

音楽情報科学研究会のページ

SIGMUS Home Page <http://www.etl.go.jp/~sigmus/>

第33回 音楽情報科学研究会のご案内

日時: 1999年12月11日(土)~12日(日)
会場: 早稲田大学 理工学部 (東京都新宿区大久保3-4-1)
* 毎年恒例のインターカレッジコンサートも同時開催致します。
問合せ先: 研究発表: 松島 俊明 matusima@is.sci.toho-u.ac.jp
コンサート: 菅野 由弘 ykanno@mn.waseda.ac.jp

第34回 音楽情報科学研究会のご案内

(情報処理学会 音楽情報処理研究会 (SIGMUS), 日本音響学会 音楽音響研究会 (MA 研), 電子情報通信学会 音声研究会 (SP) 共催研究会)
日時: 2000年2月17日(木)~18日(金)
会場: NTT 武蔵野研究センター
問合せ先: 堀内 靖雄 (千葉大)
e-mail: horyu@ics.tj.chiba-u.ac.jp (できるだけこちらへ), Tel: 043-290-3300

SIGMUS 第31回 研究発表会 報告・質疑記録

1999年8月7~8日 茨城県土浦市 国民宿舎「水郷」

(1) 楽曲分析システム DAPHNE — 実際の楽譜上での自動分析

劉剣利 (筑波大), 平賀瑠美 (筑波技術短大), 五十嵐滋, 関口由浩 (筑波大)

記録: 中村文隆 (神戸山手女子短大)

楽曲分析システム DAPHNE について, システムの構成とユーザーインターフェースについての説明と, 実際の使用例デモビデオが紹介された。ユーザーによる使用感の評価についても報告があった。

Q: 荒木 (豊田中研) デモでは, 異なる版の楽譜を表示するのにディスプレイを2台並べていたが, DAPHNE 自体にそういう機能があるのか。

A: 現バージョンの DAPHNE にはその機能がある訳ではなく, デモの作成には市販のソフトウェアを使った。

Q: 西原 (NTT) DAPHNE に実装されている, 構造分析の部分, モチーフ, フレーズ, センテンスの具体的な分析のアルゴリズムについては, リーマンの「大楽節構造」を勉強すれば判るのか。また, それ以外の部分, 例えば *anacursis*, *initiative*, *desinence* に関しては自動化はまだされていないということではないか。

A: 構造分析, モチーフ, フレーズ, センテンスの部分はリーマンの「大楽節構造」をベースに自動化している。また, 構造間の関係や類似度分析の部分も自動的にできている。構造機能分析については *initiative* や *anacursis* についてのルールがまだないので音楽家に聞いてやっている。この部分は人によって結果が違ったりする。

Q: 色々な大きさの丸を書いて演奏データの視覚化をしていたが, 元になるデータは演奏から自動的に抽出したのか, 人が聞きながら採譜したのか。

A: 演奏を MIDI 化したものから行った。

Q: 通常の音楽 CD からでもできるのか。

A: できない

C: 平賀(瑠) YAMAHA の Disklavier での演奏データを用いている。

Q: 松島 (東邦大) 分析結果などのデータをファイルに保存する際の形式はどういうものか。

A: 楽譜イメージはスキャナで取りこんだ BMP, 演奏データは MIDI, と独自開発の UNI という形式, 内部では EUROPA という独自の楽譜記述言語のデータで, 出力は DAPHNE 独自の形式である。

(2) パービーブーン: 誰でもどこでもインタラクティブに使える知的ジャズ和音生成システム

平田圭二 (NTT), 青柳龍也 (津田塾大)

記録: 中村文隆 (神戸山手女子短大)

システムの名前の由来, システム構築上の音楽的仮定とシステムの構成, 推論規則についての説明があり, Java で実装されたユーザーインターフェースと実際の使用例が提示された。

Q: 堀内 (千葉大) 階層性について, 和声進行に基づいて階層を組んで行くことが現実的にできるのか, また, 階層を規定してしまうことで表現力が却って落ちているということはないか。

