

UNIXネットワークにおける音声メールおよび会話システムの 実現

6M-8

斉藤忠司 村上勝彦 中川徹 北川一
(豊田工業大学)

1. はじめに

近年、コンピュータネットワークでマルチメディアの研究が盛んであり、市販ワークステーション (WS) やパソコン (PC) にサウンド機能等が標準装備されるに至っている。しかし、いまだに低価格で実用的なアプリケーションが出回っていないのが現状である。一方、UNIX ネットワークにおけるコミュニケーション手段として、従来から mail, talk および NEWS システムが使われてきている。そこで、今回この中の mail および talk について、現有機能に音声通信機能を付加した vmail および vtalk の開発を行った。実現にあたり、特別なハードウェアを追加せず、SPARC WS に標準装備のサウンドボード並びに LAN (Ethernet) を利用する。これらの WS は既存の WAN を介して相互通信も可能である。本研究の目的は、UNIX WS の利用者がネットワークを介して文字と音声で会話しながら各種システムを開発できる実用的な tool を開発する事である。また、LAN だけでなく WAN を介しても会話できるようにする。

2. パケット網による音声通信

パケット網による音声通信の評価としては再生品質で評価するのが一般的である [3][5]。そこで、本研究においても再生品質で評価を行う。以下、LAN および WAN のようなパケット網において音声通信機能を実現するために検討すべき点について述べる。

第一に通信データ量を減らすためにデータの圧縮を行う必要がある。圧縮には元のデータを忠実に再現できる方式 (情報量を保存する圧縮) と、そうでない方式 (情報量削減による圧縮: log 圧縮, ビット削除等) がある。SPARC WS 内蔵のサウンドボードでも既に、8kHz 標本化 12 ビットデータを μ -law 変換 (log 圧縮) で 8 ビットデータに圧縮している。今回はそのデータをさらに圧縮する方式についても検討する。

第二にパケット網における音声通信では、通信データ量削減のために無音部分のケットを送出しない方式が開発されており、会話音声のオンオフパターンに関する研究がなされている [1][2][3]。

第三にパケットの網内遅延時間に揺らぎがあるので、送出側で連続している音声データが、受信側において不連続となり音声の途切れを生じる。そのためリアルタイム性が望まれる通信では、この問題を解決する必要がある [4][5]。これらパケット網を用いた音声通信の特徴を考慮にいれ、現在、vmail および vtalk の開発を進めている。以下、その実現方式について述べる。

3. UNIX WS 上での実現

UNIX WS 上での実現にあたり、(a) 現用の mail や talk と同様のユーザーインターフェースを持つようにする。また (b) vmail と vtalk で、できるだけ共通に使える方式を開発する。(c) 音声の再生品質については、品質が若干悪くなる事を覚悟する。なぜなら本研究では、UNIX 利用者が気軽に使える tool として、また、WAN を介しても通信コストが安価である事を前提として開発を進めているからである。そのため、再生音質の評価については過去に発表された論文 [3][5] とは違う判定基準を定めた。表 1 に判定基準を示す。

Table 1: 評価基準と例

段階	評価基準	例
a	美しい	FM 放送程度の音質
b	単語が良く聞きとれる	AM (中波) 放送程度の音質
c	一応全部聞きとれる	短波放送の近い局 (FEN 等)
d	ときどき分からない	短波放送の海外局
e	聞き返せば分かる	—
f	全く分からない	—

3.1 vmail の実現方式

(1) 音声データの情報量削減による圧縮

SPARC WS 内蔵のサウンドボードでは、8kHz 標本化 12 ビットデータを μ -law 変換で 8 ビットデータに圧縮している。この 8 ビットデータの下位 n ビットを削除したときの会話および音楽の音質について、6 人の被験者に主観評価を行ってもらった。評価結果は表 2、表 3 に示すが値は各段階を選んだ人数である。

この結果より、8 ビットデータの下位 4 ビットを削除しても、短波放送の海外局を受信した程度の音質が再現できる事が判った。この場合、通信データ量は半分になる。会話と音声の音質劣化を比べると音楽の方が音質劣化が少ない。これより、常に音圧レベルの高いものは音質が劣化しにくい事が判った。そこで、音声データのオンオフパターンにより削除ビット数を動的に変える方法が有効であると考えられる。

(2) 音声データのバッキング

会話音声のオンオフパターンは図 1 のように、発声部 (talkspurt)、無音部 (silence) にモデル化できる

Table 2: 評価結果 (音楽)

