

国際標準に準拠したインターネットテレビ電話構成法の検討

福田 浩司 安藤 大 林 泰仁 小谷野 浩

日本電信電話株式会社
NTT ヒューマンインタフェース研究所

現在、IP テレフォニーをはじめ IP ネットワークへの音声や画像の統合が盛んに行われインターネット上で双方向のリアルタイム音声画像通信を行うインターネットテレビ電話が注目を集めている。インターネットテレビ電話は、これまで製品ごとに独自の通信方式を採用していたため相互接続性がなかった。しかし、インターネットテレビ電話の国際標準方式として、ITU-T H.323 が勧告され、相互接続性の問題は解消される方向にむかっている。そこで、インターネット電話をより一層の普及へ導くため、標準化対応、画像通信機能を持つ「VocaLink-Soft/V」を試作した。「VocaLink-Soft/V」は、画像符号化方式にパケット損失に強い H.263 Annex N (Newpred 方式) を実装している。本稿では、「VocaLink-Soft/V」の構成法及びネットワーク帯域・パソコンのスペックに応じたパラメータ設計方法について述べる。

Architecture of an Internet Visual Telephone System based on International Standard

Koji FUKUDA, Dai ANDO, Yasuhito HAYASHI and Hiroshi KOYANO

NTT Human Interface Laboratories

Recently various sorts of audio and video services integrated over IP networks have been developed. As one of such services, Internet videophone is highlighted. At the early stage, interconnectivity between Internet videophone products was not guaranteed. However, it is globally agreed that an ITU-T standard, H.323, is used for Internet videophone, and the issue of interconnectivity is being resolved.

To achieve more wide use of Internet videophone, we developed an Internet videophone software, VocaLink-Soft/V, supporting the international standard. VocaLink-Soft/V also supports an international standard for video codec, H.263 Annex N, which is meant to compensate for packet loss caused by network congestion.

This paper describes the architecture of VocaLink-Soft/V and parameter design in accordance with network bandwidth and PC capability.

1. はじめに

現在、IP テレフォニーをはじめ IP ネットワークでの音声や画像データの利用が盛んに行われ、インターネット上で双方向のリアルタイム音声画像通信を行うインターネットテレビ電話が注目を集めている。インターネットテレビ電話は、これまで製品ごとに独自の通信方式を採用していたため相互接続性がなかった。しかし、インターネットテレビ電話の国際標準方式として、ITU-T H.323 が勧告され、相互接続性の問題は解消される方向に向かっている。

筆者らは、従来から双方向リアルタイム音声通信方式を検討しており、インターネット電話ソフト「VocaLink-Soft」を試作してきた^[1]。そこで、インターネット電話の利用価値を向上させるため、「VocaLink-Soft」に標準化対応及び画像機能を追加した「VocaLink-Soft/V」を試作した。また、画像符号化時のパラメータ設定を、使用するネットワーク及び使用パソコンの処理能力に応じて適切に設定する方法について、検討を行い実装した。本稿では、「VocaLink-Soft/V」の構成法及びパラメータ設計法について報告する。

2. 国際標準方式 ITU-T H.323

ITU-T 勧告 H.323 に準拠したインターネットテレビ電話は、主に図1のようなプロトコル構成を取る。図1の斜線部は、「VocaLink-Soft/V」の構成に関わるプロトコル部分であり、その主要プロトコル部分について機能概要を示す。

