

全方位カメラを用いたインターネット生放送における 視線ヒートマップの評価

高田真也¹ 西岡大¹ 齊藤義仰¹

An Evaluation of the POV Heat Map in Internet Broadcast using an Omnidirectional Camera

MASAYA TAKADA¹ DAI NISHIOKA¹ YOSHIA SAITO¹

1. はじめに

インターネット技術の進歩とともに多くのユーザが YouTube やニコニコ動画, Ustream などのインターネット生放送サービスを利用するようになった。スマートフォンやタブレットなどのモバイル端末が普及したことで、ユーザはいつでもこれらのサービスを利用することができるようになってきている。平成28年度の情報通信白書[1]によると、インターネット利用者のうち、約7割が動画投稿・共有サイトの利用経験があり、そのうちほとんどが今後も利用したいと回答している。インターネット生放送は、放送者と視聴者がコメント等により、リアルタイムなコミュニケーションを取ることが特徴であり、コミュニケーションツールとしても利用されている。

現在、インターネット生放送サービスにおいて全方位カメラの利用に対応する動きが強まっている。全方位カメラとは複数の広角レンズによりカメラの周囲360度の映像を撮影できるカメラのことで、これにより撮影されたパノラマ映像を360度動画と呼ぶ。360度動画はカーソル操作によって視聴方向を自由に変更することができる。動画共有サービス最大手である YouTube では2015年に360度動画のオンデマンド配信に対応し、2016年には360度動画のインターネット生放送サービスへの対応が開始された。YouTube の他にも、Twitter の Periscope[2]や Facebook の Live360[3]などでも対応が進んでおり、今後の爆発的な普及が期待される。

全方位カメラを用いたインターネット生放送では、これまでの通常の Web カメラやスマートフォンのカメラを使用した動画配信の方法とは異なり、視聴者の視線を放送者が察知できないという問題が発生する。従来の Web カメラを用いた場合は単一のレンズが矩形の撮影範囲を明示的に示しており、視聴者は必然的にそのレンズが向けられた撮影範囲内を視聴していた。それに対し、全方位カメラを用いた場合は、複数の広角レンズで同時に撮影しているため、視聴者がどのレンズによって撮影されている範囲を見ているのか、さらにはその撮影範囲の内どこを見ているのかが分からない。コミュニケーションにおける相手の視線情報は話題の中心や相手の興味を示すことが明らかにされており[4][5][6]、視聴者がどこを見ているのかを察知できない

場合、放送者と視聴者のコミュニケーションが円滑に行えない可能性がある。そこで我々は、全方位カメラを用いたインターネット生放送における、放送者と視聴者の円滑なコミュニケーションを実現することを目的に、視聴者の視線情報を放送者に対して可視化し提示する視線ヒートマップを実装した[7][8]。本稿では、視線ヒートマップの評価結果について報告する。

2. 関連サービス

これまでも、コミュニケーションにおけるノンバーバル情報の重要性に関して様々な研究が進められてきている。特に、コミュニケーションにおける視線情報はお互いの意図を伝える重要な役割を担っているとしており、相手が何に対し興味を持っているのかを知るためには視線情報が必要である。全方位コンテンツにおける視聴方向の分析については、YouTube で公開されているクリエイターツールの1つである360°ヒートマップ[9]やジョリーグッドが開発している VRCHL[10]などでも行われている。

360°ヒートマップは YouTube に投稿したヒートマップを用いて360度動画を解析することが出来る機能である。YouTube が提供するアナリティクスに追加された機能で、2017年6月にサービスが開始された。このツールでは、動画を視聴したユーザがどの方向に視線を向けていたのかを時間経過とともに確認することが出来る。YouTube のクリエイターズブログでは[11]、魅力ある360度動画の作成には、撮影範囲全体を利用した映像コンテンツの作成が必要だと述べており、そのためには360°ヒートマップのような分析手法が役に立つとしている。

VRCHL は株式会社ジョリーグッドが開発している360°VRのコンテンツ解析とユーザ行動のパターン解析を人工知能によって行うという取り組みである。視聴中の360度動画の人物や物体、話されている言語、雰囲気、シーン構成、ストーリーなどの様々な動画の構成要素をリアルタイムに解析し動画にオーバーレイ表示できる。

しかし、これらの分析手法はあくまでもオンデマンドの360度動画を分析対象としており、インターネット生放送における利用は想定されていない。そこで、インターネット生放送でも利用可能な視聴者の視聴方向を分析し、放送者に提示するシステムが必要になる。

¹ 岩手県立大学大学院 ソフトウェア情報学研究科
Iwate Prefectural University,
Graduate School of Software and Information Science

3. 全方位カメラと360度動画

ここでは、本研究で対象とした全方位カメラおよび360度動画について説明する。特に、本研究で使用した全方位カメラとインターネット生放送環境について詳しく述べる

3.1 全方位カメラ

全方位カメラはカメラを中心とした周囲360度全方向の映像を撮影することができるカメラのことである。撮影には非常に画角の広い広角レンズを用いる方法と通常のカメラを複数用いる方法がある。撮影された映像は付属しているソフトウェアなどを用いることで複数のカメラで撮影された映像の合成や歪みの補正を行い、360度動画として視聴することが可能である。

