

# 会社で働かない時代を支える 映像コミュニケーションの これまでとこれから

Past and Future of Video Communication for Working from Anywhere

鎌村星平

## Abstract

昨今の社会情勢を受けて、Web会議をはじめとする映像コミュニケーションはこれまで以上に注目され、急速に普及しつつあります。本稿では、まずネットワークの広帯域化と映像符号化技術の発展に支えられながら進化してきた映像コミュニケーションの歴史を振り返ります。続いて、多点接続装置と呼ばれる装置を用いた場合を例に、複数の参加者の間で映像や音声共有され、コミュニケーションが実現される基本的な仕組みを解説します。最後に映像コミュニケーションの今後の展望についても簡単に紹介します。

キーワード：映像コミュニケーション、Web会議、テレビ会議、働き方改革

## 1. はじめに

映像コミュニケーションとは、社会人の皆さんがネットワークにより自宅から利用するWeb会議や、学生の皆さん向けの遠隔授業を実現する手段の総称です。映像コミュニケーションそのものは古くから存在していたものの、2020年初頭から世界各国を震撼させているCOVID-19対策を契機に、その利用が急速に普及し始めました。

筆者も2020年2月から5月にかけてほぼ毎日、テレワークにより自宅ですべての仕事をしておりました。往復で約2時間強の満員電車通勤から解放されたことで時間を有効に活用でき、むしろ生産的な日々を過ごせているように感じます。

場所や時間にとらわれない柔軟な働き方として定義される「テレワーク」を実現するためのツールは世の中にたくさんありますが、中でも会議が業務の中心となるホワイトカラーにとって映像コミュニケーションは欠かせません。そもそも、コミュニケーションとは社会生活を営む人間の間に行われる知覚、感情、及び思考の伝達のことであり、その中で、映像コミュニケーションとは通

信技術を用いて双方向での音声や映像によるコミュニケーションを行うことと定義されます。電話のような音声のみによるコミュニケーションとは異なり、人の表情や仕草といった非言語の情報が伝達されることで、たとえ離れていても対面しているかのような臨場感を提供することが可能となります。

本稿では今後社会で活躍される読者の皆さんが映像コミュニケーションに関する理解を深めることで、それをより活用頂けるよう、映像コミュニケーションがこれまでたどってきた歴史と、複数の参加者間で映像や音声共有される基本的な仕組みについて解説していきます。

## 2. 映像コミュニケーションの歴史

図1に映像コミュニケーションのこれまでの歴史を示します。1970年に大阪万博でテレビ電話のデモが登場しますが、本格的な商用サービスは1980年代に当時の電電公社により開始されました。1990年代になるとISDN (Integrated Services Digital Network) と呼ばれるデジタル通信網が普及し、ISDN向けのH. 320と呼ばれるプロトコルを用いたテレビ会議が普及していきます。更に1990年代後半にはIP (Internet Protocol) 網を対象としたH. 323<sup>(用語)</sup> (1)プロトコルが勧告され、2000年代に突入するとより高品質なテレビ会議が普及していきます。例えば、それまではSD (Standard Definition) 画質と呼ばれる画素数が720×480、いわゆるDVDと同じ画質であったテレビ会議は、画素数が

鎌村星平 正員：シニア会員 日本電信電話株式会社 NTT ネットワークサービスシステム研究所  
E-mail shohei.kamamura@ieee.org  
Shohei KAMAMURA, Senior Member (NTT Network Service Systems Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Musashino-shi, 180-8585 Japan).  
電子情報通信学会誌 Vol.103 No.12 pp.1261-1265 2020年12月  
©電子情報通信学会 2020

