

ヴァイオリン演奏に基づくリアルタイム照明制御システム

福知山公立大学 情報学部情報学科

32245085 松永流翠

指導教員 橋田光代 准教授

提出日 2026年1月31日

改訂日 2026年2月7日

— 目次 —

1 はじめに	2	4.6 音響特微量と照明特微量の対応関係	11
2 関連研究	2	4.6.1 解析手法の概要	
2.1 音響信号に基づく照明演出の自動化	2	4.6.2 使用データ	
2.2 楽曲の印象・意味内容に基づく照明生成 . . .	2	4.6.3 照明特微量別の解析方法	
2.3 深層学習を用いた照明演出生成	2	4.6.4 音響特微量と照明色 (R 成分) の対応関係	
2.4 演奏者操作によるリアルタイム照明制御 . .	2	4.6.5 音響特微量と照明色 (G 成分) の対応関係	
2.5 身体動作と照明演出の連動	3	4.6.6 音響特微量と照明色 (B 成分) の対応関係	
2.6 本研究の位置づけ	3	4.6.7 音響特微量と照明の明るさの対応関係	
3 提案手法	3	4.6.8 音響特微量と照明の点滅速度の対応関係	
3.1 音響特微量に基づく照明制御	3	5 リアルタイム照明制御システムの実装	13
3.2 運弓動作に基づくライトの上下動作制御 . .	3	5.1 システム設計の基本方針	13
4 照明予測モデル	4	5.2 ハードウェア構成	13
4.1 データベースの構築	4	5.3 ソフトウェア構成	14
4.1.1 映像データの収集		5.4 全体処理フロー	14
4.1.2 音響特微量の抽出		5.5 音声入力処理	14
4.1.3 照明特微量の抽出		5.6 音響特微量抽出	14
4.2 照明予測モデル(仮)の学習	7	5.7 演奏動作取得	14
4.2.1 学習データの概要		5.7.1 ハードウェア配置	
4.2.2 入力変数(音響特微量)		5.7.2 データ通信方式	
4.2.3 出力変数(照明特微量)		5.7.3 運弓方向判定アルゴリズム	
4.2.4 学習データとテストデータの分割		5.7.4 リアルタイム処理	
4.2.5 モデルの構造と学習方法		5.8 照明特微量推定	15
4.2.6 学習の実行とモデル保存		5.9 DMX 信号生成と照明制御	15
4.2.7 照明予測モデル(仮)の位置づけ		5.9.1 DMX512 プロトコルの概要	
4.3 MIDI データの生成と特微量抽出	8	5.9.2 DMX ハードウェアインターフェース	
4.3.1 MIDI データセットの概要		5.9.3 照明特微量から DMX 値への変換	
4.3.2 MIDI データの音声変換		5.9.4 複数照明機器への同時制御	
4.3.3 音響特微量の抽出		5.9.5 演奏動作情報との統合制御	
4.3.4 1 秒単位への集約処理		5.9.6 リアルタイム出力制御と更新周期	
4.3.5 照明予測モデル(仮)による照明特微量推定		6 実演	16
4.3.6 CSV 形式での出力		6.1 実演環境	16
4.4 MIDI データを含めた照明予測モデルの学習	10	6.2 実演曲目	16
4.4.1 MIDI データ導入の目的		6.3 実演中の照明変化の様子	16
4.4.2 学習データセットの構成		7 考察	18
4.4.3 特微量構成と前処理		7.1 提案手法の有効性に関する考察	18
4.4.4 学習手法		7.2 実演環境における制約と課題	18
4.4.5 学習データとテストデータの分割		8 結び	18
4.4.6 学習の実行とモデル保存			
4.4.7 照明予測モデル(本)の位置づけ			
4.5 照明予測モデルの評価	10		
4.5.1 評価指標: 決定係数 (R^2 スコア)			
4.5.2 評価方法			
4.5.3 評価結果			

— 図目次 ——

1	ライトの配置イメージ(上面図)	3
2	上げ弓 (UP)	3
3	下げ弓 (DOWN)	4
4	照明予測モデル構築のためのデータ生成フロー	4
5	*千本桜* 演奏映像	4
6	*Sonata* 演奏映像	4
7	*スプラッシュ* 演奏映像	5
8	抽出する音響特徴量の内訳	5
9	ゼロ交差率 (ZCR) の定義式	5
10	スペクトル重心 (Spectral Centroid) の定義式	5
11	スペクトル帯域幅 (Spectral Bandwidth) の定義式	6
12	スペクトルフラットネス (Spectral Flatness) の定義式	6
13	音量 (RMS) の定義式	6
14	MFCC の一次差分および二次差分の定義式	6
15	決定係数 (R^2) の定義式	10
16	決定係数 (R^2) の結果	11
17	照明色 (R 成分) に対する音響特徴量	11
18	照明色 (G 成分) に対する音響特徴量	12
19	照明色 (B 成分) に対する音響特徴量	12
20	照明の明るさに対する音響特徴量	13
21	照明の点滅速度に対する音響特徴量	13
22	DMX 対応照明機器	14
23	リアルタイム照明制御システム構成図	14
24	照明上下動作	16
25	実演機材配置図(上面図)	16
26	実演曲目『紅蓮華』楽譜	17
27	実演中の照明変化の様子 (イントロ部分)	17
28	実演中の照明変化の様子 (A メロ～C メロ)	17
29	実演中の照明変化の様子 (サビ部分)	17
30	実演中の照明変化の様子 (サビ部分)	18

— 表目次 ——

1	動画メタ情報 (千本桜)	4
2	動画メタ情報 (ソナタ第三番)	4
3	動画メタ情報 (スプラッシュ)	5
4	各映像の音響特徴量の CSV データ (一部抜粋)	7
5	千本桜の RGB 値の CSV データ (一部抜粋)	8
6	ソナタ第 3 番の RGB 値の CSV データ (一部抜粋)	8
7	スプラッシュの RGB 値の CSV データ (一部抜粋)	8
8	千本桜の Brightness の CSV データ (一部抜粋)	8

9	ソナタ第 3 番の Brightness の CSV データ (一部抜粋)	9
10	スプラッシュの Brightness の CSV データ (一部抜粋)	9
11	千本桜の点滅速度の CSV データ (一部抜粋)	9
12	ソナタ第 3 番の点滅速度の CSV データ (一部抜粋)	9
13	スプラッシュの点滅速度の CSV データ (一部抜粋)	9

1. はじめに

舞台演出やライブパフォーマンスにおいて、照明は音楽表現を視覚的に拡張し、観客の没入感を高める重要な要素である。照明の色や明暗、動きは演奏の意図や感情を強調し、音響だけでは伝わりにくい表現を補完する。しかし、従来の照明制御では照明オペレーターが手動で操作する必要がある。

こうした背景のもと、本研究ではヴァイオリン演奏における音響信号と身体動作を組み合わせたリアルタイム照明制御システムの構築を目的とする。具体的には、マイクから入力されたヴァイオリン音響信号に対して、スペクトル重心、スペクトル帯域幅、MFCC、ZCR などの特徴量を抽出し、機械学習モデルにより照明の RGB 値、明るさ、点滅速度といった照明特徴量を予測する。また、スマートフォンのモーションセンシングアプリ ZIG-SIM[1] を用いて演奏者のヴァイオリン演奏における「弓の上下動作」を角速度として取得し、照明の上下の動きに直接反映させている。

本研究を開始するにあたり、筆者自身の「自分のヴァイオリン演奏だけで照明を直接動かしてみたい」という個人的関心が着想となっている。ヴァイオリンを対象とした理由として、筆者自身が演奏経験を持つことから、演奏者の立場からシステムの反応性や表現性を直接評価できる利点がある。

以上より本研究では、自身のヴァイオリン演奏の音響・身体動作解析に基づく照明特徴量生成を統合したリアルタイム照明制御システムを開発し、実演することを目的とする。

2. 関連研究

音楽ライブやコンサートにおける照明演出は、観客の没入感や一体感を高める重要な要素であり、従来は照明デザイナーやオペレータの経験や感性に基づいて設計されてきた。しかし、小規模公演や即興性の高い演奏では十分な準備時間や人員を確保することが難しく、音楽情報処理技術を用いた照明演出の自動化が有効な手段として注目されている。

2.1 音響信号に基づく照明演出の自動化

音響信号に基づく照明制御に関する研究として、月東らは楽曲音源からリズムの強調箇所を検出し、それを照明演出へ反映する手法を提案している [2]。この研究では、楽曲をドラム、ベース、ギター、ボーカルの各パートに音源分離し、それぞれのオンセットエンベロープを算出した上で、複数パートで同時に強いオンセットが観測されるタイミングを「キメ」と定義している。検出されたキメを基点として、照明のフラッシュや明滅といった瞬間的な演出を付与し、主観評価実験によりリズムの一体感や迫力の向上

を確認している。

この手法は高い時間解像度でリズム的アクセントを捉えられる点に特徴がある一方で、照明制御はキメの有無に依存した離散的なものに留まっており、演奏中の表現変化を連続的に反映することは難しい。また、演奏者の身体的動作は入力として扱われていない。

2.2 楽曲の印象・意味内容に基づく照明生成

楽曲全体の印象や意味内容を照明演出に反映する研究も多く報告されている。神野らは、歌詞および曲調の印象に基づいて照明演出を自動生成するシステムを提案している [3]。

この研究では、歌詞を形態素解析した上で分散表現ベクトルに変換し、言語イメージスケールを用いて照明色候補を選択する。さらに、楽曲の長短調やテンポに基づいて色の優先度や配置を決定し、BPM や歌詞のネガティブ・ポジティブ度を用いて明度を調整することで、照明の色、配置、明るさを総合的に決定している。この手法により、照明デザイナーの感性に依存しない一貫性のある演出が実現されている。

