

研究会推薦論文

仮想ジャズセッションシステム：VirJa Session

後藤 真孝^{†,☆} 日高 伊佐夫^{†,☆☆} 松本 英明^{†,☆☆}
黒田 洋介[†] 村岡 洋一[†]

本論文では、すべてのプレーヤーが対等な立場でインタラクションし、即興演奏するジャズセッションシステムについて述べる。本研究の目的は、人間と計算機とが影響を与え合いながら演奏する新しい統合演奏環境を実現することである。ジャズではプレーヤー間のインタラクションが重要であるが、従来の多くのシステムでは、人間のソロ演奏に対して、計算機が他のプレーヤー全員の演奏を伴奏としてまとめて生成していた。本論文では、計算機内のプレーヤー同士も、人間同様にお互いの演奏を聞き合っているシステムを提案する。そして、その発展形として、各プレーヤーがお互いの姿を見ることもできる仮想ジャズセッションシステム VirJa Session を提案する。本システムでは、計算機プレーヤーの姿やジェスチャーがコンピュータグラフィックスで視覚化され、計算機プレーヤーが他のプレーヤーのジェスチャーをカメラ等を通じて認識できる。こうして、全プレーヤー間のマルチモーダルインタラクションを実現することで、従来の音だけのセッションシステムに比べ、より臨場感のあるセッションが達成できる。現在の実装では、ジャズのピアノトリオを対象とし、人間がピアニスト、計算機がベーシストとドラマーを担当する。両計算機プレーヤーを独立したプロセスとして複数の計算機上に実装し、実験を行った結果、提案したジャズセッションが実現できたことを確認した。

A Virtual Jazz Session System: VirJa Session

MASATAKA GOTO,^{†,☆} ISAO HIDAKA,^{†,☆☆} HIDEAKI MATSUMOTO,^{†,☆☆}
YOSUKE KURODA[†] and YOICHI MURAOKA[†]

This paper presents a jazz session system in which each player is independent and can interplay with all other players. The purpose of this research is to build a new performance environment that facilitates interplay among humans and computers. Although interaction among players is important in jazz, computer accompaniment parts of most of the previous systems were collectively generated as a single task and only reacted to the human soloist's performance. This paper proposes a system that enables computer players to listen to other computer players' performances as well as the human players' performances and to interact with each other. This paper moreover proposes an advanced virtual jazz session system called *VirJa Session* which also enables all players to see each others' gestures. In our system, the bodies and gestures of computer players are visualized on computer graphics and each computer player can recognize other players' gestures. Thus, we can achieve multimodal interaction among all players. In our current implementation, the system deals with a jazz piano trio consisting of a human pianist, a computer bassist, and a computer drummer. Both computer players have been implemented as separate processes on a distributed environment of multiple workstations. Through our experiments, we verified that our proposed objectives have been achieved.

1. はじめに

ジャズセッションにおいて重要なのは、すべてのプレーヤーが対等な立場で、お互いの演奏を聞き合いながら即興演奏をすることである。これはインタープレ

† 早稲田大学理工学部
School of Science and Engineering, Waseda University
☆ 現在、電子技術総合研究所
Presently with Electrotechnical Laboratory
☆☆ 現在、ソニー株式会社
Presently with Sony Corporation

本論文の内容は、1996年2月の音楽情報科学研究会にて報告され、同研究会主査により情報処理学会論文誌への掲載が推薦された論文である。

イ (interplay) とも呼ばれ、プレーヤーが主従関係を持たずに、相互に相手の音に耳を傾けて音で反応する。本研究は、このようなインタープレイを人間と計算機上の複数のプレーヤー間で達成するために、新しい統合演奏環境を提供するジャズセッションシステムを実現することを目的とする。

従来のジャズセッションシステム^{1)~3)}や自動伴奏システム^{4)~10)}の多くでは、人間のソロ演奏に対して、計算機が他のすべてのプレーヤーの演奏を伴奏としてまとめて生成していた。そこでは、人間の演奏に対していかに適切な伴奏をするかが研究の中心となっており、多くの場合、人間が主で計算機が従の関係にあった。また、計算機が担当するプレーヤーは人間の演奏しか聞けなかったため、計算機のプレーヤー同士はインタラクション^{**}できなかった。一方、文献 11) では、マルチエージェントモデルによる計算機のプレーヤー同士のインタラクションに言及しているが、音楽聴取過程の実現が中心であり、実際の演奏の生成は行っていなかった。演奏生成も行う演奏環境を提供するシステムという観点からは文献 12) の研究も関連するが、インタラクティブな作曲環境という性格が強く、人間同士のジャズセッションを人間と計算機間で達成するものではなかった。

本論文では、すべてのプレーヤーが対等な関係にあり、計算機が担当するプレーヤー同士もインタラクションできるジャズセッションシステムを提案する。本システムでは、計算機のプレーヤーは人間の演奏に反応するだけでなく、ソロを弾くなどの自己主張も行う。さらに、実際の人間同士のセッションにおいて各プレーヤーが自分以外のすべてのプレーヤーの演奏を聞くように、計算機内のプレーヤー同士も人間の演奏と同様にお互いの演奏に耳を傾け、影響を与え合って演奏することができる。

本論文ではさらに発展形として、各プレーヤーがお互いの姿を見ることができ、仮想ジャズセッションシステム VirJa Session (Virtual Jazz Session System) も提案する。本システムでは、計算機のプレーヤーが演奏する姿は、コンピュータグラフィックス (CG) により表示される。これにより、人間は音を聞くだけで

なく相手の動きも見ることができ、人間同士のジャズセッションに近い、より臨場感のあるセッションを行える。同時に、計算機の各プレーヤーはカメラを通じて人間が演奏する姿を見ながら、通信によって仮想的に、他の計算機プレーヤーの姿やジェスチャー (身体的な合図) も見る。こうして、すべてのプレーヤー間で、演奏の終わりなどを指示するジェスチャーを交換することができ、聴覚と視覚の両方を通じたマルチモーダルインタラクションを実現できる。

現在の実装では、本システムはジャズのピアノトリオの演奏を対象とし、人間がピアニスト、計算機がベーシストとドラマーを担当する。そして、テンポが一定で演奏される 4ビートのジャズスタンダード曲を扱う。計算負荷の分散のために、本システムをネットワークに接続された複数の計算機上に実装した。その際、実装を容易にし拡張性を高くするために、ベーシストとドラマーを独立したプロセスとして実現し、さらにこれらの演奏理解・生成部と CG 表示部も別々のプロセスとした。実験の結果、実際にプレーヤーが聴覚と視覚を通じてインタラクションしながら即興演奏をするセッションを達成できた。