現在、全方位カメラは様々な製品が登場している。本研究ではYouTubeのインターネット生放送サービスで、公式がサポートしているRICOHのTheta S[12]を用いた。Theta Sはdouble fish eyesと呼ばれる2つの魚眼レンズで撮影した魚眼映像が横に並んだ映像が撮影可能である。この映像は1280×720(pixel)のHDサイズになっており、直径640ピクセルの魚眼映像1つで180度強の映像をカバーする。また、マイクも内蔵されているため、インターネット生放送での利用時はTheta Sを中心とした360度映像および音声を利用することが可能である。

3.2 360度動画

360度動画とは、動画プレイヤー上でマウスによるドラッグ操作や指によるフリック操作により視聴方向を自由に変更することが可能な動画のことである。スマートフォンやタブレット端末などでは、ジャイロセンサを用いることで端末の傾きを検知し、動画の視線を端末の動きに合わせて変化させることが可能である。OculusRift[13]やGoogle Cardboard[14]のようなヘッドマウントディスプレイを利用する場合は顔を向けた方向をヘッドマウントディスプレイやスマートフォン側でセンシングし、視聴方向の変更を行う。映像ソースの入力はdouble fish eyes形式の他にも、エクイレクタングラー形式の映像によっても行われる。エクイレクタングラー形式とは世界地図のメルカトル図法のように緯線と経線が直交するように平面に映像を展開した形式である。

パノラマコンテンツは古くから認知されているコンテンツの1つだが、ハードウェアやネットワークの発達により、普及してきている。2012年に発表されたOculusRiftの発売から、HMDやパノラマコンテンツには関心が集まっていたものの、コンテンツの少なさから普及には時間を必要としていた。しかし、2015年にYouTubeやFacebookなどのスマートフォンアプリでも利用できるようになったことで、様々なデジタルコンテンツへの対応が見込まれる。

360度動画に関する研究において、その活用法を模索する研究は近年多く進められている[15][16]。主に、カメラフレームの枠を超えた様々な表現の可能性や臨場感向上について論じられたものが多く、コミュニケーションにおける利用に関する研究は少ない。特に、全方位コンテンツにおける視聴方向変更機能とコミュニケーションにおける視線の役割に着目した研究は今のところ報告されておらず、今後のコミュニケーション分野の研究において必要になると考える。

3.3 全方位カメラを用いたインターネット生放送環境

全方位カメラを用いたインターネット生放送サービスは既にYouTubeなどでも提供されている。しかし、システム実装と評価のためには、双方向のコミュニケーションが行える低遅延であることと、機能追加が行えることが必要だったため、全方位カメラを用いたインターネット生放送システムを我々が構築、使用することとした。

本研究では、映像と音声の取り扱いに関して、従来のFlash Playerを利用した動画配信の仕組みを使用しなかった。2016年にGoggle社がChromeブラウザにおけるFlash Playerサポート廃止を発表したこと[17]を受けて、動画配信・共有サービスの最大手であるYouTubeがFlash PlayerからHTML5への乗り換えを発表している[18]。ニコニコ動画に関しても、2016年10月27日より、一部ユーザから順次HTML5版と題したHTML5対応のプレイヤーの利用が進められている[19]。動画配信・共有サービスで進められているHTML5への対応に合わせ、本研究にでもHTML5で再生可能なWebRTC技術を採用し、WebRTCの動画配信に対応したKutrento Media Server[20]を利用した実装を行った。

また、全方位カメラを用いたインターネット生放送の実装には入力映像を360度映像へ加工する必要がある。この点に関しては、Theta_GL[21]というライブラリを使用した。THETA_GLはインフォコム社が公開しているライブラリで、THETA Sで撮影されたdouble fish eyes映像などをWebGLの3D空間における球体表面にマッピングすることで、360度動画としてWebブラウザ上で再生することができる。再生する映像のソースをWebRTCサーバから取得することでライブストリーミングにも対応することができる。

Theta_GLでは、WebGLの3D空間上に球体オブジェクトとカメラを設定している。球体の内側表面には360度動画の映像がマッピングされており、球体の中心点にカメラ位置が設定されている。カメラの方向は球体表面に緯度経度のように設定された(x, y)の座標によって表現されており、球表面とカメラとの距離を示すz座標を付加した(x, y, z)のベクトル情報として管理される。本研究では、このベクトル情報を視聴者の視線情報として収集し、利用することにした。

4. 提案システム

ここでは、視聴者の視線情報を放送者に可視化して提示する視線ヒートマップについて述べる。また、本研究におけるユースケースと期待される効果についても述べる。

4.1 ユースケース

本研究において、全方位カメラを用いたインターネット生放送サービス利用のユースケースを、従来のインターネット生放送サービスの利用動向調査の結果[22]を基に定義した。調査では、インターネット生放送サービスの視聴および放送、両方の経験がある500人に対して放送内容に関するアンケートを行った。

