

TR-ACR-0001

007

VoIP アプリケーションソフトウェアの適応機能が
主観品質に及ぼす影響

The Effect of Adaptability of VoIP Application Software
on Subjective Quality

若林 一磨, 種田 和正

Kazuma Wakabayashi and Kazumasa Oida

2001.10.31

(株)国際電気通信基礎技術研究所
適応コミュニケーション研究所

〒619-0288 京都府相楽郡精華町光台二丁目2番地2

Tel: 0774-95-1501 Fax: 0774-95-1508

Advanced Telecommunications Research Institute International
Adaptive Communications Research Laboratories

2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Telephone: +81-774-95-1501 Fax: +81-774-95-1508

©2001 (株)国際電気通信基礎技術研究所

©2001 Advanced Telecommunications Research Institute International

VoIP アプリケーションソフトウェアの適応機能が主観品質に及ぼす影響

若林 一磨

種田 和正

概要

VoIP アプリケーションソフトウェアは、通信品質変動に適応する機能を複数備える。アプリケーションユーザは、これらを適切に併用することで、音声品質の劣化を抑制する。筆者らは、これら機能の有効性を定量化し、その利用指針を明確にする。このため多様な通信品質のネットワークを設定し、音声品質の主観評価試験を行った。その結果、送信側で冗長な音声情報を付加する Forward Error Correction (FEC) を用いた場合、すべてのネットワークにおいて音声品質の劣化が防げた。一方、失われた音声情報を、直後の音声情報で代替する Packet Repetition (PR) は、それが効果的に機能するネットワークは限られた。特に、通信品質が不安定なネットワークにおいて FEC と PR を併用すると、主観品質を低下させるおそれがあることを確認した。本報告書では、これら適応機能と音声品質の関係を述べる。

目次

1	まえがき	2
2	VoIP アプリケーションの適応機能	2
2.1	RAT における音声データの流れ	3
2.2	ゆらぎ吸収バッファ	3
2.3	Forward Error Correction (FEC)	3
2.4	Packet Repetition (PR)	4
3	評価用ネットワークの模擬方法	4
3.1	ネットワークモデル	5
3.2	RAT を接続するノード	5
4	音声主観品質評価方法	6
5	実験結果	7
5.1	ネットワーク通信品質の評価	7
5.2	アプリケーション適応機能の評価	7
5.3	音声主観品質の評価	10
6	むすび	11

1 まえがき

Voice over IP (VoIP) は、音声をパケットにして送受信することで、音声通信を実現する技術である。この媒体となるインターネットは、通信品質を保証する仕組みを持たない。このため、パケットの転送は、背景負荷トラフィックの影響で遅延する。またふくそうが激しくなると、ノードのバッファがオーバーフローし、パケットは廃棄される。このように通信品質が変動することで、音声品質も劣化する [2, 4]。この音声品質に対する不安が、VoIP 普及の障害となっており技術課題である。その対応策として、VoIP アプリケーションソフトウェア（これ以降、VoIP アプリケーションと略す）は、通信品質変動に適応するように、ゆらぎ吸収バッファや Forward Error Correction (FEC) などの機能を備えている [2, 3, 4, 5, 6]。アプリケーションユーザは、これらを適切に併用することで、音声品質劣化を抑制する。本報告書では、これら機能の有効性を定量化することで、その利用指針を明確にする。

一般に、これら機能の有効性を計測する試験は、実ネットワークの代替に、パケット転送を確率的に遅延させたり、廃棄する環境で実施する。筆者らが試験に用いる VoIP アプリケーションにおいても、パケットを 20-30 % の割合でランダム廃棄させた状況で、FEC が Mean Opinion Score (MOS) を向上させることが示されている [5]。ただ、このように単純な通信品質モデルは、複雑なネットワークの品質として現実的でないと考えた。そこで我々は、ネットワーク模擬装置を開発した。本装置は、ネットワークシミュレーションに基づき End-to-End の音声通信を模擬する。これにより、現実のネットワークに近い、パケットの遅延変動と紛失の関係をモデル化する。この装置を用いて、背景負荷とノードバッファ容量をネットワークパラメータとした条件（以降、NW 条件と略す）の評価用ネットワークを作った。そして、このネットワークにおける VoIP アプリケーションの音声主観品質を評価した。

