

インターネット上での音声会話ツール

田中裕之* 平原正樹** 荒木啓二郎*

* 奈良先端科学技術大学院大学情報科学研究科

**九州大学工学部情報工学科

近年の計算機の進歩や計算機ネットワークの回線容量の増大をうけて、インターネットを通して音声や画像を用いて会話をするツールが作られている。本稿ではこれらのツールの中から音声会話ツールについてとりあげ、既存の音声会話ツールが抱える問題について述べる。既存の音声会話ツールが抱える最も重要な問題は、回線容量の変動が原因となって生じる音声の途切れである。本稿では、回線の状態に応じて適度な遅延を音声の処理に与えることによって既存のツールよりも円滑な会話ができる音声会話ツールを作成することが可能であることを示した。

Voice Conference Tool across the Internet

Hiroyuki Tanaka* Masaki Hirabaru** Keijiro Araki*

* Advanced Institute of Science and Technology, Nara Graduate School of Information Science
8916-5, Takayama, Ikoma 630-01, JAPAN

** Department of Computer Science and Communication Engineering, Kyushu University
Hakozaki 6-10-1, Higashi-ku, Fukuoka 812, JAPAN

Increased computational resources and bandwidth have made interactive voice and video conference tools available between workstations across the Internet. This paper describes some existing voice conference tools and their problems. The most serious problem of these tools is fluctuating load of networks causes interruption of conferences. In order to avoid this interruption, we delay processing voice packets according to load of networks appropriately. This paper describes this delay makes more smoother conference across the Internet.

1 背景

計算機技術の発達に伴い計算機の利用は多様化しており、現在では理工学分野以外の分野にも広く計算機が利用されるようになった。それに伴って計算機ネットワークを介したコミュニケーションの利用が広がる傾向にあり、その重要性がしだいに認識されるようになった[1]。現在学術研究機関で広く普及している計算機ネットワークであるインターネットでは、talk や phone のようなリアルタイム性の高い会話ツールとともに電子郵便や電子掲示板のようなメッセージ蓄積型のリアルタイム性の低いメッセージ伝達ツールが提供されている。

これらのツールは現在文字を媒体として使用しているが、近年文字だけではなく音声や映像を媒体に使用したツールが試作されるようになった。例えば、NeXT station の Voice Mail は音声を用いた電子郵便の実装例で、MIME(Multipurpose Internet Mail Extensions)[2] は音声や映像を用いた電子郵便の実装例である。また vat(Visual Audio Tool)[3] や Nevot(Network Voice Terminal)[3] は音声を用いた会話ツールの実装例である。本稿では、上述のような音声や画像を用いたコミュニケーションツールの中から、vat や Nevot のような音声会話ツールに的を絞って考察する。2 節では、現在のインターネットでの代表的な音声会話ツールである vat を例にとって既存の音声会話ツールの問題点について示し、3 節ではこれらの問題点の解決策について考察する。4 節では 3 節で提示した解決策の実装及び評価について述べる。

2 現状と問題点

2.1 vat の概要

vat は Sun Microsystems の SPARC Station 上で利用できる X-Window ベースの音声会話ツールである。vat は SPARC Station に標準で搭載されている音声入出力機能を使用して音声の入出力を実現して

いるので、特別な機器を特に必要とすることもなく手軽に利用、実験を行なうことができる。vat のおもな特長は以下の通りである。

- 多人数での会話が可能
IP-Multicast[4] を利用することにより多人数での会話を可能としている。
- 音声データを圧縮して伝送可能
vat では 64Kb/s 8bit μ -law 対数圧縮 8KHz PCM で入力される音声データを圧縮して伝送することにより通信回線にかかる負担を軽減している。また数種類の圧縮率の違う圧縮方式を用意しており、利用者が適宜切替えることによってその時点の回線状態に応じた圧縮方式を使用することができる。
- データ転送時に無音部を省略可能
通常の会話は有音部 (talkspurt) と無音部 (silence) から構成されている (図 1 参照)。通常の会話では、無音部は全体の約 60% を占めているので有音部のみを伝送することによって通信回線にかかる負担を軽減することができる [6]。