一方で、照明変化は小節やフレーズ単位といった比較的長い時間スケールで行われるため、演奏中の瞬間的な表現変化や演奏者の身体動作を直接反映することは想定されていない。

2.3 深層学習を用いた照明演出生成

近年では、照明演出を生成タスクとして捉え、深層学習モデルを用いて照明パラメータを自動生成する研究も行われている。月東らの研究内でも言及されている Zhao らの手法では、楽曲特徴量系列を入力として照明の色や明るさの時系列データを生成する生成モデルが提案されている [4]。

このようなアプローチは、表現力や多様性の高い照明演出を実現できる可能性を示しているが、学習データへの依存度が高く、生成過程の解釈が困難であるという課題がある。また、演奏者の身体動作は入力として扱われていない。

2.4 演奏者操作によるリアルタイム照明制御

演奏者自身が照明演出に直接関与する研究として、瀧口らはドラム音や MIDI フットコントローラーを用いて演奏中に照明を操作可能なシステムを提案している [5]。この研究では、バスドラム音の音量が閾値を超えた際に照明を点灯させる手法や、足元のコントローラー操作によって照明の色や輝度、ストロボ効果を変更する仕組みが実装されている。

この手法により、即興演奏に柔軟に対応できる照明演出が実現されている一方で、演奏者が照明操作を意識する必要があり、演奏負荷が増加する可能性がある。また、演奏動作そのものを自動的に取得・利用する仕組みは含まれて

いない。

2.5 身体動作と照明演出の連動

演奏者の身体動作と照明演出を連動させる研究として、浅田らはオンラインライブ空間において、演奏者の身体情報を用いて照明や振動装置を制御するシステムを提案している [4]。この研究では、ギター演奏者の腕の角度に応じてムービングライトの向きを変化させるなど、身体動作を照明の物理的な動きに対応付けることで、演奏者と観客の一体感を高めている。

しかし、音響特徴量と身体動作を統合的に扱う設計にはなっておらず、音楽表現と照明表現の対応関係は限定的である。

2.6 本研究の位置づけ

以上の関連研究から、音響特徴量に基づく照明制御、楽曲印象に基づく照明生成、演奏者操作型および身体動作連動型の照明演出はそれぞれ発展してきたものの、音響情報と演奏者の身体動作を同時に統合し、連続的かつリアルタイムに照明を制御する研究は十分に行われていないことが分かる。

本研究では、ヴァイオリン演奏音から抽出した音響特徴量に基づいて照明の色や明るさ、点滅速度を制御すると同時に、演奏者の運弓動作を照明の上下動作として直接反映することで、音楽表現と身体表現が一体となった照明演出を実現する点に特徴がある。

3. 提案手法

本研究では、ヴァイオリン演奏に応じてステージ上の照明がリアルタイムに変化する照明制御システムを提案する。



図 1: ライトの配置イメージ (上面図)

本システムは、特に盛り上がりのあるヴァイオリン楽曲

を入力として想定している。その理由として、筆者自身が舞台演奏においてエネルギー感や高揚感を伴う楽曲を演奏したいという意図を持っているためである。

本手法は、演奏者を中心地に四方向に配置した4台のライトを対象とし、演奏音の特徴量および演奏動作（運弓方向）を解析することで、ライトの色（RGB）、明るさ、点滅速度、上下動作を瞬時に制御する点に特徴がある。図1に示すように、ライトはステージの四隅に配置され、床から天井に向けて四方から演奏者に対して光が寄り集まるような演出を生み出す構成となっている。

以下、本システムの主要な2つの制御手法について記述する。

3.1 音響特徴量に基づく照明制御

まず、マイクから入力されたヴァイオリン演奏音を1秒ごとに解析し、音色・音量・スペクトル構造に関する複数の音響特徴量を抽出する。抽出される特徴量には、MFCC、RMS、スペクトル重心、スペクトルコントラストなどが含まれる。これらの特徴量は機械学習モデルに入力され、ライトの「RGB値（色）」、「Brightness（明るさ）」、「点滅速度」といった照明特徴量が予測される。

3.2 運弓動作に基づくライトの上下動作制御

本手法では、スマートフォンアプリ ZIG-SIM を利用し、演奏者の右腕の角速度をリアルタイムで取得し、弓の動作方向を推定する。

推定した運弓方向はライトの上下動作に直接割り当てられ、以下のように制御される。

- ・ 上げ弓 → ライトの光軸が上方向へ移動 (図 2)
- ・ 下げ弓 → ライトの光軸が下方向へ移動 (図 3)

これにより、演奏者の身体的アクションが照明の動きへと反映される。

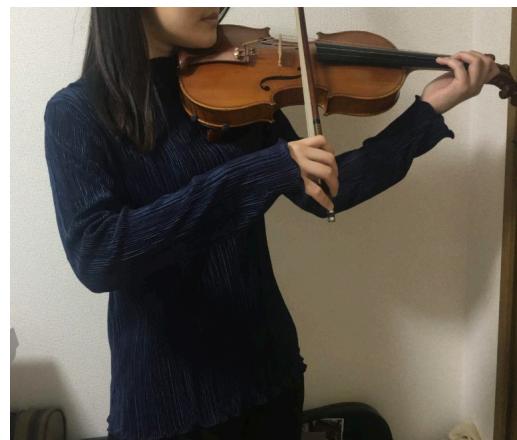


図 2: 上げ弓 (UP)

[6]



図 3: 下げ弓 (DOWN)

[6]

4. 照明予測モデル

本章では、ヴァイオリン演奏音から照明特徴量を予測する機械学習モデルの構築方法について詳述する。本研究では、音響特徴量を入力とし、それらの照明特徴量を予測するモデルを用いて照明の制御を行っている。最終的に構築されたモデルは、リアルタイムの演奏音を入力として照明を自動制御する「照明予測モデル」として機能する。図 4 に、照明予測モデル構築のためのデータ生成フローを示す。この図に基づき、各ステップを詳細に説明する。

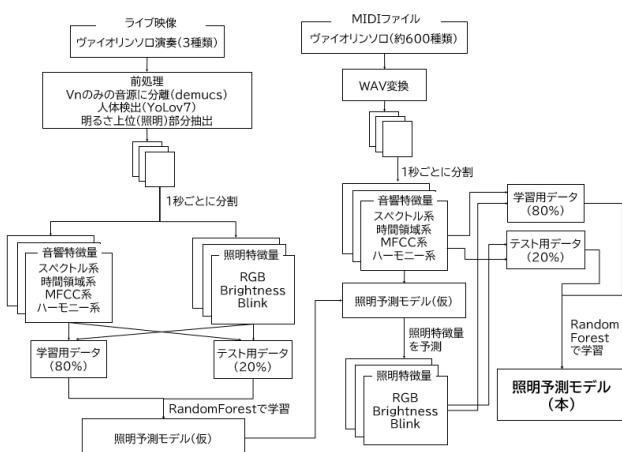


図 4: 照明予測モデル構築のためのデータ生成フロー

4.1 データベースの構築

まずは、モデルを学習させるためのデータベースの構築を行う。

4.1.1 映像データの収集

照明予測モデルの学習には、音響特徴量と対応する照明特徴量のペアが必要である。そのためのデータソースとして、実際のヴァイオリン演奏におけるライブ映像データを用意した。これらは実際の照明を伴う演奏映像である。各映像の詳細は以下の通りである。

ライブ映像データの 1 つ目は、ヴァイオリン演奏曲「千本桜 [7]」の演奏を記録した映像(図 5)である[8]。



図 5: *千本桜* 演奏映像

動画メタ情報を解析した結果、以下のような特性を持つことが確認された(表 1)。

表 1: 動画メタ情報(千本桜)

項目	値
フレームレート (fps)	29.85fps
総フレーム数	5940 フレーム
動画の長さ	約 199 秒

2 つ目のライブ映像データは、ヴァイオリンソロによるクラシック楽曲「ソナタ第三番 [9]」を演奏した映像(図 6)である[10]。



図 6: *Sonata* 演奏映像

動画メタ情報を解析した結果、以下のような特性を持つことが確認された(表 2)。

表 2: 動画メタ情報(ソナタ第三番)

項目	値
フレームレート (fps)	30.0fps
総フレーム数	2776 フレーム
動画の長さ	約 92.53 秒

3つ目のライブ映像データは、ヴァイオリンソロによる楽曲「スプラッシュ [11]」を演奏した映像(図 7)である[12]。



図 7: *スプラッシュ* 演奏映像

動画メタ情報を解析した結果、以下のような特性を持つことが確認された(表 3)。

表 3: 動画メタ情報(スプラッシュ)

項目	値
フレームレート (fps)	24.97fps
総フレーム数	5814 フレーム
動画の長さ	約 233 秒

これらの映像に対して、音響特徴量と照明特徴量を抽出するための処理を行い、照明予測モデルの学習データを生成する。

4.1.2 音響特徴量の抽出

(1) ヴァイオリン音源の分離

ライブ映像には、会場の反響音、観客ノイズ、伴奏音など、ヴァイオリン以外の音が含まれる。音響特徴量は不要な音が混入したまま学習するとモデルの精度劣化を招く。そこで音源分離モデル Demucs[13] を用い、音声からヴァイオリン成分を抽出した。

(2) ヴァイオリン特徴量の抽出

抽出したヴァイオリン音声に対して、1秒ごとに音響特徴量を算出した。特徴量の内訳は図 8 の通りである。本節では、その具体的な処理内容を述べる。

まず、音声ファイルを Python[14] の librosa[15] ライブライアリを用いて読み込む。この際、サンプリングレートは 44.1 kHz、チャンネル数はモノラルに統一する。読み込んだ波形データおよびサンプリングレートから音声全体の長さ(秒)を算出し、その長さに応じて 0 秒から終端まで 1 秒刻みの時間インデックスを生成する。