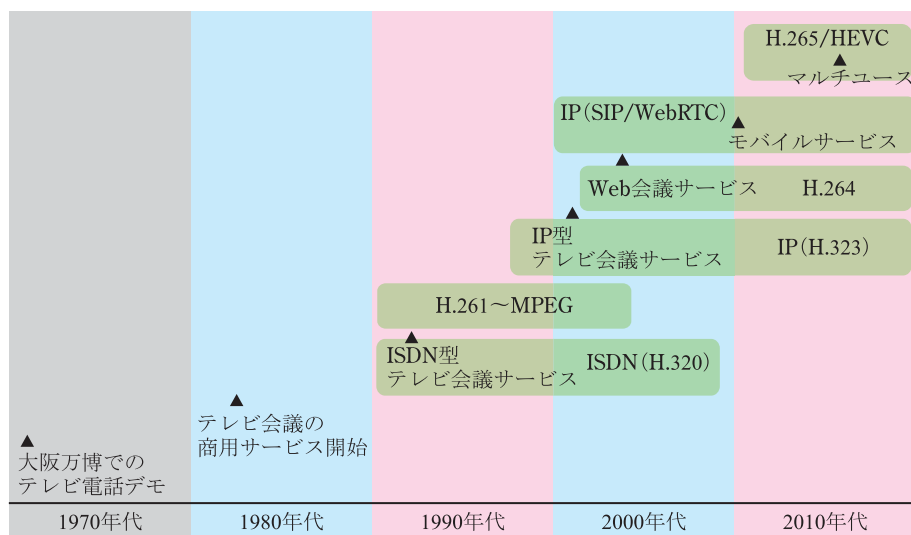


図1 映像コミュニケーションの歴史 1970年のテレビ電話のデモから始まり、ネットワーク技術と映像符号化技術の発展に支えられながら本格的なテレビ会議やWeb会議へと発展。映像コミュニケーションは時代のニーズに合わせて、様々なデバイスで、様々なユースケースで使われてきました。

1,280×720であるHD (High Definition) 画質へと変化してきました。

2000年代の中ほどになると、専用端末を用いたいわゆるテレビ会議だけではなく、PCを用いてインターネットを経由して利用するWeb会議サービスが開始します。その後のWeb会議においては、専用のアプリケーションを用いるものだけでなく、後述するWebRTC<sup>(2),(3)</sup>と呼ばれるWebブラウザ上で動作するものも登場してきました。また、2000年代後半にスマートフォンが爆発的に普及し始めたことで、2010年代に突入するとLINEなどのモバイル向けサービスが流行し始めました。IPネットワークの広帯域化もますます進んでいき、現在では画素数が1,920×1,080でブルーレイ並みの高画質となるフルHDによる映像コミュニケーションも可能となっています。このような映像品質の向上やデバイスの多様化を受けて、2010年代中頃からは、映像コミュニケーションは会議のためにある、という考え方も変わってきています。例えば、IT重説と呼ばれるオンラインでの不動産契約、医療過疎地に対するオンライン診療、オンラインでの英会話や授業といった教育向けの利用など、会議の枠を超えた使い方への期待が次第に高まってきました。

また、映像コミュニケーションの発展は、IPに代表

## 用語解説

**H. 323** IP網でリアルタイムの音声・動画通信を行うためにITU-Tで勧告。ITU-TにおけるHシリーズはオーディオビジュアル及びマルチメディアシステムに関する。

されるネットワークの大容量化だけでなく、映像や音声の符号化技術の進展によって支えられてきました。符号化技術とは、簡単に言うと映像の視聴品質をなるべく落とさずにデータサイズを圧縮する技術となります。映像コミュニケーションに関連すると、まず1990年に誕生したH. 261がISDNによるテレビ会議サービスで使われました。次に2003年に成立したH. 264/AVC<sup>(4)</sup>がIPによる映像コミュニケーションサービスで使われ、現在も主流な映像符号化技術となっています。また2013年にはH. 265/HEVC<sup>(5)</sup>が提案され、映像コミュニケーションの更なる高精細化を実現する次世代規格として期待されています。

このように映像コミュニケーションはネットワーク技術と映像符号化技術の発展に支えられながら、時代のニーズに合わせて、様々なデバイスで、様々なユースケースで使われてきました。3.では、この映像コミュニケーションがどのような仕組みで動いているか、その基本的な仕組みを解説していきます。

## 3. 映像コミュニケーションを支える技術

### 3.1 多地点接続の仕組み

映像コミュニケーションでは電話のように一対一の通信を行うのではなく、複数の参加者が同時に接続して通信する仕組みが必要になります。最も単純な仕組みとして、全ての参加者が互いに接続するいわゆるフルメッシュの接続方式が考えられます(図2(a))。しかし、この接続方式は全ユーザ間で映像トラフィックの送受信が発生するため、ユーザ端末の負荷が大きくネットワーク帯

