

## PHSによる蓄積メディア伝送時のメディア同期性能評価

加藤 正美<sup>†</sup>      白井 憲義<sup>††</sup>      田坂 修二<sup>†††</sup>

### Performance Evaluation of Stored Media Synchronization in PHS

Masami KATO<sup>†</sup>, Noriyoshi USUI<sup>††</sup>, and Shuji TASAKA<sup>†††</sup>

あらまし 本研究は、PHSの二つの通信チャネルを使用し、蓄積されたH.263圧縮ビデオ情報をARQで、蓄積されたADPCM音声情報を無手順で伝送しながら再生する場合を考える。このとき、H.263の符号発生量とPHSの伝送容量がほぼ等しいため、再送による遅延はビデオのメディア内同期を乱す。特に、蓄積されたビデオの伝送では、通信状態に応じた符号発生量の制御が難しく、メディア内同期をいかに制御するかが課題となる。更に、別の通信チャネルで同時に音声も伝送する場合、ビデオのメディア内同期の乱れは、音声とビデオのメディア間同期の乱れとしても問題になる。そこで本研究は、ビデオや音声のメディア内同期や両者のメディア間同期を維持するために、送信側においてBブロック廃棄制御方式を提案する。更に、受信側において出力開始遅延制御を適用すると共に、スライド制御方式を提案する。そして、ランダム誤りとバースト誤りの各環境下で、実際のビデオ情報と音声情報を使用したシミュレーションを行い、メディア同期の性能評価、および主観評価により本提案方式の有効性を示す。

キーワード PHS, ビデオ伝送, 蓄積メディア, メディア同期

### 1. ま え が き

第2世代移動通信システム[1]の一つであるPHS(Personal Handy Phone System)[2]は、32 kbit/s音声(通話)サービスに加えて、32 kbit/s非制限デジタルベアラサービスを実現した。これによりPHSは、音声だけでなく、ビデオを含むデジタルデータの伝送も可能となり、マルチメディアに対応できる無線インフラストラクチャとなった。更に、複数の通信チャネルを利用すれば、音声とビデオの統合伝送も実現可能であり、PHSによるモバイル・ビジュアル・コミュニケーションへの期待が大きくなりつつある。

PHSでビデオ伝送を行うには、H.263[3]等の低ビットレート対応の動画像圧縮符号化方式を利用する。この場合、H.263の符号発生量とPHSの伝送容量がほぼ等しくなる。そのため、誤りに弱い圧縮ビデオ情報の伝送にARQ(Automatic Repeat Request)を適用すると、再送によ

る遅延の発生が、送信側でのビデオ発生間隔と受信側でのビデオ表示間隔との同期、いわゆるメディア内同期の乱れとして問題になる[4]。

例えばライブ伝送では、ビデオの空間的・時間的解像度制御により、通信状態に応じて符号発生量を制御すれば[5]、メディア内同期を良好に維持できる。しかし、ビデオ・オン・デマンドのように蓄積された圧縮ビデオ情報を伝送しながら再生する場合、もはや空間的解像度制御は困難である。更に、不用意に時間的解像度制御を行えば、フレーム間予測符号化された画像は連続して乱れてしまう。このように、蓄積された圧縮ビデオ情報をPHSで伝送しながら再生する場合、どのようにメディア内同期を制御するかが課題となる。

更に、PHSの一つの通信チャネルでARQによるビデオ伝送を行う場合には、別の通信チャネルで音声を伝送する必要があることが多い。そこで、音声サービスの利用を考える。これは、通話用に設けられたコーデック等の回路を共用することで、コスト削減および小型化を実現するための判断である。このとき、32 kbit/s ADPCM(Adaptive Differential Pulse Code Modulation)で符号化かつ蓄積された音声情報が、無手順で伝送されるため、音声のメディア内同期は乱れない。しかし、再送によるビデオのメディア内同期の乱れが、音声とビデオのメディア間同期の乱れとして問題になる。

<sup>†</sup> 三洋電機株式会社ハイパーメディア研究所, 岐阜県  
Hypermedia Research Center, SANYO Electric Co., Ltd., Gifu-ken,  
503-01 Japan

<sup>††</sup> 三洋電機株式会社情報通信開発研究所, 岐阜県  
Information & Telecommunications Development Center, SANYO  
Electric Co., Ltd., Gifu-ken, 503-01 Japan

<sup>†††</sup> 名古屋工業大学電気情報工学科, 名古屋市  
Department of Electrical and Computer Engineering, Nagoya Institute of  
Technology, Nagoya-shi, 466 Japan

特にH.263では、低ビットレート化のためにフレーム内符号化フレーム(1フレーム)を周期的に挿入しておらず、1フレーム直前の複数ビデオフレームを一度に廃棄してメディア内同期を回復することはできない。すなわち、ひとたびビデオのメディア内同期が大きすぎてしまうと、もはやビデオを音声に再同期できない。このように、蓄積されたビデオと音声のメディア間同期をどのような制御するかも大きな課題である。