その結果、FEC を用いた場合、すべての NW 条件において音声品質の劣化が防げた。また、パケット紛失により失われた音声情報を、直後の音声情報で代用する Packet Repetition は、効果的に機能できる NW 条件に限られた。特に、通信品質が不安定となる条件では、FEC と PR を併用すると、主観品質を低下させるおそれがあることを確認した。本報告書は次の構成で、適応機能と音声主観品質の関係について考察する。まず、2 では、VoIP アプリケーションの適応機能を明確にする。次に 3 で、ネットワーク模擬装置とそれがモデル化するネットワークを説明し、4 で、音声主観品質の評価方法を示す。その評価結果を 5 で述べる。

2 VoIP アプリケーションの適応機能

VoIP アプリケーションには、インターネット電話として利用されている Robust Audio Tool (RAT)[5, 8] を使用した。ここでは、まず図 1 を用いて RAT における音声データの流れを説明し、次に適応機能について述べる。

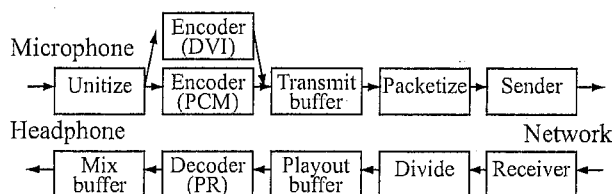


図 1: RAT 内部での音声データの流れ

2.1 RAT における音声データの流れ

1. 送信側処理

マイクやオーディオ形式ファイルから入力された音声は、Unitize にて固定長の音声ユニットに分割される。音声ユニットは、符号化器 Encoder (PCM) にて、符号化ユニットに変換される。Packetize にて、符号化ユニットを組み合わせ RTP[7] ヘッダを付けた音声パケットとし、ネットワークへ送信する。

2. 受信側処理

Receiver に到着した音声パケットを、Divide が符号化ユニットに分解する。各ユニットの再生予定時刻は、RTP ヘッダに記録された送信時刻より決定し、Playout buffer に格納する。再生時刻となった符号化ユニットは、Decoder にて復号される。音声ユニットは混合され、ヘッドホンなど出力装置が音声を再生する。

2.2 ゆらぎ吸収バッファ

音声パケットのネットワーク遅延時間は、中継ノードバッファの混雑状況や通過経路により変動する。これにより Receiver への到着間隔がゆらぐ。これに対処するため、ゆらぎ吸収バッファである Playout buffer は、符号化ユニットの復号処理を保留し、ユニットの時間関係を再構成する。この保留時間より遅れて到着したユニットは、紛失したとみなされる。このユニットを延着ユニットと呼ぶ。

2.3 Forward Error Correction (FEC)

FEC は、送信側で冗長な情報を付加することで、パケット転送時の紛失に備える、送信側主体の復元方法である。RAT が採用している、Media Specific FEC[6] の仕組みを説明する。

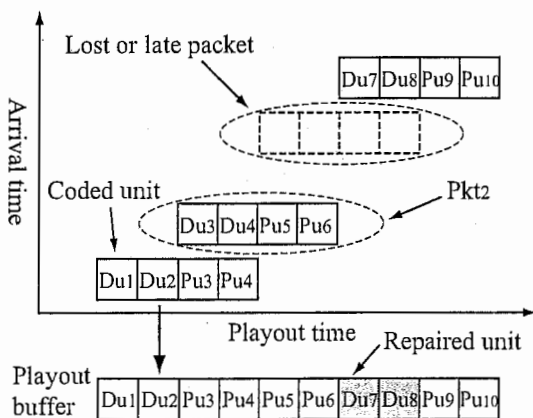


図 2: FEC による紛失パケットの修復方法

図 1 より、 i 番目の音声ユニット Au_i は、2 種類の符号化器 Encoder (PCM) と Encoder (DVI) にて、符号化ユニット Pu_i と冗長なユニット Du_i に変換される。なお、PCM 及び DVI は符号化方式であり、それぞれ PCM μ -law と DVI ADPCM である。パケットの符号化ユニット収容数は、FEC の利用有無とパケット送信間隔に依存する。本報告書では、送信間隔を RAT の初期設定値 40 msec としている。この場合、図 2 の

ように Pkt_k は, Du_{j-2} , Du_{j-1} と Pu_j , Pu_{j+1} の4ユニットで構成される (ただし, $j = 2i + 1$, $k > 0$). 符号化ユニットを Playout buffer に格納する際は, Du_i より高品質な Pu_i を優先する. ただし, Pkt_k が紛失したり延着した場合には, Pkt_{k+1} の Du_{j-2} , Du_{j-1} が復号される.