次に、生成した各時間インデックスに対して以下の処理を行う。

まずは各 1 秒セグメントごとに extractAdvancedFeatures 関数を用いて音響特徴量を算出する。

以下、各特徴量の算出方法を詳述する。

カテゴリ	特徴量名	説明
時間領域系	ZCR	波形がゼロをまたぐ回数。音の荒さ・バーカッショを示す。
	RMS	振幅の2乗平均の平方根。音の強さ(音量)を表す。
スペクトル系	Spectral Centroid	スペクトルの重心。値が高いほど高音成分が多い。
	Spectral Bandwidth	スペクトルの広がり。音の豊かさ・ノイズ感を表す。
	Spectral Contrast	各周波数帯の強弱差。倍音構造やハーモニクスの特徴を示す。
	Spectral Flatness	スペクトルの平坦性。1に近いとノイズ的、0に近いとトーン的。
	Spectral Rolloff	高周波エネルギーの95%が含まれる周波数。高音の強さ指標。
MFCC 系	MFCC	人間の聴覚特性を模倣した周波数領域の要約特徴。低次は音色、高次はノイズ傾向を示す。
	ΔMFCC	MFCC の時間変化速度。動的特徴を捉える。
	Δ^2MFCC	MFCC 变化の加速度。滑らかさ・発音の安定性を表す。
ハーモニー系	Chroma	12 音階(C~B)ごとの強度。音楽的な調性・和音情報を表す。
	Tonnetz	音楽理論に基づく和声空間座標。調性や転調関係を反映。

図 8: 抽出する音響特徴量の内訳

(2-1)ゼロ交差率 (ZCR)

ゼロ交差率 (ZCR) は、時間領域信号が正から負、または負から正へ符号反転する回数を表す特徴量であり、音の粗さやノイズ成分の多さを示す指標である。フレーム長を N、時間領域信号を $x[n]$ とすると、ZCR は次式で定義される。

$$ZCR = \frac{1}{2N} \sum_{n=1}^N |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|$$

図 9: ゼロ交差率 (ZCR) の定義式

ここで $\text{sgn}(\cdot)$ は符号関数である。本研究では、librosa の librosa.feature.zero_crossing_rate() を用いて ZCR を算出し、各フレームにおける平均値を特徴量として採用した。

(2-2)スペクトル重心 (Spectral Centroid)

スペクトル重心は、周波数スペクトルにおけるエネルギー分布の重心位置を示す特徴量であり、音の明るさや鋭さと関連がある。短時間フーリエ変換 (STFT) によって得られた振幅スペクトルを $|X(k)|$ 、周波数ビンを k とすると、スペクトル重心 C は次式で定義される。

$$C = \frac{\sum_k f_k |X(k)|}{\sum_k |X(k)|}$$

図 10: スペクトル重心 (Spectral Centroid) の定義式

本研究では、librosa の librosa.feature.spectral_centroid() を用いてスペクトル重心を算出し、各フレームにおける平均値を特徴量とした。

(2-3)スペクトル帯域幅 (Spectral Bandwidth)

スペクトル帯域幅は、スペクトル重心を中心とした周波数成分の広がりを表す特徴量であり、音色の拡散度合いを示す指標である。次数 p=2 の場合、スペクトル帯域幅 B は次式で表される。

$$B = \left(\frac{\sum_k |f_k - C|^p |X(k)|}{\sum_k |X(k)|} \right)^{1/p}$$

図 11: スペクトル帯域幅 (Spectral Bandwidth) の定義式

本研究では、librosa の librosa.feature.spectral_bandwidth() を用いて算出した値の平均を特徴量として用いた。

(2-4)スペクトルコントラスト (Spectral Contrast)

スペクトルコントラストは、周波数帯域ごとのピーク成分と谷成分の差を表す特徴量であり、倍音構造の明瞭さや音色のコントラストを表現する。librosa の librosa.feature.spectral_contrast() を用いて複数帯域に分割したスペクトルコントラストを算出し、その平均値を特徴量として使用した。

(2-5)スペクトルフラットネス (Spectral Flatness)

スペクトルフラットネスは、スペクトルの平坦さを示す特徴量であり、音が純音的かノイズ的かを判別する指標である。振幅スペクトルを $|X(k)|$ とすると、スペクトルフラットネス F は次式で定義される。

$$F = \frac{\left(\prod_k |X(k)|\right)^{1/K}}{\frac{1}{K} \sum_k |X(k)|}$$

図 12: スペクトルフラットネス (Spectral Flatness) の定義式

本研究では、librosa の librosa.feature.spectral_flatness() を用いて算出した。

(2-6)スペクトルロールオフ (Spectral Rolloff)

スペクトルロールオフは、全スペクトルエネルギーの一割合が含まれる周波数を示す特徴量である。本研究では、全エネルギーの 85% を含む周波数をロールオフ周波数として定義し、librosa の librosa.feature.spectral_rolloff() を用いて算出した。

(2-7)音量 (RMS)

音量 (RMS) は、音声信号の振幅エネルギーを表す特徴量である。時間領域信号 $x[n]$ に対して、RMS は次式で定義される。

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N x[n]^2}$$

図 13: 音量 (RMS) の定義式

本研究では、librosa の librosa.feature.rms() を用いて算出した。

(2-8)メル周波数ケプストラム係数 (MFCC)

MFCC は、人間の聴覚特性を考慮したメル周波数尺度に基づく特徴量であり、音色を表現する代表的な指標である。本研究では、音声信号に対してメルフィルタバンクを適用し、対数パワースペクトルに離散コサイン変換 (DCT) を施すことによって MFCC を算出する。librosa の librosa.feature.mfcc() を用いて 20 次元の MFCC を算出し、各次元の平均値を特徴量として採用した。

さらに、時間的変化を捉えるため、MFCC の一次差分および二次差分を以下の式で定義し、librosa.feature.delta() を用いて算出した。

$$\Delta c_t = c_t - c_{t-1}, \quad \Delta^2 c_t = \Delta c_t - \Delta c_{t-1}$$

図 14: MFCC の一次差分および二次差分の定義式

(2-9)クロマ特徴量 (Chroma)

クロマ特徴量は、12 音階ごとのエネルギー分布を表す特徴量であり、音高や和声的特徴を反映する。本研究では、librosa の librosa.feature.chroma_stft() を用いてクロマ特徴量を算出し、その平均値を特徴量として使用した。

(2-10)トーナルネットワーク特徴量 (Tonnetz)

トーナルネットワーク特徴量 (Tonnetz) は、調性や和声構造を表現する特徴量である。調波成分を抽出した音声信号に対して、librosa の librosa.feature.tonnetz() を用いて 6 次元の Tonnetz 特徴量を算出し、各次元の平均値を特徴量として採用した。

以上の手順で算出されたすべての音響特徴量は一つの辞書にまとめられ、最終的に「1 行を 1 秒セグメント、1 列を 1 種類の音響特徴量」とする特徴量テーブルを生成する。

この特徴量テーブルは CSV 形式で保存され、学習用 CSV データとして構築される。表 4 に、各ライブ映像から抽出された音響特徴量の CSV データの一部を示す。

4.1.3 照明特徴量の抽出

続いて、各映像から RGB 値(色)、Brightness(明るさ)、点滅速度の 3 種類の照明特徴量を 1 秒ごとに抽出し、学習用 CSV データとして構築した。

本節では、照明特徴量の抽出手法、CSV データ構造について詳述する。

表 4: 各映像の音響特徴量の CSV データ (一部抜粋)

千本桜										
Time [s]	Centroid	Contrast	Bandwidth	Rolloff	ZCR	RMS	MFCC_1	Delta_MFCC_1	Chroma_1	Tonnetz_1
1.0	8894	14.5	6149	16606	0.28	0.001	-948	-4.54	0.67	0.03
2.0	6722	17.9	3928	11385	0.23	0.005	-661	4.44	0.76	0.12
3.0	5257	23.2	3046	8144.0	0.18	0.033	-385	0.52	0.57	0.20
4.0	4459	21.7	2653	6389.1	0.17	0.023	-398	2.95	0.75	0.26
5.0	3987	19.5	2593	5936.9	0.12	0.028	-358	-1.11	0.28	0.32
6.0	3678	18.3	3031	7210.9	0.16	0.050	-347	-0.45	0.37	0.20
7.0	3521	17.8	2742	7049.2	0.17	0.039	-410	-0.38	0.30	0.27
8.0	3400	16.7	2896	7186.3	0.14	0.008	-471	-2.29	0.43	0.18
ソナタ第三番										
1.0	1947	9.04	1705	3958.3	0.03	0.001	-1068	3.25	0.04	-0.05
2.0	2898	20.1	3562	4256.3	0.07	0.000	-809.0	1.43	0.12	-0.20
3.0	1695	19.9	2284	2615.9	0.03	0.001	-675.9	1.79	0.35	-0.07
4.0	1560	20.1	1951	2630.2	0.03	0.002	-618.0	-0.26	0.43	0.27
5.0	1728	19.1	1909	2767.1	0.05	0.002	-606.8	-0.07	0.56	-0.04
6.0	1887	19.5	1940	2967.1	0.05	0.002	-604.9	-0.06	0.28	0.02
7.0	2113	19.0	2080	3132.7	0.06	0.002	-585.3	0.05	0.22	0.03
8.0	2475	19.9	2305	3629.7	0.07	0.002	-602.9	-0.41	0.08	-0.01
スプラッシュ										
1.0	1002	0.00	579.2	1750.0	0.00	0.000	-1131	0.00	0.08	-0.01
2.0	3582	8.98	3553	8076.6	0.00	0.008	-1055	1.65	0.27	0.07
3.0	1603	2.24	1542	3564.3	0.00	0.001	-1109	-0.29	0.16	-0.02
4.0	1564	15.0	2032	2707.4	0.03	0.032	-774.5	4.52	0.23	0.14
5.0	1798	18.4	2226	2553.0	0.05	0.054	-641.0	1.15	0.37	0.31
6.0	1893	20.1	2182	2354.2	0.05	0.055	-667.7	-0.62	0.39	0.27
7.0	2060	20.1	2190	2775.3	0.05	0.057	-634.1	-0.16	0.30	0.33
8.0	759.7	19.8	2239	3193.1	0.05	0.062	-621.3	-1.14	0.25	0.17