無線通信のメディア同期に関するこれまでの研究として、[6]はシェアドメディア型の無線LANによる蓄積メディア伝送において、ビデオと音声のメディア同期性能を報告している。また、[4]は帯域保証された回線交換型のPHSによるライブメディア伝送において、ビデオのメディア内同期性能を報告している。しかし、PHSによる蓄積メディア伝送において、ビデオのメディア内同期性能、更にビデオと音声のメディア間同期性能は明らかになっていない。一方、文献[5]はPHSによるビデオと音声のライブ伝送において、通信状態に適応したビデオの空間的・時間的解像度制御方式を検討している。しかし、メディア同期の検討や、蓄積メディア伝送におけるビデオの解像度制御方式の検討はなされていない。

そこで本研究は、蓄積されたH.263圧縮ビデオ情報と32 kbit/s ADPCM音声情報を、それぞれPHSの一つの通信チャンネルを使用し、サーバからPHS基地局を介してPHS移動局に伝送しながら再生する場合を考える。特に、ビデオはARQで、音声は無手順で伝送する(図1)。

このとき、ビデオや音声のメディア内同期や、両者のメディア間同期を維持するために、送信側において、通信状態により双方向予測符号化情報を廃棄するBブロック廃棄制御方式を提案する。また受信側において、メ

ディア内同期の一時的な乱れをバッファで吸収する出力開始遅延制御を適用する。更に、ビデオのメディア内同期の状態により、音声への影響を考慮しながら遅れたビデオに音声を再同期するスライド制御方式を提案する。

そして、ランダム誤りとバースト誤りの各環境下で、実際のH.263圧縮ビデオ情報と32 kbit/s ADPCM音声情報を使用したシミュレーションを行い、メディア同期の性能評価、および主観評価により本提案方式の有効性を示す。

本論文の構成は次のとおりである。2.では提案方式について述べる。3.ではシミュレーションの仮定について述べる。4.では数値結果と性能評価を報告する。

## 2. メディア同期制御方式

本論文は、メディア同期、すなわちビデオや音声のメディア内同期、および両者のメディア間同期を検討の対象とする。

ここで、文献[6]に従い、各メディアは、その出力単位であるメディアユニット(MU)から構成されているものと仮定する。特に、ビデオは1ビデオフレームを1ビデオMUとし、音声は一定の情報量をもって1音声MUを構成すると仮定する。そして、メディア発生源*i*における*n*番目のMUの発生時刻を $T_n^{(i)}$ 、*n*番目と*n+1*番目のMUの発生間隔を $\sigma_{n,n+1}^{(i)} (= T_{n+1}^{(i)} - T_n^{(i)})$ と定義する(図2)。なお、以後*i*=1はビデオ、*i*=2は音声を示す。

このとき、メディア内同期は、単一メディア*i*を構成するMU間の発生間隔 $\sigma_{n,n+1}^{(i)}$ を、メディア出力先においても出力間隔として維持することである。更に、メディア間同期は、メディア1とメディア2の間で対応するMUの発生間隔 $T_n^{(1)} - T_m^{(2)}$ を、メディア出力先において同様に維持することである。

### 2.1 メディア内同期

まず、メディア内同期について考える。ここで、メ

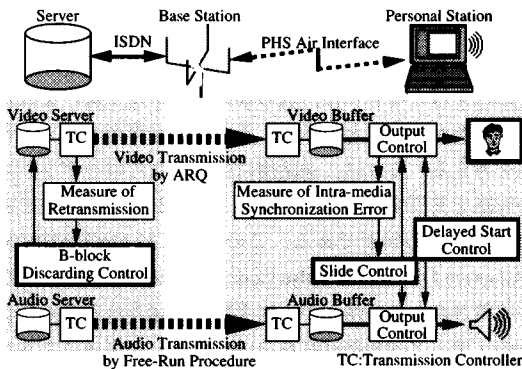


図1 PHSビデオ・音声伝送システム構成図

Fig. 1 Block diagram of video & audio transmission system for PHS.

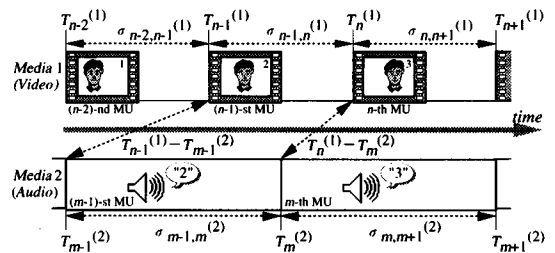


図2 ビデオMUと音声MUの時間関係

Fig. 2 Temporal relationship between video and audio MUs.

ディア出力先におけるメディア*i*の*n*番目のMUの目標出力時刻を $t_n^{(i)}$ , 到着時刻を $A_n^{(i)}$ , 出力時刻を $D_n^{(i)}$ , 到着してから出力されるまでの出力待ち時間を $\tau_n^{(i)} (= D_n^{(i)} - A_n^{(i)})$ と定義する[6].