音声パケット長は, FEC を利用する場合 500 bytes, 利用しない場合 332 bytes である.

2.4 Packet Repetition (PR)

受信者主体の復元方法としては, Silence Substitution (SS), Packet Repetition (PR) など [6] がある. これらは, FEC と異なりネットワークに負荷をかけないという長所がある. SS は, 紛失した音声ユニットを無音で代用することで, 音声の再生間隔を維持する. PR は, 図3のように, 紛失直後の音声ユニットを前方へ複写し, 音声の途切れを隠蔽する方法である.

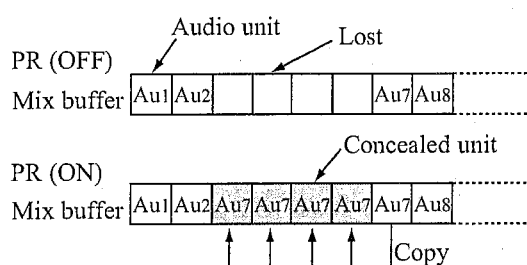


図3: PRによる欠落ユニットの隠蔽方法

RAT では, 最大4回の複写が行われる. この複写により付加されたユニットを隠蔽ユニット (図中 Concealed unit) と呼ぶ. ランダムなパケット廃棄率が20%となる試験環境にて, PRの音声明瞭度は, SSより約15ポイント優れていることを文献 [4] は示している.

表1: RAT条件とその呼称

適応機能		FEC	
		ON	OFF
PR	ON	dual	pr
	OFF	fec	none

これら適応機能を表1の組み合わせで動作させる. この動作条件をRAT条件と呼ぶ.

3 評価用ネットワークの模擬方法

評価用ネットワークにRATを接続し, その適応機能を評価する. この評価用ネットワークを経済的に準備するため, ネットワーク模擬装置を開発した. 本装置は, 次の手順で, 評価用ネットワークの2ノードに接続しているよう通信を模擬する.

1. RATにて音声通信を行い, 音声パケットの発生時刻と長さを計測する.
2. シミュレーションネットワークの2ノード間に, 計測した特徴を持つ疑似音声パケットを発生させる. この疑似音声パケットの遅延と紛失を模擬シナリオとして記録する.

3. 模擬シナリオに基づき、音声パケットの転送を任意の時間保留したり、特定の音声パケットを廃棄する。

3.1 ネットワークモデル

評価用ネットワークは、米国の NSFNET を規範とし、図 4-(a) の構造である。

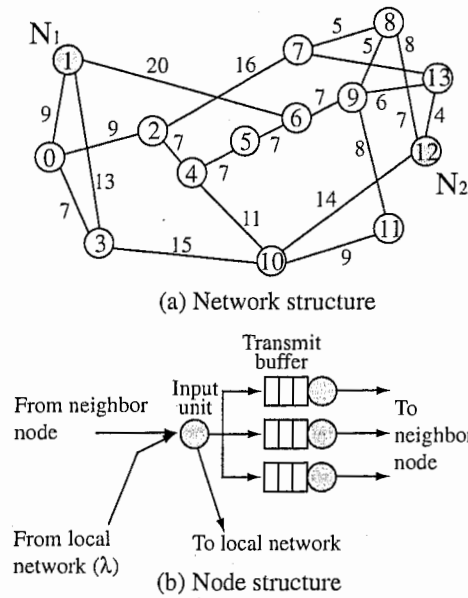


図 4: 評価用ネットワーク

ネットワークは、ノードとリンクで構成される。ノードは、伝送速度 1.5 Mbps の全二重リンクに接続される。リンクには、図のように伝搬遅延時間 (msec) を割り当てる。ノードの内部構造は、図 4-(b) である。各ノードは、入力装置と伝送装置により構成される。伝送装置は、キュー構造のバッファを持ち、その容量は、 C である。評価用ネットワークの通信経路は、シミュレーション開始時に、伝搬遅延時間をコストとして、その最短コストパスを計算し設定する。このネットワークには、定常的に表 2 の背景負荷トラフィックが流入している。背景負荷パケットは、 λ (パケット/2.73 msec) の生起率でローカルネットワークから発生する。この発生時刻を T_a とする。入力装置は、隣接ノードから次ノードを決定する。伝送装置は、パケットを次ノードへ転送する。宛先へ到着したパケットは、評価用ネットワークから消滅する。この消滅時刻を T_d とする。パケットのネットワーク遅延時間は、 $T_d - T_a$ であり、パケット廃棄は、伝送バッファがオーバーフローすることで発生する。

