

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

小川 剣二郎¹ 中村 聡史¹

概要：軽音楽団体でのライブ映像配信では、撮影・配信のための機材が揃っていないために一視点から撮影することが多く、またこういった映像は臨場感が伝わりづらく離脱者が多い。以前の調査では、音楽に関する漫画・アニメで用いられている視覚的表現をライブ映像の特徴的なシーンに付与することが、視聴維持率、印象の向上に有効であることを明らかにした。ここで、視覚と聴覚の情報の一致による認知、感性への効果が報告されており、映像表現に合わせて音響表現を加えることでその効果を高められるのではないかと考えた。そこで本研究では、映像・音響表現の付与による、ライブ映像視聴時の音楽体験への影響を、没入感、社会的存在感、演者への興味度を比較し調査した。その結果、どの評価軸においても映像表現を付与することが有効であることが明らかになった。また、一視点型ライブ映像への満足度が高い場合は、映像表現と音響表現を組み合わせることにより音楽体験を向上させられる可能性が示唆された。

キーワード：ライブ映像配信, 映像表現, 音響表現, 臨場感, 没入度, 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したい際には、対面ライブと並行してオンライン配信ライブを行うことが多い。オンライン配信ライブは気軽に周知を行うことができ、新しい入会者の招集やその団体の活性化などを目的とした場面で開催される。

こうした配信では、リアルタイムであることが重要であり、視聴者の体験に影響する。ライブ映像視聴時の感動と感情を共有している感覚には相関関係があり、オンラインライブ視聴時は事前に録画されたライブ視聴時よりも感情を共有している感覚が高まることが明らかになっている [1]。そのため、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られる。

軽音楽団体でのオンライン配信ライブでは、主なプラットフォームとして Instagram や YouTube などの SNS が用いられる。しかしこれらの配信では、視聴者がすぐ離脱を

してしまい、多くの人を留まらせておくことが難しい。このようなことが起こる要因として、一視点から撮影された映像を配信していること、利用端末としてスマートフォンが多く [2] 視聴時の画角が小さくなり臨場感が薄くなること [3] の 2 つが挙げられる。山根ら [4] は、場面に合わせてカットとズームインを組み合わせることが没入感を高めると述べており、さらに適した頻度のカットが視覚的興味を維持することを示唆している。カット頻度による印象に関する研究は他にも見られ [5]、映像に変化を持たせることが視聴体験を向上させると考えられる。

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を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表現を付与する手法を提案した [6]。臨場感を向上させる音楽表現として漫画・アニメのものを参考にし、ズーム、集中線、画面分割の 3 種類の表現を採用した。複数のライブ映像を視聴してもらう実験を行い、映像表現無し、ズームのみ付与、全ての映像表現を付与の 3 条件で比較した。その結果、提案手法によって映像表現を付与することにより視聴時間、印象が向上することを明らかにした。

ここで、より臨場感を向上させる方法として、音響表現を用いることが考えられる。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 [7] [8]、映像表現と同じ対象を音響表現でフォー

¹ 明治大学
Meiji University

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。軽音楽団体においてライブ映像配信から入会や団体の活性化に繋げるには、そのライブ自体や演奏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。そこでより臨場感を向上させその団体に興味を持ってもらうことに着目し、映像表現に合わせた音響表現を付与することにより音楽体験の向上を目指す。

そこで本研究では、映像・音響表現の付与、さらにそれらを組み合わせることによる、ライブ映像視聴時の音楽体験への影響を調査し、よりライブの魅力伝えられる手法を模索する。具体的には、それぞれの条件において没入感、社会的存在感、演者への興味度を測ることで、詳細な影響を調査する。

2. 関連研究

2.1 ライブ映像の臨場感向上に関する研究

特有のデバイスを用いることでオンライン配信ライブの臨場感を向上させる研究は複数行われている。Munozら [9] は、HMDと小型脳波レコーダーを利用し、観客の脳の状態を共有できるフィードバックシステムを提案している。VR (Virtual Reality) コンサートに参加しているユーザの脳波を測定し、 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一体感を増強させている。柳沢ら [10] 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。