2.2 音声会話実験

広域ネットワーク上で行なわれた vat を利用した音声会話の実験の 1 つに 1992 年 11 月にワシントン D.C. で行なわれた IETF の会議内容を vat で放送する実験をあげることができる [3]。この時の実験では同時に国内参加組織間での会話実験も行なわれた [5]。この実験で我々は、現在のインターネットで vat を使用する際には以下の 2 つの点が問題になると認識した。

1. 会話相手との接続手段の不備
2. 音声の途切れの発生

vat には会話したい相手を呼び出したり、他の人の会話に自分が途中参加の意志表示をするための機能が実装されていない。このため、会話をするために

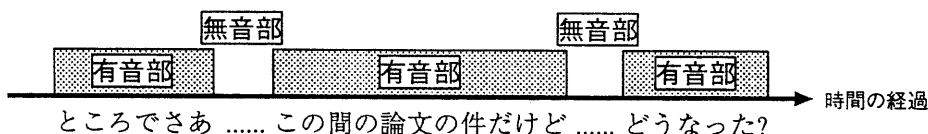


図 1: 有音部 (talkspurt) と無音部 (silence)

は phone や電子郵便などのコミュニケーションツールを併用して相手を呼び出したり参加意志を伝えなくてはならない。音声会話ツールを使用する際に他のコミュニケーションツールを併用した場合、音声会話ツールだけを使用する時と比べ操作が複雑になり利用者に負担がかかる。従って、音声会話ツールの実用化には相手を呼び出したり会話への参加意志を伝える機構を組み込む必要がある。

このような機構を音声会話ツールに組み込む際には、talk や phone にある会話相手を呼び出す機能を実現する技術を音声会話ツールに応用することが可能である。またこの問題は、インターネットを通じた音声会話そのものの問題ではなく人間とツールとのインターフェイスの問題なので本稿では特に議論しない。

もう1つの問題点である「音声の途切れ」とは、話者の言葉が発言の途中で途切れて会話内容を聞きとることが困難になる現象を指す。この現象は、Nevot などの他の音声会話ツールでも vat とほぼ同じ音声通信、処理方式を採用しているため同様に発生する。会話内容を聞きとることが困難になると、円滑に会話を進めることができなくなるのでコミュニケーションツールとして使用することはできない。従って既存の音声会話ツールを実用化するには、音声の途切れが発生しないようにする改良が必要となる。そこで次章では、既存ツールの音声通信、処理の方式を紹介し、音声の途切れが発生する原因とその解決手段について議論する。

3 音声の途切れ

3.1 発生の原因

既存ツールでの音声通信及び音声処理の仕組みを図2に示す。図2で、横軸は時間の経過、縦軸は処理の進行を表している。また、有音部のデータは斜線で、無音部のデータは空白で表している。Ts は音声 packets を分割する時の単位時間である。

発言者が発生した音声は、まずデジタルデータに変換されて計算機に取り込まれる。計算機はそれを Ts 時間おきに packets にして相手に伝送する。受けとられた packets は元の連続したデータの形に再構成する作業が終了した後、直ちにアナログデータに変換して再生される。以上が発言者が発生した音声が開き手の計算機で再生されるまでの処理の流れである。

この一連の処理を経て話者の会話が聞き手に伝わるまでの時間は一定ではなく変動している。この変動は、インターネット上を音声の packets が伝達する時間が一定ではないことが原因となって発生する。音声 packets が伝送されるインターネット上では、音声会話ツール以外のアプリケーションが使用している packets も同時に伝送されている。FTP のような短時間で大量の packets を発生させるアプリケーションを音声会話ツールと同時に使用してネットワークが混雑した場合、音声データの伝送にかかる時間は平常時よりも長くなり音声 packets に遅延が生じる。図2は、このような音声 packets の遅延によって再生される音声に途切れが発生することを示している。