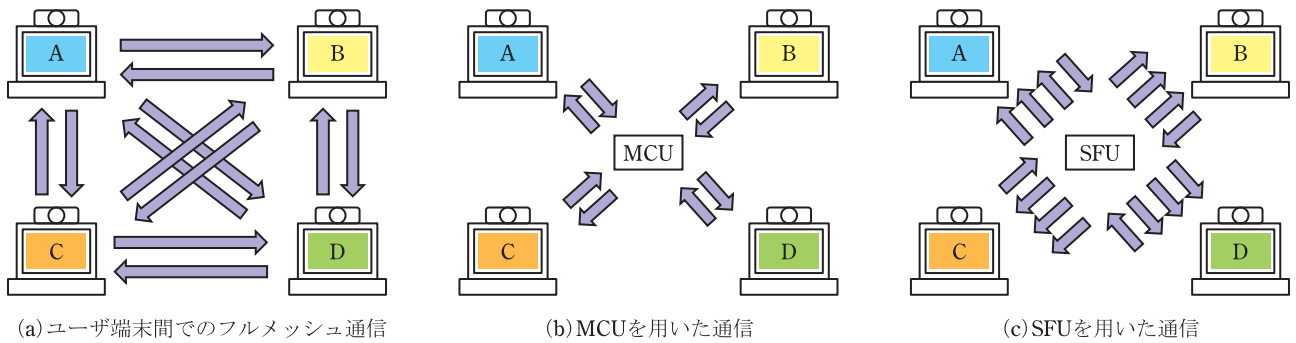


図2 多地点接続の仕組み (a)の方式では全ユーザ端末間でフルメッシュに映像と音声を通じます。(b)の方式では、MCUと呼ばれる多地点接続装置が各ユーザ端末から送信される映像の合成処理を行った上で、各ユーザ端末に合成映像を送り返します。(c)の方式ではSFUと呼ばれる装置がユーザ端末から送信される映像と音声を複数のユーザ端末へ代行して配信します。

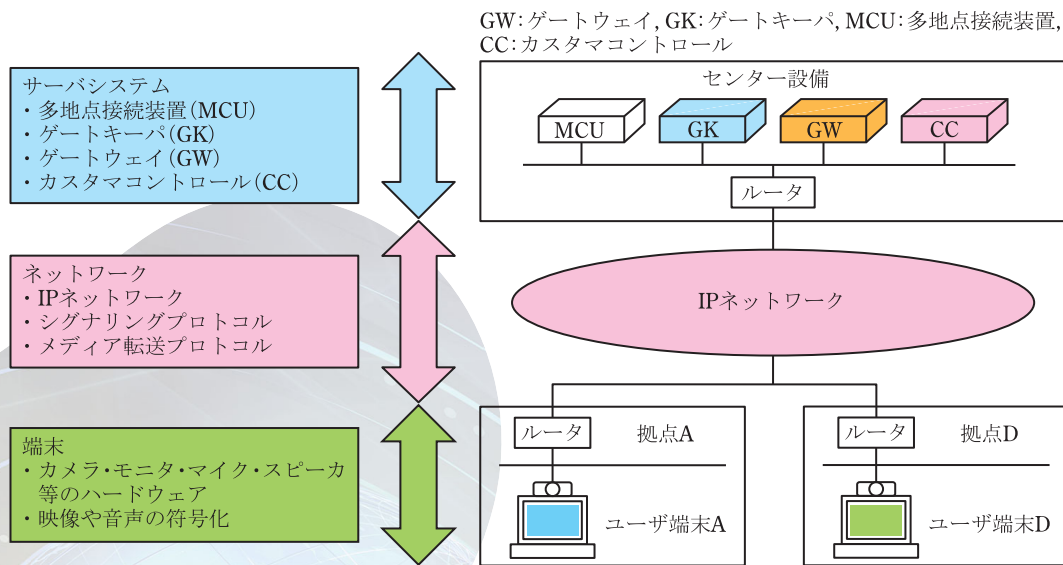


図3 多地点接続装置を用いた場合の構成技術 MCUと呼ばれる多地点接続装置を用いた映像コミュニケーションは、大きく分類するとサーバシステム、ネットワーク、及び端末に分類される技術で構成されます。

域の利用量も増加します。この問題を解決するためにサーバを間に置いて通信処理を代行する方式が提案されています(図2(b), (c))。図2(b)には、MCU (Multi-point Control Unit) と呼ばれる多地点接続装置を介して参加者が通信を行う仕組みを示しています。MCUは各ユーザ端末から送信される映像と音声の合成処理を行った上で、各ユーザ端末に合成映像を送り返します。このためMCUが設置されるサーバに対しては高い処理能力が要求されますが、ユーザ端末の負荷は削減されネットワークの帯域利用量も削減することが可能です。図2(c)には、SFU (Selective Forwarding Unit) と呼ばれる装置がユーザ端末から送信される映像を複数のユーザ端末へ配信代行する仕組みを示しています。ユーザ端末は全端末からの映像を受信しますが、映像の送信はSFUに対してのみ行われます。このためMCUを用

