援システムとして発展させるためには ,認識された反復単位を演奏者が演奏中に自由に切り替えられるようにする必要がある . しかし ,実際のセッションでは切り替えに際して ,ジェスチャーやアイコンタクトなど ,音以外の情報を用いている場合が多い . これをどのように実現するか ,また音情報のみで切り替えを行うことができるかについて , 検討を行う必要がある .

現在は ,入力楽器としてピアノ ,エレクトリックピアノを中心に開発を行っている . エレクトリックギターも ,多少コードの認識率が低下するものの ,セッションが可能であることを確認している . ギター系の楽器は MIDI 化が難しいとされており ,その普及も鍵盤楽器に比べて極めて低い . 引き続き ,これらの楽器を視野に入れながらシステムの改良を行っていく必要がある .

謝辞　実験に際して ,サンプルデータの録音場所を提供して頂いた大阪芸術大学 ,及び多大なご助力を頂いた大阪芸術大学講師志村哲氏 , 同大学副手池渕隆氏に感謝します .

参考文献

[1] Robert Rowe: *Interctive Music Systems Machine Listerning and Composing*, The MIT Press, 1993.

[2] Masako Nishijima and Kazuyuki Watanabe: *Interactive music composer based on neural networks*, Proc. of the 1992 International Computer Music Conference, pp. 53-56, 1992.

[3] Bruce Pennycook, Dale R. Stammen, and Debbie Reynolds: *Toward a Computer Model of a Jazz Improviser*, Proc. of the 1993 International Computer Music Conference, 1993.

[4] Dale R. Stammen and Bruce Pennycook: *Real-time Recognition of Melodic Fragments Using the Dynamic Timewarp Algorithm*, Proc. of the 1993 International Computer Music Conference, 1993.

[5] Fred Lerdahl and Ray Jackendoff: *A Generative Theory of Tonal Music*, The MIT Press, 1983.

[6] 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.

[7] 後藤真孝 , 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.

[8] 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.

[9] 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.

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

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

[12] 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.

では1小節ごとに)再度それらを組み合わせることによって基本演奏パターンを生成する(図9参照).再構成を行う際 , 音の選択や発音時間 (リズム) の選択に自由度を持たせることにより ,演奏は作り出されるたびに微妙に異なってくる . しかし基本演奏パターン中の音数は , アクセントパターン中のアクセントの数に依存するので , 音の隙間 (休符) が多いことがある . このような隙間を埋めるために装飾音が付加され , 最終的な演奏となる . 装飾音は , そのときの和声的制約を受けながら ,システムに実装された音楽的知識を用いて自動生成される . 音楽的知識とは , あるコードのもとで使用できる音の ,使い方や組み合わせ方に関する妥当性と頻度を記述したものである .以上の処理の結果 , コンピュータの出力する演奏は , シーケンサ的な画一的なものとは異なり即興性豊かなものとなる .

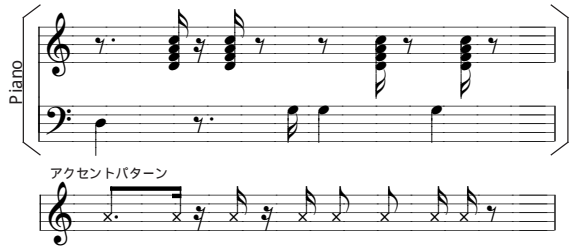


図 8 : ピアノ演奏からのアクセントパターン抽出の例

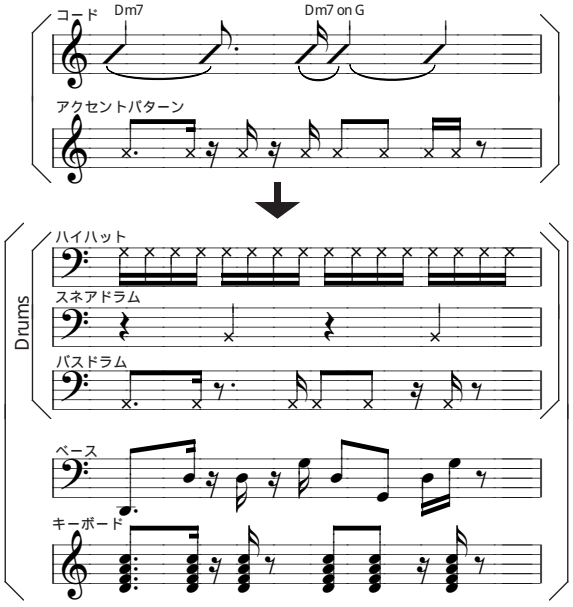


図 9 : 基本演奏パターン生成の例

7. 実験と結果

本研究で提案するセッションシステムの性能は ,音程およびコード名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 ,音程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 ,アコースティックピアノを用い , 防音のされていない一般的な部屋と , 録音等に用いる防音室で一度ディジタルオーディオテープ (DAT) にそれぞれ録音し , そのテープ再生音をシステムに入力する形で行った . 防音のされていない部屋では131個の和音を , 防音室では145個の和音を入力し , システムが正しく認識した数を集計した .実験の結果を表1に示す .

