

インターネット電話の品質評価について

高野 正次 KOUNO Shouji

NTT サービスインテグレーション基盤研究所

アブストラクト

インターネットやLANなどのデータ系のネットワークを使って、内線電話のような音声アプリケーションを運用することが広がっている。これには技術的な成熟と優れた経済性にその理由があるが、利用者が実際に耳で聞いたときに違和感を覚えないうように音声を適切に転送できる必要がある。音声の品質を保証するために、電話端末間での遅延や揺らぎをトラヒックモデルによって評価し、ネットワーク設計を行った事例について述べる。

1 はじめに

電話サービスをLANやインターネットといったデータ系のネットワークの上へ統合し再構築することが注目されている [1], [3]。これを実現するための技術は、VoIP (Voice over IP) 技術と総称されるが、この技術を導入する契機としては、音声系のネットワークをデータ系のネットワークの上に構築し直してコストを下げようとする場合と、既存のデータ系ネットワークの新しいサービスとして、安価に電話サービスを提供しようという場合があるが、本稿では、CATV 網 (ケーブルテレビネットワーク) で加入者に対して新しく電話サービスを提供する場合の音声品質について検討する。

以下では、通常の電話を指す場合には、一般電話と呼び、VoIP 技術を利用してデータ系のネットワークを利用して提供される電話サービスを、IP 電話と呼んで区別する。

本検討では、CATV 網を使った IP 電話サービスの End-End 音声品質を、トラヒック設計に基づいて評価・定量化し、一般電話と比較する。特に、本稿では、CATV 加入者が一般電話へ IP 電話をかける場合の音声品質について報告する。

2 VoIP

2.1 導入メリット

VoIP 技術が注目されている理由は、以下のように明確である。

(1) 通話帯域圧縮による優れた経済性

新しい音声符号化方式を利用すれば、通話当りに使用する通信帯域を従来よりも大幅に圧縮することができる。その圧縮の程度は、音声符号化方式に依存するものの、通常と比べて4分の1から8分の1以下にすることができる。また、他のアプリケーションで利用する通信帯域との統計多重効果を期待できるようにもなるので、通信回線全体の有効利用をさらに高めることになり、コスト削減メリットが大きい。

(2) アプリケーションとの親和性

音声通話機能を必要とするシステムでは、VoIP 技術を適用する方がより効率的に高機能のシステムを構築することができる場合がある。具体例を挙げると、通信販売や顧客サポートで使われる顧客対応システムやコールセンターのシステムなどは、最近ではパソコン上の Web ブラウザを表示用のアプリケーションとして作られることが多いから、音声機能の部分は VoIP 技術を利用した方が開発効率がよい。

また、多くの拠点間を結ぶシステムでは、一斉通報機能が重宝されるが、こちらも VoIP 技術を応用すると、IP プロトコルのマルチキャストにより実

ITU-T 勧告 (方式)	符号化速度
G.711(PCM)	48k,56k,64k
G.726(ADPCM)	16k,24k,32k,40k
G.729/G.729A(CS-ACELP)	8k
G.723.1(ACELP/MP-MLQ)	5.3k/6.3k

表 1: 主な音声符号化方式

現できる。このように、新しい通信機能をアプリケーションに追加したい場合、VoIP 技術に乗り換えることで実装が比較的容易になる場合がある。

2.2 課題は音声品質の確保

電話の音声品質に対して、ユーザはとても敏感に感じとっているので、音声品質を現状程度に確保した上で、コスト削減効果を期待している。しかし、これが必ずしも容易ではない。その理由は以下である。

(1) 高速回線と低速回線の混在

LAN が 10Mbps あるいは 100Mbps といった高速回線であるのに対して、WAN への接続・中継回線の多くが 64kbps、128kbps 程度から数 Mbps の低速回線であることが多い。そのため、低速回線の部分がネットワークのボトルネックになりやすい。

(2) データ系の長いパケットの存在

データ系のアプリケーションによって使用されるサイズの大きなパケットは、転送されるときに送出処理時間がかかなり大きくなる。低速の回線で特に顕著となるが、この影響により、同じ通信回線で転送される音声パケットのバッファリング遅延時間とそのゆらぎ（分散・ばらつき）を非常に大きくする。

(3) IP レイヤでの QoS 技術が未確立

音声パケットとデータ系のパケットが同じ通信回線で転送されることにより、データ系トラヒックの特徴であるバースト性に起因して輻輳が発生しやすい。しかし、このバースト性を制御したり、パケットの損失確率や遅延時間を保証するような QoS 技術が IP レイヤで確立されていない。

