

2025年度 修士学位請求論文

一視点固定型ライブ映像における
臨場感向上に関する研究

明治大学大学院先端数理科学研究科
先端メディアサイエンス専攻

小川 剣二郎

Master's Thesis

A Study on Enhancing the Sense of Presence in
Single-Viewpoint Live Video

Frontier Media Science Program,
Graduate School of Advanced Mathematical Sciences,
Meiji University

Kenjiro Ogawa

概要

軽音楽団体によるオンライン配信ライブでは、撮影機材の制約から一視点固定の映像が用いられることが多く、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。本研究は、このような一視点固定型ライブ配信において、ライブ映像中の特徴的なシーンを自動検出し、適切な映像効果を付与することで、臨場感の向上と視聴体験の改善を目指す。

まず、漫画・アニメ等に見られる音楽表現を参考に、ズーム、集中線、画面分割等の映像効果をライブ映像へ付与するプロトタイプシステムを実装し、映像効果付与が視聴維持率や印象に与える影響を調査した。その結果、映像効果を付与した条件で視聴維持率、エンゲージメントが向上し、一視点固定型ライブ映像の単調さを緩和することによる影響を確認することができた。

次に、映像効果と音響効果を組み合わせて適用する手法を構築し主観評価実験を行った。没入感・社会的存在感・演者への興味度を指標として評価を行い、映像効果付与によりこれらの評価が向上することを明らかにした。一方、音響効果の明確な効果は確認されず、個人差や提示方法に依存する可能性が示唆された。

最後に、リアルタイム運用を想定したシステム「OneCamLiveFX」を実装した。意図的遅延を導入し発音前後動作量を参照可能とすることで、文脈に基づく効果付与を実現した。ログに基づく性能評価とライブ現場での実証実験を通して、リアルタイム環境における演出強化と追従性のトレードオフを明確化し、運用可能性と今後の改善指針を示した。加えて、視聴者アンケートから臨場感や満足度の指標が評価された一方で、誤適用が体験を損ね得ることや、主体性やパーソナライズが継続利用に繋がる可能性が示唆された。

以上より、一視点固定型ライブ配信に対して、自動的に映像効果を付与することで視聴体験を改善し得る可能性を示した。今後は、映像効果適用タイミングの判定の精度向上に加え、視聴者・演者の選好を反映できるインタラクティブな操作機能の導入により、より安定したリアルタイム運用と体験価値の向上を目指す。

Abstract

Online live streams by light music groups are often recorded with a single fixed camera due to limited filming equipment, which can make the video visually monotonous. Moreover, many viewers watch on smartphones with small screens, so the sense of presence tends to be diminished. This study aims to improve the viewing experience of such single-view live streams by automatically detecting characteristic moments and applying appropriate video effects.

First, inspired by musical expressions in comics and animations, we examined how video effects such as zooming and split-screen layouts influence viewing retention and impressions. The results suggested that video effects can improve viewing retention, while simply increasing the variety or intensity of effects does not necessarily yield further benefits, indicating the importance of careful timing and frequency design.

Next, we combined video effects with audio effects and conducted a subjective evaluation using immersion, social presence, and interest in performers as metrics. Video effects tended to improve these metrics, whereas the impact of audio effects was not consistently observed.

Finally, we implemented a real-time system, OneCamLiveFX, which introduces an intentional delay to enable context-aware effect control around onsets. Log-based evaluation and an in-the-wild demonstration clarified a trade-off between richer presentation and real-time responsiveness. Audience questionnaires also indicated generally positive experiences, while highlighting risks of misapplied effects and demands for better operability and personalization. Overall, the study shows the potential of automatic video effects for enhancing single-view live streams, and suggests improving performer identification and providing interactive controls as future work.

目次

第1章	はじめに	1
1.1	ライブの開催形態	1
1.2	軽音楽団体におけるライブ映像配信の問題点	1
1.3	本研究の目的	2
1.4	本稿の構成	2
第2章	関連研究	4
2.1	臨場感へ影響を及ぼす要素に関する研究	4
2.2	メディアにおける臨場感の生成	5
2.3	ライブ映像の臨場感向上に着目した研究	5
第3章	映像効果の付与による影響の調査	7
3.1	実験概要	7
3.2	映像効果の検討	7
3.3	プロトタイプシステムの実装	9
3.4	実験設計	11
3.5	実験手順	12
3.6	結果	13
3.6.1	映像効果の適用条件ごとの視聴時間	13
3.6.2	映像効果の適用条件ごとのいいね数	14
3.6.3	アンケート	15
3.7	考察	16
第4章	映像・音響効果の付与に対する主観評価実験	18
4.1	実験概要	18
4.2	システムの改良と音響効果の検討	19
4.2.1	映像・音響効果の付与手法	19

4.2.2	演者の動作量算出の実装	21
4.2.3	映像効果の実装	21
4.2.4	音響効果の実装	22
4.3	実験設計	23
4.4	実験手順	25
4.5	結果	25
4.5.1	適用条件ごとの評価値	26
4.5.2	アンケート項目ごとの評価値	28
4.5.3	一視点固定型ライブ映像への満足度による評価	29
4.5.4	実験後アンケート	29
4.6	考察	30
第5章	リアルタイム配信のためのシステムの実装と評価	32
5.1	実験概要	32
5.2	リアルタイム映像効果付与システム	32
5.2.1	設計指針	32
5.2.2	システム概要	33
5.2.3	バックエンド処理の実装	34
5.2.4	配信・視聴インタフェース	35
5.3	実験設計	36
5.4	実験手順	37
5.5	結果	38
5.5.1	システム処理時間	38
5.5.2	FPS・フレームドロップ・Buffering events	40
5.5.3	視聴者アンケートによる主観評価	41
5.5.4	演者アンケートによる主観評価	42
5.6	考察	42
5.6.1	ログによる客観指標の評価	42
5.6.2	アンケートによる主観指標の評価	43
第6章	総合考察と今後の展望	45
6.1	総合考察	45

6.2 制約と今後の展望	46
第7章 まとめ	48
謝辞	49

第1章 はじめに

1.1 ライブの開催形態

軽音楽団体での活動では、定期的にライブを行い、またそれに向けて練習をすることが多くを占めている。ライブでは、演奏者自身が楽しむことや、その団体の人達に披露し盛り上げることが活動のモチベーションとなっている。その一方で、自分達の活動内容を広くアピールすることも重要であり、その際には対面ライブと並行してオンラインライブを行うことも多い。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。また軽音楽団体においては周知の役割が大きく、より多くの人に興味を持ってもらうことで団体の活性化へ繋がるものとなっている。

ここでライブは予測不能性が強く [1]、即時的である [2] 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 [1, 3, 4]。これらはオンラインライブでも同様であり、ただ録画した映像をSNSで配信するのではなく、リアルタイムでライブを伝えることでより視聴者にその魅力を伝えることができる。

1.2 軽音楽団体におけるライブ映像配信の問題点

手軽にライブ配信できるプラットフォームとして、Instagram や YouTube といったSNSが挙げられ、軽音楽団体によるオンラインライブでよく利用されている。しかしこれらのライブ映像配信において、多くの視聴者を留まらせることは難しい。視聴者が離脱してしまう要因としてはまず、軽音楽の部活やサークルでは撮影機材が乏しく、一視点から撮影された映像を配信することが多いため、映像に変化が無く単調となることが挙げられる。場面に応じたカットおよびズームの操作が臨場感の構成要素である没入感 [5] に影響を及ぼすこと [6, 7]、またカット頻度は視聴者の視覚的関心の維持に関係することが示されており [8]、一視点からの映像そのままではすぐに飽きられてしまう可能性がある。また別の要因としては、利用端末としてスマートフォンが多く用いられており [9]、結果

として視聴時の画角が小さくなり臨場感が薄くなることが挙げられる。臨場感は観察画角とともに単調増加し、 20° 以下の狭い画角では臨場感もかなり弱く、 60° 以上になると飽和傾向があることが明らかになっており [10]、スマートフォンでの視聴では画角は小さいために臨場感が薄くなってしまう。

オンラインライブの臨場感の向上を目的とした研究として、スマートウォッチを用いて心拍数を計測しそれをもとに精神状態を視覚化するもの [11, 12] や、スマートフォンの周囲にLED点灯装置を設置し映像に同期して点灯させるもの [13] などがある。これらは特有のデバイスを用いることで他に無い体験を提供しているが、そのデバイスを用意できるユーザは少ない。デバイスに依存しない手法で臨場感を向上させることが、より多くのユーザに視聴してもらう上で重要となる。以上のことをまとめると、一視点からのライブ映像配信において、視聴端末がスマートフォンのような小さい画面でも臨場感が伝わるような仕組みを、他のデバイスに頼らず実現する必要がある。

1.3 本研究の目的

本研究では前節で述べた、軽音楽団体でよく見られる一視点から撮影されたライブ映像配信では臨場感が乏しく視聴が維持されないことが多いという問題に着目した。そこで、ライブ映像の特徴的なシーンを自動で検出し、適した効果を付与することで臨場感を向上させる手法を提案し、その効果を実証することを目的とする。

具体的にはまず、初期検討として自動で映像効果を付与する手法を検討し、用意したライブ映像に適用することで視聴時間やエンゲージメントにどのような影響があるかを調査する。次に、映像に加えて音声への効果付与の可能性の検討と、視聴者の主観評価による検証を行う。さらに、これらの調査を踏まえ、リアルタイムで運用可能なシステムを実装し、その性能評価とライブ現場での実証実験を行う。最後にこれらの結果をもとに、今後の軽音楽団体におけるオンラインライブでの運用に向けた展望について考察をする。

1.4 本稿の構成

本稿は、本章を含む全7章から構成される。まず本章では、軽音楽団体におけるオンラインライブの意味合いとその問題点について述べた。これ以降、2章では、関連研究について述べ、本研究の位置付けを明確にする。3章では、映像効果の自動付与手法を検討するとともに、プロトタイプシステムを作成し、視聴時間やエンゲージメントへの影響に

関する調査の結果および考察について述べる。4 章では、映像効果に加えて音響効果を組み合わせることを検討し、主観評価実験を実施した結果および考察について述べる。5 章では、リアルタイムで運用可能なシステムの説明と、ライブ現場での実証実験を行った結果および考察について述べる。6 章では本研究の総合的な考察と今後の展望を述べ、7 章で本稿のまとめを行う。

第2章 関連研究

2.1 臨場感へ影響を及ぼす要素に関する研究

臨場感の概念として、Lombard [5] は、メディアを介した臨場感の概念として、気配、リアリティ、場所、没入感、メディア内の人をどう捉えるか、メディア自体をどう捉えるかの6つの見方があると述べている。また Lee [14] は、メディアを介した経験を、単に知覚する、対象を操作できる、相互作用できるという3つのレベルで整理し、レベルに応じて経験の性質が変化すると述べている。その上で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 [15] 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、身体反応として現れる臨場感の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な要素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 [16] 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 [17] 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 [18] とファン度合いが有効であると述べている。

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。

2.2 メディアにおける臨場感の生成

Wirthら [19]は、メディアにおける空間的臨場感が成立する過程を段階的な認知プロセスとしてモデル化している。それによれば、視聴者はまず映像や音響といったメディア上の手がかりに注意を配分し、そこから提示空間の心的モデルを構築する。次に、その空間を今自分が基準としている場所として認識すると、空間的臨場感が成立する。すなわち臨場感の成立には、注意の持続と空間モデル構築という認知的過程に基づくものと解釈できる。

さらに、メディアには、媒介性を消す、逆にメディア性を前景化するという相反する2つの志向が同時に存在し、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いる [20]。加えて、メディアにおける存在感を感覚情報の豊かさと相互作用性による主観的経験として定義している研究 [21] や、臨場感を感覚運動的一貫性や因果的整合性の2つの独立した側面として整理している研究 [22] があり、視覚的リアリズムが高くなるともこれらは成立し得ると示されている。

さらに映像の分野においてLangら [23]は、映像のカットやズーム、音量変化といった構造的特徴その頻度が自動的な注意反応を引き起こすと説明している。またブラー [24, 25] や放射構造 [26] といった視覚的特徴は視線を強く誘導する要因となるとされている。Mayerら [27]は、関連情報が空間的に近接して提示されることで理解が促進されることを示している。

本研究では、特徴的なシーンに対して映像・音響効果を付与することで、注意段階への介入、因果的整合性の強化、および視線の誘導や演者間の関係構造の可視化を通じて、臨場感形成過程に作用することを目指す。

2.3 ライブ映像の臨場感向上に着目した研究

特有のデバイスを用いることでオンラインライブの臨場感を向上させる研究は複数行われている。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 [28, 29, 12]。これらの研究では、 α 波と β 波の比率によって様々な視覚効果を映像に映し出すことで観客の楽しさ、一体感を増強させている。柳沢ら [13]は、オンラインライブ配信に同期してLED点灯装置の制御を行う「Immersive Online Live System」を提案している。このシステムはライブ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ている。「空間の充填」と呼ばれる手法を用いることで、

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させ、強い非日常感覚を体験可能としている。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ため、デバイスに依存しない手法で臨場感を向上させることが、より多くのユーザに視聴してもらう上では重要となる。

また、映像・音響処理によって臨場感を向上させる研究も多数存在する。大石ら [30] は、映像の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を提案している。アンケートの結果、提案手法によるズームの前後で臨場感に関する主観評価の参加者間平均値が向上することが明らかになっている。Shin ら [31] は、3D サウンドが視聴者の体験に与える影響を社会的存在感、パラソーシャル・インタラクション、楽しさ、経済的支援意図の4つの観点から調査をしている。その結果、3D サウンドが視聴者の社会的存在感を高めること、さらにそれに付随してその他の項目も評価が向上することを明らかにした。

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処理も加えて音響処理を施す手法を提案し、臨場感への影響を調査する。

第3章 映像効果の付与による影響の調査

3.1 実験概要

ライブ映像への映像効果付与手法が臨場感の増加に繋がり得るかを調査するため、複数のライブ映像を視聴してもらう実験を行う。具体的には、提案手法により映像効果が付与された動画とそうでない動画をランダムで実験参加者に視聴してもらい、視聴時間やいいねの数の比較を行うことで提案手法の有用性を検証する。

3.2 映像効果の検討

軽音楽団体におけるライブ映像配信では一視点からの撮影、スマートフォンでの視聴が多く、視覚的関心の維持 [8] や画角の小ささ [10] の観点から、臨場感を伝えることが難しい。他の研究において、臨場感は視覚的リアリズムのみで成立するものではないこと [21, 22] や、メディアには媒介性を消去する志向と逆にメディア性を前景化する志向という二重の側面があること [20] が示されている。一視点から撮影されたライブ映像では、画角的制限や画質の問題があるため、メディア性を前景化する方向から臨場感を高める可能性を検討する。ここで、メディア性が顕著に現れる表現の一例として、漫画やアニメに着目する。漫画やアニメでは、出来事の印象を視覚的に誇張する表現が用いられている。これらは出来事の印象を増幅するための演出的操作として機能している [32, 33]。本研究ではこうした強調的表現を参考にし、ライブ映像に導入することで臨場感の向上を目指す。

音楽に関する漫画・アニメで使われる表現としては、ズーム、画面分割、画面の一部に視線を集中させる表現（集中表現）(図1)に加えて、抽象的な映像による感情や音の伝播の可視化などがある。この中で、ズーム、画面分割、集中表現は技術的に実現可能であり、汎用的であると考えた。ここで映像の分野において映像のカットやズームが自動的な注意反応を引き起こすことが示されており [23]、断続的なズームは視覚的注意を持続させる効果を持つと考えられる。また、放射構造のような視覚的特徴は視線を中心へ誘導することが報告されており [26]、集中線は注意誘導効果を持つ可能性がある。さらに、



図 1: 音楽に関するアニメでの映像効果

© 小玉ユキ「坂道のアポロン」, © はまじあき「ぼっち・ざ・ろっく!」, © アミュー「この音とまれ!」



図 2: 適用する映像効果

複数の情報を空間的に近接して提示することは理解を促進することが示されており [27], 画面分割は演者間の関係や同時性を可視化する手法として機能し得る. そこで本手法では, ズーム, 集中線, 画面分割の 3 種類の映像効果を採用し (図 2), 一視点固定型ライブ映像 (元映像) へ付与することによる印象を調査する.

ここで, ライブ映像へ映像効果を付与する際に, どのタイミングでどのような映像効果を施すかが重要となる. Slater [22] は, 臨場感の成立には, 提示される出来事が文脈的・因果的に整合していることが重要であると指摘している. 本研究ではこの観点を踏まえ, 演者の動作量やオンセット情報といった指標に基づき適用する設計とした. これにより, 構造的変化が音楽的出来事と整合し, 因果的整合性を保ちながら強調を行うことを目指す. ズームでは, 演者の動作量に着目し, より大きな動きをしている演者にフォーカスを移す. また, オンセットに同期して焦点対象を切り替えることで, 構造的変化を断続的に挿入し, 注意の持続を図る. 画面分割は, バンド全体の動作量が高い場面において適用する. これは複数演者の同時的活動を可視化し, 演奏の同時性や関係性を明示することを目的とする. 集中線は, 音量変化量が大きい場面に適用し, 視覚的焦点を強調する効果を期待する.

3.3 プロトタイプシステムの実装

本システムの処理の主な流れを図 3 に示す。主な処理は以下のとおりである。まず映像から人物検出を行い、演者を特定する。次にオプティカルフローを用いて演者の動作量を算出し、その情報に基づいて、演者の動きによるパフォーマンスを重視したシーン検出を行う。一方、音声データからはオンセットを検出することで、演奏に沿ったタイミングを取得する。最後に、これらから得られたタイミングを特徴的なシーンとして抽出し、当該シーンに特定の映像効果を付与する。以下では、それぞれの処理について順に説明する。

映像からは、まず物体検出ライブラリ YOLOv7 を用いて人物検出をする。試験的に、ライブ映像は演奏ステージの真正面のやや上から撮られたものの中から演者 3 人で構成されているバンド（3 ピースバンド）に限定し、それぞれの演者に特定の ID を割り当てた。この際、演奏者は画面上部に、観客は画面下部に映される。そのため、上から順に 3 人を演者として認識した。ここで、検出結果が 3 人未満になっていたり、別の演者に ID が移ってしまったりと思い通りに検出されないケースが複数確認された。そのため、演者 3 人が正常に検出されたフレームをストックしておき、上手く検出されなかったフレームではストックしていたフレーム情報を用いて補完することで、ほぼ全てのフレームで演者 3 人を認識することを可能とした。具体的には、ストックしていたフレームの検出結果と、現在のフレームの検出結果の位置関係を比較することで、ID を引き継がせる。検出されなかった演者に関してはストックしていたフレームでの位置とすることで、演者全員が上手く検出されていない場合でも、ある程度の演者の位置を推定することを可能とした。

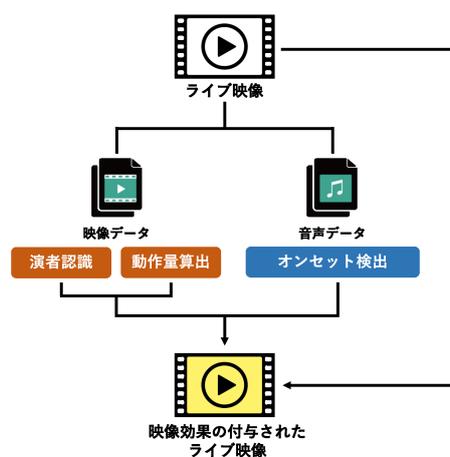


図 3: プロトタイプシステムの処理の流れ

次に、演者の動作量を算出する。毎フレームでの算出は微小揺れや検出誤差の影響を受けやすく、また計算量も増大するため、5フレーム間隔で特徴点を抽出し、前回の特徴点とのオプティカルフローを算出した。その移動量を動作量とし、演者ごとに保存する。ここで求めた動作量は、ズームと画面分割を適用する際に使用する。

音声からは、音声処理ライブラリである LibROSA を用いてオンセット検出を行った。ここで、ライブ映像によって音の強弱が激しい場合や緩やかな場合があるため、映像に合わせて発音されたと判断する閾値を設定した。求めたオンセットは、ズームと集中線を映像に適用する際に使用する。なお、オンセット時の音量差も保存しておく。

そして、映像データ、音声データから求めた演者の動作量、オンセットから、適用する映像効果を選定する。ズームは演奏シーンを通して適用され、演奏のテンポに合わせてより動きの大きい演者にフォーカスする。そのためオンセットにおいて、一番動きのある演者にズームすることとした。また、画面分割は演者全員にフォーカスが当たるため、各演者に動きがあるシーンで付与する必要がある。そのため、検出された演者全員の動作量が一定の値を超えた部分で適用することとした。一方、集中線は迫力や音圧を伝えるため、オンセットにおいてより音量差が大きいときに適用することとした。

それぞれの映像効果付与の条件を以下に示す。

- ズーム：オンセット検出(上位 30%)のタイミングで、動作量増加割合が一番大きい演者に適用
- 画面分割：全演者の動作量が、人物幅との相対距離で 10%以上増加した際に適用
- 集中線：オンセット検出(上位 10%)のタイミングで適用

なお、画面分割と集中線は出来事を強調するための効果であり、装飾的な視覚要素が過剰に付与されると認知負荷が増加する可能性が指摘されているため [27]、映像中での適用回数が過度にならないよう制限した。

ここで、ライブ映像に映像効果を付与するうえで、複数の映像効果付与の条件が同時に成立してしまう場合があり得る。映像を通して適用されるズームは、装飾的な効果として使用される画面分割や集中線よりも適用される頻度は高くなるため、適用することへの優先度は画面分割や集中線の方が高い。そのため、映像効果付与の優先順位はズーム、集中線、画面分割の順に高くなるように設定した。

3.4 実験設計

本実験の目的は、提案手法によるライブ映像への映像効果の付与が、視聴体験に与える影響を検証し、臨場感向上に寄与し得るかを明らかにすることである。Wirthらのモデル [19] では臨場感形成は段階的な認知プロセスとして説明され、その第一段階としてメディアへの注意配分が前提となっている。その上で、提示された空間の理解が進み、自己の基準座標がメディア空間へ移行することで臨場感が成立するとされている。したがって臨場感の形成には、注意がメディアに向けられている状態が継続され、空間理解が進むことが重要である。そこで本実験では、臨場感形成の前提段階にあたるメディアへの注意に着目し、視聴時間を注意の持続の指標として、またいいねの数をエンゲージメントの指標として用いる。以上を踏まえ、以下の仮説を立てた。

仮説1 映像効果を自動付与したライブ映像は、付与しない場合と比べて視聴者の注意がより持続される。

仮説2 映像効果を自動付与したライブ映像は、付与しない場合と比べて視聴者のエンゲージメントが高い。

実験には、ある軽音楽サークルでのライブ映像から2分程度の動画を46本用意した。このうち、3ピース編成のバンドに該当する動画6本（動画1～6）を分析の対象とした。残りの動画はダミー刺激として用い、分析対象動画の間に挿入することで、視聴者が分析対象の動画のみへ意識することを防ぐようにした。

映像効果を付与するうえで、採用した映像効果の中でもズームとその他では主な役割が異なる。ズームは常に映像に変化をもたせるのに対して、画面分割や集中線はアクセントとして部分的に必要となる。これらのことを踏まえ、映像効果の適用条件は以下の通りとした。

- None：映像効果無し
- Zoom：ズーム
- Mix：ズーム+画面分割+集中線

動画1～6からそれぞれ2本ずつ各条件に割り当て、適用条件と動画の組み合わせは実験参加者ごとにランダムになるようにした。また提案手法によって映像効果を付与する対象動画は一定の間隔で表れるようにし、動画の視聴時間を条件ごとに比較する。さら

に、視聴時にはいいねボタンを押せるようにすることで、その場での主観評価を得られるようにした。

3.5 実験手順

実験にはスマートフォンを使用し、比較的静かな部屋でイヤホンをして実施してもらった。実験システム(図4)は、画面を上下にスワイプすることで前後の動画に進むような仕様になっている。いいねボタンは動画の右下に配置し、他の動画と比べて良いと感じた場合に押しってもらうように指示した。また視聴する際、一つ一つの動画を最後まで見てもらう必要はなく、満足したり気に入らなかったりした場合は、次の動画に進んでもいいように指示をした。

実験ではまずチュートリアルとして、分析対象を含まない、20本の動画を視聴してもらった。これは、ライブ映像を新鮮に感じ、序盤の映像を眺めに視聴することが考えられたため、チュートリアルの段階でライブ映像を視聴すること、またいいねボタンを押すことに慣れてもらう狙いがある。その後、本実験として分析対象を含む残りの26本の動画を視聴してもらった。最後の動画をスワイプした後は、実験後アンケート(表1)に答えてもらった。ここでは、映像効果無しの動画(None条件)と有りの動画(Mix条件)を視聴してもらいどの程度迫力を感じたかを答えてもらった。臨場感という概念は多次元

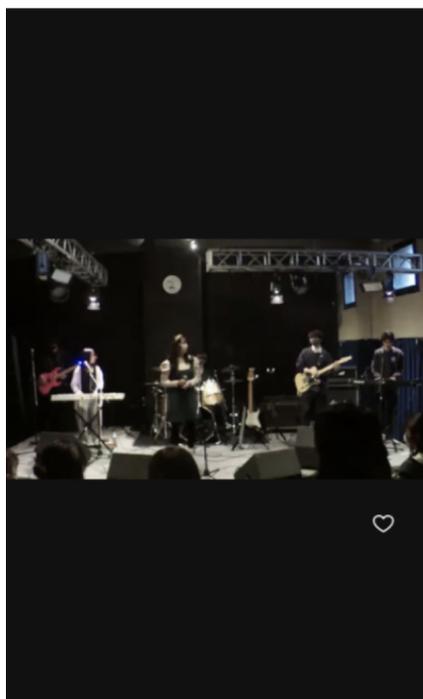


図4: 視聴実験システム

表 1: 実験後アンケート項目

番号	質問	回答形式
Q1	音楽やライブ映像を視聴することは好きですか	5段階 (1-5)
Q2	このライブ映像 (映像効果無し) を視聴して、どの程度迫力を感じましたか	5段階 (1-5)
Q3	このライブ映像 (映像効果有り) を視聴して、どの程度迫力を感じましたか	5段階 (1-5)
Q4	これらのライブ映像を視聴して感じたことがありましたら教えてください	自由記述

的であり、解釈の幅が広い。そのため単一質問で評価した場合、回答のばらつきが大きくなる可能性がある。一方、迫力は出来事の刺激強度や印象の強さに焦点を当てた比較的具体的な評価概念であり、参加者にとって判断しやすい指標であると考えられる。また、刺激強度の知覚は注意喚起や出来事の顕著性に関わることから、臨場感形成過程にも関連すると考えられる [19, 22]。そのため本研究では、臨場感の一側面であり、かつ形成過程に関与し得る指標として迫力を測定した。

3.6 結果

実験は男性 17 名、女性 7 名の合計 24 人に行ってもらった。以下では、実験の結果について述べる。

3.6.1 映像効果の適用条件ごとの視聴時間

図 5 は、動画 1~6 における映像効果の適用条件ごとの視聴時間を箱ひげ図にして比較したものである。横軸が映像効果の適用条件、縦軸が視聴時間である。図中のオレンジの線は中央値であり、緑色の三角形は平均値である。条件ごとの平均視聴時間は、None 条

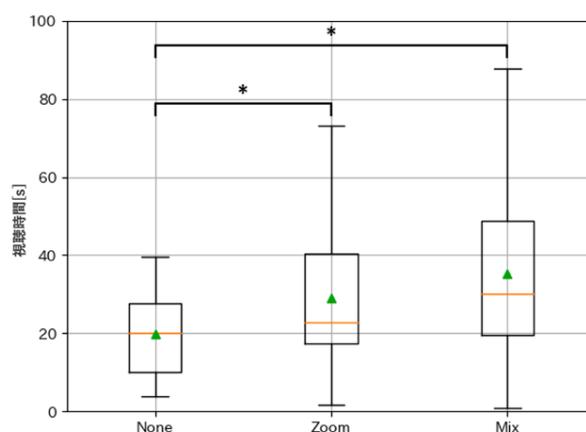


図 5: 映像効果ごとの視聴時間

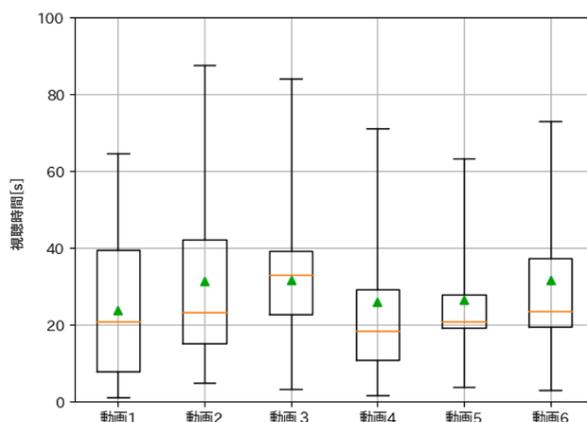


図 6: 動画ごとの視聴時間

件が 19.8 秒、Zoom 条件が 29.1 秒、Mix 条件が 35.6 秒となり、映像効果が増えるごとに増加している。また標準偏差は None 条件が 10.4、Zoom 条件が 19.1、Mix 条件が 22.8 となり、映像効果を付与することでより大きく散らばるようになっている。分散分析を行ったところ 3 条件の平均の差に有意差が認められた ($p < 0.05$)。続いて多重比較を行った結果、None 条件と Zoom 条件間、None 条件と Mix 条件間で平均の差に有意差が認められた ($p < 0.05$)。そのため、「映像効果を自動付与したライブ映像は、付与しない場合と比べて視聴者の注意がより持続される」という仮説は支持された。

また図 6 は、動画の種類ごとの視聴時間を比較したものである。分散分析を行った結果、平均値の差に有意差は見られなかった。しかし比べる動画によっては 7.9 秒ほどの差が見られることから、多少の差はあることがわかる。

3.6.2 映像効果の適用条件ごとのいいね数

表 2 は、動画 1～6 における映像効果条件ごとのいいね数を比較したものである。映像効果条件ごとの合計は、None 条件が 10 個、Zoom 条件が 19 個、Mix 条件が 25 個と映像効果が増えるごとにいいね数も増えており、「映像効果を自動付与したライブ映像は、付与しない場合と比べて視聴者のエンゲージメントが高い」という仮説も支持された。動画の種類別に比べると、動画 4 が 14 個と 1 番多く、順に動画 6 が 10 個、動画 1、動画 2 が 9 個となっている。また全体だと合計 54 個のいいねが付与されていた。

表 3 はいいねの有無ごとの視聴時間平均視聴時間を比較したものである。全てのライブ映像を平均するといいねされている場合は 40.2 秒、されていない場合は 22.4 秒となっており、いいねされている映像の方がより視聴時間が長くなっていることがわかる。動画

表 2: 条件ごとのいいね数

	None	Zoom	Mix
動画 1	1	3	5
動画 2	3	3	3
動画 3	0	2	3
動画 4	2	5	7
動画 5	1	2	4
動画 6	3	4	3
合計	10	19	25

表 3: いいねの有無ごとの平均視聴時間

	いいねされている	いいねされていない
動画 1	49.4	17.9
動画 2	47.5	24.0
動画 3	20.2	31.2
動画 4	30.5	13.8
動画 5	46.8	20.7
動画 6	46.8	26.6
平均	40.2	22.4

3ではいいねされていない場合の方が11.0秒長くなっているが、その他の映像ではいいねされている場合の方がより視聴されており、動画1では31.5秒長くなっている。また、いいねの有無による視聴時間の平均に有意差が認められた ($p < 0.05$)。

3.6.3 アンケート

Q2, Q3の結果に関しては、映像効果有りの方が高い数値を得られ迫力を感じられたという結果になったが、動画を見てもらう順番を映像効果無しから有りの固定にしていたため、正しく測れていない可能性が考えられる。またQ4の自由記述欄では、良い印象のものとして「1人1人を大きく映すことでより迫力が出たように感じた」、「カメラのアンクルや効果線などバンドメンバーによりフォーカスしている映像が見ていて迫力を感じた」、「全体を映しているだけだとただ漠然と見てしまうが、アップしていたりすると細かいところに注目しやすい」などの意見が得られた。一方、悪い印象のものとして、「過剰な編集は飛ばしたくなった」、「集中線はライブ映像に使うと安っぽくなるので無い方がいいと思った」、「演者1人1人にスポットを当てる編集は使いすぎて見にくいと感じた」などの意見が見られた。これらのように映像効果の付与に対しての印象の良し悪しは個人差があることがわかる。

3.7 考察

映像効果の適用条件ごとの平均視聴時間において分散分析、多重比較を行った結果、None 条件と Zoom 条件間、None 条件と Mix 条件間で有意差が見られた。このことから、提案手法によるライブ映像への映像効果の付与は視聴時間の向上に有効であることがわかる。しかし、Zoom 条件と Mix 条件間で有意差が見られず、付与する映像効果を組み合わせることによる視聴時間への効果は大きくなかった。これは、Zoom 条件においても映像として満足するものになっており、Zoom 条件と Mix 条件間で満足度に大きな差が無かったことが原因として 1 つ考えられる。今回、適用の優先度の都合で、ズームが適用されているタイミングでも画面分割や集中線を適用可能にしていた。そのため、映像効果の適用頻度としては Zoom 条件と Mix 条件で大きな差は無かった。ある程度映像効果を適用すると満足度は変わらなくなることも考えられるため、映像効果の適用頻度による影響も今後調査していく予定である。

映像効果の適用条件ごとにいいね数を比較した結果、ライブ映像に適用する映像効果が増えるごとにいいね数も増加していた。また、いいねボタンが押されている場合の方が押されていない場合と比べ全体で平均視聴時間が 17.8 秒長くなっており、両者の間には平均視聴時間に有意差が認められた。これらのことから、提案手法により映像効果が付与されたライブ映像は視聴者の注意が持続するだけでなく、エンゲージメント向上にも繋がっていることが考えられる。

アンケート設問の Q4 では、ズームや画面分割により演者 1 人 1 人を大きく映すことによって注目しやすく迫力を感じたといった意見や、効果線などによって演者にフォーカスしている映像が見ていて迫力を感じたという意見がみられた。しかし、良い印象の意見が得られた反面、提案手法への印象があまり好ましくない意見も多くみられた。これらの多くは、映像効果の付与が過剰であると感じたことや、映像効果が馴染んでいないということが原因であった。今回、適切な映像効果の付与のための動作の変化量やオンセットを検証できていなかったため、過剰に映像効果を付与してしまっていることが考えられる。映像効果の適用条件ごとの平均視聴時間においても、映像効果を付与した Zoom 条件、Mix 条件では標準偏差が大きくなっており、視聴時間に散らばりがあることがわかる。これらの結果も考慮すると、提案手法によって映像効果を付与したライブ映像は人による印象の差が大きく、参加者の傾向から、普段からライブ映像を多く視聴する人からの印象は悪く、逆にそうでない人からの印象は良いことがわかった。つまり、ライブ慣れしていない人に向け、新規の客を開拓するという意味では効果があると考えら

れる。また、今回使用した動画の中には、曲調が激しいものや緩やかなものなど多様な楽曲があった。採用した映像効果は、演者の動きによるパフォーマンスや演奏の勢いを伝えるようなものであり、曲調によっては適用するのが難しい場合もあり得る。そのため、適用する映像効果の種類を楽曲の曲調を考慮しつつ、再度検討する必要がある。

一方、これらの結果は映像効果に慣れていないことによる視聴時間や評価の上振れの影響を含んでいる可能性が考えられる。チュートリアルでは、実験システムやライブ映像の視聴への慣れにのみフォーカスを当てており、映像効果そのものへの慣れは考慮できていなかった。今後の実験では、公平に比較するために、実験設計を見直す必要がある。

第4章 映像・音響効果の付与に対する主観評価 実験

4.1 実験概要

前章では、オンラインライブの臨場感向上に関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効果を付与するプロトタイプシステムを作成した。また、映像効果を付与することにより視聴時間、エンゲージメントが向上することを明らかにした。つまり、臨場感形成に関する Wirth らのモデル [19] における前提条件であるメディアへの注意が、提案手法によって向上したということである。

ここで、より詳細な主観評価実験を行うことで、メディアへの注意から臨場感の形成に進んでいるかを確認する必要がある。臨場感は広い意味合いを持っており、構成要素として没入感や社会的存在感、インタラクション性、リアリティなど多くが含まれている。Lombard [5] は、臨場感の概念を包括的に整理し、気配、リアリティ、場所、没入感、メディア内の人をどう捉えるか、メディア自体をどう捉えるかの6つの見方があるとまとめている。また Lee [14] 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、演者への興味度も視聴者の行動に関わってくる。演者の存在を感じることもこうした配信での臨場感において重要なことが明らかになっており [15]、軽音楽団体でそれぞれの演者に興味を持ち、さらにその団体全体に興味が行くことで、入会やその団体の活性化に繋がると考えられる。感情的関与度が高まることで没入感が向上するという研究もあり [17, 34]、視聴体験への影響も示唆される。これらの指標の中で、軽音楽団体のオンラインライブにおける音楽体験では、没入感、社会的存在感、演者への興味度が特に直接的に関わってくると考えられる。本研究ではこれら3つの指標をまとめた概念を臨場感と定義し、評価を行う。

また映像効果だけでなく、音響的な表現を用いることも、臨場感を向上させる方法として考えられる。視覚と聴覚の情報が一致していることが認知、感性的に良い影響を与えるという研究は複数見られ [35, 36]、映像効果と同じ対象を音響効果でフォーカスを当てることでよりそれぞれの演奏に興味を持ってもらえることが期待される。軽音楽団体に

において、ライブ映像配信を通じて入会や団体の活性化に繋げるには、そのライブ自体や演者に興味を持ってもらい、臨場感のある音楽体験をしてもらう必要がある。そこでより視聴者の没入感を高め演者への関心を引きつけることに着目し、映像効果に合わせた音響効果を付与することにより音楽体験の向上を目指す。

4.2 システムの改良と音響効果の検討

4.2.1 映像・音響効果の付与手法

本手法の流れを図7に示す。ライブ映像を映像データ、音声データに分けて処理を行い、映像効果と音響効果の付与をしている。以下では、映像効果、音響効果の付与手法について説明する。

前章において、ライブ音源のオンセットと演者の動作量を用いて特徴的なシーンを検出し、適切な映像効果を付与する手法を提案した。その際のアンケートでは、集中線の表現が映像に合っておらず過剰であるという意見が得られた。そのため、同じように画面の特定の箇所に視線を集中させる効果を持つラディアルブラーを採用した。さらに画面分割の効果を増強させるために、動作量の増加割合が大きい順に時間差表示するように変更した。これらの変更のうえ、一視点固定型ライブ映像（元映像）へのズーム、ラディアルブラー、画面分割の3種類の映像効果（図8）を付与することによる印象を再度調査する。また、映像効果を付与するタイミング、対象に関しても、それぞれの表現で手法が混在していたため改善した。具体的には、音源からオンセットを検出しそこで映像をブロック分けし、ブロック間でより増加割合の大きい演者にフォーカスを当てるように映

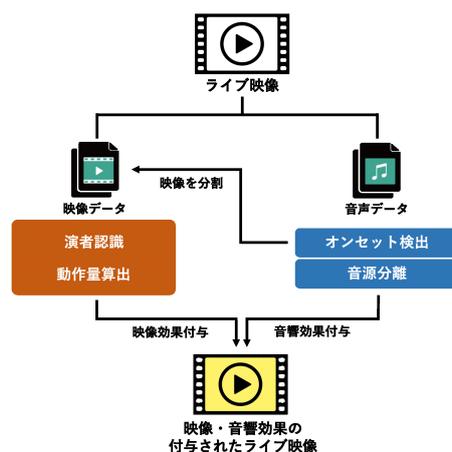


図 7: 映像・音響効果付与システムの処理の流れ



図 8: 映像効果付与による画面パターン

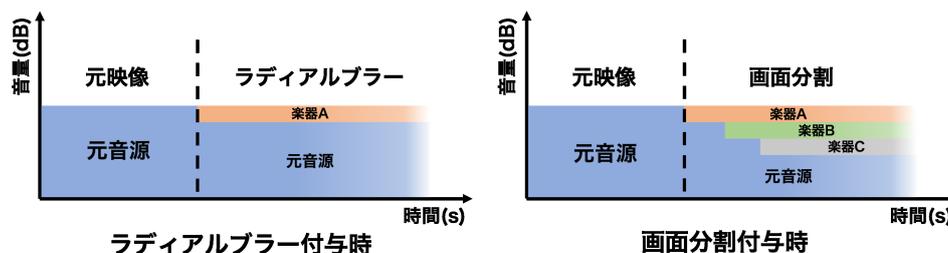


図 9: 音響効果イメージ

像効果を付与する。

音響において特定の楽器（周波数）を際立たせる手法としては、音源分離やイコライジング処理がある。今回は、より要所的に適用するために、ライブ音源を楽器ごとに分離し、元のライブ音源に重ねることで、特定の楽器を際立たせる。ただ音源分離したものを重ねるだけでは、ユーザの意図しない音量変化が起こり印象に差が出てしまうことが考えられる。そのため、音響効果の前後での音量を等しくし、際立たせることのみによる効果を調査する。また、頻繁に音響が変わることはユーザに違和感を与えてしまう可能性があるため、音響効果を付与するタイミングは映像効果を付与するタイミングより少ない方が望ましい。そのため、映像効果において選定していた、比較的付与頻度の少ないラディアルブラー、画面分割を付与するタイミングと同一の部分に限定する。ラディアルブラーの付与時は、フォーカスを当てている対象の楽器の音源を重ねることで、さらなる集中効果を狙う。また画面分割の付与時では、時間差表示によってそれぞれの演者がフォーカスを当てられるため、その対象の楽器の音源が追加で重なるようにした(図9)。この、映像効果の付与と音響効果の付与を組み合わせることで、一視点固定型ライブ映像の没入感、社会的存在感、演者への興味度を高める。

4.2.2 演者の動作量算出の実装

まず librosa を用いて音源からオンセットを検出し、それをもとに映像を分割しブロック分けする。オンセットを用いることで、サビの開始地点など、音源的にタイミングの良い箇所への表現の付与を可能となる。オンセット検出に関する設定については初期設定のままとした。ここで、ライブ演奏では音量に波があり、部分的に適したオンセット検出の設定が変わってくる。そのため、8 次多項式近似により平滑化し、音量の全体的な波を抑えてからオンセットを検出した。映像からは、前章同様物体検出ライブラリ (YOLOv10) を用いて人物を検出する。続いて、OpenCV を用いて演者付近の特徴点を検出し、その移動量をもとにブロックごとの動作量を求める。

4.2.3 映像効果の実装

オンセットによって分けた箇所の前後のブロック同士で演者の動作量を比較することにより映像効果を付与するタイミング、対象を決定する。ブロックごとに各演者の単位時間あたりの動作量を求め、前回のブロックとの増減割合を比較し、以下の基準で映像効果を適用する。

- ズーム：演者の動作量が 1.2 倍以上増加した際に適用
- ラディアルブラー：演者の動作量が 2.0 倍以上増加した際に適用
- 画面分割：全演者の動作量が 2.0 倍以上増加した際に適用

このようにすることで、ズームである程度の画面遷移を担保しつつ、さらにラディアルブラーで動的なシーンの強調、画面分割で演者間の相互作用を強調している。

次に、それぞれの映像効果での画像処理について説明する。ズームでは OpenCV を用い、対象の演者を中心にフレーム画像を切り取り拡大している。切り取る際は、画面の縦横比を保ちつつ、対象の演者の検出結果の横幅に対して切り取る枠の横幅が 4 倍になるようにしている。ラディアルブラーでは、ズームでの処理に加えて、画像のサイズを縮小し配列に保存する処理を n 回繰り返す、それらの画像を平均化することで実装している。また、マスク処理を用いて外側から内側にかけて透明度が 0 に近づくようにし、その画像を元のフレーム画像に重ねることで、中心が鮮明に映るようにしている。画面分割では、各演者を中心にそれぞれフレーム画像を切り取り拡大して、左からベース、ドラム、ギターの順に配置している。各演者の表示スペースは画面の縦幅はそのままで横

幅を 1/3 とし、表示スペースの縦横比かつ縦幅が各演者の検出結果の縦幅と同じになるようにし切り取る。また、それぞれの演者により視線を向けさせるため、動作量が多い演者ごとに 1 秒ずつ時間差表示をしている。

4.2.4 音響効果の実装

ライブ音源を音源分離し、元の音源に重ねて強調をするのだが、音源分離を行ったところ、ギターとベースがお互いに誤判定される場合が見られた。そこで、まずはボーカル、ドラム、弦楽器に音源分離をし、その後弦楽器の音源を高い周波数帯と低い周波数帯でイコライジング処理をすることで、ギターとベースに分離させた。具体的には、ギターの場合は弦楽器音源の高い周波数のレベルを上げ低い周波数のレベルを下げ、ベースの場合はその逆の処理をすることで、綺麗に分離した。本手法においては、250Hz を境として、なだらかにイコライジング処理を施している。音源分離には、Demucs v4 に含まれる軽量モデル mdx extra q を用いた。なお、音源分離に関する設定については初期設定のままとした。

音源分離をした音源を元に、音響効果を実装した。主な操作は pydub を用いて行った。ラディアルブラーではその対象の楽器の音源、画面分割では表示する順にその対象の楽器の音源の音量を 1.5 倍にし、元の音源に重ねることで強調している。この時、音響効果の前後で音量が違っていると、印象に差が生じることが考えられる。そのため、音響効果を付与した部分の平均 RMS が、付与する前の平均 RMS と等しくなるように調整をした。画面分割においても、映像効果の時間差表示に合わせて対象の楽器の音源を重ねていき、各部分の平均 RMS が等しくなるように調整をしている。

音源分離を行った音源を用いて音響効果を実装した。主な操作は pydub を用いて行った。ラディアルブラーでは対象楽器の音源を、画面分割では表示順に対応する楽器の音源を 1.5 倍に増幅し、元音源に重ねることで強調した。一方、単純に音量を増加させると全体ラウドネスが上昇し、音量差そのものによる印象変化が生じる可能性がある。そこで本研究では、音響効果付与前後で平均 RMS が等しくなるよう調整を行い、全体ラウドネスを一定に保った。本処理の目的は音量を増大させることではなく、映像効果によって注目された楽器を聴覚的にも前景化し、視聴覚的整合性を高めることである。

4.3 実験設計

本実験の目的は、主観評価アンケートを実施し、ライブ映像への映像・音響効果の付与が臨場感を向上させるかを明らかにすることである。前章では、提案手法による映像効果の付与が、視聴者の注意の維持やエンゲージメントを向上させることを明らかにした。そのため、Wirth らのモデル [19] における前提条件はクリアしており、空間理解が進み臨場感が成立することが期待される。また、視覚と聴覚の情報的一致による認知、感性への影響 [35, 36] も臨場感に繋がると考えられる、そこで本実験では、以下の仮説を立てた。

仮説 1 映像効果を自動付与したライブ映像は、付与しない場合と比べて臨場感が向上する。

仮説 2 映像・音響効果を自動付与したライブ映像は、映像効果のみを付与したものとは比べて臨場感が向上する。

実験用の動画としては、ある軽音楽サークルでのライブ映像の 1 バンドから、30 秒程度の動画を 12 本作成した。楽曲のジャンルは邦楽のロックであり、比較的アップテンポなもの 2 曲から作成したため、それぞれの動画の印象は同程度であると考えられる。また、その楽曲を知っているか否かによって評価が変わってくるのが考えられる。そのため、一般的にあまり知られていない楽曲を演奏しているものを対象とした。動画に付与する映像・音響効果の適用条件は、以下の通りとする。

- None: 映像・音響効果無し
- Visual: 映像効果のみ付与
- Audio: 音響効果のみ付与
- Both: 映像・音響効果付与

各参加者ごとに、12 本の全ての映像をランダムな順番で視聴してもらおう。この時、表現の適用条件は映像 3 本ずつ割り振った。

本研究では、軽音楽団体のライブ映像配信において重要な要素として、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。ここで、それぞれの指標を測る手法に関して述べる。Sandstrom ら [18] は、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) を提案している。視覚的イメージ、意識の変容、音楽への注意の集中の 3 つの因子があり、それぞれに関する複数の質問の回

答の平均値によって評価することができる。また Swarbrick ら [17] は、コンサートでの没入感が観客体験とどう関係しているかを調査する際に、AIMS の質問項目をピックアップし音楽吸収を測定している。社会的存在感に関しては、Lombard ら [37] がメディア環境における存在感を測る指標として TPI(Temple Presence Inventory) を提案している。さらに、特定の対象に対する個人のファン度合いを評価するために、Fanship Scale が開発されている [38]。これらの評価指標を参考にし、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 4)。

また一視点から撮影されたライブ映像を視聴する際の満足度に、参加者それぞれにばらつきがあることが考えられる。そのため、実験前タスクとして別に用意した 3 本のライブ映像を視聴してもらい、満足度を測るアンケートをそれぞれ実施した。項目は、Patwardhan ら [39] がメディア満足度を評価するために開発した ABMSS(Affect-Based Media Satisfaction Scale) を参考にし選定している (表 5)。この結果をもとに、参加者それぞれの特性を踏まえた分析を行う。

表 4: 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 5: ライブ映像の満足度に関するアンケート

番号	質問
Q1	このライブ映像を視聴して、楽しさを感じましたか。
Q2	このライブ映像に興味を持ちましたか。
Q3	このライブ映像に引き込まれる感覚がありましたか。
Q4	このライブ映像に全体的に満足しましたか。

4.4 実験手順

実験にはスマートフォンを使用し、参加者には比較的静かな部屋で、イヤホンを装着して行なってもらった。図 10 のように動画、アンケートを表示し、それぞれの動画进行评估してもらった。動画再生画面では、動画の下に配置したボタンによって再生・停止を操作できるようにし、視覚的な影響を抑えている。アンケートは 6 段階のリッカート尺度で回答してもらった。実験ではまず、実験前タスクとして 3 本のライブ映像に対して、視聴とアンケート回答をして参加者の特性を測った。その後、表現の付与対象である 12 本の動画に対しても、視聴とアンケート回答をしてもらい、それぞれの動画进行评估してもらった。最後には全体アンケートとして、映像・音響効果、またその他の意見を答えてもらった。

4.5 結果

実験参加者は、普段から音楽をよく聴く人を対象にし、大学生・大学院生の合計 24 名 (男性 15 名, 女性 9 名) であった。また、そのうち楽器経験者は 14 名, 楽器未経験者は 10 名であった。以下では、実験の結果について述べる。なお、グラフの縦軸はそれぞれのアンケート項目の評価値, 横軸は表現の適用条件であり, 図中のオレンジ色の線は中央値, 緑色の三角形は平均値を表している。



図 10: 主観評価実験画面

4.5.1 適用条件ごとの評価値

各評価指標について Shapiro-Wilk 検定を用いて正規性を検証した結果、没入感において Both 条件 ($p = 0.004$)、社会的存在感において Visual 条件 ($p = 0.027$) および Both 条件 ($p = 0.003$)、演者への興味度において Visual 条件 ($p = 0.034$) および Both 条件 ($p = 0.014$) で正規性が棄却された。しかし、すべての項目で Levene 検定による等分散性は保持されていた ($p > 0.05$)。これらを踏まえ、各指標に対して分析を行う。

図 11 は、各試行における没入感の評価値を適用条件ごとに箱ひげ図にして比較したものである。各条件ごとの平均値は None 条件が 3.71、Visual 条件が 3.98、Audio 条件が 3.57、Both 条件が 3.90 であり、映像効果を付与した際に評価が高くなっている。また標準偏差は None 条件が 0.99、Visual 条件が 0.96、Audio 条件が 1.05、Both 条件が 1.00 であった。一元配置分散分析 (ANOVA) を行ったところ、4 条件の平均値の差に有意差は認められなかった ($F = 2.37, p = 0.071, \eta^2 = 0.03$)。

また図 12 は、各試行における社会的存在感の評価値を比較したグラフである。各条件ごとの平均値は None 条件が 3.76、Visual 条件が 4.22、Audio 条件が 3.66、Both 条件が 4.23 であった。また標準偏差は None 条件が 0.83、Visual 条件が 0.76、Audio 条件が 0.82、Both 条件が 0.82 であった。分散分析を行ったところ、4 条件の平均値の差に有意差は認められた ($F = 9.63, p < 0.001, \eta^2 = 0.10$)。続いて Tukey の HSD 多重比較検定を行った結果、None-Visual ($p = 0.005, d = 0.58$)、None-Both ($p = 0.004, d = 0.58$)、Visual-Audio ($p < 0.001, d = 0.71$)、Audio-Both ($p < 0.001, d = 0.70$) 間で有意差が認められた。特に Visual-Audio、Audio-Both 間は $p < 0.0005$ であり、映像効果は没入感を向上させているが、音響効果の効果は見られないことがわかる。

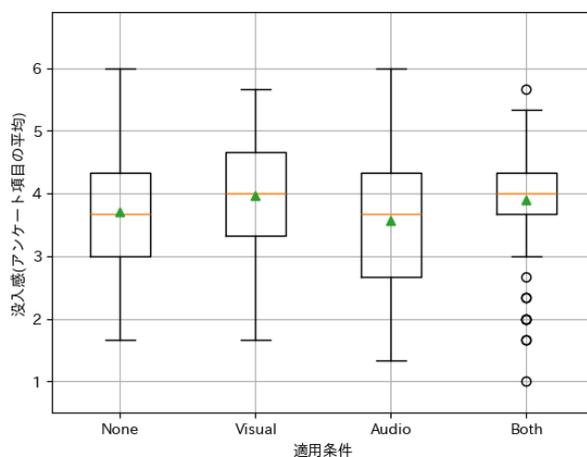


図 11: 適用条件ごとの没入感の評価値

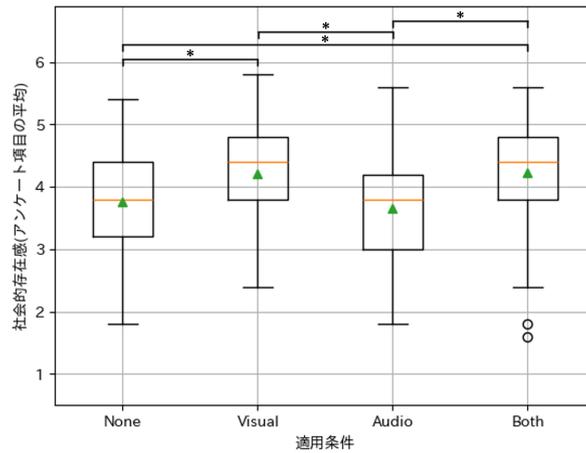


図 12: 適用条件ごとの社会的存在感の評価値

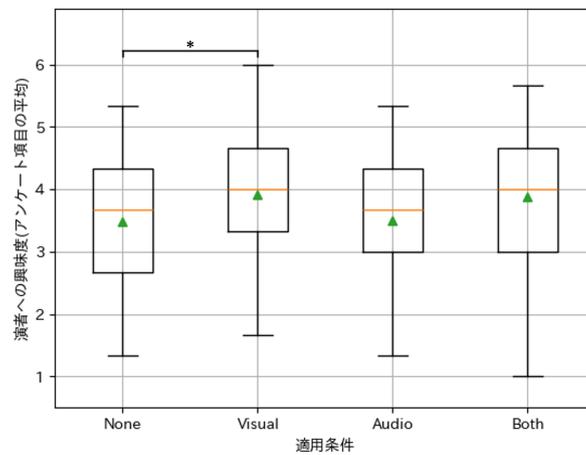


図 13: 適用条件ごとの演者への興味度の評価値

図 13 は、各試行における演者への興味度の評価値を比較したグラフである。各条件ごとの平均値は None 条件が 3.49、Visual 条件が 3.92、Audio 条件が 3.51、Both 条件が 3.89 であった。また標準偏差は None 条件が 0.93、Visual 条件が 0.93、Audio 条件が 0.98、Both 条件が 0.94 であった。分散分析を行ったところ、4 条件の平均値の差に有意差は認められた ($F = 4.26, p < 0.01, \eta^2 = 0.04$)。続いて多重比較を行った結果、None–Visual 間で有意差が認められた ($p = 0.040, d = 0.46$)。

図 14 は、各試行における臨場感（没入感・社会的存在感・演者への興味度を平均した合成指標）の評価値を比較したグラフである。各条件ごとの平均値は None 条件が 3.65、Visual 条件が 4.04、Audio 条件が 3.58、Both 条件が 4.01 であった。また標準偏差は None 条件が 0.82、Visual 条件が 0.80、Audio 条件が 0.85、Both 条件が 0.85 であった。分散分析を行ったところ、4 条件の平均値の差に有意差は認められた ($F = 5.72, p < 0.001, \eta^2 = 0.06$)。

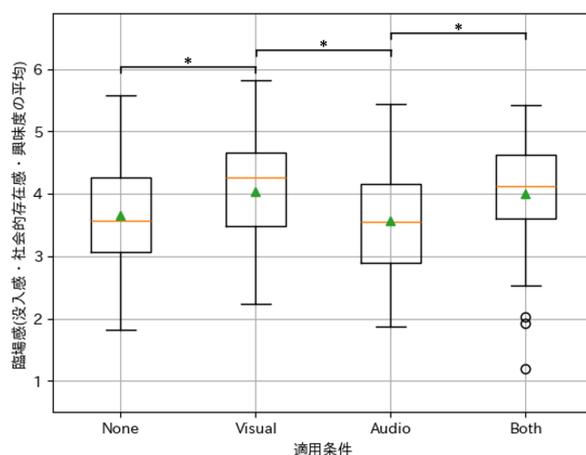


図 14: 適用条件ごとの臨場感の評価値

続いて Tukey の HSD 多重比較検定を行った結果, Audio-Visual ($p = 0.007, d = 0.56$), Audio-Both ($p = 0.013, d = 0.51$), None-Visual ($p = 0.032, d = 0.48$) 間で有意差が認められた. このことから, 「映像効果を自動付与したライブ映像は, 付与しない場合と比べて臨場感が向上する」という仮説は支持されたが, 「映像・音響効果を自動付与したライブ映像は, 映像効果のみを付与したものと比べて臨場感が向上する」という仮説は支持されない結果となった.

また, アンケートでの没入感, 社会的存在感, 演者への興味度に関する項目における Cronbach's α は, それぞれ 0.76, 0.75, 0.72 であり, 十分な信頼性を持っているとみなせる.

4.5.2 アンケート項目ごとの評価値

Kruskal-Wallis 検定を行った結果, Q2, Q4-Q8, Q10 で有意差が確認された ($p < 0.05$). 特に Q4-Q8 は $p < 0.01$ であり, 社会的存在感に関する項目で多くの差が見られた. 続いて Steel-Dwass 検定を行うと, 音響効果に関係なく, 映像効果の有無によって有意差が見られた ($p < 0.05$).

また Q6 の「演奏を感情的 (躍動的) に感じましたか」という質問では, 有意差はないが, Visual 条件よりも Both 条件が, Audio 条件よりも None 条件の方が評価が高くなっている. さらに, Q7 の「演奏に現実味がありましたか」, Q10 の「もし演者が演奏を成功・失敗すると心が揺らぎますか」という質問においては, Visual 条件よりも Both 条件の方が値が高い.

4.5.3 一視点固定型ライブ映像への満足度による評価

実験前タスクで視聴してもらった3本のライブ映像の評価値から算出した満足度をもとに分析を行う。表6, 7は、参加者全体の中央値よりも高い人, 低い人で分けて, 評価値をまとめたものである。表6を見ると, 一視点から撮影されたライブ映像への満足度が高い参加者は没入感と社会的存在感において, None条件からVisual条件の増加は小さいが, Visual条件からBoth条件でも値が増えていることがわかる。一方表7では, 一視点から撮影されたライブ映像への満足度が低い参加者は, None条件からVisual条件の増加は大きく, Visual条件からBoth条件では値が下がっている。これらのことから, 元々のライブ映像に満足している場合は, 映像効果と音響効果を組み合わせることでより没入感と社会的存在感を向上させることのできる可能性があると考えられる。

4.5.4 実験後アンケート

映像効果に対しては, 「演者一人一人の表情が見えるような編集は演奏に入り込めて良かった」「ある演者にフォーカスが行くと, その人のパフォーマンスに注目するようになったと思った」のように演奏に注目するようになったという意見が得られた。特に画面分割に対しては好印象の意見が多かった。一方で, 映像効果に違和感を感じたり, タイミングがずれている場合に不自然に感じた人が複数見られた。

また音響効果に関しては, 7人の参加者から「集中できなかった」「違和感があった」などマイナスな意見が見られた一方で, 「映像効果と合わさって, 集中対象がわかる新しい経

表 6: 一視点固定型ライブ映像への満足度が高い参加者の評価値

適用条件	没入感	社会的存在感	演者への興味度
None	4.02	4.20	3.96
Visual	4.09	4.39	4.15
Audio	3.97	4.08	3.97
Both	4.13	4.50	4.15

表 7: 一視点固定型ライブ映像への満足度が低い参加者の評価値

適用条件	没入感	社会的存在感	演者への興味度
None	3.40	3.33	3.04
Visual	3.89	4.08	3.74
Audio	3.20	3.27	3.08
Both	3.69	3.98	3.64

験ができた」「個々の演者にも注目がいくようになるので良いと思った」と好印象の意見も得られた。また音響効果に気づかない人もおり、参加者の個人差が大きく結果に出た。

4.6 考察

没入感、社会的存在感、演者の興味度の全てにおいて、音響効果に関係なく、映像効果を付与することで評価が高くなっていた。このことは、映像効果を付与することが視聴者の音楽体験の向上に寄与する可能性を示唆している。そのため、実際のライブ映像配信においても映像的变化を付与することが視覚的関心の維持に寄与する可能性を示唆するものである。

一方、音響効果による影響は確認されず、アンケートにおいても評価が大きく分かれていた。今回の実験では、楽器経験者が 14 人、未経験者が 10 人であり、楽器経験の有無による影響が結果に反映されていた可能性がある。「集中できなかった」「違和感があった」など、音響効果に対してマイナスな意見を述べていた参加者の内の多くが楽器経験者であり、逆に楽器経験者は好印象な印象を持ち、中には音響効果に気づかない人もいた。そのため、楽器経験の有無が音響効果の認識や印象に影響を与えることが考えられる。またこれはユーザの意図しないタイミングで切り替わる違和感が顕著に現れたと思われる。一方、一視点から撮影されたライブ映像への満足度が高いユーザには少し効果が見られ、新しい体験ができたという意見も得られた。

ここで、各試行をまとめて分析した場合上記のような結果になるが、その中でもより参加者の好みにマッチした映像と表現の組み合わせがあると考えられる。そこで各参加者それぞれの結果を再度分析し、どの条件の動画が最も評価が高いのかを比較する。表 8 は、各評価軸において、参加者ごとに最も評価が高かった試行の適用条件をカウントしまとめたものである。没入感、社会的存在感においては、Both 条件が最も多くなっており、演者への興味度においては Visual 条件が最も多くなっている。また全ての評価軸をまとめた際には、Both 条件が最も高くなっているため、動画の種類によっては映像効果

表 8: 各参加者における最も評価が高い試行での適用条件

適用条件	没入感	社会的存在感	演者への興味度	全評価軸
None	6	4	2	3
Visual	6	8	10	7
Audio	4	2	3	3
Both	8	10	9	11

と音響効果を組み合わせることで音楽体験を向上させる可能性がある。したがって、適用のタイミングを慎重に選定し音響効果を取り入れることは有効な方法の一つとなり得る。また、取り入れるには異なるアプローチが必要となる。例えば、視聴者のインタラクションに応じて映像効果と音響効果を適用することは今後の方針として考えられる。

しかし、本手法を実際のライブ映像配信に適用するには、リアルタイム処理に関する技術的課題が依然として残されている。そこで、現状のシステムの性能を図るため、12本のライブ映像に対する処理速度を測定した。処理環境には、GPU (NVIDIA GeForce RTX 4090, 24GB) を搭載した、GALLERIA シリーズのデスクトップ PC を使用した。その結果、30 秒の動画に対して平均 49.77 秒かかっていた。このうち、音源分離には 3.60 秒、オンセット検出には 0.94 秒、映像効果付与には 1.05 秒、音響効果付与には 0.01 秒、演者認識には 14.60 秒、動作量算出には 13.81 秒かかっていた。

ライブ配信では通常数秒から数十秒の遅延が発生するが、大きすぎる遅延はリアルタイム性に影響してくると考えられるため、1 分以内の遅延に抑えることを目指している。そのため、実用化するには演者認識、動作量算出の処理速度を上げる必要がある。今のシステムでは 1 フレームごとに演者を検出しフレーム間で特徴量を比較することで動作量を求めている。処理の間隔を広げ、間のフレームでの情報を補完することで処理を軽くできることが見込める。さらに、演者認識と動作量算出をするフレームをずらし並列処理にすることで、さらなる処理時間の削減が見込める。

また現状では音響に関する処理（オンセット検出、音源分離）は最初にバッチ処理をしており、ストリーミング処理で実装する必要がある。オンセット検出においては、瞬間的な音量差を元にした処理を模索する予定である。音源分離のリアルタイム適用においては、ライブにはリハーサルの時間が設けられるため、このタイミングで事前に正解データを取得することで、演者ごとの音響特性を把握し、ライブ配信中のリアルタイム処理を実現できる可能性がある。

加えて、リアルタイムでの実装にむけて映像・音響効果の処理を改良する必要がある。現状では映像をオンセットによって分割し、そのブロック単位でフレームや音声に表現を付与しているため、リアルタイムを想定した際に遅延時間にプラスで処理時間が加わってしまう。逐次的にフレーム間での動作量を比較し表現を付与するように改良していく。

しかし、本研究では効果の付与が演奏内容と文脈的に整合していたことが評価向上の要因であるかどうかは直接検証していない。すなわち、効果の存在自体が評価を高めた可能性も否定できない。今後は、映像効果の適用タイミングや対象をランダム化した条件を設けることで、文脈整合性の寄与を明確化する必要がある。

第5章 リアルタイム配信のためのシステムの実装と評価

5.1 実験概要

3章, 4章において, 一視点から撮影されたライブ映像に自動で映像効果を付与するシステムが視聴時間, 臨場感を向上させることを明らかにした. 一方, 音響効果の主な影響は確認されなかった. ライブは予測不能性 [1] や即時性 [2], 同時性 [1, 3, 4] の観点から, リアルタイムであることの重要性が示されている. そこで本実験では, 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.

5.2 リアルタイム映像効果付与システム

5.2.1 設計指針

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システム「OneCamLiveFX」を構築する. 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 [40], 視聴中心のケースでは数秒の遅延は許容される [41] 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



図 15: 適用する映像効果

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、前章ではズーム、画面分割、ラディアルブラーを採用していた。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15)。

5.2.2 システム概要

全体的な処理の流れを図16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。

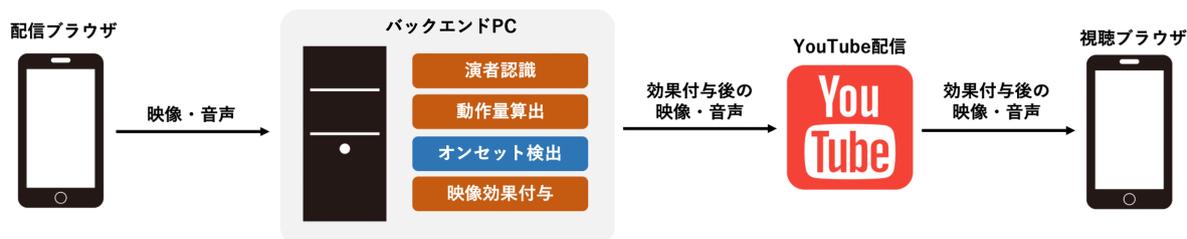


図 16: リアルタイム映像効果付与システムの処理の流れ

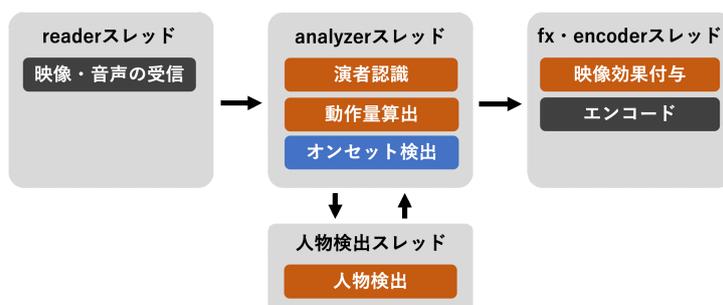


図 17: バックエンド内の処理の流れ

5.2.3 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 17 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し、YouTube Live に配信する。

まず reader スレッドでは、配信ブラウザから RTMP¹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／音声チャンクの 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

¹Real-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。

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。

5.2.4 配信・視聴インターフェース

ユーザは配信端末 (図 18) 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することができる。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、



図 18: 配信端末画面



図 19: 視聴端末画面

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 19)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。

5.3 実験設計

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。本研究では以下の仮説を立てた。

- 仮説 1：本システムはリアルタイムでの運用可能性が見込める
- 仮説 2-1：本システムによる配信は視聴者の臨場感を向上させ、リアルタイム性がその効果を助長する
- 仮説 2-2：本システムは視聴者にとって満足度が高く、継続利用が見込める
- 仮説 3：本システムは演者にとっても評価が高く、継続利用が見込める

本実験では、システムログによる客観指標と、アンケートによる主観指標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

表 9: 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)

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。

また主観指標として、視聴者アンケートにより没入感、社会的存在感、演者への興味度を測定した。これらは前章と同様のアンケートを使用した。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 [42]。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 9)。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した。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答えてもらった。

5.4 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を操作し、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。

5.5 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10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。

5.5.1 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11, 12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

表 10: 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 11: 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 12: 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

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 11 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 12 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を表 13 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。

バックエンドでは、処理の追従性を評価するため、fx・encode スレッドへの順番待ちキューの滞留量を記録した。表 14 より、FX:OFF 区間では fx・encode スレッドへの順番待ちキューの中央値は 0 フレームであり、p90 も 1.0 フレームと小さかった。また、100 フレーム以上の滞留が生じたサンプルの割合は 0.2% であった。一方で最大値は 147 フレームであり、短時間の滞留は存在していたことが分かる。表 15 より、FX:ON 区間では fx・encode スレッドへの順番待ちキューの滞留が顕著に増加した。中央値は 12 フレーム、p90 は 136.0 フレームとなり、100 フレーム以上の滞留が生じた割合は 16.5% であった。特に

表 13: 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 14: キュー滞留量 (FX:OFF)

配信	中央値	p90	最大	100 フレーム以上の割合 (%)
1 回目	0	1.0	142	0.3
2 回目	0	1.0	147	0.2
3 回目	0	1.0	141	0.1
全体	0	1.0	147	0.2

表 15: キュー滞留量 (FX:ON)

配信	中央値	p90	最大	100 フレーム以上の割合 (%)
1 回目	1	117.9	300	13.0
2 回目	1	110.6	300	11.8
3 回目	41	170.0	300	24.4
全体	12	136.0	300	16.5

3 回目では中央値が 41 フレーム、p90 が 170.0 フレームとなっており、他の回と比べて追従性が低下している。

5.5.2 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 であり安定していた。

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 16 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 17 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

表 16: 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

表 17: 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

ではドロップは観測されなかった。

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 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。

5.5.3 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 18 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。さらにこれらの指標をまとめた臨場感では平均 4.22(SD=0.76)であった。中点との差は 0.72 であり、Wilcoxon 符号付順位検定の結果、中点を有意に上回った ($p = 0.0043$)。

次に、満足度及び継続利用可能性の結果を表 19 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

表 18: 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 19: 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 20: 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	5.42	0.92	.002**
継続利用意思	5.18	1.35	.011*

注: * $p < .05$.

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。

5.5.4 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 20 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。

5.6 考察

5.6.1 ログによる客観指標の評価

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。このことから、「本システムはリアルタイムでの運用可能性が見込める」という仮説は支持されたとと言える。

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にく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。

ここで、FX:ON 区間において、遅延が 13s 以上となる区間（意図的遅延 5s に対してその他遅延が約 8s 以上）を抽出し、モジュール処理時間およびキュー滞留との同期を確認した。その結果、エンコード入力キューの急増と強く同期しており、この区間ではキュー滞留の中央値が 242–269 フレームまで増加していた。また、映像効果付与（FX）の 1 回あたり処理時間が平均で約 2.54–4.22 倍に増加していた。一方で、YOLO+トラッキングは平均で 1.01–1.04 倍程度の増加に留まり、エンコードも平均で約 1.07–1.14 倍の増加であった。以上より、FX 合成処理の一時的な高負荷を起点としてエンコード入力キューに滞留が蓄積し、その結果遅延の増加やフレームドロップが発生したと考えられる。

5.6.2 アンケートによる主観指標の評価

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。さらに、これらの項目をまとめた、臨場感でも中点を有意に上回っていた。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、「本システムによる配信は視聴者の臨場感を向上させる」という仮説は支持された。

視聴者の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。そのため、「本システムは視聴者にとって満足度が高く、継続利用が見込める」という仮説は支持されなかった。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、人物とそれ以外と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。このことから、「本システムは演者にとっても評価が高く、継続利用が見込める」という仮説は支持されたと言える。また、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。

第6章 総合考察と今後の展望

6.1 総合考察

本研究は、軽音楽団体で多く見られる一視点固定型のオンラインライブにおいて、映像が単調になりやすく臨場感が乏しいという課題に対し、ライブ映像中の特徴的なシーンを自動検出して映像（および一部音響）効果を付与することで、視聴体験の向上を目指したものである。具体的には、映像効果付与の基礎検討と影響調査、映像・音響効果の組み合わせに関する主観評価、リアルタイム運用可能なシステム実装と現場実証、という段階的な検討を行った。

まず第3章の基礎検討では、ライブ映像にズーム、画面分割、集中線の映像効果を付与することが視聴者の注意の持続やエンゲージメントにどう影響するのかを調査した。その結果、映像効果を付与することで視聴時間、エンゲージメントが増加した。一方で映像効果の種類の高さには影響が見られず、適用頻度による影響が強く反映されている可能性がある。またアンケートでは、映像効果付与に対する印象に差が見られたため、映像効果の種類や楽曲の曲調を考慮する必要がある。

次に第4章の主観評価実験では、臨場感の関連概念として没入感・社会的存在感・演者への興味度を指標として挙げ、その全てにおいて、音響効果の有無にかかわらず映像効果を付与することで評価が高くなっていた。これは、一視点固定型映像においても、提案手法が位置視点固定型ライブ映像の視覚的興味を維持し、体験価値を高め得ることを示している。一方で音響効果については、明確な影響が見られず評価が分かれたことから、効果の知覚には個人差が大きく、適用タイミングや効果の設計をより慎重に行う必要がある。

さらに第5章では、リアルタイム運用を前提とした「OneCamLiveFX」を実装し、性能評価と現場実証を行った。ログに基づく評価では、FXを無効化した場合は安定していた一方、FXを有効化すると遅延時間やフレームドロップの不安定さが見られた。またモジュール内訳から、人物検出と映像効果付与が主要ボトルネックとなり、負荷増大区間で追従性が低下しやすいことが分かった。実効出力処理率はFX:ONで分布が大きく拡

大し、特に映像効果付与を起点として遅延の増加やフレームドロップが発生していたことから、演出強化はQoSとトレードオフ関係になり得る。主観評価は、視聴者アンケートにおいて没入感・社会的存在感・演者への興味度はいずれも中点を上回り、特にリアルタイムでの運用により社会的存在感が強く増した可能性が示された。一方で、満足度は中点を明確に上回ったものの、継続利用意思は相対的にばらつきが大きく、自由記述では「自分でズーム対象を決めたい」など主体性やパーソナライズを求める意見が多かった。また「演奏に関係ない部分にズームされていた」という指摘も見られ、映像効果の誤適用がQoEを損なうリスクを示しており、人物検出や映像効果付与の精度向上が体験安定化に直結することが考えられる。一方演者側では、満足度・継続利用意思が中点を上回り、両者に強い正の相関 (Spearman $\rho = 0.89$) が見られた点から、自分（または演者）が強調されることの価値が大きい可能性が示唆された。

以上を総合すると、本研究の提案手法は視聴維持や視聴体験の向上に寄与し得る一方で、リアルタイム運用では処理負荷増大による遅延・フレームドロップが起こり得るため、演出強化とQoSのトレードオフ関係を意識して設計する必要があるという結論に整理できる。

6.2 制約と今後の展望

本研究には以下の制約がある。

まず、対象としたライブ映像の編成・撮影条件が限定されており、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。そのため、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誤検出・誤適用が増える可能性がある。特に意図しない対象へのズームはQoEを下げるリスクが示唆されており、精度の改善が必要となる。

次に、音響効果については効果が一貫して確認できず、個人差が大きかった。これは表現の選定・強度に依存する可能性があり、映像効果に比べて「気づかれにくい」「不快にもなり得る」といった側面を持つ。今後は、音響効果の知覚しやすさを担保する設計、音楽構造に沿った適用といった工夫が必要である。

さらに、本稿で行った3つの実験では、用いた映像の長さが実験ごとに大きく異なっている。3章では2分程度の映像を視聴者の好きな箇所まで見てもらい、4章では30秒程度の映像を最後まで視聴してもらった。さらに5章では約20分のライブ映像配信をリアルタイムで視聴してもらった。このような視聴時間の差が、結果に影響を与えている可能性がある。そのため、それぞれの結果の比較や解釈には十分な注意が必要である。

また、提案手法では様々な映像効果を付与しているが、付与頻度を計測していなかったため、その妥当性は検証できていない。今後は付与頻度による印象への影響を調査し、適切な割合で映像効果を付与できる仕組みを検討していく。

加えて、効果付与の自動化は視聴体験向上に影響する一方で、演者の表現意図とのズレや制御不能性といった面の問題も考えられる。5章において、受動的な自動演出だけでなく、主体性やパーソナライズに焦点を当てることが継続利用に繋がる可能性が示唆された。バーチャルコンサート視聴においては、能動的な視点切替が臨場感を高め、満足度やコンサート参加意図を向上させることも明らかになっている [43]。そのため将来的には、視聴者が見たい演者を選べるような、インタラクティブな設計が有望である。また演者側の満足度と継続意図が強く関連するという結果を踏まえると、演者が配信前に「強調されたい瞬間」や「見せ場」を簡単に指定できる仕組みも改善案となり得る。

第7章 まとめ

本研究では、軽音楽団体における一視点固定型オンラインライブ配信の映像の単調さに起因する臨場感の不足に対し、ライブ映像中の特徴的シーンを自動検出して映像（および一部音響）効果を付与することで視聴体験を向上させる手法を検討した。

まず、ライブ映像への映像効果の自動付与が、平均視聴時間を増加させ、視聴維持に寄与する可能性を示した。一方で、人による印象の差が大きく、適用頻度や演出設計の最適化が課題であることを示した。次に、主観評価実験を通して、没入感・社会的存在感・演者への興味度の観点から、映像効果付与が視聴体験を向上させる傾向を確認した。音響効果については影響があまり見られず、提示条件や適用タイミングを含めた設計改善が必要であることを示した。そして、リアルタイム配信に対応した「OneCamLiveFX」を実装し、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した。その結果、FX有効化時に遅延の増加やフレームドロップが確認された一方で、視聴者の没入感・社会的存在感・演者への興味度は中点を上回り、リアルタイムにおいても提案手法が視聴体験を向上させることが示唆された。また、演者の満足度、継続利用意思も高く、両者に強い正の相関が見られたことから、演者にとっても価値の高いものであることを確認することができた。

以上より、一視点固定型ライブ映像に対して、自動的に映像効果を付与することが視聴体験向上に有効である可能性を示すとともに、リアルタイム運用における演出強化とQoSの両立という課題を明確化した。今後は、映像効果の適用タイミングの精度向上、処理負荷に応じた制御、および視聴者・演者の関与を取り込むインタラクティブな設計を進めることで、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。

謝辞

本研究を行うにあたり，多くのご指導をいただいた中村聡史先生に感謝申し上げます。
また，様々なアドバイスをくれた研究室の皆様方，ありがとうございました。

参考文献

- [1] Claudia Georgi. *Liveness on stage: Intermedial challenges in contemporary British theatre and performance*, volume 25. Walter de Gruyter GmbH & Co KG, 2014.
- [2] Melanie Wald-Fuhrmann, Hauke Egermann, Anna Czepiel, Katherine O’Neill, Christian Weining, Deborah Meier, Wolfgang Tschacher, Folkert Uhde, Jutta Toelle, and Martin Tröndle. Music listening in classical concerts: Theory, literature review, and research program. *Frontiers in Psychology*, 12, 2021.
- [3] Kelsey E Onderdijk, Dana Swarbrick, Bavo Van Kerrebroeck, Maximillian Mantei, Jonna K Vuoskoski, Pieter-Jan Maes, and Marc Leman. Livestream experiments: the role of covid-19, agency, presence, and social context in facilitating social connectedness. *Frontiers in Psychology*, 12, 2021.
- [4] Sophie T Kulla, Hande Sungur, and Sindy R Sumter. The role of social presence in live and recorded concert viewing: Effects on enjoyment and emotional well-being. *Computers in Human Behavior Reports*, 14:100409, 2024.
- [5] M Lombard. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.
- [6] Pietro Modestini and Christian Weining. Affordances and experiential dimensions of digital concerts: Closer the eye, farther the crowd. *Music Perception: An Interdisciplinary Journal*, 42(3):269–285, 2025.
- [7] 山根将太, 池谷絵理子, and 佐藤美恵. 感性コンテンツにおける印象抽出と評価—ズームとカットを併用した手法によるコンテンツ制作と感性的評価. 映像情報メディア学会技術報告, 28(27):37–40, 2004.
- [8] Mark Ludwig and Christoph Bertling. The effect of cutting rates on the liking of live sports broadcasts. *International Journal of Sport Communication*, 10:359–370, 09 2017.

- [9] 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. 参照 2025-12-09.
- [10] 畑田豊彦, 坂田晴夫, and 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*, 33(5):407–413, 1979.
- [11] Tianyi Wang and Shima Okada. Heart fire for online live-streamed concerts: a pilot study of a smartwatch-based musician-listener interaction system. *Frontiers in Computer Science*, 5:1150348, 2023.
- [12] Ryota Horie, Minami Wada, and Eri Watanabe. Participation in a virtual reality concert via brainwave and heartbeat. In *International Conference on Applied Human Factors and Ergonomics*, pages 276–284. Springer, 2017.
- [13] 柳沢豊, 小野圭介, 上田健太郎, 出田怜, 吉池俊貴, and 藤本実. 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ーション (DCC)*, 2021(2):1–5, 2021.
- [14] Kwan Min Lee. Presence, explicated. *Communication theory*, 14(1):27–50, 2004.
- [15] Dixi Zhong, Qiongxin Li, and Xueting Zhai. The digital transformation of music events: from being there to the feeling of being there. *Journal of Hospitality and Tourism Technology*, 16(5):1193–1212, 08 2025.
- [16] Tara Venkatesan and Qian Janice Wang. Feeling connected: The role of haptic feedback in vr concerts and the impact of haptic music players on the music listening experience. *Arts*, 12(4):148, 2023.
- [17] D. Swarbrick, R. Martin, S. Høffding, N. Nielsen, and J. K. Vuoskoski. Audience musical absorption: Exploring attention and affect in the live concert setting. *Music & Science*, 7:1–19, 2024.
- [18] Gillian Sandstrom and Frank Russo. Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, 41:216–228, 03 2013.
- [19] Werner Wirth, Tilo Hartmann, Saskia Böcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Ribeiro Gouveia,

- et al. A process model of the formation of spatial presence experiences. *Media psychology*, 9(3):493–525, 2007.
- [20] Jay David Bolter and Richard Grusin. *Remediation: Understanding New Media*. The MIT Press, Cambridge, MA, 1999.
- [21] Jonathan Steuer et al. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93, 1992.
- [22] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [23] Annie Lang. The limited capacity model of motivated mediated message processing. *The SAGE handbook of media processes and effects*, 11:193–204, 2009.
- [24] Wayne S Smith and Yoav Tadmor. Nonblurred regions show priority for gaze direction over spatial blur. *The Quarterly Journal of Experimental Psychology*, 66(5):927–945, 2013.
- [25] Rizwan Ahmed Khan, Eric Dinet, and Hubert Konik. Visual attention: Effects of blur. In *2011 18th IEEE international conference on image processing*, pages 3289–3292. IEEE, 2011.
- [26] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.
- [27] Richard E Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.
- [28] Ángel Muñoz-González, Shohei Kobayashi, and Ryota Horie. A multiplayer vr live concert with information exchange through feedback modulated by eeg signals. *IEEE Transactions on Human-Machine Systems*, 52(2):248–255, 2022.
- [29] Kota Ide and Ryota Horie. Improving a sense of unity via brainwaves in a virtual concert where performances in real space are watched in a virtual space. In *ICAT-EGVE (Posters and Demos)*, pages 25–26, 2022.

- [30] 大石あい and 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, 121(23):7–12, 2021.
- [31] Mincheol Shin, Stephen W Song, Se Jung Kim, and Frank Biocca. The effects of 3d sound in a 360-degree live concert video on social presence, parasocial interaction, enjoyment, and intent of financial supportive action. *International Journal of Human-Computer Studies*, 126:81–93, 2019.
- [32] Neil Cohn. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. A&C Black, 2013.
- [33] Thomas Lamarre. *The anime machine: A media theory of animation*. U of Minnesota Press, 2009.
- [34] James J Cummings and Jeremy N Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media psychology*, 19(2):272–309, 2016.
- [35] Hayeon Kim and In-Kwon Lee. Studying the effects of congruence of auditory and visual stimuli on virtual reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2080–2090, 2022.
- [36] Patrycja Delong and Uta Noppeney. Semantic and spatial congruency mould audiovisual integration depending on perceptual awareness. *Scientific Reports*, 11(1):10832, 2021.
- [37] Matthew Lombard, Theresa B Ditton, and Lisa Weinstein. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence*, pages 1–15. International Society for Presence Research Los Angeles, CA, 2009.
- [38] Stephen Reysen and Nyla R. Branscombe. Fanship and fandom: Comparisons between sport fans and non-sport fans. *Journal of Sport Behavior*, 33(2):176–193, 2010.
- [39] Padmini Patwardhan, Jin Yang, and Hemant Patwardhan. Understanding media satisfaction: Development and validation of an affect-based scale. *Atlantic Journal of Communication*, 19:169–188, 07 2011.

-
- [40] Karin Van Es. Liveness redux: on media and their claim to be live. *Media, Culture & Society*, 39(8):1245–1256, 2017.
- [41] Jinyang Li, Zhenyu Li, Ri Lu, Kai Xiao, Songlin Li, Jufeng Chen, Jingyu Yang, Chunli Zong, Aiyun Chen, Qinghua Wu, et al. Livenet: a low-latency video transport network for large-scale live streaming. In *Proceedings of the ACM SIGCOMM 2022 Conference*, pages 812–825, 2022.
- [42] Anol Bhattacharjee. Understanding information systems continuance: An expectation-confirmation model. *MIS quarterly*, pages 351–370, 2001.
- [43] Shih-Yu Lo and Chih-Yuan Lai. Investigating how immersive virtual reality and active navigation mediate the experience of virtual concerts. *Scientific Reports*, 13(1):8507, 2023.

研究業績

- [1] 小川 剣二郎, 青木 柊八, 中村 瞭汰, 山中 祥太. PCとスマートフォンにおけるプログレスバーと周辺視野への視覚刺激の提示による体感時間短縮効果の調査, 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , Vol.2023-HCI-202, No.55, pp.1-7, 2023.
- [2] 三山 貴也, 小川 剣二郎, 青木 柊八, 中村 瞭汰, 中村 聡史, 山中 祥太. 周辺視野への視覚刺激の提示が読み込み中のページ離脱率に及ぼす影響, 信学技報, Vol.123, No.188, HCS2023-63, pp.35-40.
- [3] 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.
- [4] 小川 剣二郎, 中村 聡史. 一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 研究報告エンタテインメントコンピューティング (EC) , Vol.2025-EC-75, No.45, pp.1-8, 2025.
- [5] Kenjiro Ogawa, Satoshi Nakamura, Shota Yamanaka, Ryota Nakamura. Reducing Perceived Waiting Time with Peripheral Visual Motion: Directional and Device-Specific Effects, 37th Australian Conference on Human-Computer Interaction (OzCHI 2025), pp.122-135, 2025.
- [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.