

アコースティックピアノを用いた セッションシステムの開発

青野 裕司[†] 片寄 晴弘^{**} 井口 征士[†]

[†]大阪大学基礎工学部システム工学科

^{**}(財) イメージ情報科学研究所

E-mail: aono@inolab.sys.es.osaka-u.ac.jp

近年のコンピュータの性能向上, 利用者層の広がりをはじめましいものがある。それを受けて, ヒューマンインタフェースに代表される, コンピュータと人間の接点を主題とした研究が盛んである。中でも, コンピュータと人間のインタラクションには, 常に高い関心が寄せられている。音楽の演奏, 特にセッションはインタラクションの重要性が高く, 研究の素材として多く利用されている。本研究は, 筆者が以前より行っている作曲支援機能を持つセッションシステムのインタフェースに, アコースティックピアノを導入する試みである。アコースティックピアノを用いるためには, その音響信号から音階, 和音の情報を導き出す必要がある。本研究で用いた, 倍音構成を利用した音解析の手法は, 特に和音名の判別に良好な結果を示し, 実時間処理にも適していることが分かった。

Development of a Session System with an Acoustic Piano

Yushi Aono[†] Haruhiro Katayose^{**} Seiji Inokuchi[†]

[†]Fac. of Engineering Science, Osaka University

^{**}Laboratories of Image Information Science and Technology

E-mail: aono@inolab.sys.es.osaka-u.ac.jp

In recent years, the progress of computers' performance is remarkable. The users of computers spread widely. Many researches such as "human interface" consider the contact of human with computers. These researchers are interested in the interaction between human and computers. Playing music, especially session, is good motif for these researches because the interactions are very important in it. This research tries to use an acoustic piano as an interface for the session system we are continuously developing. To use an acoustic piano, we need to know note names and chord names by analyzing the acoustic signals. The techniques of note analysis we use in this research show good result for chord-name extraction. We find these are suit for real-time processing.

1. はじめに

近年、音楽情報科学の分野において、人間とコンピュータのインタラクションに注目したセッションシステムの研究が盛んである^{[1][2][3][4]}。この背景には、コンピュータの性能のめざましい向上と、普及の広まりがある。利用者層の拡大はユーザビリティの高いコンピュータを求め、またコンピュータもそれらを満たす性能をもちはじめている。ユーザビリティを左右する要因として、インタフェースのインタラクティブ性があげられる。一方、音楽の演奏、特にセッションは演奏者間のインタラクションが演奏生成に必要不可欠である。人間とコンピュータの間でいかにインタラクションを保つか、インタラクションがユーザにとってどのような影響を与えているかといった研究や興味に対して、セッションは格好のモデルといえる。

過去のセッションシステムのインタフェースは、そのほとんどがMIDIと呼ばれる標準規格を持つ電子楽器である。MIDI楽器はコンピュータと接続することを前提として開発された楽器であり、音程・音量などの演奏情報をコンピュータに正確に伝達することができる。しかしそれらの音質や演奏の感触などを、アコースティック楽器（非電子楽器）と比較すると、確実に近づいてはいるが、いぜん差異があり、それらを別種の楽器と考えて使い分けているミュージシャンも多い。本研究は、我々が研究開発を行ってきたセッションシステムの入力楽器として、アコースティック楽器、特にアコースティックピアノを用いる。この試みはセッションシステム研究の新たな展開であると同時に、音響工学的にも興味もたれる。アコースティック楽器を用いるためには、それらの音響信号から音程、音量を抽出する必要がある。楽器の音程抽出の研究は、音響工学の分野を中心に行われてきたが、単旋律を対象としたものは実時間での処理が^[5]、採譜システムのように和音を含む場合はオフラインでの処理^[6]が可能になっている。セッションシステムにおいては、コードに代表される複合音をリアルタイムに知る必要があるため、従来の方法を用いることはできない。本研究では、セッションに用いるという前提のもとに、その際に必要となる情報を中心に解析抽出するという制約を設け、複合音の実時間解析を行った。具体的には、コード名、ベースノートと思われる低い周波数の音の音階、リズムパターンなどは、セッション

において重要である。一方、正確な発音時間や音長など、アーティキュレーションに含まれる情報の解析は行っていない。

2. セッションシステム概要

システムの概要を図1に示す。まず、ピアノ音響に対してFFTを行う事により、音響信号に含まれる周波数スペクトルを抽出する。その際、ピアノ等、楽器の特性に注目した抽出手法を考案した。この手法によって、周波数分解能と処理速度を共に向上することが可能になる。詳しい内容については後述する。

次に、FFTの結果を、Subharmonic Summation (SHS)と呼ばれる倍音構造を利用した解析によって、複合音に含まれる音階の候補を選ぶ。それらの候補に対して、トニック・ドミナントペア検索を行い、コードを決定する。また過去のコードの流れから次のコードを予測し、その結果をフィードバックする事により、コード決定の精度を上げる。

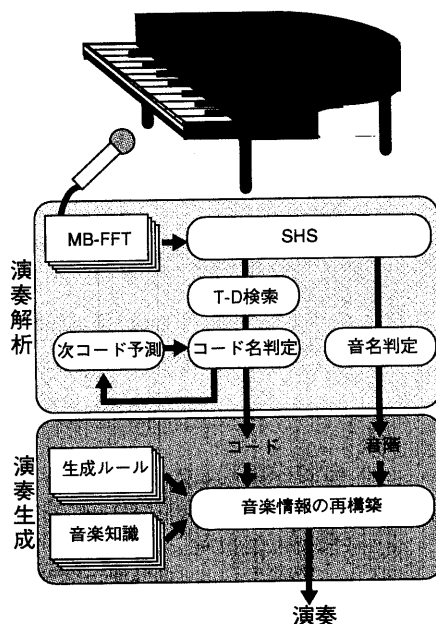


図1 システム概要

3. 演奏解析部

3.1 Multi Band FFT (MB-FFT)

図1に示される処理の中でも、スペクトル解析は後に続く全ての処理に影響を与えるという意味で非常に重要である。ピアノという楽器は他の楽器に比べて音域が広くまた倍音構成も複雑である。よってスペクトル解析には高い周波数分解能と、広い周波数帯域が同時に要求される。具体的には最も低い音程の周波数は40Hz弱、最も高い音程は4KHzであるが、さらにその倍音成分を考慮に入れると10KHzを超える。また最も低い半音階の差は、約2Hzである。10KHzを超えるバンド幅に対し、2Hz以下の分解能をえるためには、FFTでは実時間処理が難しくなる。そこで本研究では、Multi Band FFTという手法を提案する。これは異なる帯域を受け持つ複数のFFTを並列で行うもので、各帯域によってサンプリング周波数が異なるので、同じ1024点のFFTでも得られる周波数分解能が異なる(図2参照)。楽器音は周波数が高くなるにつれて半音階の差も大きくなり、均一な周波数分解能を必要としないことに着目している。この方法により、最低音部では0.5Hzの周波数分解能と、10KHz強の周波数帯域を同時に獲得することが可能となった(図3参照)。MB-FFTにおいて、もっとも重要なことは、低い周波数帯域を受け持つFFTのエイリアシングを防ぐことである。本研究では、ハードウェア的に構成された10次のローパスフィルタを用いて、これに対処している。

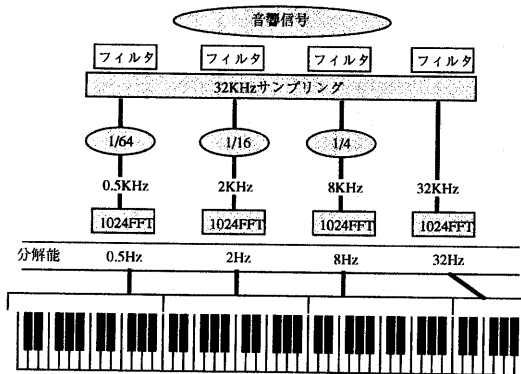


図2 MB-FFT

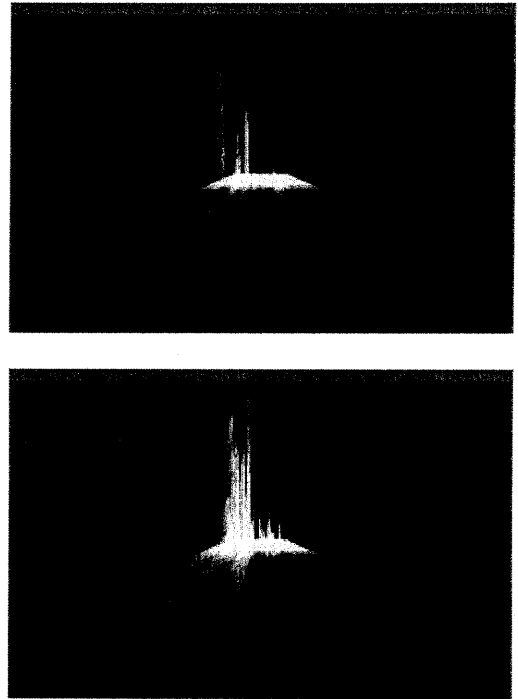


図3 MB-FFTによるパワースペクトル
(上:単音 下:コード)