(1) 人体検出による照明領域の特定

照明特徴量の抽出において最も重要な前処理は、YOLOv8[16] を用いた人物領域の除外である。人物領域を含んだまま RGB や Brightness を算出すると、肌色や衣服の色、演奏者の動きによる輝度変化が混入し、照明そのものの変化を正確に捉えることができない。

そこで本研究では、人物検出後に膨張処理を施し、演奏者周囲も含めて背景から除外した。そのうえで、背景画素のうち明度上位 10 %のみを抽出し、照明器具に最も近い画素群を対象として RGB および Brightness を算出している。

(2) RGB 値の算出と各映像の csv データ

映像のフレーム画像を B,G,R チャンネルに分解し、背景マスクかつ上位 10% 領域に該当する画素を抽出し、それぞれの平均値を計算した。

これらの値をフレーム単位でバッファに蓄積し、1 秒ごとに平均を取ることで、1 秒単位の RGB 値の照明特徴量として CSV に書き出した (表 5,6,7)。

(3) Brightness の算出と各映像の csv データ

背景照明の明るさは、グレースケール画像（または HSV の V 成分）から、同じく上位 10% の明度領域を抽出して平均値を求めた。また、RGB 値と同様に、1 秒ごとに平均を取ることで、1 秒単位の Brightness 値として CSV に書き出した (表 8,9,10)。

(4) 点滅速度の算出と各映像の csv データ

照明の点滅速度を抽出するために、1 秒ごとの平均 Brightness 値を時系列信号として蓄積し、前後 5 秒の局所 FFT を行うことで点滅に対応する周波数成分を推定し、csv に書き出した (表 11,12,13)。

以上により、各映像から音響特徴量と照明特徴量のペアを 1 秒ごとに抽出し、学習用 CSV データとして構築した。

4.2 照明予測モデル (仮) の学習

4.1 節で収集した学習用 CSV データを用いて、照明予測モデルの学習を行った。

4.2.1 学習データの概要

照明予測モデル (仮) の学習には、4.1 節にて生成した

音響特徴量と照明特徴量の CSV のデータを 1 つのデータセットとして統合し、使用した。

4.2.2 入力変数（音響特徴量）

照明予測モデルへの入力として使用した音響特徴量のデータセットは 4.1.2 節で述べた通りである。

4.2.3 出力変数（照明特徴量）

照明予測モデルの出力として使用した照明特徴量のデータセットは 4.1.3 節で述べた通りである。

4.2.4 学習データとテストデータの分割

構築したデータセットは、モデルの汎化性能を評価するため、全体の 80% を学習用データ、20% をテスト用データに分割した。分割はランダムに行い、乱数シードを固定することで実験の再現性を確保した。

- 学習データ：モデルの学習に使用
- テストデータ：モデルの性能評価に使用

4.2.5 モデルの構造と学習方法

照明予測モデル（仮）には、回帰モデルとして Random Forest Regressor[17] を採用した。Random Forest は複数の決定木を組み合わせたアンサンブル学習手法である。決定木の本数は 100 とし、学習精度と計算コストのバランスを考慮した。

4.2.6 学習の実行とモデル保存

学習用データを用いて Random Forest モデルを学習し、音響特徴量と照明特徴量の対応関係を獲得した。学習が完了したモデルは、再利用を目的として joblib を用いてファイルとして保存した。

4.2.7 照明予測モデル（仮）の位置づけ

本節で構築した照明予測モデル（仮）は、ライブ映像由来の実測データのみを用いて学習された点に特徴がある。一方で、演奏種類やデータ量には限界があるため、本研究ではこのモデルを「照明予測モデル（仮）」と位置付けている。

4.3 MIDI データの生成と特徴量抽出

本節では、4.2 節で構築した照明予測モデル（仮）を用いて、MIDI 形式のヴァイオリン楽曲データから音響特徴量および照明特徴量を生成する手法について述べる。

MIDI データは実演奏の録音と異なり、演奏指示情報のみを保持する形式である。そのため、本研究では MIDI データを音声信号へ変換した上で、ライブ映像データと同一の音響特徴量を抽出し、照明予測モデル（仮）によって照明特徴量を推定するという処理フローを採用した。

4.3.1 MIDI データセットの概要

本研究で使用した MIDI データは、ヴァイオリン演奏において代表的かつ教育的価値の高い楽曲群から構成されている。具体的には、「Paganini[18]」、「Kayser[19]」、「Wohlfahrt[20]」の作品を中心としたヴァイオリンソロの MIDI データを用いた。

表 5: 千本桜の RGB 値の CSV データ（一部抜粋）

Time [s]	R	G	B
1.0	90.1	45.7	62.1
2.0	93.9	60.2	74.1
3.0	89.8	60.7	70.0
4.0	84.2	42.5	53.6
5.0	84.8	32.2	45.8
6.0	130.6	16.2	31.1
7.0	125.8	25.1	33.2
8.0	124.7	35.9	38.0

表 6: ソナタ第 3 番の RGB 値の CSV データ（一部抜粋）

Time [s]	R	G	B
1.0	50.8	45.1	68.7
2.0	52.3	45.6	69.7
3.0	54.5	46.9	71.6
4.0	53.9	46.2	70.7
5.0	65.0	58.0	81.3
6.0	120.8	119.2	131.2
7.0	128.8	126.4	136.2
8.0	139.3	137.1	145.3

表 7: スプラッシュの RGB 値の CSV データ（一部抜粋）

Time [s]	R	G	B
1.0	91.6	70.0	76.5
2.0	54.1	36.1	50.5
3.0	17.8	11.4	23.9
4.0	45.1	36.0	66.9
5.0	239.0	238.1	242.0
6.0	189.4	201.3	208.0
7.0	191.6	223.9	233.7
8.0	200.0	203.9	199.8

表 8: 千本桜の Brightness の CSV データ（一部抜粋）

Time [s]	Brightness
1.0	59.8
2.0	69.3
3.0	87.0
4.0	53.5
5.0	47.7
6.0	53.3
7.0	56.7
8.0	63.3

表 9: ソナタ第 3 番の Brightness の CSV データ (一部抜粋)

Time [s]	Brightness
1.0	50.2
2.0	49.8
3.0	52.9
4.0	51.1
5.0	63.2
6.0	124.0
7.0	130.9
8.0	140.5

表 10: スプラッシュの Brightness の CSV データ (一部抜粋)

Time [s]	Brightness
1.0	77.5
2.0	35.3
3.0	19.2
4.0	43.4
5.0	242.3
6.0	200.0
7.0	215.8
8.0	202.8

表 11: 千本桜の点滅速度の CSV データ (一部抜粋)

Time [s]	Blinking Speed
1.0	0.40
2.0	0.33
3.0	0.29
4.0	0.25
5.0	0.44
6.0	0.40
7.0	0.40
8.0	0.20

表 12: ソナタ第 3 番の点滅速度の CSV データ (一部抜粋)

Time [s]	Blinking Speed
1.0	0.40
2.0	0.33
3.0	0.29
4.0	0.25
5.0	0.22
6.0	0.20
7.0	0.20
8.0	0.20

表 13: スプラッシュの点滅速度の CSV データ (一部抜粋)

Time [s]	Blinking Speed
1.0	0.40
2.0	0.33
3.0	0.29
4.0	0.25
5.0	0.22
6.0	0.20
7.0	0.20
8.0	0.20

Paganini の作品は、高度な演奏技巧や急激な音高変化、速いパッセージを多く含み、演奏表現の幅が非常に広い。一方で、Kayser および Wohlfahrt の練習曲は、音階練習や基本的な運弓・音程感覚の習得を目的とした楽曲が多く、比較的安定した音高変化と明瞭なフレーズ構造を持つ。

MIDI データは、演奏音の音高、音価、発音タイミングなどを厳密に数値情報として保持しているため、音響特徴量抽出において再現性が高く、楽曲間の比較にも適している。特に、これらの練習曲・技巧曲は、音楽教育の現場でも広く使用されており、ヴァイオリン演奏の典型的な音域・音色変化・フレーズ構造を網羅している点で、本研究の補助的データとして有用である。

4.3.2 MIDI データの音声変換

MIDI データは直接音響特徴量を算出することができないため、まず音声信号へ変換する必要がある。本研究では、ソフトウェアシンセサイザー FluidSynth[21] を用い、MIDI ファイルを WAV 形式へ変換した。

MIDI から WAV への変換は各ファイルごとに自動的に実行される。

4.3.3 音響特徴量の抽出

生成された WAV ファイルに対し、Python ライブライブラリ librosa を用いて音響特徴量を抽出した。抽出する特徴量の種類および構成は、5.3 節で構築した照明予測モデル（仮）の学習時に用いたデータセットと完全に一致させている。

4.3.4 1 秒単位への集約処理

抽出された音響特徴量はフレーム単位のデータであるため、そのままでは照明制御モデルとの時間スケールが一致しない。そこで、本研究では 1 秒単位で特徴量を平均化する処理を行った。

具体的には、サンプリングレートおよび hop length から 1 秒あたりのフレーム数を算出し、該当フレーム範囲の特徴量を平均することで、1 秒ごとの特徴ベクトルを生成した。

この処理により、ライブ映像データと同様に「1 秒 = 1 照明フレーム」という時間対応が可能となる。

4.3.5 照明予測モデル（仮）による照明特徴量推定

1 秒単位に集約された音響特徴量は、4.2 節で構築した 照

明予測モデル（仮）に入力される。本モデルは、ライブ映像データから学習した「音響特微量と照明特微量の対応関係」を保持しており、MIDI 由来の音響特微量に対しても照明特微量を推定することが可能である。推定される照明特微量は RGB 値、Brightness、点滅速度の 3 種類である。