その結果、放送の内容では、視聴者参加型放送、動物、エンタテイメント、音楽などが上位に挙げられた。視聴者参加型放送は、1人の放送者が多数の視聴者とのコミュニケーションを行う放送である。本研究ではコミュニケーションツールとして利用されている視聴者参加型の放送を対

象とする。また、調査結果から個人ユーザによって行われている放送では1回の放送における視聴者数が100人を超えるものがごく少数であることが分かっている。よって、視聴者は数十人程度の不特定ユーザとした。放送者については、個人ユーザによる利用を想定しているため、1～3人程度のごく少数であるとし、放送者の前方中心に全方位カメラを固定するものとする。

4.2 視線ヒートマップ

全方位カメラを用いたインターネット生放送における視聴者の視線情報を放送者に可視化し提示する、視線ヒートマップについて説明する。視線ヒートマップは、インターネット生放送サービスと同時に利用するシステムで、視聴者が放送において視聴している方向を集計し、全方位カメラを中心とした仮想の球体表面にヒートマップとして表示することで、視線集中の程度を可視化する。視線ヒートマップのモデル図を図1に示す。

球体状のヒートマップは非常に新しい概念で、既存の手法にはない利点を持っている。球体状ヒートマップを用いることで3次元空間上における特定の座標を中心とした複数のベクトルに重み付けを行いながら可視化することができる。入力データとして視聴者の視線情報、すなわち視聴方向のベクトル情報を与えることにより、全方位カメラを中心とした視聴者の視線情報を可視化することができる。

本研究において視線ヒートマップはAR機能を用いて実装を行うことにした。図2に実装した視線ヒートマップの画面を示す。この理由として、視線ヒートマップのキャリブレーションの容易さが挙げられる。前述の通り、視聴者が視聴する映像はWebGL空間上の仮想球体にマッピングされたものであり、球体状ヒートマップに入力されるベクトル情報もWebGL空間における座標を元に取得される。しかし、放送者が視線ヒートマップを確認する際にWebGL空間上の座標が現実空間の座標とどのように対応するのか分からなければ、視線ヒートマップによって可視化された視線が実際に何に向けられているのかを把握することができない。そこでAR技術を用いることで、WebGL空間という仮想の空間を現実を重ね合わせてキャリブレーションを行うことにした。具体的には球体状のヒートマップと基準方向を指し示す矢印をARオブジェクトとして表現する。ARマーカーを動かす、矢印が全方位カメラの基準方向に向くように調整することでキャリブレーションが完了する。

また、キャリブレーションを行った座標の同期状態を保持するために、ARマーカーの移動が行えなくなる。ARマーカーを動かさない場合、その特性上ARオブジェクトの背面方向を確認することが難しくなってしまうため、ARオブジェクトの回転機能と回転リセット機能も実装した。

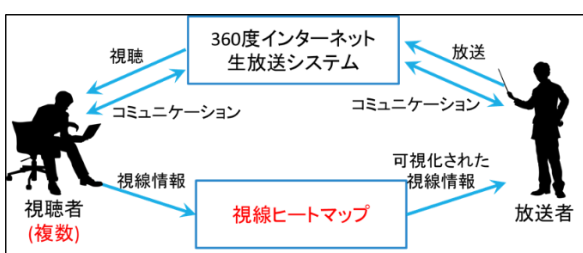


図1：視線ヒートマップのモデル図

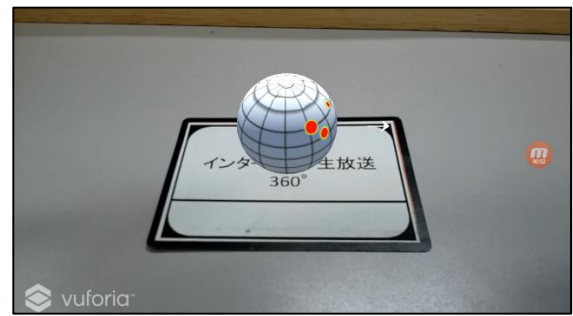


図2：視線ヒートマップの表示例

4.3 期待される効果

放送者は全方位カメラの撮影方向に対応付けられたヒートマップを見ることで視聴者の多くがどこに関心を示しているのか、集中した視線と異なる視線を向けている視聴者がどこに関心を示しているのかを直観的に理解することができる。視聴者から送信されるコメント情報にこの視線情報を付加して考えることで、指示代名詞を含んでいたり省略が行われていたりするコメントであっても、その発言意図や話題の中心となっているものが何なのかを放送者は把握することができるようになる。さらにコメントを送信せずに視聴しているコミュニケーションに対して消極的な視聴者の視線を可視化することにより、それらのユーザを巻き込んだ放送を行うことも可能になる。

5. 評価

実装した視線ヒートマップについて、放送者と視聴者の間で行われるコミュニケーションの円滑化を行うことができるかを評価するために実験を行った。ここでは、実施した評価実験の目的と概要、その結果について述べる。

