

# OneCamLiveFX：一視点固定型ライブ映像配信のための映像効果付与システムとその運用

小川 剣二郎<sup>1</sup> 渡邊 健斗<sup>1</sup> 中村 聡史<sup>1</sup>

**概要：**軽音楽団体によるオンラインライブでは、一視点から配信することが多く映像が単調になりやすい。また視聴端末としてスマートフォンが多く画角が小さいことから、臨場感が薄れやすいという問題がある。我々はこれまで、ライブ映像に対して映像効果を自動で付与するシステムを提案し、視聴時間や主観評価により臨場感向上に繋がり得ることを明らかにしてきた。これらの結果を軽音楽の分野に活用するには、ライブの価値の観点からリアルタイムで運用可能なシステムを実装する必要がある。そのため、リアルタイムでライブ映像配信に映像効果を自動付与するシステム「OneCamLiveFX」を構築し、その性能評価と、視聴者・演者へのアンケートを行なった。その結果、リアルタイムで運用可能であるが撮影環境によって一部遅延が見られること、本システムは臨場感を向上させリアルタイムであることがその感覚を強化していることを明らかにした。

**キーワード：**ライブ映像配信, 映像効果, リアルタイム, 臨場感, 没入感, 社会的存在感, 興味度, 人物認識

## 1. はじめに

軽音楽団体でのライブでは、対面ライブと並行してオンラインライブを行うことで活動の幅を広げているケースが多い。オンラインライブは気軽に参加できることが特徴であり、スケジュールの都合上参加できない人や対面ライブに参加することに壁を感じている人にもライブを届けることができる。また、オンラインライブは軽音楽団体側にとっては周知の意図もあり、より多くの人に興味を持ってもらうことで団体を活性化へ繋がるものとなっている。

しかし、このようなオンラインライブにおいて多くの視聴者を留まらせることは容易ではない。視聴者が離脱してしまう要因として、まず軽音楽の部活やサークルでは撮影機材が乏しいことが多く、結果として一視点から撮影された映像配信になり、映像に変化が無く単調となってしまうことが挙げられる [1,2]。また別の要因として、視聴者が利用する端末としてスマートフォンが多く [3]、結果として視聴時の画角が小さくなり、臨場感も低くなる事が挙げられる [4]。これらの要因から、スマートフォンのような画面が小さい端末での視聴においても、視覚的興味を維持できるような工夫をする必要がある。

我々はこれまでの研究において、オンライン配信ライブ

の臨場感向上に対する初期検討として、ライブ映像の特徴的なシーンを検出し自動で適した映像表現を付与する手法を提案してきた [5]。また、より詳細な主観評価による臨場感形成の確認と、映像効果に合わせて音響効果を付与することによる影響を調査し、効果を明らかにしてきた [6]。しかし、これらの研究は事前に用意した映像を対象としており、ライブ現場での検証は行うことができていない。

ここでライブは予測不能性が強く [7]、即時的である [8] ことが、その場限りの体験となり臨場感に繋がっている。また、他の誰かが同じ体験を同じタイミングでしていることがライブの価値を構成する重要な要素とされている [7,9,10]。これらはオンラインライブでも同様であり、リアルタイムで運用可能なシステムを実装し実際に軽音楽団体での活動に活用可能かを調査する必要がある。さらに、メディア性を強調することがリアルさを増幅させる場合があるとされており [11]、提案手法による臨場感の向上が期待される。

そこで本研究では、リアルタイムでライブ映像配信へ映像効果を自動付与可能なシステムを実装し、その性能評価と現場実証を行う。また実際の利用に基づくログとアンケートから評価を行い、「本システムはリアルタイムでの運用可能か」「映像効果の付与によりオンラインライブの臨場感が増加するか」という問いのもと、それぞれの観点から議論する。

<sup>1</sup> 明治大学  
Meiji University

## 2. 関連研究

### 2.1 ライブ映像の臨場感向上に関する研究

ライブ映像の臨場感を向上させるため、VR コンサートの分野では、HMD (Head Mounted Display) と小型脳波レコーダーを組み合わせることで脳の状態を共有するシステムが複数提案されている [12,13]。これらの研究では、生体信号をフィードバックとして提示することで、参加感や一体感に関する主観評価が向上する可能性が示されている。一方、柳沢ら [14] は、ライブ映像配信を閲覧するスマートフォンを高輝度 LED で囲い、映像と同期して点灯させることで臨場感を高める手法を提案している。小さな画面での視聴を行った場合でもディスプレイを大きく感じることができ、非日常的な体験を提供可能としている。

これらのような特有のデバイスを用いることは他に無い体験を提供している一方、そのデバイスを用意できるユーザは少ないため汎用性に欠ける。本研究で構築したシステムでは、これまで提案してきた映像効果自動付与をリアルタイムで処理し、かつ配信・視聴の両方をスマートフォンのみでできるため、誰でも簡単に臨場感のあるオンラインライブを行うことができる。