4.3.6 CSV 形式での出力

各 MIDI ファイルについて、以下の情報を 1 つの CSV データとして出力した。

- ・ 時間(秒)
- ・ 音響特微量
- ・ 照明予測モデル(仮)による推定照明特微量

これにより、MIDI データから得られた音響特微量と対応する照明特微量を一元的に管理できるようになった。

4.4 MIDI データを含めた照明予測モデルの学習

4.3 節で生成した MIDI データ由来の特微量データを加え、照明予測モデルを再学習する手法について述べる。本節で構築されるモデルは、ライブ映像データに加えて MIDI データも含めた拡張データセットを用いて学習されたモデルであり、以降「照明予測モデル(本)」と呼ぶ。

4.2 節では、ライブ映像データのみを用いた学習によって、音響特微量と実際の照明演出との対応関係を獲得できることを示した。しかし、ライブ映像データは演奏数や演奏スタイルに限りがあり、機械学習モデルとしての汎化性能には一定の制約が存在する。そこで本研究では、MIDI データを用いて演奏パターンの多様性を拡張し、より汎用的な照明予測モデル(本)の構築を試みた。

4.4.1 MIDI データ導入の目的

MIDI データを学習データに加える目的は、以下の点にある。

第一に、演奏音のバリエーションを大幅に増加させることである。MIDI データには、音高変化、フレーズ構造、音価の違いなど、実演奏に近い音楽的情報が含まれており、ライブ映像データだけでは不足しがちな演奏パターンを補完できる。

第二に、モデルが特定の演奏映像や照明環境に過度に依存することを防ぎ、未知の演奏音に対しても安定した照明予測を行えるようにする点である。

4.4.2 学習データセットの構成

本研究では、4.2 節で生成した MIDI データ由来の特微量データを、4.1 節で構築したライブ映像データ由来の特微量データに追加し、拡張された学習データセットを構成した。

MIDI データについては、実際の照明映像を伴わないため、4.2 節で学習した照明予測モデル(仮)を用いて照明パラメータを推定し、擬似的な教師データとして扱っている。

4.4.3 特徴量構成と前処理

入力特徴量および出力変数の構成は、4.2 節で使用した

ものと同一であるため、本節では詳細な説明を省略する。

- ・ 入力：音響特微量
- ・ 出力：照明特微量

4.4.4 学習手法

学習手法としては、4.2 節と同様に Random Forest Regressor を採用した。モデル構造や主要なハイパーパラメータも、同一に設定している。

4.4.5 学習データとテストデータの分割

統合されたデータセットは、4.2 節と同様に学習用データとテスト用データに分割した。分割比率および分割方法は同一であり、実験条件の公平性を保っている。

4.4.6 学習の実行とモデル保存

拡張された学習データセットを用いて Random Forest モデルの学習を実行し、音響特微量と照明特微量の対応関係を再度獲得した。学習が完了したモデルは、joblib を用いてファイルとして保存し、リアルタイム照明制御システムに直接組み込んだ。

4.4.7 照明予測モデル(本)の位置づけ

本節で学習された照明予測モデル(本)は、ライブ映像由来の実測データと、MIDI データによって補完された多様な演奏パターンの両方を内包している。

このモデルを、本研究における最終的な照明予測モデルと位置付ける。

4.5 照明予測モデルの評価

本節では、4.4 節で構築した MIDI データを含めた照明予測モデル(本)の性能評価について述べる。評価には、機械学習における回帰問題で一般的に用いられる決定係数 (R^2 スコア) を採用した。

4.5.1 評価指標：決定係数 (R^2 スコア)

本研究では、照明予測モデル(本)の性能を定量的に評価するため、決定係数 R^2 (Coefficient of Determination) (図 15) を用いた。 R^2 スコアは、モデルの予測値が実測値の分散をどの程度説明できているかを示す指標であり、以下の式で定義される。

$$\text{決定係数 } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

y_i : 実測値_i \hat{y}_i : 予測値_i \bar{y} : 実測値の平均値

図 15: 決定係数 (R^2) の定義式

R^2 スコアは 1 に近いほど予測精度が高く、0 は平均値予測と同等、負の値は平均予測より劣ることを示す。

4.5.2 評価方法

評価には、学習時に使用していないテストデータを用いた。テストデータには、ライブ映像由来データを中心に含めることで、実際の照明演出との整合性を重視した評価を行った。

評価は、RGB 値、Brightness、点滅速度といった複数の照明特徴量を同時に含む多出力回帰問題として実施し、全出力を総合した R^2 スコアを算出した。

4.5.3 評価結果

最終照明予測モデル(本)に対する評価の結果、 R^2 スコアは 0.85 を示した(図 16)。

✓ モデルを lighting_model_violin.pkl として保存しました。
モデル精度 (R^2 スコア) : 0.8533

図 16: 決定係数 (R^2) の結果

この結果は、音響特徴量から照明特徴量への変換において、モデルが実測照明データの分散の約 85%を説明できていることを意味する。すなわち、照明の色、明るさ、点滅といった複合的な要素を、演奏音のみから高精度に推定できていることが確認された。

4.6 音響特徴量と照明特徴量の対応関係

本節では、最終的に構築した照明予測モデル(本)が、音響特徴量と照明特徴量の間にどのような対応関係を学習しているかを分析する。前節まででは、 R^2 スコアを用いた定量的評価によりモデル全体の性能を示したが、本節ではさらに踏み込み、「どの音響的要素が、どの照明要素に強く影響しているか」を明らかにすることを目的とする。

音と照明の関係は本質的に主観的であり、一意に定義されるものではない。しかし、機械学習モデルが内部的に利用している特徴量の重要度を解析することで、モデルがどのような音響的判断基準に基づいて照明を制御しているかを客観的に考察することが可能となる。

4.6.1 解析手法の概要

音響特徴量と照明特徴量の対応関係を解析するため、本研究では Random Forest における特徴量重要度 (feature importance) を用いた。

Random Forest では、各決定木において分岐に使用された特徴量がどれだけ予測誤差の低減に寄与したかを基に、特徴量重要度が算出される。この値は、モデルが予測を行う際にどの特徴量を重視しているかを示す指標である。

本研究では、照明特徴量ごとに個別の Random Forest 回帰モデルを学習し、それぞれの出力に対して音響特徴量の重要度を算出した。これにより、照明の色・明るさ・点滅といった各要素が、どの音響的特徴と強く結びついているかを独立に分析できる。

4.6.2 使用データ

解析には、4.1 節および 4.3 節で生成された CSV データを使用した。これらの CSV ファイルには、1 秒単位で集約された音響特徴量と対応する照明特徴量が含まれている。

4.6.3 照明特徴量別の解析方法

「RGB 値」「Brightness」「点滅速度」の各照明特徴量に対して、それぞれ独立に解析を行った。

各照明特徴量を目的変数とし、音響特徴量を説明変数として Random Forest 回帰モデルを学習した。学習後、特徴量重要度を算出し、寄与度の高い上位 5 特徴量を抽出した。

結果の可視化には円グラフを用い、各特徴量が予測にどの程度寄与しているかを直感的に把握できるようにした。

4.6.4 音響特徴量と照明色 (R 成分) の対応関係

図 17 は、照明色の R 成分 (Average R) を目的変数として学習した Random Forest 回帰モデルにおける、音響特徴量の重要度上位 5 項目を円グラフで示したものである。本図は、モデルが赤色成分の強度を予測する際に、どの音響特徴量をどの程度重視しているかを可視化した結果である。

本解析の結果、RMS が最も高い重要度を示し、次いで Tonnetz_6、MFCC_5、MFCC_19、SpectralContrast の順となった。

Top 5 Important Features for Average R

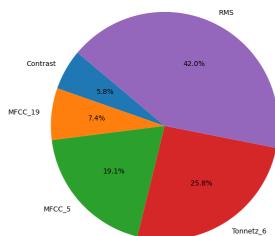


図 17: 照明色 (R 成分) に対する音響特徴量

RMS (Root Mean Square) は音響信号のエネルギー量、すなわち音量を表す特徴量である。本結果において RMS が最も高い寄与を示したことは、音量の大きさが照明の赤成分の強度に強く反映されていることを示している。

2 番目に高い重要度を示した Tonnetz_6 は、音楽の和声構造や調性感を表す特徴量である。これが R 成分の予測に大きく寄与していることから、単なる音量だけでなく、和声的・音楽的な性質が照明色に反映されていることが示唆される。

また、MFCC_5 および MFCC_19 といった MFCC 系特徴量も比較的高い重要度を示している。MFCC は音色の特徴を表す指標であり、これらの結果は、音色の違いが照明色の変化に影響していることを意味している。

特に中低次および高次の MFCC 成分が同時に重要な要素である点から、モデルは音の明るさだけでなく、倍音構造の微妙な違いも考慮して照明色を決定していると考えられる。

れる。

Spectral Contrast は、周波数帯域ごとのピークと谷の差を表す特徴量であり、音の鋭さや迫力を反映する指標である。本結果では重要度は比較的低いものの、上位 5 特徴量に含まれている。

これは、音の輪郭が明瞭であるかどうかが、赤色成分の強調に一定の影響を与えていていることを示している。

4.6.5 音響特徴量と照明色 (G 成分) の対応関係

図 18 は、照明色の G 成分 (Average G) を目的変数として学習した Random Forest 回帰モデルにおける、音響特徴量の重要度上位 5 項目を示した結果である。本図は、音響信号のどの要素が緑色成分の強度に影響を与えていているかを可視化したものである。

