

# 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

小川剣二郎<sup>1</sup> 中村聡史<sup>1</sup>

**概要**：アーティストや軽音楽団体にとってライブは活動の軸となっている。対面ライブの他にもオンライン配信で行う場合もあり、オンラインライブはライブの実施形態として定着している。軽音楽団体がオンラインライブをする際、主にスマートフォンやビデオカメラを用いて一視点から撮影されることが多く、その場合映像に変化がなく単調なものとなってしまい、視聴者に離脱されやすい問題がある。そこで我々は、画角が固定でのオンラインライブ配信の映像に漫画・アニメにおける映像表現をリアルタイムに自動で付与することで、より臨場感のある映像を配信可能とする手法を提案する。本稿では映像データ、音声データから適切な映像表現を自動で付与し、システムとして実装するとともに、映像表現が付与されていないライブ映像と比較実験を行った。実験の結果、提案手法によって映像表現が付与されたライブ映像は視聴維持率、満足度が向上することを明らかにした。

**キーワード**：オンラインライブ、映像表現、人物認識、オンセット検知、臨場感、視聴時間、軽音

## 1. はじめに

アーティストや軽音楽団体は、定期的にライブを行うことを活動の軸としている。ライブは演奏者自身が楽しむことの他に知名度や認知度を高めることやその団体の活性化の役割をもっており、大衆に魅力を伝えられる機会となっている。開催形態としては対面で行う以外にもオンライン配信や、両者を組み合わせたハイブリッドな形でのライブの開催も一般的となり、ライブ開催の選択肢は増えている。ぴあ総研[1]によると、2022年における有料オンラインライブの国内市場規模は466億円程であり、ライブの実施方法の一形態として定着している。

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、日程的に都合が合わ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、オンラインライブは周知の役割が大きく、より多くの人に見てもらふことによる新規の客の開拓を行い、活性化することを目的としている。こうした配信においてはリアルタイムであることが重要であり、Swarbrickら[2]は、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を明らかにしている。

オンラインでのライブを手軽に配信できるプラットフォームとして、XやInstagram、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。しかしこれらによるオンライン配信において、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SNSにおいて最も利用されているスマートフォン[3]の画面が小さいことが挙げられる。畑田ら[4]は、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

和傾向があることを明らかにしており、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ふ。また別の要因としては、軽音楽の部活やサークルでは撮影機材が乏しく、一視点からの映像となることが多いため、映像に変化が無く単調なものとなり、臨場感が欠如してしまうことが挙げられる。オンライン配信ライブの臨場感の向上を目的とした研究として、HMD (Head Mounted Display) と小型脳波レコーダーを組み合わせることで観客の脳の状態を共有するもの[5]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[6]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。そのため映像処理のみで臨場感を向上させることで、より多くのユーザに視聴してもらう必要があると考えられる。

ここで、音楽にまつわるコンテンツの臨場感を増加させているものに、漫画やアニメでの特殊な表現がある。音楽に関する漫画・アニメでは、実際にその楽器を用いて演奏している音が聞こえるわけではないため、演奏の臨場感を伝えるには工夫が必要である。具体的には、演奏に合わせて場面を切り替え、演奏が激しいシーンなどで効果線やオノマトペの付与、画面分割を施すなどして、臨場感を増加させている。このような音楽の表現は、演奏の臨場感を伝えるという点において、軽音楽におけるライブ映像配信に対しても適用することが可能なのではないかと考えた。

そこで本研究では、オンラインライブの映像に表現を自動付与することで、より臨場感のある映像を配信可能とする手法を提案する。本研究では、初期検討として撮影したライブ動画に自動で映像表現を付与する手法を模索する。また、映像表現にはズームによる画面切り替え、画面分割、集中線を採用し、これらの映像表現を適用することがライブ映像の印象にどう影響するのかを調査する。

<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

Angel ら[5]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。また 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、 $\alpha$ 波と $\beta$ 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。一方柳沢ら[6]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。これらの研究では、特有のデバイスを用いることで他にない体験をユーザに提供している。しかし、そのデバイスを用意できるユーザは比較的に少なく汎用性に欠ける。本研究で提案する手法は映像処理での臨場感の向上を図るため、特有のデバイスを用意する必要がなく、より多くのユーザに使用してもらえることが期待できる。