A: 階層, すなわち, コード進行をどう解釈するか, ということはユーザーに任されている。問い合わせを作成する時にケードンツリーを作る訳だが, それはユーザーが与えられたコード進行をそのようなツリー構造として解釈する, ということを表している。

Q: そのツリーというのは何層にも深くなるものなのか。

A: 今サンプルとして使った *Autumn Leaves* などの曲では6段か7段, といったところである。

Q: コンテキストツリーでは, コード名が継承されるのか。

A: コード名が継承される訳ではない。

Q: Im7 といった表記を継承して見えているように見えたが,

A: 例えば BbM7 → GM7 といったコード進行であれば, それ

を代表するコードを導出していき、これを繰り返して行く。1曲全部に対してこの操作が終われば1曲を代表するコードが得られる。それは大体の場合トニックである。

Q: つまり、そういったイメージで階層を作る、ということか。
A: そうだ。コード進行の階層イメージを作るのがユーザーの解釈であり、ケーデンスツリーを上手く編集することが各ユーザーの好みの演奏を作るということにつながる。そのためユーザーインターフェースを作成した。

Q: 大矢(長野高専)演奏事例ベースのソースは何か。
A: Autumn leaves はハービー・ハンコック、All the things you are は自分、Misty は Jazz Life 誌の無名のアレンジャー。

Q: 岸田(立命館大)階層構造の意味は? 上とか下、という意味はあるのか。
A: そうではない。例えば、Autumn Leaves の最初の8小節を例に取ると、このコード進行をどう解釈するか、ということである。例えば、最初の Cm7→F7→BbM7 で1つケーデンスを作り、これを BbM7 が代表している、と解釈する。こうしてまとめられた BbM7 のグループは、次の EbM7 と一緒にまとめられて IM7→IVM7(ここでは BbM7→EbM7) という進行となり、IM7(key Bb) としてまとめる。同じように後半4小節は GM7 で代表されるようにまとめることができ、こちらも IM7(key G) としてまとめられる。結局、前半は BbM7、後半は GM7 で代表され、この8小節は IM7(key Bb)→IM7(key G)(コードでは BbM7→GM7)、という進行としてグルーピングされる。

Q: そうすると、同じ IM7 と書かれているものも場所によって違うのか。
A: そうである。

(3) 主旋律の特徴を生かした不干渉な対旋律の生成手法
伊藤響, ○小澤玲子, 犬塚信博, 伊藤 英則(名工大)
記録: 中村文隆(神戸山手女子短大)
主旋律からリズム補填や模倣を用いて対旋律を生成する手法についての説明と結果の評価法、実際の生成例について報告された。

Q: 長嶋(ASL)予稿集図7について。発表したルールからは2小節目の最後の D|| (F→F→D||) がなぜ出てきたのかわからない。
A: 単純に2小節目前半の D→D→C の進行をそのまま平行移動したもの。

Q: 聞いた感じはどうか。
A: よくないと思う。

C: 私はよいと思う。対旋律というのも1つのメロディである。生成された対旋律自身もメロディとしてよくなければならないと思う。不自然な跳躍があるのはよくないのではないか。そういう評価基準を入れるとすれば、この D|| が次に E に半音上がるならよいと思うが、G に上がってしまったのでメロディとしてよくない。次はそういった基準を取り入れて欲しい。

A: わかりました。
Q: 平田(NTT) このシステムの中でさいころを振っている部分はありますか。
A: ない。具体的にはどういったことか。
Q: 例えば、同じ主旋律なら毎回同じ結果になるのか。
A: そうである。

Q: すばらしい。対旋律の対旋律が主旋律に戻るようなメロディはあるか。
A: 試したことがないのでわからない。
Q: 平野 曲の主観評価とは人間が聞いた評価のことか。
A: そうである。

Q: 人によってかなり違うことはありうるのか。
A: ありうる。

Q: 有意差の検定のような統計的な評価はまだやっていないのか。
A: まだだが、いずれやりたいと思う。

Q: 不干渉度と主観評価に相関があるかどうか。例えば、不干渉度が73%で主観評価が×のものもあれば46%で×のものもある。ある程度この2つに相関がある、と考えている理由は何か。
A: 主観評価には旋律の独立性なども入れているのでその辺が影響していると思う。

Q: 例えば曲番号6で不干渉度が73%でも主観評価が悪い、という事例があるが。
A: これはカウンターラインを生成した時点で音高が常識の範囲を越えてしまい、聞いた感じがよくなかったため失敗した。

(4) 物理モデルによるピアノ音とギター音のモーフィングの検討
引地孝文, 小坂直敏(NTT)
記録: 南高純一(カシオ)
物理モデルによるピアノ、ギター音とそれらの滑らかなモーフィング音についての発表とデモ。

Q: 西原(NTT) 多次元尺度法とはどのようなものか?
A: 対比較試験により各サンプル間の主観的距離を得て、この距離から各サンプルの配置を復元する逆問題を解く手法である。

C: 荒木(豊田中研) モーフィング途中の音の中にギターハーモニクスのように聞こえるものがあった。
Q: 後藤(電研研) スペクトルセントロイドに着目した理由は。
A: 今回のモデル(物理モデル)では妥当。

Q: 平田(NTT) アタック部と定常部にわけたらどうか。
A: 分けて比較した例が過去にあるが、両者同じような結果だった。

Q: 大矢(長野高専) 各モデルにおける数式は、どれが既知でどれがオリジナルか。
A: どの数式も割と一般的なものだが、共鳴系を追加して改善し、さらに励起モデルにおいて打弦と撥弦の両方を表現できるように改善した。

Q: 中村(神戸山手女子短大) 応用を考えると持続音の方がよいのではないかと思われるが。
A: これから。

Q: 青野(NTT) パラメーターは、アタック部、持続部など、どの辺にきくのか。
A: 共鳴のほう。

(5) Windows 上の音合成システム「おっきんしゃい」
小坂直敏(NTT)
記録: 南高純一(カシオ)
Windows 上の波形エディタ。正弦波合成モデル、モーフィングなどを、直感的に操作できるようにデモ。

Q: 岸田(立命館大) 最近のポップスでよくあるモーフィングとの違いは?
*ポップスのモーフィング: シェールの「ビリーブ」など。
C: 長嶋(ASL) Kyma は、リアルタイムでのモーフィングができる。

(6) MIDI 音源の発音遅延と音源アルゴリズムに関する検討
長嶋洋一(ASL)
記録: 南高純一(カシオ)
ハード、ソフト音源について、発音遅延とばらつき計測結果と音源アルゴリズムに関する検討。計測結果は研究に使う音源を選ぶときに有効。

Q: 平賀(蔵) (図書館情報大) ソフトシンセで遅延が生じない理由は?
A: ベロシティ値の小さいものは捨てていると推測される。

(7) 3D 音響ブラウザの実装と駅発車音への適用
増井誠生(富士通研)
記録: 南高純一(カシオ)

画像と音響の3D空間への配置。実際にブラウジングアメモを示してくれた。

(8) 招待講演: 「コンピュータ音楽とユーザー・インターフェイス」
久保田晃弘 (多摩美大)

記録: 上田健太郎 (阪大)

Q: 後藤 (電総研) アフォーダンスの話やイメージスキーマの話では視覚的なものや身体的なものが多いと思うが、音に関する議論としてはどのようなものがあるのか?

A: 音に対する議論はほとんどなされていないと思う。というのも、ギブソンの仕事は視覚だけであったし、後の認知理論、つまりは言語、音響や音といった事の議論はほとんどされていなくて、音とインタフェースというのは「こんなものかな」といったレベルでしか使われていない。そう言った観点からは音と3Dとか、音によるフィードバック、あるいは音をどういう風に使ってあげればいいのかというのはこれからきちっと議論していかなければならない問題である。先ほどのギブソンにしても、デザイン、インタフェースの重要な2つの要件として挙げたのは、1つはわかりやすい概念を用いることと、もう1つは視覚化すること(ビジュアルイゼイション)である。