2. すべてのプレーヤーが対等なジャズセッションシステム

ジャズセッションでは、すべてのプレーヤーがお互いの即興演奏を聞き合い、固定した主従関係を持たずに対等な立場でインタラクションすることが重要である。そこで以下では、まずこのようなジャズセッションのモデルを提示した後、そのモデルをジャズのピアノトリオへと適用する。

2.1 ジャズセッションのモデル

すべてのプレーヤーが対等なジャズセッションシステムは、図 1 のようなモデルに基づいている。ここではカルテット (4 人) の場合を例示したが、一般に n 人 (n は 2 以上の整数) の場合に適用できる。プレーヤーは自分以外のすべてのプレーヤーの演奏を聞き、状況に応じて音で反応したりソロを演奏したりする。

各プレーヤーを人間が担当するか計算機が担当するかはモデルとしては自由である。たとえば、2 人の人間がプレーヤー 1 とプレーヤー 2 を担当するときには、計算機はプレーヤー 3 とプレーヤー 4 を担当する。さらに、すべてのプレーヤーを人間が担当してもよいし、すべてのプレーヤーを計算機が担当してもよい。すべてを人間が担当する前者の場合は、複数箇所にいるプレーヤー間でのセッションなどに有効である。これは、6 章で後述するようにコンピュータネットワークを介

* ジャズセッションシステムにおいて演奏を「聞く」とは、他者の演奏情報を分析・理解して、自分の即興演奏に反映させるための情報を抽出することを意味する。

** 本論文での「インタラクション」という用語は、ヒューマンインタフェース研究における「インタラクション」とは相互作用という観点で共通であるが、演奏時の意図の理解と表現をとまなう、音楽的なより高次の相互作用という意味も含んでいる点異なる。

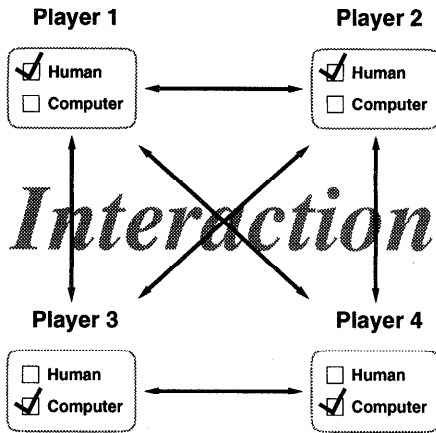


図1 すべてのプレイヤーが対等なジャズセッションのモデル
Fig. 1 A jazz session model for interplay among all players.

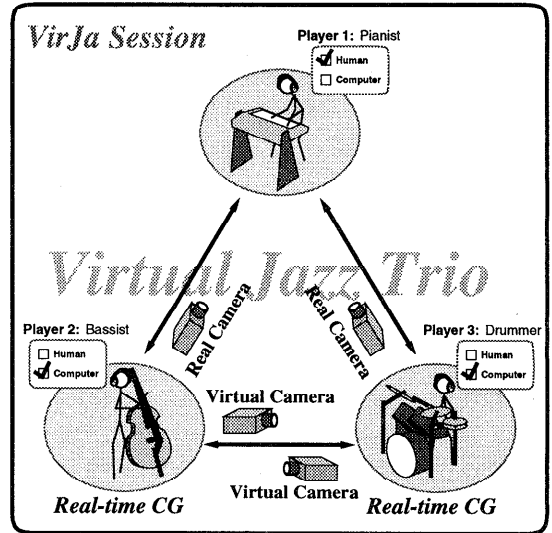


図3 VirJa Session のモデル
Fig. 3 A model of VirJa Session.

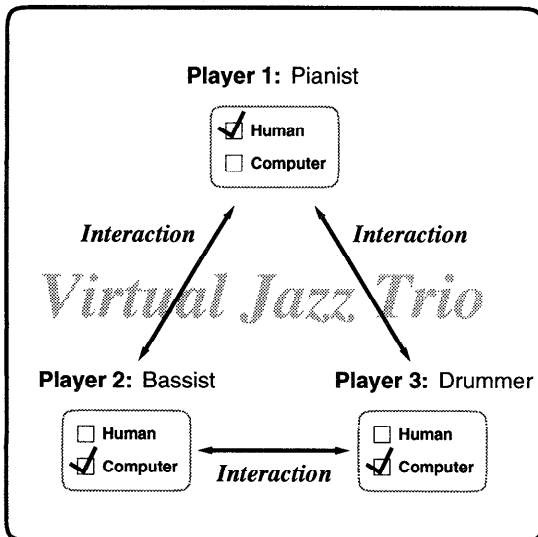


図2 ピアノトリオのジャズセッションのモデル
Fig. 2 A jazz session model for piano trio.

して演奏情報をやりとりすることで実現できる。すべてを計算機が担当する後者の場合は、様々な個性を持たせたプレイヤーや異なる人が設計したプレイヤー同士と一緒に演奏させるセッションなどに有効である。そこでは、計算機上のプレイヤーがいわば設計者の分身となり、それらの分身同士のインタラクションが行われる。

2.2 ジャズのピアノトリオへの適用

本研究では、以上のモデルをジャズのピアノトリオへ適用したシステムを実現する(図2)。プレイヤー1のピアニストを人間が担当し、プレイヤー2のベーシスト、プレイヤー3のドラマーを計算機が担当する。

ここで重要なのは、計算機がプレイヤー2とプレイヤー3をともに担当する場合にも、それらは独立したプレイヤーとして処理されることである。プレイヤー2は、プレイヤー1の人間の演奏だけでなく、プレイヤー3の計算機の演奏にも反応して演奏する。つまり、プレイヤー1の演奏だけを解釈して、プレイヤー2, 3の演奏を従属的に決定するのではない。ここでは、プレイヤー2(ベーシスト)とプレイヤー3(ドラマー)が演奏によって影響を与え合うことができる。

3. 姿の見える仮想ジャズセッションシステム: VirJa Session

図2のモデルを発展させた、各プレイヤーがお互いの姿を見ることが出来るジャズセッションシステム VirJa Session のモデルを図3に示す。2章で述べたプレイヤーは、頭脳以外には演奏を聞くための耳と演奏するための手足だけしか持っていなかった。ここで提案する VirJa Session のプレイヤーは、さらに他のプレイヤーを見るための目と自分の状態を表す体全体を持つ。これにより各プレイヤーは演奏音以外にも視覚情報を交換することができ、より人間同士のジャズセッションに近付いた、聴覚と視覚の両方による臨場感のあるインタラクションを実現できる。また、VirJa Session の演奏を観客として楽しむ立場からは、すべてのプレイヤーの姿が見えることで、ライブ演奏を見ているときのようにプレイヤーの存在を感じることができる。