臨場感を向上させるため映像処理を用いる研究もあり、一視点型ライブ映像に対して、オプティカルフローの変化が大きいフレームで画面中央にズームすることで、周辺視野に刺激を与え臨場感を向上させる手法が提案されている [15]。また、実験により臨場感向上に寄与していることを示唆している。

本研究では、映像のオプティカルフローの変化に加えてオンセット情報を用いることで、より音楽的に適した箇所でのカットを行う。さらにズーム以外の映像効果も加えた上で、リアルタイムで運用可能なシステムを提案する。

### 2.2 臨場感へ影響を及ぼす要素に関する研究

Lee [16] は、臨場感を物理的存在感、社会的存在感、自己存在感の3つに分類している。また、デジタル化した音楽イベントにおける臨場感について Zhong ら [17] は、視聴者が感じる臨場感は観客 (他視聴者) と一緒にいる感覚、演者が目の前にいる感覚、空間的にその場にいる感覚、身体反応の4つの概念から成ると説明しており、その中でも特に観客や演者と繋がっている感覚が重要であると述べている。これらを踏まえると、没入感、社会的存在感、演者との繋がりは、臨場感が成立するうえで重要であることがわかる。

没入感、社会的存在感、演者との繋がりを指標を用いて音楽体験を評価している研究は複数存在する。Venkatesan ら [18] は、音楽鑑賞時の振動触覚フィードバックの有無による影響を調査するために、アーティストへの共感、パラ

ソーシャル関係、ファン度合い、社会的存在感、孤独感を指標として採用している。また Swarbrick ら [19] は、ライブでの没入感を測る尺度について、音楽への没頭尺度である AIMS [20] とファン度合いが有効であると述べている。

上記の指標の中で特に、音楽・ライブへの没入感、他の視聴者や演者との間の社会的存在感、演者やそのライブにどのくらい興味を持っているかは、軽音楽団体のオンラインライブにおいて臨場感のある音楽体験をしてもらうため重要であると考えられる。そのため本研究では、軽音楽団体におけるオンラインライブの臨場感を測る指標として、没入感、社会的存在感、演者への興味度を採用し評価を行う。

## 3. OneCamLiveFX

我々はこれまでの研究 [5,6] において、映像効果自動付与手法がライブ映像の臨場感を向上させることを明らかにしてきた。これらの結果を軽音楽団体でのライブに活用するためには、リアルタイムで運用可能にする必要があるが、これまで実現してきたシステムは一旦オフライン処理を行う必要があった。

そのため、一視点から撮影されたライブ映像に対して、演者の認識と動作量推定、および発音タイミング検出をし、その結果から映像効果をリアルタイムに付与して配信する手法「OneCamLiveFX」を提案する。なお、メディアにおけるライブ性について、視聴者がリアルタイムであると認識することが重要であり [21]、視聴中心のケースでは数秒の遅延は許容される [22] とあるため、本手法ではリアルタイムを10秒程度の遅延は許容するものとし、その範囲で配信および自動的に映像効果を付与することを目指す。

本手法では、汎用的に利用してもらうため、スマートフォンのみで配信・視聴のできるような環境の構築を目指す。ここでは撮影を開始すると、バックエンドで特徴的なシーンを抽出および映像効果付与し、配信者と視聴者はそれぞれのブラウザでその配信を視聴することを可能とする。

バックエンドではまず、映像から演者認識、動作量算出を行い、音声からオンセット検出をする。また、オンセットのタイミングで映像効果の対象と映像効果の種類を決定し付与する。このオンセットを付与タイミングとすることで、音楽的に適した箇所画面を切り替えることが可能となる。さらに動作量を基に付与対象を決定することで、演奏やそれに付随したパフォーマンスが特徴的なシーンを検出する。ここで、映像効果付与の前に意図的な遅延を導入する。これにより同じフレームに対する動作量算出、映像効果付与のタイミングを数秒ずれることで、オンセットの前後の動作量を参照することができるようになり、文脈に沿った映像効果の付与が可能となる。

映像効果の種類について、以前の研究では視覚における心理的側面 [23,24] から、ズーム、ラディアルブラー、画面



図 1: 適用する映像効果

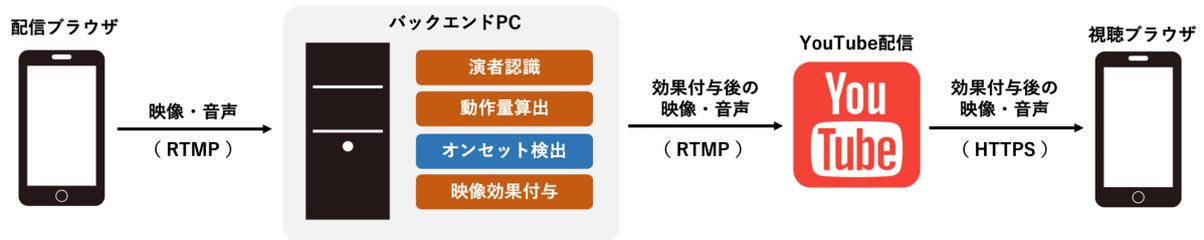


図 2: リアルタイム映像効果付与システムの処理の流れ

分割を採用していた。しかし、ラディアルブラーはリアルタイム処理ではが重く、遅延の要因になると判断した。そこでラディアルブラーのかわりに、同様の集中効果を持つビネットを採用した。これらの映像効果はそれぞれ、ズームは1人の演者にフォーカスを当てることによる画面切り替え、ビネットは集中効果による映像への引き込み、画面分割は複数の演者にフォーカスを当てることによるバンドの一体感や個々の演奏を伝える効果をもっている（図1）。

