

P2P/ブロードバンド時代の 新・TCP/IP 入門

村上 健一郎 法政大学ビジネススクール イノベーション・マネジメント研究科 教授

第4回 リアルタイム通信での音声や映像の送り方は？

連載3回目は、インターネットの仕組みはリアルタイム通信に設計されていないということに触れました。しかし、私たちの生活の中では現実インターネットを使った音声や映像のリアルタイム通信が行われています。今回は、それらはいったいどのような仕組みでやりとりされているのかについて、見てみます。

[Q1]

リアルタイム通信で音声や映像情報をTCP/IPでどのように送るのでしょうか？

[A1]

音声も映像も圧縮して転送するのは同じ

音声にしる映像にしる、インターネットを使用して通信を行う場合には、できるだけ使用する帯域を小さくするために情報を圧縮します。帯域とは必要な回線速度のことで、ビットレートといいます。

この圧縮には、

- (1)人間の視覚や聴覚の特性を利用して気づきにくい部分を省略する方法
- (2)毎回全部の信号を送るのではなく、前に送った信号との差分だけを伝えることによってデータ量を抑える方法などが使われます。ここで重要なのは、アプリケーションを区別して考えること

です。

リアルタイム型の通信では、図1に示すように、インターネット電話のような会話型通信と、一方的に相手に音楽や映像を送る非会話型通信に分類できます。

前者のインターネット電話では、伝えるのが音声です。この場合、人間の音声特有の性質を利用して音声データを大幅に圧縮し、転送するビットレートを下げています。一方、後者の場合には、それが人間の音声だけではないことに注意を払わなければなりません。音楽では、楽器などのように人間の音声とは性質の異なるものがあり、音声と同じ方式で圧縮したのでは著しく品質が低下してしまいます。ですから、音楽では音声とは異なる方式で圧縮を行うのです。

インターネット特有の「遅延」

会話型と非会話型では、送信したデータが相手に届くまでの時間、つまり遅延の影響についての違いも考えなければなりません。

インターネット電話のような会話型の通

信では、遅延の問題が大きなポイントとなります。遅延が大きいと、相手がまだ話していないと思って話しはじめたところ、相手からの音声到着して両者の音声がかぶってしまい、相手が何を言っているのかわからなくなります。

一方、映像配信のような非会話型の通信では、遅延はそれほど重要な要素にはなりません。時報を送る場合などには、それがずれてしまうという問題はありませんが、一方向に映像を送る映画やニュースでは、少々の遅延によって品質が劣化することにはなりません。

なお、ここでは、RTP(Real-time Transport Protocol)を使用して連続的にリアルタイムでデータを送るストリーミング方式に限定して話をしています。TCPを使用してファイルをダウンロードした後でローカルに音楽やビデオを再生する方法については、リアルタイム通信の範疇に入れていません。

さて、図2は音声や音楽、映像などがどのように転送処理をされているのかについて示しています。

図2に示すように音楽や映像のデータ

には、まずデジタル変換と圧縮が行われます。映像やその音の場合には、さまざまな圧縮方式があります。

たとえば、映像の場合、画像圧縮の国際規格である MPEG(Moving Picture Experts Group)方式による圧縮が利用されます。MPEG には、それぞれ圧縮率(ビットレート)の違う MPEG-1(ビデオCDなどで利用)や MPEG-2(デジタル放送などで利用) MPEG-4(携帯電話のテレビ電話などで利用)などの方式があります。また、国際規格ではない、マイクロソフト社の Windows Media Video 特有の方式やアップルコンピュータ社の QuickTime 方式が使用されることもあります。

弱も判別しているのですが、人間の耳に聞こえる最も小さい音と最も大きな音の差は1兆倍とも言われています。これはダイナミックレンジと呼ばれています。音楽CDの場合には、このダイナミックレンジをほぼカバーしていますが、電話の場合にはその半分程度のダイナミックレンジになります。

一方、映像信号は、たとえばアナログ

放送のテレビ映像の場合、必要な周波数帯域幅は4MHzから6MHz程度になります。これは、音声と比較すると約1000倍という非常に大きな値です。また、画面の明暗を表現する輝度のダイナミックレンジ(画像の黒から白までの濃度が変化する幅)は、音のダイナミックレンジの半分以下になっています。

