

画像情報を MIDI 情報として用いた電子音楽の制作

熊谷 武洋

Composition of Synthetic Music that uses Bitmap Images as MIDI Note Number

KUMAGAI Takehiro

(Received September 30, 2011)

Key words: MIDI, Digital Music, DTM

1. はじめに

画像情報を音楽情報や音響情報として活用する試みは、これまでに数多くあった。

前者の代表的な例としては、作曲家ヤニス・クセナキスにより1970年代に開発された UPIC システム、またその発展型である IanniX システムが挙げられる。図形情報を楽譜として記譜し、グラフィカルな音楽情報を生成するというものである。

後者の代表的な例としては、METASynth として商品化されているスペクトログラムシンセサイザー等が挙げられる。複合信号を窓関数に通して周波数スペクトルをサウンドスペクトログラム情報に変換したり、通常の画像情報をサウンドスペクトログラムとして見なして音声合成処理をおこなうものである。

本稿で論じるシステムにおいては、これらのアプローチとは異なり、ファイル上の画像情報やリアルタイムに生成された画像情報を任意のタイミングで組み合わせ、その画像情報の中から音楽的要素を抽出して即興フレーズを生成するというものである。

本システムは、画像情報から出力された情報から音楽の構成要素として有用と思われる箇所を制作者自身が感性と主観によって恣意的に取捨選択し、音楽的な情報として積極的に活用するという演奏スタイルである。よって、モチーフ自体が持つ構造を分析し、自動作曲化していくアルゴリズムコンポジションの方法論ではなく人為的な修正や二次加工を前提とした手動作曲の方法論である。

このように本システムは既成曲を模倣して演奏するためのものではなく、単純な操作によってフレーズを即興で創出するということを目的としている。生成されるフレーズは音楽的に整ったものではないが、画像でフレーズを奏でる、という新しい演奏スタイルの応用分野としては、音楽療法、知育玩具、エデュテイメントソフトなどが考えられる。

なお本システムを用いて制作したデモンストレーション曲は、作品名「Beating FM」として芸術科学会主催第8回芸術科学会展2010 第2部門 デジタルミュージック部門にて最優秀賞という客観的な評価を得た。このことにより、部分的ではあるが本方法の有効性と可能性を示すことが出来た。

2-1. 映像の音声信号化

本システムでは、画像情報から音響・音楽的な情報へマッピングするための情報を取得するにあたり以下の処理過程を経ている。

- ① 実数型のマトリックスのデータを音声信号変換してオーディオ化し、WAV データを生成
- ② WAV データを高速フーリエ変換処理を行い、ピッチ情報を取得
- ③ ピッチ情報から MIDI ノートへ変換

音声信号変換処理を行う際、jit.peek オブジェクトを用いた。

jit.peek オブジェクトは内部にバッファを持っており、latency アトリビュートで指定した msec 単位の時間に相当するバッファが生成される。この時間内に受け取ったすべてのマトリックスのデータを変換する (図 2)。

音声信号は、MSP のサンプリング・レートに従って常に出力されている。サンプリング・レートが 44.1kHz なら、1 秒間に 44,100 個のオーディオ・データが再生されることになる。このことからソースとなる画像情報は時間軸を持っているほうが多彩な変化を生み出すことができる。

図 4-8 は本システムの jit.peek オブジェクトの一部の様子である。read で読み取った静止画やムービー画像をシグナルに変換して、波形を scope オブジェクトで確認しながら音として聴くことができる。

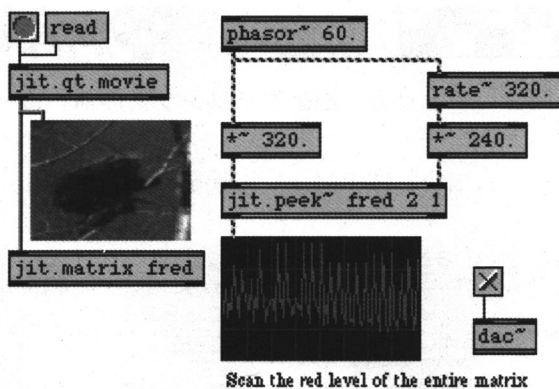


図 2 画像の音声信号変換画

2-2. 画像情報ソースによる変化

次に read によって読み込む画像や映像によって出力される情報がどのように変化するか対照実験を行った。タイリング密度の異なるチェッカー画像をそれぞれ読み込ませた結果、出力情報は密度が高くなる順に複雑になり、最低限音楽的な要素を感じる事の出来るような情報が出力された (図3)。