3.2 RAT を接続するノード

ネットワークの品質変動が最も大きいと予想されるノード N_1 , N_2 間にて音声通信を行う。これは、全ノードの組み合わせから、 N_1 , N_2 間の最短コストパスのホップ数とコスト合計 P_d が大きくなる組み合わせを選択した。その結果、 N_1 , N_2 は、それぞれノード 1, 12 であり、通信経路は $1 \rightarrow 6 \rightarrow 9 \rightarrow 13 \rightarrow 12$, P_d は 37 とする。

本報告書では、表 2 の背景負荷 λ と伝送バッファ容量 C の組み合わせで、NW 条件を設定する。これ以降、

NW 条件を, C_n 若しくは $C_n(\lambda)$ で表す.

表 2: NW 条件

パラメータ	設定値
パケット長の分布	負の指数分布 (平均 512 bytes)
パケットの生起分布	ポアソン分布 (平均 λ)
発信元と宛先の分布	一様分布
伝送バッファ容量 C	$C_n = 512 \times 10^n$ (bytes) $n = 1, 3$

4 音声主観品質評価方法

片方向試験で, RAT の適応機能が音声品質に及ぼす影響を測定する.

1. 試験音声と受信音声

ノード N_1 から送信する音声を試験音声と呼び, N_2 の RAT が受信した音声を受信音声と呼ぶ. 試験音声として, ことわざなどの単純な文をオーディオ形式ファイルに録音する. この文間隔は約 1 秒である. RAT 起動と同時に試験音声を送信する. 送信開始後 15 秒からの 15 秒間が評価時間 E_t である. この E_t における受信音声の品質を評価する. このように E_t を設定したのは, ネットワーク及び RAT のバッファが, 適度に満たされた状態にて評価を行うためである.

2. 標準音声と評価音声の評価尺度

受信音声の品質は, 電話系の総合品質評価に用いられる主観品質評価法 [1] で評価する.

最初に, 標準系における受信音声を標準音声とし, その標準音声の品質を絶対品質尺度 (以降, ACR と略す) にて評価する. この評価方法は, 複数の被験者が, 標準音声の品質を日常使用している電話の品質に照らし, どのように感じるかを表 3 の ACR の評定語で評価し評点を記録する. これら評点の平均が MOS である. ACR の MOS を A-MOS と呼ぶ. 標準系とは, NW 条件が $C_3(0.0)$ であり, RAT 条件が dual という組み合わせのシステムである.

次に, 評価系における受信音声を評価音声とし, その品質を妨害品質尺度 (以降, DCR と略す) で評価した. この評価方法は, 評価音声は標準音声に対し, どの程度劣化していると感じるかを表 3 の DCR の評定語で評価する. DCR の MOS を D-MOS と呼ぶ. 評価音声を DCR で評価する理由は, 試験音声の発音などで決まる送話品質や受話者の受聴能力などで決まる受話品質による影響を除去し, ネットワークの通信品質による音声品質への影響のみを計測するためである.

3. 被験者への音声提示方法と集計方法

被験者に音声を, 標準音声, 2 秒間の無音, 評価音声という順序で提示し, 直後 5 秒の無音区間において, DCR の評点を記入させる. 被験者は, 男性 29 名, 女性 22 名であり, 全条件の組み合わせで 3 種類の試験音声を評価させた. 各条件において, 被験者毎に評点の平均値を求めた. 異常値として, これら値の最大及び最小からそれぞれ約 5% 区間に含まれる被験者の平均値を除去した. したがって, 各条件の D-MOS は, 被験者 45 名 \times 試験音声数 3 の平均値である. また, A-MOS も同数の被験者と試験音声で評価した.