映像や音声を処理することによる臨場感向上について、米澤ら[7]は、視聴者が楽器演奏の生配信カメラの遠隔制御を可能とするシステムを提案している。視聴者はコメントにコマンドを打つことでカメラ切り替え、ズームイン処理、パン処理、回転処理を行うことができる。提案手法によって視聴者と配信者間、および視聴者間でこれまでにない非言語コミュニケーションを創発しており、ライブ配信の演出効果を高め臨場感のある映像を実現している。また大石ら[8]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。

本稿での提案手法では、映像データと音声データから処理するタイミングを求め、さらに複数の映像表現を組み合わせることでより臨場感を向上させる。

### 2.2 特徴的なシーンの自動検出に関する研究

Ido ら[9]は、手持ちカメラ (スマートフォンやポケットサイズのビデオカメラなど) での複数視点からの同時撮影におけるカットシーンの自動推定手法を提案している。各シーンにおけるカメラの注意の対象をランク付けしカットするフレームを決め、各カメラをノードとして様々な編集パターンを計算することで、最適なカットとカメラ切り替えを行う。また佃ら[10]は、再生時刻に沿って付与されたコメントを用いることで、映像に登場する人物が視聴者の注目を集めているシーンと、それぞれのシーンにおける各登

場人物の注目度合いを推定する手法を提案している。登場人物の名前を含むコメント数とその周辺の感情を含むコメント数から注目度合いを求めることで、比較的高い精度で各登場人物が注目されているシーンを特定することを可能としている。

ライブ映像の中の特徴的なシーンを自動検出するため、Prarthana ら[11]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。Mikko ら[12]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。

これらの研究では、映像データや音声データを解析することで特徴的なシーンを自動検出している。本研究でも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表現を適用することで臨場感の向上を図る。

### 2.3 映像における印象評価に関する研究

映像の印象評価について、劉ら[13]は、市販の映像作品の一部を用いて音楽と映像それぞれの印象と調和感の連続測定実験を行い、調和感が形成される心的過程についての検討を行った。類似した印象の音楽と映像を組み合わせると高い調和感が形成され、異なる作品の音楽と映像を組み合わせても、音楽と映像の印象は類似しないことを述べている。また藤田ら[14][15]は、ディゾルブ、ワイプ、カット、ズームの4種類のトランジションが視聴者の印象にどう影響するかを比較実験している。実験では、4種類の映像素材、4種類のフレームサイズの映像を用意し、それぞれを組み合わせた映像の印象を評価してもらった結果を因子分析したうえで、3つの因子「好感度」「インパクト度」「明晰度」を抽出している。その結果、同じ映像素材においても異なるトランジションを使用することで視聴者の印象が変わることを明らかにした。また、撮影技法による印象評価に関しても調査を行っており、6種類のカメラムーブ(ズームアップとズームバック、左パンと右パン、ティルトアップとティルトダウン)と4種類の情景カテゴリー(図形、自然、人口、人間)を組み合わせた映像を印象評価してもらっている。その結果、同一の映像素材を映した場合でもカメラムーブメントの違いによって与える印象が変化することを明らかにした。

本研究ではライブ映像に適用する映像表現としてズームによる画面切り替え、画面分割、集中線を採用し、それらが臨場感、視聴時間にどう影響するかを調べる。



図1：音楽に関する漫画・アニメで使われる映像表現

©小玉ユキ「坂道のアポロン」、©はまじあき「ぼっち・ざ・ろっく!」、©阿久井真「青のオーケストラ」



図2：映像表現を適用したライブ映像

### 3. ライブ映像への映像表現付与

#### 3.1 提案手法

オンライン配信でのライブ映像において、単一方向から撮影された映像では臨場感を伝えることが難しい。そこで、1章でも述べた通り、ライブ映像に音楽に関する漫画・アニメで使用されているような映像表現を自動付与する手法を提案する。ここで、ライブ映像へ映像表現を付与する際に、どのタイミングでどのような映像表現を施すかが重要となる。音楽に関する漫画・アニメでは、演奏のテンポが合ったシーンや演奏の激しいシーンなどで画面を切り替え、映像表現を加えることで臨場感をより伝えている。このことから、ライブ映像中の演者の動き、発音タイミングから自動で適切な映像表現を付与する。

