

TCP/IP上でのマルチメディア通信とその性能

坂本 泰久 桑名 栄二
NTTソフトウェア研究所
{sakamoto,kuwana}@mickey.ntt.jp

概要

端末間のパーソナルなコミュニケーションを支援することを目的としたオーディオ/ビデオ情報のリアルタイム通信環境の実現方式について報告する。通信プロトコルとして既存のコンピュータネットワーク上で最も普及しているTCP/IPを用いる点に特徴がある。静止画圧縮機能付きビデオボードとオーディオ符号化デバイスを内蔵したワークステーションを用いてプロトタイプを構築し、その試用を通じて得られたいくつかの問題点のうちビデオ品質についての性能評価実験を行った。その結果、今回対象とした二端末間のビデオ通信では、ネットワークの伝送能力ではなくハードウェアによるイメージ圧縮処理が主なボトルネックになっていることを明らかにした。

Performance Evaluation of Multimedia Communication on TCP/IP

Yasuhisa Sakamoto and Eiji Kuwana
NTT Software Laboratories
1-9-1 Kohnan Minato-ku Tokyo, 108 Japan
{sakamoto,kuwana}@mickey.ntt.jp

Abstract

In this paper, we describe an architecture of audio/video real-time communication environment to support inter-personal communication. This architecture is characterized by using TCP/IP protocol which is used most widely on existing computer networks. We made a prototype by use of workstations with a built-in videoboard which has image compression function and a built-in audio device and made a performance evaluation experiment about the video quality in some problems which were found through its experimental use. It was clarified that the bottleneck of video communication between two terminals were not network transmission capacity but image compression process in videoboard.

1. はじめに

従来のデータ通信システムは主として形式情報を正確に高速に効率的に伝送することに主眼が置かれてきた。形式情報とは例えばテキストの16進表現やイメージのドット情報などのような0,1のビット列情報である。一方、人間と人間のコミュニケーションでは他人に何かを伝える時、送り手は伝えたい意図を何らかの形式情報に変換して伝達する。受け手は受けとった形式情報を自分の知識に基づき解釈してその意味を理解する。今後の計算機通信システムは、従来の形式通信のみならずこのような意味レベルのコミュニケーションを支援する方向に向かう必要がある[1]。

遠隔地にいる人間の意味レベルのコミュニケーションを支援する方法の1つとして、マルチメディア通信がある。遠隔コミュニケーションにおいては表情、ゼスチャー、話し方などのいわゆる非言語情報をいかにして伝達するかが課題の1つである。これらの非言語情報は人間のふるまいを映像情報、声を音声情報として通信することによりある程度の伝達が可能である。従来の計算機ネットワークのコミュニケーション支援範囲はテキスト通信による言語情報の伝達に限られていたが、CPUパワーの増大とネットワーク通信技術、データ圧縮技術の発達によりオーディオ/ビデオ通信を含めたマルチメディア通信をパーソナルな計算機間で行うことが可能になってきている。重要なのは、各個人の端末にマルチメディア通信機能が備わることである。このパーソナルなコミュニケーション環境は、個人のアクティビティとの個人間のコミュニケーションの連続性が高まるという点で、TV会議室などにはないメリットがあると考えられる。

本報告ではパーソナルマルチメディアコミュニケーションの基盤技術である計算機ネットワーク上のマルチメディア通信の一実現方式について報告する。2章で映像の符号圧縮機能と音声の符号化機能を備えたワークステーションによるマルチメディアLANの構想を述べ、3章で映像通信における性能評価実験の結果について報告する。

2. LAN上のマルチメディア通信環境

2.1 背景

CS-PCM衛星放送、次世代CATV網をはじめとしてオーディオ/ビデオ放送分野のデジタル化が進んでいる。デジタル化のメリットとしては伝達における品質の保証の他に、情報操作の自由度が増すことが挙げられる。例えば、画面サイズの変更や飛ばし見など再生時の空間的/時間的な操作をよりフレキシブルに行ったり、データ中に曲のタ