また、ネットワークの混雑による音声 packets 転

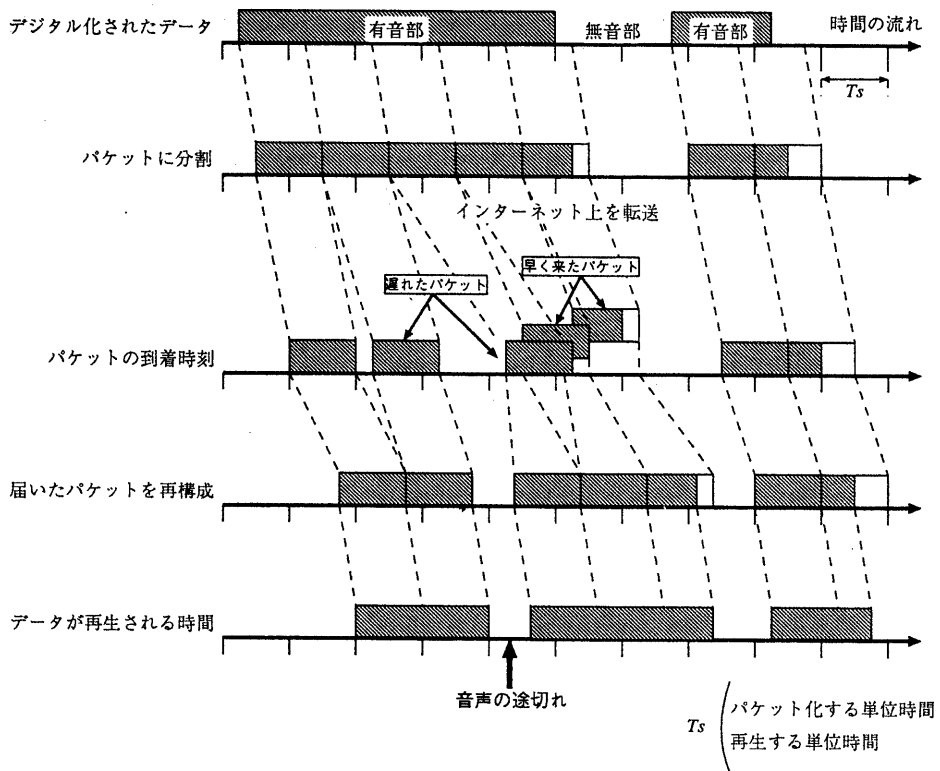


図 2: インターネットを通じた音声伝達

送の遅延によって、送信側の送信バッファには、送信待ちのデータが蓄積される。送信待ちのパケットの量がバッファの限界を越えた場合には、一部のパケットが損失する。この損失も再生時に音声に途切れが発生する原因となる。

以上のことから、音声再生時に途切れなく再生するためには、パケットの遅延が発生しないようにするか、パケットの遅延が発生しても音声再生時に途切れなくなるような工夫をすることが必要であることがわかる。以降では、再生時の音声の途切れを回避するための方策についてこの2つの側面から考える。

3.2 回線容量の確保

音声パケットがインターネット上に伝送される時に生じる遅延は音声パケットの伝送のために必要な回線容量が常に確保されないために発生する。そこで、インターネット上でアプリケーションが必要とする回線容量を確保するための機構を実現して、音声会話ツールに利用することが考えられている[7]。例えばvatやNevotではST-II(Experimental Internet Stream Protocol, Version 2)[8]を用いて必要な回線容量を確保して会話を行うことが可能である。

このような解決方法をとることによって音声の途切れは完全に解消されるが、そのためには回線容量を確保するための新技術の導入が必要となる。しかしながら我々は、実用的な音声会話ツールの実現の際に音声会話ツール以外の新しい技術を導入することはコミュニケーション手段としての音声会話ツールの利用に制限を加えることになると考えている。したがって、本稿では回線容量を確保する技術の導入による音声会話ツールの実現についてはこれ以上触れず、既存のインターネットでの通信技術を元にした音声会話ツールの実現について考える。

3.3 回線の状態を反映した音声伝達

必要な回線容量が十分に確保できるとは限らない現在のインターネット上で音声の伝達を行なう場合、回線の状態を反映しない既存のツールの音声伝達方式では音声の途切れが発生して、会話を円滑に進めることができない。そこで我々は、音声の途切れの発生を防ぐために、以下の機構を既存のツールに追加して現在の回線の状態を音声の伝達に適切に反映することを提案する。

- パケットの遅延状況に応じて音声の圧縮方式を切替える
- パケットの遅延状態に応じて有音部の再生に遅延を与える
- パケットの遅延状況をパケットの送信に反映する