この結果によると ,いずれの環境においても高い正解率を示しており ,本研究が提案する一連のコード名認識の処理が十分に機能していることを示唆している .

一方 , 誤って認識したコードについて考えてみると , 一様に高い音域で演奏を行った場合に多く見られた .ピアノの音は音域が高くなるにつれて倍音が出にくくなる傾向があり ,倍音を積極的に利用するSHS のメリットが十分に生かされなかったためではないかと考えられる .

表 1 : コード名認識実験の結果

録音場所	非防音室	防音室	合計
データ数	131	145	276
正解数	127	138	265
正解率	96.9%	95.2%	96.0%

8. おわりに

当初本システムでは ,あらかじめパターン化した倍音パワースペクトルのテンプレートと ,SHS スペクトルとの間でマッチングをとることによって , 音程・コード名の認識を行っていた . 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 , リアルタイム性を獲得することができた . 一方リズム認識の時間分解能については , MIDI を用いた場合に比べ , 現段階では不十分といわざるを得ない . 試作システムを用いた予備実験では ,向上の余地があることが分かっており , 今後改善していく予定である .

また , 曲の主調と相関係数を併用することによって , 正確に反復単位を認識することが可能となった .今後作曲支

5.4　反復単位の認識

本研究がモデルとして取り上げたセッションの形態では，主導的演奏者の提示したコード進行が，曲を構成するための唯一の情報となる．そのため，規則性に乏しいコード進行は敬遠される傾向にあり，一般的には4小節や8小節といった単位で繰り返されることが多い．さらに，2ないし3種類の反復単位を切り替えて用いることにより，更に複雑な演奏を行うこともある．反復単位の切り替えは，セッション曲に構成を与え，簡易的な作曲作業となる．セッションシステムにこのような作曲支援の機能をもたせるためには，最小反復単位を的確に認識すること，演奏者の意図に従って反復単位を切り替えることが必要となる．

本システムでは，相補的な2つの方法で反復単位の認識を行っている．まず1つは，曲の主調に基づく認識である．演奏者が用いているスケールの傾向から主調を推定し，その主調における主和音をもって反復単位の終止とするこの手法は，認識における時間遅れがないという特徴をもつ．逆に，反復単位中に起こった転調により，正確に主調が推定できなかった場合は，終止位置を誤認識しやすいという欠点も存在する．もう1つは，相関係数を用いた方法である．この方法は，認識に必ず2回以上の反復単位を必要とするが，正確性は高い．この2段階の方法は，人間の認識プロセスと類似している．

本システムでは，4，8，12，16小節の幅で相関係数の計算を行っている．図6では，長さの異なる3種類の反復単位を入力した際の，相関係数の推移を示している．

6.　演奏生成部

6.1　ミュージカルジェスチャーの概念

特にジャムセッションなど，即興演奏を生成する場合に有効な手法として我々が提唱する，「ミュージカルジェスチャー」の概念について説明を行う[9][10]．ミュージカルジェスチャーとは，和音列やリズムパターン，フレーズの集合を総称するもので，それらの時系列データが単一の意味・性質を持つように適度に分離された音楽要素ということができる．人間が即興演奏を行う場合，過去の演奏情報を利用する事によって新たな演奏を生成する．このとき過去の演奏情報を楽曲として完全な形で記憶しているとは考えにくい．ほとんどの場合が，リズムはリズムパターンとして，コードはコード進行として，個別にパターン化し記憶していると考えられる．そしてそれらを再度組み合わせることにより新たな演奏を生成する(図7参照)．この再構成の過程において生じる微妙な変化が演奏に多様性をもたらし，即興演奏の基本となる．ミュージカルジェスチャーの再構築による演奏生成は，人間の演奏生成過程をモデルとした即興演奏に適した手法ということができる．