3.2 Subharmonic Summation (SHS)

SHSは図4に示すように、FFTによって得られた周波数軸パワースペクトル $P(f)$ を

$$s = \log_2 f$$

でログ周波数軸に変換した後、

$$H(s) = \sum_{n=1}^N P(s + \log_2 n)$$

を用いて倍音的にシフト(1オクターブずつ圧縮)し、その総和を取ることによって、優位な周波数成分を検出する方法である¹⁷⁾。この方法は、Dick J. Hermesによって提唱された方法であるが、その時は単音の音程抽出に用いられていた。本研究では、この手法を和音のパワースペクトルに適用し、和音構成音の同定に利用している。

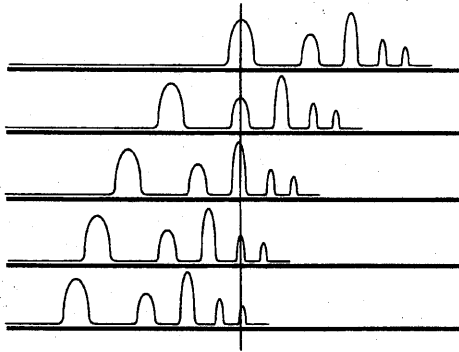


図4 SHSによるスペクトル圧縮

3.3 トニック-ドミナントペア検索

単音のパワースペクトルの大きさが、倍音の次数が大きくなるに従って減少すると仮定すると、SHSスペクトルのもっとも大きなものと、その次に大きなものがコードのルートと5度の音である可能性が非常に高い¹⁰⁾。実際にFFTのスペクトルを観察すると、上記の仮定はほぼ満たされていることが分かった。そこで、もっとも「有力」なルートと5度のペアを捜すことによって、ルート音を決定する手法を採用している。具体的な計算方法を以下に示す。

SHSでえられたスペクトルの優位なものから順に、0で始まるインデックスを与える。ある音階 N がそのときのコードのルートであると仮定し、ルートのインデックスの値を $I_{N(\text{Tonic})}$ 、 N に対して5度の音階のインデックスの値を $I_{N(\text{Dominant})}$ と表す。さらにこれらの和 Sum 、差の絶対値 $Diff$ 、その二乗和 $Weight$ を導入する。

$$Sum = I_{N(\text{Tonic})} + I_{N(\text{Dominant})}$$

$$Diff = |I_{N(\text{Tonic})} - I_{N(\text{Dominant})}|$$

$$Weight = Sum^2 + Diff^2$$

SHSのスペクトルにおいて上位4つの音階をルート音の候補とし、各音階について $Weight$ を算出する。このとき $Weight$ がもっとも小さいものをルートとして認定する。このことは、その音階とその5度の音階が、共にそろって大きいことを意味し、その音がルートである可能性が高いと考えられる。

3.4 コード進行の予測

本システムでは、二つの方法でコード進行の予測を行っている。一つは過去に入力されたコード進行に基づく予測、もう一つは音楽の傾向を利用した予測である。後者は、現段階では、「3度と7度がスムーズにつながるコード進行を好む」「ルートがII-Vの進行になる」など、ごく簡単なルールを用い、また判定に対する影響度も軽微に抑えてある。影響度を大きくすると、演奏者とシステムの意図の不一致が生じた場合、正確な演奏入力の妨げになる。ルールの選定や影響の度合いについては、今後もさらに検討を行う必要がある。

4. 演奏生成部

演奏生成は、過去に我々が開発したセッションシステムと同じく、「ミュージカルジェスチャーの再構成」によって行う¹⁰⁾。ミュージカルジェスチャーとは、それらの時系列データが単一の意味を持つように適度に分離された音楽要素を意味し、和音や旋律、リズムパターンなどを示す。演奏情報をミュージカルジェスチャーに分離した状態でシステム内に保存し、再合成することによって演奏を生成する。このことは非能率的に見えるが、即興性を導入しやすいという大きなメリットがある。

