

# QoE を考慮した多視点ビデオ・音声伝送の Web/HTTP への適応

代表研究者 布目敏郎 名古屋工業大学 大学院工学研究科 准教授

## 1 はじめに

本研究では、YouTube やニコニコ動画といったサービスにおいて現在広く用いられている Web/HTTP の枠組みをベースとして、ユーザが複数の視点の映像から希望する視点の映像を選択して視聴することができる MVV-A (Multi-View Video and Audio) を、高い QoE (ユーザ体感品質: Quality of Experience) が得られるよう実現するための要素技術を検討する。

MVV-A においては、ユーザが見たい視点を要求してから、その映像がユーザに提示されるまでの遅延 (視点切替遅延) がユーザの知覚品質である QoE に大きく影響する。Web/HTTP の枠組みは、もともとトランスポートプロトコルとして TCP を用いることを基本に構築されており、リアルタイム情報の伝送に適さなかった。一方、最近ではそのようなリアルタイム情報の伝送に向かないという弱点を克服するために、WebSocket (RFC 6455) や WebRTC API (w3c WebRTC 1.0) などのような新しい枠組みが登場してきている。また、Web 上での動画配信では、MPEG-DASH (Dynamic Adaptive Streaming over HTTP) という仕組みも登場してきている。これらに代表される Web/HTTP の世界で提案されているリアルタイム通信を視野に入れた枠組みを用いることにより、多視点ビデオ・音声の伝送を Web/HTTP ベースで実現する。そして、ユーザの体感品質である QoE の高いサービスの実現を目指す。

## 2 本研究での検討事項

本研究では、MPEG-DASH を用いた MVV-A 伝送システムを実装し、その QoE を評価した。また、従来研究を行っていた HTTP によらない MVV-A 伝送システムを利用して、Web/HTTP 伝送に応用可能な要素技術の QoE 評価を行った。ここでは前者を中心に報告する。

### 2-1 MPEG-DASH を用いた MVV-A 伝送システムの実装と QoE 評価

#### (1) まえがき

従前の HTTP/TCP による動画配信では、ファイルを単位としてリクエスト及び伝送が行われていた。これに対し、近年、ネットワークの環境に応じて配信するストリーミングデータの質を動的に変化させることができるアダプティブストリーミングが注目されている。

2008 年ごろから、主な IT ベンダは、HTTP によるアダプティブストリーミングを独自の技術により行っている。Microsoft 社の Smooth Streaming, Apple 社の HTTP Live Streaming (HLS), Adobe 社の Dynamic HTTP Streaming がその例である [1]。これらは、各ベンダによってシステムの仕様が異なるため、サービスごとに動画コンテンツやソフトウェアモジュールなどを必要としていた。その中で、HTTP によるアダプティブストリーミングの配信方式の統一を意図して MPEG-DASH (Dynamic Adaptive Streaming over HTTP) [2] の標準化が行われた。この標準規格により、動画配信サービスのためのインフラや動画再生クライアント向けの技術を開発しやすくなる。特に、配信サービスを手掛ける事業者にとっては、動画配信サービスと動画再生クライアント間の互換性が向上する他、既存のビデオデータを再利用しやすい利点があり、市場の成長を促進する効果が期待されている。

これまでに MPEG-DASH に関する研究が数多く行われている。文献 [3] では、メディアバッファが予め設定された大きさであるとき、送信されるビデオデータの符号化ビットレートの切り替えが起きないようにすることで QoE 向上を図るアルゴリズムの評価が行われている。文献 [4] では、初期遅延、フレーム停止、符号化ビットレートの変動の 3 つがユーザ体験にどのように影響するかを調査するために主観評価実験を行っている。しかし、文献 [3] や文献 [4] は、単視点のストリーミングに関する研究であり、MVV-A システムを考慮した QoE 評価がされていない。

また、多視点ビデオを DASH で扱う試みに関して、文献 [5] では、RTP/RTSP (Real-time Transport Protocol/Real Time Streaming Protocol), HTTP Progressive Download と DASH の 3 つの伝送方式に対して、多視点ビデオの伝送方法を提案している。そして、それぞれに対して、帯域使用量や視点切り替え遅延など