解析の結果、RMS が全体の約 86.8% と極めて高い重要度を示し、他の特徴量 (Tonnetz_6、Chroma_7、MFCC_5、SpectralContrast) はいずれも 1~7% 程度に留まった。

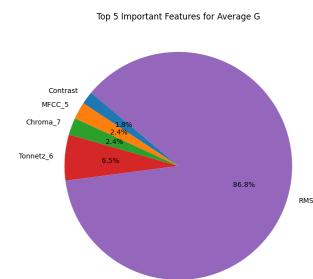


図 18: 照明色 (G 成分) に対する音響特徴量

G 成分において RMS が圧倒的な重要度を示したことは、音量が緑色照明の強度をほぼ支配的に決定していることを意味している。

この結果は、G 成分が他の色成分に比べて、音楽的な細かなニュアンスよりも音圧の変化を直接反映する成分として扱われていることを示している。

RMS に次いで Tonnetz_6 (約 6.5%) が高かった。この結果は、緑色成分においても、音量以外に和声的要素が一定の影響を及ぼしていることを示唆している。

また、Chroma_7 や MFCC_5 はそれぞれ音高クラス分布および音色に関連する特徴量であり、これらが上位に含まれている点は、緑成分が完全に単純な音量依存ではないことを示している。ただし、これらの寄与率は小さく、あくまで補助的な要素であると解釈できる。

R 成分の解析結果では、RMS に加えて Tonnetz や MFCC が比較的高い割合を占めていたのに対し、G 成分では RMS の寄与が著しく大きくなっている。この違いは、照明色ごとに音響特徴量との対応関係が異なることを示している。

4.6.6 音響特徴量と照明色 (B 成分) の対応関係

図 19 は、照明色の B 成分 (Average B) を目的変数として学習した Random Forest 回帰モデルにおける、音響特徴

量の重要度上位 5 項目を示した結果である。本図は、音響信号のどの要素が青色照明の強度に影響を与えていているかを示している。

解析の結果、RMS が約 70.2% と最も高い重要度を示した。次いで MFCC_19 (約 14.9%) が比較的高い寄与を示し、Chroma_7、Tonnetz_6、MFCC_5 はそれぞれ約 5% 前後であった。

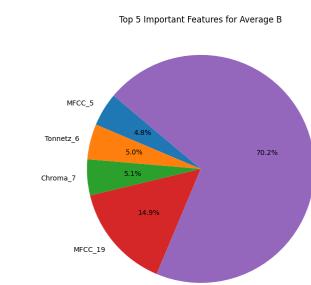


図 19: 照明色 (B 成分) に対する音響特徴量

B 成分においても RMS が最も支配的な特徴量であることから、青色照明も音量変化に大きく依存していることが分かる。音量が大きい場面では青色成分が強調され、音量が小さい場面では抑制される傾向をモデルが学習していると考えられる。

ただし、G 成分と比較すると RMS の割合はやや低下しており、B 成分では音量以外の要素も一定程度考慮されていることが示唆される。

MFCC_19 が約 15% と比較的高い重要度を示した点は、B 成分が音色や高次スペクトル成分に影響されやすいことを示している。

Chroma_7 および Tonnetz_6 が上位に含まれていることから、B 成分は音量や音色だけでなく、音楽的構造（音高分布や和声的関係）とも関連していることが分かる。

特に Tonnetz が青色成分に影響している点は、静的・緊張感のある和声や、落ち着いた調性が青系照明として表現されやすい可能性を示唆している。

これらの結果から、RGB 各成分は一様に音量に反応するのではなく、それぞれ異なる音響的役割を担っていることが明らかとなった。

4.6.7 音響特徴量と照明の明るさの対応関係

図 20 は、照明の明るさ (Background Brightness) を目的変数として学習した Random Forest 回帰モデルにおける、音響特徴量の重要度上位 5 項目を示した結果である。本図は、音響信号のどの要素が照明全体の明るさに影響を与えているかを明らかにするものである。解析の結果、RMS が約 77.8% と最も高い重要度を示した。次いで Tonnetz_6 (約 10.4%)、MFCC_19 (約 5.3%)、MFCC_5 (約 3.7%)、SpectralContrast (約 2.8%) が続いた。

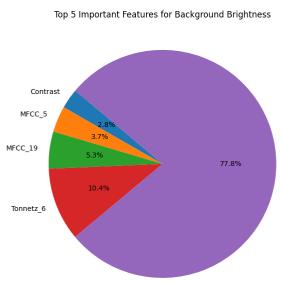


図 20: 照明の明るさに対する音響特微量

RMS が約 8 割近い寄与率を示したことから、照明の明るさは主に音量に基づいて決定されていることが明らかとなった。

本モデルでは、演奏音が強くなるにつれて照明が明るくなり、弱くなるにつれて暗くなるという対応関係が学習されている。

RMS に次いで Tonnetz_6 (約 10.4%) の重要度が高かった。この結果は、照明の明るさが単純な音量変化だけでなく、音楽的な構造要素にも影響を受けていることを示している。

また、MFCC_19 および MFCC_5 が上位に含まれていることから、照明の明るさは音量だけでなく、音の質感や倍音構成にも影響されていることが分かる。

RGB 各成分と照明の明るさの関係を比較すると、明るさは RGB 成分以上に RMS への依存度が高いことが分かる。これは、明るさが照明全体の基盤的なパラメータであり、色成分よりも直接的に音量変化を反映する役割を担っているためである。

一方で、Tonnetz や MFCC が一定の寄与を示していることから、明るさも完全に単純な制御ではなく、音楽的文脈を考慮した調整が行われていると解釈できる。

4.6.8 音響特微量と照明の点滅速度の対応関係

図 21 は、照明の点滅速度 (Blinking Speed) を目的変数として学習した Random Forest 回帰モデルにおける、音響特微量の重要度上位 5 項目を示した結果である。本図は、音響信号のどの要素が照明の時間的变化、すなわち点滅の速さに影響を与えていているかを可視化したものである。

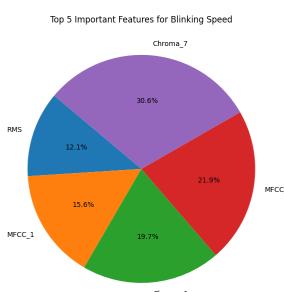


図 21: 照明の点滅速度に対する音響特微量

解析の結果、Chroma_7 が約 30.6% と最も高い重要度を示し、次いで MFCC_5 (約 21.9%)、Chroma_1 (約 19.7%)、MFCC_1 (約 15.6%)、RMS (約 12.%) という順となった。

点滅速度において最も大きな影響を与えているのは Chroma 系特微量であり、特に Chroma_7 および Chroma_1 が全体の約 50% を占めている。この結果は、照明の点滅が音量の大小よりも、音楽の和声的・音高的構造に強く依存していることを示している。

Chroma の変化が大きい場面では、照明の点滅が速くなり、音楽的な動きや緊張感を視覚的に強調する演出が行われていると考えられる。

MFCC_5 および MFCC_1 も比較的高い重要度を示しており、点滅速度が音色の変化にも影響を受けていることが分かる。

RMS は点滅速度においても一定の影響を持つものの、その重要度は約 12.1% に留まっている。これは、照明の明るさとは異なり、点滅という時間的变化の制御においては、音量そのものよりも音楽的構造や音高変化が重視されていることを示唆している。

RGB 各成分および照明の明るさでは RMS の重要度が支配的であったのに対し、点滅速度では Chroma や MFCC といった音楽的特微量が主導的な役割を果たしている。この違いは、照明パラメータごとに参照される音響情報の種類が異なることを明確に示している。

5. リアルタイム照明制御システムの実装

本章では、本研究で提案したヴァイオリン演奏に基づくリアルタイム照明制御手法を実現するために構築したシステム構成について、ハードウェア構成、ソフトウェア構成、処理フロー、通信方式、リアルタイム性確保の工夫の観点から詳細に述べる。本システムは、演奏者の音響的表現および身体的動作を入力とし、それらを機械学習モデルによって照明表現へ変換する演出システムである。

5.1 システム設計の基本方針

本システムの設計にあたり、以下の方針を重視した。

- ・ 演奏中に人手操作を必要としないこと
- ・ 遅延を極力抑え、リアルタイム性を確保すること
- ・ 実演が可能な安定した動作を実現すること

特に本研究では、筆者自身がヴァイオリン演奏を行なながら照明演出を体験することを目的としているため、操作負荷の少ない自律的なシステム構成が不可欠であった。

5.2 ハードウェア構成

本システムは以下のハードウェアで構成される。

- ・ PC(Windows 11)
- ・ コンデンサマイク (音声入力)
- ・ スマートフォン (ZIG-SIM 用)

- USB-DMX インタフェース
 - DMX 対応照明機器 4 台 (図 22)
- マイクは演奏音をリアルタイムで取得するために使用し、スマートフォンはヴァイオリン演奏時の運弓動作を取得するために使用する。照明機器はステージ四方に配置され、演奏者を中心に照射する構成とした。



図 22: DMX 対応照明機器

5.3 ソフトウェア構成

本システムは Python で実装され、すべての処理を单一プログラム上で統合している。使用した主なライブラリを以下に示す。

- sounddevice[22] : 音声ストリーム取得
- librosa : 音響特徴量抽出
- numpy[23], pandas[24] : 数値処理、データ管理
- scikit-learn[25] : 機械学習モデル推論
- socket[26] : UDP 通信
- ftd2xx[27] : DMX 信号出力

これにより、外部ソフトウェアを必要とせず、Python のみでリアルタイム照明制御を実現している。

5.4 全体処理フロー

システムの処理は、演奏開始から照明出力まで以下の流れで進行する。また、図 23 にシステム全体の構成図を示す。

- 音声入力ストリームの開始
- 音声バッファへの蓄積
- 一定時間ごとのフレーム分割
- 音響特徴量抽出
- 照明特徴量推定
- 運弓方向推定
- DMX データ生成
- 照明機器への信号送信