更に本研究では、 $D_n^{(i)}$ を $t_n^{(i)}$ から意図的に遅らせる時間量としてスライド時間 $\Delta S_n^{(i)}$ を明示的に定義する。これは、ビデオのメディア内同期の状態に応じて、音声への影響を考慮しつつ行われるものであり、2.2.2のスライド制御において導入される。また、総スライド時間を $S_n^{(i)} = S_{n-1}^{(i)} + \Delta S_n^{(i)}$ と定義する。このとき、 $t_n^{(i)}$ ,  $D_n^{(i)}$ , およびメディア内同期ずれ時間 $\Delta_n^{(i)}$ は、次式で表される。

(a) MUの目標出力時刻

$$t_1^{(i)} = A_1^{(i)} + \tau_1^{(i)} \quad (n=1)$$

$$t_n^{(i)} = t_{n-1}^{(i)} + \sigma_{n-1,n}^{(i)} + S_{n-1}^{(i)} \quad (n \geq 2)$$

(b) MUの出力時刻

$$D_1^{(i)} = t_1^{(i)} \quad (n=1)$$

$$D_n^{(i)} = t_n^{(i)} + \Delta S_n^{(i)} \quad (n \geq 2, A_n^{(i)} \leq t_n^{(i)} + \Delta S_n^{(i)})$$

$$D_n^{(i)} = A_n^{(i)} \quad (n \geq 2, t_n^{(i)} + \Delta S_n^{(i)} < A_n^{(i)})$$

(c) メディア内同期ずれ時間

$$\Delta_1^{(i)} = D_1^{(i)} - t_1^{(i)} = 0 \quad (n=1)$$

$$\Delta_n^{(i)} = D_n^{(i)} - t_n^{(i)} = \Delta S_n^{(i)} \quad (n \geq 2, A_n^{(i)} \leq t_n^{(i)} + \Delta S_n^{(i)})$$

$$\Delta_n^{(i)} = D_n^{(i)} - t_n^{(i)} = A_n^{(i)} - t_n^{(i)} \quad (n \geq 2, t_n^{(i)} + \Delta S_n^{(i)} < A_n^{(i)})$$

例えば図3において、1番目のMUは到着時刻 $A_1^{(i)}$ から $\tau_1^{(i)}$ 経過後に出力される。*n*番目のMUは、到着時刻が $A_n^{(i)} \leq t_n^{(i)}$ の場合(CASE 1), 若しくは $t_n^{(i)} < A_n^{(i)} \leq t_n^{(i)} + \Delta S_n^{(i)}$ の場合(CASE 2), 時刻 $t_n^{(i)} + \Delta S_n^{(i)}$ に出力される。また $t_n^{(i)} + \Delta S_n^{(i)} < A_n^{(i)}$ の場合(CASE 3), 時刻 $A_n^{(i)}$ に出力される。特に $\Delta S_n^{(i)} = 0$ の場合、到着時刻が $A_n^{(i)} \leq t_n^{(i)}$ では時刻 $t_n^{(i)}$ ,  $t_n^{(i)} < A_n^{(i)}$ では時刻 $A_n^{(i)}$ に出力される。すなわち、メディア内同期制御とは、メディア内同期ずれ時間 $\Delta_n^{(i)}$ が最小になるように制御することである。

### 2.1.1 ビデオのメディア内同期

本研究では、PHSの32 kbit/s非制限デジタルベアラサービを利用し、H.263圧縮ビデオ情報をARQで伝送することを考える。特に、ビデオのメディア内同

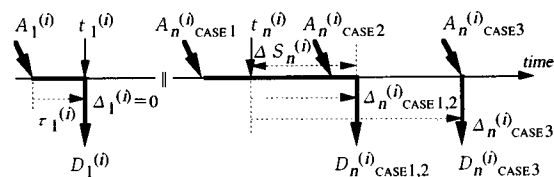


図3 メディア内同期ずれ

Fig. 3 Definition of intra-media synchronization error.

期を制御するために、送信側においてBブロック廃棄制御を提案すると共に、受信側において出力開始遅延制御を採用する(図1).

#### (1) Bブロック廃棄制御

PHSの伝送容量とH.263の符号発生量がほぼ等しいため、再送による遅延を回復するには、送信側において伝送すべき情報量を削減することが効果的である。

一般的に、圧縮符号化されたビデオフレームは、適用した予測符号化方式によりフレーム内符号化フレーム(Iフレーム), フレーム間予測符号化フレーム(Pフレーム), 双方向予測符号化フレーム(Bフレーム)に分けられる[7]. 特にH.263は、PフレームとBフレームを一つのビデオフレーム(PBフレーム)として符号化する。すなわち、Bフレームに関する情報をBブロックとしてPフレームの中に埋め込み、ヘッダ情報等を削減している。

ここでビデオフレームの廃棄を考える。H.263では、符号量の大きなIフレームを周期的に挿入せず、マクロブロック単位で132回に1回以上のフレーム内符号化を規定している。従って、画像の乱れが発生しないようにIフレームの直前で一度に複数のビデオフレームを廃棄することはできない。また、PフレームやPBフレームを廃棄すれば、画像の乱れが連続したビデオフレームに現れてしまう。しかし、Bフレームに相当するBブロックは他のビデオフレームから参照されないため、廃棄しても画像は乱れない。

そこで本研究では、Bブロックを送信側で廃棄することを考える。すなわち、送信側において再送の発生回数を計測し、再送に要した情報量に相当する量のBブロックを、これ以後に送信するPBフレームから廃棄する。これをBブロック廃棄制御と呼ぶ。これにより伝送すべき情報量は削減され、メディア内同期の回復が期待できる。

#### (2) 出力開始遅延制御

最初のMUの受信処理を開始した受信開始時刻から、その受信処理を完了する到着時刻 $A_1^{(i)}$ までの時間を $a_1^{(i)}$ と定義する。このとき受信開始時刻から最初のMUの出力時刻 $D_1^{(i)}$ までの出力開始遅延時間 $S_0^{(i)}$ は次式で与えられる。

$$S_0^{(i)} = a_1^{(i)} + D_1^{(i)} - A_1^{(i)} = a_1^{(i)} + \tau_1^{(i)}$$

なお文献[6]と同様に、再送による遅延の最大値の見積値を $\Delta_{max}$ と定義するならば、 $\tau_1^{(i)} = \Delta_{max}$ と考えられる。

ここで、受信側におけるビデオ情報のバッファリングについて考える。容易にわかるように、 $S_0^{(i)}$ を大きく設定し、ビデオ情報のある程度バッファリングしてから出力を開始すれば、一時的なメディア内同期の乱

れをバッファにより吸収できる。これを出力開始遅延制御と呼ぶ。

特に、PHSの伝送容量とH.263の符号発生量がほぼ等しいため、再送が発生しなくても、各ビデオフレームサイズの大小がメディア内同期に影響を及ぼす。またPBフレームでは、先に表示するBフレームの目標出力時刻 $t_n^{(1)}$ までに、PBフレーム全体が到着( $A_{n+1}^{(1)} \leq t_n^{(1)}$ )しなければならない。従って、出力開始遅延制御の適用は必須と考える。

### 2.1.2 音声のメディア内同期

本研究では、PHSの32 kbit/s音声サービスを利用し、32 kbit/s ADPCMにより符号化された音声情報を、通話と同様に無手順で伝送することを考える。従って、必ず $A_n^{(2)} \leq t_n^{(2)}$ となり、常にメディア内同期は維持される。

しかし、2.2.2に示すスライド制御により、出力時刻 $D_n^{(2)}$ をスライド時間 $\Delta S_n^{(2)}$ だけ意図的に遅らせた場合、音声のメディア内同期は $\Delta S_n^{(2)}$ だけ乱れる。

## 2.2 ビデオと音声のメディア間同期

### 2.2.1 メディア間同期

つぎに、メディア間同期について考える。ここで、メディア $i$ を構成するMUの単位時間当りの発生数(MU発生レート)を $M_R^{(i)}$ と定義し、各メディアのMU発生レートは異なるが、固定値であると仮定する。このとき、メディア1と2のメディア間同期ずれ時間は、メディア1の $n$ 番目のMUの出力時刻 $D_n^{(1)}$ と、このMUの直前に発生したメディア2の $m$ 番目のMUの出力時刻 $D_m^{(2)}$ とを比較した次式で与えられる。

$$\Delta_n^{(1-2)} = D_n^{(1)} - D_m^{(2)} - \{(n-1)/M_R^{(1)} - (m-1)/M_R^{(2)}\}$$

$$\text{但し、} m = \text{int}\{(n-1) \times M_R^{(2)}/M_R^{(1)}\} + 1, n \geq 1$$

$\text{int}(x)$ は $x$ の整数値を求める関数を示す。

### 2.2.2 スライド制御

ビデオをARQで、音声を無手順で伝送する場合、再送によるビデオのメディア内同期の乱れは、音声とビデオのメディア間同期の乱れとしても問題となる。

特にH.263では、周期的なIフレームの挿入がないため、画像を乱さないようにIフレームの直前で一度に多くのビデオフレームを廃棄することができない。従って、ひとたびメディア内同期が大きすぎてしまうと、一度に多くのビデオフレームを廃棄してビデオを音声に再同期させることはできない。

そこで本研究では、逆に、遅延したビデオに音声を再同期し、音声のメディア内同期を一時的に乱してでもメディア間同期を維持することを考える。すなわち、ビデオをメディア間同期の基準となるマスターメディア、

音声をマスターメディアに再同期しながら出力するスレーブメディアと考える。そして、ビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ に応じて、出力時刻 $D_n^{(0)}$ を目標出力時刻 $t_n^{(0)}$ からスライド時間 $\Delta S_n^{(0)}$ だけ遅らせ、遅延したビデオに音声を再同期させる。これをスライド制御と呼ぶ。

