

## 多目的な音声伝送システムの設計

岸田 崇志 河野 英太郎† 前田 香織† 天野 橋太郎‡  
E-mail: takashi@nets.ce.hiroshima-cu.ac.jp

広島市立大学大学院 情報科学研究科  
〒731-3194 広島市安佐南区大塚東 3-4-1  
† 広島市立大学情報処理センター  
‡ 広島市立大学 情報科学部

### 概要

広帯域ネットワークを用いた音声伝送が一般的となってきたり、遠隔講義や遠隔講演も頻繁に実施されるようになってきた。さらに、遠隔合唱も可能になってきている。このような使用状況のなかで音声伝送システムに要求される条件も様々に異なる。本研究では様々な場面ごとに音声伝送に要求される条件を検討し、それぞれの状況に対応できる音声伝送システムの設計・開発をしている。開発中のシステムは、Mbone のツールの一つである RAT (Robust Audio Tool) を拡張し、低遅延伝送や Reed-Solomon 符号を用いた FEC によるエラー回復機構などを実装している。

## Design of a Multipurpose Audio Transmission System

Takashi KISHIDA Eitaro KOHNO Kaori MAEDA Kitsutarō AMANO

Graduate School of Information Sciences, Hiroshima City University  
3-4-1 Ozuka-Higashi, Asa-minami, Hiroshima, 731-3194, Japan

### Abstract

It is becoming easy to use audio communications on the Internet by broadband networks. For example, distance lecture, distance chorus and so on. In our study, we consider conditions of audio communications required in various situations and design a multipurpose audio transmission system called as "MRAT". We enhance RAT (Robust Audio Tool) that is one of the Mbone tools, and add a few new options for keeping lower delay between end nodes and using a Forward Error Correction mechanism with Reed-Solomon code to recover packet losses.

### 1. はじめに

近年、ADSL や光ファイバにより急速に広帯域ネットワークの普及が進んでおり、音声や映像を用いたマルチメディア通信も一般的になってきている。このような通信環境を用いて、遠隔合唱や遠隔合奏などの遠隔地間での音楽セッションを行うといった試みも可能になっている。遠隔合唱を行うにあたって、伝送される音声は高品質なものが求められ、遅延が小さくなければならない。さらに、音声伝送が一般的になってくるにつれ遠隔講演、遠隔講義など

多様な使用場面も増えてきており、そういう状況下で要求される音声の品質はについても考慮しなければならない。

一般に遅延を少なくするためにはジッタ緩衝バッファや再生バッファなどのバッファをできるかぎり小さくするためロバスト性(パケット損失などに対する耐性)が減る。逆に、ロバスト性を増すと処理時間が増大し、遅延時間が増えるといったトレードオフがあり、両方を同時に実現することは難しい。

この問題を解決するために、使用場面(モー

ド)に応じて音声伝送方式を使い分けることが考えられる。我々はモードに応じて最適な状況で使えるように Mbone[1]のツールの一つである音声伝送システム RAT(Robust Audio Tool)[2]を拡張した。

双方向音声会議システムは RAT, VAT[3], SpeakFreely[4], Netmeeting[5]など数多く開発されている。RAT はマルチキャストに対応しており、さらにロバスト性も高く悪条件なネットワークでも音声伝送を行うことができる優れたツールである。さらに、その他の音声会議システムで装備していない多くのコーデックをサポートしており無圧縮 PCM での高音質な伝送も行える。送信するコーデック、チャンネル数、サンプリング周波数など様々なオプションの設定も可能である。そこで本研究では、RAT を元に遅延を少なくしたモード、ロバスト性を強化したモードなどを追加し、使用状況に応じて使えるようにした "MRAT (Multipurpose RAT)"を開発した。

本稿では、2章で多目的音声伝送システムの必要条件の考察とシステムの概要を述べ、3章でシステムの実装について詳述する。さらに4章で評価と MRAT を用いた実践実験を示し、5章で今後の課題についてまとめた。