5.1 評価実験の目的

評価実験の目的は、視線ヒートマップを用いた視線情報提示についての有効性評価である。期待される具体的な結果は、視聴者が送信したコメントにおいて指示代名詞が用いられた場合や省略が行われた場合にコメントの補完と理解にかかる時間の短縮と正確性の向上、視聴者に対して行われる注目指示などの視線を用いたコミュニケーションにおける円滑化である。これらを確認するため、評価では放送者と視聴者の間で行われるコミュニケーションにおける応答や次の発言までの時間の計測や、コミュニケーション内容の分析、アンケートによるコメント伝達精度に関する主観評価を行うことにした。

5.2 評価実験の概要

評価実験は岩手県立大学コミュニケーション学講座研究室にて、2017年8月2日、2017年10月23日、2017年11月16日、2017年12月21日の計4回実施した。評価実験の協力者は1回あたり6名で、放送者1名、視聴者5名に分けて実装環境での放送を実施し、放送内で行われたコメントや視線情報に関するアンケートを行った。放送時間は19分で、放送内容はあらかじめ室内に用意した設置物に関する視聴者との雑談である。

用意した設置物は「ノートPC、ぬいぐるみ、ゲーム機、ラケット」の4種類について色や用途などの条件を変えたものを2つずつ用意し、それぞれ2つの設置物が同時に放送者の視界に入らないように配置した。4種類の設置物と

放送者の位置関係を図4に示す。この他にも、「プログラミング教本、ヘッドマウントディスプレイ、プロジェクター、タブレットPC、ライトスタンド、音楽CD」の6種類を設置した。これらの設置物と放送者の位置関係を図5に示す。

評価実験では、放送者と視聴者の間で行われるコミュニケーションの内容を分析するために、放送者から視聴者へのコミュニケーションと視聴者から放送者へのコミュニケーションを切り分けてそれぞれに対する評価項目を用意した。評価項目とその評価方法の一覧を表1に示す。

放送者から視聴者へのコミュニケーションの分析では、室内に用意した6種類の設置物のうち指定した1つの設置物に対しての注目指示を放送者から視聴者に対して行い、視聴者が全員注目したと感じた時点で注目を指示した設置物についての雑談を行ってもらうというタスクを用意した。このとき、放送者が話し始めるタイミングがどの程度早かったのか、または遅かったのかを、注目指示から次に話し始めるまでの時間計測と放送者と視聴者それぞれに対してアンケートによる主観評価を実施することで評価した。

視聴者から放送者へのコミュニケーションの分析では、2つずつ用意した設置物4種類に関するコメントを指定した視聴者に指定した文言で行い、放送者に対する正しい応答を行うまでの時間計測とコメントがどの程度正しく伝わった、または理解できたと感じたかについて放送者と視聴者それぞれにアンケートで主観評価を実施することで評価した。

また、アンケートを用いて視線ヒートマップを用いた放送と用いない放送のどちらがコミュニケーションをとりやすいと感じたか、視線情報を用いたコミュニケーションに関する自由な感想を聞いた。

放送ははじめに1分の放送操作確認時間を用意し、その後視聴者から放送者へのコミュニケーションとして、指示代名詞を含むコメントを3分おきに4回送信し、放送者から視聴者へのコミュニケーションとして、注目指示を2分おきに3回実施する、計19分の放送を実施する。

視聴者から放送者へのコミュニケーションとして、「このゲーム機では何が遊べるの?」というように、設置物の種類と指示代名詞を含んだコメントを視聴者が送信する。このとき、送信するコメントは手順書において指定しており、送信タイミングも口頭で「指定コメントの送信をお願いします」と指示を行った。放送者は送信されたコメントを元に室内の設置物のうち、どれに対してのコメントなのかを判断し、視聴者へと応答する。視聴者が何を視聴しながらコメントしているのかを把握できている場合、応答までにかかる時間は短くなることが予想される。

表1: 評価項目と評価方法

評価対象	評価項目	評価方法	対応する仮説
視聴者から放送者	視聴者のコメントから応答までの時間	ログを元に計測	視線情報を可視化することで視聴者のコメントを理解しやすくなる
	視聴者のコメントへの理解度	アンケート(主観評価)	
放送者から視聴者	放送者の指示から次の発言までの時間	ログを元に計測	視聴者の視線情報を可視化することで視聴者の反応を容易に確認できる
	発言タイミングがどの程度遅れたか	アンケート(主観評価)	
放送の満足度	放送に対する視聴者の満足度	アンケート(主観評価)	視線情報によるコミュニケーション支援で放送に対する満足度を向上できる

放送者から視聴者へのコミュニケーションでは、放送者に指定された設置物を指差してもらい、視聴者に「皆さん、これを見てください」と指示を出してもらおう。このとき、放送者には「視聴者の視線が集まったタイミングで注目を指示した設置物に関する雑談を行ってください」と指示をしている。放送者が視聴者の視線方向の動きを把握できている場合、視聴者の視線が集まり終わると同時に雑談を開始できることが予想される。