#### 4. システムの実装

全体的な処理の流れを図2に示す。まず配信デバイスのブラウザで取得した映像・音声の中継サーバを介して、バックエンドPCへ送られる。バックエンド側では演者認識、動作量推定、発音タイミング検出を逐次実行し、特徴的なシーンで映像効果を付与した後、YouTube Liveへ送出する。ここで、ライブ映像には演者のみでなく観客が同時に映ってしまうことが予想されるため、画面上部に映る演者と下部に映る観客を区切る基準線を配信者が指定可能とする。YouTube Liveに出力された映像は、視聴者、配信者ともにブラウザから視聴可能である。本章では、バックエンドの実装と配信・視聴インタフェースについて述べる。

##### 4.1 バックエンド処理の実装

バックエンドでは、リアルタイム性の確保、ストリームのスムーズな受信・送信が重要となる。そのため図3のように複数のスレッドに分けて処理を行う。readerスレッドでは、配信ブラウザから受け取ったストリームを読み込み、次にanalyzerスレッドで演者認識や動作量算出、オンセット検出の処理を行う。得られた特徴量をもとにfx・encoderスレッドで映像効果を付与したうえでエンコード

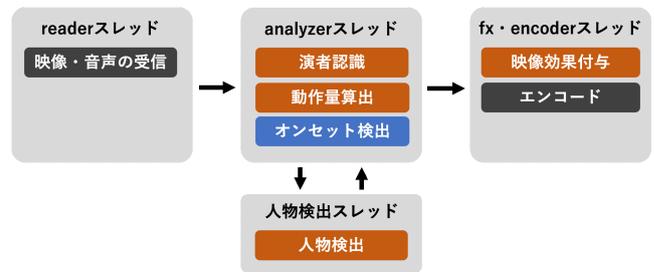


図 3: バックエンド内の処理の流れ

し、YouTubeLiveに配信する。

まずreaderスレッドでは、配信ブラウザからRTMP\*1で送出されたストリームを受信し、映像、音声、映像framehash、音声framehashの4つのパイプ出力に分岐させるようにした。またframehashからは映像フレーム/音声チャプクのPTS(時刻)情報を取得した。以降の処理では、このPTS情報を用いて映像・音声の時刻基準を揃え、同期を保ったまま出力するようにした。

analyzerスレッドでは演者認識、動作量算出、オンセット検出を行うようにした。ここで、人物検出は特に計算コストが高く、同一スレッドで行うとリアルタイム性を損ないやすい。そのため、人物検出スレッドを別に設けた。人物検出スレッドではYOLOv8nを用いて人物検出とID追跡を行い、基準線より上の人物を演者として認識した。また、演者ごとの動作量を一定時間分履歴として保存するようにした。オンセット検出では、その瞬間のRMSを長期的なRMSで割ることで相対的な値に正規化し、平均と標準偏差を用いた動的な閾値を設定することで、演奏の音の大きさのトレンドに左右されることなくオンセットを検出

\*1 Real-Time Messaging Protocol. ストリーミング向けの通信プロトコルであり、配信者から配信サーバへの映像・音声の送信で広く用いられている。



図 4: 配信端末画面



図 5: 視聴端末画面

できるようにした。さらに、意図的遅延の追加は演者認識、オンセット検出後行うようにした。ここでは意図的遅延を5秒に設定した。

fx・encoder スレッドでは、映像はオンセット検出のタイミングで動作量履歴を参照して映像効果を決定・付与するようにした。ここで、映像効果を付与する時点で既に遅延があるため、前後の動作量を参照可能となる。これらの映像効果の付与については、以下の基準で適用した。

- ズーム：演者の動作量が 1.2 倍以上増加した際に適用
- ビネット：演者の動作量が 2.0 倍以上増加した際に適用
- 画面分割：複数人の動作量が 2.0 倍以上増加した際に適用

最後に、映像・音声をエンコードし、PTS に基づく出力時刻を付与したうえで、最終的に同一コンテナへ多重化し、YouTube Live Streaming API を用いて YouTube Live へ RTMP 送出した。

なお、PC には GPU (NVIDIA GeForce RTX 4090, 24GB) を搭載したものを用了。

#### 4.2 配信・視聴インタフェース

ユーザは配信端末 (図 4) の中央下部のボタンを押すことで、配信の開始・停止を操作することができる。また基準線を赤く表示しており、撮影環境に応じてドラッグアンドドロップにより調整できる。配信中は、左右下部に2つのボタンがあり、映像効果を適用 (FX:ON) か不適用 (FX:OFF) を切り替えることと、撮影カメラの映像と視聴端末の映像を切り替えることができる。この映像効果に関するボタンを用いることで、配信者は映像効果付与機能の使い分けをすることができる。一方、視聴端末 (図 5) では、音声の ON/OFF 操作に加え、配信の最新部分へジャンプできるボタンを配置し、最新のライブを視聴できるようにした。

### 5. 実験

#### 5.1 実験目的

本実験の目的は、提案システムを実環境 (軽音楽サークルのライブ) で運用し、リアルタイム配信として成立するか、また視聴者・演者の主観評価が高く継続利用が見込め

るかを検証することである。

#### 5.2 実験設計

本実験では、システムログによる客観評価と、アンケートによる主観評価を併用して評価を行った。

客観評価について、各項目と評価の対象を以下に示す。

- 処理時間：どのくらいの遅延で配信されているか、安定して配信できているか
- FPS・フレームドロップ数：映像自体の安定性
- Buffering events：ストリームの送受信の安定性

処理時間は一定時間ごとに計測し、パイプライン全体の処理時間に加えて主要処理の内訳も記録した。

また主観評価として、視聴者、演者にそれぞれアンケートを行った。本研究では、臨場感の関連概念として、没入感、社会的存在感、演者への興味度に着目し評価を行うため、没入感を含むより広範な体験を指す音楽吸収に特化した評価指標である AIMS (Absorption in Music Scale) [20]、メディア環境における存在感を測る指標である TPI (Temple Presence Inventory) [25]、特定の対象に対する個人のファン度合いを評価する指標である Fanship Scale [26] を参考にした。質問項目を、本実験の内容に合うように調整したのち、没入感に関する3項目、社会的存在感に関する5項目、演者への興味度に関する3項目を含んだアンケートを作成した (表 1)。

さらに本システムの満足度および継続利用可能性を評価

表 1: 臨場感に関するアンケート

番号	没入感を測る質問
Q1	演奏に夢中になりましたか。
Q2	時間の感覚が短く感じましたか
Q3	気を散らさずに集中できましたか。
社会的存在感を測る質問	
Q4	感覚的に演者をより近くに感じましたか。
Q5	演者の存在感はありましたか。
Q6	演奏を感情的 (躍動的) に感じましたか。
Q7	演奏に現実味がありましたか。
(Q1)	(演奏に夢中になりましたか。)
演者への興味度を測る質問	
Q8	演者に興味を持ちましたか。
Q9	演奏を積極的に視聴しましたか。
Q10	もし演者が演奏を成功・失敗すると心が揺らぎますか。

表 2: 満足度及び継続利用可能性に関するアンケート

番号	満足度を測る質問
Q1	とても不満だーとても満足している
Q2	とても不快だーとても心地よい
Q3	非常にいらいらしたー非常に満ち足りていた
Q4	最悪だったー最高だった
継続利用を測る質問	
Q5	今後もこのライブ配信システムで視聴してみたい。
Q6	他の視聴方法があったとしてもこのライブ配信システムで視聴してみたいと思う。
Q7	もしできるならこのライブ配信システムでの視聴をやめたいと思う。(逆転項目)

するため、情報システムの継続利用を説明する期待確認モデルに基づく尺度を参考にした [27]。本研究では満足度と継続利用意図の 2 指標を重要視し、先行研究の尺度項目を本システムの文脈に合わせて表現を調整しアンケートを作成した (表 2)。満足度は 7 件法の両極形容詞対で測定し、継続利用意図は 7 件法リッカート尺度で測定するようにした。

視聴者には当日、臨場感、満足度及び継続利用可能性に関するアンケートに回答してもらった。また、演者にも配信された映像を後日確認してもらい、満足度及び継続利用可能性を評価するためのアンケートに回答してもらった。

### 5.3 実験手順

実験は、ある軽音楽サークルのライブにおいて実際に撮影・配信を行う形で実施した。配信は合計 3 回行い、視聴者はそのいずれか 1 回を視聴してもらうこととした。

実験ではまず、それぞれのライブが始まる直前で著者が配信デバイス进行操作し、カメラの角度や基準線の設定をした後、配信を開始した。視聴者には自由にその配信に参加してもらい、そのライブが終了するまで (約 20 分程度) 視聴してもらった。また配信終了後、視聴者には、臨場感に関するアンケート (没入感, 社会的存在感, 演者への興味度) と満足度及び継続利用可能性に関するアンケートに答えてもらった。さらに後日、演者には自身が出演していたライブの配信を確認してもらった後、満足度及び継続利用可能性に関するアンケートに答えてもらった。