計算機が担当するベーシストとドラマーの演奏する

姿・ジェスチャーは、三次元CGアニメーションによってリアルタイムに表示され、人間は演奏中にこれらを見ることができる。一方、計算機の両プレイヤーは、図3のように実際のカメラにより人間が演奏する姿をとらえ、ジェスチャーを認識する。同時に、通信で実現される仮想的なカメラにより他方の計算機プレイヤーの姿を見る^{*}。こうして、各プレイヤーはお互いにジェスチャーを交換することができ、演奏音だけでは伝えにくい情報を介した協調が可能になる。

3.1 CGによるプレイヤーの可視化

CGの画面上にはプレイヤーとその演奏する楽器（ベースまたはドラムス）が表示され、プレイヤーは演奏音に合わせた演奏動作を行う。さらに両計算機プレイヤーは、他のプレイヤーとのコミュニケーションをとるために、次節で述べる2種類のジェスチャーの動作やうなずく動作をしたり、他のプレイヤーに視線を向けたりする。また曲のビートに合わせて体を揺らし足踏みなども行う。

なお、CGによってミュージシャンの姿を可視化する試みとしては、文献13)や文献14)があげられるが、ギタリストの手の動きの生成が中心となっていた。また、文献15)では提案だけにとどまっていた。

3.2 カメラによる視覚を通じたインタラクション

VirJa Sessionでは、テーマ演奏部・各プレイヤーのソロ演奏部・フォーバース（four verses）部を組み合わせた演奏進行表を事前にシナリオとして与えることができる。たとえば、次のようなシナリオに沿って演奏できる：テーマ演奏部 ⇒ ピアノソロ演奏部 ⇒ ベースソロ演奏部 ⇒ フォーバース部（ピアニストとドラマーが交互に4小節間ソロを演奏）⇒ テーマ演奏部。そして、シナリオ中の各部の繰返し回数（何コーラス演奏するか）は、最大回数を制限するだけで固定とはせず、ジェスチャーなどを用いた演奏中のインタラクションによって動的に決定できるようにする。一般に人間同士のセッションにおいても、シナリオは事前に打合せをして決めておくが、このシナリオ中の各部の繰返し回数は演奏時に決定されることが多い。

本システムではそのための合図として、次のプレイヤーのソロ演奏部に移るように指示するジェスチャーと、テーマ演奏部に移る（戻る）ことを指示するジェスチャーの2種類を用いる。前者は体を左右に傾けるジェスチャーで、傾けた側にいるプレイヤーにソロを指示する。後者は手で頭を指すジェスチャーにより指

示する。ただし、3人のプレイヤーはお互いに向き合うように三角形の各頂点に配置され、この配置はあらかじめ与えておく。

計算機プレイヤーは、人間のこのようなジェスチャーをカメラ画像から認識して判断する。また計算機プレイヤー同士は、通信によりジェスチャーを交換して判断する。一方、人間はCGによって可視化された計算機プレイヤーのジェスチャーを見て判断する。現在の実装では、計算機プレイヤーは、シナリオ中のソロ演奏部やフォーバース部で最大繰返し回数に達したときに、次へ移ることを人間に知らせるためにジェスチャー動作を行う^{**}。

計算機の担当するベーシストとドラマーは視線を持っており、シナリオ中の各部において終わりに近付くと、再び繰り返すか次に移るかを判断するために頻繁に他のプレイヤーへ視線を向ける。人間がジェスチャーを計算機プレイヤーに伝えたいときには、視線が自分を向いている間に行わなければならない。ジェスチャーを受け取って了解した計算機プレイヤーは、相手に対して確認の意味でうなずく動作を返す。

4. 分散環境での実装

提案したシステムを実現するには、大別して以下の6種類の処理が必要となる。

- (1) 計算機プレイヤーが演奏を理解・生成する処理（計算機プレイヤーの頭脳）
- (2) 演奏全体のテンポを管理する処理
- (3) 計算機プレイヤーの演奏を人間が聞こえるように音として出力する処理
- (4) 人間の演奏を計算機プレイヤーに伝えるために入力する処理
- (5) 計算機プレイヤーのCG画像を表示する処理（計算機プレイヤーの体全体）
- (6) カメラにより人間のジェスチャーを検出する処理（計算機プレイヤーの目）

以上の処理を単一プロセスで実現するのは、計算量が多いために現実的でない。そこで本研究では、これらの各処理を担当する複数のプロセスをサーバ・クライアント・モデルに基づいて実装し、演奏情報などをお互いに通信することでシステムを実現した。これは、

- 各プロセスはそれぞれに特化した処理に専念すればよい実装が容易になる。

^{*} カメラによるジェスチャー認識の結果に相当する情報を通信する。

^{**} まだ実現はされていないが、たとえば、人間のピアノソロを計算機プレイヤーが低調だと判断したときには、計算機の方から次のベースソロ演奏部へ移るジェスチャー動作をするように設計してもよい。

- 6章で考察するように拡張性が高く、提案したモデルを実現するうえで適している。

という利点がある。

現在の実装では、ネットワーク (Ethernet) に接続された3台のワークステーション (SGI Indigo2 Impact/Extreme) 上でシステムは動作する。演奏の入力と出力はMIDI (Musical Instrument Digital Interface) により行い、プロセス間の通信には次に述べるRMCP (Remote Music Control Protocol)^{15)~17)}を用いた。

4.1 RMCP

RMCPは、MIDIとLAN (Local Area Network) を融合した分散協調システムにおける音楽情報処理のためのネットワークプロトコルであり、シンボル化された音楽情報をネットワークを通じて共有するために設計された。下位レイヤーとしてUDP/IPを用いた、コネクションレス型のプロトコルである。RMCPでは、様々な音楽情報はRMCPパケットという単位で伝送され、分散した計算機上の複数のプロセスによって共有される。RMCPはマルチサーバ・マルチクライアント・モデルに基づいており、すべてのプロセスは、RMCPパケットのブロードキャスト (同報送信) だけを行うRMCPクライアントと、RMCPパケットを受信できるRMCPサーバに分類される。RMCPパケットはクライアントから全サーバへ片方向に送信され、サーバからクライアントへの返答は原則としてない。ブロードキャストを採用したことにより、演奏情報などをネットワーク上の様々なサーバで効率良く共有して同時に活用できる。さらに、サーバを単に追加するだけで新たな機能を実現できるため、システムの実装や拡張が容易になるという特長を持つ。

RMCPパケットは、ヘッダーとそれに続くメッセージ本体で構成される。ヘッダーには、メッセージの種類、サーバ番号、クライアント識別子 (ユーザ識別子)、タイムスタンプ情報、ヘッダー長、メッセージ長などが含まれている。本研究で用いるメッセージの種類には、以下の4種類がある。

- MIDI情報
演奏情報であるMIDIメッセージを伝送するために用いる。メッセージ本体がそのままMIDIメッセージとなっている。
- ビート情報
四分音符の時刻と現在のテンポを伝送して、テンポを同期させるために用いる。
- ジェスチャー情報
「体を左に傾けた」、「体を右に傾けた」、「手で頭を

表1 RMCPサーバ・クライアント一覧
Table 1 RMCP servers and clients.

処理内容	サーバ・クライアント名
1. 演奏理解生成	RMCP Player Server (サーバ)
2. テンポ管理	RMCP Beat Provider (クライアント)
3. 演奏音出力	RMCP Sound Server (サーバ)
4. 演奏入力	RMCP MIDI Receiver (クライアント)
5. CG出力	RMCP Animation Server (サーバ)
6. カメラ入力	RMCP Camera Analyzer (クライアント)

指した」というジェスチャー内容を伝送するために用いる。

- アニメーション情報

ジェスチャー動作指示、うなずく動作指示、視線方向指示を伝送して、CGを生成・制御するために用いる。

またヘッダー中のタイムスタンプ情報 (1ms単位の絶対時刻) により、サーバは、タイムスタンプの時刻以前に到着したパケットを、伝送中の遅延の揺らぎ等の影響を受けずに、その時刻どおりに処理することができる。ヘッダー中の他の項目は今回の実装では重要でないため、説明は文献16)に委ねて省略する。

前述の6種類の必要な処理を担当するRMCPサーバ・クライアントの一覧を表1に示す。これらのサーバ・クライアントの運用例について、4.2節、4.3節で順に説明する。

4.2 すべてのプレーヤーが対等なジャズセッションシステムの実装

すべてのプレーヤーが対等なジャズセッションシステムをRMCPにより実装した構成図を図4に示す。これは、表1の1~4で構成される。前述したRMCPパケットが伝送する4種類の情報のうち、ここではMIDI情報とビート情報が用いられる。人間や計算機のプレーヤーの演奏情報は、図の中央のネットワーク (Ethernet) 上にMIDI情報のRMCPパケットとしてブロードキャストされる。このパケットをすべてのサーバが受信することで、演奏情報を共有できる。これは、全プレーヤーがお互いの演奏を聞き合って共有することに相当する。

人間がピアニストとしてMIDI楽器の鍵盤を弾くと、そのMIDIメッセージはMIDI情報のRMCPパケットとして、RMCP MIDI Receiverによりネットワーク上にブロードキャストされる。ベーシストとドラマーを担当する2つのRMCP Player Serverは、このMIDI情報を受信しながら (演奏を聞きながら)、自分のMIDI情報をブロードキャストする (自分の楽器を演奏する)。この際、他方のRMCP Player ServerのMIDI情報も同様に受信して考慮する。これらの

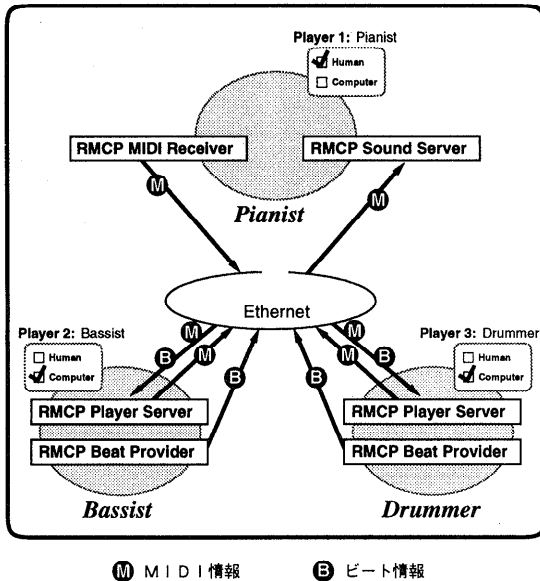


図4 RMCPによるジャズセッションシステムの構成図

Fig. 4 System configuration of our jazz session system on RMCP.

MIDI 情報を含む RMCP パケットを RMCP Sound Server が受信すると、パケット中の MIDI メッセージを MIDI 楽器に送ることにより実際の音として出力する。両 RMCP Player Server はテンポに合わせて演奏を理解・生成しなければならないが、これは RMCP Beat Provider が送信するビート情報に基づいて演奏生成することで実現する。

4.2.1 RMCP Beat Provider

RMCP Beat Provider は、各拍の位置（四分音符の時刻）を表すビート情報を順番に RMCP パケットとしてブロードキャストする。このパケットは、タイムスタンプとして四分音符の時刻を持ち、実際の時刻の 1 拍前までに送信される。RMCP Player Server は、この時刻に基づいて、他のプレイヤーの演奏を分析し自分の演奏音の時刻を決定する。

現在の実装では演奏のテンポは一定としたため、本サーバはつねに同一間隔のビート情報を送信する。ベーシストとドラマーのそれぞれに RMCP Beat Provider を用意してあるが、両者は演奏開始時に同期をとり、同じ内容のビート情報を送信する。

4.2.2 RMCP Player Server

ベーシスト用とドラマー用のそれぞれの RMCP Player Server は、他の 2 人のプレイヤーの MIDI 情報と自分用のビート情報の RMCP パケットを受信し、1 拍ごとにその内容を過去の MIDI 情報とともに分析する。そして、分析結果に基づいて次の MIDI 情報を

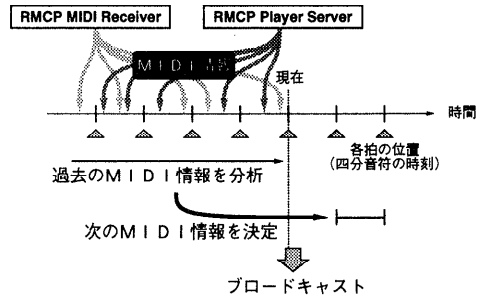


図5 RMCP Player Server による MIDI 情報の送受信状況
Fig. 5 Receiving and sending MIDI information by RMCP Player Server.

決定し、それを実際に鳴る時刻の 1 拍前までにブロードキャストする（図 5）。この際、各 RMCP パケットには、それが実際に処理されるべき時刻を表すタイムスタンプを付加する。

本サーバは大きく分けて、演奏理解部と演奏決定部、演奏生成部で構成される。演奏理解部では、各プレイヤーの演奏をその楽器の演奏特性に応じた方法で分析して、プレイヤーの主張度を求める。演奏決定部では、次の主張度などを予測しながら、自分がその後の演奏においてどれくらいの主導権を握るかを決定する。その際、シナリオ中のどの部分を現在演奏しているかも考慮する。最後の演奏生成部では、音程・音量・リズムそれぞれについてのパターンのデータベースから、決定した主導権を実現できるようなパターンを選出して組み合わせ、実際の MIDI 情報（MIDI メッセージ）を生成する。各部における具体的な処理は、文献 18) に詳しく述べられており、本論文の範囲を超えるためここでは議論しない。

4.3 VirJa Session の実装

姿の見える仮想ジャズセッションシステム VirJa Session を RMCP により実装した構成図を図 6 に示す。図 4 と比べると、2 つの RMCP Animation Server と 1 つの RMCP Camera Analyzer が追加されている。RMCP Camera Analyzer を 2 つ用意する実装も考えられるが、現時点では正面から人間のジェスチャーを検出すれば十分なので、1 つのビデオカメラを両計算機プレイヤーが共用している。以上の追加にとともに、RMCP パケットとして、MIDI 情報とビート情報に加えてジェスチャー情報とアニメーション情報も伝送される。ジェスチャー情報は、RMCP Camera Analyzer から両 RMCP Player Server へ送信されるか、両 RMCP Player Server 間で通信される。後者は、仮想的なカメラを通じて他方の計算機のジェスチャーを見ることに相当する。アニメーション情報は、

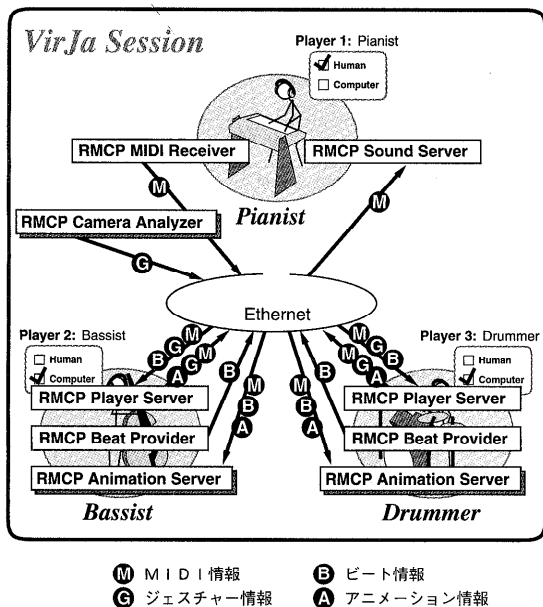


図6 RMCPによるVirJa Sessionの構成図

Fig. 6 System configuration of VirJa Session on RMCP.

プレイヤーの可視化用に演奏音以外の動作情報を伝達するために用いられ、RMCP Player Serverから対応するRMCP Animation Serverに送られる。これは、計算機プレイヤーの頭脳から体全体への動作指令に相当する。

VirJa Sessionのために、4.2節で述べたRMCP Player Serverを次の3点で拡張する。

- RMCP Camera Analyzerや他方のRMCP Player Serverが送信するジェスチャー情報を受信し、シナリオ中の各部の繰返し回数を動的に決定する。そして、確認のためのうなずく動作をアニメーション情報として送信する。
- ソロ演奏部やフォーバース部で最大繰返し回数に達したときに、次へ移るためのジェスチャー情報を送信するとともに、そのジェスチャーに相当するアニメーション情報も送信する。
- シナリオ中の各部において終わりに近付くと、頻繁に他のプレイヤーへ視線方向を切り替えるためのアニメーション情報を送信する。また、他のプレイヤーの主張度が高いときなどに視線を向けることも行う。

4.3.1 RMCP Animation Server

RMCP Animation Serverは、MIDI情報とアニメーション情報、ビート情報のRMCPパケットを受信し、その内容に応じてリアルタイムに変化するプレイヤーの姿を三次元CGで表示する(図7)。対応

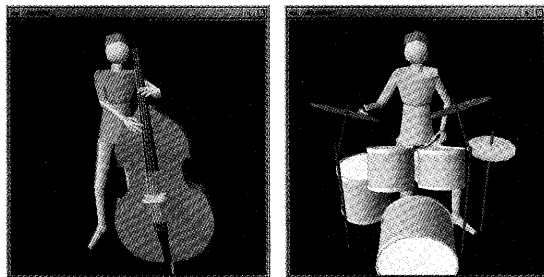


図7 RMCP Animation Serverによるベーシストとドラマーの画面表示例

Fig. 7 An example of computer graphics output of bassist and drummer by RMCP Animation Server.

するRMCP Player ServerからMIDI情報を受信すると、その音を自分の楽器で演奏する動作を生成する。アニメーション情報を受信すると、その指示に応じてジェスチャー動作やうなずく動作、視線を向ける動作を生成する。また、RMCP Beat Providerから受信したビート情報により、体を揺らしたり足踏みをしたりする動作を生成する。

本サーバは大きく分けて、プレイヤー共通のRMCP処理部と画像生成部、プレイヤー個別の動作決定部とモデリングデータからなる。RMCP処理部では、パケット中のタイムスタンプの時刻に実際の動作が起きるように、時間管理等を行う[☆]。動作決定部では、ベーシストの場合には、左手のポジションと右手で弾く弦を、前の音からの指の移動量などを考慮しながら決定する。ドラマーの場合には、足の動作はMIDI情報から一意に決定し、手の動作はある打楽器を左右どちらの手で叩くかという曖昧性があるため動的に決定する。本サーバの具体的な処理は、文献19)、20)に詳しく述べられている。

4.3.2 RMCP Camera Analyzer

RMCP Camera Analyzerは、演奏する人間の上半身を正面からビデオカメラで撮影して解析し、2種類のジェスチャーを検出する。そして、この結果得られるジェスチャー情報をブロードキャストする。

本サーバは、頭部位置推定部とその結果を用いるジェスチャー検出部からなる。頭部位置推定部では、まず事前に取り込んだ背景画像との差分を二値化し、x軸(水平軸)方向のヒストグラムを求める。次に、

[☆] ただし、現在の実装では、ソロを指示する相手の方向へ体を傾けるジェスチャーではなく、その相手の方向へ指やドラムスティックを向けるジェスチャーを行う。

^{☆☆} RMCP Player Serverによるパケットは1拍前までにブロードキャストされるため、手を振り上げるなどの予備動作を行ってから実際の発音タイミングで楽器を弾くといった動作が生成できる。

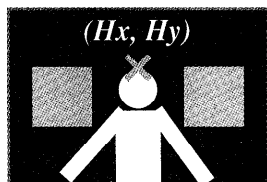


図8 RMCP Camera Analyzerによるジェスチャーの検出
Fig. 8 Gesture recognition by RMCP Camera Analyzer.

