or aural stimulus, *Journal of Research in Music Education*, Vol.42, pp.306–317 (Jan. 1994).
- 3) Goebel, W. and Palmer, C.: Synchronization of timing and motion among performing musicians, *Music Perception*, Vol.26, pp.427–438 (May 2009).
- 4) 五十棲隆勝, 赤地一彦, 平田 勝, 金子健二, 梶田秀司, 比留川博久: ヒューマノイ

- ドロボット HRP-2 の開発 , 日本ロボット学会誌 , Vol.22, No.8, pp.1004–1012 (Aug. 2004).
- 5) 水本武志 , 辻野広司 , 高橋 徹 , 駒谷和範 , 尾形哲也 , 奥乃 博 : テルミンの音高・音量特性のモデルに基づくテルミン演奏ロボットの開発 , 情報処理学会論文誌 , Vol.51, No.10, pp.2008–2019 (Oct. 2010).
  - 6) Dannenberg, R.B.: An On-Line Algorithm for Real-Time Accompaniment, *Proc. International Computer Music Conference*, pp.193–198 (1984).
  - 7) Horiuchi, Y. and Tanaka, H.: A Computer Accompaniment System With Independence, *Proc. International Computer Music Conference*, pp.418–420 (1993).
  - 8) Grubb, L. and Dannenberg, R.B.: Enhanced Vocal Performance Tracking Using Multiple Information Sources, *Proc. International Computer Music Conference*, pp.37–44 (1998).
  - 9) Raphael, C.: Orchestra in a Box: A System for Real-Time Musical Accompaniment, *Proc. International Joint Conference on Artificial Intelligence*, pp.5–10 (2003).
  - 10) Dannenberg, R.B. and Raphael, C.: Music score alignment and computer accompaniment, *Comm. ACM*, Vol.49, No.8, pp.44–48, ACM (Aug. 2006).
  - 11) Otsuka, T., Nakadai, K., Takahashi, T., Ogata, T. and Okuno, H.G.: Real-Time Audio-to-Score Alignment using Particle Filter for Co-player Music Robots, *EURASIP Journal on Advances in Signal Processing*, Vol.2011, Article ID 384651, p.13, Hindawi Pub (2011).
  - 12) Weinberg, G. and Driscoll, S.: Robot-human interaction with an anthropomorphic percussionist, *SIGCHI*, pp.1229–1232 (2006).
  - 13) 村田和真 , 中臺一博 , 武田 龍 , 奥乃 博 , 長谷川雄二 , 辻野広司 : 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, No.7, pp.793–801 (Sep. 2009).
  - 14) 糸原達彦 , 大塚琢馬 , 水本武志 , 高橋 徹 , 尾形哲也 , 奥乃 博 : 視聴覚統合ビートトラッキングを用いた音楽ロボットとギターとの合奏システム , 情報処理学会第 73 回全国大会 , 4-235-236, 1ZB-2 (Mar. 2011).
  - 15) Itohara, T., Mizumoto, T., Otsuka, T., Ogata, T. and Okuno, H.G.: Particle-filter Based Audio-visual Beat-tracking for Music Robot Ensemble with Human Guitarist, *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2011)*, accepted, IEEE, RSJ, San Francisco (Sep. 2011).
  - 16) Weinberg, G., Raman, A. and Mallikarjuna, T.: Interactive jamming with Shimon: a social robotic musician, *HRI*, pp.233–234 (2009).
  - 17) Solis, J., Chida, K., Taniguchi, K., Hashimoto, S.M., Suefuji, K. and Takanishi, A.: The Waseda flutist robot WF-4RII in comparison with a professional flutist, *Computer Music Journal*, Vol.30, No.4, pp.12–27 (2006).
  - 18) Petersen, K., Solis, J. and Takanishi, A.: Development of a real-time instrument tracking system for enabling the musical interaction with the Waseda Flutist Robot, *IROS-2008*, pp.313–318 (2008).
  - 19) Overholt, D., Thompson, J., Putnam, L., Bell, B., Kleban, J., Sturm, B. and Kuchera-Morin, J.: A multimodal system for gesture recognition in interactive music performance, *Computer Music Journal*, Vol.33, No.4, pp.69–82 (2009).
  - 20) 白鳥貴亮 , 中澤篤志 , 池内克史 : モーションキャプチャと音楽情報を用いた舞踊動作解析手法 , 電子情報通信学会論文誌 , Vol.J88-D2, No.8, pp.1662–1671 (Aug. 2005).
  - 21) 堀内靖雄 , 西田昌史 , 市川 翼 : プレスによる合図を検出する伴奏システム , 情報処理学会論文誌 , Vol.50, No.3, pp.1079–1089 (Mar. 2009).
  - 22) 堀内靖雄 , 橋本周司 : 自動伴奏システム , 情報処理 , Vol.35, No.9, pp.815–821 (Sep. 1994).
  - 23) Wanderley, M., Vines, B., Middleton, N., McKay, C. and Hatch, W.: The musical significance of clarinetists' ancillary gestures: An exploration of the field, *Journal of New Music Research*, Vol.34, No.1, pp.97–113 (2005).
  - 24) Bolles, R.C. and Fischler, M.A.: A RANSAC-based approach to model fitting and its application to finding cylinders in range data, *IJCAI-1981*, pp.637–643 (1981).
  - 25) Duxbury, C., Bello, J.P., Davies, M. and Sandler, M.: A combined phase and amplitude based approach to onset detection for audio segmentation, *WIAMIS*, pp.275–280 (2003).
  - 26) Brossier, P.M.: Automatic Annotation of Musical Audio for Interactive Applications, Ph.D. thesis, Queen Mary University of London (2006).
  - 27) Mills, D.: Network Time Protocol (Version 3) Specification, Implementation and Analysis (1992).
  - 28) Gilden, D.L., Thornton, T. and Mallon, M.W.: 1/f noise in human cognition, *Science*, Vol.267, p.1837 (1995).
  - 29) Mizumoto, T., Otsuka, T., Nakadai, K., Takahashi, T., Komatani, K., Ogata, T. and Okuno, H.G.: Human-Robot Ensemble between Robot Thereminist and Human Percussionist using Coupled Oscillator Model, *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2010)*, pp.1957–1962, Taipei (Oct. 2010). DOI:10.1109/IROS.2010.5650364.
  - 30) 武田 龍 , 中臺一博 , 駒谷和範 , 尾形哲也 , 奥乃 博 : 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, No.7/8, pp.782–792 (2009).