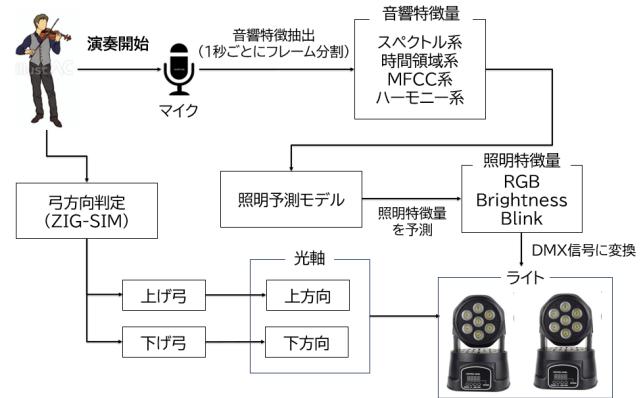


図 23: リアルタイム照明制御システム構成図

5.5 音声入力処理

音声入力は sounddevice.InputStream を用いて行う。サンプリング周波数は 22,050 Hz、モノラル入力とし、一定数のサンプルが蓄積されるとコールバック関数内でリングバッファに格納される。

リングバッファを採用することで、音声の欠落を防ぎつつ、安定したフレーム分割が可能となっている。バッファ内のデータが 22,050 サンプルに達した時点で、1 秒分の音声フレームとして切り出される。

また、音声フレームごとに RMS 値を算出し、あらかじめ設定した閾値以下の場合は無音区間と判定する。無音区間では照明更新を行わず、直前の照明状態を保持する。

5.6 音響特徴量抽出

各音声フレームからモデル構築の際に用いた特徴量と同一かつ同様の手法 (4.1.2 節参照) での特徴量を抽出する。

5.7 演奏動作取得

本システムでは、演奏者の弓の上下運動（上げ弓・下げ弓）を取得するために、スマートフォンを ZIG-SIM センサとして利用している。スマートフォンにはジャイロセンサが搭載されており、X 軸方向の角速度を取得することで弓の前後運動を検出する。

5.7.1 ハードウェア配置

スマートフォンは演奏者の右腕に装着し、弓を持つ手首付近に位置させる。これにより、弓の上下運動が腕の回転として反映されるため、ジャイロセンサで容易に検出可能となる。

5.7.2 データ通信方式

スマートフォンから取得したジャイロデータは UDP 通信を用いて PC に送信する。UDP は TCP に比べて通信遅延が少なく、リアルタイム性を重視した設計である。送信ポートは 55555 番を使用し、PC 側は非同期で受信する専用スレッドを設け、常時データを待ち受ける。

受信データは JSON 形式で送信され、ジャイロ情報は

”sensordata”:”gyro”:”x”: 値 という構造になっている。PC 側では、受信データをデコードして Python 辞書型に変換後、X 軸の角速度値を取り出す。

5.7.3 運弓方向判定アルゴリズム

弓方向の判定は単純な閾値判定ではなく、以下の 2 つの条件を組み合わせて精度を向上させている。

- ・ 角速度の絶対値による判定
閾値 THRESH_GX = 0.01 を設定

$gx > \text{THRESH_GX}$ なら上げ弓、 $gx < -\text{THRESH_GX}$ なら下げ弓

- ・ 角速変化量による判定

前フレームとの変化量 $\text{delta} = gx - \text{prev_gx}$ を計算

閾値 THRESH_DELTA = 0.009 を超えた場合も上げ弓／下げ弓と判定

この 2 つを併用することで、微小な振動や演奏の微妙なタッチでも誤判定を避け、弓の動作を正確に捉えることができる。

5.7.4 リアルタイム処理

ZIG-SIM データ受信スレッドは 10 ms 周期で UDP パケットを受信し、取得した角速度に基づいて弓方向を更新する。最新の弓方向はグローバル変数 bow_direction に保存され、照明制御スレッドで参照される。

bow_direction の値は以下の 3 状態を持つ。

- ・ 1：上げ弓
- ・ -1：下げ弓

これにより、演奏中の弓の動きに応じてライトの上下動作を同期させることができる。

このように、本システムではジャイロセンサによる演奏動作取得をリアルタイムで行い、音響特徴量と組み合わせて照明制御に反映している。

5.8 照明特徴量推定

抽出された音響特徴量を用いて、4.4 節で構築した照明予測モデル(本)により照明特徴量を推定する。

5.9 DMX 信号生成と照明制御

本節では、機械学習モデルおよび演奏動作推定によって得られた照明特徴量を、実際の照明機器に反映させるための DMX 信号生成および制御方法について詳細に述べる。DMX512 は舞台照明分野において広く用いられている通信規格であり、本研究ではこの標準規格を用いることで、実際のステージ環境に近い条件での実演を可能にしている。

5.9.1 DMX512 プロトコルの概要

DMX512 は、最大 512 チャンネルの制御データを 1 フレームとして送信する一方向通信プロトコルである。各チャンネルは 0~255 の 8bit 値を持ち、照明機器の色、明るさ、点滅速度を制御することができる。

5.9.2 DMX ハードウェアインターフェース

DMX 信号の出力には、USB 接続の DMX インタフェースを使用した。本研究では FTDI 社製チップを搭載したデバイスを用い、Python から ftd2xx ライブラリを介して制御を行っている。

このインターフェースでは、DMX 信号送信の前にブレーク信号を発生させる必要がある。プログラム内では、setBreakOn() および setBreakOff() を用いてブレークを明示的に制御し、その後に 512 チャンネル分のデータを一括送信する構成とした。

5.9.3 照明特徴量から DMX 値への変換

照明特徴量 (RGB 値, Brightness, 点滅速度) は、DMX チャンネルに対応する値へ変換される。

これらは連続値として出力されるため、DMX チャンネルに割り当てる際には 0~255 の整数値へ正規化・クリッピング処理を行う。

具体的には、各値に対して以下の処理を適用する。

- (1) 正規化：各特徴量の最小値・最大値を基に 0~1 の範囲にスケーリング
- (2) スケーリング：0~1 の値を 255 倍して 0~255 の範囲に変換
- (3) クリッピング：小数点以下を切り捨て、整数値に変換

これにより、モデルの出力値が異常な場合でも照明機器に過度な負荷がかからないよう安全性を確保している。

5.9.4 複数照明機器への同時制御

本システムでは、4 台の照明機器を同時に制御する構成とした。各照明機器は異なる DMX アドレスに割り当てられており、同一の照明特徴量を用いて各機器へ同時に信号を送信する。

DMX 配列は 513 バイトで構成され、照明機器ごとに以下のようないちごとにチャンネル割り当てを行っている。

- ・ チャンネル 1：未使用
- ・ チャンネル 2：上下動作
- ・ チャンネル 3：点滅速度
- ・ チャンネル 4：R 成分
- ・ チャンネル 5：G 成分
- ・ チャンネル 6：B 成分
- ・ チャンネル 7：明るさ

5.9.5 演奏動作情報との統合制御

音響特徴量から推定された照明特徴量とは別に、ZIG-SIM によって取得した運弓方向情報を DMX 制御に直接反映させている。

具体的には、運弓方向に応じて上下動作の DMX チャンネル (チャンネル 2) の値を変更し、照明の上下動作を演奏者の弓の動きに同期させている。

- ・ 上げ弓：DMX 値を 120 に設定 (照明上昇) (図 24 参照)
- ・ 下げ弓：DMX 値を 0 に設定 (照明下降) (図 24 参照)

とすることで、演奏者の身体動作と照明の物理的動きを

同期させている。この設計により、音だけでなく演奏動作そのものが照明演出に影響を与える構造となっている。

5.9.6 リアルタイム出力制御と更新周期

DMX 信号の送信は約 10ms 周期で実行される。これは、人間が遅延として知覚しにくい範囲であり、演奏と照明がほぼ同時に変化しているように感じられる。

また、照明出力処理は独立したスレッドで動作し、音声処理やモデル推論の遅延の影響を受けにくい設計となっている。これにより、安定した照明制御が可能となった。



図 24: 照明上下動作

6. 実演

本章では、本研究で構築したリアルタイム照明制御システムを用いて実施した実演について述べる。実演では、実際の演奏環境においてシステムを稼働させ、ヴァイオリン演奏に対して照明がリアルタイムに反応する様子を確認した。

6.1 実演環境

実演は、京都府福知山市の新町商店街にある「Tsunaga Room[28]」にて実施した。室内には演奏者を中心配置し、その周囲四方向に DMX 対応照明機器を設置した。照明は床方向から天井に向けて照射される構成とし、四方から演奏者を照らすような光環境を作り出した。

システムは以下の機器構成で運用した。また、各機材の配置図(上面図)は図 25 に示す。

- ・ ヴァイオリン(演奏者：筆者)
- ・ コンデンサマイク(音声入力)
- ・ Windows 11 PC
- ・ スマートフォン(ZIG-SIM 用)
- ・ USB-DMX インタフェース
- ・ DMX 対応照明機器(4 台)

図 25 におけるそれぞれの配線はケーブルを用いたもの

を実線で示し、無線通信は点線で示している。

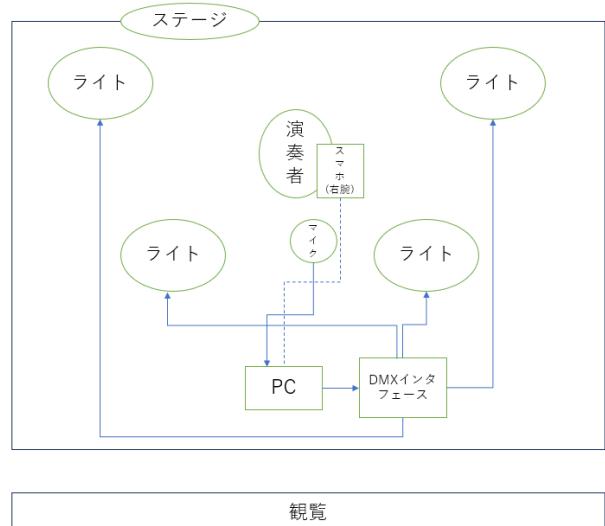


図 25: 実演機材配置図(上面図)