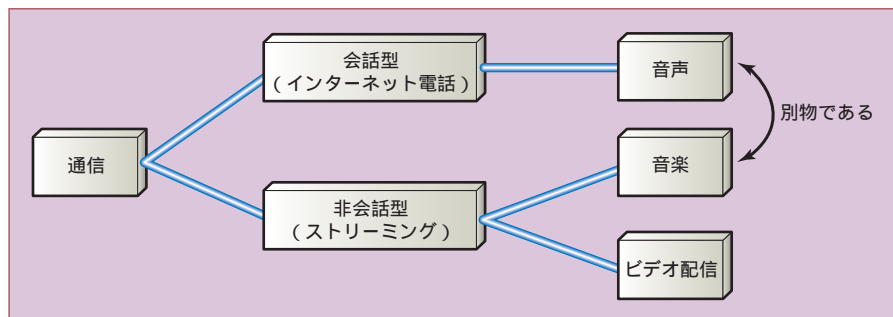


図1 リアルタイム通信の分類

[Q2]

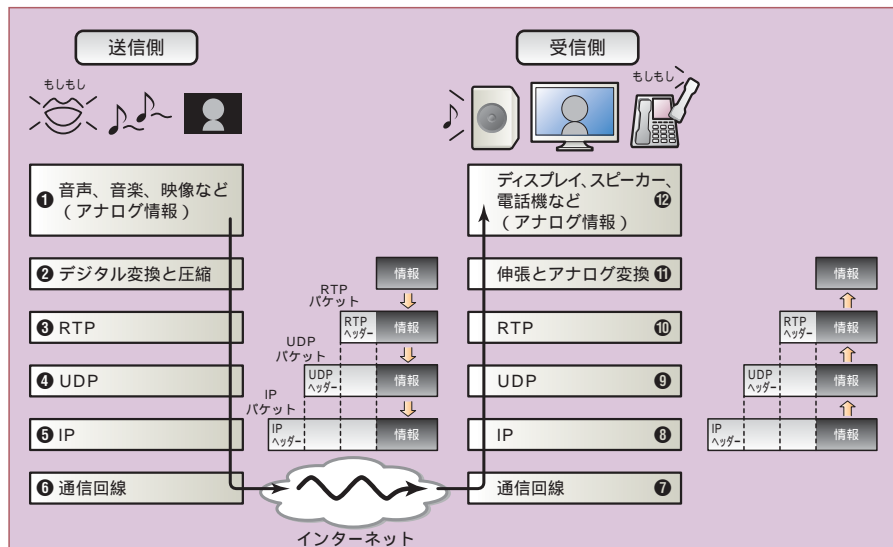
音声と映像の性質はどのように違うのでしょうか？ また、それらの送り方はどのように違うのですか？

[A2]

カバーする周波数帯の違い

一般に音声信号と映像信号の性質には次のような違いがあります。まず、音声ですが、人間の聞くことのできる周波数は20Hzから20kHzという範囲です(図3)。たとえば音楽CDの場合、このほとんどの周波数帯をカバーしています。しかし、電話の場合には、このうち300kHzから3.4kHz程度の範囲の音だけを送っています。このような狭い範囲の音だけでも、会話には支障がありません。つまり、送信する音の周波数範囲を限定することによって、送るデータ量を圧縮しているのです。