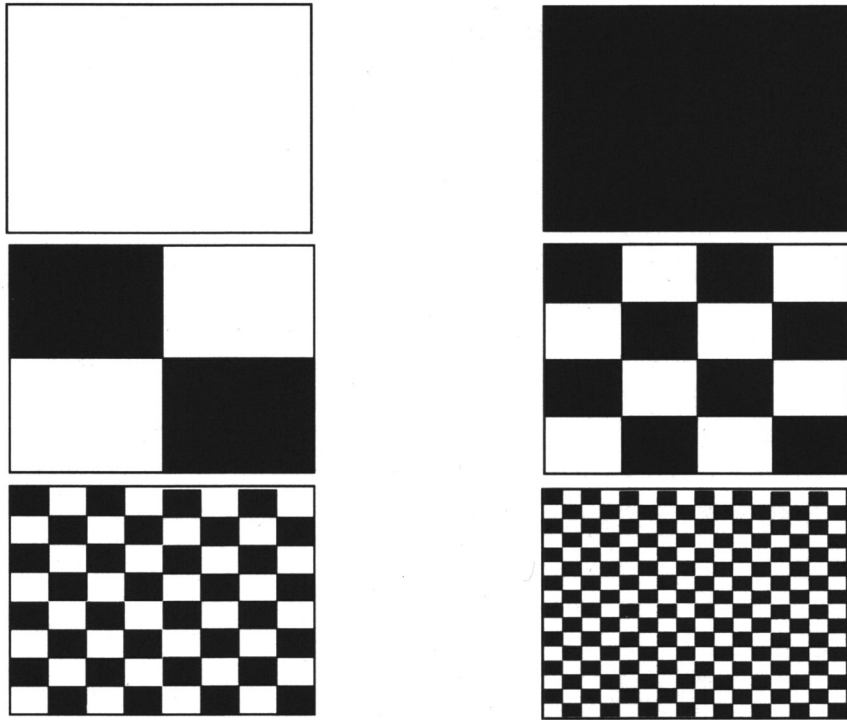
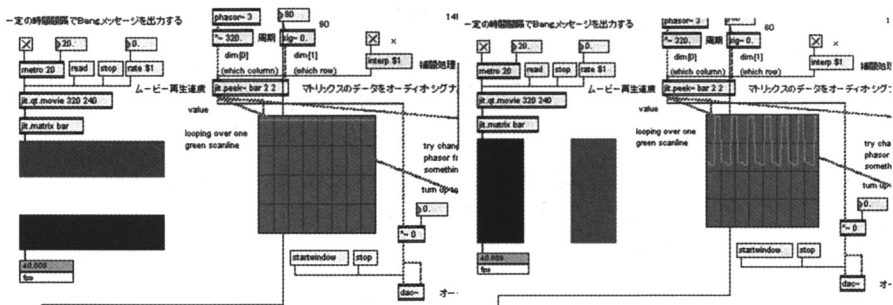