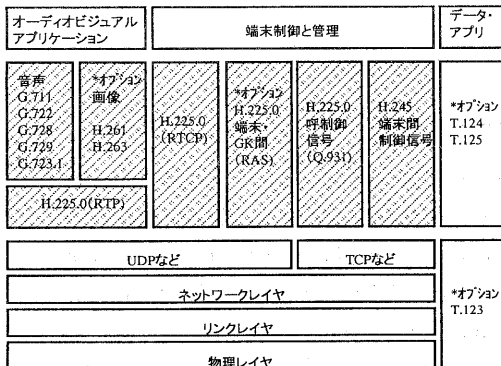


図1 H.323 プロトコル構成図

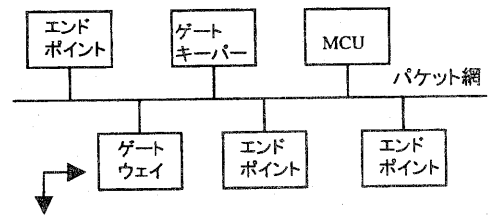
(a) RAS プロトコル

H.323 システムの構成要素を図2に示す。RAS プロトコルは、主として、ゲートキーパーと呼ばれるサーバとエンドポイント(端末)間の通信用プロトコルとして規定されている。このRASメッセージの通

信では、UDP プロトコルが用いられる。

通常は、エンドポイントが起動時にゲートキーパーに登録を行い、ゲートキーパーがエンドポイントや通信の管理/制御を行う。ゲートキーパーの具体的機能としては、アドレス変換、アクセス制御、帯域管理等の機能がある。また、エンドポイントは、通信開始前に相手エンドポイントのアドレスをゲートキーパーに問い合わせることも可能である。

なお、RAS プロトコルは、H.323 においてはオプションである。



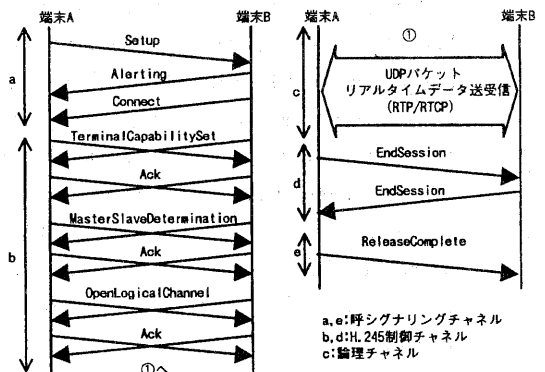
回線交換網

図2 H.323 システムの構成要素

(b) 通信制御

エンドポイント間の通信開始時のセッション確立などを制御する部分においては接続の確実性が要求されるためTCP プロトコルを使用する。エンドポイント間の接続シーケンス例を図3に示す。

まず第一のTCP コネクションは、H.225.0(Q.931)により規定される呼シグナリングチャネル(図3:a)を使用し、発呼端末と被呼端末との間の呼を確立する。次に、第二のTCP コネクションは、H.245により規定されるH.245制御チャネル(図3:b)を使用し、会議制御及び能力交換等を行う。その後、論理チャネル(図3:c)を開きリアルタイムデータ通信状態に移移する。切断処理は、H.245制御チャネル(図3:d)を解放した後、呼シグナリングチャネル(図3:e)を解放する。



a, e: 呼シグナリングチャネル
b, d: H.245制御チャネル
c: 論理チャネル

図3 エンドポイント間の接続シーケンス例

(c) リアルタイムデータ制御

論理チャネル上で行われる音声及び画像等のリアルタイムデータ通信は、H.225.0 で規定される RTP/RTCP プロトコルを使用する。RTP/RTCP プロトコルは、UDP プロトコル上で使用し、リアルタイム性を確保しつつ通信制御を行う。

(d) 音声能力

音声の符号化方式は、必須である G.711 とオプションの低レート符号化方式 G.723.1, G.729 などが規定されている。

(e) 画像能力

勧告 H.323 では、画像能力はオプションであるが、画像能力を持つ場合の符号化方式は、必須である H.261 とオプションの H.263 が規定されている。

3. 標準化対応を考慮した検討内容

(a) RAS プロトコル

従来の「VocaLink-Soft」では、通信相手選択を、一般的な Web ブラウザから行える電話帳サーバ^[2]があった。また、他にエイリアス変換機能に独自のアドレスサーバを設け、変換していた。そのアドレスサーバは、エイリアス変換機能のほかに転送、ボイスメール等の付加サービス機能も提供していた。