人間は、音の周波数だけでなく音の強



- ① 音声や音楽、映像のアナログ情報を取り込む。
- ② ①のアナログ情報をデジタル情報に換え、圧縮(符号化)処理を行う。
- ③ ②のデジタル情報をRTPのペケットに詰める。この時、タイムスタンプ(時刻印)が押される。これは、受信側がペケットの逆転現象を検出してペケットを正しい順番に並べ替えたり、指定された間隔でディスプレイやスピーカーからの再生を行うために使用される。
- ④ RTPペケットはUDPペケットの情報部分に入れられる。UDPでは、チェックサム(検査合計)を計算してペケットに入れておき、受信側で転送中にペケットが損傷を受けていないかどうかを判別できるようにしておく。
- ⑤ UDPペケットはさらにIPペケット内に入れられる。ここでは、転送先のパソコンのIPアドレスが指定されている。
- ⑥ IPペケットは、インターネットなどの通信回線からインターネットへ送り出される。インターネット上ではルーターがIPペケット内にあるIPアドレスを参照して、それを指定されたパソコンまで転送する。
- ⑦~⑩ 受信側では送信側の処理と逆の順序で処理される。IPペケットからUDPペケットを抜き出し、さらにRTPペケットを抜き出す。
- ⑪ ⑩の圧縮されている情報を伸張処理などを行った後、アナログ化する。
- ⑫ アナログ化された情報は、スピーカーを鳴らしたり、ディスプレイに表示されたりする。

図2 音声や映像情報の転送処理

音声は どのように送られている？

普段使っている固定電話でのアナログの音声は、

- (1) 受話器からのアナログ信号を一定時間ごとに振幅をサンプリング(標本化)
 - (2) (1)を一定のビット数(例:電話では8ビット)のデータに置き換える(量子化)
 - (3) アナログ信号をデジタル信号に変換する(符号化)
- の3段階でデジタル化して交換機を通して送っています。

インターネット電話では、デジタル化されるまでは同じですが、デジタル化後は、音声データをRTPパケットに入れて転送する仕組みとなっています。

また、インターネット電話や携帯電話では、固定電話のような帯域が確保できないために、さらに音声を圧縮しています。たとえば、日本の携帯電話の方式PDC(Personal Digital Cellular)では、約6.7kbpsの低いビットレートに圧縮して音声を送っています。

ここで、基本となる固定電話で使われているG.711標準の、音声信号の送り方について見ていきましょう。

前述したように、電話音声の場合は、300Hzから3.4kHzまでの音を送れば通話に支障がありません。この約4kHzまでの音を受信側で再生するためには、受話器で拾ったアナログ音声を8kHzで振幅をサンプリングしてデジタル化します。8kHzでサンプリングするのは、元のアナ

ログ波形に含まれる最大周波数の2倍の周期でデジタル化すれば情報の損失が起こらない(元の音に復元できる)という「シャノンの標本化定理」によっています。

このように、ビットレートは、8ビット×8kHz = 64kbpsとなります。

音声と音楽は別物

電話は音声に特化した圧縮をしているため、楽器の音などを送ると不自然な音になります。また、音楽では、電話より速いビットレートが必要となります。たとえばCDは、44.1kHzの周期でサンプリングを行っており、音の振幅はより解像度の高い16ビットで表現されます。このため、CDの音をそのまま送った場合、44.1kHz × 16ビット × 2 = 1.4Mbpsのビットレートとなります。

最近利用されているMP3は、この音楽の情報量を1/12から1/13程度に圧縮しています。MP3とはISOで規定されている音声圧縮規格のMPEG-1 Audio Layerのことで、MPEG-3のことはありません(誤解のないように)。

音楽では、通常は128kbps程度のビットレートを必要とし、電話のように6.7kビットなどの低ビットレートでは音質が非常に悪くなってしまいます。

MP3では、音声を圧縮する場合、音声の場合と同じようにいくつかの人間の聴覚特性を利用しています。まず、人間の可聴範囲(20Hzから20kHz)が均一ではなく、3～4kHzをピークにして低い音と

高い音の感度が急激に低下することを利用して、聞こえない範囲の音は削除します。また、たとえばBGMなどでほかの耳障りな雑音を聞こえなくする「マスキング効果」も考慮して圧縮しています。

さらに、高域や低域では音源の位置がわかりにくいことを利用して、音の時間差だけを残してモノラル化します。そして、出現頻度の高い量子化された値は「短いビット数」で、出現頻度の低い値は「長いビット数」で符号化することによってビットレートを下げて圧縮します。これはハフマン符号化と言われています。

