

研究報告

色情報による CV 信号の生成・オーディオ信号による映像生成 及び相互フィードバック

Generating CV signals based on color information and video based on audio signals, And implementation mutual feed

野垣 美歩

Miho NOGAKI
東京電機大学大学院
Tokyo Denki University

斎藤 博人

Hiroto SAITO
東京電機大学
Tokyo Denki University

概要

本研究では、音と映像が相互に影響して音と映像を自動生成するシステムを提案する。このシステムでは、音（CV 信号）と映像（色）がフィードバックループによりリアルタイムに生成され、不確実性および再現不可能性を持っているのが特徴である。生成する映像は、CV 信号によるシンセサイザの演奏を含めた全ての演奏音から特徴量を抽出し、そのパラメータで出力映像を制御する。さらに、出力された映像の特徴量に基づき、新たに音のパラメータ（CV 信号）を決定する。演奏開始時の CV 信号の初期値は、ステージ上にある物体の色をカラーセンサによって取得し生成することによって、不確実性と再現不可能性を実現している。本研究は新たなメディアアートの手法の要素技術へ、または映像編集技術の専門的知識を有していないミュージシャンを対象とした映像自動生成システムへの応用が期待できる。

In this research, we propose a system in which sounds and video interact with each other to automatically generate sounds and video. The feature of this system is sounds (CV signals) and video (colors) are generated in real time by feedback loops, which has uncertainty and non-reproducibility. Parameters are extracted from the feature values of all performance sounds including the synthesizer performance by the CV signals, and the output video is controlled by the parameters. Furthermore, sound parameters (CV signals) are newly determined based on the feature values of the output video. The initial value of the CV signals at the start of the performance is uncertain and unreproducible by acquiring and generating the color of the object on the stage with a color sensor. This

research is expected to be applied to elemental technology of new media art method, and automatic video generation system for musicians who do not have specialized knowledge of video editing skills.

1. 背景

一般的にシンセサイザとは、電気信号で音を生成する電子楽器であるオーディオシンセサイザを指す。合成音に用いられる電気信号の規格の一種である CV(Control Voltage) 信号は、ユーロラック規格のモジュラシンセサイザの再評価により近年需要が高まっている。オーディオシンセサイザは 100 年以上続く歴史の中で数多くの製品が発売されてきた。一方で、映像を生成するビデオシンセサイザはオーディオシンセサイザと比べて多くの製品は出回っていない。そして、製品化されたものの多くはオーディオシンセサイザメーカーの製品である。そのため、出力時に音楽との連携機能がある一方、オーディオ入力により映像を生成できる製品は市場にはあまり流通していない。2010 年代になると、ゲーム会社である Atari から 1977 年に発売されたオーディオ信号の強弱や音色から映像を出力するビデオシンセサイザ『Atari Video Music』が再評価され始めた(ATARI 1977)。現在ではニューヨークの電子楽器メーカーである Critter and Guitari から、オーディオ・MIDI 入力に対し映像を生成するビデオシンセサイザ『EYESY』が発売されている(Critter & Guitari 2021)。

また、視覚芸術と音楽は、これまでの歴史の中で互いに影響を与え合っていた。アイザック・ニュートンは著書『光学』において、色とオクターブの 7 音階を関連付けることで白色光のスペクトルを 7 色に分類した(S, I, Newton. 1704)。20 世紀美術における表現主義

は、目に見えないものを描写する抽象絵画が特徴的である。中でも音楽に造詣の深いワシリー・カンディンスキーやパウル・クレーは音楽をモチーフにした作品や、音楽的なリズムと連続性を持った作品で有名である。カンディンスキーは著書において楽器の音色と色を具体的に対応させたうえで抽象芸術を論じている(W, Kandinsky. 1911)。一方で作曲家のジョン・ケージは、曲線と色彩で構成された記譜法による『Aria』(1958)などの図形譜を数多く残している(J, Cage. 1958)。ケージは、ヘンリー・デイヴィッド・ソロー著『日記』からドローイングを楽譜へ引用した『連歌』(1975-1976)においては、図形楽譜を用いる理由について「現代の楽器でも十八世紀の楽器でも、または東洋の楽器を加えても演奏できるようにするため」と言及している。また、ドローイングと実際の演奏との関係については「両者の論理的な結びつきを探することはあっても、詩的な関係は期待しない」とも残している。(J, Cage. 1977・1975) ニューウェイヴバンドのハルメンズは、アートワークを手掛けた作家である太田螢一を準メンバーとしている。