イトルなどの情報を付加して音声と同時に表示したりすることが容易に実現できる。情報源(ソース)がデジタルになってくると、従来からデジタル通信を行ってきたコンピュータネットワークをその通信インフラストラクチャとして利用するのは自然な発想である。以下、本稿ではオーディオ情報、ビデオ情報を合わせてマルチメディア情報と呼ぶ。

複数のワークステーションが接続されているLANをベースとしてマルチメディア情報のリアルタイム通信システムを構築する場合、最も現実的な方法はこれらをLANとは別の媒体(例えば同軸ケーブル)上にアナログ情報として流し、ワークステーションはその制御を担当するというやり方である。これは、(1)マルチメディア情報のデータ量が非常に大きくEthernetを媒体としたLANには負荷が高すぎる、(2)連続したアナログ信号、特にビデオ信号をリアルタイムでデジタル化/圧縮化する能力が計算機に備わっていないこと、などの理由による。

しかし(1)に対してはFDDI/CDDIやATMを用いたLAN技術の開発により計算機ネットワークの伝送能力はますます大きくなっている。また(2)に対してはJPEG,MPEGなどの画像の符号圧縮に関する国際標準規格の確立を受けて、専用のLSIを計算機に内蔵する方向で解決されてきている。これらの要素技術の進展をみるとマルチメディア情報をLAN上でデジタル通信できる状況が着々と整いつつあるといえる。

またマルチメディア情報をリアルタイム通信するための基盤技術としてリアルタイムカーネル、リアルタイム同期プロトコル、メディアサーバなどの研究が盛んになってきている[2][3]。

2.2 プロトタイプの構成

図1,2に示す例は、特別なCODECなどの装置を利用することなく、既存のコンピュータネットワーク上の通信プロトコル(TCP/IP)を用いて、ビデオ通信、オーディオ通信、テキスト通信を実現したプロトタイプである。図1にシステム全体の構成を示す。図2は本システムをネットワークアプリケーション(共有エディタ)と組み合わせて使用しているイメージである。

ビデオ通信は、まずカメラからのアナログビデオ信号をワークステーション上のビデオボードに取り込み自分のウインドウに表示する。同時に、ビデオボードに付属しているJPEG圧縮チップによりその画像を圧縮したのち、コンピュータネットワークプロトコル(TCP/IP,Xウインドウ)を用いて相手のワークステーション上に転送し表示する。この静止画通信を連続的に実行することによりデジタル動画像通信を実現している。また相手のワ

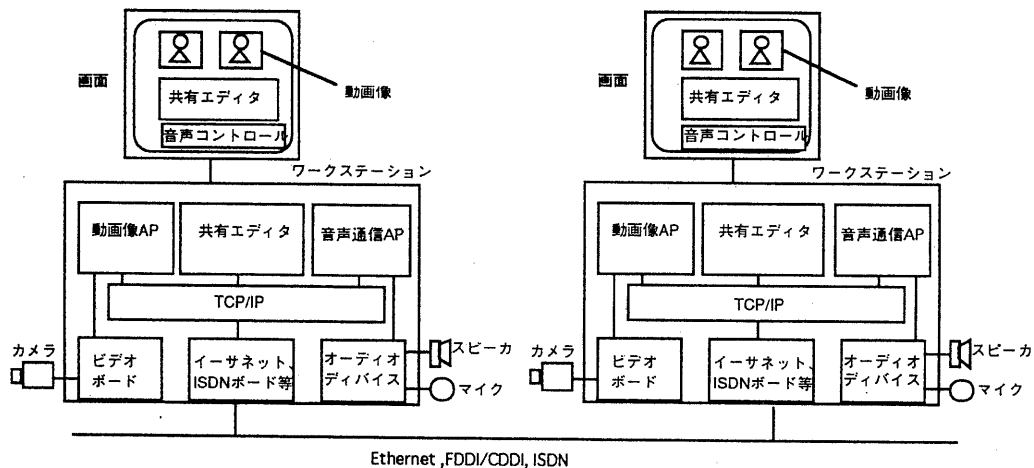


図1：システム構成

ークステーションでも同じ処理を行い、双方のワークステーション上に双方のカメラの動画を表示することができる。画像データの圧縮率はパラメータの変更により調整可能である。