## 6. 結果

視聴者として、普段から音楽をよく聴く人を対象に募集をし、14 人 (男性 7 人, 女性 7 人) が実験に参加した。また、演者として筆頭著者が所属する軽音サークルの 12 人 (男性 6 人, 女性 6 人) が実験に参加した。配信は 3 回行ったが、そのそれぞれの参加人数は表 3 の通りであった。なお、参加した視聴者、演者には所要時間に応じ、謝金を支払った。

以下では、実験の結果について述べる。

### 6.1 システム処理時間

配信端末から視聴端末にストリームがどのくらいの遅延で届くのか、提案システムが安定して配信できているかを調べるために、システム処理時間のログを分析した。

表 4, 5 は、本システムにおけるバックエンドにおけるパイプラインの処理時間 (遅延時間) を FX の OFF, ON で分けてまとめたものである。ここでの処理時間は意図的な遅延 (本実験では 5 秒) を含んでいる。ここでは、極端な外れ値の影響を抑えつつ遅延が増加した際の傾向を見るために、全体の 90 パーセント目に当たる数値である p90 も示している。表 4 より、FX を OFF にした場合のパイプラインの処理時間は 3 回の配信で概ね安定しており、全体の中央値は 6.11 秒であった。しかし、3 回目の配信において最大値は 9.30 秒となっており、一部の区間で一時的な遅延増加が生じていた。一方、表 5 より、FX を ON にした場合は中央値こそ 6.53 秒と FX:OFF と同程度であるものの、p90 は 10.66 秒まで増加し、最大値は 17.08 秒に達した。

次に、FX:ON 時におけるモジュール別の処理時間内訳

表 3: 各配信の参加人数

配信	視聴者 (人)	演者 (人)
1 回目	5	3
2 回目	4	4
3 回目	5	5
全体	14	12

表 4: パイプライン全体の処理時間 (FX:OFF)

配信	中央値 (s)	p90 (s)	平均 (s)	最小 (s)	最大 (s)
1 回目	6.13	6.16	6.13	5.00	6.71
2 回目	6.07	6.11	6.07	5.00	6.54
3 回目	6.12	6.15	6.12	5.00	9.30
全体	6.11	6.15	6.11	5.00	9.30

表 5: パイプライン全体の処理時間 (FX:ON)

配信	中央値 (s)	p90 (s)	平均 (s)	最小 (s)	最大 (s)
1 回目	6.18	10.01	7.35	6.10	17.08
2 回目	6.13	9.89	7.27	6.04	16.12
3 回目	7.53	11.77	8.21	6.11	17.00
全体	6.53	10.66	7.62	6.04	17.08

表 6: モジュール別処理時間の内訳

モジュール	平均 (ms)	p90 (ms)	頻度 (Hz)	割合 (%)
YOLO+トラッキング	97.57	109.84	6.15	42.5
映像効果付与 (FX)	17.42	54.85	20.76	25.6
エンコード (H.264)	6.33	8.68	30.65	13.7
JPEG デコード	3.32	3.74	30.75	7.2
オンセット検出	0.17	0.30	48.06	0.6
動作量集計	0.02	0.03	6.15	0.0
その他 (キュー/IO 等)	4.75	6.99	30.65	10.3

表 7: バックエンド実効出力処理率 (FX:OFF)

配信	中央値 (fps)	p10 (fps)	p90 (fps)	平均 (fps)
1 回目	29.93	29.32	30.81	30.35
2 回目	29.91	29.33	30.82	30.35
3 回目	29.92	29.34	30.79	30.15
全体	29.92	29.33	30.81	30.25

を表 6 に示す。この表は、各モジュールの 1 回あたりの処理時間に加え、呼び出し頻度を考慮して映像 1 秒あたりの処理時間を算出し、その割合を示したものである。その結果、YOLO+トラッキングが 42.5% と多くの割合を占めており、次いで映像効果付与が 25.6% となった。また、エンコードは 13.7%、JPEG デコードは 7.2% であった。

## 6.2 FPS・フレームドロップ・Buffering events

映像の安定性を調べるために、FPS とフレームドロップを分析した。

まず配信端末から送出する時の FPS は、中央値 30.00 fps、平均 29.99 fps であった。

バックエンドでは、YouTube Live へ送出する段階において、1 秒間に出力できたフレーム数を「実効出力処理率」として記録した。本指標はストリームのフレームレートではなく、fx・encoder スレッドへの順番待ちキューから YouTube Live への送出へ到達できたフレーム数を表している。そのため、処理遅延によりキューに滞留したフレームを短時間に処理して追いつく場合、30 fps を超える値を取りうる。表 7 を見ると、FX:OFF 区間では、実効出力処理率の中央値は 29.92 fps、平均は 30.25 fps であり、p10 は 29.33 fps、p90 は 30.81 fps であった。一方、表 8 を見ると、FX:ON 区間では、中央値は 29.52 fps、平均は 29.99 fps であったが、p10 は 14.41 fps まで低下し、p90 は 56.82 fps まで増加した。