近年では、『耳をすまして－美術と音楽の交差点』(2011 茨城県近代美術館)『絵と音の対話』(2012 東京国立近代美術館)『アートと音楽－新たな共感覚をもとめて』(2012-2013 東京都現代美術館)などの視覚芸術と音楽に焦点をあてた展示が開催された。このように視覚芸術と音楽の関連性において、筆跡や色に対する点と線や構図、装飾音や音色に対する楽式や和声・対位法のような形相対質料に似た構造を両者が含んでいることが背景にあると考えられる。

本研究においても、音と映像の両方に作用する点に着目し、ハードウェアを導入させたリアルタイムフィードバックシステムの制作を試みる。

2. 関連作品・研究

視覚芸術による楽曲生成においては、山本らが動画をもとにしたマルコフ過程による自動作曲システムを音楽心理学的アプローチで制作し、後に音楽的場面転換の閾値をユーザがリアルタイムに変更することでユーザ適応も可能にした(山本, 宝珍, 野宮 2009)(山本, 宝珍, 野宮 2010)。また、バンドシーンではニューウェーブバンドである P-MODEL が、フラクタル状のヤシの木の画像データをもとにトラックを作成した(P-MODEL 1995)。

音楽と関わりを持つ映像表現としては、ドイツのアニメーション作家であるオスカー・フィッシングはサイレント映画『Radio Dynamics』において、線や図形といったモチーフの動きや色の変化のみで音楽を表

現した(O, W, Fischinger 1942)。また、中村滋延は具体音の意味性を音楽と映像の両方に作用することで感覚様相のズレを描いていく手法を見出した(中村 1994)。そして、中村隆行らは4つのフィードバックユニットの相互作用から生じたノイズを出力し、3D 映像としてビジュアライズするシステムを制作した(中村, 柴山 2013)。

視覚と聴覚の統合において、1960 年以降のドイツでのミュージック・シアターでは、演奏者の演劇的運動が視覚的要素でありながらも音楽に組み込まれていくような表現の試みがあった(L, Nono 1960-61)。

3. コンセプト

3.1. 不確定性と感覚様相による不定形

不確定性を音楽に取り入れるために、色や映像によって変化する CV 信号を生成する。具体的には、モジュラーシンセサイザの入力端子 CV/GATE の電圧値を制御して、再現不可能性を有する CV 信号を出力する。

映像生成部では、縮小/縮小による拡張性を提示する機能を構築する。「木目を長いこと見続けていると徐々に見え方が変わってくことに気付く。咀嚼中にディスプレイを見ると画素だけが揺れているように見えることに気付く。インターネットの関与しない鏡の中にもヴァーチャルの世界は存在することを思い出す。」という筆者自身の感覚的経験から、感覚器官と空間への問題意識が生まれた。テクノロジーを用いた視覚芸術は自己言及の脆弱性を均したユーモアとしての拡張性を描くことしかできないのだろうか。各々の人体へと回帰して固有の感覚を暴露し、差異や視点に鋭敏になることで新たな拡張性を提示することはできないだろうかと考えている。

3.2. 相互フィードバック

文化はしばしば誤解や誤用を始点とするパラダイムシフトが起きてきたと解釈している。コンピュータが加速度的に誤解を促すことで新たな表現が生まれていくことを期待している。

4. 制作物の概要

4.1. システム構成

システムの構成を図1に示す。まず、CV 信号の初期値には、ユーザが物体をカラーセンサに近づけて取得した色情報を使用することでランダム性を持たせる。色情報は、M5 Stick C から M5 Stack Basic へと WebSocket 通信を介して送信される。M5 Stack Basic 上で色情報から CV 信号を生成し、オーディオシンセサイザへ出

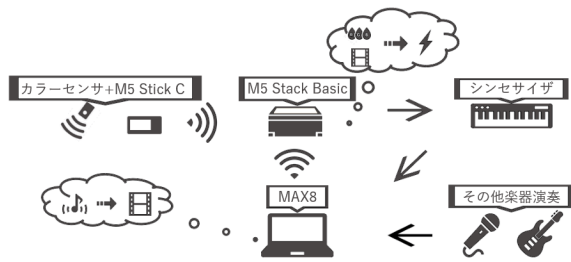


図 1: システム概要図

力される。CV 信号を入力したシンセサイザを含めた楽曲演奏全体のオーディオ信号を Cycling'74 MAX に入力し、特徴量を抽出する。次に、オーディオ信号の特徴量を用いてあらかじめ作成したオブジェクトを制御する。そして、生成した映像の特徴量を抽出し Node.js を用いた WebSocket 通信によって MAX から M5 Stack Basic へ送信される。映像の特徴量を用いて CV 信号を生成し、以降 CV 信号と映像が相互にフィードバックする。ユーザは物体の色情報をカラーセンサに検出させた後、CV 信号を入力したシンセサイザや他の楽器を演奏し、演奏全体のオーディオ信号を MAX に入力することを想定している。M5 Stack Basic をサーバとした WebSocket 通信によるそれぞれの双方向通信を実現している。

4.2. 使用特徴量

制作物において使用される色情報の特徴量は、赤・緑・青・明度である。オーディオ信号の特徴量は、音圧・周波数成分・スペクトル重心・SNR・調性・BPM である。一般的に SNR とは通信信号等の受信信号と背景雑音のレベルの差を表す値である (Cisco)。本研究では、振幅最大値を取る周波数のレベルとその他楽器とのレベルの差として使用する。映像の特徴量は、彩度平均・明度平均・コントラスト・画像重心の座標・オプティカルフロー・物体の示す向き（傾き）である。

5. 色情報による CV 信号の生成

5.1. 色情報による CV 信号の生成

M5 Stick C（入力部）及び M5 Stack Basic（出力部）を使用し小型化とステージ上でのユーザの可動範囲の広さを実現する。システムの詳細な構成を図 2 に示す。処理手順は以下の通りである。

1. M5 Stick C に接続したカラーセンサ (TCS34725) から取得した値を WebSocket 通信によって Json 形式で M5 Stack Basic（出力部）へ送信

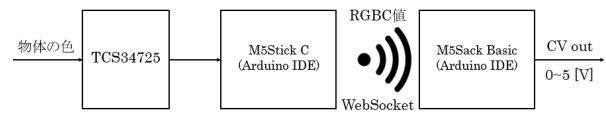


図 2: ランダム CV ジェネレータの構成

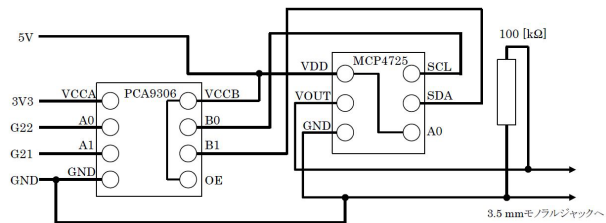


図 3: M5 Stack Basic に実装した DAC 回路

2. 連続的に取得した RGB 値を I2C バス用双方向電圧レベル変換モジュール (PCA9306 搭載) によって 0~5 [V] の出力電圧に対応させ、DA コンバータ (MCP4725) 搭載モジュールを用いて 12bit DA 変換
3. M5 Stack Basic とシンセサイザ(本研究では BASTL INSTRUMENTS KASTLE v1.5 を使用) をパッチケーブルで接続し、M5 Stack Basic から CV 信号を入力

以上を 20 シーケンス分行う。M5 Stick C においては、カラーセンサから検出された RGB 値を 500 ミリ秒ごとにディスプレイ上に表示し、M5 Stack Basic へ送信される。M5 Stack Basic においては、常時受信した RGB 値をディスプレイ上に表示させ、色情報出力電圧に対応させた。CV 信号の出力において、ユーロラックモジュラーシンセサイザとの互換性を考慮し、3.5 [mm] モノラルジャックを用いて実装した。また、M5 Stack Basic に実装した DAC 回路を図 3 に示す。動作している通信部の様子を図 4 に示す。

5.2. 色情報と CV 信号の対応関係

対応関係については図 5 及び式 1 に示す。本研究では経時変化を持つエンベロープ CV が出力される。また、CV 信号をピッチに入力して演奏する際は、12 平均律に基づいて設計をした方がより実用的である。しかし、本研究ではランダム性を重視し、ドローンサウンドが特徴的なシンセサイザを使用するため微分音も発生するように設計した。

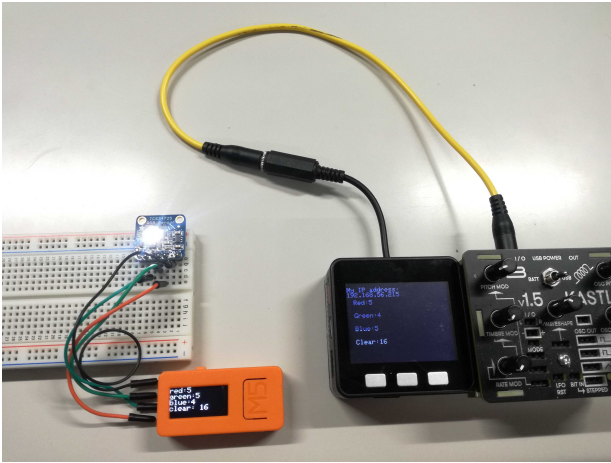


図 4: 動作している通信部の様子

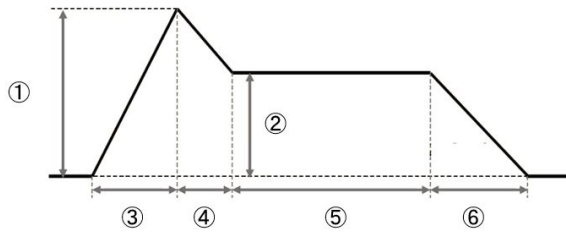


図 5: RGB 値と出力電圧の対応

$$\begin{aligned}
 \text{Attack volt} &= B \times 70 \times 1.22 \text{ [mV]} \\
 \text{Sustain volt} &= (R \times 70 - C \times 3) \times 1.22 \text{ [mV]} \\
 \text{Attack time} &= \frac{B \times 70}{G \times 10} \text{ [msec]} \\
 \text{Decay time} &= \frac{C \times 3}{G} \text{ [msec]} \\
 \text{Sustain time} &= \frac{R + G + B}{2} \text{ [msec]} \\
 \text{Release time} &= \frac{(B \times 70 - C \times 3) \times 4}{G + B + 80} \text{ [msec]}
 \end{aligned} \tag{1}$$

6. オーディオ信号による映像の生成

6.1. オーディオ信号の特徴量抽出

本制作物において使用するオーディオ信号の特徴量は、音圧・周波数成分・スペクトル重心・SNR・調性・BPMである。中でも、調性とBPMはMAX上での分析ではなくユーザー自身による入力を想定した。また、本来SNRは通信信号等の受信信号と背景雑音のレベルの差を表す値である。本研究では、振幅最大値を取る周波数のレベルとその他楽器とのレベルの差として

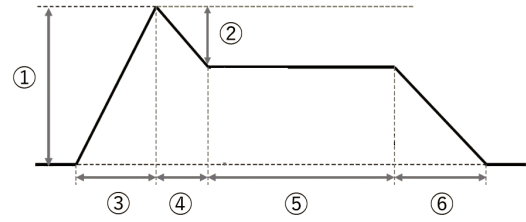


図 6: 映像の特徴量と出力電圧の対応

使用した。スペクトル重心 $f_c(t)$ は、以下の式 2 を用いて算出した。

$$f_c(t) = \frac{\sum_{f=0}^{\frac{f_x}{2}} X(t, f) \cdot f}{\sum_{f=0}^{\frac{f_x}{2}} X(t, f)} \tag{2}$$