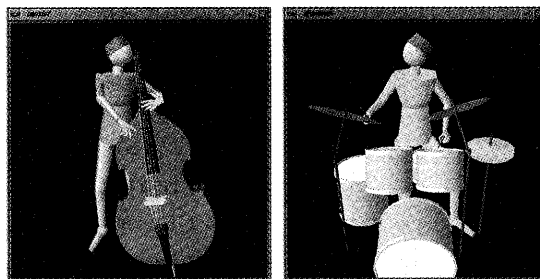
ヒストグラム中で最大値をとる x 座標を頭部の x 座標 (Hx) とし、その最大値自体を頭部の y 座標 (Hy) と推定する (図8)。ジェスチャー検出部では、 Hx が左右に大きく変動すれば、体をその方向へ傾けるジェスチャーをしたと見なす。また、(Hx, Hy) の座標があまり動かずに安定であり、しかもその左右の領域の二値化画像に大きな変化があれば、手で頭を指すジェスチャーをしたと見なす。

5. 実験結果

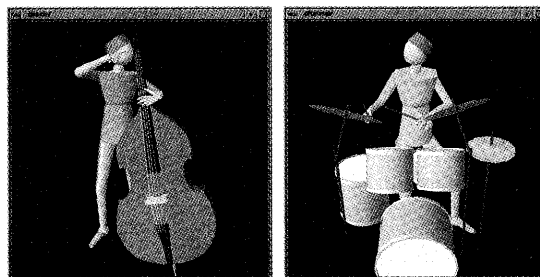
実装したシステムを運用して、提案したジャズセッションが実際に可能であることを確認する実験を行った。ピアニストの演奏用の鍵盤および音源にはシンセサイザー (Korg 01/W FD) を用い、この鍵盤は2台のワークステーション (SGI Indigo2 Impact/Extreme) の画面の前に設置された。それぞれの画面には、ピアニストから見て左側にベーシスト、右側にドラマーの姿が表示されており、3者が向かい合うように演奏できる配置とした。ピアニストの正面には、上半身の映像をRMCP Camera Analyzerが取り込めるようにビデオカメラが設置され、その出力は別のワークステーション (SGI Indigo2 Extreme) で処理された。実験中はこれら3台のワークステーションを占有使用し、Ethernetのトラフィックは低い状態に保っていた^{*}。

実験では、ジャズピアノ歴5年 (ピアノ歴21年) のピアニストが被験者となり、4ビートのジャズのスタンダード曲「Take the "A" Train」を即興演奏した。この曲の調やコード進行は事前にシステムに与えておいた。シナリオとしてはベースソロ演奏部やフォーバース部 (ピアノとドラムスのソロ) を含むものを設定し、曲のテンポは、187~230 M.M. (Mälzel's Metronome) の間で、様々なテンポを設定して実験した。

^{*} トラフィックの低い状態でのRMCPによる遅延時間は、平均値0.30msと非常に小さいことが確認されている¹⁶⁾。RMCP Player Serverによるバケットは1拍前までにブロードキャストされていることから、トラフィックの増大などにより数十msのネットワーク遅延が発生したとしても、RMCP Sound Serverによる演奏音出力には基本的に影響しない。



(a) ドラマーを見るベーシスト (b) ベーシストを見るドラマー



(c) テーマ演奏部に戻ることを指示するベーシストのジェスチャー動作 (d) ベーシストにソロ演奏部に移ることを指示するドラマーのジェスチャー動作

図9 実験中のベーシストとドラマーの画面表示例

Fig. 9 Examples of computer graphics output of bassist and drummer in our experiments.

シナリオ「テーマ演奏部 ⇒ ピアノソロ演奏部 ⇒ ベースソロ演奏部 ⇒ フォーバース部 ⇒ テーマ演奏部」を例にとり、ジェスチャーを用いたインタラクションによって、曲がどのように進行していったかを述べる。まず、人間同士のセッションのようにドラマーがスティックを用いてカウントをとり、ベーシストとドラマーとともにピアニストは演奏を開始した。演奏中を通じて、ベーシストとドラマーは図7のようにピアニストへ視線を向けるだけでなく、自分の楽器へ視線を向けたり、図9(a)と(b)のように他方の計算機プレイヤーへ視線を向けたりした。次に、ピアノソロ演奏部において、ピアニストがベースソロ演奏部に移るジェスチャー動作をしたところ、ベーシストはピアニストに向けてうなづく動作をし、次のコーラスの頭からソロ演奏を正しく開始した。場合によっては図9(d)のように、ドラマーがベースソロ演奏部に移るジェスチャー動作をした。その場合は、ベーシストがドラマーに向けてうなづく動作をしてソロ演奏を開始した。また、フォーバース部では、ピアニストが自分の望むところでテーマ演奏部に戻るジェスチャー動作をしたところ、ベーシストとドラマーがピアニストに向けてうなづく動作をし、次のコーラスの頭から全員でテーマ演奏に戻ることができた。このときも、場合によって

は図9(c)のように、ベーシストがテーマ演奏部に戻るジェスチャー動作をした。そして、ドラマーがベーシストに向けてうなづく動作をした後に、全員でテーマ演奏に戻った。

実験の結果、ピアニストが主導権を握るだけでなく、他のプレーヤーも主導権を握ってソロ演奏などを行うといった、すべてのプレーヤーがより対等なセッションを実現できた。そして、ベーシストの演奏に応じてドラマーの演奏の盛り上がり方も変わるなど、計算機プレーヤー同士のインタラクションが起きたことも確認された。たとえば、ベーシストが音高が高く音数が多いような主張度の大きい演奏をしたのに反応して、ドラマーがクラッシュシンバルやスネアドラムなどを多用した盛り上がる演奏を行った。

VirJa Sessionによる計算機プレーヤーの姿が見えるセッションを演奏音だけのセッションと比較すると、相手の姿が見えることで臨場感が増し、ピアニストは実際にライブ演奏をしているときより近い感覚を得られた。従来のジャズセッションシステムが演奏音だけを扱い、あたかも各プレーヤーが別々の部屋で演奏している状況であったことを考えれば、これは大きな前進である。また、ピアニストのジェスチャーが計算機プレーヤーによって適切に認識され、計算機プレーヤーもジェスチャーをピアニストに送るといった、シナリオが演奏中に動的に切り替わるセッションも実現できた。人間同士のセッションでは頻出するこうしたシナリオの動的な切替えは、ジャズの魅力の1つであり、それが人間と計算機との間で実現できたことの意義は大きい。

以上により、演奏音とジェスチャーを介したプレーヤー間のマルチモーダルインタラクションが達成でき、提案したジャズセッションを実現できたことが確認された。また、分散環境で実装した本システムが有効に機能し、ジャズセッションのための新しい演奏環境を提供できたことも示された。

6. 拡張性に関する考察

提案したモデルを分散環境で実装したことによって、どのように拡張性が高くなったかを具体的に考察する。

● 図1のモデルの実現

2.1節で述べたモデルにおいて、各プレーヤーを人間と計算機のいずれが担当してもよいことを述べた。たとえば、図2のモデルにおいてベーシストの担当を計算機から人間に変更するには、図4の計算機担当用(ベーシスト用)の2つのプロセス(RMCP Player Server, RMCP Beat Provider)を人間担当用の2つ