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 [11] 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を明らかにした。Shinら [12] 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た。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、ライブ音源の発音タイミング、演者の動作量をもとに自動で映像・音響処理を施す手法を提案し、音楽体験への影響を調査する。

2.2 視覚・聴覚の情報の一致による印象に関する研究

空間的に視覚・聴覚の情報を一致させることによる印象を調査している研究は多数存在する。Kimら [7] は、VR環境において聴覚と視覚の刺激の一致・不一致が、ユーザーの体験にどのような影響を及ぼすかを調査した。聴覚と視覚の刺激が一致しており、感覚の整合性が取れているとVR環境に適応しやすく、没入感と臨場感が向上することを明らかにした。DeLong [8] らは、視覚と聴覚の意味的および空間的一致性が、視聴覚統合にどのように影響するかを分析した。その結果、どちらの意味合いにおいても、一致性が視聴覚統合を促進することが確認された。

本研究では、ライブ映像において映像・音響表現の付与、さらにそれらを組み合わせることによる音楽体験への影響を、没入度、社会的存在感、演者への興味度を測ることで比較する。

2.3 特徴的なシーンの自動検出に関する研究

映像において特徴的なシーンを自動検出する手法を提案している研究は複数存在する。Sharmaら [13] は、マルチレベルの時空間双方向ネットワークとマルチスケール転移学習を組み合わせることにより、時空間情報を考慮したニューラルネットワーク (MSB-Net) を提案している。物体認識や映像解析の精度を向上させることを目的としており、既存の顕著オブジェクト検出モデルと比較して、精度と計算効率の面で優れていることが実証された。Sevilmicら [14] は、映像内の顕著なオブジェクトとその空間的な関係を自動検出する手法を提案している。提案手法により、映像内の意味的に重要なシーンの抽出が可能になり、物理解やコンテンツ解析の精度が向上した。画像認識技術と組み合わせることで、映画のハイライトシーンの自動抽出や、スポーツ映像のキーイベント検出に応用可能であると述べている。

さらに、ライブ映像における特徴的なシーンの自動検出手法も多く提案されている。Shresthaら [15] は、音楽のライブ中に複数人のアマチュアの人によって撮影された映像をまとめて、1つの映像に自動生成するシステムを提案している。複数カメラによる映像を同期し、ファーストフィットと呼ばれるアルゴリズムを使用することで自動編集を実現している。Roininenら [16] は、ライブ映像を対象にプロが編集したカットタイミングを分析してモデルを作成し、オーディオ変化点分析およびカット偏差モデルと組み合わせることで、暫定的な編集を自動的にしてくれるアルゴリズムを提案している。

本研究でもこれらの研究と同じように、映像と音声の両方から特徴的なシーンを自動検出し、さらにそのタイミングで映像・音響表現を適用することで臨場感の向上を図る。

3. 映像・音響表現の付与手法

3.1 提案手法

本手法の流れを図1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、映像表現と音響表現の付与をしている。以下では、映像表現、音響表現の付与手法について説明する。

我々はこれまでの研究 [6] において、ライブ音源の発音タイミングと演者の動作量を用いて特徴的なシーンを検出し、適切な映像表現を付与する手法を提案してきた。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。これらの変更により、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

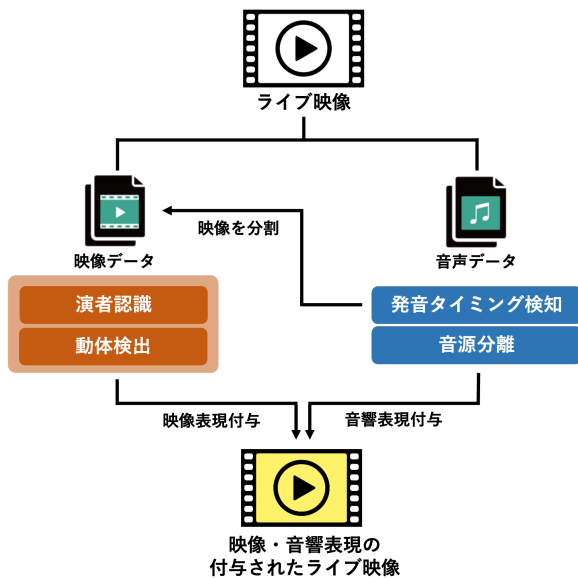


図1: 処理の流れ



図2: 映像表現付与による画面パターン

の3種類の映像表現(図2)を付与することによる印象を再度調査する。また、映像表現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、音源から発音タイミング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像表現を付与する。

音響的に特定の楽器(周波数)を際立たせる手法として、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。しかし、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、音響表現を付与するタイミングは映像表現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、映像表現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3)。

この、映像表現の付与と音響表現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。

3.2 音響表現の実装

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、音響表現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した。

まず、ライブ音源を Demucs を用いて音源分離をし元の音源に重ねて強調をするが、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。ラディアルブラーではその対象の楽器を、画面分割では表示する順にその対象の楽器を元の音源に重ね、全体の音量を調節することで強調している。

3.3 付与タイミングの実装

まず音源から発音タイミングを検出しそれをもとに映像を分割することで、音源的にタイミングのいい箇所をブロック分けする。発音タイミングを用いることで、サビ

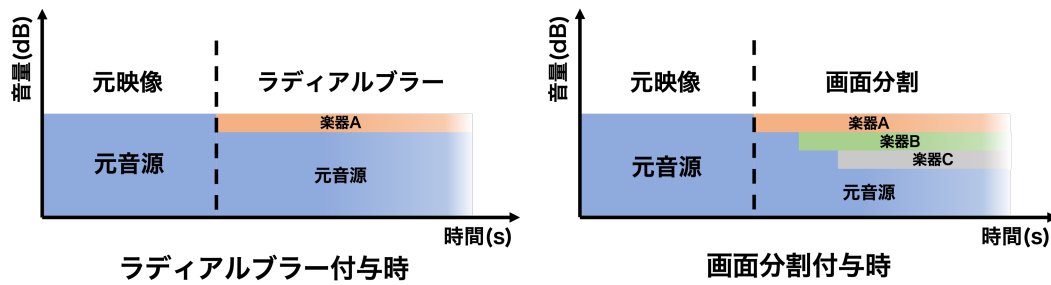


図 3: 音響表現イメージ

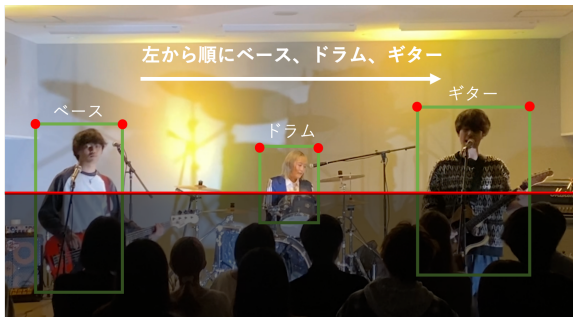


図 4: 演者の認識

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する。映像からは、物体検出ライブラリ YOLOv10 を用いて映像の中の人物を検出する。ここで映像は、ステージの真正面のやや上から撮影されたものとし、演奏者 3 人で構成されているバンド (3 ピースバンド) に限定している。この際、演奏者は画面上部に、観客は画面下部に映される。また、演奏者のパートの位置は通常固定であり、3 ピースバンドの場合観客側から見て左からベース、ドラム、ギターの順となる。そのため、上から順に 3 人を演奏者とし、左から順にパートを振っていった (図 4)。ここで、上手く演奏者が検出されないフレームもある。そのため、上手く検出された際のフレーム情報をストックしておき、再度検出された際に情報を更新するといった処理を挟むことで、ほぼ全てのフレームにおいて正しい演奏者の位置を検出可能としている。続いて、OpenCV を用いて演奏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。

さらに、発音タイミングによって分けた箇所の前後のブロック同士で演奏者の動作量を比較することにより表現を付与するタイミング、対象を決定する。動作量の増減の割合を求め、前回のブロックと比べ全ての演奏者の動作量が 2.0 倍以上になっていたら画面分割、1 人でも動作量が 2.0 倍以上になっていたらその中で特に動作量が多い演奏者にラディアルブラー、それ以外で 1 人でも動作量が 1.2 倍以上になっていたらズームを付与している。その際、画面分割ではよりそれぞれの演奏者に視線を向けるため、動作量が多い演奏者ごとに 1 秒ずつ時間差表示をしている。

4. 実験

4.1 実験目的

本研究での提案手法による、映像・音響表現の自動付与が、ライブ映像視聴時の音楽体験へどのような影響を与えるのか調査するため、複数のライブ映像に対してアンケートに答えてもらう実験を行なった。アンケートは没入感、社会的存在感、演奏者への興味度に関するものであり、それらの結果を比較することで詳細に分析する。

4.2 評価基準の選定

エンタテインメントにおけるユーザ体験を測る指標として、臨場感が挙げられる。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、インタラクティブ性、リアリティなど多くが含まれている。Lombard [17] は、過去の研究から臨場感の概念を包括的に整理し、存在感やリアリティ、没入感、社会性に関する 6 つの概念としてまとめている。その中で、ライブ映像視聴時の音楽体験では、没入感、社会的存在感が特に直接的に関わってくると考えられる。

また、演奏者への興味度も視聴者の行動に関わってくる。それぞれの演奏者に興味を持ち、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。また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 [18] [19]、視聴体験への影響も示唆される。

これらを踏まえ、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、演奏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 [20] 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 [18] 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

プシ音楽吸収を測定している。社会的存在感に関しては、Lombard ら [21] 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 [22]。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 1)。

4.3 実験設計

実験用の動画として、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。動画に付与する映像・音響表現の適用条件は、以下の通りとする。

- None: 映像・音響表現無し
- Visual: 映像表現のみ付与
- Audio: 音響表現のみ付与
- Both: 映像・音響表現付与

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う。この時、表現の適用条件は映像 3 本ずつ割り振った。

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。Patwardhan ら [23] 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、楽しさ、興味深さ、没入感、満足度を測る 4 項目を含んだアンケートを作成した (表 2)。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 1: 実験アンケート

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。

4.4 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参加者にはイヤホンをして行なった。図 5 のように動画、アンケートを表示し、それぞれの動画を評価してもらった。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、それぞれの動画を評価してもらった。最後には全体アンケートとして、映像・音響表現、またその他の意見を答えてもらった。

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 人であった。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。

5.1 適用条件ごとの評価値

図 6 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

表 2: ライブ映像の満足度を測るアンケート

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。



図 5: 実験画面

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, 映像表現を付与した際に評価が高くなっている. 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なかった ($p < 0.05$).

また図 7 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. 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 ($p < 0.05$). 続いて多重比較を行なった結果, None-Visual, None-Both, Visual-Audio, Audio-Both 間で有意差が認められた ($p < 0.05$). 特に Visual-Audio, Audio-Both 間は $p < 0.001$ であり, 映像表現は没入感を向上させているが, 音響表現の効果は見られないことがわかる.

図 8 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. 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 ($p < 0.05$). 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 ($p < 0.05$).

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 Alpha の値はそれぞれ 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.

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. 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, 音響表現に関係なく, 映像表現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.