図3 チェッカー画像タイリング密度別一覧

次に、チェッカー画像以外に国旗画像で同様に実験を行った (図4)。



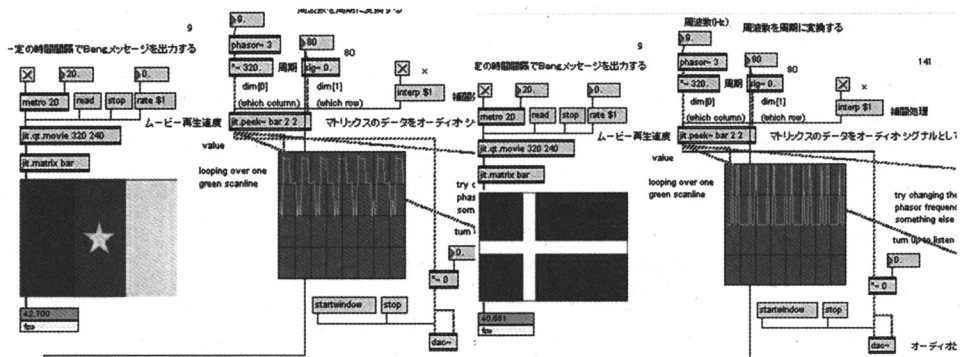


図4 国旗画像

実験の結果、横方向に複雑な画像は周期に顕著な変化が表れていることがわかる。

しかしながら、音声信号自体は倍音成分の少ないサイン波のような音であった。

次に周波数成分の複雑な標本化された画像、具体的にはデジタルカメラによって撮影をした自然景観の画像について同様に実験を行った。



図5 標本化された画像

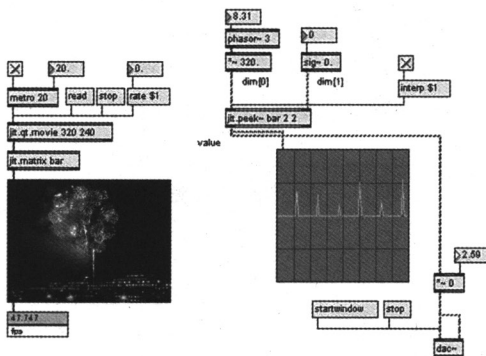


図6 標本化された画像による音声信号

出力される音声信号は不規則なランダム成分やノイズが入り混じる大変不快な音であった(図6)。

このように出力された情報をそのまま波形合成のための変数として扱うと、耳障りであり音楽的な目的としては用いることができない。そこで、本システムでは、ピッチのみを検出して、MIDI のノートナンバーとして外部音源にアサインするという方法を採用した。

2-3. ピッチ抽出

高速フーリエ変換処理を行うため、fiddle[™] オブジェクトを用いた。fiddle[™] オブジェクトは、入力された音声信号からピッチ、アタック、ピーク・レベルを検出するオブジェクトである。fiddle[™] オブジェクトの主な機能を以下に挙げる。

- ①単音のオーディオに対して基本パッチの検出を行う
- ②ピッチは周波数 (Hz) 単位ではなく、MIDIノート・ナンバーの数値で出力される
- ③誤差や誤認識を低減した加工されたピッチと、そうした処理を行わない未加工のピッチを出力する
- ④ピーク・レベルも検出可能である

本システム内の fiddle[™] オブジェクト内の窓関数のサイズとして1024を設定した。この数値はサンプル単位の整数で、256、512、1024、2048といった2のべき乗でなければならない。数値が大きくなるにつれて、ピッチ検出の時間間隔は長くなるが、検出可能な最低ピッチが下がる。256、512、1024、2048等を全て試した結果、1024のサイズが最適であった。

音楽表現として成立させるには、それが自動にせよ手動にせよ何らかの二次加工が必要になってくる。まず音楽性に大きく関わるのはノートとスケールのマッピングである。この一次処理で大きく完成度や音楽性が決定される。

本システムでは、MIDI規格に準拠しているため、微分音程といった音高はスキップし、いわゆるドレミファソラシドのピッチを取得し、処理を行った (図7)。

音名	C	C#	D	D#	E	F	F#	G	G#	A	A#	B	C
周	261.	277.	293.	311.	329.	349.	369.	391.	415.	440	466.	493.	523.
波	6255	1826	6647	1269	6275	2282	9944	9954	3046		1637	8833	2511
数	6530	3097	6791	8372	5691	3143	2271	3598	9757		6151	0125	3060
	0598	6872	7407	2080	2869	3003	1634	1749	9945		8089	6124	1197
	6	1	6	9	9	9	4	26	1		9	1	2

図7 周波数と音名の対応

3. 実証応用作品事例

3-1. 制作目的

本システムのツールとしての表現力や機能を実証する目的としてデモンストレーション曲を制作した。

“何かが生成される時の不安感と期待感”を表現テーマに曲を構成した。しかしながら、画像情報を用いるということから、先に曲調など音楽的な意図や設計があって、それを具体的な音に変えていく、というアプローチは採ることができない。加えて画像情報を MIDI 情報に変換したとしても音楽表現的に有益な情報が含まれているとは限らない。

そこで、まずはモチーフとなる様々な断片的フレーズを画像情報から生成し、そこから取捨選択して構成していくというアプローチを採った。

前作と同様に、原始的な信号のような響きの中に複雑な音色変化による表現手法によって展覧感を形成した。

このように本作品は、本システムの処理結果を基にしてアプリケーションソフトウェアと DAW ソフトウェアを用途別に使い分け、半自動作曲による方法で作曲した。自動作曲ソフトが出力する MIDI ファイルを基に DAW ソフト上で音楽的調整を加えて完成とした。

3-2. 作品概要

作品時間： 2分40秒

使用ソフトウェア： WindowsXP (OS)、MAX/MSP (開発環境)、CUBASE (DAW)
FM7 (VSTi)、SY77 (ハードウェアシンセサイザー)

3-3. 制作過程

まず、画像情報の中に含まれる情報を音声情報として取得する。その場合、あらかじめ画像情報をクイックタイム形式でマージし、フィジカルインターフェイスもしくは、画面上の GUI でフレームをランダムアクセスし、フレーム単位で静止や動画を呼び出す。このときの画像の取捨選択具合が、あたかも映像による VJ の即興演奏のようにプレイヤーの個性やセンスの差が顕著に出てくる部分である。今回のデモンストレーション曲制作では対照実験の結果から、周期性や反復性がありながらも音高の跳躍や乱れを生じさせる画像をあらかじめ選定しておいたものを用いた。