表 3: 音声品質評価尺度

ACR の評定語	DCR の評定語	評点
非常によい	劣化がまったく認められない	4
よい	劣化が認められるが気にならない	3
まあよい	劣化がわずかに気になる	2
悪い	劣化が気になる	1
非常に悪い	劣化が非常に気になる	0

5 実験結果

NW 条件と RAT 条件を組み合わせる音声主観品質評価実験を行った。この実験結果を、ネットワーク、アプリケーション、ユーザの観点で評価する。

5.1 ネットワーク通信品質の評価

通信品質は、評価時間 E_t に、ノード N_1 から N_2 へ送信した疑似音声パケットの遅延と紛失率にて評価する。

そのシミュレーション結果が、図 5 である。疑似音声パケットのネットワーク遅延時間の平均が、図 5-(a) であり、標準偏差が図 5-(b) である。平均と標準偏差は、宛先へ到着したパケットを算出対象とした。図 5-(c) は、疑似音声パケットの紛失率である。これらグラフの横軸は、背景負荷の大きさである。ネットワークの通信品質は、受信後処理である PR の影響を受けない。このため、RAT 条件の fec と none のグラフを表示した。

C_1 では、遅延の平均と標準偏差は、背景負荷及び FEC 利用有無に関わらず、ほぼ一定である。また、すべての背景負荷でパケットが紛失し、緩やかに増加する。

$C_3(0.55)$ から $C_3(0.7)$ では、遅延は増加し、標準偏差も大きい。また、紛失率は、 $C_3(0.6)$ から $C_3(0.7)$ へ急速に増加する。したがって、 $C_3(0.55)$ から $C_3(0.7)$ は、ネットワーク処理能力の限界付近かつ不安定な状態である。

5.2 アプリケーション適応機能の評価

ノード N_2 に接続した RAT が処理した音声ユニット数で、適応機能による効果を計測した。これら指標と計測結果を述べる。

1. 受信ユニット

RAT の受信処理部である Receiver (図 1) へ到着したパケットに含まれる符号化ユニットを受信ユニットと呼ぶ。この数も、PR による影響を受けないため、fec と none について説明する。理想的な通信環境において時間 E_t に受信するユニット数は、パケット受信数とパケットのユニット収容数より、fec では $1500 (= 15/0.04 \times 4)$ 、none では $750 (= 15/0.04 \times 2)$ である。なお、理想的な通信環境とは、 E_t におけるパケット受信数がパケット送信数と同一となるネットワークとする。

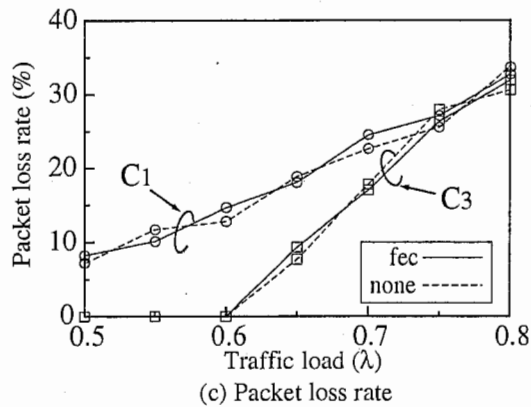
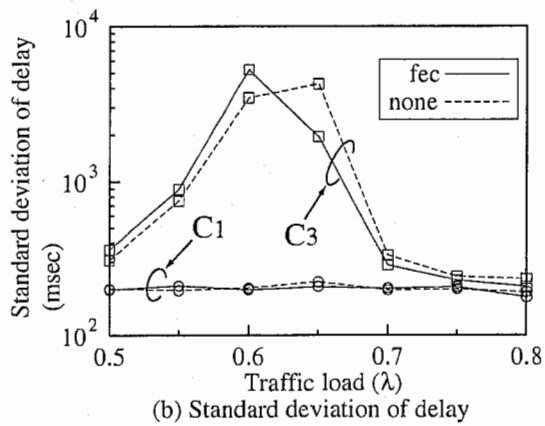
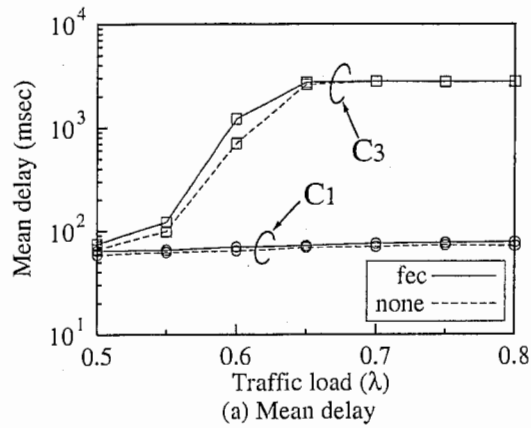