多くの人が例えば、アフォーダンスとって使っていることは、単なるミスユーズであって、ビジュアルフィードバックしか考えていないので、音の問題も専門家の意見をふまえた上で考えるべき問題である。

(9) アゴークルールの適用による自動演奏生成システムの構築

小池宏幸 (筑波大), 平賀瑠美 (筑波技術短大), 五十嵐滋, 水谷哲也, 塩雅之 (筑波大)

記録: 青野裕司 (NTT)

Q: 長嶋 (ASL) 「じだいに」という演奏を表現する際に01から04のどれをどのように選択するのか。

A: 人間がヒューリスティックに行っている。

C: これらの関数の形状はあまりよくないので、指数関数や三角関数などを適用してみようか。

C: 問題が複雑になる可能性もあるが、アゴークの関数を決定するときには、ダイナミックからの影響というものを考慮に入れるべきではないか。そのためには、アゴークとダイナミックの両方を扱うべきではないか。

Q: 平賀(讀) (図書館情報大) トリルの生成はどのようにしているのか。

A: 元の演奏データを用いている。

(10) 重回帰分析を用いた演奏の表情付け

石川修 (阪大), 片倉晴弘 (LIST), 井口征士 (阪大)

記録: 青野裕司 (NTT)

Q: 長嶋 (ASL) 発表にあった演奏は、人間の演奏データを元に、システムが自動的に表情付けを行ったものか。

A: 説明変数の付与や、その際に必要となる構造解析などにおいては、手動で行っている。

Q: 西原 (NTT) 教師部分と同じ部分を自動生成しているのか。

A: 曲の前半部分を教師曲として用い、後半部分を未知曲としている。

Q: 澤田 (香川大) 未知曲の構造解析の方法は?

A: 前半部分の教師曲と一緒にまとめて、人手で行っている。

Q: 将来的にはどのような形になるのか。

A: 構造解析部分まで自動化することが目標である。

C: 長嶋 (ASL) 筑波大の手法のように関数によるものは、演奏に違和感を感じる。一方この手法は、構造解析まで自動化できれば、非常によいものであると思う。表情付け研究においては、GTTM (Generative Theory of Tonal Music) を学ぶことが欠かせない。

Q: 平田 (NTT) 重回帰分析を行うに当たってグループ分けをしているが、それでは情報量が少なすぎるのが懸念される。

A: 現状では問題ないと思う。

Q: 澤田 (香川大) グループ間のつなぎ目の部分は大丈夫か。

A: システム内部で補正している。

C: グループ分けを行った方がむしろ、データを大量に用意するよりもフィッティングしやすいように感じるが、そのときにグループのつなぎ目が不連続になるように思われる。

A: つなぎ目の問題は確かにある。

Q: 出口 (攻玉社工科大) 教師曲と適用曲が似ているが、他の曲には適用できないのか。

A: やってはみたが、あまりお聞かせできない。

(11) 自動表情付システム MorPH

真栄城哲也, 樋口直史, 中口孝雄 (ATR)

記録: 青野裕司 (NTT)

(12) 自動表情付システム MorPH の実装

中口孝雄, 樋口直史, 真栄城哲也 (ATR)

記録: 青野裕司 (NTT)

備考: (11)(12) はまとめて発表されたので質疑も混在している。

Q: 上田 (阪大) 曲の構造を解析して得られたフレーズの情報はどのように利用しているのか。

A: サクソフォンの演奏の表情付けで用いている。

Q: 他の楽器には利用していないのか。

A: ギターのソロ部分などにも用いている。

Q: フレーズの解析は時間的な要素をもとに行っているようだが、コード進行などハーモニー的なものは考えなくてもよいのか。

A: 現状で、良好な結果が得られているので、考慮に入れていない。今後扱っていく。

Q: サクソの表情付けはバリエーションに欠けるのではないか。

A: 表情付けのルールを増やしていく。ユーザーが入力できる形にしたい。

Q: 長嶋 (ASL) ATR のホームページではアクセスしたユーザーの曲データが自動的に保存されていることを明記しているのか。著作権に絡む問題が生じるのでは。

A: オープンにした場合は問題になるだろうが、現在は外には出さないようにしている。

Q: ギターのコードフォームの解析を行っているが、曲中の和音が必ずしもギターを意図してボイスンされているとは限らないのではないか。

A: そのような場合のために、DP マッチングを用いている。また、ユーザーはあるトラックをギターとして処理するかどうかが選択することができる。ギターの種類の判別は、プログラムナンバーをもとに行っている。