圧縮率が高い音声の圧縮方式では復元される音声の音質を落すことによって圧縮率を高めている。従って、高い圧縮率の圧縮方式を使った場合低い圧縮率の圧縮方式を使った時と比べて回線にかかる負担は小さくなるが、再生される音声の音質は低下する。既存のツールでは、利用者が現在の回線状態と会話に必要な音声の音質を判断して使用する圧縮方式を決定する。このため、常に現在の回線状態を正しく反映した圧縮方式を使用した音声伝達をしているとは限らない。そこで、回線が混雑していない時は圧縮

率の低い圧縮方式を使って音質の良い音声会話サービスを提供するが、回線が混雑してパケットが遅延してきた場合には圧縮率の高い圧縮方式に切替えることによって回線の混雑を緩和するような機構を既存のツールに追加する。この機構の追加によってパケットの遅延が緩和されて音声の途切れの発生を抑えることができる。ただし、圧縮方式を切替える場合には利用者の定める最低限の音質を確保するように配慮する必要があるため、回線の混雑の緩和には限界がある。

有音部の再生に遅延を与える機構とは、音声パケットの遅延に応じて有音部の再生開始時刻に適切な遅延を与えることによって音声パケットが遅延しても再生時刻に間に合うようにする機構である。この機構を追加することによって音声の再生に遅延が発生するが、音声の途切れは発生しなくなる。(図3参照)。パケットの遅延が著しいために再生に大幅な遅延が必要な時は、1つの有音部の音声パケットがすべて再構成されるまで再生せずに待つことによって音声の途切れない再生を行なうことができる。パケットの遅延が著しい場合は、受信側で前述のような処置をすると同時に、パケットの送信間隔を広げてパケットの損失を防ぐことによって、音声の途切れが発生しないようにすることも必要である。

3.4 遅延の問題

音声の途切れないようにするために音声の再生開始時刻に遅延を与えると会話の進行に遅延が生じることになる。音声の伝達の際に話者と聞き手との間に生じる遅延はコミュニケーションの障害になるので、音声会話ツールの製作時にはできるだけ遅延を生じさせてはならない。しかしながら、コミュニケーションで最も重要なことは相手への意志伝達である。従って遅延を作ることによって会話内容が正確に相手に伝わるのなら遅延が生じるのもやむを得ない処置であると我々は考える。

また、遅延の発生を許容することによって既存のツールでは音声の途切れて会話が困難になるような

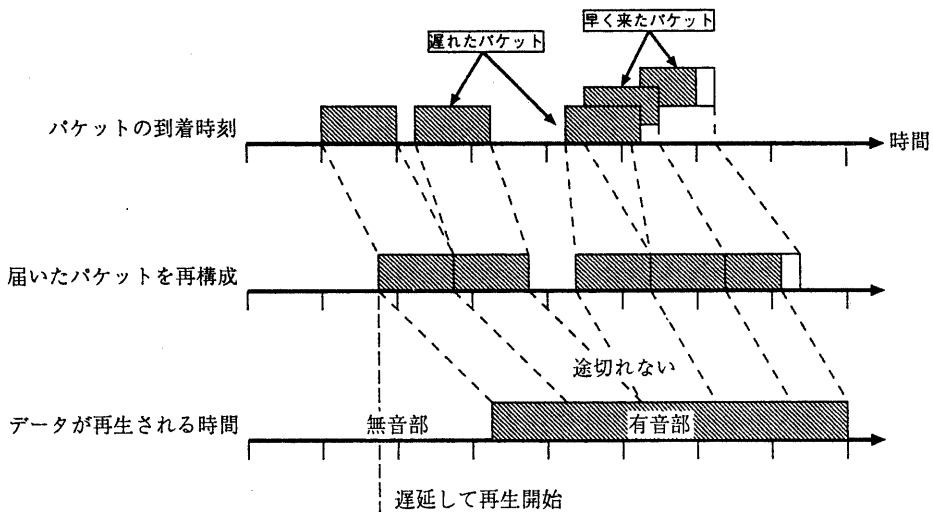


図 3: 有音部の再生に遅延を与える

回線容量の小さい回線を使用した場合でも遅延が発生するという制限つきながら会話が可能となる。これは、既存のツールと比較するとコミュニケーションが可能となる範囲が広がるという点で大きな利点になっていると我々は考えている。