映像は どのように送られている？

ここでは、MPEGを例として映像の圧縮や送り方について簡単に見てみます。

映像通信では、1秒間に30回程度、つまり30コマの静止画像を切り替えることによって、人間が動画として感じられるようにします。これは、少しずつ違う漫画を紙に書き、パラパラとめくるとまるで動いているように感じるのと同じ原理です。ただし、映像の転送では、完全な画像を毎回送るのではなく、完全な元の画像1枚を送った後は、その画像との差分だけを数回送ることで、転送すべき情報量を圧縮します。具体的には、図4のようになります。

「遅延」は会話型の音声で問題

ここでは、深刻な問題が発生しがちなインターネット電話の遅延について考えてみます。

現在の固定電話では、8ビットのデータが125マイクロ秒(1/8000秒)という定期的な周期で送られます。つまり、通話相手との間に設定された回線では、1秒間に8000回の転送が正確に行われ、これが遅れたり早まったりすることはありません。しかし、インターネットでは、パ

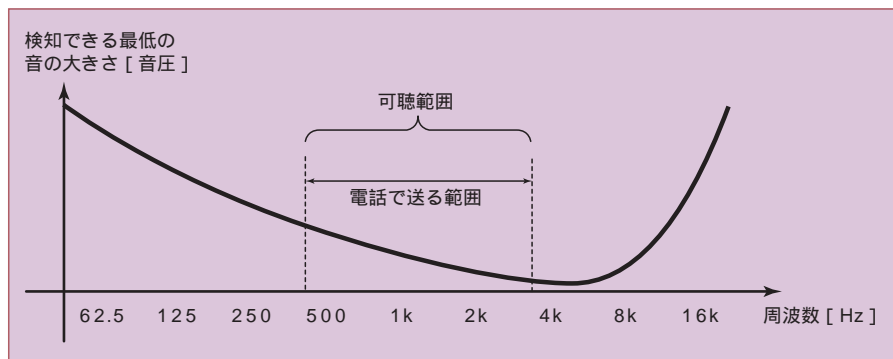


図3 可聴周波数と音圧の関係(ロビンソン・タドソンの等ラウドネス曲線)

ケットが相手まで到着する時間は保証されていません。また、複雑な圧縮をすればするほど、その処理にかかる時間が増え、遅延が増加してしまいます。

ITU-T(国際電気通信連合の電気通信標準化部門)の勧告 G.114 では、送信側から受信側までの全体の遅延を 150 ミリ秒以下に規定しています。これ以上の遅延では会話に支障をきたし、400 ミリ秒以上では利用不可能としています。

実際の遅延はどのくらい？

送信側では、アナログ信号をサンプリングしてデジタル化した後に符号化器による圧縮を行います。これは、定期的に RTP のパケットに入れられて受信側まで転送されます。また、受信側では、送信側と逆の処理が行われ、最終的に受話器から音を出します(図2参照)。

ここで、インターネット電話でよく使用される規格の1つ「G.729」を例として遅延を計算してみましょう。この G.729 のビットレートは 8kbps です。サンプリング周波数は 8kHz で 125 マイクロ秒ごと、つまり 1/8kHz ごとにサンプリングを行いますが、10 ミリ秒分のサンプリングデータをブロック(固まり)として処理を行います。

この処理結果は 10 バイト(80kbps × 10 ミリ秒 = 80 ビット)のデータ量になります。ただし、処理には次のブロックも参照する必要があるため、10 ミリ秒では処理は完了せず、圧縮が完了するまでには 15 ミリ秒の遅延が発生します。注意しなければならないのは、圧縮されたデータごとにパケット化するとオーバーヘッド(負荷)が大きくなることです。なぜならば、IP ヘッダーが 20 バイト、UDP ヘッダーが 20 バイト、RTP ヘッダーが 12 バイトの計 52 バイトのヘッダーが、わずか 10 バイトの音声データを転送するために必要になってしまうからです。

そこで、複数の処理結果をまとめて入れる方法がとられますが、1ブロック増え

るごとにその処理完了時間である 10 ミリ秒の遅延が増えるため、最大 3 ブロック分のデータを入れるのが限度とされています。その場合、パケット化までに 30 ミリ秒の遅延が発生します。