特に、スライド制御が音声に及ぼす影響を考慮し、2段階のスライド制御を適用する。すなわち、音声への影響を小さくするために、 $D_n^{(0)}$ を少しずつ変更して徐々にメディア間同期を修復する方法と、音声への影響は大きいけれども、 $D_n^{(0)}$ を一度に大きく変更し、早急にメディア間同期を修復する方法とを採用する。このとき、それぞれの場合のスライド時間を $\theta_1$ 、および $\theta_2$ と定義する。また、いずれのスライド制御を実行するかを決めるために、 $\Delta_n^{(1)}$ に対し2個のしきい値 $T_{h1}$ と $T_{h2}$ を定義する。

具体的には、まず、最初のビデオフレームの $D_1^{(1)}$ と、最初の音声MUの $D_1^{(2)}$ を一致させ、ビデオと音声のメディア間同期をとる。その後、 $\Delta_n^{(1)}$ に応じて、以下のスライド制御を行う。

(a)  $\Delta_n^{(1)} = 0$ の場合：メディア内同期、およびメディア間同期とも維持されているため、そのまま継続する。

(b)  $0 < \Delta_n^{(1)} \leq T_{h1}$ の場合：Bブロック廃棄制御によるメディア内同期の回復を期待し、そのまま継続する。

(c)  $T_{h1} < \Delta_n^{(1)} \leq T_{h2}$ の場合：ビデオと音声の $D_n^{(0)}$ を、 $t_n^{(0)}$ より1秒間に $\theta_1$  [ms]の割合で少しずつ後ろにスライドし、徐々にメディア間同期を修復する。なお、1ビデオMUに対し1回のスライド制御しか実行されず、 $\Delta S_n^{(0)} = \Delta S_m^{(2)} = \theta_1$ となる。

(d)  $T_{h2} < \Delta_n^{(1)}$ の場合：ビデオと音声の $D_n^{(0)}$ を、 $t_n^{(0)}$ より1分間に $\theta_2$  [ms] ( $\geq \theta_1$ )の割合で一度に大きく後ろにスライドし、一度にメディア間同期を修復する。なお、(c)と同様に $\Delta S_n^{(0)} = \Delta S_m^{(2)} = \theta_2$ となる。

このスライド制御により、音声の出力は $\Delta S_m^{(2)}$ だけ停止し、音声のメディア内同期は乱れる。但し、出力ずみの音声情報をスライド時間に相当する量だけバッファに残し、スライド制御時に音声を巻き戻して出力すれば、音声の中断を回避できる。このようにすれば、音声のエコーが聞こえることになるが、スライド時間があまり大きくなければ、不自然さは生じない。

一方、ビデオは $\Delta S_n^{(1)}$ が $\Delta_n^{(1)}$ より大きければ、 $\Delta S_n^{(1)} - \Delta_n^{(1)}$ だけビデオの出力が停止する。しかし、その間に受信したビデオ情報はバッファリングされ、メディア

内同期の乱れを吸収するのに役立つ。

### 3. シミュレーションの仮定

#### 3.1 伝送制御方式の仮定

各メディアの伝送に使用する伝送制御方式に関し、以下の仮定を置く。

##### (1) ビデオ情報の伝送

PHSの32 kbit/s非制限デジタルベアラーサービスにおいて、SR-ARQ (Selective Repeat - Automatic Repeat Request)を採用する[8]。ARQフレームの構造は640 bitを一つのARQフレームとし、ARQ制御情報とユーザデータを常に組み込むピギーバック方式(図4)とする。なお、ARQフレームの同期は通信開始時に確立するものとする。

ユーザデータ領域(584 bit)には、データ長(DL)で指定したバイト数の有意データを設定し、CRCによりARQフレームの誤りを検出する。

ARQ制御情報は、フレーム種別識別子(FD)、送信フレーム番号(TN)、要求フレーム番号(RN)で構成する。ACKやNACKはFDにおいて指定される。また、TNは送信情報フレームの順序番号、RNは受信情報フレームの順序番号を示し、モジュロは64である。

##### (2) 音声情報の伝送

ビデオとは別の通信チャンネルにおいて、32 kbit/s音声サービスを利用して、ADPCM音声情報を無手順で伝送する。但し、メディア間同期のために、音声情報の開始位置を明確にする必要がある。また、スライド制御による受信バッファのオーバフローを回避するために、フロー制御が必要である。そこで、音声情報と識別可能な160 bitのユニークコードを定義する。そして、音声情報を伝送しないときには、TDMAフレームの情報チャンネルTCH (Traffic Channel: 160 bit)単位でこれを送信し、受信側で読み捨てるものとする。

#### 3.2 通信環境の仮定

ランダム誤り環境およびバースト誤り環境におけるシミュレーションを行うにあたり、以下の仮定を置く。

##### A1. ビット誤りは、基地局と移動局間の無線部にお

ける情報チャンネルTCHのみに発生する。ビデオサーバと基地局間の有線部におけるビット誤りは無視する。

A2. ランダム誤り環境におけるビット誤りの発生は、幾何分布に従う。

A3. バースト誤り環境として2状態マルコフ連鎖モデル(ギルバートモデル)を適用する[9]。なお、バースト状態から良好状態への遷移確率 $\gamma$ 、良好状態からバースト状態への遷移確率 $\sigma$ は、フェージング周波数8 Hz(歩行時に相当)における各受信レベルとビット誤り率(BER)、平均バースト長の測定結果から算出した[4]。また、バースト状態と良好状態間の状態遷移の発生は、幾何分布に従う。

A4. ビデオサーバと基地局間の伝搬遅延時間、およびARQフレーミングによる処理遅延時間の合計を40 msと仮定する。また、基地局と移動局間の無線部における電波伝搬遅延は無視する。

A5. ACK/NACK等のARQ制御情報には誤りが発生しないものとし、また再送要求の廃棄は行わない。

#### 3.3 蓄積メディアの仮定

シミュレーションで使用するH.263圧縮ビデオ情報とADPCM音声情報の作成に関し、以下の仮定を置く。

A1. H.263で規定された四つのコーディングオプションのうち、PBフレームモードのみを適用する。

A2. 1ビデオフレームを1ビデオMUとする。またMUサイズ $M_s^{(1)}$ は可変長とし、MU発生レート $M_R^{(1)}$ を15 MU/sの固定値とする。これにより連続したMU間の相関が強くなり、多くのPBフレームの発生を期待できる。

A3. ビデオの符号発生レートを29.2 kbit/sとし、空間的解像度制御により符号発生量を制御して、H.263圧縮ビデオ情報を作成する。すなわち、符号発生レートとMU発生レート $M_R^{(2)}$ により定まるMUサイズの平均値と、ある時点までに符号化されたビデオフレームのMUサイズの平均値とを比較し、その差に応じて量子化特性を制御する。

A4. ビデオの発生時刻 $T_n^{(1)}$ は、符号化時に各ビデオフレームに設定された $TR_n$  (Temporal Reference)と $TR_B$  (TR for B-picture)により、次式で与える。但し、最初のビデオフレームの $TR_1$ を0とする。

$$T_n^{(1)} = TR_n / 30 \quad (I, P \text{ フレーム})$$

$$T_n^{(1)} = (TR_{n+1} - TR_B) / 30 \quad (B \text{ フレーム})$$

A5. 音声の各MUサイズ $M_s^{(2)}$ を3,200 bit、MU発生レート $M_R^{(2)}$ を10 MU/sの固定値とし、各MUの発生時刻 $T_n^{(2)}$ を次式で与える。

$$T_n^{(2)} = (n-1) / M_R^{(2)} \quad (n \geq 1)$$

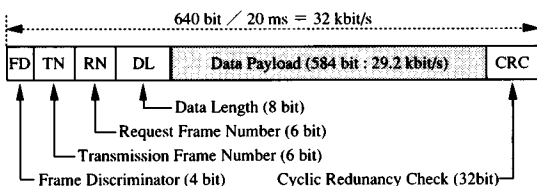


図4 ARQフレームの構成  
Fig. 4 Structure of ARQ frame.

A6. 使用した蓄積メディアの諸元を示す.

- (a) メディアソース 映画「ラビリンス」のラストシーン
- (b) 画像サイズ sub-QCIF (128×96 ピクセル)
- (c) ソース時間 122.333 s
- (d) I, P, PBフレーム数 1,102フレーム
- (e) PBフレーム数 734フレーム
- (f) ビデオ情報量 3,583,656 bit
- (g) Bブロック占有率 13.1% (470,748 bit)
- (h) 平均量子化特性 11.2

ビデオフレーム長の分布を調べると(図5), 2 kbitと4 kbitにピークがある(□でプロット). 2 kbitのピークはPフレーム(○でプロット)のピークであり, ビデオの平均MUサイズである1.947 kbitにはほぼ等しい. また4 kbitのピークはPBフレーム(△でプロット)のピークであり, ビデオの平均MUサイズの倍の値にはほぼ等しい. なお, MUサイズの2乗平均値は2.993 (kbit)<sup>2</sup>である.

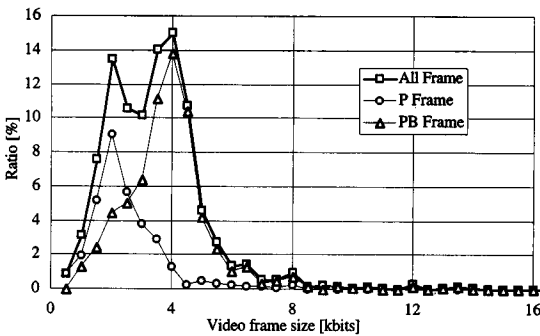


図5 ビデオフレーム長の分布  
Fig. 5 Distribution of video frame size.

3.4 メディア同期制御の仮定

メディア同期制御に関し, 以下の仮定を置く.

A1. スライド制御のしきい値 $T_{h1}$ によるスライド時間 $\theta_1$ は, 100 msとする. この値は, スライドにより巻き戻された音声と違和感の少ないエコーに聞こえ, かつ音声のMUサイズに相当するようを選んだ. またしきい値 $T_{h1}$ は, 処理を簡単にするため, 音声のMUサイズに相当する100 msとする.

A2. スライド制御のしきい値 $T_{h2}$ は2 sとする. これは, ハンドオーバー時に通信チャンネルが約1 s中断することを考慮した値である. またスライド時間 $\theta_2$ は4 sとする. これにより1回のスライドによるビデオの停止時間 $\theta_2 - \Delta_n^{(1)} (\leq \theta_2 - T_{h2})$ は2 s以下になる.

A3. 出力開始遅延時間 $S_0^{(i)}$ は, その待ち時間をユーザが許容できるように5 sとする.

3.5 評価システム

図6に実際のH.263圧縮ビデオ情報とADPCM音声情報を使用する評価システムのソフトウェア構成を示す. ビデオの情報ファイルは, 各ビデオフレームのリファレンス番号, ピクチャ種別, サイズ, Bブロックサイズ等を含む. ワークステーションで通信プロトコルやエラー環境等のシミュレーションを行い, ビデオの情報ファイルをもとにメディア内, およびメディア間同期ずれ時間を求め, 各メディアの出力情報ファイルを生成する. 更に, ADPCM音声情報にエラー環境に応じたビット誤りを付加する. そして各メディアは, これらの出力情報ファイルをもとに出力制御されて再生される. なお, H.263圧縮符号化と情報ファイルの生成, および復号化と再生には, H.263ソフトウェアエンコーダ/デコーダ[10]を一部改造して使用した.

4. 数値結果と性能評価

ランダム誤り環境とバースト誤り環境の両方の通信環境におけるメディア同期の性能を評価する.

4.1 ランダム誤り環境

ランダム誤り環境における評価結果を以下に示す. まず, SR-ARQによるビデオ通信チャンネルのスループット特性は図7のとおりである. なおスループットは, 単位時間当りに送信が完了した情報量をPHSの伝送容量(32 kbit/s)で正規化したものとする.

(1) Bブロック廃棄制御

図8にビデオ情報廃棄率とBERの関係, 図9にビデオの平均MU出力レートとBERの関係, および図10にビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ の平均値とBERの

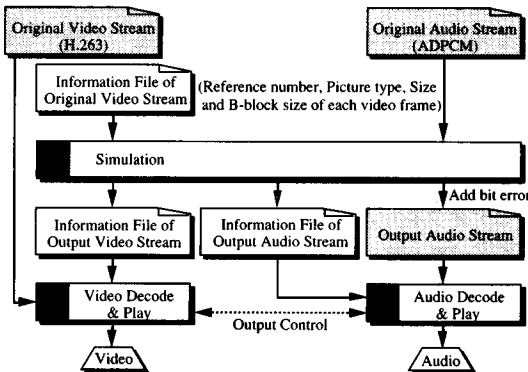


図6 評価システムの構成  
Fig. 6 Block diagram of the evaluation system.

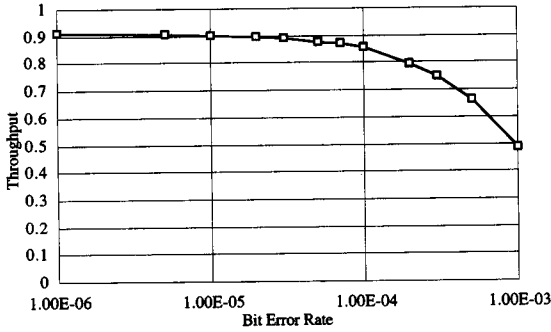


図7 スループット (ランダム誤り環境)  
Fig. 7 Throughput in random error environment.

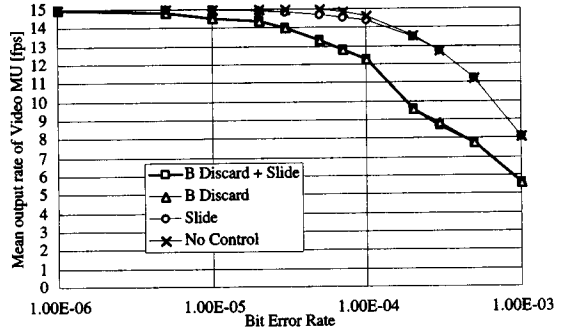


図9 ビデオの平均MU出力レート (ランダム誤り環境)  
Fig. 9 Mean output rate of video MU in random error environment.

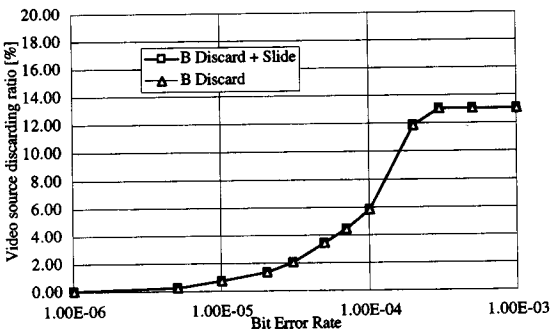


図8 ビデオ情報廃棄率  
Fig. 8 Video source discarding ratio.