の比較、評価を行っている。しかし、文献[5]では、QoEに関する評価は行われていない。

一方、MVVのIP伝送におけるQoEを議論している研究として、文献[6]や文献[7]が挙げられる。これらの研究では、MVVに音声を加えた多視点ビデオ・音声(MVV-A: Multi-View Video and Audio)システムを扱い、QoE評価を行っている。しかし、文献[6]、文献[7]では、HTTP/TCPを用いたMVV-Aシステムは検討されていない。

以上のことから、本研究では、HTTP/TCPによるMVV-Aシステムが、ユーザ満足度にどのような影響を及ぼすか定量的に調査する。MPEG-DASHを用いたMVV-Aシステムを実装し、1視点のみを伝送する場合と4視点選択することができる場合を扱い、主観評価実験を行う。この際、ネットワークに負荷トラフィックを流す。種々のネットワーク状態を作り出すことで、ネットワーク負荷状況がQoEにどのような影響を及ぼすか考察する。

## (2) HTTP/TCPを用いたMVV-Aシステム

MVV-Aシステムとは、複数台のビデオカメラで様々な視点から撮影されたビデオを、ユーザが任意の視点を選択しながら視聴できるシステムである。見たい視点を選択して視聴できることで、ユーザに対し自由度の高いサービスを実現できる。

MPEG-DASHは、ISO/IEC 23009-1として知られるHTTPによるストリーミング技術の標準規格である。MPEG-DASHでは、サーバから受信されるビデオデータを動的に変更するアダプティブストリーミング配信を実現するため、1本のコンテンツに対して、画面サイズや符号化ビットレートが異なるビデオデータ群をWebサーバに格納しておく。各ビデオデータは、それぞれSegmentと呼ばれる小さな塊に分割される。伝送路の状況に応じて、クライアントが最適な符号化ビットレートのSegmentをHTTPでサーバへ要求し、サーバはその要求に応じてSegmentを送信する。

この配信の仕組みを管理するために用いるデータがMPD(Media Presentation Description)ファイルである。Webサーバに格納されたコンテンツの構成に関する情報をXML(eXtensible Markup Language)形式の階層構造で記述する。MPDファイルに記述する情報は、ビデオデータを格納したサーバのURL(Uniform Resource Locator)や、格納しているビデオデータ群の圧縮方式、画面サイズや符号化ビットレート、音声データの圧縮方式、音声の言語情報などである。これらを、Period、AdaptationSet、Representationといった複数の構造体による階層構造で記述する。Periodは、番組やコンテンツを構成する1つの単位で、同期をとった一組の音声やビデオなどの情報のことである。AdaptationSetは、コンテンツを構成するビデオ、異なるチャンネルや音声の音声、異なる言語の字幕などの単位を表現し、クライアントが再生する時に選択するコンテンツの構成要素となる。Representationは、メディアファイルの保存先を示すURLやビデオの画像サイズ、ビデオや音声の符号化ビットレートなどを記述する。

本研究では、多視点用MPDファイルを作成し、JavaScriptによる視聴プログラム[8]を拡張する。さらに、視点要求機能を実装することにより、HTTP/TCPによるMVV-Aシステムを実現する。MPDファイルには、音声、ビデオそれぞれにAdaptationSetがあり、ビデオのAdaptationSetには用意した符号化ビットレートごとにRepresentationがある。多視点用MPDファイルを作成するために、予め複数の視点について、複数の符号化ビットレートのビデオを用意し、視点ごとのMPDファイルを作成する。これを基に、多視点用MPDファイル内には、視点ごとに音声、ビデオのAdaptationSetを用意する。そして、AdaptationSet内には、Representationに加えて、視点情報であるViewpointを追加する。このViewpointを使うことにより、要求視点の判断、切り替え、データの伝送を行う。これにより、1つのMPDファイルによって多視点の異なる符号化ビットレートのメディアファイルの管理を行うことが可能となる。

ここで、実装したMVV-Aシステムの視点切り替えプログラムを説明する。まず、クライアントがMPDファイルをサーバへリクエストする。クライアントは、MPDファイル、及び視点切り替えプログラムをサーバから受信し、コンテンツの全体像を把握する。次に、MPDファイルの記述に従い、各AdaptationSetにおけるそれぞれのRepresentationが示すURLにあるメディアファイルのヘッダとSegment情報を含むCueリストを取得する。この情報を用い、初期視点の音声・ビデオのSegmentをサーバへ要求し、視聴する。視点切り替え時には、要求された視点の再生に必要なSegmentをクライアントが決定し、それをサーバへリクエストすることで、要求された視点の視聴を行う。

## (3) 実験方法

本実験で用いたネットワーク構成を図1に示す。実験システムは、メディアサーバ、メディアクライアン

ト、Webサーバ、Webクライアントの4端末と、ルータ2台から構成される。メディアサーバのOSはCentOS 5.3であり、メディアクライアントのOSはWindows 7である。2台のルータにはAlcatel Lucent社(旧RiverStone Networks社)製のRS3000を用いる。全ての回線は100Mbpsの全二重Ethernet回線で接続されている。メディアサーバからメディアクライアントに向けて音声及びビデオを、WebサーバからWebクライアントに向けてWebStone2.5[9]によるWebトラフィックを伝送する。また、Webサーバ、メディアサーバとともにApache2.2[10]を用いる。

メディアクライアントでのビデオ視聴に用いるWebブラウザはGoogle Chromeを使用する。MPEG-DASHによるビデオ再生のために、ビデオプレイヤーとして、webm-dash-javascript[8]を用いる。HTML5のvideo要素を用いてビデオの制御を行う。MPEG-DASH伝送を行うため、音声・ビデオのコンテナフォーマットとしてWebMを用いる。このため、ビデオの符号化方式はVP8、音声の符号化方式はVorbisを採用する。ビデオ変換ツールとして、ffmpeg, libwebm[11], webm-tools[12]を利用する。ffmpegは、WebM形式の音声・ビデオファイルを作成するためにlibvpx[13]とlibvorbis[14]をサポートしているものとする。libwebmは、sample\_muxerを使用し、伝送される符号化ビットレートの切り替えを可能にするようWebMファイルを整える。webm-toolsは、MPDファイルの作成に使用する。

表1に実験で用いた音声・ビデオの仕様を示す。本実験では、使用するビデオを500kbps, 1000kbps, 1500kbpsの3種類の符号化ビットレートでエンコードした。ネットワークの負荷状況に対して、500kbps, 1000kbps, 1500kbpsの間で、配信するフラグメントをシームレスに変えることができる。

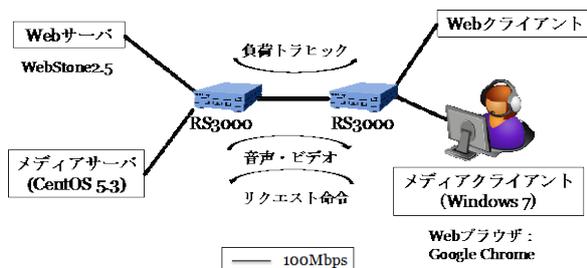


図1: 実験環境

表1: 音声・ビデオの仕様

|        | 項目            | 内容            |
|--------|---------------|---------------|
| 音声     | 符号化方式         | Vorbis        |
|        | フレームレート [fps] | 25            |
|        | ビットレート [kbps] | 32            |
|        | チャンネル         | モノラル          |
| ビデオ    | サンプルレート [kHz] | 8             |
|        | 符号化方式         | VP8           |
|        | フレームレート [fps] | 29.97         |
|        | GOP長          | 15            |
| 音声・ビデオ | ビットレート [kbps] | 500,1000,1500 |
|        | 解像度           | 640×480       |
|        | コンテナ          | WebM          |
| 音声・ビデオ | 時間 [s]        | 630           |

主観評価実験では、図2の環境で、8の字に配置されたプラスチックのレール上を動く電車の玩具を視聴してもらう。単視点音声・ビデオ伝送の場合は、視点1の映像のみ視聴することができる。MVV-A伝送の場合、被験者は4つある視点を切り替えながら視聴する。視聴するビデオは予め録画されたものである。使用するカメラは全4台である。

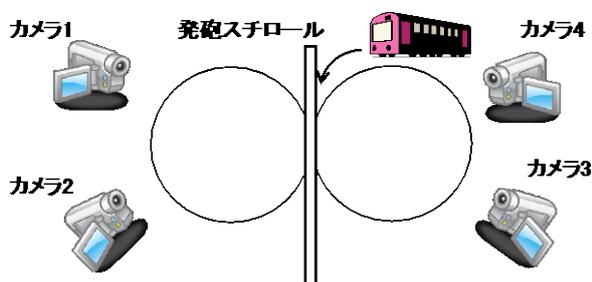


図2: 電車のカメラ配置

表2: 主観評価実験で用いる形容詞対

|     |                   |
|-----|-------------------|
| ビデオ | ビデオがごちゃごちゃしない-滑らか |
|     | ビデオが見にくい-見やすい     |
| 同期  | ビデオと音声はずれた-一致した   |
| 心理  | 束縛的な-自由な          |
| 総合  | 不満な-満足な           |

本実験では、多数の形容詞対を用いて多次元的なQoE評価を行う。形容詞対とは、相反する意味を持った2つの形容詞のことである。表2に主観評価実験で用いる形容詞対を示す。形容詞対は全9個あり、六つのカテゴリーに分類される。ビデオに関する形容詞対が三つ、音声に関する形容詞対が一つ、ビデオと音声の同期に関する形容詞対が一つ、視点切り替えの応答性に関する形容詞対が一つ、心理的要因に関する形容詞対が二つ、総合満足度に関する形容詞対が一つである。各形容詞対には、v1~o1までの略

称をつける。これらの形容詞対を5段階（評点1～5）の評定尺度法[15]で評価してもらう。評点1には負の印象を持つ形容詞（表2において左側に書かれているもの）を、評点5には正の印象を持つ形容詞（表2において右側に書かれているもの）を設定する。得られた評点から MOS（Mean Opinion Score）を求め、QoE 尺度値とする。

評価対象は配信方式2種類、Webクライアント数5種類を組み合わせた10通りである。Webクライアントの値は、10、20、30、40、50である。1回の試行は20秒である。被験者は、20代男子学生20人である。1回の試行が終了するごとに各々の形容詞対で表現される品質を評価する。1人当たりの総評価時間は、約15分である。なお、被験者は実験を行う前に、負荷トラヒックのない状態で、作業に慣れるまで練習を行った。

#### （4）実験結果と考察

本報告では、形容詞対「ビデオがぎこちない - 滑らか」、「ビデオが見にくい - 見やすい」、「束縛的な - 自由な」、「不満な - 満足な」について評価結果を図3から図6に示す。本節で示すグラフは、横軸がWebクライアント数で、縦軸がMOSである。

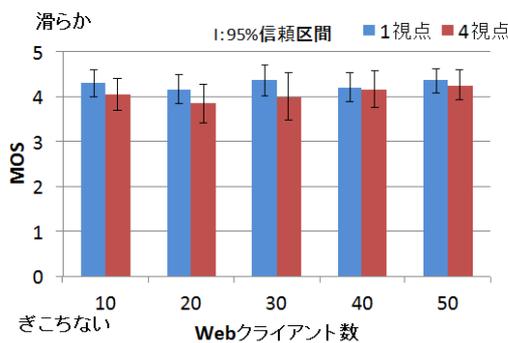


図3：ビデオがぎこちない - 滑らか

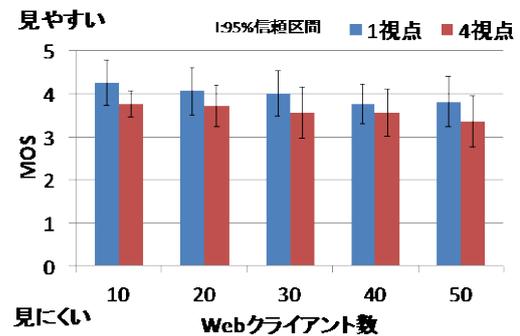


図4：ビデオが見にくい - 見やすい

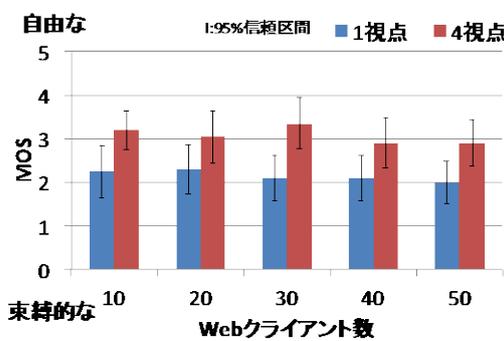


図5：束縛的な - 自由な

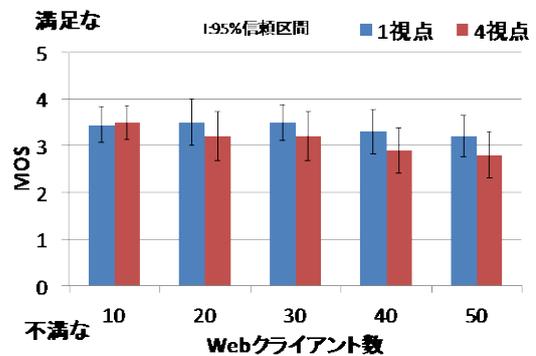


図6：不満な - 満足な

「ビデオがぎこちない - 滑らか」のMOSを図3に示す。すべてのWebクライアント数において、単視点音声・ビデオ伝送がMVV-A伝送よりもMOSが高くなっている。視点切り替えが起きる際、切り替わった視点の符号化ビットレート群の中で最も小さい符号化ビットレートのデータをサーバから取得しようとするため、データの再バッファリングが発生する。このため、ビデオがかくつき、MVV-A伝送のMOSが小さくなっている。

「ビデオが見にくい - 見やすい」のMOSを図4に示す。すべてのWebクライアント数において、単視点音声・ビデオ伝送がMVV-A伝送よりもMOSが高くなっている。MVV-A伝送では、視点切り替えが起きる際、切り替わった視点の符号化ビットレート群の中で最も小さい符号化ビットレートのデータを再生され、映像のきれいさが失われる。また、データの再バッファリングの発生によりビデオの滑らかさが低下する。一方、単視点音声・ビデオ伝送では、時間の経過とともに高い符号化ビットレートを再バッファリングすることな

く視聴することができる。このため、ビデオの見やすさにおいて単視点音声・ビデオ伝送が MVV-A 伝送よりも MOS が高くなっている。

「束縛的な - 自由な」の MOS を図 5 に示す。すべての Web クライアント数において、MVV-A 伝送が単視点音声・ビデオ伝送よりも MOS が高くなっている。視点切り替えの有無が、ユーザが感じるシステムの自由度に影響を与え、MOS を高めている。

「不満な - 満足な」の MOS を図 6 に示す。Web クライアント数 10 において MVV-A 伝送の方が、わずかながら単視点音声・ビデオ伝送よりも高い MOS を示し、それ以外の Web クライアント数では低い MOS を示している。本研究で用いた MVV-A システムは、ビデオのなめらかさや見やすさにおいて MOS は低くなっているが、システムの自由度においては高い MOS を示す。本実験で行った作業では動く電車を見てもらった。ネットワーク負荷が高い状態においては、視点の切り替えを行った際に、再バッファリングが発生し、ビデオがぎこちなくなってしまう。しかし、MVV-A システムでは視点を切り替えることができることにより自由度が高くなる。被写体が動く状況下において、視点切り替えにより被写体を追いかけて見ることができるということはユーザの満足度を高める。だが、ネットワーク負荷が高い状態においては、視点切り替え時に再バッファリングが発生するため、視点切り替え速度が遅くなり、ビデオがぎこちなくなるため、単視点音声・ビデオ伝送の方がユーザの満足度が高くなった。以上の結果から、視点切り替え時の再バッファリングによる影響を緩和し、視点切り替え速度の向上を図るような制御を検討する必要があると考えられる。

## (5) むすび

本報告では、MPEG-DASH を用いた MVV-A システムを実装し、QoE に与える影響を定量的に調査した。1 視点のみを伝送する場合と 4 視点選択することができる場合を主観評価実験により比較した。その結果、ネットワーク負荷が小さい場合、HTTP/TCP を用いた MVV-A 伝送は QoE を向上させることがわかった。HTTP を用いた MVV-A システムにおいても、視点切り替えが可能であることによるシステムの自由度がユーザの満足度の向上につながることをわかった。

今後の課題として、視点切り替え速度の向上を図り、QoE を向上させる伝送制御システムを構築することが挙げられる。異なるコンテンツを用いた実験により、コンテンツの違いによる QoE への影響の調査する。

## 2-2 Web/HTTP 伝送システムに適用可能な要素技術の QoE 評価

従来研究を行っていた HTTP によらない MVV-A 伝送システムを利用して、Web/HTTP 伝送に応用可能な要素技術の QoE 評価を行った。これらの検討は、大きく以下の三つに分類される。

### (1) ビデオピクチャパターンの違いによる画質と視点切替の応答性とのトレードオフ

ビデオの配信方式とユーザインタフェースが QoE に及ぼす影響を調査した。4 視点のビデオを扱う場合を考え、合計のビットレートが一定となる三つの配信・表示方式（1 画面 1 ストリーム方式、1 画面 4 ストリーム方式、および 4 画面 4 ストリーム方式）がユーザに与える影響を QoE の観点から比較する。主観評価実験の結果より、視点切り替えが速く行える状況では 1 画面 1 ストリーム方式が最もユーザの満足度が高く、視点切り替えが遅くなってしまう状況では視点切り替えを行わなくても全ての視点の映像を視聴できる 4 画面 4 ストリーム方式で高いユーザ満足度を得られた。

### (2) ビデオの提示方式や転送方式に関する検討

ビデオの配信方式とユーザインタフェースが QoE に及ぼす影響を調査した。4 視点のビデオを扱う場合を考え、合計のビットレートが一定となる三つの配信・表示方式（1 画面 1 ストリーム方式、1 画面 4 ストリーム方式、および 4 画面 4 ストリーム方式）がユーザに与える影響を QoE の観点から比較する。主観評価実験の結果より、視点切り替えが速く行える状況では 1 画面 1 ストリーム方式が最もユーザの満足度が高く、視点切り替えが遅くなってしまう状況では視点切り替えを行わなくても全ての視点の映像を視聴できる 4 画面 4 ストリーム方式で高いユーザ満足度を得られた。

また、複数視点同時配信方式と視点切替ユーザインタフェースとの組み合わせが QoE に及ぼす影響を評価した。視点切替の応答性とビデオ画質とのトレードオフを考慮した三種類の配信方式を扱い、三種類の視点切替ユーザインタフェースを QoE の観点から比較した。その結果、ユーザインタフェースが QoE に大きく影響すること、ならびに画質を優先した配信方式がユーザに好まれることを明らかにした。

### (3) 音声切り替え可能な MVV-A システム (MVV-SA)

視点切替において、ビデオのみが切り替え可能な場合とビデオとともに音声も切り替え可能な場合とで QoE に及ぼす影響の違いを評価するため比較実験を行った。QoE を SD 法により多次元的に評価した。主観評価実験により得られた QoE 尺度値に対して因子分析を行い、主に三つの因子（音声の臨場感、ビデオの出力品質、および視点切替の簡単さ）が QoE に影響を及ぼすことを示した。

## 3 おわりに

本研究では、現在広く用いられている Web/HTTP の枠組みを用いて、多視点ビデオ・音声 (MVV-A: Multi-View Video and Audio) 伝送を QoE 高く実現するための技術検討を行った。Web/HTTP を介して、多視点ビデオ・音声の伝送ができるよう、既存の MPEG-DASH (Dynamic Adaptive Streaming over HTTP) によるビデオ伝送のためのオープンソースプレーヤをベースとして、サーバおよびクライアントソフトウェアの実装を行い、基本的な QoE 評価を行った。一方、Web/HTTP での伝送に応用可能な要素技術の QoE 評価を行った。これらの技術は、今後、MPEG-DASH を用いて本研究で開発したシステムへ順次適用し、QoE 向上に適したシステム構築につなげていく予定である。

## 【参考文献】

- [1] 平林 光浩, “4. 次世代動画配信技術「MPEG-DASH」技術概要と標準化・関連技術動向”, 映像情報メディア学会誌, vol. 67, no. 2, pp. 109-115, Feb. 2013.
- [2] ISO/IEC 23009-1, “Dynamic adaptive streaming over HTTP (DASH) Part1: Media presentation description and segment formats, ” May 2014.
- [3] Y. Cao, X. You, J. Wang and L. Song, “A QoE friendly rate adaptation method for DASH,” Proc. IEEE BMSB 2014, pp. 1-6, June 2014.
- [4] Y. Liu, S. Dey, D. Gillies, F. Ulupinar and M. Luby, “User experience modeling for DASH video,” Proc. Packet Video Workshop (PV 2013), pp. 1-8, Dec. 2013
- [5] H. Zhang, X. Gu and R. Ishibashi, “Seamless and efficient stream switching of multi-perspective videos,” Proc. Packet Video Workshop (PV 2012), pp. 31-36, May 2012
- [6] E. Jimenez Rodriguez, T. Nunome and S. Tasaka, “QoE assessment of multi-view video and audio IP transmission,” IEICE Trans. Commun., vol. E92-B, no. 6, pp. 1373-1383, June 2010.
- [7] 山本 誠, 布目 敏郎, 田坂 修二, “多視点ビデオ・音声 IP 伝送のための複数視点同時配信方式の QoE 評価”, 信学技報, vol. 112, no. 10, pp. 81-86, Apr. 2012.
- [8] “webm-dash-javascript,” <https://chromium.googlesource.com/webm/webm-dash-javascript/>.
- [9] Minecraft Inc, “WebStone benchmark information,” <http://www.minecraft.com/webstone/>.
- [10] “Apache HTTP SERVER PROJECT,” <http://httpd.apache.org/>.
- [11] “libwebm,” <https://chromium.googlesource.com/webm/libwebm>.
- [12] “webm-tools,” <https://chromium.googlesource.com/webm/webm-tools>.
- [13] “libvpx - The WebM Project,” <http://www.webmproject.org/code/>.
- [14] “Ogg Vorbis - Xiph.org,” <http://xiph.org/vorbis/>.
- [15] 田中 良久, “心理学的測定法 第 2 版”, 東京大学出版会, 1977.

〈発表資料〉

| 題名  | 掲載誌・学会名等   | 発表年月                    |
|---|--|-------------------------|
| Multidimensional QoE of multiview video and selectable audio IP transmission                                      | The Scientific World Journal   | in Press (2015年3月Web掲載) |
| MPEG-DASH を用いた多視点ビデオ・音声 (MVV-A) 伝送の QoE 評価  | 電子情報通信学会技術研究報告   | 2015年3月                 |
| The effect of user attributes and video presentation methods on QoE of multi-view video and audio IP transmission | 電子情報通信学会 2015 年総合大会 講演論文集  | 2015年3月                 |
| Video transmission and presentation methods for multi-view video and audio IP transmission                        | Proc. 8th International Conference on Signal Processing and Communication Systems (ICSPCS 2014)        | 2014年12月                |
| QoE assessment of multi-view video and audio simultaneous IP transmission: The effect of user interfaces          | Proc. 5th International Conference on Information and Communication Technology Convergence (ICTC 2014) | 2014年10月                |
| 多視点ビデオ・音声 IP 伝送における時間空間品質のトレードオフが QoE に及ぼす影響  | 電子情報通信学会 2014 年ソサイエティ大会講演論文集   | 2014年9月                 |