いた方式と比較してサーバの処理負荷を抑えつつ、フルメッシュ通信と比較してユーザ負荷やネットワーク帯域利用量を削減可能な方式となります。

### 3.2 多地点接続装置 (MCU) を用いた場合の構成技術

次に代表的な方式であるMCUと呼ばれる多地点接続装置を用いた場合を例に、映像コミュニケーションを構成する技術を解説していきます。本方式は非常に多岐にわたる要素技術から構成されますが、大きく分類すると、図3に示すようにサーバシステム、ネットワーク、及び端末で構成されています。

サーバシステムは図3上部に示すように多地点接続装置、ゲートキーパ、ゲートウェイ、及びカスタマコントロールから構成されます。多地点接続装置は先ほど説明

したとおり、ユーザ端末から送信される映像と音声を作成し、ユーザに送り返す映像コミュニケーションの要となる技術です。その隣のゲートキーパは、その名が示すとおり門番的な受付制御機能を提供します。例えば図3に示すユーザ端末AとDのみがゲートキーパによって登録されている場合、その他のユーザ端末は多地点接続装置へアクセスできないため、会議へは参加できず通信のセキュリティを保つことが可能となります。ゲートキーパは一般的にはこの受付制御以外にも、映像コミュニケーション特有のIDとIPアドレスとをひも付けるアドレス解決機能や、ユーザ端末の接続帯域を調整する帯域制御機能も提供します。次にゲートウェイを説明します。ゲートウェイはプロトコル変換機能を提供することで、様々なユーザ端末やアプリケーションの相互接続を実現する技術です。プロトコルとは通信を行うための共通手順のことで、プロトコルが異なるアプリケーションは本来お互いに通信できません。そこでゲートウェイが間に入って共通のプロトコルによる通信に見せかけることで、異なる端末やアプリケーション間での相互接続を可能としています。最後にカスタマコントロールを説明します。カスタマコントロールはカスタマ、すなわちユーザへ会議の制御権を提供する機能です。例えば、会議の開始や終了、画面レイアウトや音声ミュートの制御、会議の予約による自動開催といった機能を、ユーザが親しみやすいWebなどのユーザインタフェースで提供します。

次にネットワークの説明に移ります。ユーザ端末とサーバシステムは原則IPネットワークで接続されます。IPネットワークそのものの解説は省略しますが、特に映像コミュニケーションにおいてはIPレイヤの上位レイヤで動作するシグナリングプロトコルとメディア転送プロトコルが重要な役割を果たします。シグナリングプロトコルの役割は、通信を行う端末間でセッションと呼ばれる接続関係を確立することです。代表的なプロトコルとしてテレビ会議で利用されるH.323、IP電話やWeb会議で利用されるSIP (Session Initiation Protocol)<sup>(6)</sup>が挙げられます。シグナリングプロトコルによって端末間の接続関係が確立された後は、メディア転送プロトコルによって音声や映像がリアルタイムに伝送されます。メディア転送には国際標準となっているRTP (Real-Time Transfer Protocol) や RTCP (Real-Time Transfer Control Protocol)<sup>(7)</sup>が使われることが主流です。RTPは音声や映像といったデータストリームをリアルタイムに配信するプロトコルであり、処理が単純で遅延が少ない特徴があります。このため、データを配信する際にデータの送達確認やフロー制御を行いたい場合には、RTCPを補助的に併用することが可能です。

次にユーザ端末側の技術を説明します。ユーザ端末はテレビ会議における専用端末だけでなく、PCやスマー

トフォン、タブレットといった汎用的なコンピュータを利用可能です。その際、カメラ、モニタ、マイク、スピーカといった映像コミュニケーションのためのハードウェアが備わっていることはもちろんですが、内部的には2.で述べた映像符号化技術が大きな役割を担っています。符号化技術とは、音声や映像を圧縮したり伸張したりする技術です。ネットワークへメディアを転送する際に容量を圧縮する処理をエンコード、逆に圧縮されたメディア情報を人が視聴できるように伸張する処理をデコードと呼びます。特に映像に関しては、品質良く効率的に圧縮する技術は重要な研究テーマとなっています。例えば、映像を送信する際に面的に隣り合った画素が似ている点に着目したフレーム内予測、また、時間的な動きベクトルを予測するフレーム間予測といった技術によって、データを丸ごと送るのではなくその差分のみを送ることで、データ圧縮を実現しています。