4 実装及び評価

回線の状態を反映させる機構を実装した音声会話ツールを作成するための前段階として有音部単位での再生を行なう音声会話ツールを試作した。この試作ツールを用いて、遅延が会話へ及ぼす影響と既存ツールでの音声の途切れによる会話への影響とを比較した。試作ツールは、ソースリストが公開されている Nevot のプログラムの一部を書き換えて製作した。また、試作ツールは Sun Microsystems の SPARC Station 上で実行し評価を行なった。

評価は、ローカルエリアネットワーク (10Mbps)

及び広域ネットワークで接続された 2 台の SPARC Station 間で会話実験をおこなって、vat と試作ツールとの比較実験を行なった。広域ネットワーク上での実験は奈良先端科学技術大学院大学と九州大学の間で行なった。この 2 組織間の回線で最も回線容量の小さい部分は 192Kbps である。また、vat については音声の圧縮方式を切替えても音声途切れがないか確認した。

実験の結果、Ethernet を使用した場合には vat と試作ツールの両方で音声の途切れのない円滑な会話が可能となった。しかしながら、試作ツールでは音声に遅延があるため会話の展開が vat と比較した場合遅くなる。広域ネットワーク上での実験では、vat を使用した場合は音声の途切れが激しく会話は頻繁に中断した。一方試作ツールでは音声の途切れはなく vat と比較すると円滑な会話を行うことができた。

またこの実験で、試作ツールを使用した会話において利用者が遅延の状態を把握できないと会話内容

に食い違いが生じて混乱が発生することがわかった。致します。

このことから、回線の状態に応じて会話に遅延を発生させる場合には、その時の音声の遅延状態を利用者に何らかの形で提示するような機構を実装する必要があることが必要だと考えられる。

5 今後の課題

本稿では、既存の音声会話ツールを実用化する場合の問題点を明確にしてその解決策を提示した。また試作ツールを使用した実験によって、回線状態に応じた遅延を挿入する方式の実装によって円滑な会話が可能となる音声会話ツールを実現できることを確認できた。今後は今回提示した解決策に基づいて、パケットの遅延状態による音声再生の遅延時間の決定方法について考察して、今回提示した解決策を用いた多人数での音声会話のための手順を確定させたいと考えている。

また本稿では音声のみを媒体とした会話ツールに絞って考察してきたが、コミュニケーションの手段として会話ツールを考えた場合、音声だけではなく映像や文字などの媒体も並行して使用可能であることが望ましいと我々は考えている。音声会話と並行して映像や文字を使ったコミュニケーションが可能となるツールでは、音声伝達に今回提示した解決策を用いた場合に音声再生時の遅延とその他の媒体の伝達の同期をとる必要が生じる。今後は音声だけではなく映像や文字を併用してコミュニケーションする場合についても検討を重ねて、さまざまな媒体が並行して使用可能な会話ツールの実現についても考えていきたい。

謝辞

本研究を行なうに当たって、多くの貴重な御助言、御批判を頂いた WIDE プロジェクトの皆様にご感謝致します。また、音声会話実験に協力して頂いた九州大学工学部情報工学科の皆様と研究室の皆様にご感謝

参考文献

- [1] WIDE プロジェクト, “WIDE プロジェクト研究報告書”, 1991.
- [2] N.Borenstein and N.Freed, “MIME(Multipurpose Internet Mail Extensions)”, RFC 1341, Jun. 1992.
- [3] Stephen Casner and Stephen Deering, “First IETF Internet Audiocast”, *ACM SIGCOMM Computer Communication Review*, pp.92-97, Jul. 1992.
- [4] Network Working Group, “Host Extensions for IP Multicasting”, RFC 1054, May. 1988.
- [5] 楠本博之, 藤本真吾, 中村素典, 田中裕之, 堀良彰, “IETF(Internet Engineering Task Force) オーディオコンファレンス実験レポート”, *WIDE Project 12 月研究会報告レポート*, Dec. 1992.
- [6] 宮原秀夫, 尾家祐二 “コンピュータネットワーク”, 森北出版, May. 1992.
- [7] Network Working Group, “Client Requirements for Real-Time Communication Services”, RFC 1193, Nov. 1990.
- [8] CIP Working Group, “Experimental Internet Stream Protocol, Version 2 (ST-II)”, RFC 1190, Oct. 1990.