フレームドロップについては、FX:OFF 区間では全キューで 0 件であったのに対し、FX:ON 区間では、fx・encoder スレッドへの順番待ちキューにおいて合算 215 フレーム (1 回目 136 件、2 回目 20 件、3 回目 59 件) のドロップが発生した。一方、その他のキューではドロップは観測されなかった。

表 8: バックエンドの実効出力処理率 (FX: ON)

配信	中央値 (fps)	p10 (fps)	p90 (fps)	平均 (fps)
1 回目	29.70	14.10	46.59	29.99
2 回目	29.67	14.96	48.93	29.97
3 回目	26.69	14.38	66.20	30.01
全体	29.52	14.41	56.82	29.99

送受信の安定性を評価するためにバックエンドの受信・送信箇所での PTS を基に Buffering events の数をカウントしたところ、どの回においても FX の ON/OFF に関わらず、Buffering events は確認されなかった。

## 6.3 視聴者アンケートによる主観評価

没入感、社会的存在感、演者への興味度の結果を表 9 に示す。各指標は 6 件法に基づく複数項目から構成し、参加者ごとに項目平均を算出して分析した。また補助的に、中点を上回るかについて片側の 1 標本の Wilcoxon 符号付順位検定を実施した。その結果、全ての指標において中点を上回る傾向が見られた。

次に、満足度及び継続利用可能性の結果を表 10 に示す。満足度は 7 件法の両極形容詞対により測定し、参加者ごとに項目平均を満足度スコアとして算出した。継続利用意思は 7 件法リッカート尺度により測定し、同様に項目平均を算出した。その結果、満足度に関する評価では、全体として中点を上回る傾向が見られた。一方、継続利用意思は平均 4.26 (SD=1.50) であり、Wilcoxon 符号付順位検定の結果、有意な差は見られなかった ( $p = 0.276$ )。

なお、満足度と継続利用意思の順位相関は強くなく (Spearman の  $\rho = 0.18$ ,  $p = 0.539$ )、両者の関連は明確ではなかった。

表 9: 視聴者の臨場感に関する主観評価

指標	平均	SD	p 値
没入感	4.00	0.92	.041*
社会的存在感	4.53	0.79	.004**
演者への興味度	4.14	1.00	.027*

注: \* $p < .05$ , \*\* $p < .01$ .

表 10: 視聴者の満足度および継続利用意思の主観評価

指標	平均	SD	p 値
満足度	5.02	0.75	< .001***
継続利用意思	4.26	1.50	.276

注: \*\*\* $p < .001$ .

表 11: 演者の満足度および継続利用意思の主観評価

指標	平均	SD	p 値
満足度	5.18	1.35	.011*
継続利用意思	5.18	1.35	.011*

注: \* $p < .05$ .

## 6.4 演者アンケートによる主観評価

演者に対しても、視聴者と同様に満足度および継続利用意思を測定し、参加者ごとに項目平均を算出した。その結果を表 11 に示す。この表より、満足度、継続利用意思のどちらの指標においても中点を上回る傾向が見られた。

また、満足度と継続利用意思の関連を検討したところ、両者には強い正の相関が認められた (Spearman の  $\rho = 0.89$ ,  $p < 0.001$ )。参考として単回帰分析を行った結果も同様の方向性を示し (傾き  $b = 1.36$ )、満足度が高い演者ほど継続利用意思も高い傾向が見られた。

## 7. 考察

### 7.1 ログによる客観的な評価

パイプライン全体の処理時間の結果から、映像効果の有無に関わらず平均値、中央値ともに 10 秒以内であり、リアルタイムの要件は満たしていることがわかる。そのため本システムはリアルタイムでも運用可能であると考えられる。

一方 FX:ON では、遅延に関して中央値は FX:OFF の時と大きく変わらないものの、p90 や最大値が増加していた。これは、FX を有効化した際に、処理負荷の増加により一部区間でリアルタイム処理に追従できず、遅延のばらつきが発生したことを示唆する。また、3 回目の配信では中央値が 7.530 秒と他の回より大きくなっていった。さらに、モジュール別で比較をすると、YOLO+トラッキングが 42.5%、映像効果付与 25.6%と多くの割合を占めていた。これは 3 回目の配信では演者数が 5 人と一番多く、人物検出や映像効果付与の負荷が最もかかりやすかったことが原因として考えられる。つまり、人物検出・追跡と映像効果処理が主要な計算ボトルネックであり、これらの負荷が増大する区間では遅延の増加や追従性低下を引き起こしやすいと考えられる。また、演者数などの環境条件によってリアルタイム追従性が変動する可能性がある。

配信端末から送出する時の FPS は、中央値 30.00 fps、平均 29.99 fps であり、配信端末から送出する処理自体は安定しており、ボトルネックになりやすいことが確認できた。一方、バックエンドの実効出力処理率に関しては、FX を有効化した際に分布が大きく拡大していた。これは、人物検