実験では、この放送手順について視線ヒートマップを用いた放送と用いない放送のそれぞれを1回ずつ、計2回実施する。また、2回目の放送では先に行った放送により放送の進行や視聴者の発言の理解に一定レベルでの慣れが発生し、放送におけるコミュニケーションが視線情報の有無に関係なく円滑に行われる可能性があるため、視線ヒートマップを用いた放送を先に行う実験と後に行う実験の2パターンを実施することにした。最初の2回が視線ヒートマップを用いた放送を後に行ったものであり、後の2回が視線ヒートマップを用いた放送を先に行ったものである。

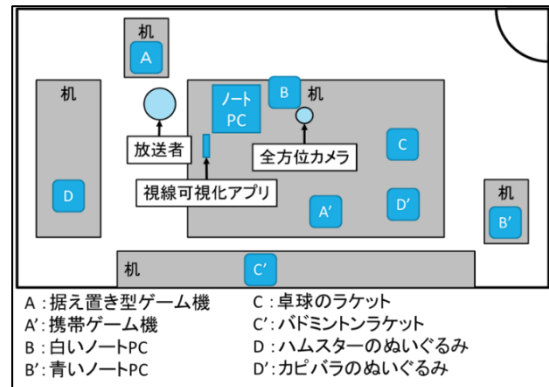


図4: 視聴者から放送者へのタスクで用いる設置物の配置

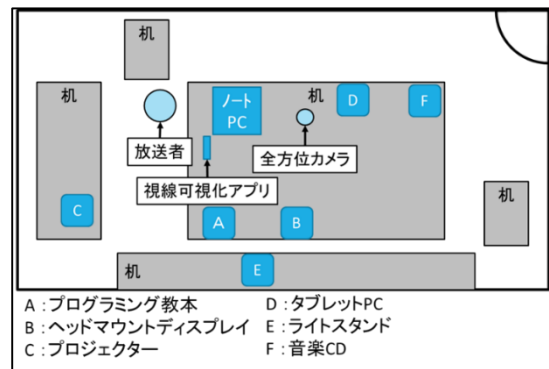


図5: 放送者から視聴者へのタスクで用いる設置物の配置

5.3 評価実験の結果

ここでは評価実験の結果について、放送者から視聴者へのコミュニケーションの結果、視聴者から放送者へのコミュニケーションの結果、2回の放送の比較結果の順で結果を述べる。

5.3.1 放送者から視聴者へのコミュニケーションの結果

放送者から視聴者へのコミュニケーションでは、注目指示から次の発言までの時間を計測するとともに、放送者には発言タイミングが掴めたかについて5段階のアンケート評価を、視聴者には放送者の発言タイミングが早いまたは遅いと感じたかについて「早い(1点)、少し早い(2点)、適

切(3点), 少し遅い(4点), 遅い(5点)」の5段階によるアンケート評価を行ってもらった。時間計測の結果を表2に、放送者に対するアンケートの結果を表3に、視聴者に対するアンケートの結果を表4に示す。

時間の計測結果について、視線ヒートマップを用いた放送と用いていない放送を t 検定によって検定したところ、計測時間に有意な差が見られた。放送者が注目指示を行ってから次に発言するまでの待機時間について、視聴者の注目完了までにかかる時間よりも短い場合は視聴者が注目する前に話を始めてしまい、話題についていけなくなる視聴者が発生する可能性がある。逆に時間が長くなってしまうと、視聴者が注目してから無駄な待機時間が発生してしまう。視線ヒートマップを用いて注目状況を確認しながら待機した放送では待機時間が平均 16.67 秒であるのに対し、視線ヒートマップを用いていない放送では平均 23.167 秒と約 7 秒の無駄な待機時間が発生してしまっていることが確認された。

放送者へのアンケートの結果から、視線ヒートマップを用いた放送のほうが、どの放送者についても発言タイミングが掴みやすかったと回答しており、放送者への支援は十分に行えていることが確認できた。一方で、視聴者に対するアンケートの結果からは約 7 秒の無駄な待機時間が発生していたにも関わらず、視線ヒートマップの有無による有意な差は見られなかった。視聴者へのアンケートの分析では最も評価点の高い 3(適切と感じた)に対する絶対値を求めて分析を行った。

これらの結果から、視線ヒートマップによるコミュニケーション支援は視聴者に対する影響が少ないことと、放送者と視聴者の間には時間に対する価値観の違いがあることが考えられる。

表 2 : 待機時間の計測結果

対象	対象物	実験1	実験2	実験3	実験4	対象平均	全体平均
視線ヒートマップ無し	設置物A	33秒	35秒	14秒	10秒	23秒	23.167秒
	設置物B	26秒	22秒	36秒	22秒	26.5秒	
	設置物C	34秒	16秒	12秒	18秒	20秒	
視線ヒートマップ有り	設置物A	10秒	18秒	24秒	22秒	18.5秒	
	設置物B	23秒	15秒	20秒	20秒	19.5秒	
	設置物C	12秒	11秒	13秒	12秒	12秒	

表 3 : 放送者に対する発言タイミングのアンケート結果

対象	実験1	実験2	実験3	実験4
視線ヒートマップ無し	掴めなかった(1)	どちらともいえない(3)	掴めなかった(1)	少し掴めた(4)
視線ヒートマップ有り	少し掴めた(4)	掴めた(5)	少し掴めた(4)	掴めた(5)

表 4 : 視聴者に対する発言タイミングのアンケート結果

対象	実験1	実験2	実験3	実験4	全体平均
視線ヒートマップ無し	3.2 (+0.2)	3.4 (+0.4)	2.6 (-0.4)	2.8 (-0.2)	3.0
視線ヒートマップ有り	3.4 (+0.4)	2.8 (-0.2)	2.6 (-0.4)	3.0 (±0)	2.95
絶対値差分	-0.2	+0.2	±0	+0.2	+0.05

5.3.2 視聴者から放送者へのコミュニケーションの結果

視聴者から放送者へのコミュニケーションでは、放送者が視聴者からの指定コメントを受けてからコメントに対しての正しい応答を行うまでの時間を計測するとともに、放送者には対象のコメントがどの程度理解できたかについて5段階のアンケート評価を、視聴者には対象のコメントがどの程度放送者に伝わったと感じたかについて5段階によるアンケート評価を行ってもらった。時間計測の結果を表5に、放送者に対するアンケートの結果を表6に、視聴者に対するアンケートの結果を表7に示す。

コメントから応答までの時間が長いと、視聴者のコメントに対する応答が遅れてしまい、放送に対する視聴者の満足度低下に繋がる。逆に時間が短い場合は視聴者に対する応答が早くなり、放送に対する視聴者の満足度が向上すると考えられる。今回の実験では予想に反して視線ヒートマップを用いた放送での計測時間は視線ヒートマップを用いていない放送に比べて 10 秒程度の増加が見られた。

放送者に対するアンケートの結果では伝達精度に向上が見られ、視線ヒートマップを用いた放送の方が正しくコメントの意味が理解できていると考えられる。一方で視聴者に対するアンケートの結果では視線ヒートマップを用いた放送のほうが評価点は高いものの、t 検定による有意差は確認できなかった。この結果は放送者から視聴者への評価と同様であり、視線ヒートマップによる影響は視聴者に対しては低いことが考えられる。

これらの結果から、コメントの伝達精度を向上させる効果は確認できたものの、視線ヒートマップの確認が放送者に対して負担になってしまっていることが考えられる。視線ヒートマップのインタフェースをより直感的で分かりやすいものに改良することで、確認にかかる時間を短縮することができればこの点についても解決可能であると考えられる。視線ヒートマップの確認にかかる時間の短縮が行えれば、視聴者のコメントの伝達精度とレスポンス速度、両者に対する向上が可能であると考えられる。

表 5 : 正確な応答までの時間の計測結果

対象	対象物	実験1	実験2	実験3	実験4	対象平均	放送平均
視線ヒートマップ無し	設置物A	21秒	22秒	6秒	24秒	18.25秒	20.875秒
	設置物C	9秒	39秒	13秒	26秒	21.75秒	
	設置物D	43秒	33秒	12秒	19秒	26.75秒	
	設置物B	4秒	24秒	19秒	20秒	16.75秒	
視線ヒートマップ有り	設置物B'	45秒	39秒	46秒	38秒	42秒	30.875秒
	設置物A'	29秒	26秒	29秒	16秒	25秒	
	設置物C	35秒	32秒	24秒	32秒	30.75秒	
	設置物D'	29秒	33秒	13秒	28秒	25.75秒	

表 6 : 放送者に対するコメント理解度のアンケート結果

対象	対象実験	設置物A	設置物B	設置物C	設置物D	実験平均	全体平均
視線ヒートマップ無し	実験1	1	2	1	3	1.75	2.8125
	実験2	3	3	3	2	2.75	
	実験3	1	2	3	2	2.0	
	実験4	4	5	5	5	4.75	
視線ヒートマップ有り	実験1	1	4	4	3	3.0	3.5625
	実験2	5	3	2	3	3.25	
	実験3	4	4	5	3	4.0	
	実験4	5	4	4	3	4.0	

表 7：視聴者に対するコメント伝達度のアンケート結果

対象	対象 実験	設置物A 平均	設置物B 平均	設置物C 平均	設置物D 平均	実験 平均	全体 平均
視線 ヒートマップ 無し	実験1	3.8	3.4	4.0	3	3.55	3.6625
	実験2	3.2	4.2	3.8	4.2	3.85	
	実験3	3.2	4.0	3.6	4.0	3.7	
	実験4	3.0	3.6	3.8	3.8	3.55	
視線 ヒートマップ 有り	実験1	3.8	4.4	4.0	3.6	3.95	3.8875
	実験2	4.2	3.8	4.4	4.6	4.25	
	実験3	3.4	3.8	4.8	3.2	3.8	
	実験4	2.4	5.0	3.4	3.4	3.55	

5.3.3 実施した 2 回の放送の比較

放送者には、視線ヒートマップを用いた放送と用いてない放送どちらの放送の方が視聴者とコミュニケーションを取りやすかったかを「視線ヒートマップ有りの放送、どちらかと言えば視線ヒートマップ有りの放送、どちらともいえない、どちらかと言えば視線ヒートマップ無しの放送、視線ヒートマップ無しの放送」の 5 段階で評価してもらった。放送者の主観評価の結果を表 8 に示す。視聴者に対しては 2 回の放送のうちどちらの放送のほうがコミュニケーションを行いやすいと感じたかを「1 回目の放送, 2 回目の放送, どちらともいえない」の 3 択で評価してもらった。視聴者評価を表 9 に示す。視聴者の主観評価結果において、視線ヒートマップの有無による差はあまり見られなかった。しかし、放送者の主観評価結果から、視線ヒートマップを用いた放送の方がコミュニケーションの取りやすさについて、評価点が高かった。評価の理由について各放送者に尋ねたところ、「注目指示に対する反応を見る際などに有用であったが、普段からずっと見続けるのは負担が大きいと感じた」という意見が多かった。これらの結果から、視線ヒートマップの有無による影響は放送者に対しては大きい、視聴者に対しては少ないということが分かった。また、視線ヒートマップによってコミュニケーションの円滑化を行うことができるということも分かった。ただし、今後は視線ヒートマップの確認方法について、負担を軽減する方法を検討し、インタフェースの改良とともに実装と評価を行っていく必要がある。

表 8：放送者に対する 2 回の放送の主観による比較結果

対象	実験1	実験2	実験3	実験4
放送者	どちらかと 言えば有り (4)	どちらかと 言えば有り (4)	どちらとも 言えない (3)	どちらかと 言えば有り (4)

表 9：視聴者に対する 2 回の放送の主観による比較結果

対象	実験1	実験2	実験3	実験4	全体
視線ヒートマップ 無し	1人	2人	2人	2人	7人
視線ヒートマップ 有り	4人	2人	2人	0人	8人
どちらとも いえない	0人	1人	1人	3人	5人

6. 考察

評価実験の結果、注目指示後の待機時間については短縮が見られた。また、アンケートを用いた主観評価においても放送者が発言タイミングを掴みやすくなっていることが分かる。これは、視線ヒートマップを用いることにより、各視聴者の視線が注目指示を行った設置物に向けられたタイミングを放送者が把握できるようになったためであると考えられる。放送者から視聴者に対して行う視線情報を用いたコミュニケーションの評価には、放送者が作業完了までにかかる時間の推測が困難な状況を設定することで、より明確な差が得られると推測される。

視聴者からのコメントに対する応答時間については、予測に反して大幅な増加が見られた。増加の原因は視線ヒートマップの確認にかかる時間が、トライアンドエラーで設置物を列挙する方法よりも長かったためだと考えられる。

1 回目の実験と 3 回目の実験の計測結果における 10 秒未満のデータは、放送者が設置物に対して「当てずっぽう」の反応を返した結果、正しい対象物に対する反応であったため発生したものである。逆に 4 回目の実験における計測結果では、反応時間には他の実験に比べ大きな増加が見られるが、これは推測が外れたために発生したと考えられる。それに対し、視線ヒートマップを用いた放送全体における計測時間には大きな差が見られないことから、視聴者からのコメントを受けてから視線ヒートマップの表示を確認し、応答するまでのプロセスにかかる時間が一定であることが考えられる。応答時間については増加してしまっていたが、主観評価については放送者主観、視聴者主観どちらについても上昇の傾向が見られた。特に放送者主観では有意差も確認できたことから、視線ヒートマップによる支援の有効性については証明できた。

ただし、2 回の放送に対してどちらの方がコミュニケーションを取りやすいかという質問について、視線ヒートマップを用いてない放送を挙げる協力者が一定数おり、視線ヒートマップの有無に対しての大きな差が見られなかったことから、現状のままでは実用性については課題があると考えられる。この原因は視線ヒートマップの操作に放送者が慣れていないためであると推測される。前述のように視線ヒートマップの表示方法の改善を行うことで、視線ヒートマップに対して不慣れな放送者であっても十分に利用することができるようになり、この課題についても改善が可能であると考えられる。

評価全体を通して、放送者の主観評価では 2 回の放送に対して有意差が見られたのに対し、視聴者の主観評価は 2 回の放送に有意差が見られなかった。特に放送者から視聴者へのコミュニケーションではこの傾向が大きい。視線ヒートマップは放送者に対する支援を目的としているため、放送者の主観評価で評価点が向上したことは目的に対して合致していると言えるが、視線ヒートマップ導入による影響が視聴者に対して低いことは今後の課題になりえると考えられる。

6.1 今後の展望

今後の課題として、本研究において実装を行った全方位カメラを用いたインターネット生放送システムに関する性能評価を実施する必要がある。さらに、視線ヒートマップ

の表示方法や、視線情報の提示方法の検討について行う必要もある。また、モバイル端末の普及によって屋外でインターネット生放送を行うユーザも多く、今回定義した以外のユースケースについても視線ヒートマップを適用することが可能か、導入に必要な機能や技術は何かについても調査を行っていく必要があると考える。