Q: MIDI のドライバには何を使用しているのか。

A: ローランドのシリアルを使用している。

Q: 青野 (NTT) すでに収集した数曲の中で、ドラムパターンが入っていない曲があったか。

A: どの曲にも入っている。

C: ならばドラムパターンの自動生成は必要ないのでは。

Q: 入力される MIDI データは、ノートオンやデュレーションに関してクオンタイズされている必要があるのか。

A: 一拍 480 の分解能で±20 で補正を行っているため、クオンタイズの必要はない。

Q: 金森 (阪大) ドラムの演奏パターンは、手数が多すぎて人間離れしているのではないか。

A: 演奏生成においては、両手両足の制約を入れてある。相当にうまく演奏家と思っていただきたい。

Q: 彌富 (ヤマハ) 曲を収集する目的は。

A: 別途研究を行っている、編曲システムのテストデータを用意するため。

(13) 人間の演奏制御モデルの推定 (2) ~ テンポ変化時における人間のふるまい ~ 堀内靖雄 (千葉大)

記録: 園田智也 (早大)

Q: 長嶋 (ASL) ピアノよりも、打楽器系の演奏者とかでタッピングでやってみるとより傾向がわかるのではないかな。

A: 打楽器の場合、ピアノに比べて被験者を集めるのが困難であり、また、伴奏システム実現を考慮に入れた場合、ピアニストでデータを集めておいた方が効果的であると考えられる。

Q: 平田 (NTT) テンポを10とか15とかに変更させると、変化はあるのか。

A: テンポ変化が大きいときには顕著であるが、テンポ変化が5の場合はテンポ安定時とほぼ同様の傾向を示した。また、被験者によって若干の差があった。

Q: 引地 (NTT) 30 ミリ秒を閾値としているのは、なぜか。

A: 本研究では経験的に30 ミリ秒としたが、値を多少変えても大きな変化は見られなかった。

Q: 西原 (NTT) 同じピアニストでも、伴奏する人、ソリストに違いがあり、注意が必要ではないか。

A: ほとんどの音大生は伴奏を日常的に行っているので、とくに意識する必要はないと考えた。

Q: ある部分は、ソリストに追従、ある部分は、そうでないときがあるのではないかな。

A: 状況に応じて追従方法を変更する機能は、既に実装した。

Q: 志村 (大阪芸大) 必ずしも、ソリストに較べ伴奏者の能力が低い、というわけではない。伴奏がしっかりできる人を探すほうが難しい場合もある。

A: レベルの高い被験者を探すのは難しいので、普通の音大生レベルで実験を行っている。被験者探しに協力していただけるとうれしい。

(14) 古い録音メディアからの演奏情報の抽出

下条敬洋 (阪大), 片寄晴弘 (LIST), 井口征士 (阪大)

記録: 園田智也 (早大)

C: 長嶋 (ASL) 原音響、ノイズ除去音響、SMF 演奏データだけでなく、SMF データを作成するのに使用した表情付けのための感性情報データも加えてはどうか。今はむきピアニストの音を聞いたり、そこに自分の解釈を加味した演奏も作れる。

Q: 西原 (NTT) 音量の大小によって、アタック判定の閾値に影響があるのではないかな。

A: 手法を摸索している段階である。

Q: 後藤 (電総研) ピアノは対象楽器として最も簡単な部類だが、提案手法はそれに特化しているのか。

A: そうだが、最終的には他の楽器もカバーしたい。

Q: デモの曲は簡単だったが、和音を含む曲や早い曲には対応可能か。

A: 早い曲なども実験し、よい結果を得ている。

C: 楽譜を用いて分析するアプローチとして、1998年に出版された本「Computational Auditory Scene Analysis」中のEric Scheirerの論文 (IJCAI-95 Workshop on CASA で発表) や、日本音響学会1998年春季研究発表会での柏野らの発表も参考にするとよい。