本試作では、これら独自サーバへの IF 機能に加え、H.225.0(RAS)を実装し、サーバとして、H.323 準拠のゲートキーパーを利用可能とした。

(b) 通信制御プロトコル

従来の「VocaLink-Soft」では、呼制御部分に独自の通信制御プロトコルを使用していた。独自の通信制御プロトコルは、シンプルで接続時間が短いなどの利点があったが、本試作では標準化方式の相互接続性及び機能拡張性を優先し、H.323 エンドポイントにおける 1対1 通話に必要な標準化プロトコル H.225.0(Q.931), H.245 を実装した。

(c) リアルタイムデータ制御プロトコル

従来の「VocaLink-Soft」では、UDP 上にリアルタイムデータ用プロトコルを実装することが有効であると考え、RFC1889, RFC1890 で規定された RTP/RTCP プロトコルを使用していた。これは内容的には H.225.0(RTP/RTCP) で規定された方式と同等であったため、変更の必要はなくそのまま利用した。

(d) 音声符号化方式

音声符号化には、必須である G.711(64kbps)、低

レート符号化方式としてインターネット電話でよく使用される G.723.1(6.3kbps), G.729(8.0kbps) を実装した。また、DualSpeech 方式^[3](7.4kbps) も非標準ではあるが低レートで高音質かつパソコンにおける処理量が少ないことから実装した。

(e) 画像符号化方式

インターネットなど必ずしも帯域を保証しないネットワークではパケット損失が起こりやすい。音声データはパケット損失があっても次の音声データへの影響は少ないが、画像データは、1つのフレームにエラーが生じると次フレーム以降にもエラーが伝播される。安定した画像通信を維持するためには、パケット損失に強い画像通信方式を採用する必要がある。そこで、エラーの伝播を防止するため、NTT 他で ITU-T に提案し、H.263 AnnexN として勧告化された NEWPRED 方式^[4]を搭載した。また、対応画像サイズは、CIF サイズ(352×288)、QCIF サイズ(176×144)である。

(f) 使用環境に応じた符号化パラメータ設計

一般に、符号化画像データのデータ量は、ネットワークの帯域に比べ、極めて膨大な量になる。そのため、対象となるネットワークの帯域に応じ、符号化パラメータを適切に設定する必要がある。そこで、ユーザの利便性をも考慮し、14.4kbps 以下(音声のみ)、14.4kbps ~28.8kbps, 28.8kbps ~64kbps, 64kbps ~128kbps, 128kbps 以上(LAN)という5段階のネットワーク帯域を設定し、ユーザが設定したネットワーク帯域内で最も品質のよい符号化画像を作成する符号化パラメータを自動的に設定する仕様とした。

また、現在の VocaLink-Soft/V では、音声の圧縮/伸張および画像の圧縮/伸張とも、全てプログラムで処理を行う。そのため、使用するパソコンの処理能力によっては、同じネットワーク帯域であっても十分に処理できない場合が発生する。そこで、ネットワーク帯域と共に、使用パソコンの CPU スペックに応じて目標となるレートを定め、そのレートを実現する符号化パラメータを設定する仕様とした。図4に使用パソコンの CPU スペックおよびネットワーク帯域による設計目標のフレームレート及びビットレート例を示す。

パソコンの CPU 使用率が 100%ギリギリの状態では、バースト的に符号化データ量が増えた場合に、パソコンがハングアップする、または他のアプリケーションに影響を与える危険性があるため、おおむね 80%以下の使用率となるよう設計した。

| CPUスペック | | ペンティアム 200MHz | MMXペンティアム 233MHz | ペンティアム II 300MHz |
|----------|----------------|------------------|---------------------|---------------------|
| ネットワーク帯域 | フレームレート(f/s) | 0.1 | 0.2 | 0.3 |
| | ビットレート(kbit/s) | 25 | 25 | 25 |
| 28.8kbps | フレームレート(f/s) | 1.2 | 2.8 | 4.2 |
| | ビットレート(kbit/s) | 60 | 60 | 60 |
| 64kbps | フレームレート(f/s) | 1.2 | 4.3 | 8.4 |
| | ビットレート(kbit/s) | 130 | 150 | 220 |
| LAN | フレームレート(f/s) | | | |
| | ビットレート(kbit/s) | | | |