#### 3.2 映像表現の種類

音楽に関する漫画・アニメで使われる映像表現として、ズーム、画面分割、集中線の付与を採用した(図1)。これらの表現は多くの音楽に関する漫画・アニメで使用されており、ライブ映像に適用しやすいと考えたためである。それぞれの特徴を以下に示す。

- **ズーム**: 主に演奏シーンを通して使用されており、画面切り替えにおいてメインとしての役割をもっている。ここでフォーカスを当てる対象は特徴的な動きをしているキャラクターであり、演奏のテンポに合わせて切り替えられている。
- **画面分割**: 演者の人数分画面を縦分割し、ズームした演者を並べて表示する表現であり、バンドの一体感や演者個々の演奏を伝えるために使用されている。
- **集中線**: 画面全体に対して真ん中に向けた線を引く表

現であり、動きというよりも、バンド全体の勢いや音圧を伝えるために使用されている。

こうしたズームや画面分割、集中線は特に注目してほしいシーンで使用されることが多く、アクセントとしての役割をもっている。ライブ映像に上記の映像表現を適用した結果は図2の通りである。

これらの特徴を考慮して、映像表現を付与する対象・タイミングを決める。ズームでは演者の動きの大きさに着目し、より大きな動きをしている演者にフォーカスを当てる。また、演奏のテンポに合わせてズームをする対象を切り替える。画面分割はバンドの一体感や演者個々の演奏を伝えるために使用されるため、バンド全体として演奏が激しいシーンで付与することでよりこの映像表現を活かすことができると考えられる。そのため検知された演者全員の動作量が一定の値を超えた部分に適用する。集中線は、バンド全体の勢いや音圧を伝える役割をもっているため、より音量の変化量が大きいシーンで適用する。

### 4. 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。まず映像からは人物検出を行って演者を特定し、オブティカルフローを用いることで演者付近の動きを抽出する。これによっ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。また音声データからは発音タイミングを検知することで、演奏に沿った映像表現・タイミングを抽出する。

映像からの演者の検出方法については、まず物体検出ライブラリYOLOv7を用いて映像の中の人物を検出する。ここでライブ映像は、まずは簡単のため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。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)。ここで認識結果が、演奏者が3人未満になっている、演奏者に違うパートが振られているなどの問題もあったため、演奏者3人が正常に検知されたフレームをストックしておき、上手く検知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奏者3人を検知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での演奏者の位置より、現在のフレームで検知されている演奏者がどのパートなのかを求める。検知されなかったパートの位置はストックしていたフレームでの位置とすることで、演奏者全員が上手く検知されていない場合でも、ある程度の演奏者の位置を推定することを可能とした。

次に演奏者の動きについては、OpenCVを用いて特徴点からオプティカルフローを求め、検出した演奏者の位置の周りの特徴点の移動量から演奏者個々の動きを計算した。これらの求めた演奏者の動きはズームによる画面切り替えと画面分割を映像に適用する際に使用する。

発音タイミングについては、音声処理ライブラリであるLibROSAを用いて検出を行った。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基準を設定した。これにより求めた発音タイミングは、ズームによる画面切り替えと集中線を映像に適用する際に使用する。なお、発音タイミングに合わせて、その音量も保存しておくこととした。

次にこれら映像データ、音声データから検出した演奏者の動作量、発音タイミングから、それぞれの映像表現を適用するタイミングを求める。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奏者にフォーカスする。そのため発音タイミングにおいて音量差が小さいときに、一番新しく検知された演奏者の動作量の中で一番動きのある演奏者にズームすることとした。また、画面分割はバンド全体として演奏が激しいシーンで付与したいため、検知された演奏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、集中線はより音量の差が激しいシーンで付与したいため、発音タイミングにおいて音の強度差が強いときに適用することとした。なお、画面分割と集中線は、アクセントとしての役割で使われるため、映像の中で適用回数が多くなりすぎないようにした。

映像表現の付与において、今回3ピースのバンドを対象としたため、画面分割では画面を3分割し、演奏者の立ち位置通り左からベース、ドラム、ギターが映るようにフォーカスを当てている。また、集中線の適用においては、バン

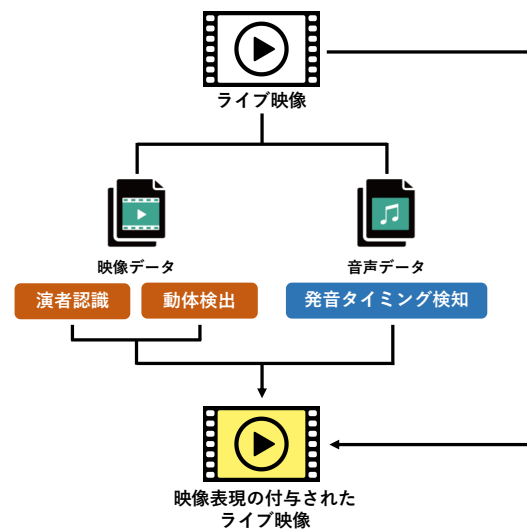


図3: ライブ映像への映像表現付与の処理の流れ

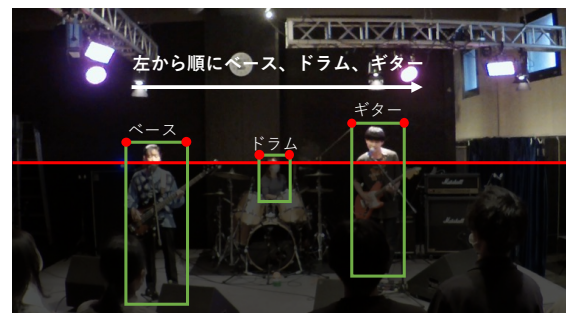


図4: YOLO v7を使った演奏者検知

ド全体の演奏を際立たせたいため、ズームはせずに集中線のみを適用することとした。

ライブ映像に映像表現を付与するうえで、ズームの条件と画面分割・集中線の条件が同時に成立してしまう場合もあり得る。映像を通して適用されるズームはアクセント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。そこで、画面分割や集中線が適用されなくなってしまうことを防ぐため、ズームが適用されている際に画面分割や集中線の適用条件が揃った場合でもこれらの映像表現が適用されるようにした。

## 5. 実験

### 5.1 実験目的

本研究で提案するライブ映像への映像表現付与手法が映像の臨場感を増加させるのに効果的であ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った。具体的には、提案手法により映像表現が付与された動画とそうでない動画をミックスして実験協力者に視聴してもらい、視聴維持率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。



## 5.2 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を46種類用意した。この内、まずチュートリアルとして動画を20本視聴してもらった。また動画を見ながら特に良いと感じた動画にはハート型のいいねボタンを押してもらった。最初の方はライブ映像を新鮮に感じ、長めに映像を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った。その後、本実験としてチュートリアルでは使用していないライブ映像を26本視聴してもらった。このうち、提案手法によって映像表現を付与する対象となる動画は6本(動画1~6)とし、それ以外を何も映像表現を付与していないものとした。

映像表現を付与するうえで、採用した映像表現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、映像表現の適用条件は以下の通りとした。

- None : 映像表現無し
- Zoom : ズーム
- Mix : ズーム+画面分割+集中線

各条件に対して動画1~6からそれぞれ2本ずつ見ってもらうこととし、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割り振った。また提案手法によって映像表現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。

## 5.3 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もらった。実験システム(図5)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デザインにしておき、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つたりしたら、次の動画に進んでもらうように指示をした。そ

の後、実験後アンケートに答えてもらった。アンケートの設問内容を表1に示す。ここでは、映像表現無しの動画と有り(Mix条件)の動画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。なお、臨場感ではなく迫力について質問した理由は、迫力の方が臨場感に比べて回答しやすく、迫力は臨場感と強く関係していると考えたためである。

## 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。

### 6.1 映像表現の適用条件ごとの視聴時間

図6は、映像表現を適用した各ライブ映像における映像表現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。横軸が映像表現の適用条件、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条件が19.8秒、Zoom条件が29.1秒、Mix条件が35.6秒とな

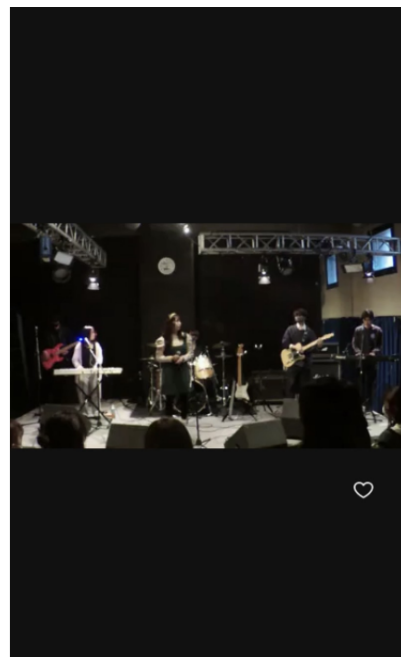


図5：実験システム

表1：アンケート設問内容

設問番号	質問事項	回答形式
Q1	氏名を回答してください	自由記述
Q2	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)
Q3	このライブ映像①(映像表現無し)を視聴してどの程度迫力を感じますか	5段階(1~5)
Q4	このライブ映像②(映像表現有り)を視聴してどの程度迫力を感じますか	5段階(1~5)
Q5	これらのライブ映像①②を視聴して感じるがありましたら教えてください	自由記述