オーディオ通信については、ワークステーションにビルトインされたオーディオ用デバイスを使いマイクからの入力をデジタル情報化し、それをTCP/IPのUDPプロトコルを用いて通信する。サンプリング周波数8KHz、 μ -LAW、8bit量子化方式でデジタル化したデータを非圧縮で送受信している。このビルトインオーディオデバイスを使ったアプリケーションがすでにいくつか開発されている。例えばマサチューセッツ大学のNevot[4]というアプリケーションでは多者間の同時音声通信を実現し

ており、コンピュータネットワークを利用して音声会議を行うこともできる。

テキスト情報やグラフィックス情報はネットワーク上のアプリケーション共有ツールによりリアルタイムに共同編集することができる。

2.3 共通プロトコルの意義

Internetに代表されるコンピュータネットワークではファイル転送、リモートファイルアクセス、電子メールなどさまざまなネットワークサービスがTCP/IPプロトコルを使って世界規模で実現され普及している。これに加えて映像/音声などのマルチメディア情報を同一プロトコルで通信する意義として次のようなことが挙げられる。ただし、これらはプロトコルを共通にした場合の一般的な意義でありTCP/IPに固有なものではない。

- ・コンピュータネットワークさえあれば、ビデオ通信、オーディオ通信もその上で可能になる。したがってシステム拡張の際に新たにマルチメディア通信の線を増設する必要がない。

- ・ネットワーク媒体を選ばない。現在のLANはEthernet、FDDIなどさまざまな媒体で実現されているが、どのような媒体であっても上位プロトコルとしてTCP/IPが確立していれば通信が可能になる。ATM-LANなどの将来的なネットワークにも適用できる。

- ・複数のメディアを制御するアプリケーションの生産性を高めるために、統一的インタフェースでマルチメディアを扱える言語の開発が進められている。マルチメディア情報の入出力の共通のインタフェース（本システムの場合はTCP/IPのソケット

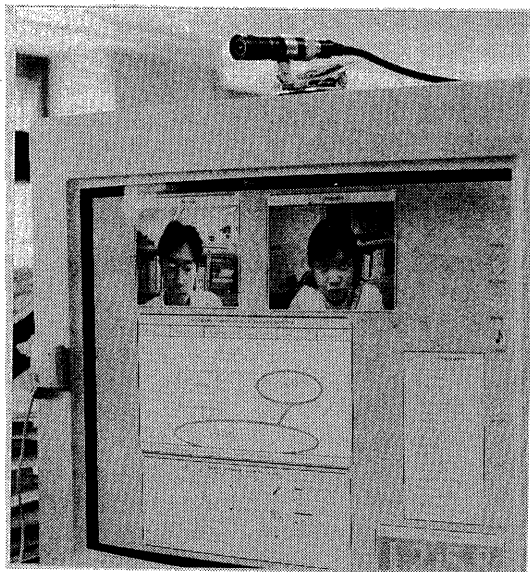


図2：使用画面の例

トインタフェース)はこのような言語を実装するための必要条件となる。

2.4 試用と考察

2.2のシステムを二者間の分散会議に試験的に利用し以下のような項目について問題点を考察した。

(1) 映像と音声の同期

現在は映像と音声は別々のアプリケーションにより独立したデータとして伝送されており、映像一音声間の同期をとるための工夫はなされていない。したがって両者の遅延時間の差によるずれが常に発生する。そのため話している人の唇の動きと音声がずれる現象が起こる。MHEGなどで議論されている映像と音声の同期制御に関する標準方式を早く利用できることが望まれる。

一方、対面コミュニケーションでは表情/態度などの非言語情報の一部を伝えることが大きな目的の1つであり会話内容はオーディオ情報だけから理解することができるので、唇の動きを正確に読み取ることまでは通常要求されない。したがって唇の動きと音声の不一致が多少あっても音声通信および映像通信はそれぞれそれなりの役割を果たしているという見方もできる。

(2) 音声の遅延

TCP/IP通信のオーバーヘッドによりどうしてもある程度の遅延は避けられない。正確な測定は行っていないが、音声の場合少なくとも0.5秒以上の遅れがあると思われる。したがって衛星中継のような感覚になり、コミュニケーションを正常に行うためには相手の発言が終わるまで待つなど、ある程度の慣れが必要となる。対話のようにリアルタイム性が要求される通信において、コンピュータネットワークを通した音声通信が浸透していくかどうかは遅延の低減にかかっていると思われる。

(3) ネットワークに対する負荷

既存のネットワークに大量のマルチメディア情報が流れることにより、他の通信サービス(例えばファイル転送、リモートファイルアクセス、ウィンドウシステム)に障害を与えるという可能性がある。2台のワークステーション間で通信を行っている状況では、めだつた障害は発生していないが、すでにネットワークの利用率はEthernetで平均25%にも達している。将来的にビデオ情報のブロードキャスト通信が複数実行される状況になると、ネットワークの伝送能力が低い場合には、他の通信セッションの中断やネットワークの輻輳などの

深刻な障害が発生すると推測される。

(4) 映像の品質

映像通信の品質は、空間的な品質と時間的な品質の2種類の側面で表現できる[2]。対面会議における顔映像情報は高画質性(空間的な品質)はあまり要求されないがフレーム表示速度(時間的な品質)については表情の変化やゼスチャが伝わる程度の品質が必要となる。一方、資料の提示のような静的な情報の場合には空間的な品質が優先される。映画のように映像そのものが伝達対象である場合には両者ともにできるだけ高い品質が望まれる。対面会議における顔映像については、ほぼ実用に耐える品質が得られたが、机上の資料などのデスクトップ映像やTV番組などの放送映像では満足な品質は得られなかった。

このようにアプリケーションの特性により品質に対する要求条件は変わってくるので、システムには空間的な品質と時間的な品質を使用形態に応じて調整できる機能が求められる。特に多数のノードで多数のプロセスが動作しているコンピュータネットワーク上の通信ではネットワークトラフィックやCPU負荷など資源の変化への対応も重要である。

音声については通常の会話には支障のない音質が得られるが、音楽の再生用としては不十分である。高音質化、ステレオ化が望まれる[5]。

3. 映像通信の性能評価実験

2.4の考察のうち(4)映像の品質の問題に着目し、映像情報のトータルな性能向上と品質調整機構の実現に向けた性能評価実験を行った。そこでは品質のボトルネックとなる要因および空間的な品質と時間的な品質の関係を明らかにするための基礎データ測定に主眼をおいた。

3.1 評価システム構成

評価実験システムの構成を図3に、構成要素の一覧を表1に示す。

JPEG圧縮機能付き動画表示用フレームバッファ(XVideo[6])を備えた2台のUNIXワークステーション間でLAN上のデジタルビデオ通信を行なう。画像の圧縮/伸長はフレームバッファ内のハードウェアで行われる。通信はサーバークライアント方式、通信プロトコルはビデオ用に拡張されたXプロトコルである(TCP/IPの上位レイヤとして実装されている)。クライアントは入力されているアナログビデオ信号を静止画フレームの列として符号化し、次に圧縮用ハードウェアに対してJPEG圧縮の要求を出す。そして圧縮されたフレームデー

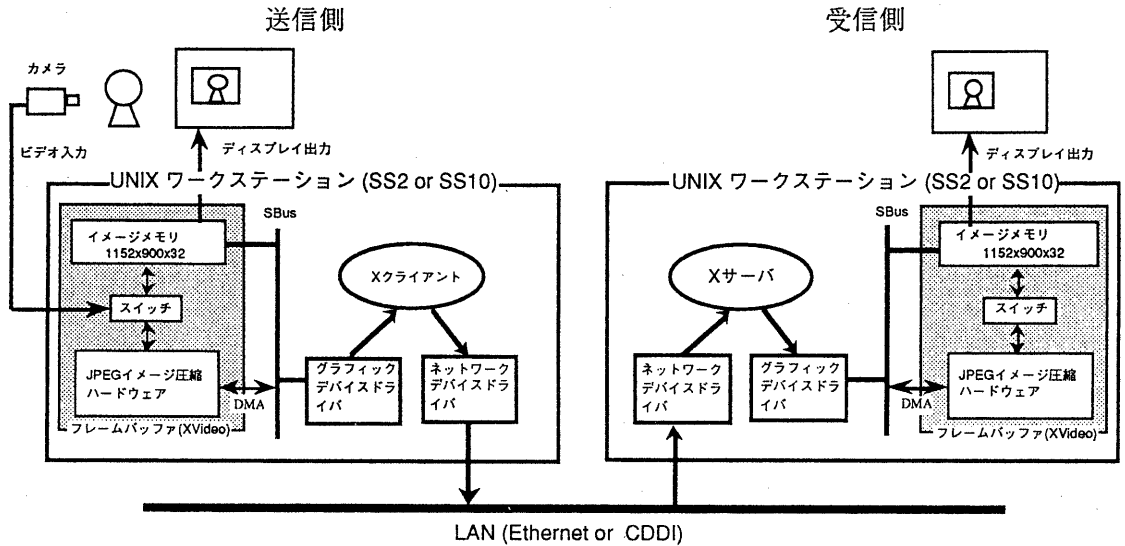


図3：評価実験のシステム構成

タをXサーバに対して送信する。サーバは表示要求を受けるとハードウェアによる伸長およびフレームバッファへの表示を実行する。

ワークステーションの性能比は、
 SS2:SS10=1:2
 ネットワークの伝送速度比は
 Ethernet:CDDI=1:10
 である。

表1：評価実験のシステム構成要素

ワークステーション	SPARC station2(以下SS2)
	SPARC station10(以下SS10) (Sun Microsystems社)
通信用OS	SunOS4.1.3
表示用ソフト	OpenWindow3.0 (XVideo対応)
フレームバッファ 画像圧縮／伸長	XVideo-24SVC-VIO(Parallax社)
ネットワーク	Ethernet
	CDDI

3.2 測定方法および測定値

測定はクライアント(送信側)とネットワークの両方で行い、相互につき合わせをしてデータを検証した。クライアントで送信フレーム数、送信データサイズおよび送信時間を記録した。この測定値

からフレームレート(1秒あたりのフレーム更新数)、フレーム処理時間、およびネットワークのスループットが算出される。そしてネットワーク上に接続したプロトコルアナライザを用いてデータサイズ、スループットなどをモニタリングし上記の値を検証した。送出された全パケットは時刻情報も含めて保存した。さらに双方のマシンで送受信したパケット数をカウントし確認を行った。サーバ(受信側)での測定は行わなかったが、予備実験により展開および表示に要する時間は圧縮に要する時間よりも充分小さいことが明らかになったので、送信側の算出値をそのまま全体の値とした。

通信形態は今回は1対1のホスト通信のみを対象とした。また同一LAN上の通信を対象とし、LANを越えた通信(WAN通信)についてはゲートウェイや接続回線などの外的要因が加わってくるので除外した。

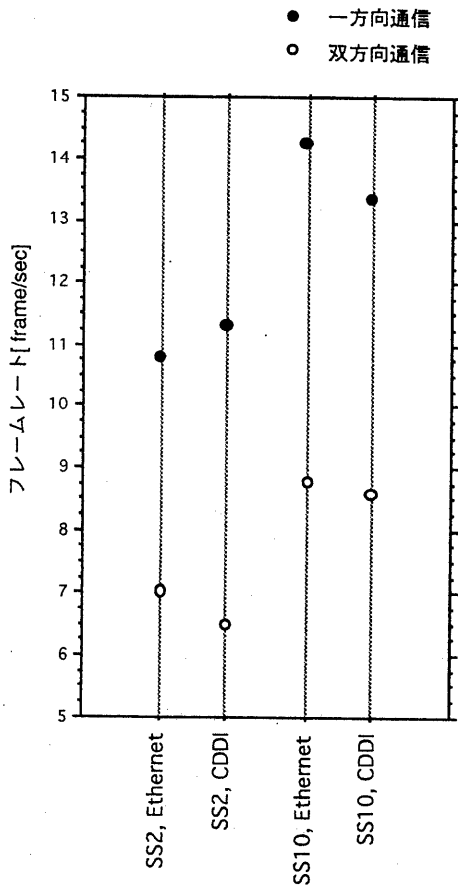
入力ソースは通信対面会議を想定して会話中の人間の顔映像とした。入力信号はNTSCアナログビデオ信号で、各実験で差がでないように8mmビデオテープに記録したものを用いた。

3.3 測定結果

(測定1) ネットワークおよびマシンによる違い

画面サイズを固定し、ネットワークインタフェースとマシンの種類を変えてフレームレート(1秒あたりのフレーム更新数)を測定した結果を図4に示す。通信形態は双方向と一方向の2種類である。双方向通信は画面上の2つのウィンドウに自分と相手の顔映像が表示される状態、一方向通信は1つのウィ

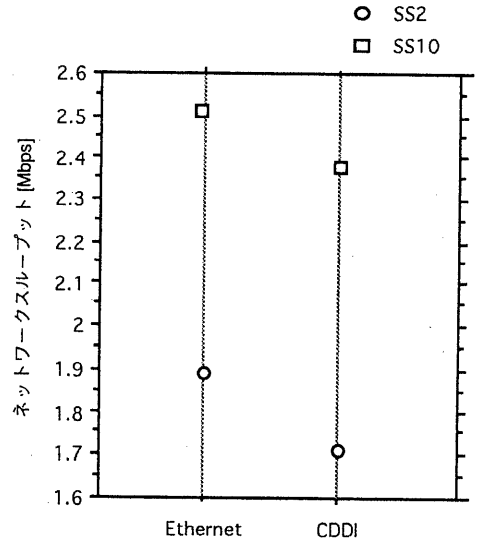
ドウにどちらか一人の顔画像だけが見える状態である。測定時間は1回あたり10秒間、測定は同一条件で5回ずつ行い平均を値とした。



画面サイズ 横320x縦240 (クォーターサイズ)
 JPEG圧縮パラメータ Q=50
 圧縮率5.5%
 (平均フレームサイズ:圧縮前307KB,圧縮後16.8KB)

図4: 平均フレームレート

アプリケーションによる測定と同時にプロトコルアナライザにより測定したネットワークスループットの値を図5に示す(双方向通信の場合)。



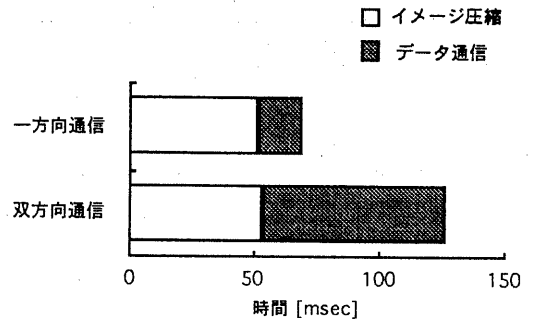
双方向通信
 画面サイズ 横320x縦240 (クォーターサイズ)
 JPEG圧縮パラメータ Q=50
 圧縮率5.5%
 (平均フレームサイズ:圧縮前307KB,圧縮後16.8KB)

図5: ネットワークスループット

フレーム送信処理は次の4つのサブルーチンの繰り返しである。

- (1) 画面の静止
- (2) 画面イメージの圧縮
- (3) サーバとのデータ通信
- (4) データ消去

マシンがSS10, ネットワークがCDDIの場合について、フレーム処理時間(1フレームを送信するのにかかる時間)とそのサブルーチンの内訳を図6に示す。



マシン SS10
 ネットワーク CDDI
 画面サイズ 横320x縦240 [pixel]
 JPEG圧縮パラメータ Q=50
 圧縮率5.5%
 (平均フレームサイズ:圧縮前307KB,圧縮後16.8KB)

図6: フレーム処理時間の内訳

(測定2) 圧縮率による違い

画質とフレームレートの関係を見るために、圧縮率を変化させてフレームレートの測定を行なった。圧縮率は圧縮前後のフレームデータサイズの比を表している。圧縮率が高くなるほど圧縮後のデータサイズは小さくなるが再生画質は粗くなる。圧縮率とフレームレートの関係を図7に示す。

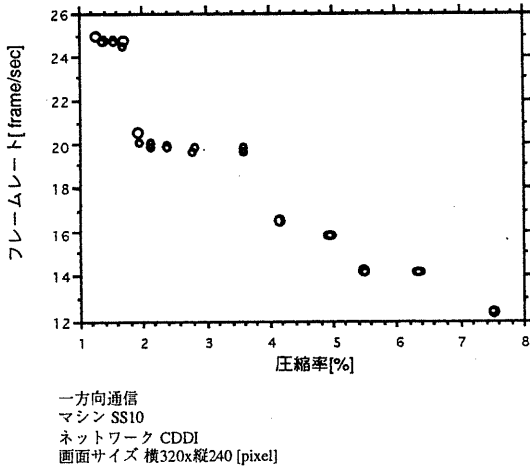


図7：圧縮率とフレームレートの関係

3.5 評価

(1) デジタル映像通信のボトルネック

ネットワークがEthernetの場合よりもCDDIの方が性能が上がることを予想していたが、図4により両者を比較すると予想に反して大きな性能の違いが見られなかった。逆にCDDIのほうが性能が落ちるケースが多かった。図5よりネットワークのスループットは最大で2.5 Mbps程度であり、ネットワークの最大伝送速度 (Ethernetで10Mbps, CDDIで100Mbps) には達していない。またOSのTCP/IP通信部のスループット (ftpによるファイル転送で少なくとも8Mbps) にも達していない。したがってネットワークの伝送能力はボトルネックにはなっていない。CDDIの方が性能が劣化する要因として、ネットワークインタフェースドライバのつくりの問題などが考えられる。

図6よりフレーム処理時間のほとんどはイメージの圧縮およびデータの通信に費やされていることがわかる。通信形態が異なってもイメージ圧縮に要する時間はほぼ同じなのに対し、データ通信に要する時間は大きな差がある。イメージ圧縮はハードウェアで行っているため、データ通信時間が仮に0になったとしてもフレーム処理時間は50

msec以下にはならない (すなわちフレームレートは20 frame/sec以上にはならない)。一方向通信の場合はイメージ圧縮とデータ通信の比率が約3 : 1であり、ハードウェアによるイメージ圧縮がもっとも大きなボトルネックになっている。双方向通信の場合は、データ通信処理もかなりの時間を要している。ただし双方向通信の場合のデータ通信時間中にはサーバのデータ受信および伸長/表示に要する時間 (予備実験により約20msec) も含まれている。データ通信時間に関してはCPU性能の向上によりアプリケーションおよびOSの処理速度が上がれば短縮される可能性がある。

現在さまざまな動画の圧縮方式が開発されている。MPEG[7]など多くの方式ではフレーム間の圧縮も行うので、静止画連続圧縮にくらべて大幅にデータの圧縮率が高くなるはずである。データ圧縮率が高くなればデータ通信時間は短縮される。しかし上記の考察から明らかのように、リアルタイム通信でトータルな性能を向上するためにはデータ圧縮率だけでなく、圧縮に要する時間も考慮に入れなければならない。圧縮率を高めるための工夫は圧縮時間をも大きくする可能性がある。すなわち圧縮率と圧縮時間は相反する関係にある。今後は動画圧縮対応のモジュールが開発されデジタル動画通信システムの核として普及していくと思われるが、常に最良の映像品質を得るためには、圧縮時間と圧縮率を可変にしコンピュータネットワークの性能に応じて最適な組み合わせを選択する機能が必要になる。これにより、例えば伝送能力の高いネットワークの場合に通常は圧縮率よりも圧縮時間を優先した組み合わせを採用して品質を上げ、ネットワークの混み具合にしたがって圧縮率優先に変えていくといった調整が可能になる。

(2) 時間的品質と空間的な品質の関係

映像品質には空間的な映像品質と時間的な映像品質の2種類がある。今回の評価における測定値と品質の関係は表2のようにになる。ただし測定値と品質の関係は線形ではない。評価実験では空間的な品質(圧縮率)を与えて時間的な映像品質(フレームレート)を変化させた。

表2 映像品質と測定値の関係

品質	測定値
時間的品質	フレームレート
空間的品質	フレーム圧縮率
	表示サイズ

この観点から図7のグラフを見ると、縦軸が時間的な品質を横軸が空間的な品質を表わしている。空間的な品質を上げるにしたがって時間的な品質は劣下していくが、その減衰の仕方は階段状になっていることがわかる。その原因として圧縮時間の非線形的な変化などが考えられるがはっきりとはわかっていない。

