



# SIGMUS 第9回 研究発表会 質疑記録

1995年2月23日 情報処理学会 会議室(芝浦)

(1) PSYCHE から:フレーズ表情の多角形,音楽構造に基づく表情付け,自動連弾システム

五十嵐滋, 彌富あかね, 小川大典, 清野桂子(筑波大)  
記録: 小坂直敏(NTT)

Q: 長嶋(LIST) フレーズ演奏の表現を多角形で試みているが,音楽はフレーズ間の関係でより大きな構造が成立している.例えば3次元の螺旋構造の表現による比較の方が適しているのではないか?

A: 清野 その通りである.

Q: 半径が速度変化成分を表している.その結果テンポの感じ方が速い人と遅い人とでは,相似形で面積が異なる多角形が得られる.フレージングの比較をするのであれば,半径を正規化するべきでは?

A: フレージングの比較については指摘の通りである.しかし,ここでは絶対的なテンポも明確に表せる方法をとった.

C: テンポという要因が全体にわたり存在し,それにフレーズ表現として局所テンポがある,とする捉え方の方が適するように思う.

A: 五十嵐 指摘されたことも含め試してみたいアイデアは数多くある.(もう少し検討が)進歩してから議論してほしい.

C: 田口のミュージズでは棒グラフによるフレーズの表現をとり,上位構造も扱っている.個々のフレーズでは多角形の方がわかりやすい点もあるため,それと本手法の両方の利点をとり,より上の階層のモデル化に期待している.

C: 今回の図形による表現形態と,音楽心理的評価の対応をつけてほしい.その結果は自動演奏などの音楽生成にも役にたつ.

Q: (音楽構造を反映した表情付けのルール)ここでのルールは,平均音高と音量の関係を設定しているが,同じ平均音高であっても,和声により音量が変わってくるのではないか?

A: 彌富 コード進行については早く採り入れたい.

A: 五十嵐 マズルカ7番では分析済みである.その結果現段階では和声を取り込まなくとも十分であることがわかった.しかし,いずれ必要となることは事実である.

C: MIDI ピアノは一律0.5秒遅れるわけではなく,音量により鳴り出す時の遅れが異なる.すなわち,ベロシティに対応して遅れを調整せねばならない.

A: (本研究で使用した)YAMAHAの製品の納入時の設定で,0.5秒あればあらゆる音量の音が鳴り出すことが分かっている.ここでは,その問題に対処せず,0.5秒の遅れののちに発音する,という考えで演奏ソフトを作成している.

Q: 平田(NTT)最後のデモについて,実時間で音量調整するシステムの技術的に難しい点は何か?

A: マチューズの電磁ドラムによる演奏にヒントを得て,まずマウスによる演奏制御を試みた.マウスでは連続的なテンポの変化の制御はできるが,急激なテンポの変化の制御ができない.マウスの代わりに鍵盤による制御を行なった.その結果,鍵盤の位置を制御情報とすることにより,連続的なテンポだけでなく,急激なテンポの変化,つまり離散的な制御が容易となった.音量制御も同様である.これが本質的な困難を克服した.ただし,鍵盤を用いての演奏情報の制御は,プロのピアニストからみると実際に演奏した方がいいので,煩わしい方法ではある.ここでのデモは,指揮者が指示した情報を素人が演奏している.

(2) ジャズの独奏の変化に対応する自動伴奏システム

日高伊佐夫, 後藤真孝, 村岡洋一(早大)

記録: 小坂直敏(NTT)

Q: 五十嵐(筑波大)クラシックではテンポのゆらぎが重要である.この意味でジャズとは補間的であろう.質問だが,最初の譜面に対して独奏者が新たに小節を付加することがあるのか,それともそれはないという前提なのか?

A: まったくなく,指定したコード進行により演奏される,という前提に立っている.

Q: 長嶋(LIST)事前に解析した小節の単位があるとして,システムは今どこを演奏している,という情報を(見出しして)使っているのか?

A: 内部では簡略した実装をしており,繰り返しに対して全てのデータを持っている.

C: 独奏者がジャズ演奏にストレスをもたずに演奏を継続するためには,伴奏側の先読みが必要不可欠である.盛り上がり予測する,またその終了など,実際のセッションでは必ず行なわれている.先読みのないシステムでは,盛り上がりを終了させるためには,独奏者は演奏をやめなければならない.研究と

しても数多くのエージェントの自己主張が強く、先読みを行なうシステムとして George Louis 報告事例などもあり、参考にされたい。

Q: 小坂 (NTT) 演奏者の意図は一つではなく、複数あり得、またどのような意図なのか本人も第三者も特定できない場合があると思うが、意図の検出はどうしているか?