出や映像効果付与による負荷増大により、処理が一時的に遅れて出力が低下する区間と、その遅れを取り戻すために短時間でまとめて処理が進み出力が増える区間が生じていたことを示している。またフレームドロップも発生しており、実質的な FPS は部分的に下がっていると考えられる。

### 7.2 アンケートによる主観的な評価

視聴者アンケートでは、Wilcoxon 符号付順位検定の結果、没入感・社会的存在感・演者への興味度のいずれも中点を有意に上回り、特に社会的存在感は効果量  $d = 1.31$  と大きかった。特に社会的存在感の評価が高かったため、リアルタイムであることが演者との繋がりを強化した可能性が高い。これらのことから、本手法は一視点からのオンライン配信ライブの臨場感向上に効果的にアプローチできていると解釈できる。

満足度は中点を明確に上回った一方、継続利用意思は有意差が確認されなかった。ここで自由記述においては「自分でズーム対象を決めるなどの操作をしたい」という意見が多く見られた。このことは、受動的な自動演出だけでは体験の決定要因になりにくく、主体性やパーソナライズが継続利用に繋がる可能性を示している。また、「演奏に関係ない部分にズームされていた」という意見も見られ、演出の誤適用が満足度を損なうリスクを示しており、人物検知や映像効果付与の精度向上が重要であると考えられる。今回の撮影では、演者の配置や照明が比較的固定である状況を前提としていた。演者の動きが激しかったり照明条件が厳しかったりする環境では、本実験より誤検出・誤適用が増える可能性があるため精度の改善が必要である。そのため、色調補正やコントラスト調整などにより、人物とそれ以外の境界を明確にする必要があると考えられる。

一方、演者アンケートでは満足度、継続利用意思がいずれも中点を有意に上回り、さらに両者には強い正の関連が観測された。これは視聴者と異なる結果となっており、演者にとって「自分 (または演者) が強調される」ことの価値が大きく、そのことから継続意図も高くなっていることを示唆している。そのため、今後システムを改善していくうえで、まずは視聴者の QoE 向上を目指すことが、視聴者・演者の両者にとって満足度の高いものとするために重要であると考えられる。

なお、視聴者、演者共に人数が少なく十分に足りていたとは言えないため、一般化には追加データが必要である。今後はより大規模な環境で実験を行っていく予定である。

## 8. まとめ

本研究では、軽音楽団体におけるオンラインライブの臨場感が薄くなってしまっている問題に着目し、リアルタイムで映像効果を自動付与、スマートフォンのみで配信・視聴のできるシステム「OneCamLiveFX」を実装した。本

システムでは、演者の動作量、オンセットから特徴的なシーンを検出し、そのタイミングで映像効果の付与をすることで、臨場感の向上を狙っている。

実験では、実際のライブでの現場検証を行いログとアンケートから評価を行った。その結果、リアルタイムでの運用は可能であるが、一部区間で遅延の増加やフレームドロップが見られ、撮影環境によってリアルタイム追従性が変動する可能性があることが示唆された。またアンケートから、視聴者の臨場感向上に効果的にアプローチできていること確認した。一方で、継続利用意思は視聴者と演者で差が見られたため、視聴者の QoE 向上を目指すことがシステム全体の改善に繋がると考えられる。

今後は、色調補正やコントラスト調整による映像効果付与の精度向上や、視聴者・演者の関与を取り込むインタラクティブな設計による影響の調査をすることで、軽音楽団体の配信活動における実運用上の価値向上を進めていく。