図5: 疑似音声パケットの振る舞い

図6-(a)は、背景負荷と受信ユニット数の関係を表している。背景負荷が大きくなるに従い、受信ユニットは理想数から減少する。全般に、 C_1 の受信ユニットは C_3 より少ない。したがって、受信ユニット数は、遅延よりパケット紛失による影響が大きいと判断する。

2. 再生ユニット

図1のMix bufferから出力した音声ユニットを再生ユニットと呼ぶ。この再生ユニットの数は、受信ユニット数と適応機能の効果で決まる。時間 E_t における再生ユニットの理想的な数は、パケット受信数とパケットの再生対象ユニット数より $750 (= 15/0.04 \times 2)$ となる。この数は、全RAT条件で同一である。

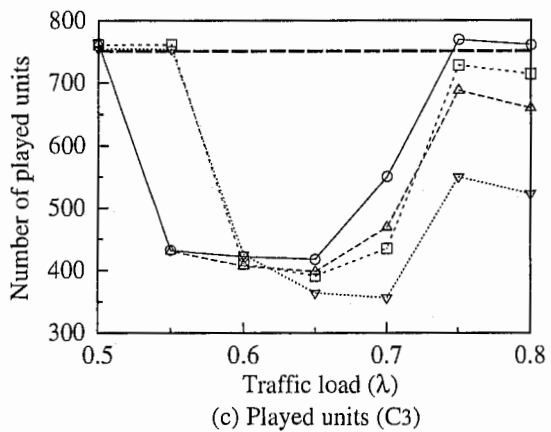
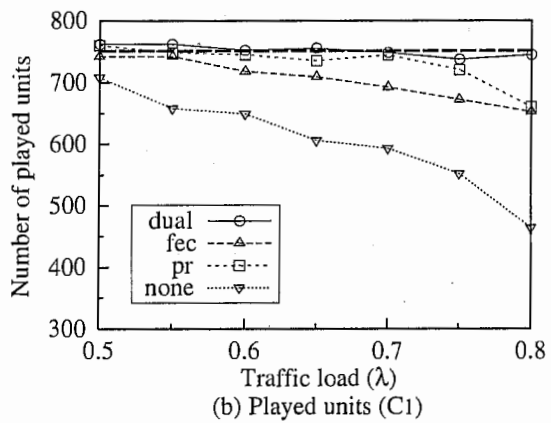
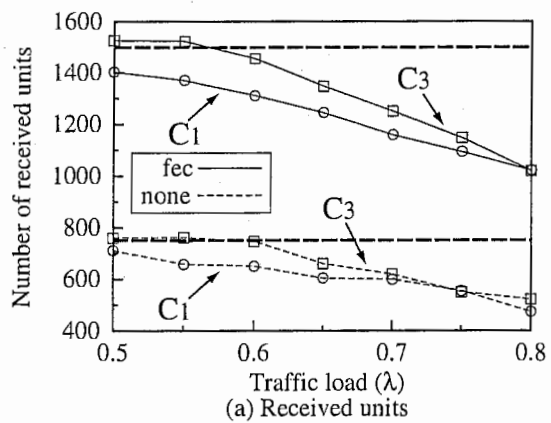


図 6: 音声ユニットと背景負荷の関係

図 6-(b), (c) は、背景負荷と C_1 , C_3 に接続した RAT の再生ユニットの関係を表す。なお、凡例は同じである。 C_1 では、再生ユニット数の減少抑制効果は、 $\text{none} < \text{fec} < \text{pr} < \text{dual}$ である。fec より pr が抑制効果が高い理由は、FEC は冗長なユニットだけで修復するのに対して、PR は 1 つのユニットから最大 4 個の隠蔽ユニットを作成するからと考える。dual は背景負荷が大きくなっても、再生ユニット数を理想数付近に保っている。一方、 C_3 では dual に設定しても理想数から大きく劣化する条件がある。これは、 $C_3(0.55)$ から $C_3(0.7)$ の条件である。特に $C_3(0.6)$ では、RAT 条件に関わらず、再生ユニットは理想数の 60% 弱である。