## 2. 多目的音声伝送システム

### 2.1 音声伝送の利用場面と必要条件

双方向音声伝送の利用場面とそれぞれに要求される条件について考察する。主な音声伝送の利用場面として、遠隔講演や遠隔講義、遠隔会議(会話)、遠隔合唱や遠隔合奏の3通りを想定した。

遠隔講演や遠隔講義は主に通信形態は一对多であり、一方的に音声を流す時間が多く、双方向での会話は少ない。そのため多少の遅延はあってもパケット損失などで音声が聞こえないということは避けたい。

遠隔会議は多対多であり双方向の会話(議論)が多い。ある程度のネットワークの状況であれば、パケット損失が起こり音声が聞き取りにくい場合でも、聞き返すといった行動もとれ、双方で状況を把握できるため、ある程度の対処は行える。遠隔講義のうちゼミ形式のものはこちら含まれる。そこで、この場面では、適度なロバスト性と違和感なく会話が行える程度の遅延時間が確保できればよいと考えた。

遠隔合唱や遠隔合奏は通信形態は多対多で、合唱や合奏の特性上、それぞれの拠点間で音声

のタイミングがずれることは許されず、同期をとることが最優先される。

ITU-T G.114[6]により一方向の伝送時間のガイドラインを参考に、音声会議における遅延は 100ms から 400ms 程度に収まるものとし、遠隔講演や遠隔講義は 400ms 以上でも許容できるとする。遠隔合唱・遠隔合奏は遠隔会議のような会話よりも許容遅延時間が小さくなければならない。予め遠隔地間で合唱する実験を何度か繰り返した際、何度かの実験により、遅延時間が 100ms を越すとお互いの音声にあわせることができず合唱が成り立たなかったため 100ms 未満であると設定した。以上より、表 1 に音声伝送システムの利用場面とその特徴を示す。

表 1. 利用場面とその要求条件

利用場面	遠隔合唱 遠隔合奏	遠隔会議 (議論が主)	遠隔講演 遠隔講義 (一方向が主)
通信形態	多対多	多対多	一对多
許容遅延時間	100ms 未満	400ms 未満	任意
ロバスト性	低	中	高
主たる要求条件	音声のタイミングを同期させること	会話に支障がないこと	確実に音声を届けること
MRAT でのモード	Chorus	Conversation	Broadcast

### 2.2 システム概要

MRAT は、表 1 の 3 つの利用場面に対応できるように開発した音声伝送システムである。MRAT は遠隔合唱や遠隔合奏に用いる「Chorus モード」、遠隔会議や会話に用いる「Conversation モード」、遠隔講義や遠隔講演に用いる「Broadcast モード」を持つ。それぞれをエンド-エンド間の遅延時間とロバスト性の観点から図示したものが図 1 である。図 1 の縦軸はネットワーク転送やジッタ等による遅延と PC による処理遅延を足し合わせたエンド-エンド間のトータルの遅延時間を示し、横軸はパケット損失など悪条件なネットワークにどれだけ耐えうるかというロバスト性を示す。ロバスト性と低遅延とはトレードオフの関係にあり、遅延時間を小さくするためには、なるべく処理遅延を減らし、ジッタ緩衝バッファなどの各種バッファを極力小さく見積もりたい。しかし、そうすることによってロバスト性は減少する。逆に、音声を確実に伝送するためにロバスト性を増すと処理遅延が増えることとなり、全体としての遅延時間は増加する。

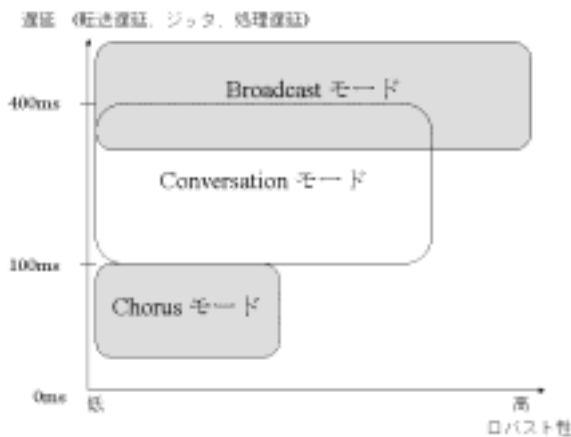


図1．遅延とロバスト性

### 3．MRATの実装

#### 3.1 モードの概要

##### (1) Chorus モード

Chorus モードは、前述したように遠隔合唱や遠隔合奏の利用場面に対応したモードである。遅延を最小減にすることを優先している。

音声を伝送する時に生じる遅延は、ネットワーク上で生じる転送遅延と音声をマシンで処理する際に生じる処理遅延を加えたものとなる。処理遅延の要因となる主なものには、ネットワークでデータを送受信する際に必要となるバッファリング時間、オーディオデバイスで処理する際に必要となるバッファリング時間がある。特に後者が大きいため、それを小さくすることを試みた。オーディオデバイスに過度のオーディオデータが流れ込むとオーディオデバイスのバッファ中のスケジューリングによって音声の途切れが生じる。RAT ではそれを解消するために、マシンの負荷状況に適應してオーディオデバイスのバッファリング時間を最適化するスケジューリングを行っている[7]。

RAT 開発時に比べ、ネットワークも広帯域化し、パケット損失も小さくなっていること、また、CPU の処理性能が向上していることを考慮にいと、現行の RAT で用いられているバッファ処理のスケジューリングでは Chorus モードの遅延の許容時間に対応できない。そこで、MRAT では RAT で用いられているマシン負荷に適應してオーディオデバイスのバッファサイズを調節するパラメータを、現行のマシン仕様や広帯域ネットワークでの使用を対象として改訂した[8]。Chorus モードで使用するパラメータは表 2 に示したものであ

る。

表2．各モードでの設定項目

設定項目	実験での設定値
エンコード方式	Linear-16 (無圧縮)
サンプリングレート	32kHz
チャンネル	モノラル
送信側のパケット損失回復機構	なし
受信側のパケット損失回復機構	パターンマッチング方式
追加再生バッファ時間に対する制限	Chorus あり min=max=0ms
	Conversation あり min=0ms,max=1000ms
	Broadcast なし

##### (2) Conversation モード

Conversation モードは、遠隔会議等双方向で会話を行う際に使用するモードで、遅延時間は表 1 に示すように会話に耐えうる 400ms 以内に設定する。これは、表 2 の元々の RAT のパラメータ設定のみで実現できる。

##### (3) Broadcast モード

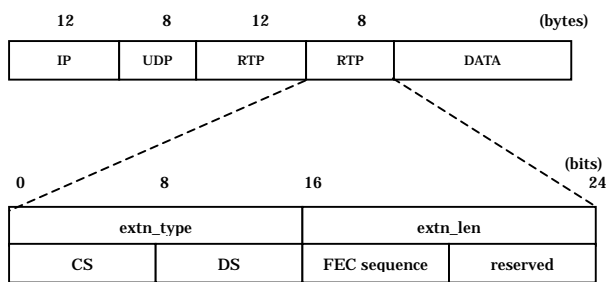
Broadcast モードは、遠隔講演など一地点から多地点へ向かって一方向から送信する際に使用するモードである。MRAT では途中で音声途切れたり、聞こえなかったりすることを回避するようにした。RAT には Lecture モードというオプションが指定でき、このオプションを指定しない場合に比べて少し多めにジッタ緩衝バッファなどのバッファ時間を見積もる。この Lecture モードそのものには Broadcast モードのようなパケット損失回復機構は無いが、Media Specific FEC (Forward Error Correction)[9]のオプションを指定することでエラー回復が図られる。しかし、これでは、インターネットで生じるバーストエラーには対応できない。そこで MRAT の Broadcast モードはバーストエラーに強い Reed-Solomon 符号を用いた FEC を実装し、より高いロバスト性を持つ。

Reed-Solomon 符号はブロック符号の一つであり、送信すべき情報を M 個の連続した複数ビット単位(シンボル)に分割し、このシンボルを単位として符号化、複合化を行う誤り訂正符号である。1 ブロックは  $N(N=2^M-1)$  個のシンボルによって構成される。そのうち K 個がデータシンボルであり、 $N-K$  個が冗長シンボルである場合、 $(N, K)$  の冗長度を持つといい  $(N, K)$  の冗長度をもつ Reed-Solomon

符号を RS ( N , K ) 符号と呼ぶ。また総シンボル数とデータシンボル数の割合 ( N - K ) / N は冗長率と呼ばれ、冗長率を高めると余分な情報量が増加するが誤り訂正能力が高くなる。同じ冗長率でより訂正能力の高い符号がよいとされている。ネットワーク通信においては、バースト損失が発生する傾向があり、バースト誤りに強いとされる Reed-Solomon 符号を用いた FEC を実装することは有益である。冗長度を増やすと処理遅延も増すため、N を 15 とすれば十分な訂正能力があると考え N を 15 としして実装を行っている。Broadcast モードで使用するパラメータは表 2 に示す通りである。

### 3.2 パケットフォーマット

Chorus モードと Conversation モードは、RAT と同じパケットフォーマットを使用する。Broadcast モードのパケットフォーマットを図 2 に示す。RAT は IP/UDP 上で RTP を使っており、Broadcast モードでは、Reed-Solomon 符号を用いた FEC を実装するために RTP 拡張ヘッダを付加し FEC ヘッダとして定義している。extn\_type, extn\_len フィールドは RTP 拡張で定義された固定フィールドであり、extn\_type には 1 を使用している。これに CS, DS, FEC シーケンスフィールドを追加して FEC ヘッダとした。



CS: code symbol size DS: the number of data symbols

図 2. パケットフォーマット

### 3.3 実装状況

表 3 に示した PC 上で MRAT の実装を行っている。Chorus モード, Conversation モードの実装を終え、Broadcast モードはエンコード方式 Linear-16 固定で実装を進めている。そのためペイロード長は 320byte 固定である。今後、RAT に実装されている全てのエンコード方式にも対応する予定である。

表 3. PC の仕様

CPU	Pentium III 1.0GHz
Memory	512 MB
OS	Vine Linux 2.1
Soundcard	SoundBlaster Live! Value

## 4. 評価

### 4.1 遅延時間測定

実際に遅延時間の測定を行い遠隔合唱に耐えうる遅延時間になっているかを測定した。実験環境を図 3 に示す。Host A の仕様は Pentium 600MHz, OS は Vine Linux2.5, Host B の仕様は Pentium 1GHz, OS は Vine Linux2.1 である。



図 3. 実験環境

Host A で 1 秒間に 1 回のテンポでメトロノームの音を入力し、その音を Host B で受信しスピーカーから出力する。録音用 PC では、メトロノームの音とネットワークを経て Host B 経由で戻ってきた音を録音し、それを波形編集ソフトである spwave[10]で波形の時間軸のずれを比較し測定した。測定結果を表 4 に示す。Broadcast モードは冗長度(15,11), (15,12), (15,13)それぞれに対して遅延時間を測定した。遅延時間は波形をランダムに 10 個選んだ平均を表示している。

表 4. 遅延測定結果

モード	片側の遅延時間 (ms)
Broadcast (15,11)	143
Broadcast (15,12)	138
Broadcast (15,13)	138
Conversation	132
Chorus	72

表 4 より、Conversation モードは片側の遅延時間が 132ms, Chorus モードは 72ms となり、表 1 で定義した条件を満たしている。

Broadcast モードでも、いずれの冗長度でも表 4 で示す許容遅延時間よりも小さく実装できている。この測定値は十分に速く条件のよいネットワーク上での結果で、現実装における MRAT の処理遅延に近い。パケット損失等のあるネットワーク上でも同様の実験をしていく。

#### 4.2 Broadcast モードのエラー回復

Reed-Solomon 符号を用いた FEC を使用した場合どれだけパケット損失が回復できるかを測定した。測定環境を図 4 に示す。表 5 は測定に使用した機器である。

##### (a)測定方法

測定方法は MRAT(Broadcast モード)を使用し Host A から Loss generator[11]経由で Host B へ伝送する。その際、Loss generator では 1, 2, 4, 6, 8, 10%の確率でパケット損失を発生させた。N は 15 固定とし、K を 11, 12, 13 と変化させ、それぞれの場合について FEC デコード後のパケット損失率を測定した。以下、ユニット内のデータパケット数の組み合わせを(N, K)と示す。

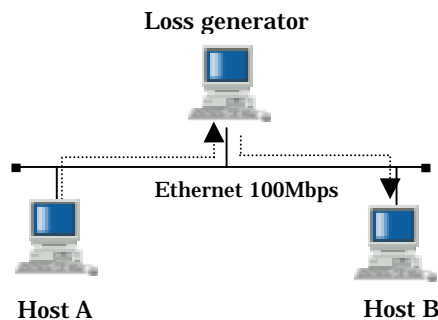


図 4. 測定環境

表 5. 測定に使用した機器

PC	CPU	Memory	OS
Host A	PentiumIII 600MHz	384MB	Vine Linux 2.5
Loss generator	PentiumII 300MHz	384MB	Vine Linux 2.1
Host B	PentiumIII 1GHz	512MB	Vine Linux 2.1

##### (b)理論値の算出

FEC 復元後のパケット損失率の理論値 P は次のようになる[12]。データパケット数の組み合わせを(N, K)とし、各パケットの損失は独立であり個々のパケットの損失率を p とする。i はデータパケット K 個中 FEC デコード後に

損失しているパケット数である。

$$P = \frac{1}{K} \sum_{i=0}^K i \binom{K}{i} p^i (1-p)^{K-i}$$

$$\sum_{j=t}^{N-K} \binom{N-K}{j} p^j (1-p)^{N-K-j} \quad (1)$$

$$t = \begin{cases} 0 & \text{if } N-K < i \leq K \\ N-K-i+1 & \text{if } 0 \leq i \leq N-K \end{cases}$$

##### (c)測定結果

パケット損失率の測定結果とともに式(1)から算出したそれぞれの場合における理論値を図 5 に示す。横軸は FEC 復元前のパケット損失率、縦軸は FEC 復元後のパケット損失率である。図 5 に示したように Reed-Solomon 符号を用いた FEC を用いることによってパケット損失率を少なく抑えることができた。また、十分なデータ数(10 万パケット以上)を採取し、理論値と実測値はほぼ一致していることがわかる。

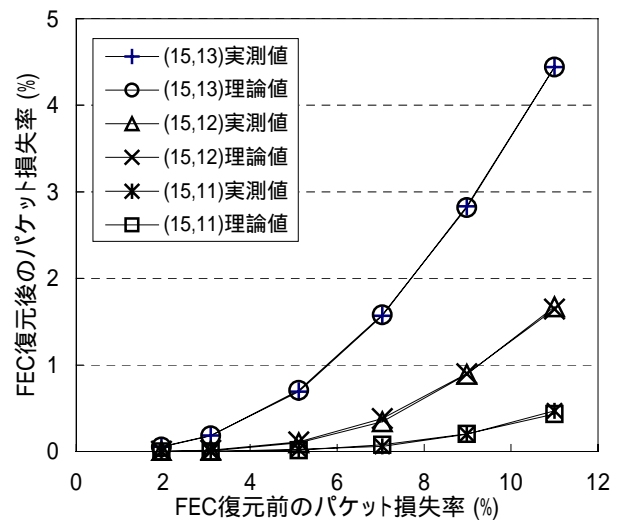


図 5. Broadcast モードにおける FEC 性能測定

#### 4.3 遠隔合唱の実験実験

Chorus モードを用いて 2 拠点間で遠隔合唱を行った。この合唱は広島市立南観音小学校と広島市立白島小学校で「ママ de がんすプロジェクト」[13]の実証実験ネットワークを用いて行った。遠隔合唱を行う 2 拠点のうち一方で伴奏をすると、その伴奏音と相手に伝送される伴奏音がずれるため、伴奏は合唱をしない第 3 の地点(広島市立大学)から送信した。各合唱地点(白島小, 南観音小)は伴奏を聞きながらそれぞれアルトパート、ソプラノパートに分