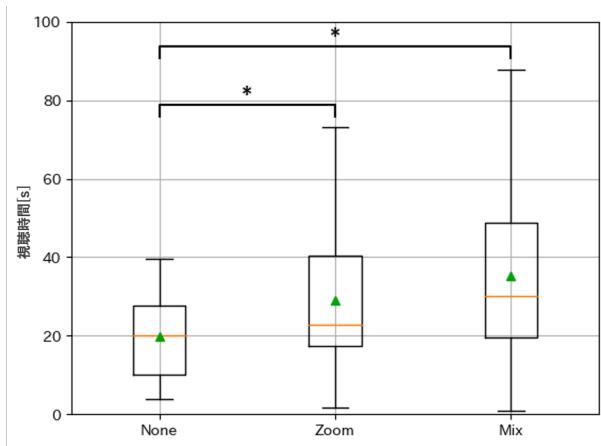


図 6：映像表現ごとの視聴時間

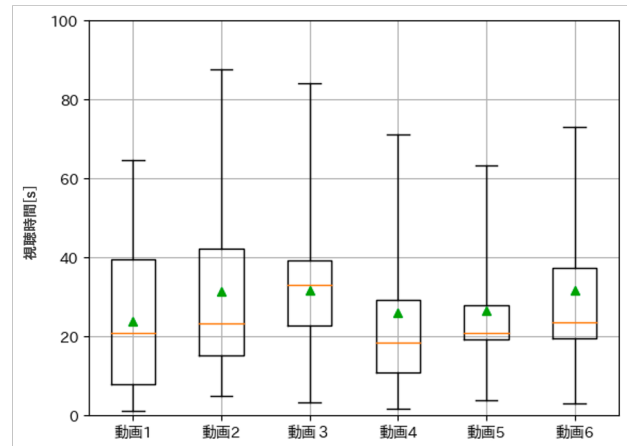


図 7：動画ごとの視聴時間

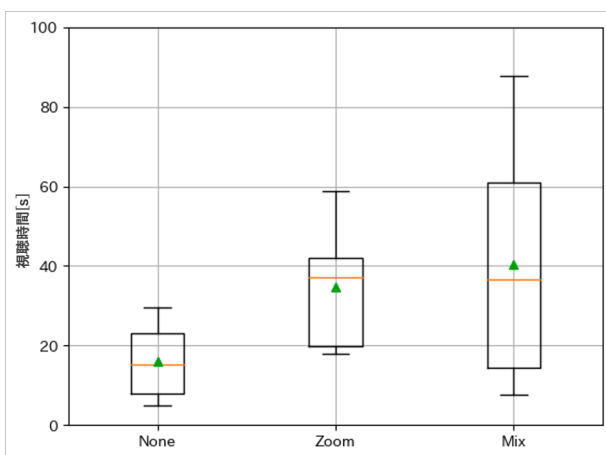


図 8：動画 2 における映像表現ごとの視聴時間

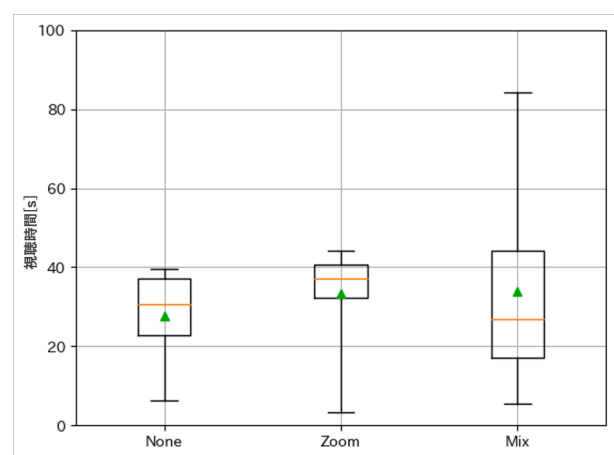


図 9：動画 3 における動画ごとの視聴時間

り、映像表現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、映像表現を付与することでより大きく散らばるようになっていた。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 ( $p < 0.05$ )。

また図 7 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。ここで、図 8、図 9 はそれぞれ動画 2、動画 3 における映像表現の適用条件ごとの視聴時間を比較したものである。動画 2 は映像表現の適用条件間での視聴時間の差が全動画の中で 1 番大きく、None 条件と Zoom 条件間で 18.5 秒ほどの差が見られた。動画 3 は映像表現の適用条件間での視聴時間の差が全動画の中で 1 番小さく、Zoom 条件と Mix 条件間で 0.7 秒ほどであった。

## 6.2 映像表現の適用条件ごとのいいね数

表 2 は、映像表現を適用した各ライブ映像における映像表現条件ごとのいいね数を比較したものである。映像表現

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 個となり、映像表現が増えるごとにいいね数も増えている。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 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 ( $p < 0.05$ )。

## 6.3 アンケート

Q3, Q4 の結果に関しては、映像表現有りの方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表現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。Q5 の映像表現無しのライブ映像①と映像表現有りのライブ映像②を

表 2：映像表現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

視聴してもらいどう感じたかの結果は、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」、「カメラの角度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表現の付与に対しての印象のよし悪しは個人差があることがわかる。

## 7. 考察

映像表現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表現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、付与する映像表現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、映像表現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。ある程度映像表現を適用したら満足度は変わらなくなることも考えられるため、映像表現の適用頻度による影響も今後調査していく予定である。

映像表現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表現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、いいねボタンが押されている動画はきちんと長い時間視聴されていると考えられる。

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

アンケート設問のQ5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、映像表現の付与が過剰であると感じたことや、映像表現が馴染んでいないということが原因であった。今回、適切な映像表現の付与のための動作の変化量や発音タイミングを検証できていなかったため、過剰に映像表現を付与してしまっていることが考えられる。映像表現の適用条件ごとの平均視聴時間においても、映像表現を付与した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表現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ひとに向け、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。採用した映像表現は、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表現の種類を楽曲の曲調を考慮しつつ再度検討するとともに、適用する頻度による影響を調査することで、さらなるライブ映像への印象の向上に繋がると考えられる。

提案手法によるシステムを実運用していく上で、演者認識、動体検出、発音タイミング検知の精度は大きな課題となる。実験で使用した映像ではほぼ全てのフレームにおいて正確に演者認識ができた。しかし今回は3ピースバンドのみを対象としており、それらの映像の中でも演者を認識するのが難しいものもあった。実際に提案手法を使用するうえで、演者が3人以上いる場合や、演者の立ち位置が入れ替わる場合も多いため、演者認識の処理を再検討する必要がある。また発音タイミングに関しては、今回映像に合

わせて発音されたと判断する基準を手動で設定していた。そのため実運用していくうえでは、映像によって自動でその基準を定めなければならない。また、もう一つの課題として映像処理の実行速度の向上が挙げられる。今回実験で使用した映像の処理には元の映像の時間の倍以上かかっており、リアルタイムで配信するにはその処理を大幅に短縮する必要がある。