3. 延着ユニット

$C_3(0.6)$ や $C_3(0.7)$ の none における、受信ユニットと再生ユニットの差が、再生時刻に間に合わず廃棄した延着ユニットの数である。 $C_3(0.75)$, $C_3(0.8)$ では、受信ユニットが減少しているにも関わらず、再生ユニットが増加している。これは、延着ユニット数が減少したからと考える。

5.3 音声主観品質の評価

標準音声の品質を評価した結果、A-MOS は 2.64 であった。これは、アナログ電話の A-MOS が 2.7[1] であることから、VoIP のスコアとして適当と判断する。

NW 条件と RAT 条件を組み合わせ、D-MOS を評価した。前述のとおり、D-MOS は、A-MOS に対する劣化の知覚である。図 7 は、 C_1 , C_3 における背景負荷と D-MOS の関係を表している。 C_1 では、D-MOS の劣化抑制効果は、 $\text{pr} < \text{none} < \text{dual} < \text{fec}$ である。これは、再生ユニット数の減少抑制効果の順序と一致していない。また、 C_3 では、D-MOS の劣化抑制効果の順序が乱れる。これら適応機能の単体及び併用時の有効性を説明する。

1. PR による音声品質への影響

図 8-(a) を用いて、PR の有効性を説明する。横軸は背景負荷の大きさであり、縦軸の D_{pr} は、PR を利用した条件から利用しなかった条件の D-MOS の差である。 D_{pr} が 0 より大きいと、PR が有効に機能していると判断する。ここでは、pr と none の差である pn に着目する。図示された NW 条件では、PR が有効に機能していない条件 $C_1(0.6)$, $C_1(0.7)$ などがある。したがって、PR は音声品質の劣化抑制に、常に有効であると言えない。

2. FEC による音声品質への影響

図 8-(b) より、FEC の有効性を説明する。 D_{fec} は、FEC を利用した条件から利用しなかった条件の D-MOS の差である。 D_{fec} が 0 より大きいと、FEC が有効に機能していると判断する。ここでは、fec と none の差である fn に着目する。すべての NW 条件で D_{fec} は 0 より大きい。また、これら値は、 D_{pr} と比較して、正側に分布している。したがって、FEC は音声品質の劣化抑制に有効で、PR より効果的である。

3. PR と FEC の併用による音声品質への影響

適応機能を併用した dual と単独に利用した条件との D-MOS の差で、併用による有効性を検証する。具体的には、dual と fec の差及び dual と pr の差であり、それぞれ df, dp と記述する。図 8-(a) の df では、ほとんどの NW 条件で D_{pr} が 0 以下である。特に、 $C_3(0.6)$, $C_3(0.7)$ では約 -0.6, -0.8 と劣化が大きい。したがって、PR を併用したことで FEC の有効性が阻害されたことになる。図 8-(b) の dp は、ほとんどの NW 条件で 0 以上であるが、 $C_3(0.6)$ において約 -0.9 と劣化が大きくなる。つまり、この NW

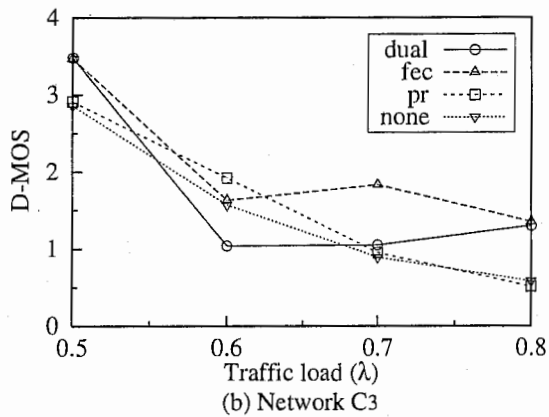
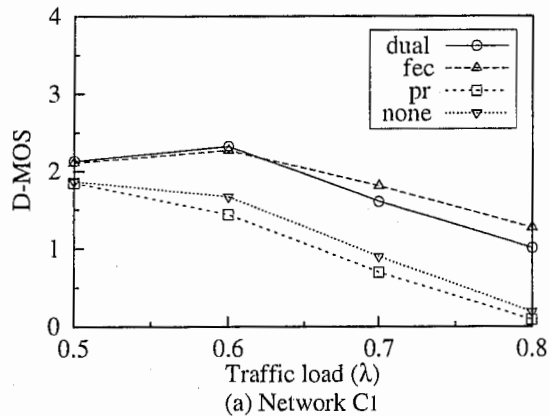