かれそれぞれ相手の歌声をネットワーク経由で聞きながら合唱した。実験構成図を図 6 に示す。各拠点間はマルチキャストで伝送し、MRAT(Chorus モード)で使用した帯域は各地点それぞれ 1Mbps であった。両校の映像は MPEG2 伝送システムの MPEG2TS[14]を用いた。MPEG2TS は MRAT の動作する PC 上で動き、映像に 4Mbps を使用した。図 7 はそのときの様子である。RAT では、片側の遅延が大きく合唱を行えなかったが、MRAT を用いることで遅延は片側 70~75ms 程度となり、わずかなエコーを感じる程度に収まった。この時のネットワークのジッタは約 7ms(RAT の統計情報による表示)であった。各小学校約 100 名の児童による合唱では、同時に MPEG2 伝送をしている PC 上で MRAT を用いても転送遅延及び、マシンでの処理遅延を感じることなく違和感のない合唱が可能だった。



図 6 . 実験構成図



図 7 . 遠隔合唱の様子

## 5 . まとめ

本稿では、使用場面に応じて音声伝送方式を使い分ける音声伝送システムの設計と実装について述べた。今後は、Chorus モードや Broadcast モードを用い遠隔合唱や遠隔講演などの実践実験に使用し、実用性の評価や、主観評価による音質の評価も行いたい。

## 謝辞

本研究の一部は広島市立大学特定研究費(平成 13 年度 1803, 平成 14 年度 2206), 通信・放送機構 JGN プロジェクト(P3410005, G13013)の支援を受けて実施されている。ここに記して謝辞を記す。

- [1]Thomas Maufer, "Deploying IP multicast in the enterprise", Prentice Hall PTR, pp.147-153(1998)
- [2]"Robust Audio Tool", <http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/index.html>
- [3]"VAT", <http://www-nrg.ee.lbl.gov/vat/>
- [4]"Speak Freely", <http://www.speakfreely.org>
- [5]"Netmeeting", <http://www.microsoft.com/japan/windows/NetMeeting/default.htm>
- [6]"ITU-T G Series", <http://www.itu.int/publibase/itu-t/ItutAllBySeries.asp?serie=G>
- [7]I.Kouvelas and V.Hardman, "Overcoming Workstation Scheduling Problems in a RealTime Audio Tool", in Proceedings of Usenix Annual Technical Conference, Anaheim, California. pp.235-242 (1997)
- [8]岸田 崇志, 河野 英太郎, 前田 香織, "多地点における高品質な遠隔合唱の実現", 広島地域の学校における高度マルチメディア通信に関する研究プロジェクト 研究成果報告書, pp93-98(2002)
- [9]C.Perkins, I.Kouvelas, O.Hodson, V.Hardman, M.Handley, J.Bolot, A.Vega-Garcia and S.Fosse-Paris, "RTP Payload for Redundant Audio Data", IETF Audio/Video Transport Working Group, RFC 2198 (1997)
- [10]"spwave", <http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/spLibs/spwave/>
- [11]近堂 徹, "MPEG2 over IPv6 システムの性能評価", 広島大学工学部第 2 類 卒業論文(2001)
- [12]大塚 玉記, 西村 浩二, 相原 玲二, 前田 香織, "FEC を用いた MPEG2 over IP システムの開発と評価", 情報処理学会 DSM 研究会, pp.43-48(2001)
- [13]マメ de がんすプロジェクト, <http://www.csi.ad.jp/activity/MAMEdeGansu>
- [14]近堂 徹, 大塚 玉記, 西村 浩二, 相原 玲二, "MPEG2 over IP 伝送システム mpeg2ts の開発と性能評価", DICO 2002 シンポジウム論文, pp.157-160(2002)