Q: 平賀(瑞) 楽譜の版による違い、さらには演奏者の違い (ベダルなど) があるのではないかな。

A: 記述されている曲情報は用いたが、版の違いなどは考慮していない。

Q: 川本 (埼玉大) MIDI になったときの、ベロシティが下の音が上の音と比較して、強調されていたのではないかな。

A: 音響信号レベルの比較・修正で、まだ十分に処理ができていない恐れがある。

Q: 志村 (大阪芸大) このシステムによって音楽データベースが構築されることは、音楽研究者にとって非常に有意義である。そこで、楽譜と違う音が演奏家によって弾かれた場合はどう処理されるのか。また、MIDI で再現演奏する意味は

どこにあるのが問題になると思う。

A: 楽譜と違うものは、無視している。

(15) 箏曲異種楽譜の比較による歌の旋律分析

出口幸子 (攻玉社工科大), 白井克彦, 小原啓義 (早大)

記録: 園田智也 (早大)

Q: 長嶋 (ASL) 箏曲のピッチは MIDI ベースの12音で表現できるのか。

A: 現在は、平均律で教育している。おそらく、昔はそうではなかったと思われる。

Q: バリエーションを取り去った、元々の旋律を再現するのが、目的か。

A: 家元があれば、それぞれの歌い方がある。しかし、7割くらいは同じである。目的として、オリジナルの旋律を復元してみたいというのが動機にある。

Q: 研究として、秘伝の旋律を公開するのは、流派から破門されないのか。

A: 純粋な研究目的として、行っている。しかし、演奏データは残したい。

(16) 実世界の音楽音響信号を対象としたメロディーとベースの音高推定

後藤真孝 (電総研)

記録: 園田智也 (早大)

Q: 平賀(瑞) (図書館情報大) 音モデルはどういうものか。単一か。

A: ガウス分布を高調波構造上に並べたモデルで、現段階では単一。ただし音高推定の目的では、音とモデルとが厳密に一致しなくても有効に機能している。

Q: 個別楽器・音声の高調波構造を使えば、性能が向上するか。

A: 原稿中に今後「音記憶を導入」と書いたように、特化した音モデルを導入すれば性能向上の可能性はある。提案したのは特定の音モデルに依存しない一般的な枠組みなので、拡張の余地は多い。

Q: メロディーが途切れるところで伴奏パートを追跡してしまうのは、パワー差を考慮することで防げるのでは。

A: パワー差に依存すると他の問題が生じる。本質的には、音源同定して解決すべきだと考えている。

Q: 子音は落ちているはずなのに、歌詞がなんとなくわかるような気がするのだが。

A: 子音は落ちているが、子音が聞こえる気がするのは、子音前後のフォルマント形状などが音韻情報にある程度伝えているためだろう。

Q: 長嶋 (ASL) ジャズベースの早いパッセージに追従していたが、メロディー・ベースと判定する「持続した」という時間はどのくらいのオーダか。

A: 数十ミリ秒。

Q: 園田 (早大) EM アルゴリズムの収束判定は、

A: リアルタイム性を重視し、一定回数で打ち切るが、通常ほぼ収束している。

Q: (質問者不明) 音出力の方法は、

A: あくまで基本周波数推定が主眼で音出力は単なる確認用だが、現実装では、推定した高調波構造上のパワーに基づいて、正弦波重量モデルで合成している。

Q: 青野 (NTT) なぜリアルタイムでいきなり実装したのか。

A: リアルタイムに組むのが好きだから (笑)。インタラクティブな応用のためにはその方がよく、非リアルタイムシステムを後からリアルタイムにするのは二度手間だから。リアルタイム実装が容易な RACP のプラットフォームを構築後に着手したため、実装は比較的容易だった。

Q: 評価方法はどうか。

A: 難しい問題だが、今考えているのは、メロディーラインとベースラインを手作業でラベリングして、それと定量的に比較する方法である。