図 7: D-MOS と背景負荷の関係

条件においては、FEC を併用したことで PR の作用が妨害された。

6 むすび

従来の試験環境では、ネットワーク処理能力限界付近の不安定なネットワーク状況を作ることが難しい。本報告書では、独自の模擬装置を用いて、このような状態を作り、適応機能の有効性を定量化した。その結果、以下が判明した。

- $C_3(0.55)$ から $C_3(0.7)$ の NW 条件では、ネットワーク処理能力の限界付近かつ通信品質が不安定な状態であった。
- PR は、再生ユニット数で評価すると有効に機能するように思えたが、D-MOS で評価すると、効果的に利用できる NW 条件は限られていた。なお、有効に機能するのは、通信品質が不安定な状態である。
- FEC を利用すると、パケットが長くなるため、ネットワーク処理能力の限界付近で、遅延の平均や標準偏差が増大する。しかしながら、すべての NW 条件で D-MOS の劣化を抑制し、PR より効果的であった。
- 通信品質が不安定な状態において FEC と PR を併用すると、各々を単独で利用した場合と比較して劣化することがあった。

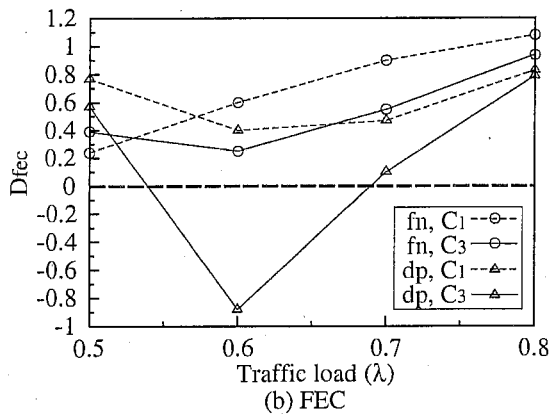
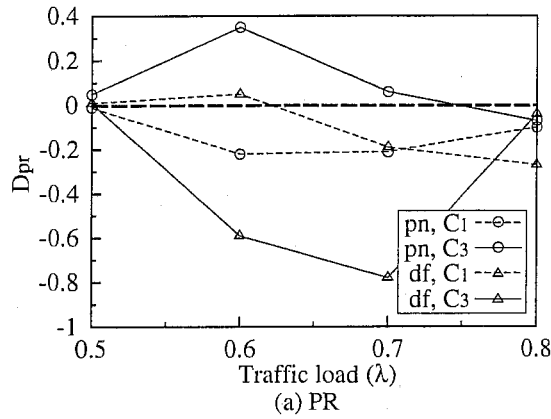


図 8: 適応機能の有効性

つまり、通信品質に応じて機能の効果が異なり、併用することによる劣化もあり得るという結論を得た。これらを踏まえて、今後、VoIP アプリケーションが、RTP[7] の制御パケットである RTCP を用いてネットワーク状態を監視し、自律的に機能の組み合わせを制御することで、MOS の低下を抑制する手法を検討する。

参考文献

- [1] 浅谷耕一, 通信ネットワークの品質設計, (社) 電子情報通信学会, 東京, 1993.
- [2] J. C. Bolot, H. Crepin and A. V. Garcia, "Analysis of Audio Packet Loss in the Internet," Proc. 5th International Workshop on Network and Operating System Support for Digital Audio and Video, pp. 163-174, Durham, New Hampshire, Apr. 1995.
- [3] J. C. Bolot and A. V. Garcia, "The Case for FEC-Based Error Control for Packet Audio in the Internet," ACM Multimedia Systems, Sept. 1996.
- [4] V. Hardman, M. A. Sasse, M. Handley and A. Watson, "Reliable Audio for Use over the Internet," Proc. INET'95, pp. 171-178, Honolulu, Hawaii, Jul. 1995.
- [5] V. Hardman, M. A. Sasse and I. Kouvelas, "Successful Multiparty Audio Communication," Communications of the ACM, Vol. 41, No. 5, pp. 74-80, May 1998.
- [6] C. Perkins, O. Hodson and V. Hardman, "A Survey of Packet-Loss Recovery Techniques for Streaming

Audio," IEEE Network Magazine, Sept./Oct. 1998.

- [7] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC1889, Jan. 1996.
- [8] "User Guide for RAT v3.0.33," University College London, URL: <http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/>, Mar. 1999.