図4 PCスペックによる設計目標レート

4. 「VocaLink-Soft/V」の概要

筆者らは、上記検討結果に基づいて「VocaLink-Soft/V」の開発を行った。

「VocaLink-Soft/V」は、Windows95をOSとするアプリケーションプログラムで、ハードウェアとしては、DOS/V パソコンと汎用のサウンドカード、ネットワークカード、キャプチャーカードとカメラを使用し、特別なハードウェアは必要としない。

「VocaLink-Soft/V」のソフトウェア構成を図5に示し、「VocaLink-Soft/V」相互の通信中画面を図6に示す。

NEWPRED 方式は、エラーの伝播が行われていないことがわかる。

また、音声については、従来の「VocaLink-Soft」同様、一般電話と比較すると多少遅延があると感じるが、G.711, G.723.1, G.729, DualSpeech とも会話には特に支障はなく、良好な音質で通話が行え、充分利用できるレベルであった。

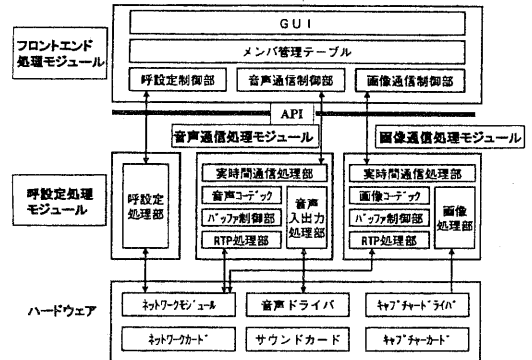


図5 「VocaLink-Soft/V」ソフトウェア構成図

5. 「VocaLink-Soft/V」の接続及び品質評価

(a) 「VocaLink-Soft/V」の相互接続性

相互接続性を確認するため、「VocaLink-Soft/V」と他社製のH.323準拠インターネットテレビ電話ソフトとの相互接続において、通信の開始終了および音声画像通信ができることを確認した。また、市販されているH.323準拠ゲートキーパーに対し登録や、ゲートキーパー経由の発呼ができることを確認した。

(b) LAN 環境での通信品質

10Base-T を使用し2台のパソコンを接続した LAN 環境で、VocaLink-Soft/V の品質を検証した。画像サイズとしては、QCIF サイズを用いた。LAN 環境の場合には、ネットワークに関わる帯域制限は特に設けないが、使用するパソコンのスペックにより、ある程度通信レートが決まってしまう。今回はCPUがペンティアム II/300MHz のものを用いた。この場合、実際の通信レートは約200kbpsであった。この環境では、画像の画品質・動きとも十分実用レベルであった(8~10枚/秒)。

また、NEWPRED 機能の効果を確認するために、ネットワーク内に大量の packets データを送出しネットワーク負荷を与え、故意に packets 損失を発生させた。図7に、NEWPRED 方式を使用した場合と使用しない場合のエラー出力後次画面を示す。図7から

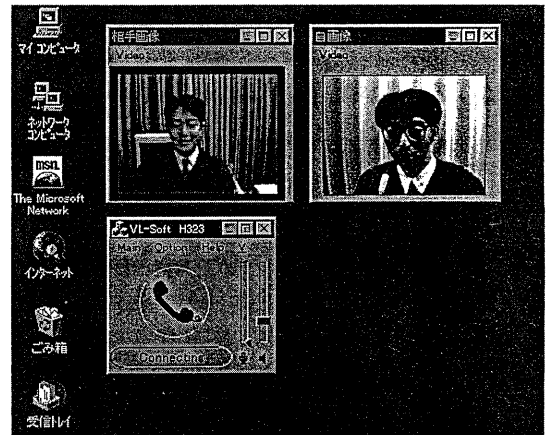


図6 「VocaLink-Soft/V」の通信中画面

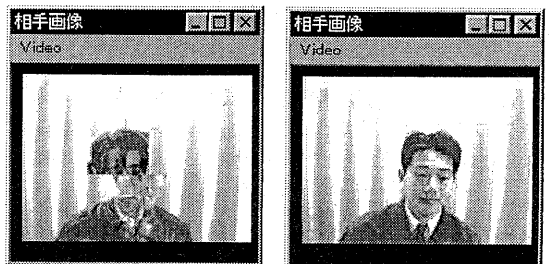


図7 エラー出力後の次画面

(c) 64kbps 設定時の通信品質

LAN 環境においては、非常に良好な品質で、音声画像通信が可能であった。しかし、端末として、外部の電話回線よりダイヤルアップ接続した端末を考えた場合、その部分で帯域が大きく制限されてしまう。現在のインターネット環境で、比較的ダイヤルアップ時の帯域が高いのは、ISDN を用いた場合である。そこで、帯域を 64kbps に制限した場合の品質について検証を行った。

インターネットテレビ電話においては、音声・画像データとも、データ圧縮後、IP パケット化して送信している。そのため、実際のビットレートは、符号化の圧縮レートより大きくなってしまふ。例えば、音声符号化に G.729 方式を選んだ場合、音声データは 8kbps に圧縮されるが、これに IP ヘッダ等が付加されてしまうため、実際のビットレートは約 15kbps になってしまう。画像データ用には、残りのビットレートが割り振られるが、これにも IP ヘッダ等のオーバーヘッドが付加されるため、ISDN 回線交換用テレビ電話端末 (H.320 端末) に比べると、若干伝送効率が低下してしまうことを考慮し、パラメータ設計を行った。

CPU がペンティアム II/300MHz のパソコンから、ISDN 用 TA を経由して、64kbps でダイヤルアップ接続を行い、LAN のパソコンとの間で、音声画像通信を行った。通信時のフレームレートは約 4~5 枚/秒であった (QCIF サイズ)。また、この時のビットレートは、おおむね 50~60kbps であった (音声含む)。画品質的にも、十分動画と感ぜられる品質であった。この状態の通信であれば、パソコンも安定して動作しており、設計したパラメータが有効に機能していると考えられる。

(d) 28.8kbps モデム使用時の通信品質

ダイヤルアップ環境として 64kbps の環境においては、良好な品質で、音声画像通信が可能であった。しかし、現実のインターネットの世界では、まだアナログモデムを用いてダイヤルアップ接続をしているユーザも多いと考えられる。そこで、28.8kbps のモデムを用いた場合についても検証を行った。

インターネットテレビ電話では、音声の実際のビットレートは約 15kbps になってしまう (G.729 使用時) ため、28.8k モデムを用いた通信環境では、画像データ用には残りの約 14kbps 程度しか使えない。このような条件下では、フレームレートを出すために画質を落とすと、実用上耐えられない品質となるため、送信枚数を少なくし、その分ある程度の画質を確保する方が有効と考えられる。

実際に 28.8k モデムを用いた環境では、フレームレートは数秒に 1 枚程度 (QCIF サイズ) であれば、安定した通信が可能であり (ペンティアム II

/300MHz) パラメータ設計の効果があつたと考えられる。このように 28.8kbps の場合は 64kbps の場合に比べて、動きがある程度制限されてしまふが、相手の顔を確認する程度には、十分使用可能であった。

また、実際のインターネット接続を利用した場合には、接続経路によっては、パケット損失などのエラーが頻繁に発生する場合があるが、NEWRED 方式の効果により、安定した画質を維持できることが期待できる。

(e) 32kPHS 使用時の品質

VocaLink-Soft/V は、IP ネットワーク用アプリケーションであるので、ネットワーク上で IP パケットが通れば使用可能である。そのため、PHS によるダイヤルアップ接続環境で使用すれば、無線インターネットテレビ電話としても活用できる。

