

## ITU-T におけるオーディオビジュアル通信の標準化

大久保榮<sup>+</sup>

オーディオビジュアル通信は利用者が人間で音声になんらかの画像が加わったリアルタイム通信であり、テレビ電話、テレビ会議はその代表例である。ITU-T (国際電気通信連盟-電気通信標準化部門) ではこの 20 年間、多様なネットワークを利用したオーディオビジュアル通信システムの標準化を進めてきた。本報告では、H.300 シリーズ標準化の経過、各種ネットワークの特性を述べ、H.320, H.323, H.324/M のシステムについてシステム構成を概説し、今後の方向に触れる。

## ITU-T Standardization on Audiovisual Communications

Sakae OKUBO<sup>+</sup>

Audiovisual communication is defined as a real-time communication where human users communicate using audio plus some pictorial information. Videophone and videoconferencing are its typical instances. ITU-T (Telecommunication Standardization Sector of International Telecommunication Union) has been developing standards in this area these 20 years. This report describes the history of H.300-series standards, characteristics of various types of network and presents the architecture of H.320, H.323 and H.324/M systems. Future direction of the system standardization is also touched upon.

### 1. はじめに

オーディオビジュアル (Audiovisual) 通信システムは、利用者が人間で、音声になんらかの画像が加わったリアルタイム通信システムの総称である。音声のみの電話が発展した通信システムとして開発が進められている。

「なんらかの画像」の種類により、具体的には次のようなオーディオビジュアル通信システムが考えられてきた。

- ・ テレビ電話 (videophone) : 動画を用いる原則として一人対一人の通信。
- ・ テレビ会議 (videoconferencing) : 動画を用いる原則として複数対複数の通信。
- ・ オーディオグラフィック会議 (audiographic teleconferencing) : 主として自然画像、ファクスなどの静止画を用いる通信。

そのほかに、テレビ会議中に説明用のスライドを提示するなど、動画と静止画を組み合

わせた通信システムも実用されている。

ITU-T は ITU (International Telecommunication Union, 国際電気通信連合) の電気通信標準化部門で、電気通信に関わる広範な分野の国際標準化を進めている[1]。

本報告では、テレビ電話、テレビ会議システムを中心に、ITU-T における標準化の経過を紹介し、今後の方向を展望する。

### 2. 標準化の展開

相手の姿を見ながら電話するというテレビ電話のコンセプトは、1890 年に電話が発明されて間もなく人々の思いつくことになった。1930 年にはテレビ電話の実験がニューヨークで行われ、日本でも 1935 年の横浜復興記念大博覧会で、相手を表示する画面を見れば正面から撮像される視線一致形のテレビ電話が展示された[2]。テレビジョン技術の応用としてテレビ放送ではなくテレビ電話が最初に取り上げられたことも興味深い。

しかし、実用的なテレビ電話には、デバイス技術、広帯域通信技術の進展が必要で、1960

<sup>+</sup>通信・放送機構早稲田リサーチセンター  
Waseda Research Center, Telecommunications  
Advancement Organization of Japan