一方、受信側ではデータを伸張して元の音声に復号する必要があります。この時間は、圧縮時間のおよそ 10 % 程度です。しかし、パケットに複数の圧縮されたブロックが入っている場合、伸張時間はその数に比例します。3 ブロックの場合、3 ミリ秒程度です。また、パケット到着の揺らぎ(ジッター)によって、次に再生する音のデータが間に合わなくなると音声途切れることとなります。これを避けるために、すぐに再生するのではなく、しばらくバッファに溜めておいて、意図的に遅延をもたせなければなりません。このバッファ(蓄積メモリ)をデジッターバッファ、その遅延をデジ

ター遅延と呼びます。

たとえば、このデジッター遅延を 40 ミリ秒とすると、全体での固定的な遅延は表 1 のようになります。すでにわかっている値だけでも 88 ミリ秒となっており、先に述べた、G.114 勧告に規定されている全体の許容遅延である 150 ミリ秒の半分を超えています。数十ミリ秒から数百ミリ秒の間、あるいはそれ以上に変動するインターネットの転送遅延を考慮すると、G.114 勧告の目標は達成が危ういものであることがわかります。

内訳	遅延時間(ミリ秒)
エンコード遅延	15
パケット化遅延	30
デコード遅延	3
デジッター遅延	40
転送遅延	不定(n)
合計	88 + n

表 1 実際の遅延時間を見てもみる

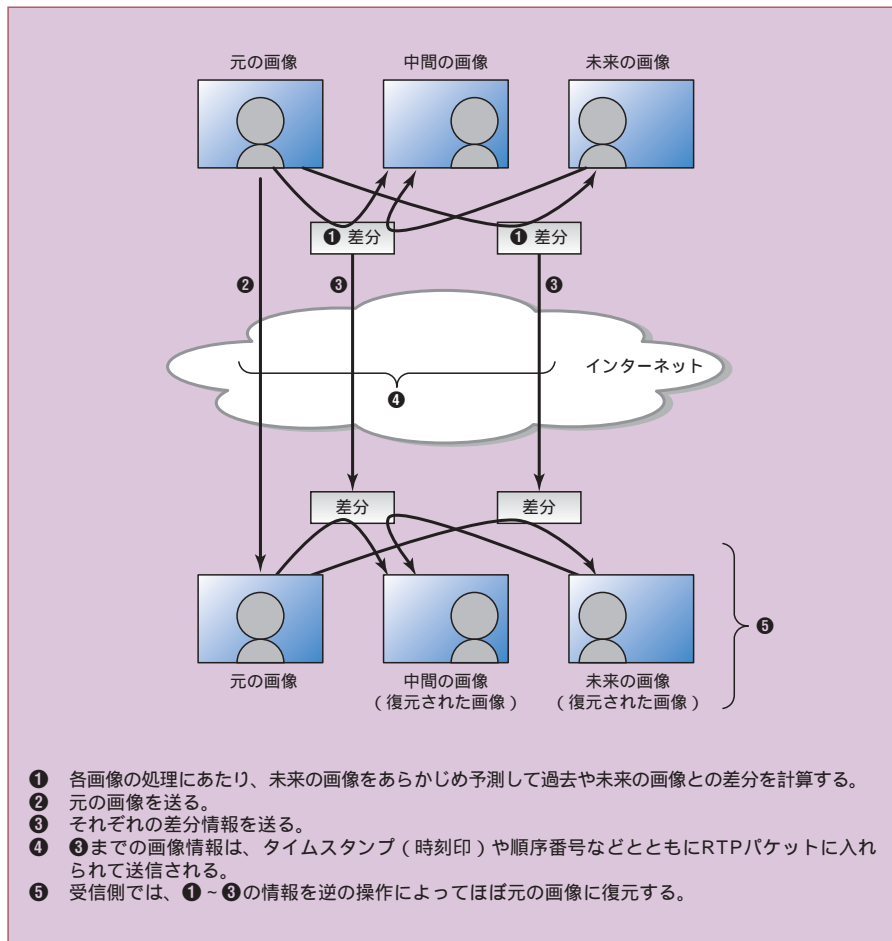


図 4 MPEGにおける画像情報の圧縮と転送



[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp