

## 光カメラ通信による映像と協調した音楽伝送法

## Color-Encoded Music Transmission Coordinated with Video via Optical Camera Communication

土屋 奈月<sup>1</sup> 中山 悠<sup>1</sup>  
Natsuki Tsuchiya Yu Nakayama

## 1. はじめに

近年、光カメラ通信(Optical Camera Communication: OCC)は、既存のディスプレイとカメラを用いた低コストな通信手段として注目されており、屋内測位やAR、IoTなど多様な応用が進んでいる[1][2]。一方、カメラのフレームレートや色再現性の制約により、連続的かつ高密度なメディアである音楽や映像の伝送は難しく、応用例は限られている。

従来の音楽伝送に関する研究では、LED光の強度変調によって音声信号(WAVファイルなど)をアナログ的に送信・再生する方式が提案されている[3]。これに対し本研究では、音楽の構成要素(音域・音価・音高)をRGB色信号に変調し、映像として提示・復号する方式を提案する。視覚と聴覚を統合した通信により、OCCの応用領域をマルチメディア伝送へ拡張する新たな可能性を示す。

## 2. 提案手法

## 2.1 概要

本研究では、音楽を構成する各音(音域、音価、音高)をN個のRGB色信号として表現し、1音につきN色の視覚信号を持つ構造を定義する。これらをディスプレイ上に時系列順に配置することで、映像による音楽表現を可能にする。各色信号は映像フレーム上に空間的に配置され、撮影したカメラ映像からフレームごとに復号されることで、音楽を再現する。本研究ではこのようなRGB色信号列による音楽表現を「色楽譜(color-encoded score)」と定義する。

## 2.2 音楽情報の色変換方式

本研究では、音楽を構成する各音をN個の色信号によって表現し、映像フレーム上に空間的に配置する。このN個の色には、それぞれ異なる音楽的意味(音域、音価、音高)を割り当て、1フレームにおける色の構成が1音に対応する「色楽譜」を形成する。本実装では $N=4$ とし、図1のようにフレームを4つの領域に分け、それぞれに音楽情報を割り当てる。カメラによる安定した色認識を実現するため、RGB空間上で明確に区別可能な赤、青、緑、黄、マゼンタ、シアン、白、黒の8色を用いる。

音高情報は1オクターブ内の12音を6音ずつに分割し、第3領域に前半(C~F)、第4領域に後半(F#~B)を割り当てている。音は常に1音ずつ順次再生されるため、同時に複数の音を鳴らすことはない。つまり、音高の指定は3・4領域のうちいずれか一方のみを使用し、他方は未使用(白または黒)とする。これにより、復号処理を簡素化するとともに、視認性と誤認識の抑制を両立している。また、未使用側に黒が指定された場合には、その音の音価を2倍に拡張する処理を加えることで、限られた色数の中でも音価情報を柔軟に表現可能とする。

定義された「色楽譜」は、映像フレーム上に順次配置され、カメラで撮影・復号されることで音楽が再現される。

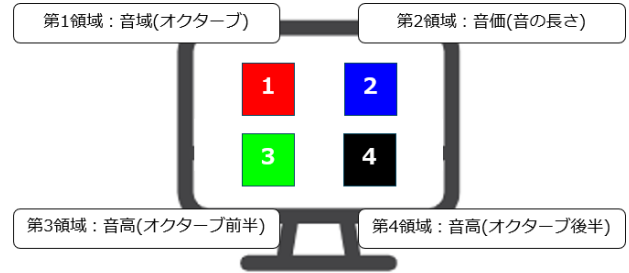


図1. 各領域と対応する音楽情報( $N=4$ のとき)

## 2.3 フレーム処理と音楽復元

本研究では、ディスプレイに表示した色信号をカメラで撮影し、各フレームから得られる色情報から音の再構成を行う。各時刻 $t$ においてカメラが取得したフレーム画像を、 $N=4$ の領域構造に従って分割し、各領域の平均色を抽出する。抽出された色情報は、次のようなベクトル形式で表現される:

$$c_t = (c_{1t}, c_{2t}, c_{3t}, c_{4t})$$

ここで、 $c_{it}$ はフレーム $t$ における第 $i$ 領域( $i=1,2,3,4$ )の平均色を表し、事前に定義された8色の色集合 $C$ のいずれかに分類される。

音楽情報の復号処理は、色情報ベクトル $c_t$ が2フレーム連続で一致した場合にのみ実行される。この判定により、一時的な色揺らぎや誤認識を抑制し、安定した復号が可能となる。また、フレームごとの内容に基づいて判断することで、カメラとディスプレイ間に明示的な時間同期を必要とせず、非同期的な構成でも信号の再現が可能となる。

色情報 $(c_1, c_2, c_3, c_4)$ は、それぞれ音域、音価、音高前半、音高後半に対応し、以下の手順で音楽パラメータに変換される。本稿では音高をMIDIノート番号で表現する。

まず、第1領域の色 $c_1$ は、相対的なオクターブ値 $o \in \{-3, -2, \dots, +3\}$ に写像される。これは写像 $f_{oct}(c_1)$ によって定義し、中央のド(C4)に対応するMIDIノート番号60を基準とした場合、基準音高 $B$ は次のように計算される:

$$B = 60 + 12 \times f_{oct}(c_1)$$

ここで、12は1オクターブあたりの半音数である。

次に、第2領域の色 $c_2$ は、音価に対応し、色に応じた拍単位の係数 $d = f_{dur}(c_2)$ を定義する。設定されたテンポBPMに対して、再生時間(秒)としての基準音価 $D_0$ は以下の式で与えられる:

$$D_0 = \frac{60}{BPM} \times f_{dur}(c_2)$$

さらに、第3または第4領域において「黒」が指定されている場合には、音価を2倍に拡張する処理を適用し、最終的な音長 $D$ を次のように定める:

$$D = \begin{cases} 2 \times D_0 (c_3 \text{ または } c_4 \text{ が黒の場合}) \\ D_0 \quad (\text{それ以外の場合}) \end{cases}$$

<sup>1</sup> 東京農工大学知能情報システム工学科

Department of Electrical Engineering and Computer Science,  
Tokyo University of Agriculture and Technology

最後に、第 3 領域の色  $c_3$  および第 4 領域の色  $c_4$  に対応する音高オフセットは、それぞれ  $f_{note1}$ ,  $f_{note2}$  によって与えられる。どちらか一方が有効な場合、そのオフセット値  $n \in \{0, 1, \dots, 11\}$  を用いて、最終的な MIDI ノート番号  $M$  を以下のように算出する：

$$M = B + n$$

ただし、両方の領域が無効(白または黒)である場合は、休止符とし、音の再生を一時停止する。

### 3. 評価実験

#### 3.1 実験環境と設定

本実験では、送信側に 3840×2160 ピクセル(60Hz)の液晶ディスプレイを用い、受信側にはソニーセミコンダクタソリューションズ製の IMX577 を使用した。IMX577 は最大 4056×3040 ピクセルの生データ撮影が可能であるが、デモザイク処理後のフルカラー画像は、原理上その 1/4 の空間解像度となる。解析にはこのカラー画像を使用した。

色分類への背景変動の影響を評価するため、2×2 ピクセル単位のランダム配色からなる背景を 3 秒毎に更新し、常に色揺らぎを伴う条件で実験を行った。この背景上に、2×2、4×4、8×8、16×16 ピクセルの信号領域を 1 つずつ重ね、領域には定義済みの 8 色のうちいずれかを割り当てた。領域サイズ、色ごとに 10000 回の色分類試行を実施し、分類精度を評価した。また、表示サイズの変更に応じてカメラの撮影条件を調整し、取得された画像内でも領域のサイズが一貫して表示サイズと対応するようにした。これにより、領域サイズと分類精度の関係を正確かつ統一的な条件下で評価可能とした。

#### 3.2 評価結果

図 2 の結果より、領域サイズが小さいほど分類精度が低下する傾向が顕現した。2 ピクセル表示では約 0.29 のエラー率を記録し、とくにマゼンタや白への誤認が顕著である。これは、色領域が小さいことで周囲の背景色と混色し、色相が崩れた影響と考えられる。一方、4 ピクセル表示ではエラー率は約 0.10 に抑えられ、青、黄、シアンなどの高彩度色においては比較的安定した分類が可能だった。また、8 ピクセルおよび 16 ピクセル表示では誤認が一切発生せず、すべての試行で正確に分類された。これらの結果から、8×8 ピクセル以上の表示サイズを確保することで、誤認のない高精度な色分類が可能であると示唆される。

#### 3.3 音楽伝送

前節の評価により、背景に揺らぎがある環境下でも 8×8 ピクセル以上の表示領域があれば、高精度な色分類が可能だと確認できた。本節では任意の楽曲を光信号として正しく伝送・復号できるかの評価実験を実施した。

表示映像には、図 3 に示す背景画像を使用した。これは夜間のイルミネーションや装飾的な照明を模したもので、色と配置にばらつきのある点光源が含まれる。画像内の各円形オブジェクトは直径 14 ピクセルで描画されており、内部に 8×8 ピクセルの正方形領域を確保できることから、前節で確認された色分類条件を満たす設計となっている。この画像を背景とし、RGB 色信号を 1 音あたり 1 フレームで構成した映像を重ねて再生した。映像全体は 1 楽曲に対

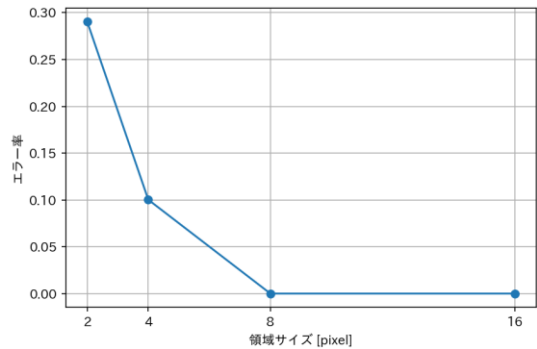


図 2. 領域サイズとエラー率の関係

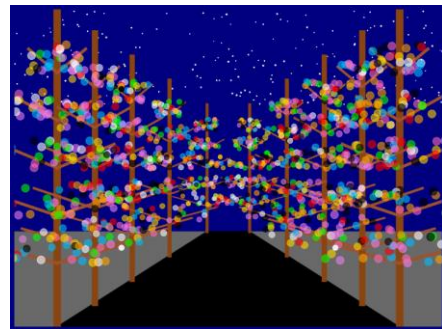


図 3. 音楽伝送に使用する画像

応するフレーム列として構成され、連続的にディスプレイ上に表示された。

撮影には IMX577 を用い、各フレームを逐次解析して色情報を抽出し、復元された音符系列を MIDI として再生する構成とした。その結果、楽曲全体が背景に依存せず完全に復号され、テンポや音階の乱れも生じなかった。このことから、本手法は背景に動的かつ多色な視覚情報が存在する環境下でも、視覚と聴覚を融合した安定した音楽伝送が可能であることが実証された。

### 4. まとめ

本研究では、RGB 色信号を用いて音楽情報を映像フレームに埋め込み、光カメラ通信により伝送・復元する手法を提案した。音域・音価・音高を 4 つの色に分担させる独自の構造を設計し、実際に表示・撮影された映像から音楽を正確に復号可能であることを示した。

評価実験では、背景色に変化する条件下でも、8×8 ピクセル以上の色領域を確保すれば高精度な色分類が可能であることを確認した。また、実際の楽曲を背景画像上に重ねて表示した実験においても、すべての音を正確に復元でき、本手法の実用性と安定性が示された。

#### 謝辞

本実験を進めるにあたり、サポートいただいたソニーセミコンダクタソリューションズ株式会社の海津俊様に感謝申し上げます。

#### 参考文献

- [1] N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends," *Physical Communication*, vol. 37, 100900, (2019).
- [2] P. Zhang, Z. Liu, X. Hu, Y. Sun, X. Deng, B. Zhu, and Y. Yang, "Constraints and recent solutions of optical camera communication for practical applications," *Photonics*, vol. 10, no. 6, pp. 1–19, (2023).
- [3] S. Sandoval-Reyes and A. Hernandez-Balderas, "Transmission of digital audio with visible light," *Research in Computing Science*, vol. 138, no. 1, pp. 61–68, (2017).