表 3: 一視点固定型ライブ映像への満足度が高めの参加者の評価値

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

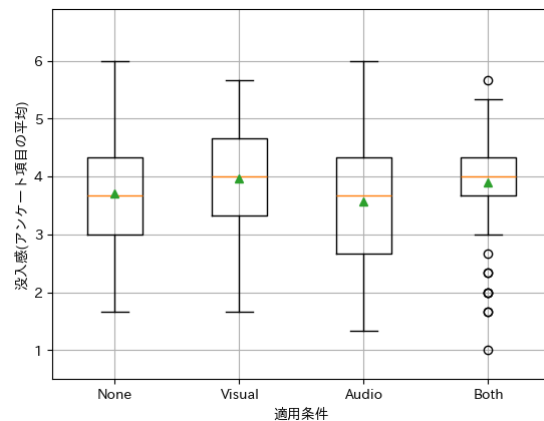


図 6: 適用条件ごとの没入感の評価値

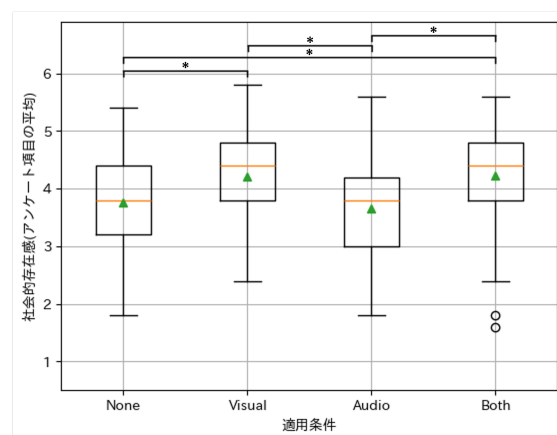


図 7: 適用条件ごとの社会的存在感の評価値

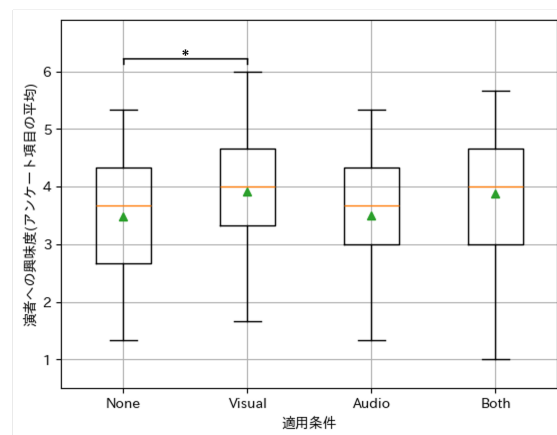


図 8: 適用条件ごとの演者への興味度の評価値

表 4: 一視点固定型ライブ映像への満足度が低めの参加者の評価値

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

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3, 4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3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めの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4では, 一視点から撮影されたライブ映像への満足度が低めの参加者は, 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, 映像表現と音響表現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。

5.4 実験後アンケート

映像表現に対しては, 「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ると思った」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, 映像表現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。

また音響表現に関しては, マイナスな意見が見られた一方で, 「映像表現と合わさって, 集中対象がわかる新しい経験ができた」「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。また音響表現に違和感を感じる人と, 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。

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, 音響表現に関係なく, 映像表現を付与することで評価が高くなっていった。このことから, 映像表現を付与することが視聴者の音楽体験を向上させることを再確認することができた。そのため, 実際のライブ映像配信においても, この手法が有効であると考えられる。

一方, 音響表現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。これには, 音響表現の質が良くなかったことも考えられるが, その他にも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。しかし, 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, また新しい体験ができたという意見も得られた。またQ6, Q7, Q10においては, 有意差はないものの映像表現と音響表現を組み合わせることによる効果が見られた。そのため, 音響表現をライブ映像に取り入れることによる音楽体験の向

表 5: 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

上可能性はあり得る。

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5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おり, 動画との組み合わせによっては音楽体験を向上させる可能性があることが考えられる。そのため付与する箇所を吟味し音響表現を取り入れることは一方針として考えられる。また, 取り入れるには別の手法にする必要がある。例えば, ユーザがインタラクションを起こしたタイミングで映像表現と音響表現を組み合わせるといったことは今後の方針として考えられる。

しかし, 本手法を実際にライブ映像配信に取り入れるには技術的制約が以前としてあり, リアルタイムでの映像・音響分析を進める必要がある。ここで音源分離に関して, ライブにはリハーサルの時間がありそのタイミングで正解データを得ることで, より正確かつリアルタイムでの処理に向けて進められるのではないかと考えている。そのため今後はリアルタイムの場合においても取り入れることのできる手法を模索し, 実際のライブ映像配信でのテストを目指していく。

7. まとめ

本研究では, 一視点から撮影されたライブ映像に映像・音響表現を付与, さらにそれらを組み合わせることによる, ライブ映像視聴時の音楽体験への影響を調査した。映像表現にはズーム, ラディアルブラー, 画面分割を採用し, 音響表現として音源分離したものを元の音源に重ねることで対象の楽器をより強調している。没入感, 社会的存在感, 演者への興味度をアンケートにより算出しそれぞれの効果を比較した結果, どの評価軸においても, 映像表現を付与することが有効であることがわかった。一方, 音響表現の効果は確認されなかったが, 一視点から撮影されたライブ映像への満足度が高い場合など一部において音楽体験を向

上させる可能性が示唆された。

ここで、実際にこれらの手法を取り入れるには技術的制約があり、リアルタイムでの処理を実現させる必要がある。そのため今後は、今回の結果を踏まえてリアルタイムの場合において取り入れることのできる手法の模索を進めていき、実際のライブ映像配信での導入を目指す予定である。

参考文献

- [1] Swarbrick, D., Seibt, B., Grinspun, N. and Vuoskoski, J. K.: Corona concerts: The effect of virtual concert characteristics on social connection and Kama Muta, *Frontiers in psychology*, Vol. 12, p. 648448 (2021).
- [2] 総務省: 令和5年通信利用動向調査の結果, 総務省(オンライン), 入手先 (https://www.soumu.go.jp/johotsusintokei/statistics/data/240607_1.pdf) (参照 2025-01-26).
- [3] 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).
- [4] 山根将太, 池谷絵理子, 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= ITE technical report*, Vol. 28, No. 27, pp. 37–40 (2004).
- [5] 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 (online), DOI: 10.1123/ijsc.2017-0046 (2017).
- [6] 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, *技術報告, 情報処理学会* (2024).
- [7] Kim, H. and Lee, I.-K.: Studying the effects of congruence of auditory and visual stimuli on virtual reality experiences, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 5, pp. 2080–2090 (2022).
- [8] Delong, P. and Noppeney, U.: Semantic and spatial congruency mould audiovisual integration depending on perceptual awareness, *Scientific Reports*, Vol. 11, No. 1, p. 10832 (2021).
- [9] Munoz-Gonzalez, A., 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).
- [10] 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).
- [11] 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).
- [12] Shin, M., Song, S. W., Kim, S. J. and Biocca, F.: 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*, Vol. 126, pp. 81–93 (2019).
- [13] Sharma, G., Singh, M. and Berwal, K.: Video Salient Object Detection Via Multi-level Spatiotemporal Bidirectional Network Using Multi-scale Transfer Learning, *IETE Journal of Research*, Vol. 70, No. 11, pp. 8077–8088 (2024).
- [14] Sevilmiş, T., Baştan, M., Güdükbay, U. and Ulusoy, Ö.: Automatic detection of salient objects and spatial relations in videos for a video database system, *Image and Vision Computing*, Vol. 26, No. 10, pp. 1384–1396 (2008).
- [15] Shrestha, P., de With, P. H., Weda, H., Barbieri, M. and Aarts, E. H.: Automatic mashup generation from multiple-camera concert recordings, *Proceedings of the 18th ACM international conference on Multimedia*, pp. 541–550 (2010).
- [16] Roininen, M. J., Leppänen, J., Eronen, A. J., Curcio, I. D. and Gabbouj, M.: Modeling the timing of cuts in automatic editing of concert videos, *Multimedia Tools and Applications*, Vol. 76, pp. 6683–6707 (2017).
- [17] Lombard, M.: At the heart of it all: The concept of presence, *Journal of Computer-Mediated Communication*, Vol. 3, No. 2 (1997).
- [18] Swarbrick, D., Martin, R., Höfding, S., Nielsen, N. and Vuoskoski, J. K.: Audience Musical Absorption: Exploring Attention and Affect in the Live Concert Setting, *Music & Science*, Vol. 7, p. 20592043241263461 (online), DOI: 10.1177/20592043241263461 (2024).
- [19] Cummings, J. J. and Bailenson, J. N.: How immersive is enough? A meta-analysis of the effect of immersive technology on user presence, *Media psychology*, Vol. 19, No. 2, pp. 272–309 (2016).
- [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 (online), DOI: 10.1177/0305735611422508 (2013).
- [21] 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).
- [22] 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).
- [23] Patwardhan, P., Yang, J. and Patwardhan, H.: Understanding Media Satisfaction: Development and Validation of an Affect-Based Scale, *Atlantic Journal of Communication*, Vol. 19, pp. 169–188 (online), DOI: 10.1080/15456870.2011.584508 (2011).