6.2 実演曲目

実演に使用した楽曲は、『紅蓮華 [29]』(作曲：草野華余子) のである。本曲はアニメ『鬼滅の刃』のオープニングテーマとして広く知られており、力強いリズムと感情豊かなメロディが特徴である。ヴァイオリンソロアレンジ版を用い、演奏時間は約 1 分 30 秒である。演奏はイントロから 1 番のサビまでをカバーしている。図 26 に実演で使用した楽譜を示す。

6.3 実演中の照明変化の様子

図 27～図 30 に、実演中の照明変化の様子を示す。これらの画像は、前述した実演環境にて、ヴァイオリン演奏曲『紅蓮華』を演奏した際の照明挙動を記録したものである。

演奏開始直後のイントロ部分(図 27)では、照明は主に青色系の低輝度な光で構成されている。これは、音響特微量のうち RMS 値が低く、音量が抑えられている区間であるためであり、照明の Brightness が低く制御されていることを示している。また、色成分についても急激な変化は見られず、楽曲導入部として静的かつ落ち着いた視覚表現が生成されている。

演奏が進行していくにつれて(図 28)、照明の色調は赤色を含む暖色系へと変化し、同時に明るさが増加している様子が確認できる。この変化は、音量の増加および音色の変化に伴い、機械学習モデルが RGB 値および Brightness を動的に更新している結果である。特に、赤系の照明が演奏者の身体およびヴァイオリンに強調的に照射されており、楽曲の緊張感やエネルギーを視覚的に表現している。さらに、サビ部分に近づくにつれて(図 29,30)、照明は高輝度状態となり、色成分も紫、黄色といった複数の色が時間的に変化している。

Violin

♪135

↑イントロ

↑Aメロ

↑Bメロ

↑Cメロ

↓サビ

図 26: 実演曲目『紅蓮華』楽譜



図 28: 実演中の照明変化の様子 (A メロ～C メロ)



図 27: 実演中の照明変化の様子 (イントロ部分)



図 29: 実演中の照明変化の様子 (サビ部分)



図 30: 実演中の照明変化の様子（サビ部分）

7. 考察

7.1 提案手法の有効性に関する考察

実演の結果、ヴァイオリン演奏音に基づく音響特徴量および演奏動作（運弓方向）を入力とした照明制御が、実環境において安定して動作することが確認された。

6 章で示した実演結果より、演奏音の音量変化に応じて照明の明るさが連動し、楽曲の盛り上がりが視覚的に強調されることが確認された。特に、RMS を主とする音響特徴量が照明の明るさや色成分に大きく寄与している点は、4 章で行った特徴量重要度解析の結果とも整合している。このことから、本システムは音響的なエネルギー変化を適切に照明表現へ反映できていると考えられる。それに加え、さらに、演奏者の運弓動作に基づくライトの上下動作制御により、音響情報だけでは表現しきれない身体的な演奏表現を照明に反映することができた。

7.2 実演環境における制約と課題

一方で、本研究の実演環境にはいくつかの制約および課題が存在する。

第一に、照明予測モデルは 1 秒単位で特徴量を集約しているため、急激な演奏変化に対してはわずかな時間遅れが生じる点が挙げられる。実演においても、非常に短いフレーズや瞬間的なアクセントについては、照明変化が完全には同期しない場面が確認された。この点は、リアルタイム性を重視する舞台演出において今後改善すべき課題である。

ある。

第二に、照明予測モデルの学習データは、特定の演奏映像および MIDI データに基づいて構築されている。そのため、演奏スタイルや楽曲ジャンルが大きく異なる場合には、照明表現が意図しない挙動を示す可能性がある。

第三に、運弓動作の取得にはスマートフォンのジャイロセンサを用いており、装着位置や演奏者の個人差によって取得データにばらつきが生じる可能性がある。実演では大きな問題は生じなかったものの、長時間の演奏や激しい動作においては、センサのずれやノイズの影響が懸念される。

これらの課題から、本システムは実用的な可能性を示した一方で、安定性および汎用性の面では改良の余地があるといえる。

8. 結び

本研究では、ヴァイオリン演奏に基づいて照明をリアルタイムに制御するインタラクティブ照明演出システムの構築を目的とし、音響信号解析、演奏動作解析、機械学習、DMX 照明制御を統合した手法を提案・実装した。

まず、ヴァイオリン演奏音から音量、音色、スペクトル構造を表す音響特徴量を抽出し、それらを入力として照明特徴量（RGB、明るさ、点滅）を推定する機械学習モデルを構築した。学習データには、実演映像から抽出した照明情報と、大量の MIDI 音源を用いた疑似照明データを組み合わせた。

さらに、演奏者の身体表現を反映させるため、スマートフォンアプリ ZIG-SIM を用いてヴァイオリン演奏時の運弓方向を取得し、照明の上下動作へ直接反映する仕組みを導入した。

提案手法を実装したリアルタイム照明制御システムは、Python 上で音響処理、モデル推論、センサ通信、DMX 制御を統合した構成となっており、マルチスレッド設計により実用的なリアルタイム性を確保した。また、京都府福知山市新町商店街での屋内環境にて、ヴァイオリン演奏曲『紅蓮華』を用いた実演を行った。

本研究の成果として、以下の点が挙げられる。第一に、音響特徴量と演奏動作を統合したリアルタイム照明制御という、新しい演出手法を実現した点である。第二に、演奏者自身が演奏によって直接照明を制御する体験を提供し、実現した点である。

以上により、本研究では、ヴァイオリン演奏に基づくりアルタイム照明制御システムの構築と実演を通じて、音響信号処理と演奏動作の融合による新たな照明制御手法を提案した。

謝辞

本論文を執筆するにあたり、多くの方々からご指導・ご支援を賜りました。ここに深く感謝の意を表します。まず、

本研究を進めるにあたり、研究の方向性から終始丁寧なご指導と貴重なご助言をくださった橋田光代准教授に心より御礼申し上げます。また、研究室の4年生の皆様には、日頃より温かく接していただき、研究の相談にも快く応じていただきました。皆様との和やかな会話は、研究生活の大きな支えとなりました。心より感謝いたします。最後に、本研究に関わってくださったすべての方々に、厚く御礼申し上げます。

参考文献

- [1] ZIG-SIMProject.
- [2] 月東菜乃, 酒向慎司: 楽曲に基づく照明演出の自動構成に向けたリズム強調箇所の推定と照明変化への反映, エンタテインメントコンピューティングシンポジウム論文集 卷 2025, p. 48-55 (2025).
- [3] 神野満里奈, 福原義久: 歌詞や曲調の印象に基づいた照明演出の自動生成システムの実現, エンタテインメントコンピューティングシンポジウム論文集 卷 2021, p. 135-143 (2021).
- [4] Zhao, Z., Jin, D., Zhou, Z. and Zhang, X.: Automatic Stage Lighting Control: Is it a Rule-Driven Process or Generative Task?, arXiv preprint arXiv:2506.01482 (2025).
- [5] 浅田風太, 南澤孝太: 演奏者と鑑賞者の身体性を接続するオンラインライブ音楽ライブ空間のデザイン, 慶應義塾大学大学院メディアデザイン研究科 (2020).
- [6] 【弓まっすぐへの道】弓だけの練習法！【1日5分】—オンラインバイオリンレッスンなら『はるか先生！』.
- [7] 「MV」千本桜 WhiteFlame feat 初音ミク - YouTube.
- [8] ストラディバリウス！今宵の『千本桜』はキレッキレ！／石川綾子 AYAKO ISHIKAWA -SENBONZAKURA on Stradivarius.
- [9] 無伴奏ヴァイオリン・ソナタ第3番ニ短調「バラード」Op. 27, No. 3 (イザイ) - ヴァイオリン 楽譜.
- [10] 【LIVE】Hikarito Violin Solo "Ysaye Sonata No.3 Ballade" /イザイ作曲ヴァイオリンソナタ第三番【ライブ映像】.
- [11] SHOGO「SPLASH」の楽曲（シングル）・歌詞ページ | 1009177709 | レコチョク.
- [12] Violinist SHOGO Solo Stage in Namba Hatch Soleil Solo Live.
- [13] Demucs - 検索.
- [14] Welcome to Python.org.
- [15] librosa — librosa 0.11.0 documentation.
- [16] YOLOv8: State-of-the-Art Computer Vision Model.
- [17] RandomForestRegressor — scikit-learn 1.8.0 documentation.
- [18] Paganini 24 Violin Caprice No. 10 in G Minor, Op. 1/10 MIDI File Download :: MidiShow.
- [19] 36 Violin Studies, Op.20 (Kayser, Heinrich Ernst) - IMSLP/ペトルッチ楽譜ライブラリー: パブリックドメインの無料楽譜.
- [20] Free sheet music - Wohlfahrt VIOLIN - FIDDLE - Download PDF, MP3 & MIDI.
- [21] FluidSynth — Software synthesizer based on the SoundFont 2 specifications.
- [22] sounddevice · PyPI.
- [23] NumPy.
- [24] pandas - Python Data Analysis Library.
- [25] scikit-learn: machine learning in Python — scikit-learn 1.8.0 documentation.
- [26] socket — 低水準ネットワークインターフェース — Python 3.14.2 ドキュメント.
- [27] ftd2xx · PyPI.
- [28] Tsunaga Room – 「お家以上、図書館未満、」 学生と企業を繋ぐ新たなコワーキングスペース.
- [29] LiSA 紅蓮華 歌詞 - 歌ネット.