## 8. まとめ

本研究では、単一方向から撮られたライブ映像の臨場感を増強させるために、漫画・アニメにおける映像表現を自動でライブ映像に適用する手法を提案した。提案手法では、ライブ動画の映像データから演者の動きを、音声データから演奏の発音タイミングを検出し、それらを用いてライブ動画にズームによる画面切り替え、画面分割、集中線といった映像表現を付与した。映像表現を付与したものとそうでないものを比較した結果、提案手法による映像表現の付与は視聴維持率や満足度を高めることを明らかにした。

今回、映像表現を組み合わせることによる効果はあまり得られなかった。これは、今回比較する上で適用頻度の大きな差が少なかったことが原因として考えられる。また、提案手法を適用するのが難しいライブ映像も多くあり、これらでは主に演者認識の不完全さが目立っている。そのため、演者の検出精度を向上させることで汎用的に使えるシステムにすることができると考えられる。一方、提案手法に対して好印象な意見もあった反面、過剰な編集はあまり印象が良くないといった意見も見られた。そのため、今後はどの映像表現をライブ映像に適用すべきか再検討したのち、適用頻度による影響を調査する予定である。また、ライブ映像への自動的な映像表現の付与手法を固めたのちに、リアルタイムで実際に行われているライブに適用して実証していく予定である。

## 参考文献

- [1] ぴあ総研：2022年のオンラインライブ市場は対前年比9.1%減の466億円に縮小，入手先  
([https://corporate.pia.jp/news/detail\\_live\\_enta20230526\\_2.html](https://corporate.pia.jp/news/detail_live_enta20230526_2.html)) (参照 2024-02-18)。
- [2] Swarbrick D, Seibt B, Grinspun N, et al. : Corona concerts: The effect of virtual concert characteristics on social connection and kama muta, *Frontiers Psychol*, Vol. 12, pp. 1-21(2021)
- [3] 総務省：令和4年通信利用動向調査の結果，入手先  
([https://www.soumu.go.jp/johotsusintokei/statistics/data/230529\\_1.pdf](https://www.soumu.go.jp/johotsusintokei/statistics/data/230529_1.pdf)) (参照 2024-02-18)。
- [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] Á. Muñoz-González, S. Kobayashi and R. Horie : 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)
- [6] 柳沢豊, 小野圭介, 上田健太郎, 出田怜, 吉池俊貴, 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イション*，Vol. 28, No. 2, pp. 1-5(2021)
- [7] 米澤拓郎, 徳田英幸：視聴者協力型ライブ演出システムの実装と評価，*情報処理学会論文誌*，Vol. 53, No. 3, pp. 1007-1016(2012)
- [8] 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*，Vol. 121, No. 23, pp. 7-12(2021)
- [9] Ido Arev, Hyun Soo Park, Yaser Sheikh, Jessica Hodgins, and Ariel Shamir: Automatic editing of footage from multiple social cameras, *ACM Transactions on Graphics*, Vol. 33, pp. 1-11(2014)
- [10] 佃洗撰, 中村聡史, 山本岳洋, 田中克己：映像に付与されたコメントを用いた登場人物が注目されるシーンの推定，*情報処理学会論文誌*，Vol. 52, No. 12, pp. 3471-3482(2011)
- [11] Prarthana Shrestha, Peter HN de With, Hans Weda, Mauro Barbieri, and Emile HL Aart: Automatic mashup generation from multiple-camera concert recordings, *Proceedings of the 18th International Conference on Multimedia 2010*, pp. 541-550(2010)
- [12] Roininen, M.J., Leppänen, J., Eronen, A.J. et al.: Modeling the timing of cuts in automatic editing of concert videos, *Multimedia Tools and Applications*, Vol. 76, pp. 6683-6707(2017)
- [13] 劉沙紀, 矢萩徹, 大西英治, 岩宮眞一郎：映像作品における音楽と映像の調和感と印象の連続測定，*音楽知覚認知研究*，Vol. 21, No. 2, pp. 73-86(2015)
- [14] 藤田良治, 山口由衣, 椎名健：映像コンテンツの印象評価における編集技法の影響—トランジションの効果について，*情報メディア研究*，Vol. 7, No. 1, pp. 1-13(2008)
- [15] 藤田良治, 山口由衣, 椎名健：映像コンテンツにおける表現技法の心理的効果：カメラムーブメントの効果について，*Vol. 54, No. 3, pp. 1-8(2007)*