2.3 遅延とゆらぎの発生

音声パケットが転送される間に、大きな遅延時間やそのゆらぎが以下のように発生する。例として、図 1 に 10 ミリ秒間隔に音声情報を符号化し、順々にひとつの音声パケットとして送出させる場合を考える。このときの音声データの大きさは、音声符号化方式に依存するものの、10～80 バイト程度になり、40 バイトの IP ヘッダを加えた音声パケット全体の大きさは 120 バイト以下となる。このように、10 ミリ秒の固定的間隔でひとつの音声情報としてパケット化して、送信側から送出された音声パケットの列は、受信側の端末でも同じ 10 ミリ秒間隔に

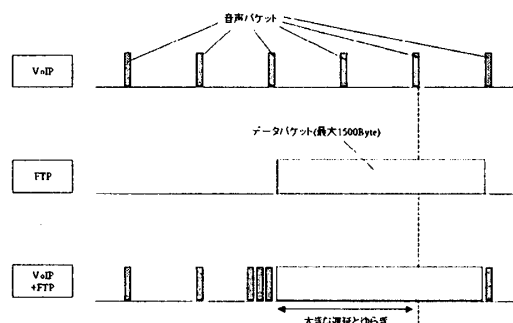


図 1: 大きな遅延とゆらぎの発生

音声パケットを受信して再生する必要がある（図 1 の VoIP）。

FTP などのデータ系のアプリケーションが使用するパケットは、IP パケットの最大サイズとして規定されている 1500 バイトであることが多いが、通信回線の速度が低速のときは送出処理にかかる時間が長くなる（図 1 の FTP）。例えば、通信回線の速度が 64kbps の場合では、1500 バイトの送出処理時間は 187.5 ミリ秒である。

10 ミリ秒の一定間隔で連続的に整列して送出されている音声パケットの列に、FTP による大きなパケットが割込んでしまうと（図 1 の VoIP+FTP）、通信回線が十分に高速でない場合には、元々の音声パケットの列の隙間である 10 ミリ秒のうちに、割り込んだパケットの送出処理が終わらなくて、後続の音声パケットの列を大きく乱してしまう。その結果として、後続の音声パケットのバッファリングによる遅延時間とそのゆらぎが大きくなる。音声品質の観点から言えば、音声パケットが本来再生されるべきタイミングで着信側端末に到着できないので、音声品質が悪くなっている。

その他にも、バースト的なデータ系パケットの連続到着によって、トラヒックの輻輳が引き起こされるときにも、同じように音声パケットの列が乱される状況が起こるので音声品質が劣化してしまう。

2.4 音声品質と会話の関係

音声品質が重要といっても、それは程度の問題といえる。会話の内容次第で、要求される音声の品質はおのずと異なってくるからである。例えば、

音声パケットの遅延時間が大きいときには、どうしても間延びした感じを受けるし、音声パケットの損失が大きいたまには、聞き取りにくいと感じる。したがって、お客様からの苦情対応窓口で音質の悪い電話を使うと、誠意が伝わらずにかえってトラブルが大きくなる恐れもある。逆に、友達同士での長距離の長電話であれば、多少の音質の悪さには目をつぶって、その分安く長く電話を楽しむ方が好ましい場合もある。そういう意味で、会話の内容と直接関連している電話に必要な品質の程度を考慮した上で、VoIP 技術によって提供する音声品質の目標あるいは目安を規定する必要がある。逆にいうと、IP 電話では想定する会話の内容次第では、設計する音質のレベルをあえて下げる代わりにコストを徹底的に削減する選択肢もあるからである。

図 2 は、会話の内容と必要な音声品質の関係を示している。

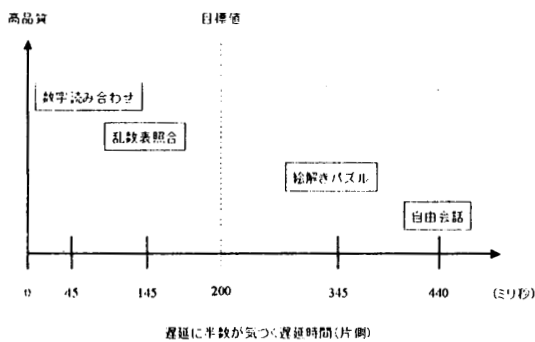


図 2: 会話と音声品質の関係

本検討では、この具体的な目標値として、片側の遅延時間が 200 ミリ秒を規定した。この数値は、一般的に VoIP サービスにおいて片側遅延時間の上限とされている。