次に画像情報から取得した値から合成される音声は使わず、音程の情報だけを MIDI のノートナンバーとして変換・取得した。

取得した情報は、MIDI 規格の音程レンジ内に収まるように写像し、音楽的な情報にリマップ処理を行うが、音価については積極的に演奏情報として意味のある情報に編集を行う必要がある。素材状態のままでは、かなり過度に不規則で不安定な情報が含まれている (図8、図9)。よってスタティックな楽曲作品として最低限の調整を施すために若干の二次加工を DAW 上で行った。

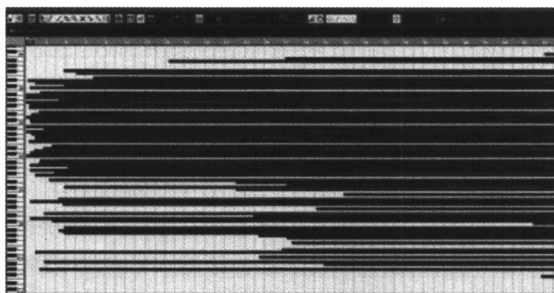


図8 素材状態のMIDIデータ（一部）



図9 MIDIデータのスコアエディタ画面

しかしながら、DAW上で二次加工を行う際、必要以上に修正加工は行わないということに注力した。その理由は、調整されたフレーズが悪い意味で文化的な音楽になってしまい、荒削り感や原始的な躍動感がなくなるためである。よって、若干のフレーズのテイク構成の編集とクオンタイズとリズム調整だけにとどめた。

なお、どのテイクを取捨選択し、どのような音色を作成・選定するかによって曲の雰囲気が大きく変わる。この過程に個性的表現の要素が大きく関与する（図10）。

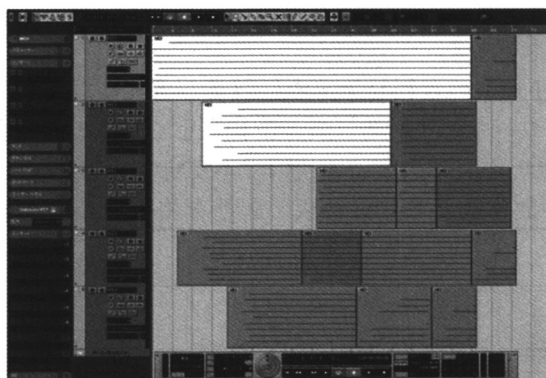


図10 DAW上の編集画面（一部）

次に MIDI のノートナンバー情報を音源側に送り、発音させる。その際、音源側で音色設定を行い、アルペジオやエフェクトを付加するなど若干の二次加工を施し、オーディオファイル化して完成となる。

今回のデモンストレーション曲制作においては、前述したように素材の取得方法に加え、最終的な曲の印象やイメージを音の高さや長さではなく、音響的な音色によって形成するという制作意図を持っている。そこで、音源には FM 合成方式によるシンセサイザーを用いた（図 11、図 12）。

FM 合成方式はすでに過去の音声合成技術となっているが、選定する理由は以下の通りである。

- ・ 現在主流である PCM 方式と異なり、様々なパラメータにアクセスできるため、曲のフレーズに合致した音色を精緻に合成する目的に適している。
- ・ あくまで主観であるが、聴感上、硬質から軟質までの幅広い音が全体になじむ
- ・ 時間的に音程が変化する音色などを作れるため、トリガーが単純でも豊かな響きを醸すことができる



図 11 FM 合成ハードウェア音源 SY77

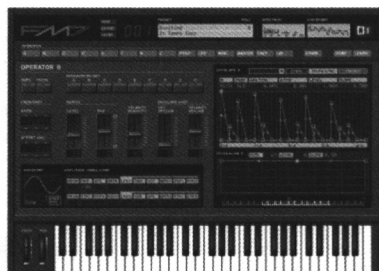


図 12 FM 合成ソフトウェア音源 FM7

4. 評価

上述した意図が効果的に反映され音色がどのように寄与しているのかを検証するために、同一の MIDI ファイルを音色を変えてミックスダウンし、ゆらぎ解析ソフトにて評価を行った。楽曲自体の出来の良し悪しは別として音色による差異があることを定量的に追認した。これらの評価方法は、前作と同様である。