7. 映像による CV 信号の生成

7.1. 映像の特徴量抽出

本制作物において使用する映像の特徴量は、彩度平均・明度平均・コントラスト・画像重心の座標・オプティカルフロー・物体の示す向き(傾き)である。画像重心の座標においては y 座標を、オプティカルフローにおいては x 方向・ y 軸方向の絶対値の平均を用いる。

7.2. 映像による CV 信号の生成

映像の特徴量を MAX から M5 Stack Basic へ WiFi を介した WebSocket 通信で常時送信する。M5 Stack Basic では、受信した特徴量をディスプレイ上に表示させ、対応した電圧値を出力する。特徴量と CV 信号の対応関係については図 6 及び以下の式 3 に示す。

$$\begin{aligned}
 \text{Attack volt} &= \text{brightness} \times 16 \times 1.22 \text{ [mV]} \\
 \text{Sustain volt} &= \text{brightness} \times 16 - \text{contrast} \times 8196 \times 1.22 \text{ [mV]} \\
 \text{Attack time} &= \frac{\text{brightness} \times 51.2}{8000 |\tan(2\pi - \text{moment})|} \text{ [msec]} \\
 \text{Decay time} &= \frac{\text{contrast} \times 2}{|\tan(2\pi - \text{moment})|} \text{ [msec]} \\
 \text{Sustain time} &= \frac{\text{centroid} \times (8 - \text{optical flow})}{40} \text{ [msec]} \\
 \text{Release time} &= \text{saturation} \times 0.5 \text{ [msec]}
 \end{aligned} \tag{3}$$

表 1: 取得された色情報

赤	緑	青	明度
56	23	17	97
57	30	23	112
56	28	21	108
55	27	20	105
55	26	20	103

表 2: 取得されたオーディオ信号の特徴量

音圧	スペクトル重心 [Hz]	最大振幅値	平均振幅値
0.64	877.58	3.56	1.69
0.63	953.59	1.70	1.66
0.57	1023.47	1.70	1.48
0.59	1075.82	1.92	1.72
0.85	1101.61	2.40	1.56

表 3: 取得された映像の特徴量

彩度平均	明度平均	コントラスト	画像重心の座標	オブティカルフロー	物体の示す向き [rad]
118.76	46.42	0.34	279.65, 285.02	6.69, 6.61	6.15
118.56	34.44	0.33	281.10, 287.64	6.64, 6.62	5.87
123.26	29.76	0.33	278.60, 280.85	6.70, 6.60	5.86
123.41	25.78	0.32	278.75, 280.27	6.64, 6.64	6.27
127.02	22.25	0.31	273.52, 280.43	6.76, 6.65	5.95

8. 動作結果と考察

8.1. 動作結果

システムはリアルタイムに動作した。(現在は通信部を再度開発中) デモンストレーション時の演奏と映像のデータを https://www.dropbox.com/sh/xjh03ewft4553n/AAB80hqRkKnwkYwycA9f3MT_a?dl=0 に示す。

取得された色情報の一例を表 1 に示す。色情報によって生成された CV 信号を正弦波のピッチに入力した際のスペクトログラムを図 7 に示す。M5 Stack Basic の起動後 CV 信号が出力されるまでに約 3 秒を要した。取得された振幅スペクトルを除くオーディオ信号の特徴量の一例を表 2 に示す。

また、振幅スペクトルを図 8 に示す。生成された映像の様子を図 9・10 に示す。再生された映像のフレームレートは約 16 fps であった。取得された映像の特徴量の一例を表 3 に示す。映像の特徴量のよって生成された CV 信号を正弦波のピッチに入力した際のスペクトログラムを図 11 に示す。