3 CATV 網での IP 電話

3.1 ネットワーク構成

CATV 網のネットワーク構成は、図 3 のようになっている。CATV 加入者は、加入者端末として Cable Modem(CM) と呼ばれる装置にパソコンや IP 電話端末を接続して IP 電話を利用する。CM はバス回線に収容されていて、数百人～1 千人程度

の加入者でそのバス回線を共有している。バス毎には、バスルータ (R1) という装置があり、上位のネットワークとの間の通信経路を制御している。各バスは、CATV 会社毎にまとめられて、外部のネットワークとつながっているが、このためのルーチング装置をヘッドエンドルータ (R2) と呼ぶ。CATV 1 社当たりの加入者数を実勢として平均約 1 万人であるので、ヘッドエンドルータはおよそ 10 本程度のバスをまとめている勘定になる。IP 電話による通信を行うためには、ヘッドエンドルータと PSTN 網 (公衆電話網) をつなぐ (図 3 の点線) 必要があり、この間でのプロトコル変換・メディア変換を Media Gateway(MG) と呼ぶ装置が行っている。ヘッドエンドルータと MG 間の回線の使用効率を上げてコストを圧縮する目的で、複数の CATV 会社のヘッドエンドルータから MG へのトラヒックをまとめる装置を中継ルータ (R3) と呼び、この中継ルータと MG を結んで (図 3 の実線) 通話を行う場合もある。

今回の検討では、中継ルータが複数の CATV 会社 (加入者数万人規模) のトラヒックをまとめるネットワークを想定した End-End 音声品質に対して評価を行う。ただし、バスルータやヘッドエンドルータの市販製品には、回線の速度やそのインタフェースの数が決まっているので、それに依存して各ルータ配下の加入者数を調整・制限することは必要になる。

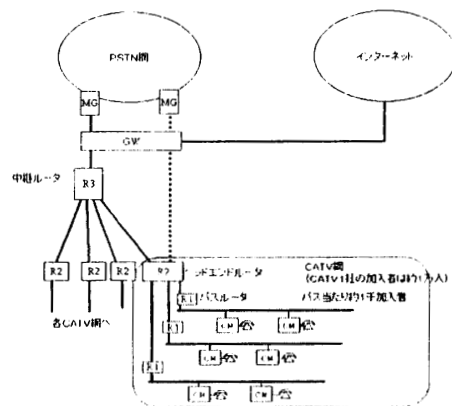


図 3: CATV 網の構成

3.2 End-End 遅延

発信側の CATV 網内端末から着信側の一般電話へ通信を行う場合の、End-End 遅延要因を説明するために図 4 を使う。CATV 網内の IP 電話端末から順番に CM、バスルータ、ヘッドエンドルータ、中継ルータ、GW (後述)、さらに、MG と PSTN 網を通過して、着信先の一般電話へつながる。MG は、IP ネットワークと PSTN 網の間で必要な変換を行っている。つまり、CATV 網から PSTN 網に向かっては、IP プロトコルを終端して、送られた音声データを復号化して、PSTN 網側の一般電話との間に通常の電話回線を接続して通信を行っている。逆向きの音声に対しては、PSTN 網からの音声を符号化・パケット化して、そのパケットを CATV 網側の IP 電話端末に IP プロトコルを使って送出している。

このときの発信側である CATV 網の IP 電話端末と着信側である PSTN 網の一般電話端末との間で生じる遅延、すなわち、End-End 遅延時間の要因は、以下である。

- (a) 符号化・復号化遅延+パケット化遅延
 - (b) 伝送遅延 (距離だけに依存する伝搬遅延)
 - (c) 遅延ゆらぎ吸収バッファにおける遅延
- End-End 遅延時間とは、上記の (a)+(b)+(c) であり、この目標値が 200 ミリ秒となる。

CATV 網の加入者は、HTTP, FTP, TELNET などのインターネットへの接続サービスも利用するが、これらのトラヒックは、GW(Gate Way) で分離されてインターネットへ流れる。この際のインターネットへのトラヒックは音声パケットの遅延時間を大きくする背景負荷として考慮する必要がある。つまり、バスルータを最大約 1 千人、ヘッドエンドルータを平均約 1 万人程度、中継ルータを数万人規模の加入者それぞれ共用することになるが、各ルータに到着した音声パケットは、他の加入者による音声パケットだけでなく、インターネットへ向かうデータパケットとも通信帯域を競合するために、パケットのバッファリングが起こり、バッファ内での滞留時間が生ずる。これも評価する。