## 参考文献

- [1] Modestini, P. and Weining, C.: Affordances and experiential dimensions of digital concerts: Closer the eye, farther the crowd, *Music Perception: An Interdisciplinary Journal*, Vol. 42, No. 3, pp. 269–285 (2025).
- [2] Ludwig, M. and Bertling, C.: The Effect of Cutting Rates on the Liking of Live Sports Broadcasts, *International Journal of Sport Communication*, Vol. 10, pp. 359–370 (2017).
- [3] 総務省：令和 6 年通信利用動向調査の結果，[https://www.soumu.go.jp/johotsusintokei/statistics/data/250530\\_1.pdf](https://www.soumu.go.jp/johotsusintokei/statistics/data/250530_1.pdf). 参照 2025-12-09.
- [4] 畑田豊彦, 坂田晴夫, 日下秀夫: 画面サイズによる方向感覚誘導効果 大画面による臨場感の基礎実験, *テレビジョン学会誌*, Vol. 33, No. 5, pp. 407–413 (1979).
- [5] 小川剣二郎, 中村聡史: 一視点固定型ライブ映像への映像表現自動付与による臨場感向上手法の検討, *情報処理学会研究報告エンタテインメントコンピューティング (EC)*, Vol. 2024-EC-71, No. 37, pp. 1–8 (2024).
- [6] 小川剣二郎, 中村聡史: 1 視点固定型ライブ映像における映像・音響表現自動付与による音楽体験拡張, *情報処理学会論文誌*, Vol. 66, No. 12, pp. 1715–1724 (2025).
- [7] Georgi, C.: *Liveness on stage: Intermedial challenges in contemporary British theatre and performance*, Vol. 25, Walter de Gruyter GmbH & Co KG (2014).
- [8] Wald-Fuhrmann, M., Egermann, H., Czepiel, A., Katherine O’Neill, Weining, C., Meier, D., Tschacher, W., Uhde, F., Toelle, J., Tröndle, M.: Music listening in classical concerts: Theory, literature review, and research program, *Frontiers in Psychology*, Vol. 12 (2021).
- [9] Onderdijk, K. E., Swarbrick, D., Van Kerrebroeck, B., Mantel, M., Vuoskoski, J. K., Maes, P.-J. and Leman, M.: Livestream experiments: the role of COVID-19, agency, presence, and social context in facilitating social connectedness, *Frontiers in Psychology*, Vol. 12 (2021).
- [10] Kulla, S. T., Sungur, H. and Sumter, S. R.: The role of social presence in live and recorded concert viewing: Effects on enjoyment and emotional well-being, *Computers in Human Behavior Reports*, Vol. 14 (2024).
- [11] Bolter, J. D. and Grusin, R.: *Remediation: Understanding New Media*, The MIT Press, Cambridge, MA (1999).
- [12] Muñoz-González, Á., Kobayashi, S. and Horie, R.: A multiplayer vr live concert with information exchange through feedback modulated by eeg signals, *IEEE Transactions on Human-Machine Systems*, Vol. 52, No. 2, pp. 248–255 (2022).
- [13] Ide, K. and Horie, R.: Improving a Sense of Unity via BrainWaves in a Virtual Concert Where Performances in Real Space are Watched in a Virtual Space., *ICAT-EGVE (Posters and Demos)*, pp. 25–26 (2022).
- [14] 柳沢豊, 小野圭介, 上田健太郎, 出田怜, 吉池俊貴, 藤本実ほか: Immersive Online Live System: ライブ配信動画に同期した演出が可能な LED 点灯制御システム, *研究報告デジタルコンテンツクリエイション (DCC)*, Vol. 2021, No. 2, pp. 1–5 (2021).
- [15] 大石あい, 上岡英史: Zooming を用いた音楽ライブ映像の臨場感改善手法, *電子情報通信学会技術研究報告; 信学技報*, Vol. 121, No. 23, pp. 7–12 (2021).
- [16] Lee, K. M.: Presence, explicated, *Communication theory*, Vol. 14, No. 1, pp. 27–50 (2004).
- [17] Zhong, D., Li, Q. and Zhai, X.: The digital transformation of music events: from being there to the feeling of being there, *Journal of Hospitality and Tourism Technology*, Vol. 16, No. 5, pp. 1193–1212 (2025).
- [18] Venkatesan, T. and Wang, Q. J.: Feeling connected: The role of haptic feedback in VR concerts and the impact of haptic music players on the music listening experience, *Arts*, Vol. 12, No. 4, p. 148 (2023).
- [19] Swarbrick, D., Martin, R., Høffding, S., Nielsen, N. and Vuoskoski, J. K.: Audience musical absorption: Exploring attention and affect in the live concert setting, *Music & Science*, Vol. 7 (2024).
- [20] Sandstrom, G. and Russo, F.: Absorption in music: Development of a scale to identify individuals with strong emotional responses to music, *Psychology of Music*, Vol. 41, pp. 216–228 (2013).
- [21] Van Es, K.: Liveness redux: on media and their claim to be live, *Media, Culture & Society*, Vol. 39, No. 8, pp. 1245–1256 (2017).
- [22] Li, J., Li, Z., Lu, R., Xiao, K., Li, S., Chen, J., Yang, J., Zong, C., Chen, A., Wu, Q. et al.: Livenet: a low-latency video transport network for large-scale live streaming, *Proceedings of the ACM SIGCOMM 2022 Conference*, pp. 812–825 (2022).
- [23] Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A. and Brandt, S. A.: A physiological correlate of the “zoom lens” of visual attention, *Journal of Neuroscience*, Vol. 23, No. 9, pp. 3561–3565 (2003).
- [24] Smith, W. S. and Tadmor, Y.: Nonblurred regions show priority for gaze direction over spatial blur, *The Quarterly Journal of Experimental Psychology*, Vol. 66, No. 5, pp. 927–945 (2013).
- [25] Lombard, M., Ditton, T. B. and Weinstein, L.: Measuring presence: the temple presence inventory, *Proceedings of the 12th annual international workshop on presence*, International Society for Presence Research Los Angeles, CA, pp. 1–15 (2009).
- [26] Reysen, S. and Branscombe, N. R.: Fanship and Fandom: Comparisons between Sport Fans and Non-Sport Fans, *Journal of Sport Behavior*, Vol. 33, No. 2, pp. 176–193 (2010).
- [27] Bhattacharjee, A.: Understanding information systems continuance: An expectation-confirmation model, *MIS quarterly*, pp. 351–370 (2001).