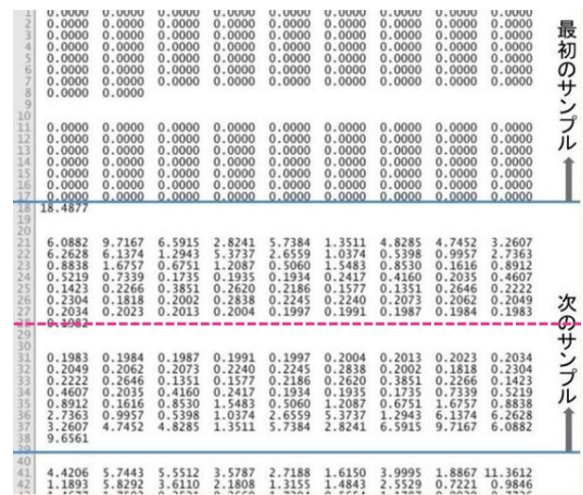


図 8: 取得された振幅スペクトル

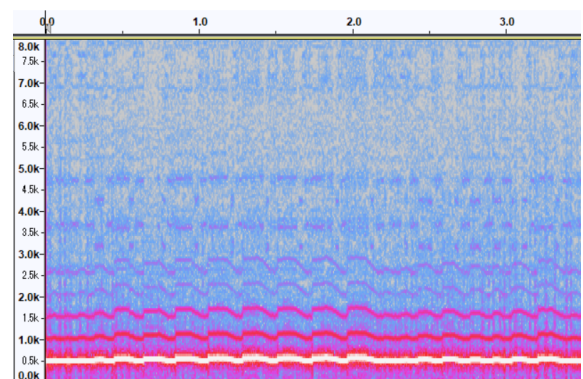


図 7: 色情報による CV 信号をピッチに入力した際のスペクトログラム

8.2. 考察

• CV 信号の生成

CV 信号の抑揚が少なく、変化の感じにくさが見られた。パラメータの重みの再検討が必要である。

• M5 Stack Basic における通信

実際に演奏者としてデモンストレーションを行った際に、M5 Stack Basic を起動するタイミングを他の楽器と合わせることが困難に感じられた。最初に M5 Stack Basic を起動すると演奏が始まる前に色情報取得の 20 シーケンス分が終わってしまい、遅めに起動すると演奏の開始に間に合わなくなってしまう。演奏開始前に M5 Stack Basic を起動し、通信を開始した状態でユーザが任意のタイミングでボタンを押すことでシーケンスが開始するように設計する必要がある。

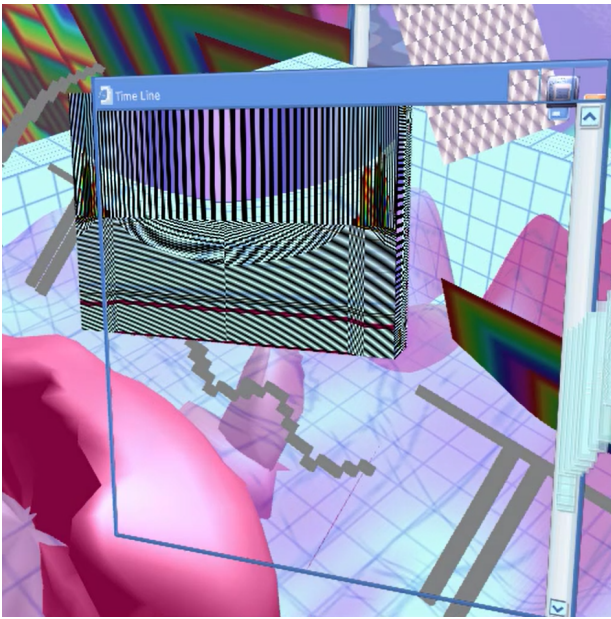


図 9: 生成した映像の様子①

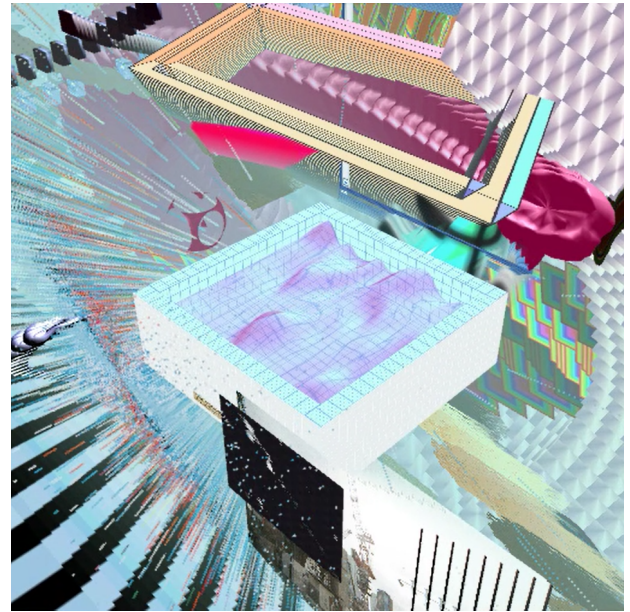


図 10: 生成した映像の様子②

- ユーザによるオーディオ信号と映像の関連付け
本研究においては制作者によってオーディオ信号と映像の関連付けを行ったが、ユーザによる使用を想定して、特徴量の関連付けを変更できる機能の実装が必要である。また、特徴量の変化が収束し変化が小さくなるループとなった場合には、収束から抜け出す仕組みの検討が必要である。
- 評価方法の検討
本研究においては制作者によるデモンストレーションの動作結果をもとに評価を行ったが、第三者による評価方法を検討する必要がある。

9. まとめ

色情報から始まり、音と映像が相互に影響し合い生成されるシステムを作成し、実装した。オーディオ信号の音圧・周波数成分・スペクトル重心・SNR・調性・BPMと映像の彩度・色相・明度・オブティカルフロー・画像重心・グラフィックの形状の指す方向を特徴量として使用した。このシステムを用いて、実際にデモンストレーションとして楽曲演奏を行った。その結果、システムは意図した通りにリアルタイムに動作した。今後は、演奏機器としての使いやすさを向上させるために、GATE信号の実装及び演奏開始用スイッチを実装する。そして、オーディオ信号と映像の関連付けをユーザによって設定できる機能を実装する。音楽的な使いやすさとコンセプトの補強として、今後は時間縮小錯覚の考え方を取り入れて、ある程度 BPM に対応

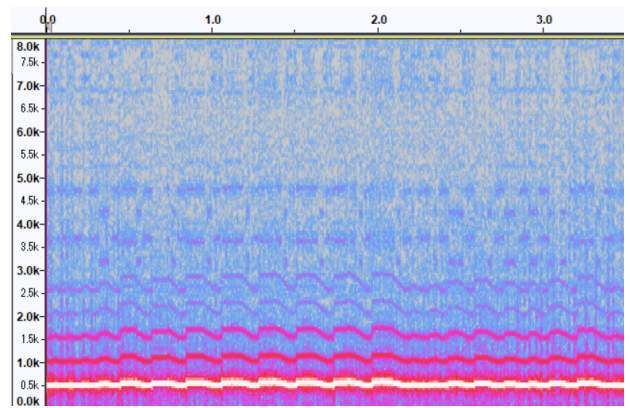


図 11: 映像の CV 信号をピッチに入力した際のスペクトログラム

する GATE 信号も実装する予定である。また、ユーザによる評価方法の検討が今後の課題である。

10. 参考文献

- ATARI Museum. 『Model #240』 2021 年 7 月 8 日閲覧. (<http://www.atarimuseum.com/videogames/dedicated/videomusic/videomusic.html>)
- Critter & Guitari. 『EYESY』 2021 年 7 月 8 日閲覧. (<https://www.critterandguitari.com/eyesy>)
- ICON. 『ビデオ・シンセサイザーの世界 ～ World of Video Synthesizer 001: ビデオ・シンセサ

イザーの基礎知識』 2021 年 1 月 25 日閲覧。
(<http://icon.jp/archives/17680>)

Sir Isaac Newton(島尾永康訳). 1983. 『光学』 東京
岩波書店.