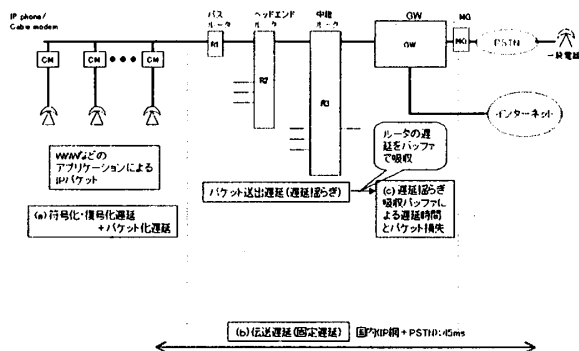


図 4: End-End 遅延要因

4 遅延を抑える仕組み

4.1 データ・フラグメンテーション機能

送信側で長いデータパケットをより短いパケットに分割することを、データ・フラグメンテーションという。パケットを送出するための処理時間が平均的に短くなるので、バッファリングによる遅延時間も短くなる。図 1 のイメージでは、データパケットが音声パケットの列に割り込んでも、データパケットの長さが短くなっているため、音声パケットの列の送出間隔の乱れが相対的に小さくなるので、結果的に、音声品質が向上する効果がある。この機能は、送信側に備えられる機能であるが、実装が比較的簡単であるためかほとんどの市販製品で利用できる。ただし、分割する度に IP ヘッダが必要になるため、小さく分割しすぎると全体として処理負荷が増大してしまうことになるので、通信回線の速度に応じて適切なサイズにフラグメンテーションすることが重要である。

4.2 音声優先制御機能

音声優先制御機能は、音声パケットと音声以外のデータパケットを識別して、音声以外のデータパケットについては、音声パケットの送出の処理が終わるまでバッファに溜めて送出を後回しにするような方法で、音声パケットのスループットを常に一定以上確保する機能である。音声パケットがバッファリングされる時間が短くなるので、IP 電話の音声品質は良くなる。ただし、標準化に時間がかかるからためと考えられるが、マルチベンダー環境で

は音声優先機能を利用することは難しい。例えば、Cable Modem の仕様には、DOCSIS1.1 という音声優先制御機能を含む仕様があるが、これに準拠した製品は 2000 年夏の時点で市場にまだ出回っていないようである。

4.3 遅延ゆらぎ吸収バッファ

音声パケットがエッジルータ、ヘッドエンドルータ、中継ルータを通過する間に、バッファリングによる送出处理待ちの時間が発生する。この待ち時間がパケット毎にちがってくるので、音声パケットの間隔は一定間隔であったものが次第にばらつきが大きくなっている。これを再び元の一定間隔にもどす機能が、遅延ゆらぎ吸収バッファの機能である。

この装置は、すべての音声パケットをいったん遅延ゆらぎ吸収バッファ内に滞留させて、元の一定間隔で再び送しているだけといえる。ただし、あらかじめ規定された遅延以上の大きい遅延時間をもった音声パケットについては破棄を行う。結局、破棄されるほど大きい遅延をもった音声パケットの確率は小さいので、ほとんどの音声パケットは元の一定の送出間隔に戻すことができる（ときどき抜けが生じるが）。これは、End-End 遅延時間とパケット損失率を犠牲にする代償として、パケット毎の遅延ゆらぎを解消する仕組みである。

実用上は、音声パケットの損失により聞き取れない音や音声の途切れが多少あっても、前後の関係から無意識のうちに補って会話を行っているので、会話に支障がおこらないことが普通である。逆に、パケットの損失を減らす目的で遅延ゆらぎ吸収バッファでの滞留時間を長くすると、End-End 遅延時間がさらに大きくなり、間延びがして会話が不自由になる。会話に支障をきたし始める音声の途切れの確率は経験的に 10^{-3} から 10^{-4} 程度とされているので、通信ネットワーク上のパケット損失率と比べると著しく大きい。したがって、遅延ゆらぎ吸収バッファ部でのパケット損失率をかなり大きくしても、End-End の遅延時間をできるだけ小さくする方が、結果として、音声の品質を向上させることができるという理屈である。

本検討では、パケット損失率は 10^{-4} と規定し、そのパケット損失率になるように遅延ゆらぎ吸収バッファでの滞留時間を求めた。

5 End-End 遅延時間の評価

5.1 設計パラメータ

図 4 での IP 電話の End-End の音声品質に影響をあたえる設計パラメータの中で、実際に変動させて評価する主要な設計パラメータについて以下に列挙する。

- 音声符号化方式
G.711(64kbps), G.729(8kbps) など
- 音声データ長
20 ~ 80 バイト程度
- データ・フラグメンテーション機能
150 バイト, 500 バイト, 1000 バイト
- 音声優先制御機能
ON/OFF
- バス当たり加入者数
数百人 ~ 1 千人
- 加入者当たり最繁時呼率 (音声のみ)
0.05 ~ 0.3 程度
- 加入者当たり IP トラフィック (音声除く)
10kbps ~ 100kbps
- 中継回線収容率
0.7 ~ 0.9
- パケット損失率目標値
 10^{-4}
- 片道 End-End 遅延時間目標値
200 ミリ秒

5.2 固定的遅延

End-End 遅延要因の内、その遅延時間分布の分散がごく小さいものは一定であると見なし、固定的遅延と呼ぶことにする。一方、その分散が無視できない遅延要因は、変動的遅延と呼ぶことにして、遅延要因を両者に分類する。そのとき、音声への符号化・復号化遅延とパケット化遅延と伝送遅延を固定的遅延に分類することができ、各ルータでの送出处理待ちのバッファリング遅延時間は、変動的遅延となる。

方式	符号化・復号化遅延	パケット化遅延
G.711	0.75 ミリ秒	5 ミリ秒
G.729	25 ミリ秒	10 ミリ秒

注) 音声データ 40 バイトの場合

表 2: 音声符号化方式とその固定的遅延

表 2 のように、音声符号化方式によって、その符号化・復号化処理の遅延時間がかなり異なる。G.711 は符号化速度が 64kbps であり、G.729 は 8kbps であるが、圧縮効率の優れた符号化方式は、差分符号化などの高度な計算をしているため、その処理時間が大きくなるからである。

また、伝送遅延は、IP 網 + PSTN 網で 45 ミリ秒とした。実際には、End-End の距離に依存して変化するが、この値は国内での片道遅延時間（衛星回線は除く）としてほぼ最大の値である。

5.3 変動的遅延とトラヒックモデル

CATV 網の各ルータでバッファリングされて待つ遅延時間の合計時間を求めるには、各ルータでのバッファリング遅延時間の分布をそれぞれ算出した上で、それらを畳み込む方法をとる。それぞれのルータは、前段の複数のルータからのトラヒックが合流するので、前後のルータ間のトラヒックの相関は無いと判断したことがその理由である。

バスルータ、ヘッドエンドルータ、中継ルータの 3 段それぞれのトラヒックモデルは、 $M/H_2/1/\infty$ とした [2]。この区間では、音声パケット（例えば、60 バイト）とインターネットへのデータパケット（例えば、500 バイト）の両方が混ざっているため、そのパケット長と割合にしたがってサービス過程を H_2 とした。MG のトラヒックモデルは、 $M/M/1/\infty$ とした。MG を通って PSTN 網に行くのは音声パケットだけなので、サービス過程を M とした。ここで、バッファサイズは、当然、有限になるが計算の簡単さのために、バッファサイズは無制限 ∞ とした。実際、パケットの損失は、たいいてい遅延ゆらぎ吸収バッファで発生して、各ルータでのバッファ溢れで損失することはほとんどない。遅延ゆらぎ吸収バッファでのパケット損失率は 10^{-4} 程度であるが、一方で、各ルータでバッファ溢れによりパケット損失となる確率は 10^{-10} 程度になるためである。

したがって、音声パケットの変動的遅延時間分布は以下の直列 4 段ルータの遅延時間分布として算出した。

- $M/H_2/1/\infty$ (バスルータ)
- $M/H_2/1/\infty$ (ヘッドエンドルータ)
- $M/H_2/1/\infty$ (中継ルータ)
- $M/M/1/\infty$ (MG)

5.4 数値例

数値例を紹介する。図 5 のようなパラメータ設定で、前述のような 4 段待ち行列モデルによる End-End 遅延時間分布を求める。その裾の確率が 10^{-4} になる値の 2 倍を、遅延ゆらぎ吸収バッファで滞留させる時間とすると、End-End 遅延時間の内訳は、表 3 のようになる。

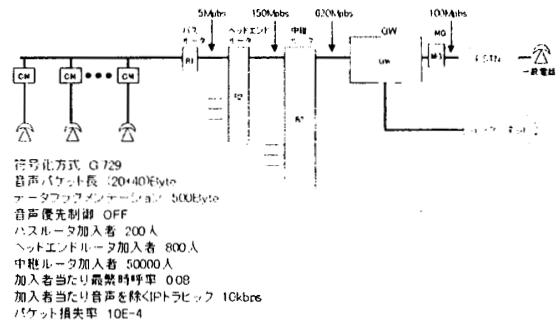


図 5: 数値例のパラメータ設定

この数値例から、現在のバスルータに加入者 200 人という条件の下では、End-End の遅延時間は 100 ミリ秒程度にしかならないので、目標値の End-End 遅延時間 200 ミリ秒と比べた場合には、十分な音声品質が確保されていると判断できる。また、4 段のルータでのバッファリング遅延のほとんど（約 99%）が 5Mbps のバスルータで発生していることに着目すると、バスルータが現在の設計パラメータ条件の下ではボトルネックであることが分かる。これは、ヘッドエンドルータなど後段のルータが十分高速であるのに比べて、バスルータが低速であるためである。しかしながら、5Mbps という回線速度がバスルータの市販品のインタフェース速度であ

パケット損失率	10 ⁻⁴
符号化・復号化 (a)	25.00
パケット化 (a)	10.00
国内 IP 網+PSTN 網 (b)	45.00
バスルータ	12.23
ヘッドエンドルータ	0.18
中継ルータ	1.11
MG	0.15
遅延ゆらぎ (4 段)	12.27
吸収バッファ(c)	24.54
片道遅延 (a)+(b)+(c)	104.54 (ミリ秒)

表 3: End-End 遅延時間の内訳

り、さらに高速にすることはできないのに対して、ヘッドエンドルータと中継ルータの回線速度は、これ以上高速にすることもできることなどを考慮すると、結局、バスルータがボトルネックとなることは不可避である。したがって、IP 電話の音声品質を決定するのは、バスルータの回線速度とそれを共用する加入者の人数などが支配的な要因であることが分かる。

6 まとめ

CATV 網で IP 電話を提供する場合の、CATV 網側から発信して PSTN 網側へ着信する通話について、音声品質を決定する End-End 遅延時間に対する評価を行った。これにより、音声品質について、一般電話と IP 電話を定量的に比較することができた。ネットワーク設計の観点からは、音声品質を左右する End-End 遅延時間を大きくする要因としては、バスルータでのバッファリング遅延時間が最大のボトルネックになることが分かった。ただし、IP 電話の品質を向上させることを目的とするならば、圧縮効率の高い音声符号化方式である G.729 などでは、ルータ全体でのバッファリング遅延時間に対して、符号化・復号化処理のための遅延時間が無視できないので、符号化処理時間を短縮する方法も有効であると考えられる。

一般電話は国内での片道遅延時間はほとんど 30 ミリ秒以内とされていることなどから、結論的には、IP 電話は一般電話よりも音声品質の面ではかなり悪いと言わざるを得ないが、衛星回線を使用す

る場合よりも良いので、コストと勘案することでサービスとしての競争力は無視できないと考える。特に、音声優先制御機能を備えた CM などの装置の価格が安くなったときには、End-End 遅延時間が短縮されて音質もかなり良くなるので、この段階で、一般電話サービスに対して IP 電話サービスが脅威となることもあり得る。

今後の課題は以下である。本稿では、発信側が CATV 網内で、着信側が PSTN 網内の場合の遅延時間を評価したが、ヘッドエンドルータでの折返し通話や、インターネットを利用して中継した発着信共に CATV 網加入者の場合の通話についても、End-End 遅延時間の評価による音声品質の定量化をする必要があると考えている。

また、バスルータ当たりの加入者数を変動させることなどで、音声品質と提供するためのコストの関係もかなり具体的に評価できた。今後は、バスのマルチアクセスによる衝突などの現象を適切に反映するように、CSMA-CD 方式などに基づくトラフィックモデルで再評価することを検討したい。

文献

- [1] 天野他, "VoFR, VoIP で音声網を再構築", 日経コミュニケーション, 1999.6.7.-1999.7.19
- [2] 藤木他, "通信トラフィック理論", 丸善, 1980
- [3] 日本フレームリレーフォーラム編, "フレームリレーネットワーク構築ガイド", 電気通信協会, 1998