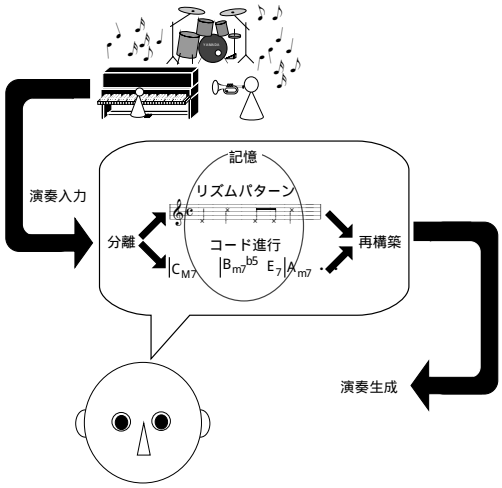


図7：ミュージカルジェスチャーを介した音楽の理解と生成

6.2　ミュージカルジェスチャーの抽出と構築

本システムが対象としているミュージカルジェスチャーは，コード列(コード進行)とリズムパターンである．このうちコード列は，演奏解析部の結果をもとにコードの切り替わりを追跡することで形成される．リズムパターンは，バンド形態の演奏ではドラムスなどの打楽器によって決定される場合が多い．しかし本システムの入力楽器はピアノやギターなどの非打楽器であるので，その演奏からリズムパターンを推定する必要がある．そこで，(1)コードの切り替わる場所　(2)相対的に音量の大きい場所　に聴感上のアクセントがあると仮定し，それらを時系列的に並べたアクセントパターンを，リズムパターンとして用いる(図8参照)．ただし，MIDI信号ではなく音響信号を用いている本システムでは，入力演奏の解析における時間分解能が，16分音符以下の精度でアクセント位置を決定するのに十分ではない．これはMB-FFT スペクトルの粗さに原因がある．たとえば120BPM(注)の曲では，1秒は16分音符8こ分に相当する．現段階では毎秒15.625本のMB-FFT スペクトルが得られているが，その2倍の32本程度得られれば十分であると予想し，システムの改良を行っている．現在は，システム内に16分音符レベルのアクセントパターンを複数用意し，曲調によって選択したそれらと演奏から抽出されるアクセントパターンを合成することによって，最終的なアクセントパターンを得ている．

得られたミュージカルジェスチャーは個別のパターンとしてシステム内に保存され，数小節ごとに(現行システム

(注) BPM = Beats Per Minute1分間の演奏に含まれる拍の数．4/4拍子で120BPMならば，1分間に4分音符が120個．

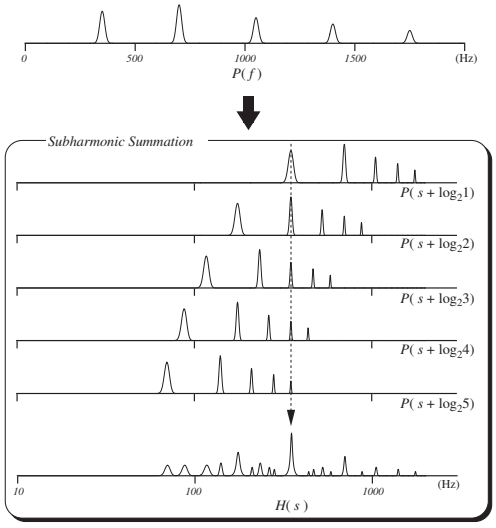
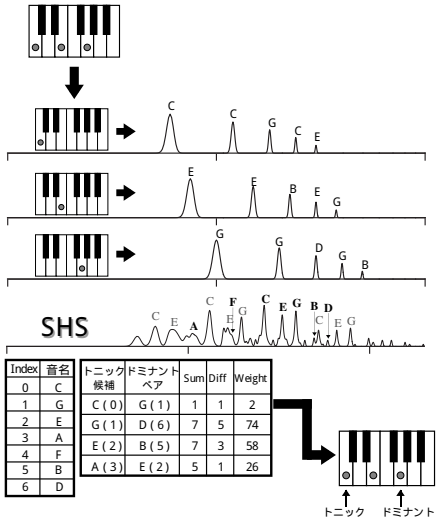


図 4 : Subharmonic Summation



Index	音名
0	C
1	G
2	E
3	A
4	F
5	B
6	D

トニック候補	ドミナントペア	Sum	Diff	Weight
C (0)	G (1)	1	1	2
G (1)	D (6)	7	5	74
E (2)	B (5)	7	3	58
A (3)	E (2)	5	1	26

図 5 : トニック - ドミナントペア検索

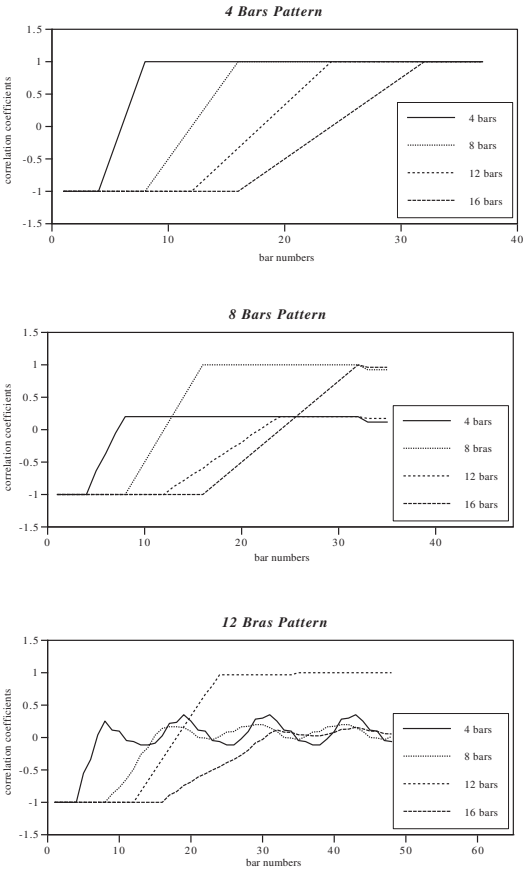


図 6 : 3 種類の反復単位に対する相関係数の推移

の音 (ドミナント , *dominant*) である可能性が非常に高い [12] . 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」なトニック - ドミナントペアを捜すことによって , コード名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 (図 5 参照) .

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 . ただし同じ音名のスペクトルについては , もっとも大きなものを選択する . ある音 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 $I_{N(Tonic)}$, N に対して 5 度の音のインデックスの値を $I_{N(Dominant)}$ 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 .

$$Sum = I_{N(Tonic)} + I_{N(Dominant)} \tag{3}$$

$$Diff = \left| I_{N(Tonic)} - I_{N(Dominant)} \right| \tag{4}$$

$$Weight = Sum^2 + Diff^2 \tag{5}$$

SHS のスペクトルにおいて上位4つの音をルート音の候補とし , それぞれの音について $Weight$ 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 , その音とその 5 度の音が , 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 .

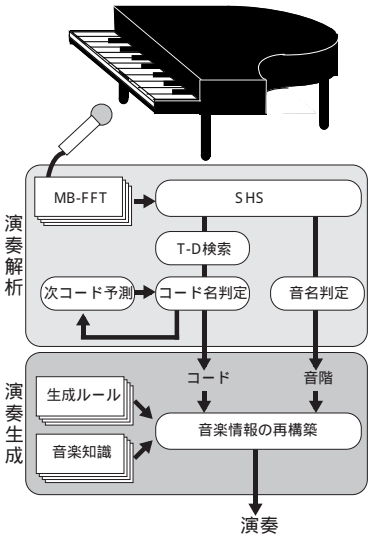


図 2 : セッションシステム概要

5. 演奏解析部

5.1 Multi Band FFT (MB-FFT)

図 2 に示した処理の中でも , 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 . また最も低い半音のピッチの差は , 約 2Hz である . 10KHz を越えるバンド幅に対し , 2Hz 以下の分解能をえるためには , 通常の 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ある . その際 , 各帯域で用いるデータのサンプリング周波数を変化させることにより , 同じ 1024 点の FFT で異なる周波数分解能を得ることができる (図 3 参照) .

本手法は , 楽器音は周波数が高くなるにつれて半音のピッチの差も大きくなり , 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 . もっとも高い 2 オクターブを受け持つ FFT では 32KHz サンプリングのデータを用いる . 1 オクターブ下がるごとに 2 倍の精度の周波数分解能が必要となるが , 逆にバンド幅は 2 分の 1 で十分であるから , その下の 2 オクターブについては 8KHz サンプリングのデータというように , データを間引きながらサンプリング周波数を擬似的に下げていく . 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 . 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 .

現段階では , 毎秒 15.625 本の MB-FFT スペクトルを出力する設定としている . また , 現行システムでは採用していないが , リアルタイムで毎秒 31.25 本のスペクトルが得られることが分かっている . 処理系として SGI 社のワークステーション (CPU: MIPS R10000 @195MHz) を用いている .