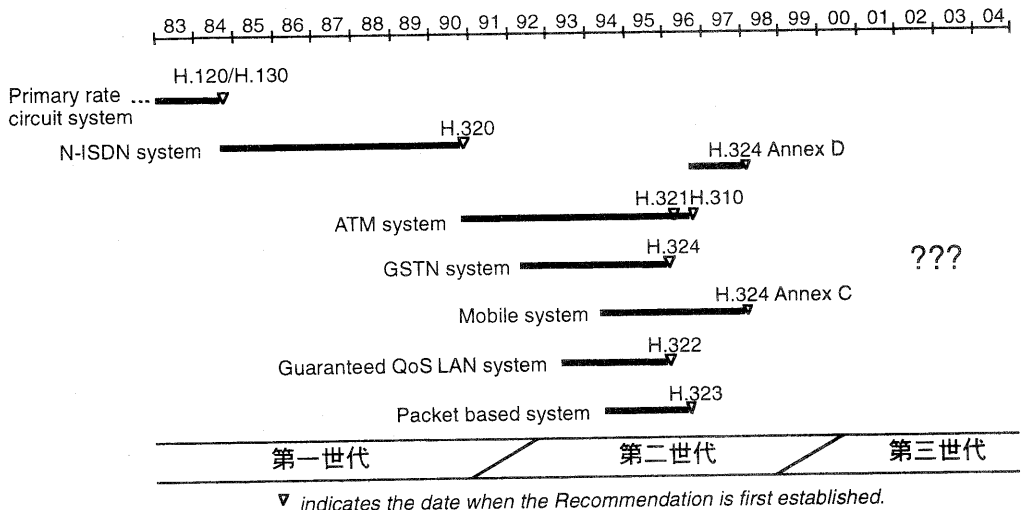


図1 ITU-TにおけるH.300シリーズ標準化の歴史

年代後半以降ようやくシステム開発が始まる。電話需要の充足に伴ってテレビ電話が着目され、1960年代後半から1970年代始めに、各国で研究、開発、商品化が行われた。しかし、テレビ電話は人々に受け入れられるに至らず、その技術はビジネスツールとしてのテレビ会議に継承された。

1980年代半ばには帯域圧縮符号化技術の進展とネットワークのデジタル化により、本格的なテレビ会議システム構築が展開された。このことが異種システム間の相互接続、国際標準化の必要性認識を生み、ITU-T（1993年3月以前はCCITT）での標準化が活発化した。

CCITT, ITU-Tにおけるオーディオビジュアル通信システム標準はH.300シリーズとして制定されている（注）。その展開を図1に示す。この図では、標準化作業の期間と初版勧告のできた時期を示している。その後も、機能拡充のための改版、付属資料の追加が継続されていることに注意が必要である。

注 - ITU-Tで作られる国際標準 Recommendation（勧告）と呼ばれ、<アルファベット1文字 + "." + 数字3桁ないし4桁の形>で表記される。勧告によっては数字のあと更に"<" + 数字1桁>がつく場合もある。最初のアルファベット1文字がおおよそその分野と作業グループを表している。

ISDN上のシステムを規定するH.320は、1990年12月に成立した。これは、システムの全体構成、システム構成要素とその相互関係を規定した最初の標準であることから、H.300シリーズの第一世代と位置づけられる。

その後、H.320を基にして、H.324（1996年3月）、H.310（1996年11月）、H.323（1996年11月）がGSTN（General Switched Telephone Network、通常の電話網）、ATM、LANなど多様なネットワークに対応した第二世代システム標準として作られた。H.324の付属資料Dは、H.324本文のシステムを基にして移動網上のシステムを規定したもので、H.324Mシステムとも呼ばれる。

図1では第三世代を記しているが、これは仮想的なもので、現時点で具体化しているわけではない。しかし、おそらく、下位のネットワークに依存しない汎用的な通信システム標準として進化するであろうこと、H.323を基にして展開してゆくであろうことが予想される。

### 3. ネットワークの特性

ネットワークはデジタル情報の転送が可能であれば、マルチメディア通信システムに利

用することができる。現在サービスが提供されている各種ネットワークとその転送速度、回線/パケット交換の別、伝送誤りなどの特性を図2に示す。この図にはH.300シリーズのオーディオビジュアル通信システムとの対応関係も示してある。

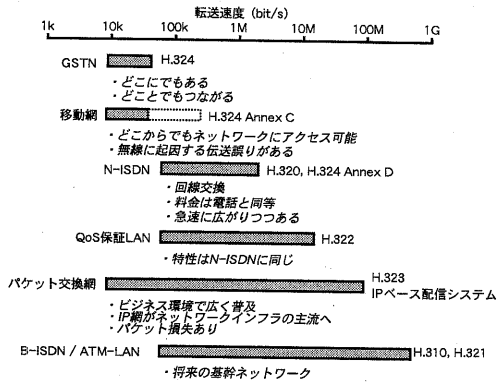


図2 各種ネットワークの特徴

#### 4. システム構成

H.300シリーズのオーディオビジュアル通信システム構成を図3に示す。大きくは端末装置、ネットワークと多地点間会議制御ユニット(MCU: Multipoint Control Unit)からなる。MCUは1対1通信のための端末装置複数をつなぎこんで多地点間の通信を実現するためのもので端末装置とは別の勧告で規定さ

れている。しかし、端末装置とMCUはネットワークに対しては同じ立場にある。図3には各要素の仕様を定めた勧告番号を合わせて記している。「H.26x」の表記は、H.261, H.262, H.263のような一連の勧告群を意味している。

端末から情報を送り出す側についてももう少し詳しく見てみる。表現メディアとしての音声(マイクロフォン出力)、映像(カメラ出力)はもとはアナログ情報であるが、符号化によりそれぞれのデジタルストリームを構成する。図2でコーデック(codec)と記されているのは符号器(coder)と復号器(decoder)を総称したものである。「データ」は、それ以外の表現メディア情報で、ホワイトボード画面情報や静止画、パソコンデータなどがその例である。システム制御情報は、端末・端末間でやりとりしてシステムをユーザの望むように動作させるための情報と、端末・ネットワーク間でやりとりして呼の接続、切断など行う呼制御情報の2種類がある。

これらの表現メディア情報は、マルチメディア多重化し一つのストリームとしてネットワークで運ばれる。その際、音声と映像を同期させてリップシンク(lip sync, 画面上の人物の口から音が出ている印象を与えること)を確保したり、画面内容に同期した文字を重ね合わせるなど、メディア間の同期をとることも「マルチメディア多重・同期」の機能である。図3の音声系に入っている「遅延」は、

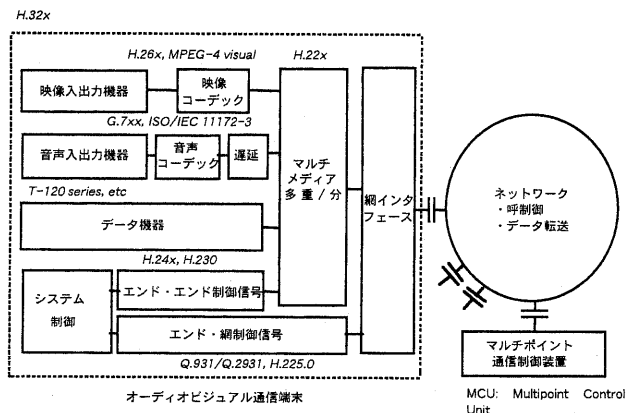


図3 一般的なシステム構成

通常映像信号処理遅延が音声信号処理遅延より大きく、リップシンクのために必要である。

マルチメディア多重情報は、利用するネットワークに応じたアダプテーション、インタフェースを介してデジタルチャネルに流れて行く。呼制御情報はこの段階でマルチメディア通信情報と多重化される。

情報を受信する側は、デジタルチャネルから逆の方向に信号が流れて行き、上記の「多重」を「分離」に、「符号化」を「復号」に置き換えればよい。

## 5. システム標準

ここでは H.300 シリーズの代表的なシステムを概説する[3]。

### 5.1 H.320 システム

N-ISDN 利用によるオーディオビジュアル通信システムと端末装置は、ITU-T 勧告 H.320 に規定されている。この勧告は、システム構成を記述するとともに、その構成要素と参照すべき勧告、各種端末タイプ、通信フェーズ、電話や他の H.300 シリーズシステムとの相互接続に関する要求条件を定めている。高品質のテレビ会議は 6B チャネル (384 kbit/s) 程度で、通常のテレビ会議は 2B チャネル (128 kbit/s) で、テレビ電話では 2B あるいは 1B チャネルで利用されるのが一般的である。なお、標準の上ではテレビ会議とテレビ電話の区別はなく、両者の端末間で相互通信可能である。

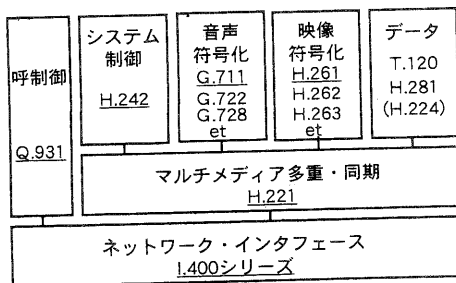


図 4 H.320 システムのプロトコルスタック

ポイント・ツ・ポイントの基本システムについては、1990 年 12 月に標準が成立し、既に 100 万台前後の製品が主にテレビ会議用として市場に導入されている。このため、H.323 を始め他のネットワークシステムに対し、N-ISDN 上の H.320 端末との相互接続は重要な要求条件となっている。

H.320 システムのプロトコルスタックを図 4 に示す。ここで下線を付した要素は必須でその他はオプションである。

### 5.2 H.323 システム

#### 5.2.1 標準化の背景

オフィス環境に広く使われている LAN は、イーサネットのように、トラフィックに応じたパケット到着遅延やパケット損失が生じうる。これは LAN がもともとコンピュータ間通信用に開発されたもので、電話網に根ざす N-ISDN とは設計思想が異なるためである。この種の LAN は、オーディオビジュアル通信の視点からは、伝送チャネルのサービス特性 (QoS) が他の利用者の存在に左右されるという意味で QoS 非保証形 LAN と呼ばれる。ネットワークを使ううえで H.320 端末が有する以上の機能が必要とする。

QoS 非保証形 LAN の上で動作し、かつ N-ISDN 上の H.320 端末との相互通信が可能なオーディオビジュアル通信システム H.323 は、たとえ通信品質が少し劣化したとしても既に設置済みの LAN が使えることから強い市場ニーズが予見され、1995 年以降集中した検討が行われた結果、1996 年 11 月に最初の勧告が制定された。標題は "Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service" であった。

#### 5.2.2 システム構成

H.323 システムの構成を図 5 に示す。端末 (T)、ゲートキーパ (GK)、ゲートウェイ (GW)、MCU である。ゲートキーパは端末、ゲートウェイ、MCU が LAN にアクセスする際のアドレス変換、アクセス制御、帯域管理

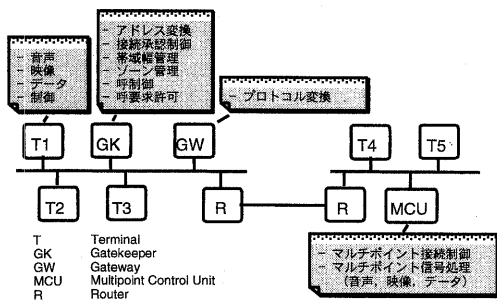


図 5 H.323 システム構成 (ゾーン)

などの機能を持つ。ゲートウェイは LAN の外側にある端末との相互通信のためのプロトコル変換を行う。MCU は、複数端末あるいはゲートウェイが参加するマルチポイント通信を可能とし、機能的には通信の制御を実行する MC (Multipoint Processor) とオーディオビジュアル信号の処理を実行する MP (Multipoint Processor) から構成される。LAN では、N-ISDN のようなポイント・ツ・ポイント接続をベースとするマルチポイントのほか、LAN に固有のマルチキャスト機能を活用したマルチポイント通信が可能なることから、このように MC/MP 機能を分離して一般化が図られている。

プロトコルスタックを図 6 に示す。トランスポートレイア以下の LAN 自体は勧告の対象外で、その上のマルチメディア情報のパケット化、同期の機能を提供する H.225.0 が新たに定められた。システム制御に関わるチャネルは、他のシステムと同様に H.225.0 呼制御チャネルと H.245 端末間制御チャネルがあるほか、H.323 システムではさらに LAN へのアクセスを制御する RAS (Registration, Admission, Status) チャネルが定義されている。LAN の代表例は、ネットワークレイアに IP、再送を含まない unreliable transport に UDP、再送を含めた reliable transport に TCP を用いるものであるが、H.323 システムではその他のプロトコルを用いる LAN でもかまわない。

H.225.0 は、RTP/RTCP (Real Time Protocol/Real Time Control Protocol) を用

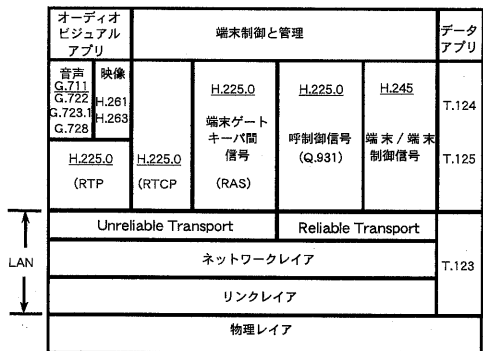


図 6 H.323 プロトコルスタック

いて音声、映像、データおよび制御ストリームのパケット化とタイムスタンプによる同期化を行い、メッセージとプロトコルを規定する。メッセージの中には、同一の LAN 上端末間、あるいはゲートウェイを介した LAN 端末 H.323 と回線交換網端末 (例えば H.320) の間を交換接続するための呼制御メッセージが含まれ、Q.931 のメッセージが用いられる。

### 5.2.3 通信の流れ

H.323 は、T/GK/GW/MCU の配置により多様なシステム構成を許容するが、一例として端末の両者が同一の GK に登録されている構成の場合の通信は、次のように進んで利用者がサービスを受けるに至る。なお所属する GK の探索は通信に先立って行われ、その手順も標準に定められている。

#### 1) フェーズ A : H.245 制御チャネルの設定

H.323 はパケット交換網における呼設定の手順を提供していることに特徴があり、図 7 にそのためのメッセージ交換を示す。

- a) 通信を開始しようとする発呼端末は、まず RAS チャネルを通じ、使用帯域幅などのパラメータをつけて GK に接続要求メッセージ ARQ を送る。
- b) GK は確認(ACF)を返す。その際、次の H.225.0 呼制御メッセージ転送に用いるチャネルのトランスポートアドレスをつける。
- c) 発呼端末は、上記トランスポートアド

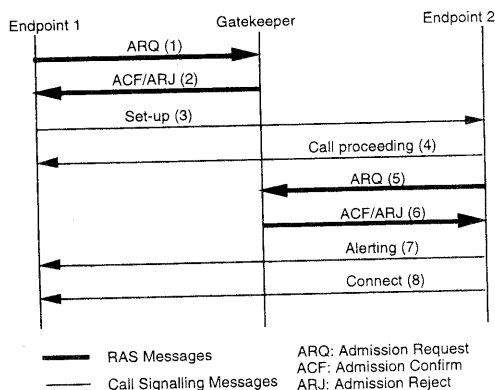


図 7 H.323 システムのフェーズ A-呼設定

レスにより、H.225.0 呼制御メッセージチャンネルを通じて、相手端末に SETUP メッセージを送る。このとき H.245 メッセージ用制御チャンネルのトランスポートアドレスが付随する。

d) 着呼端末に応答の用意があれば、RAS チャンネルを通じ GK に接続要求 ARQ を出す。

e) GK が接続を許可すれば、H.225.0 呼制御メッセージチャンネルを通じ CONNECT メッセージを発呼端末返す。その際、次の H.245 メッセージ用制御チャンネルのトランスポートアドレスが付随する。

なお、LAN 内の帯域を予約することも可能であるが、これはフェーズ A に先だって行われ、その方法は H.225.0/H.323 勧告の対象ではない。

## 2) フェーズ B：能力交換

フェーズ A で定められた H.245 制御チャンネルを通じ、両端末間で H.245 メッセージによる能力交換が行われる。

## 3) フェーズ C：オーディオビジュアル通信

能力交換に基づき、使用する通信モードが決定され、H.245 の手順に従って必要な論理チャンネルが設定される。音声、映像は、RTP パケットにより転送される。

## 4) フェーズ D：通信中の帯域変更

何らかの理由で通信中に使用帯域（ビットレート）の変更が生じると、RAS チャンネルによるパラメタの伝達、H.245 制御チャンネルによる制御メッセージのやりとりにより、該当論理チャンネルの解除、新たな論理チャンネル設定の手順が実行される。

## 5) フェーズ E：通信終了

すべてのチャンネルを解除する。

## 5.2.4 H.323 システムの進化

H.323 システムとその関連要素標準の進化を表 3 に示す。H.225.0 は呼制御とメディアの packets 化を定めた標準で H.323 と対になって改訂が行われている。インターネットサービスとそのネットワークである IP パケット交換網の急速な発展に合わせ、H.323 システム標準も急テンポの拡充が図られていることがわかる。1998 年 2 月の第 2 版からは標題も "Packet-based multimedia communications systems" と対象ネットワーク、対象アプリケーションの双方が一般化されている。H.323 が規定する範囲はトランスポート層より上の部分で、LAN プロトコルはどのようなものであっても構わないが、現在ではほぼ IP ネットワークとその上の TCP および UDP に集約されている。

H.323 システムは、当初一つのゾーン（論理的な閉域網、図 5 参照）の中で動作することに限られていたが、H.225.0 Annex G が定めるゾーン間の通信制御プロトコルにより、課金情報の受け渡しを含め、二つのゾーンにまたがって、さらにはグローバルなネットワークで動作するシステムへと進化した。

データ通信ネットワークでは、電話網と違って呼接続の概念はなく、データグラム（パケット）を必要なときに送り出すのが基本となっている。H.323 が多くの関心を集めている理由は、IP ネットワークに対する呼制御プロトコルを最初に定義したことにある。マルチメディア通信のサブセットとしてモノメディア通信、特に音声のみの電話通信があり、H.323 はそれを考慮して最初から音声のみが

表 1 H.323 標準の進化

承認時期	H.323	関連勧告
1996/11	V1	H.225.0 V1
1998/02	V2 (fast connect の追加など) Annex B: 階層符号化の利用 Annex C: ATM 利用	H.225.0 V2 H.235: セキュリティ H.246: H.323-H.320 相互接続
1998/09	Annex D: リアルタイム FAX	H.332: 大規模会議 H.450.1, 2, 3: 付加サービス
1999/05	Annex E: UDP 上の呼制御 Annex F: 黒電話 SET	H.225.0 Annex G: ドメイン間通信 H.341: MIB H.450.4, 5, 6, 7: 付加サービス
1999/09	V3 (実装に基づく明確化など)	H.225.0 V3
2000/02	Annex G: テキスト SET	H.246 Annex C: ISUP との接続 H.450.8: 付加サービス
2000/06		H.248: Media GW 制御プロトコル
2000/11	V4 (下記 Annex による機能拡張) Annex J: セキュリティ SET Annex K: http によるサービス制御 Annex L: スティミュラス信号 Annex M1: QSIG トンネリング Annex M2: ISUP トンネリング	H.225.0 V4 H.235 V2 H.450.9: 付加サービス

ISUP: ISDN User Parts of Signalling System number 7  
 QSIG: Signalling between the Q reference points  
 SET: Simple Endpoint Types (単機能端末)  
 V: Version (版)

必須、映像はオプションとして扱っている。すなわち巨大なビジネスである電話を IP ネットワーク上で展開する (VoIP, Voice over IP) プロトコルとして H.323 は着目されている。

2000 年 6 月に承認される H.248 は、IP ネットワークと既存電話網との相互接続に必要なゲートウェイの制御プロトコルを定めたものである。これにより、私設網 (LAN) から公衆網への接続が実現され、電話のサービスにおいても IP がネットワークインフラとなる道が開かれることになる。

また、テレビ会議の分野では、ISDN と IP 網両方のインタフェースを持つ H.320/H.323 複合形を実現する製品が提供されており[4]、ここでもネットワークの IP 化への対応が進められている。

### 5.3 H.324M システム

既存電話網を用いたオーディオビジュアル通信は、一方で音声符号化、映像符号化技術の進歩による信号源の低ビットレート化と、他方でモデム技術の進歩による通信チャネル

の高速化があいまって、1992 年頃から関心が高まってきた。これに伴って、ITU-T における標準化作業が進展し、1996 年 3 月、システム勧告 H.324、映像符号化勧告 H.263、音声符号化勧告 G.723.1、マルチメディア多重分離勧告 H.223 が成立した。

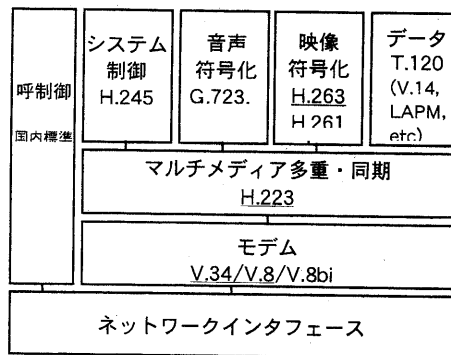


図 8 H.324 システムのプロトコルスタック

H.324 システムのプロトコル構成を図 8 に示す。電話網インタフェースの上にはパケット多重形のマルチメディア多重 H.223 が適用される。この方式は、多重化のための MUX

レイア、情報メディアストリームへのアダプテーション AL レイアから構成される。AL には、主に制御あるいはデータ情報用の AL1、音声情報用の AL2、映像情報用の AL3 が定義され、メディア情報の特性に適應するため、誤り検出、シーケンス番号、再送などの点で異なった機能を有する。H.223 の特徴の一つは、低ビットレートで特に問題となるパケット化遅延を、一つの MUX パケットの中に複数メディア情報を入れることで解決している。

電話網と同様モバイル網が提供できる回線速度は現状では 64 kbit/s 以下のビットレートである。そこで H.324 システムをモバイル網に適用することが検討され 1998 年 2 月、Annex C (H.323/M とも呼ばれる) が制定された。伝送誤りへの対策を強化するため、図 9 に示すように H.223 にいくつかの誤り耐性ツールが新たに用意されている。これらはレベルが高くなるほど強力で、上のレベルはその下のレベルの全ての機能を包含する階層的仕様になっていて、H.324/M は少なくともレベル 0 で既存の H.324M と相互通信できる。

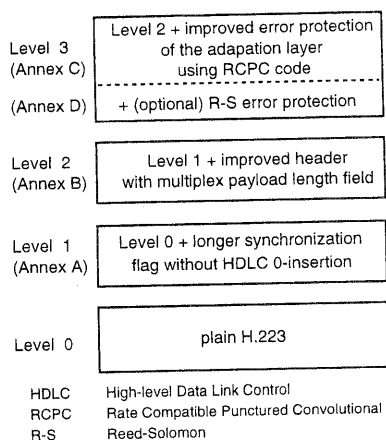


図 9 H.223 誤り耐性ツール

## 6. おわりに

ネットワーク環境の変化は急である。電話中心の回線交換ネットワークからデータ中心の IP パケット交換網へと大きく動いていて、電話すらも IP パケット交換網で送られることが進められている。標準化活動に参加して

いる機関も従来の電話関連企業からインターネット関連企業へと変わりつつある。また ITU-T 単独ではなく、IETF と協調しつつ共通の標準を作る試みも行われている。このような状況で、ITU-T におけるオーディオビジュアル通信システムの標準化も転換点にきているといえよう。

これまで多様なネットワークに最適なシステム解を求めて多数の H.300 シリーズ標準ができたが、利用者から見ればネットワークに関わらず同じサービスに見えるわけで、ネットワークには依存しないアプリケーションプロトコルの視点からの標準化が必要であろう。その際、既存のシステムとの相互通信を確保することが大切で、ゲートウェイによる解、切り替え可能な端末による解、あるいはその組み合わせなど、適切な解を用意しなければならない。

ここ 20 年間の活動で、オーディオビジュアル通信システムのインフラは整備されてきた。もう一つの標準化の方向は、遠隔教育、遠隔医療などよりアプリケーションに踏み込んだ領域でのサービス性の向上であろう。

## 参考文献

- [1] ITU-T の活動およびその標準は、<http://www.itu.ch/>で得られる。その多くは TTC で日本標準となっている。<http://www.ttc.or.jp/>参照。また、作業過程の文書は<http://standard.pictel.com/ftp/avc-site/>および<http://lbc-site/>で得られる。
- [2] テレビジョン技術史編集委員会: "テレビジョン技術史", pp.70-74 (Mar. 1971)
- [3] H.300 シリーズの標題
  - H.310 Broadband and audiovisual communication systems and terminals
  - H.320 Narrowband visual telephone systems and terminal equipment
  - H.321 Adaptation of H.320 visual telephone terminals to B-ISDN environments
  - H.322 Visual telephone systems and terminal equipment for local area networks which provide a guaranteed quality of service
  - H.323 Packet-based multimedia communications systems
  - H.324 Terminal for low bit rate multimedia communication
- [4] <http://www.vtel.com/newsinfo/news/genpress/mnpra50.htm> など