5. 実験と結果

本研究の場合、セッションシステムとしての性能は、音階およびコード名の抽出の性能に大きく依存している。正確な音楽情報がえられなければ、的確な演奏を生成することができない。

そこで、音階およびコード名が音域によってどの程度正確に抽出されるか、実験を行った。実験は、単音と3構成音の和音について行い、それぞれを各オクターブにつき100回ランダムな順番と強度で入力した。実験の結果を表1に示す。

表1 音階、和音の認識率

音程	24-35	36-47	48-59	60-71	72-83	84-95
単音	100%	100%	100%	93%	67%	63%
和音	75%	67%	100%	100%	92%	83%

この結果によると、単音の場合、音程72以上(MIDI規格による、以降も同じ)の音に関して正解率が低下している。ピアノの高音域は倍音の発生が少なくなり、ほとんど基音のみの音も少ない。今回の実験は、単音も和音も同じ方法(SHS, T-Dペア検索)で行ったが、これは本来コード名抽出のための手法である。倍音構成の単純な高音域の音はパワースペクトルのピークを検出するという、より簡単な方法で抽出することができる。

和音の場合、音程47以下の和音に関して正解率が低下している。この音域の音は非数次の倍音を多く含み、システムが誤認識を行った理由は、その倍音構成の不規則さにあると考えられる。しかしこれらの音で和音を構成すると、音が濁り人間でも判別しにくいことがあるため、演奏時に和音を構成するために用いられることは少ない。よってセッションシステムの機能に重大な障害になることはないと考えられる。

今回の実験では、入力がランダムなコード列なので、コード予測に基づくフィードバックは用いていない。よってこの機能についての評価は行っていない。

6. まとめと今後の課題

当初本システムでは、音階とコード名の決定を、倍音のパワースペクトルをパターン化し、SHSスペクトルとの間でマッチングをとることによって行っていた。その方法に比ベトニック-ドミナントペア検索による方法では、処理時間が短縮し、認識率は飛躍的に向上した。その結果、MIDI楽器を用いたシステムと比較して遜色ない、リアルタイム性、インタラクティブ性を獲得することができた。

コード進行予測については、特に音楽的傾向を利用した予測において、幾つかの問題を残す結果となった。今後も継続して、コード名判定に与える影響度の検討、システム内に備えるルールの選定と評価を行っていく必要がある。

さらに、作曲支援システムとしての機能を高めるために、改良を行ってゆく予定である。現在のシステムでも、複数のコード進行を記憶保存することは可能である。それらを演奏中、演奏後に自由に組み合わせ再現することができれば、作曲支援ツールとしてより有効に機能すると考える。

参考文献

[1] 青野裕司, 片寄晴弘, 井口征士: "バンドライクな音楽アシスタントシステムについて", 情報処理学会音楽情報科学研究会研究報告 94-MUS-8, Vol.94, No.103 (1994).

[2] 後藤真考, 日高伊佐夫, 松本英明, 黒田洋介, 村岡洋一: "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I. システムの全体構想と分散環境での実装", 情報処理学会音楽情報科学研究会研究報告 96-MUS-14, Vol. 96, No. 19 (1996).

[3] 日高伊佐夫, 後藤真考, 村岡洋一: "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II. ベーシストとドラマーの実現", 情報処理学会音楽情報科学研究会研究報告 96-MUS-14, Vol. 96, No. 19 (1996).

[4] 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演奏システム - JASPER -", 情報処理学会論文誌, Vol. 35, No. 7, pp.1469-1481 (1994).

[5] I. J. Taylor and M. Greenhough: "Neural Network Pitch Tracking over the Pitch Continuum", Proceedings of the 1995 International Computer Music Conference, pp. 432-435 (1995).

[6] 柏野邦夫・木下智義・田中英彦: "音楽情景分析の処理モデルOPTIMAの実装", 情報処理学会'95春全国大会, 6D-2, Vol. 2, p.97 (1995).

[7] Dick J. Hermes: "Measurement of pitch by subharmonic summation", Journal of Acoustical Society of America, 83(1), January 1988.

[8] Kia Ng, Roger Boyle and David Cooper: "Automatic Detection of Tonality Using Note Distribution", Journal of New Music Reserch, Vol. 25, pp. 369-381 (1996).

[9] Yushi Aono, Haruhiro Katayose and Seiji Inokuchi: "An Improvisational Accompaniment System Observing Performer's Musical Gesture", Proc. of 1995 International Computer Music Conference, pp. 106-107 (1995).