またJPEG(MPEG)圧縮を用いた映像通信の場合、サービス品質は映像ソースによって大きく変化することを考慮しておく必要がある。JPEG圧縮アルゴリズムの性質上、映像の性質により圧縮率および圧縮時間が変化するためである。

3.6. 今後の展開

(1)連続静止画圧縮と動画圧縮の性能比較

3.5で述べたように動画圧縮方式(例えばMPEG)のリアルタイム通信の場合、今回の連続静止画圧縮方式(例えばJPEG)以上に圧縮時間がボトルネックになる可能性がある。伝送能力の異なるネットワークで両者の性能を比較し、圧縮方式とコンピュータネットワークの関係について調査する。

(2)マルチポイント通信への拡張

通信形態は品質に大きな影響を及ぼす要因である。例えば、三者以上の多者間通信やブロードキャスト通信による放送的な通信になるとネットワークのトラフィックが増しデータ通信のオーバーヘッドはさらに増大すると考えられる。今回はPoint-to-Point通信のみを対象としたため通信形態と品質との関係を確認することができなかった。今後、実現方式を含めてマルチポイント通信について検討していく予定である。

4. まとめ

オーディオ/ビデオ情報のリアルタイム通信を、既存のコンピュータネットワーク上の通信プロトコル(TCP/IP)を用いて実現する方式を提案した。特別なCODECなどの外部装置を利用することなく、JPEG静止画圧縮用ハードウェア付きフレームバッファとオーディオ符号化デバイスを内蔵したワークステーションによりプロトタイプを構築した。その試用経験から、映像と音声の同期、音声の遅延、ネットワークに対する負荷、映像の品質に関する問題点を考察した。

上記問題点のうち映像品質について、ボトルネックとなる要因および空間的な品質と時間的な品質の関係性を明らかにするための性能評価実験を行った。その結果、今回対象とした1対1のホスト間通信ではネットワークの伝送能力は性能のボトルネッ

クにならないことがわかった。さらにフレーム送信処理に要する時間を分析することにより、データ通信処理と同等あるいはそれ以上の割合でハードウェアによるイメージ圧縮処理が大きなボトルネックになっていることを明らかにした。そして、コンピュータネットワーク上のリアルタイムビデオ通信において常に最良の映像品質を得るためには、動画圧縮用モジュールに圧縮率と圧縮時間を変えられる機能を持たせ、コンピュータネットワークの性能に応じて最適な組み合わせを選択する必要があることを示した。

参考文献

- [1] 桑名 栄二, 増尾 剛, "マルチメディアコミュニケーションが意味レベルで人をつなぐ", ビジネスコミュニケーション, Vol.30, No.3/4 (1993)
- [2] 徳田 英幸, 斎藤 信男, "マルチメディア統合環境プロジェクトにおけるリアルタイム処理技術", 信学技報, コンピュータシステム, CPSY92-75 (1993)
- [3] 慶大環境情報研究所, "開放型基盤ソフトウェア研究開発評価事業に係わる「マルチメディア統合環境基盤ソフトウェア研究開発」報告書" (1993)
- [4] Schulzrinne, H., "Voice Communication Across the Internet: A Network Voice Terminal", July (1992)
- [5] 篠田 晃, 石井 一彦, 桑名 栄二, "クロスプラットフォームなコンピュータネットワーク上での音声通信の実現, 情報処理学会マルチメディアと分散処理研究会 (1993)
- [6] "XVideo Software Developer's Guide", Parallax Graphics, Inc. (1991)
- [7] 渡辺 裕, "MPEG", テレビジョン学会誌技術解説 Vol.46, No.9, pp1149-1152 (1992)
- [8] 西田 竹志, "TCP/IP ~ネットワークプロトコルとインプリメント~, ソフト・リサーチ・センター (1989)
- [9] Rowe, L. and Smith, B., "A Continuous Media Player", Proc. 3rd Int. Workshop on Network and OS Support for Digital Audio and Video, San Diego CA (1992)