削除ビット数\段階	a	b	c	d	e	f
0	0	4	3	0	0	0
1	0	3	3	0	0	0
2	0	3	3	0	0	0
3	0	2	4	0	0	0
4	0	0	3	3	0	0
5	0	0	1	2	2	1

Table 3: 評価結果 (会話)

削除ビット数\段階	a	b	c	d	e	f
0	0	5	1	0	0	0
1	0	5	1	0	0	0
2	0	2	4	0	0	0
3	0	0	3	3	0	0
4	0	0	1	5	0	0
5	0	0	0	0	3	3

[1]. これを図2のようにバッキングし、発声部では音声データの前に2バイトのヘッダをつける。また無音部の前には1ビットのヘッダをつけ無音の長さを表示する。この方式は vtalk でも応用でき、その場合、通信の終わりを示す1バイトのデータを付加する。

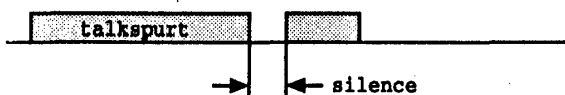


Figure 1: 音声データの構成例

(3) 音声データの情報量保存による圧縮

vmail では、音声メッセージ送出にリアルタイム性が要求されないので、メッセージを圧縮・記録した後にさらに compress コマンドで圧縮をかけて送出し、受信側で忠実にデータを再現出来るようにした。

3.2 vtalk の実現方式

(1) 通信データ量の削減

前述のデータバッキングおよび下位ビットの削除方式が vtalk にも適用できる。ここで、発声部が会話中に約 30% 出現すると仮定し [2], データバッキングおよび下位 2 ビットの削除を行うものとする。そのとき、通信データ量は約 14500 bits/sec となる。これは、19200bps のモデムを通して会話でき、ISDN (B チャネル) では 4 人同時に会話する事ができる通信データ量である。

(2) パケット網内における遅延対策

音声パケットの網内遅延による通話品質の劣化を

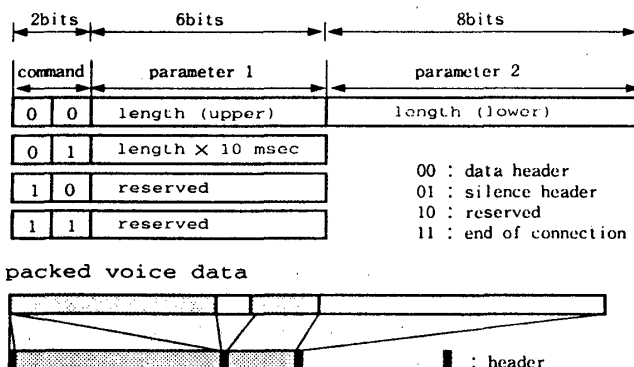


Figure 2: 音声データのバッキング例

防ぐために、受信側で遅延の変動 (遅延揺らぎ) を吸収する必要がある。そのため、受信側で先頭の packets に遅延 (揺らぎ吸収遅延) をかける方式が一般的に用いられる [4]. 本研究では、さらに 0.2 秒の遅延をかけ、LAN や WAN を介した会話が行える方式を開発中である。この方式では、無音データのみを含む packets を次の packet と連結して送出でき、packet 数を減らす事ができる一方で、受信したメッセージが常に 1 テンポ遅れて聞こえる事になる。

4. おわりに

現段階では、vmail については、情報量保存の圧縮を行える実用バージョンが実現できている。vtalk では、文字と音声で通信する基本機能は実現できており、現在、無音削除やビット削除の実装を進めている。

References

- [1] J. G. Gruber: "A Comparison of Measure and Calculates Speech Temporal Parameters Relavant to Speech Activity Detection," IEEE. Trans. Comm., COM-30, No.4, pp. 728-738 (Apr. 1982).
- [2] H. H. Lee and C. K. Un: "A Study of On-Off Characteristics Of Conversational Speech," IEEE. Trans. Comm., COM-34, No.6, pp. 630-636 (June 1986).
- [3] 田中良明, 井上潤吾, 秋山稔: 「日本語会話音声のオンオフパターン」, 信学論 (B-I), Vol. J72-B-1 No. 4, pp. 323-330 (Apr. 1989).
- [4] 鈴木三知男, 榎尾次郎, 土岐隆一, 加藤孝雄: 「音声パケット交換システムにおける遅延適応方式の提案」, 信学論 (B), Vol. J66-B No. 7, pp. 893-899 (July 1983).
- [5] 田中良明, 上野友幸, 井上潤吾, 秋山稔: 「音声パケット通信における瞬断品質評価と適応遅延制御」, 信学論 (B-I), Vol. J73-B-1, No. 7, pp. 610-617 (July 1990).