現在の PHS データ通信は、PIAFS と呼ばれる 32kbps のデータ通信が可能である。ネットワーク構成としては、図 8 の構成で確認した。

32kbps の通信では、実効的には、28.8kbps のモデム利用時と同等の帯域であるため、画品質もほぼ同等であった。そのため、品質としては、相手の顔が確認できるというレベルであるが、このように、PHS による無線データ通信と組み合わせることで、有線の電話回線が無い場所から社内パソコンへの音声画像による連絡ができるなど、新しい使い方への応用が可能となる。

また、PHS データ通信は、現在 64kbps の方式が検討中である [5]。これが、実用化されれば、PHS 無線データ通信利用時の画品質も ISDN 利用時 (64kbps) 程度まで上がるため、応用範囲も広がるものと期待される。

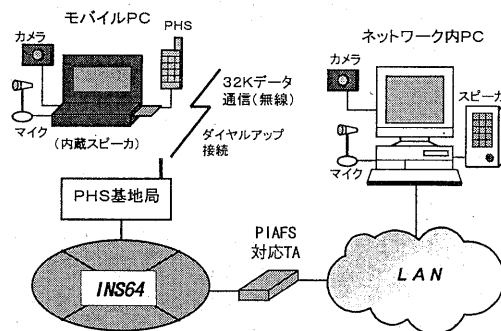


図 8 PHS を利用した場合のネットワーク構成図

6. 今後の H.323 標準化動向

ITU-T 勧告 H.323 は、現在 Version2 が勧告化されている。Version2 では、H.245 コマンドをカプセル化する高速セットアップ手順や、低レートの音声符号化に関する規定が追加されている。

H.323 のメインストリームは、Version2 でほぼ固まり、現在は、リアルタイムインターネット FAX 機能、より高速なファスターコネクト手順、機能限定音声専用端末仕様、ゲートキーパー間通信方式、付加サービス機能などについて議論されている^[6]。付加サービス機能は、H.323 ではオプションであるが、機能仕様は H.450.x シリーズで規定されている。既に、コールトランスファー(H.450.2)、着信転送(H.450.3)については勧告化が終了し、現在は、保留、通信中機器移動、コールウェイティング等の機能について、勧告化作業中である。

今後の標準化動向を注視し、必要な機能については、「VocaLink-Soft/V」への実装を検討していく予定である。

7. まとめ

H.323 標準に準拠したインターネット電話ソフトウェア「VocaLink-Soft/V」を試作し、相互接続性と NEWPRED 方式の効果を確認した。また、ネットワーク帯域およびパソコンのスペックに応じた符号化パラメータの設計を行い、その有効性について検証した。

今後は、従来から筆者らが試作してきた「VocaLink-GW」^[7]についても標準化対応を行い、国際標準方式に準拠した統合的インターネット電話システムを構築していく予定である。また、標準化動向を注視し、必要な機能について実装を図っていく予定である。さらに、品質向上として、音声遅延の低減などについても検討を進める予定である。

参考文献

- [1] 安藤他，“インターネット電話ソフトウェア構成法”，1998 信学総大 SB-11-8, 1998.
- [2] 渡邊他，“Internet 用音声通信アプリケーション用電話帳システムの構成に関する検討”，1998 信学総大 B-8-91, 1997.
- [3] 小澤，“音声符号化技術とその応用”，信学技報，SP95-110(1996-01)，pp9-16, 1996.
- [4] 小桐他，“実網を用いたエラー耐性機能を有

する画像通信方式の性能評価”，1998 信学総大 D-11-75, 1998.

[5] 杉山，“移動通信最速の 32k ビット/秒 99 年に 64k ビット/秒へ高速化”，日経コミュニケーション '98.8.17 号，pp.116-123, 1998.

[6] 安藤他，“ITU-T SG16 の動向(2)”，オーディオビジュアル複合情報処理 22-9，pp.49-53, 1998.

[7] 林他，“インターネット電話システム VocaLink の構成技術”，NTT R&D '97.9，pp.93-102, 1997.