A: 演奏者の意図は複数あるので同時に検出している。また、独奏者が一つの意図しかないときも、複数の意図を検出することがある。実験では、被験者に(意図的に)一つの意図を持って演奏してもらい、計算機がその意図を検出したか、によって判断している。

Q: 五十嵐 関連して質問する。1) 独奏者のどのような意図のときにシステムは誤った意図を抽出したか? 具体例を知りたい。2) また、(独奏者の意図が一意に特定できるというより、曖昧な側面がある、という議論を受け) 独奏者の意図をファジーに捉えていく、という考え方もできると思うがどうか?

A: 1) は、盛り上げる意図を誤って抽出する例があった。盛り上げる意図と緊張感を与える、あるいはその逆があった。2) はファジーで捉える考え方はとらない。明確に意図を検出する方法で行ないたい。

Q: つまり、アルゴリズムから統計までの範囲で行なうということか?

A: そうである。

Q: 柏野 (東大) 関連質問。5 個の意図を選んでいるが、これらの根拠は何か? その数でいいのか? また、これらは直交しているのか? これらの説明がないと、これらを選んだことの正当性が納得できない。

A: 経験から決めた。直交性の裏付けはとっていない。他の意図の可能性は考察中である。その一つとして、ジャズ特有のスイング感などを考えている。これには、テンポが一定という条件をとりさる必要がある。今後このような意図も検討していきたい。

Q: 先に話題となった「緊張感」と「盛り上げ」というのは比較的相関が強いように思う。

A: そうだ。

### (3) NeXT・IRCAM Signal Processing Workstation と Objective-C・MAX を利用した新たなコンピュータ音楽環境へのチャレンジ

片桐健順, 松田周, 菜孝之 (国立音大)  
記録: 柏野邦夫 (東大)

Q: 長嶋 (LIST) MAX で、1つのオブジェクトからいくつものオブジェクトにオーディオ出力を分配することが、画面上で線をつなぐことによってできるのか?

A: できる。

Q: 逆に、2つのオブジェクトの出力を合成して1つのオブジェクトに入力することは?

A: 信号を足し合わせる機能を持つオブジェクトを用いることによって可能である。

Q: NeXT は既に入手不可能な計算機であるが、システムを他の環境に移植することは考えられているのか?

A: MAX と ISPW をサポートする環境であることが前提である。将来的には SGI の計算機などに移植されるはずである。コンセプトの提案を行っていると理解されたい。

Q: MAX のオブジェクトの開発は工学的な話のように思われるが、音大でやることは妥当なのか。音大では、より上位の階層を扱うべきではないのか?

A: 音大に対する認識の違いと思う。

Q: 平田 (NTT) MAX のパッチが数多くたまった場合、その再利用性をどうやって試すのか?

A: 名前を入れることで対応したいが、現状の MAX ではサポートされていない。

Q: 後藤 (早大) ネットワーク越しの合奏において、バッファによる 6 秒間のずれが活用できるとの発言があったが、具体的にはどう活用するのか?

A: ミニマル・ミュージックのようなものも考えられるが、今後の課題。

Q: そのシステムの目的は、インターネットの利用なのか、それとも遠隔地で合奏を行うことなのか?

A: 現状で利用できる技術を用いて、遠隔地で合奏を行うことである。

Q: 用いたパケットは、MIDI データにタイムスタンプをつけたようなものと考えてよいのか?

A: よい。

Q: 学内で 10 ms という遅延時間はどのようにして測ったのか?

A: パケット伝送にかかる時間の概略の平均値である。

C: 菜 (国立音大) 芸術は何らかのメディアを用い、メディアの特徴・制約の中で表現を行う。今回の発表も、メディアの特徴・制約の中で何かをクリエートするというのが基本的スタンスである。

Q: 平賀 (図書館情報大) 手の画像による演奏は、明度情報を用いているのか?

A: RGB (各 24 bit) の情報のうち、赤の輝度値を用いている。

Q: 画像の変化の速さの情報は使っているのか?

A: 使うことはできるが、今回は多用していない。少しは用いている。

C: インターネットによる遅れは、実は面白いのではないかと思う。

Q: 小坂 (NTT) デジタルミキシングにおいて量子化歪みが懸念されるが、演算誤差に関して使用時の実感は何かあるか?

A: 使用感はアナログよりも格段に良い。ISPW の中では 32 bit 演算をしていると記憶している。

Q: 手の画像による演奏で、処理速度の点で、線の数は何本まで対応できるのか?

A: 直線の場合は、1 本あたり 6 ms の処理時間がかかる。このシステムはまだ出来たばかりなので、詳しくは今後実験していきたい。

Q: 平田 (NTT) 音大でプログラミング演習のような授業は行うのか?

A: C 言語入門を少しやる程度である。MAX は少し丁寧に教える。

#### (4) MIDI 音源を用いた並列プログラムの動作理解

風間一洋, 佐藤孝治 (NTT)  
記録: 柏野邦夫 (東大)

Q: 鈴木 (東京高専) 複数の音色の識別に関連して、渋谷・新宿などの JR 駅の発車ベルでは、隣のホームのベルと干渉しないように音のパターンが定められている。ピッチ、音色の他にも、識別に有効な手がかりがあるのではないか?

A: 音の並びのパターンは、現在は検討していないが、効果はあると思う。同時に発音しても識別できるような音を生成することも今後試みたい。

Q: 後藤 (早大) プロセッサ数が 64、さらに 1 万となると、提示する情報を抽出するフィルタが重要な問題となると思うが、どのようなアイデアがあるか?

A: フィルタが重要になるというのはその通りと思う。ゆっくり再生することでカバーできると思う。さらに、フォーカスの制御についても考えたい。

Q: 並列計算機の構成を考えると、通信のオーバーヘッドなどの点から、観測負荷によって計算機の状態が変わってしまうのでは?

A: どんな場合もプローブ・エフェクトは避けられない。その大きさが問題になるだろう。

Q: 今回の例題では視覚化の方が分かりやすいのではないかと感じたが、聴覚化のメリットが生きる例はどのようなものがあるのか?

A: mtcopy の例では分かるのだが、タイミング、テンポなどを用いて、関連する物事が関連している様子をうまく表現することができる。時間制約を意識すべきプログラムなどにも有効と思う。

Q: 平賀 (図書館情報大) 哲学者の問題では、飢える者が出るかどうかポイントである。従って、問題設定を逆にして、腹が減った者が音を出す (文句を言う) ようにした方が分かりやすいのではないか?

A: その通りと思う。

#### (5) ソフトウェアシンセサイザの開発 – FPD の開発 – 門田暁人 (奈良先端大), 黒田久泰 (京大), 藤井秀樹 記録: 松島俊明 (東邦大)

Q: 菜 (国立音大) リアルタイムで再生可能なのか?

A: そうだ。1 つの音色に対して 1 つの波形をメモリに読み込み、ピッチをリアルタイムで変化させている。

Q: 長嶋 (LIST) ハードウェアはサウンドブラスターみたいなものか?

A: PC9801-86 という PCM ボードだが、16 ビット PCM を再生する能力があれば何でも可。

Q: サンプリングもそのボードで行うのか?

A: そうだ。

Q: リアルタイムとは外部で演奏したものがリアルタイムで出力されるということか?

A: 既存の MIDI データを対象としている。一旦独自の演奏データへ変換を行うが、波形の合成結果をファイルに落さずに演奏するという意味でリアルタイム。

Q: 再生周波数はサンプリング周波数のことか?

A: サンプリングは 44.1kHz で行っている。再生周波数は 7 種類の中からユーザが選択可能。

Q: 内部演算は実数型?

A: 整数型で計算。音色波形データは 8 ビットで内部では 16 ビットで計算している。ただし、波形合成時の音割れ防止のため 1 音色あたり 14 ビットとしている。

C: 菜 (国立音大) DSP を使わず本体の CPU だけで行うという方向性は良い。今後注目される方向だと思う。

Q: 平田 (NTT) コルグの音源を使っているということだが、FPD 自体は Free なのか?

A: 音源データそのものではなく音色をサンプリングしたデータを使うということでした承を得た。FPD は WWW で入手可能。

Q: 坪井 (千葉職能短大) MML に拡張を行っているのか?

A: 基本的には従来のものと同じだが、多少拡張を行っている。

Q: 小坂 (NTT) 音程変更には 1 次補間を選んだ理由は?

A: 現実の CPU 速度でできる方法だから。

C: 長嶋 (LIST) ゲームでは昔からやられている方法。