最後にこれまで紹介した技術のおさらいを兼ねて、関連の深いWebRTC<sup>(2),(3)</sup>と呼ばれる技術を紹介します。

WebRTCはWebブラウザ上で映像コミュニケーションを実現する技術で、W3CとIETFによって標準化が進められています。WebRTCという単体の通信技術があるのではなく、複数の技術が合わさってWebRTCによる通信が実現されているイメージを持って頂くともいかもしれません。このため、WebRTC通信の実現方法は自由度が高いものとなります。例えば標準的なWebRTCアプリケーションの実装では図2(a)のフルメッシュ通信となりますが、もちろん、図2(b)のMCUや図2(c)のSFUを用いた実装も可能です。また、前述したゲートウェイを用いれば、プロトコルが異なるテレビ会議やWeb会議との相互接続も可能となります。WebRTCにおけるシグナリングプロトコルとしてはWebSocketと呼ばれる技術が使われることが多いですが、前述のSIPを使っても問題ありません。メディア転送にはRTPのセキュリティ機能を高めたSRTP (Secure Real-time Transport Protocol)<sup>(8)</sup>が使われます。WebRTCではメディア転送に関する規定以外に、テキスト等のデータを送信する方法も規定されており、例えば映像コミュニケーションとチャットを組み合わせたといった使い方も実現できます。映像符号化のエンジンはWebブラウザの実装に依存しますが、主にH.264/AVCやVP8、VP9といったものが使われ、ブラウザが対応しているならばセッション確立時に指定可能です。このように、WebRTCによる通信を例にしても、原則的には図3に示す要素技術で構成されていることが分かります。皆さんが普段ブラウザで利用されるWeb会議サービスにおいても、実はこのWebRTCが使われたものが数多く存在しています<sup>(9)</sup>。

## 4. おわりに

本稿では映像コミュニケーションの歴史と基本技術を解説してきました。映像コミュニケーションはネットワーク技術や映像符号化技術の進展により、映像が高精細化し臨場感や没入感が向上するとともに、昨今の社会情勢からその利用が急速に普及しつつあります。一方で、中には「やっぱり面と向かって話した方が伝わりやすいよね」と感じている方も多いと思います。筆者が最後にお伝えしたいのは、映像コミュニケーションは対面によるコミュニケーションの単なる置き換えではないという点です。今回は誌面の都合上詳しくはお伝えできませんでしたが、例えば音声認識や機械翻訳といった技術と連携することで、会議における発話内容を取得し自動的に議事録を作成する、また、英語で発言された内容を日本語に変換するといったことが可能となり、これらはサービスとしても提供され始めています。このように、映像や音声データをビッグデータとして蓄積し活用することで、実空間では実現できなかった新たな価値が提供されるよう、映像コミュニケーションは今後も進化していくと考えています。

最後になりますが、本稿をきっかけに、読者の皆さんがこれまで以上に映像コミュニケーションを活用することで柔軟な働き方を実現し、結果としてワークライフバランスの充実やQoL (Quality of Life) の向上につなげて頂ければ幸いです。

## 文 献

- (1) Recommendation ITU-T H. 323, "Packet-based multimedia communications systems, Infrastructure of audiovisual services-Systems and terminal equipment for audiovisual services," 2003.
- (2) IETF, "Real-time communication in WEB-browsers (rtcweb)," <http://datatracker.ietf.org/wg/rtcweb/>
- (3) W3C, "WebRTC 1.0: Real-time communication between browsers," <https://www.w3.org/TR/webrtc/>
- (4) ISO/IEC 14496-10, "Coding of audio-visual objects-Part 10: Advanced video coding," Recommendation ITU-T H. 264, "Advanced video coding," 2003.
- (5) ISO/IEC 23008-2, "High efficiency coding and media delivery in heterogeneous environments-Part 2: High efficiency video coding," Recommendation ITU-T H. 265, "High efficiency video coding," 2013.
- (6) IETF RFC 3261, "SIP: Session initiation protocol," 2002, <https://tools.ietf.org/html/rfc3261>
- (7) IETF RFC 3550, "RTP: A transport protocol for real-time applications," 2003, <https://tools.ietf.org/html/rfc3550>,
- (8) IETF RFC 3711, "The secure real-time transport protocol (SRTP)," 2004, <https://tools.ietf.org/html/rfc3711>
- (9) "SMART communication & collaboration cloud," [https://www.nttbiz.com/solution/vcs/service/smart\\_cc/](https://www.nttbiz.com/solution/vcs/service/smart_cc/)

(2020年5月20日受付 2020年6月3日最終受付)



かまむら しょうへい  
鎌村 星平 (正員: シニア会員)

平16 早大・理工・電子・情報通信卒。平18 同大学院国際情報通信研究科修士課程了。平25 同大学院博士課程了。平18 日本電信電話株式会社入社。以来、IP・光伝送ネットワークの研究開発に従事。現在、同社ネットワークサービスシステム研究所主任研究員。博士(国際情報通信学)。