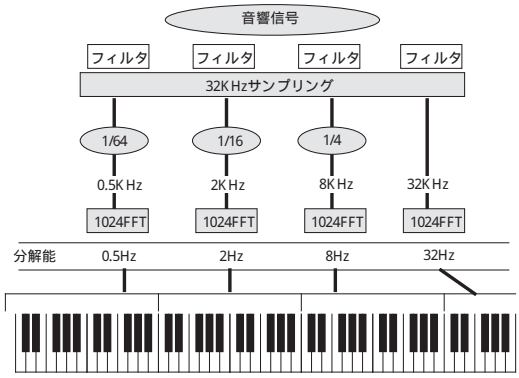


図 3 : Multi Band FFT

5.2 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 $\mathrm{P}(f)$ を

$$\mathrm{s} = \log_2 f \tag{1}$$

でログ周波数軸に変換した後 ,

$$\mathrm{H}(\mathrm{s}) = \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{P}(\mathrm{s}+\log_2 \mathrm{n}) \tag{2}$$

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 [11] . 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 , コード構成音の同定に利用している . 図 4 ではもっとも大きいスペクトルにのみ矢印を施してあるが , 実際には上位 4 つのスペクトルを和音のルート (根音) の候補として利用する .

5.3 トニック - ドミナントペア検索

単音によって励起される倍音のパワースペクトルの大きさが , 倍音の次数が高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の音 (トニック , *tonic*) と 5 度

2. セッションとセッションシステム

我々が提案するセッションシステムについて述べる前に，ジャムセッションについて簡単な説明を行い，本システムのモデルを明らかにする．

ロックやジャズなど軽音楽と呼ばれる音楽は，バンドという形態で行われることが多く，メンバー間の意志疎通と相互作用が演奏に際して非常に重要になる．さらにこういった小規模なバンドでは，メンバー間のインタラクション自体が楽しみの対象となることがある．つまり，即興的に演奏を進める中で互いの音を聴き，それに何らかのリアクションを加えてゆくことで，ひとつの楽曲を形成してゆく「セッション」と呼ばれるものがそれにあたる．普通セッションは，決まった曲を演奏するときに比べると格段に混沌とした状況下でスタートする．たとえば，曲の主調やリズムパターンだけをあらかじめ決めておいて演奏をはじめる場合や，主導的演奏者がコード進行を順次提示しそれに他の演奏者が追従する場合などがある．本研究で取り上げたのは後者のタイプのセッションで，主導的演奏者を人間が，その他の演奏者をコンピュータが担う．コンピュータは，人間から提示されたコードを解析し，それらをコード進行として記憶しながら音楽的に調和のとれた伴奏を生成する．これら一連の処理は，実際のセッションと同様に，主導的演奏者の演奏に追従する形で行われる必要がある．（図 1 参照）

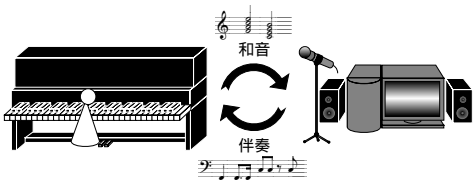


図 1：セッションシステムの入出力

3. 関連研究

Roweは，コンピュータによる音楽システムの聴取機能と生成機能について系統だった研究を行い[1]，本研究を含め多くのセッションシステム研究に影響を与えた．この研究に基づいて開発されたCypherと呼ばれるシステムは，作曲が可能であり，セッションシステムというよりも，インタラクティブコンポージングシステムとしてとらえられることが多い．Nishijimaらは，ニューラルネットワークによるセッションシステムを実現した[2]．ニューラルネットには，入力となる演奏の例とそのときに期待される出力演奏を学習させる．Pennycookらはジャズの即興演奏をモデルとしたセッションシステムを開発した．このシステムもCypherと同じく，聴取部と生成部が独立した構造をとっている[3][4]．聴取部では，LerdahlとJackendoffの Grouping Preference Rule[5]をもとに，入力された演奏をフレーズに分割する．さらに言語処理等の分野で用いられてきた Dynamic Timewarpアルゴリズムをフレーズのパターン化に利用するなどの特徴を持つ．和気らは，セッションシステムに感性量を導入し，生成する演奏の表情付けを行った[6]．この研究では，入力される演奏（音）の時間的密度や不協性をテンションパラメータと呼ばれる数値で表すことにより，演奏の盛り上がりを認識し，演奏出力に反映させる試みを行っている．後藤らは，独自に開発したネットワークを介した MIDI 通信プロトコルを用いることにより，一台のコンピュータが一つの楽器を担当する分散協調型のセッションシステムを開発した[7][8]．このシステムは，人間とコンピュータの間のインタラクションだけでなく，異なる楽器を担当するコンピュータ間のインタラクションも実現できるという特色を持つ．また，演奏者のジェスチャーを演奏制御情報として用いるという新しい試みも行っている．筆者らは，楽譜が与えられていない状況でのセッションに着目したセッションシステムを開発した[9][10]．このシステムでは人間の演奏者から与えられる和音やリズムの情報を正確に把握し，演奏生成に利用することが重要な問題となる．我々は実際のセッションにおいて演奏者が音楽を記憶する方法について考察し，即興演奏に適した音楽情報のパターン化手法を提案した．この音楽情報パターンをミュージカルジェスチャーと呼び，本システムでも利用している．