のプロセス(RMCP MIDI Receiver, RMCP Sound Server)に置き換えるだけでよい。ピアニストの担当を人間から計算機に変更する場合も、人間担当用を計算機担当用(ピアニスト用に新たに実装する必要がある)に逆に置き換えればよい。

一方、プレーヤーを新たに追加する場合には、人間担当用か計算機担当用のプロセスを追加し、各RMCP Player Serverの演奏理解部を新たなプレーヤーにも対応するように修正すればよい。ただし、現在のシステムでは、演奏環境としての枠組みは用意しているものの、実際に拡張するための個別の課題については解決していない。たとえば、一般に楽器の種類が変わると、その演奏の理解や生成の仕方は大きな変更をとまなう場合が多く、新たなプレーヤー用の演奏理解部、演奏生成部を実装する際には、様々な課題を解決する必要がある。また、最大プレーヤー数に関しては、MIDIのチャンネル数の制約に加えて、プレーヤー数の増加にともなう演奏理解部の計算量の増大という制約も考慮しなければならない。RMCPはブロードキャストを採用しているため、プレーヤー数の増加にともなうトラフィックの増大は(演奏生成量が同程度ならば)線形に近く、相対的に大きな制約とはならない。

以上の議論はVirJa Sessionの場合にも同様に適用できる。

● テンポ変化への対応

現在のシステムには、RMCP Beat Providerによる演奏テンポが一定であるという制約がある。テンポが変化する演奏に対応するよう拡張するには、他のプレーヤーの演奏のテンポ変化に追従できるようなRMCP Beat Trackerを実装し、RMCP Beat Providerと置き換えればよい。

● 複数箇所にいるプレーヤー間でのセッション

今回は1カ所で実験したが、RMCPによりネットワークを利用して実装したことで、複数箇所にいるプレーヤー間にも自然に拡張できる。この場合には、各プレーヤーのいるLAN間でRMCPパケットを中継する。

● 様々な個性を持つ計算機プレーヤー間でのセッション

RMCPをプラットフォームとして、様々な人が各自の思いを込めたRMCP Player Serverを実装すれば、それらのプレーヤーを接続していろいろな形態のセッションが行える。つまり、計算機上のプレーヤーを介した、それらの設計者同士の間接的なインタラクションも考えられる。この場合、RMCP Player Serverだけを選択・交換すれば、他のプロセスを変更する必要はない。

7. おわりに

本論文では、すべてのプレーヤーが対等なジャズセッションシステム、およびその発展形である、姿の見える仮想ジャズセッションシステム VirJa Session の、全体構想と RMCP による実装について述べた。本研究では、人間と計算機とのジャズセッションのための新たな統合演奏環境を提供するシステムを提案し、プレーヤー間に固定した主従関係のない即興演奏や、演奏音に加えてジェスチャーという視覚情報も介したマルチモーダルインタラクションを達成した。さらに、本システムを RMCP を用いて分散環境で実装したことにより、単一プロセスですべてを実現する場合に比べ、拡張性が高く実現できたことを示した。

本研究の持つ意義を以下にまとめる。

- すべてのプレーヤーが対等で、人間と計算機が入れ替わり可能な新たなジャズセッションのモデルを提案し、ピアノトリオを対象に実際にシステムとして構築した。これは、計算機とのジャズセッションをとらえ直す新たなモデルであり、6 章でも述べたような様々な発展が期待できる。
- 総合的な音楽情報処理システムを、異なる計算機上で動作する複数のプロセスで構成し、ネットワークを介して音楽情報を共有することで、効果的に負荷分散しながら拡張性高く実装する方法を示した。これは、特に本研究のように、個々の要素技術を組み合わせた統合環境を提供する際に効果的な実装方法である。
- ジャズセッションシステムの研究が今後進むべき、新たな方向性を示した。本研究では、人間同士が 1 カ所に集まってセッションをするとき、視覚情報も円滑にセッションを進めるうえで重要な役割を果たしていると考え、CG による可視化とジェスチャー認識を導入した。今後、人間と計算機との間で人間同士に匹敵するセッションを達成していくためには、本研究をさらに進め、視線一致の利用といった、人間同士のより緊密なマルチモーダルインタラクションを実現していくことが重要になる。
- ヒューマンインタフェースにおける擬人化エージェントの可能性を広げた。擬人化エージェントの研究では、音声対話にジェスチャー等を導入したマルチモーダルインタラクションが活発に研究されている²¹⁾。それに対して本研究では、音楽というノンバーバル情報を中心として、CG とジェスチャーを用いた新たなマルチモーダルインタラク

ションを達成した。人間がコミュニケーションに用いる情報は多岐にわたり、今後さらに多様なマルチモーダルインタラクションが実現されていくことが期待される。

本論文で提案したモデルと実装方法は様々な発展の可能性を持っている。そこで今後は 6 章で述べたような拡張を実際に行っていく予定である。ほかにも、画面上に仮想のライブステージを設置し、その中でピアニストの CG を含めた全員の姿を可視化することも予定している。

謝辞 VirJa Session の実験にご協力をいただき、貴重なコメントをいただいた緒方美音子氏、浅谷薫氏、大森哲也氏、黛慎一郎氏に感謝する。

参考文献

- 1) 和気早苗, 加藤博一, 才脇直樹, 井口征士: テンション・パラメータを用いた協調型自動伴奏システム: JASPER, 情報処理学会論文誌, Vol.35, No.7, pp.1469-1481 (1994).
- 2) 近藤欣也, 片寄晴弘, 井口征士: 音楽情報から奏者の意図を理解する伴奏システム—JASPER++、第 46 回情報処理学会全国大会論文集, 7Q-8 (1993).
- 3) 渡辺和之, 西嶋正子, 柿本正憲, 村上公一: ニューラルネットワークを用いたジャズセッションシステム—ニューロミュージシャン, 第 44 回情報処理学会全国大会論文集, 4R-5 (1992).
- 4) Dannenberg, R.B.: An On-Line Algorithm for Real-Time Accompaniment, *Proc. 1984 International Computer Music Conference*, pp.193-198 (1984).
- 5) Vercoe, B.: The Synthetic Performer in the Context of Live Performance, *Proc. 1984 International Computer Music Conference*, pp.199-200 (1984).
- 6) Dannenberg, R.B. and Mukaino, H.: New Techniques for Enhanced Quality of Computer Accompaniment, *Proc. 1988 International Computer Music Conference*, pp.243-249 (1988).
- 7) Baird, B., Blevins, D. and Zahler, N.: The Artificially Intelligent Computer Performer on the Macintosh II and a Pattern Matching Algorithm for Real-time Interactive Performance, *Proc. 1989 International Computer Music Conference*, pp.13-16 (1989).
- 8) 堀内靖雄, 藤井 敦, 田中穂積: 複数の人間と協調して演奏するシステム, コンピュータソフトウェア, Vol.12, No.5, pp.63-71 (1995).
- 9) 日高伊佐夫, 後藤真孝, 村岡洋一: ジャズの独奏の変化に対応する自動伴奏システム, 情報処理学