さらに、まだ研究がなされていない全方位カメラのコミュニケーション応用についても今回の知見をもとに提案を行っていく必要がある。具体的には今回の実験を通して収集した視線情報についても分析を行い、視線情報から他のノンバーバル情報を求められないかについても検討したい。

7. まとめ

本研究では、全方位カメラを用いたインターネット生放送環境において、放送者が視聴者の視線を把握できないという課題に対し、視聴者の視線を可視化する視線ヒートマップによる解決を図った。全方位カメラを用いたインターネット生放送環境と視線ヒートマップの実装を行い、評価実験を実施した。評価実験では、視線ヒートアップの有無で2パターンの放送を実施し、放送者と視聴者が行うコミュニケーションを時間計測とアンケートによる主観評価により分析した。実験の結果、視線ヒートマップの確認に対して一定の慣れが必要になるという課題が見つかったものの、視聴者の視線情報を提示することで視聴者のコメントを放送者に正しく伝達できるようになったということが分かった。今後の課題として、視線ヒートマップのインタフェース改良や視線ベクトルに対する重み付けの検討が挙げられる。

参考文献

- [1] 総務省：通信利用動向調査 平成27年調査, 2016; http://www.soumu.go.jp/johotsusintokei/statistics/data/160722_1.pdf
- [2] Periscope, <https://play.google.com/store/apps/details?id=tv.periscope.android&hl=ja>, (2017年12月参照)
- [3] Live 360 –Facebook 360, <https://facebook360.fb.com/live360/>, (2017年12月参照)
- [4] Roel Vertegaal, “The GAZE groupware system: mediating joint attention in multiparty communication and collaboration”, CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp.294-301, 1999
- [5] David M. Grayson, Andrew F. Monk, “Are you looking at me? Eye contact and desktop video conferencing”, ACM Transactions on Computer-Human Interaction (TOCHI) Volume 10 Issue 3, September 2003, pp.221-243, 2003
- [6] Hiroshi Ishii, Minoru Kobayashi, “ClearBoard: a seamless medium for shared drawing and conversation with eye contact”, CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.525-532, 1992
- [7] 高田真也, 西岡大, 齊藤義仰, “全方位カメラを用いたインターネット生放送における視線ヒートマップの提案”, 2017-GN-101, 25, pp.1-6, 2017
- [8] 高田真也, 西岡大, 齊藤義仰, “全方位カメラを用いたインターネット生放送における視線ヒートマップの開発”, マルチメディア、分散、協調とモバイル シンポジウム 2017 論文集, pp.1081-1088, (2017)
- [9] 360° 動画のヒートマップレポート, <https://support.google.com/youtube/answer/7407544/>, (2017年12月参照)
- [10] VRCHHEL, <https://vrchel.com/>, (2017年12月参照)
- [11] YouTube Creator Blog: Hot and Cold: Heatmaps in VR, <https://youtube-creators.googleblog.com/2017/06/hot-and-cold-heatmaps-in-vr.html>, (2017年12月参照)
- [12] 製品紹介 | RICOH THETA, <https://theta360.com/ja/about/theta/s.html>
- [13] Oculus, <https://www.oculus.com/rift/>
- [14] Google Cardboard(商品情報)
- [15] 笠原俊一, 暦本純一, “JackIn: 一人称視点と体外離脱視点を融合した人間による人間のオーグメンテーションの枠組み”, 情報処理学会論文誌 56(4), pp.1248-1257, 2015
- [16] 片山康太郎, 柴田啓司, 堀田裕弘, “極座標方向への画像積分による全方位動画像中の車両検出の高速化”, 電子情報通信学会技術研究報告. IE, 画像工学 113(469), pp.125-128, 2014
- [17] Intent to implement: HTML5 by Default, 2016, [https://groups.google.com/a/chromium.org/forum/#!searchin/chromium-dev/HTML5\\$20by\\$20default/chromium-dev/0wWoRRhTA_E/_E3jf400AAJ](https://groups.google.com/a/chromium.org/forum/#!searchin/chromium-dev/HTML5$20by$20default/chromium-dev/0wWoRRhTA_E/_E3jf400AAJ)
- [18] YouTube Engineering and Developers Blog: YouTube now defaults to HTML5 <video>, 2016, https://youtube-eng.googleblog.com/2015/01/youtube-now-defaults-to-html5_27.html
- [19] 「動画視聴ページ HTML5 版 (β)」提供開始のお知らせ - ニコニコインフォ, 2016, <http://blog.nicovideo.jp/niconews/ni064261.html>
- [20] Kurento, <http://www.kurento.org/>
- [21] Theta_GL, 2016, https://github.com/mganeko/THETA_GL
- [22] 津田侑, 上原哲太郎, 森村吉貴, 森幹彦, 喜多一, “インターネット生放送におけるユーザの活動の分析”, システム制御情報学会論文誌 28(10), pp.407-418, 2015