4. セッションシステム概要

システムの概要を図 2 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．西洋音楽に代表される 12 音階を基本とする音楽では，任意の音は 1 オクターブ上昇すると周波数が 2 倍になるという性質がある．本研究では，この楽器音の性質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．詳しくは後述する．

次に，FFT の結果を，Subharmonic Summation (SHS)と呼ばれる，楽器音の持つ倍音構造を利用した解析によって，複合音に含まれる音程の候補を選ぶ．この音程の候補群に対して，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．これらについても，説明を加える．

演奏生成部では，得られたコード進行の情報をもとに伴奏を自動生成する．伴奏の生成は，セッション中の人の思考をモデルとした「ミュージカルジェスチャーの再構築」という手法で行う．

アコースティック楽器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

[††]和歌山大学　システム工学部　デザイン情報学科

あらまし　簡易作・編曲支援機能を備えた,アコースティック楽器対応ジャムセッションシステムについて述べる．従来のセッションシステムは,MIDI楽器を用いるものが主流であった.我々は,セッションでは和音とリズムが重要な情報になると考え,解析の対象をそれらに絞ることによって,アコースティック楽器による演奏入力を実現した.演奏聴取部では,和音名の認識にあたって楽器音に含まれる倍音を積極的に活用している.この手法により,ジャムセッションを行うのに十分正確な認識が可能である.また,入力された演奏から最小の繰り返しパターンを自動的に認識・保存する機能があり,簡易的な作・編曲支援システムとして用いることもできる.ミュージカルジェスチャーという概念を導入した演奏生成部は，即興性の高い演奏を作り出す．

Development of a Session System with Acoustic Instruments

Yushi AONO[†]　Haruhiro KATAYOSE[††]　Seiji INOKUCHI[†]

[†] Faculty of Engineering Science, Osaka University

[††] Faculty of System Engineering, Wakayama University

ABSTRACT　This paper describes an improvisational session system for popular music. The system responds to acoustic polyphonic input without any scores. There are few systems which allow acoustic signal input of polyphonic instruments. We, however, realized a real-time acoustic listener by focusing limited information, chords and rhythms, which come to be important in session. The listener makes good use of harmonic overtones though those are often disturbances in polyphonic signal processing. The accuracy of chord recognition is fairly high. We introduce a unique concept, musical gesture, into the player of the system. The concept makes the player produce highly improvisational performances. This system has also composing ability. This system is not a simple accompanying system but a bandlike composing environment.

1.　はじめに

即興的な演奏を基本とするジャムセッションでは,演奏者間の音情報の円滑なやりとりがもっとも重要である.既存の曲を演奏する場合とは異なり，ジャムセッションでは,各演奏者が提示する断片的な曲のアイディアを膨らませてゆくことがその目的となるからである.この様なセッションを人間とコンピュータの間で実現するセッションシステムの研究が,過去幾つかなされてきた.しかしそのすべてはMIDIと呼ばれる電子楽器を，人間がコンピュータに演奏を入力する際のインタフェースとして用いている．本研究は,ピアノやギターなどのアコースティック楽器を入力インタフェースとして用いるための音響解析手法を提案し,より演奏者にとって自然なセッションシステムを実現するものである．またセッションはその即興的性質から,作曲や編曲の場として用いられることがある.我々は，セッションシステムの持つ作・編曲環境としての可能性を引き出すために,楽譜を用いることなく,入力演奏のみから楽曲を構成できるセッションシステムの開発を目指す．

† 〒560　大阪府豊中市待兼山町 1-1
TEL: 06-850-6373
E-mail: aono@inolab.sys.es.osaka-u.ac.jp