- 会研究報告, 音楽情報科学 95-MUS-9-2, Vol.95, No.19, pp.49-56 (1995).
- 10) Hidaka, I., Goto, M. and Muraoka, Y.: An Automatic Jazz Accompaniment System Reacting to Solo, *Proc. 1995 International Computer Music Conference*, pp.167-170 (1995).
 - 11) 金森 務, 片寄晴弘, 新美康永, 平井 宏, 井口 征士: ジャズセッションシステムのための音楽認識処理の一実現法, 情報処理学会論文誌, Vol.36, No.1, pp.139-152 (1995).
 - 12) Rowe, R.: *Interactive Music Systems*, The MIT Press (1993).
 - 13) Katayose, H., Kanamori, T., Kamei, K., Nagashima, Y., Sato, K., Inokuchi, S. and Simura, S.: Virtual Performer, *Proc. 1993 International Computer Music Conference*, pp.138-145 (1993).
 - 14) 亀井克之, 佐藤宏介, 片寄晴弘, 井口征士: 実写画像の編集と手の3次元モデルとによる人の動作アニメーションの生成, 情報処理学会論文誌, Vol.36, No.2, pp.374-382 (1995).
 - 15) 後藤真孝, 橋本裕司: MIDI 制御のための分散協調システム—遠隔地間の合奏を目指して, 情報処理学会研究報告, 音楽情報科学 93-MUS-4-1, Vol.93, No.109, pp.1-8 (1993).
 - 16) 後藤真孝, 根山 亮, 菊地淑晃, 村岡洋一: RMCP: Remote Media Control Protocol—時間管理機能の拡張と遅延を考慮した遠隔地間の合奏, 情報処理学会研究報告, 音楽情報科学 97-MUS-21-3, Vol.97, No.67, pp.13-20 (1997).
 - 17) Goto, M., Neyama, R. and Muraoka, Y.: RMCP: Remote Music Control Protocol - Design and Applications, *Proc. 1997 International Computer Music Conference*, pp.446-449 (1997).
 - 18) 日高伊佐夫, 後藤真孝, 村岡洋一: すべてのプレイヤーが対等なジャズセッションシステム II. ベーシストとドラマーの実現, 情報処理学会研究報告, 音楽情報科学 96-MUS-14-5, Vol.96, No.19, pp.29-36 (1996).
 - 19) 松本英明, 後藤真孝, 阿部哲也, 村岡洋一: VirStA System: 仮想ステージと仮想アクターによる分散CGアニメーションシステム III. ジャズセッションプレイヤーの実現, 第53回情報処理学会全国大会論文集, 1P-07 (1996).
 - 20) 松本英明, 後藤真孝, 村岡洋一: 音楽演奏情報による仮想演奏者のCGアニメーション生成, 情報処理学会研究報告, グラフィクスとCAD 98-CG-89-3, Vol.98, No.16, pp.11-16 (1998).
 - 21) 速水 悟, 竹澤寿幸: マルチモーダル情報統合システムの研究動向, 人工知能学会誌, Vol.13, No.2, pp.206-211 (1998).

(平成10年4月6日受付)

(平成10年12月7日採録)

推薦文

本論文は, 単に人間の演奏者に追従するのではなく, 人間あるいは他のシステムと対等に合奏(セッション)を行うシステム, VirJa Sessionの全体像を述べたものである。

研究の特長としては, 以下の点があげられる。まず音楽面においては, 人間の演奏者とコンピュータ演奏システムとの間で必要なテンポの同期処理を, MIDI情報というプラットフォーム上で実現したこと, さらにコンピュータにも必要に応じて演奏を自主的に進める機能を持たせたこと。人間とコンピュータの意志疎通のため, ジェスチャーなど, 演奏情報以外の情報を積極的に活用したこと(その結果, よりリアルな演奏環境に近づいている)。その実現のため, 演奏情報, 画像情報などを統合的に処理する総合音楽情報処理システムをLAN上の分散環境において構築したこと, さらにそのための基礎技術として, 実時間に音楽情報を交換する通信プロトコルRMCPを提案・実装したこと。

これらはいずれの点においても現在の研究水準から見てきわめて高度で斬新さが認められ, 特にRMCPに関してはその後の様々な研究で実装・応用され, 分野全体に対して大きな影響を及ぼしている。

以上の点から, 本論文は音楽情報科学分野を代表するものとして, 研究会推薦論文とするにふさわしい内容を持ったものとして強く推薦するものである。

(音楽情報科学研究会主査 平賀 謙)



後藤 真孝 (正会員)

1993年早稲田大学理工学部電子通信学科卒業。1998年同大学大学院博士後期課程修了。同年, 電子技術総合研究所に入所し, 現在に至る。博士(工学)。音楽情報処理, マルチモーダルインタラクション等に興味を持つ。1992年jvs設立10周年記念UNIX国際シンポジウム論文賞受賞。1993年NICOGRAPH'93CG教育シンポジウム最優秀賞受賞。1997年情報処理学会山下記念研究賞受賞。電子情報通信学会, 日本音楽知覚認知学会, ICMA各会員。

**日高伊佐夫**

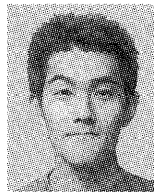
1994年早稲田大学理工学部電子通信学科卒業。1996年同大学大学院修士課程修了。同年、ソニー（株）アーキテクチャ研究所に入所し、現在に至る。音楽情報処理、認知科学等に

興味を持つ。

**松本 英明**

1996年早稲田大学理工学部情報学科卒業。1998年同大学大学院修士課程修了。同年、ソニー（株）に入社し、現在に至る。コンピュータグラフィックス、ユーザインタフェース

に興味を持つ。

**黒田 洋介**

1998年早稲田大学理工学部情報学科中退。同年、同大学大学院理工学研究科修士課程に入学。現在、情報科学を専攻中。TVゲーム「ピカチュウげんきでちゅう」およびWWW

検索エンジン「Ringring」の開発者。

**村岡 洋一（正会員）**

1965年早稲田大学理工学部電気通信学科卒業。1971年イリノイ大学電子計算機学科博士課程修了。Ph.D. この間、Illiac-IVプロジェクトで並列処理ソフトウェアの研究に従事。

同学科助手ののち、日本電信電話公社（現NTT）電気通信研究所に入所。1985年より早稲田大学理工学部教授。現在同大学メディアネットワークセンター所長。並列処理、マンマシンインタフェース等に興味を持つ。「コンピュータアーキテクチャ」（近代科学社）等著書多数。