Wassily Kandinsky(西田秀穂訳). 1979. 『カンディ
ンスキー著作集 1 改訂版 抽象芸術論—
芸術における精神的なもの—』 美術出版社.

白石美雪. 2009. 『ジョン・ケージ—混沌ではなく
アナーキー』 武蔵野美術大学出版局.

John Cage. 2009(小沼純一編・伊藤制子引用箇所訳).
『ジョン・ケージ著作選』 ちくま学芸文庫.

Y, Kato., S, Kurabayashi. 2014. “A Cross-Media
Music Retrieval System by Converting Color
Changes into Tonal Changes” in *IEEE 15th In-
ternational Conference on Mobile Data Man-
agement, Brisbane, QLD, 2014*, IEEE, Tokyo,
Japan, pp. 349-352, doi: 10.1109/MDM.2014.
51.

長嶋 洋一. 1995. 「マルチメディア Computer Mu-
sic 作品の実例報告」『日本音響学会秋季研究発
表会講演論文集』情報処理学会研究報告 Vol.97,
No.71(94-MUS-11).

NTT インターコミュニケーション・センター
[ICC]. 『テクノ・アートの総決算 第3回リ
ヨン現代美術ビエンナーレ』 2021 年 2 月 2
日閲覧. ([https://www.ntticc.or.jp/pub/
ic_mag/ic016/lyon/lyon_j.html](https://www.ntticc.or.jp/pub/ic_mag/ic016/lyon/lyon_j.html))

山本 敏生, 宝珍 輝尚, 野宮 浩揮. 2009. 「動画をも
とにした自動作曲」『平成 21 年度情報処理
学会関西支部大会論文集』 C-22.

山本 敏生, 宝珍 輝尚, 野宮 浩揮. 2010. 「動画をも
とにした自動作曲のユーザ適応について」『平
成 22 年度情報処理学会関西支部大会論文集』
E-01.

Cisco Meraki Japan. 『無線の基礎：信号対
雑音比 (SNR) と無線信号強度』 2020
年 7 月 26 日閲覧. ([https://www.cisco.
com/c/m/ja_jp/meraki/documentation/
mr/wifi-basics-and-best-practices/
wireless-fundamentals-signal-to-noise-
ratio-snr-and-wireless-signal-
strength.html](https://www.cisco.com/c/m/ja_jp/meraki/documentation/mr/wifi-basics-and-best-practices/wireless-fundamentals-signal-to-noise-ratio-snr-and-wireless-signal-strength.html))

中村 隆行, 柴山 拓郎. 2013. 「ラップトップ音楽
作品『Internal Sound Flow』の制作と表現」『先
端芸術音楽創作学会会報』 Vol.4, No.4.

11. 参考作品

Oskar Wilhelm Fischinger. 1942. 「Radio Dynamics」
P-MODEL. 1995. 「残骸の船 Saksit」 in 『舟』,
COCA-13083.

Luigi Nono. 1960-61 *Intolleranza (Handlung in 2
Teilen)*

中村 滋延. 1994 《WALK》コンピュータ音響とビ
デオのための映像音響詩

12. 著者プロフィール

野垣 美歩 (Miho NOGAKI)

1998 年生まれ. 東京電機大学大学院システムデザ
イン工学研究科在籍. 信号処理と芸術について研究を
している.

斎藤 博人 (Hiroto SAITO)

1995 東京電機大・工・電子卒, 1997 同大大学院修士
課程了. 2000 同大大学院博士後期課程単位取得退学.
同年東京電機大・工・電子助手, 情報環境学部を経て
現在東京電機大・システムデザイン工学部教授. 博士
(工学). 信号処理, 会話システム (特に話速変換) の
研究に従事.



この作品は、クリエイティブ・コモンズの表示 - 非営
利 - 改変禁止 4.0 国際 ライセンスで提供されていま
す。ライセンスの写しをご覧になるには、[http://
creativecommons.org/licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/) をご覧頂くか、
Creative Commons, PO Box 1866, Mountain View, CA 94042, USA ま
でお手紙をお送りください。