(平成 23 年 4 月 11 日受付)

(平成 23 年 9 月 12 日採録)



リム アンジェリカ（学生会員）

2008 年加国サイモン・フレーザ大学計算機科学科卒業。2009 年国費留学生として来日、京都大学大学院情報学研究科研究生。2010 年同大学同研究科知能情報学専攻入学。2008 Google Canada Anita Borg Memorial Scholarship 等受賞。IROS-2010 NTF Award for Entertainment Robots and Systems 受賞。合奏する表情豊かな音楽ロボットの研究に従事。IEEE,

日本ロボット学会各会員。



水本 武志（学生会員）

2008 年京都大学工学部情報学科卒業。2010 年同大学大学院情報学研究科知能情報学専攻修士課程修了。同年同専攻後期博士課程進学。学振特別研究員 (DC2)。主にテルミン演奏ロボットの開発と人とロボットの合奏の研究に従事。IROS2008 Award for Entertainment Robots and Systems Nomination Finalist, 情報処理学会第 71・72 回全国大会学生奨励賞。IEEE, 日本ロボット学会各会員。



大塚 琢馬（学生会員）

2009 年京都大学工学部情報学科卒業。2011 年同大学大学院情報学研究科知能情報学専攻修了。同年同専攻後期博士課程進学。学振特別研究員 (DC1)。音楽ロボットのためのペイズ手法に基づいた楽譜追跡と合奏技法の研究に従事。IEA/AIE-2010 Best Paper Award, (財) C&C 振興財団 2010 年度 C&C 若手優秀論文賞受賞。IEEE, 日本ロボット学会各会員。



古谷 ルイ 賢造 カイエ (学生会員)

2008 年仏国ニース・ソフィア・アンチポリス大学計算機科学科卒業。2009 年同大学大学院修士課程計算機科学専攻修了。2010 年京都大学大学院情報学研究科知能情報学専攻後期博士課程入学。画像情報と音響情報を統合した Audio-Visual SLAM の研究に従事。IROS-2010 NTF Award for Entertainment Robots and Systems 受賞。日本ロボット学会会員。



尾形 哲也（正会員）

1993 年早稲田大学理工学部機械工学科卒業。日本学術振興会特別研究員, 早稲田大学理工学部助手, 理化学研究所脳科学総合研究センター研究員, 京都大学大学院情報学研究科講師を経て, 2005 年より同助教授(現・准教授)。博士(工学)。JST さきがけ研究「情報環境と人」領域研究員。研究分野は人工神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学。RSJ, JSM, JSAT, SICE, IEEE 等会員。



奥乃 博（正会員）

1972 年東京大学教養学部基礎科学科卒業。日本電信電話公社, 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。博士(工学)。この間, スタンフォード大学客員研究員, 東京大学工学部客員助教授。人工知能, 音環境理解, ロボット聴覚, 音楽情報処理の研究に従事。JSAT, JSSST, RSJ, ACM, IEEE, AAAI 等会員。