FM 音源の効果を定量評価するために完成した曲を、スタインウェイ B のサンプリング音のみによる曲と FM 音源を用いた曲とで解析ソフトによるゆらぎ成分の分析と対照比較した。

この二つの曲は、当然ながら音価と音程は同じである。差異があるとするならば音色の違いということになる。この解析方法はあくまで評価方法の一つにすぎないが、音色が及ぼす効果を計る上で一つの指標になり得る（図 13）。

解析結果は FM 音源による曲の方が音楽的表現としては豊かであるが、成分的にはノイズ的な指向を持っていた（図 14、図 15、図 16）。曲のコンセプトが題名どおりにビートが主なモチーフであるため、当初の意図を達成できたと評価できよう。

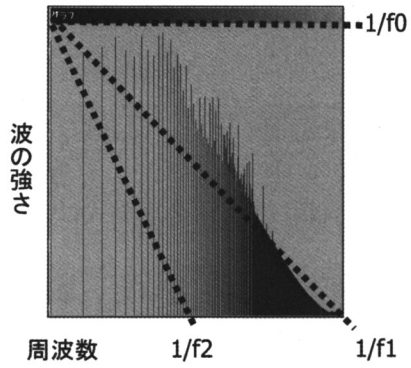


図13 「 $1/f$ のゆらぎ」では、周期の長い波ほど大きな振幅に周期の短い波ほど小さな振幅になる

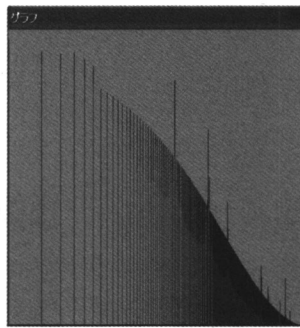


図14 サイン波のみ→適度なゆらぎ

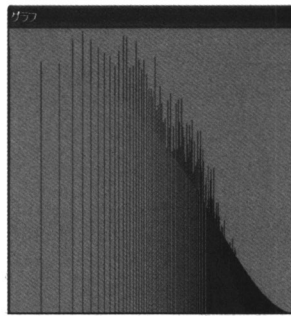


図15 スタインウェイBのピアノ音のみ→過小なゆらぎ

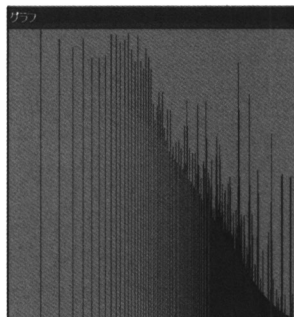


図16 デモンストレーション作品「Beating FM」→過剰なゆらぎ

5. 今後の課題

多くの改善点はあるが、画像でフレーズを生成する、という演奏スタイルによる楽曲生成の可能性を検証し、加えてデモンストレーションとして制作した曲は学会主催のコンテストにて客観的な評価も得ることもできた。よって本システムの有効性について一定の成果を得たと言える。

画像情報といういわば音楽的なものとは対極にあるものを用いて即興的かつ偶発的に、音楽的なものを形成するという本システムの特徴は様々な応用例が考えられる。

今後の課題としては、もっと多くの人たちに楽しんでもらえるために映像とフレーズの関連性が先験的に感じられるように設計し、インターフェイスと画像情報と音楽情報のリマップについて多様性を持たせることである。

参考文献

- [1] Curtis Road：コンピュータ音楽 歴史・テクノロジー・アート、東京電機大学出版局、2001
- [2] 高橋信之：コンプリート MIDI ブック、リットーミュージック、2003
- [3] 高橋信之：コンプリート MIDI プログラミング・ブック、リットーミュージック、2006
- [4] 高橋信之：コンプリート MIDI ブック、リットーミュージック、2003
- [5] 高橋信之：コンプリート SYNTH プログラミングブック、リットーミュージック、2007
- [6] リットーミュージック出版編集部：MIDI バイブル 〈1〉 MIDI1.0規格 基礎編、リットーミュージック、1997
- [7] リットーミュージック出版編集部：MIDI バイブル 〈2〉 MIDI1.0規格 基礎編、リットーミュージック、1999
- [8] ポール メシク：C++MIDI プログラミング— Windows95/98用 MIDI アプリケーションの開発 (Windows programming technique)、プレンティスホール出版、1999
- [9] 岩宮 眞一郎、音色の感性学—音色・音質の評価と創造、コロナ社、2010
- [10] 谷口 高士、音は心の中で音楽になる—音楽心理学への招待、北大路書房、2000
- [11] 中村 文隆、グラフィカル言語 PureData による音声処理、CQ 出版、2009