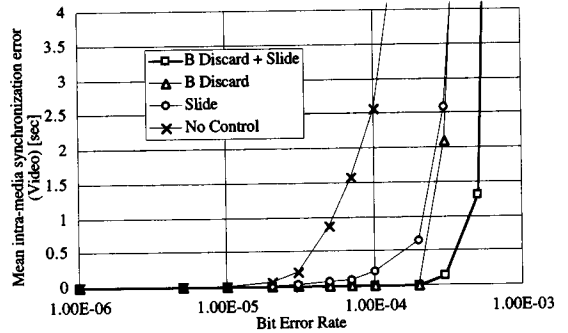


図10 ビデオの平均メディア内同期ずれ時間(ランダム誤り環境)  
Fig. 10 Mean of  $\Delta_n^{(1)}$  in random error environment.

関係を示す。ビデオ情報廃棄率は、すべてのビデオ情報のうち、Bブロック廃棄制御により送信側で廃棄した情報量の割合と定義する。平均MU出力レートは、受信側で単位時間当りに出力したMU数の平均値とし、ビデオの場合fps (frame per second) で定義する。また各図において、□はBブロック廃棄制御とスライド制御を適用した場合 (B Discard + Slide)、△はBブロック廃棄制御のみ (B Discard)、○はスライド制御のみ (Slide) を適用した場合、×はいずれも適用しない場合 (No Control) を示す。但し、No Controlの場合でも、目標出力時刻 $t_n^{(0)}$ と到着時刻 $A_n^{(0)}$ の関係に応じたメディア内同期制御を行うものとする。従って、いずれの場合においても $\Delta_n^{(1)} \geq 0$ となるため、メディア内同期の性能は、平均2乗誤差でなく平均値で比較を行うこととする。

図8によれば、ビデオ情報廃棄率は $BER = 3.0 \times 10^{-4}$ 以上で一定になる。これは、この領域ではすべてのBブロックが廃棄されていることを意味する。すなわち、Bブロック廃棄制御の有効適用範囲は、 $BER$ が $3.0 \times 10^{-4}$ 以下と言える。

一方、ビデオの平均MU出力レート(図9)は、Bブロック廃棄制御の適用により低下する。例えば、 $BER = 2.0 \times 10^{-4}$ において、Bブロック廃棄制御を適用しなければ13.5 fps (○, ×でプロット) であるが、Bブロック廃棄制御を適用すれば9.5 fps (□, △でプロット) と30%削減される。しかし、ほとんどのBブロックを廃棄しても9.5 fpsを確保できるため、主観評価においても比較的なめらかなビデオ表示に感じられた。

更に、ビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ の平均値(図10)は、Bブロック廃棄制御の適用により小さく制御できる。例えば、No Control (×でプロット) の場合、 $BER$ が $1.0 \times 10^{-4}$ において $\Delta_n^{(1)}$ は平均2.6 sである。しかし、Bブロック廃棄制御を適用(□, △でプロット)すれば、 $BER$ が $2.0 \times 10^{-4}$ 以下において $\Delta_n^{(1)}$ を0にできる。従って、Bブロック廃棄制御は、ビデオのメディア内同期を維持するのに、非常に効果的である。一方、スライド制御を適用すれば(□, ○でプロット)、メディア内同期は更に改善される。これは、目標出力時刻を後ろにスライドすることで、 $\Delta_n^{(1)}$ が小さくなるためである。

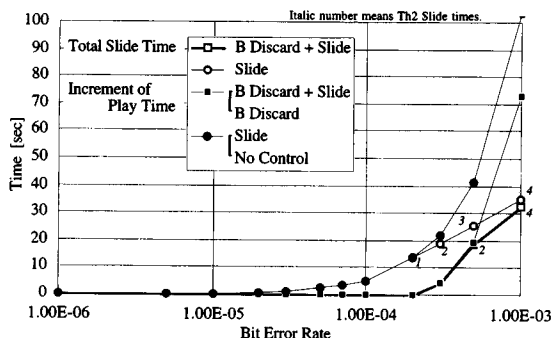


図11 総スライド時間と再生増加時間  
Fig. 11 Total slide time & increment of play time.

(2) スライド制御

図11に目標出力時刻の総スライド時間 $S_n$ およびビデオの再生増加時間とBERの関係を示す。ビデオの再生増加時間は、ビデオの再生時間とビデオソース時間の差と定義する。なお、図中のシンボル(□,○)に付した数字は、ビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ がしきい値 $T_{h2}$ (=2 s)を超えたために発生したスライド制御の回数を示す。

まず、総スライド時間 $S_n$ について考える。スライド制御だけを適用(○でプロット)した場合、BERが $1.0 \times 10^{-5}$ 以上で目標出力時刻のスライドが発生する。これは、スライド制御だけでは、 $\Delta_n^{(1)}$ が頻繁にしきい値 $T_{h1}$ (=100 ms)を超えてしまうためである。一方、Bブロック廃棄制御とスライド制御を適用(□でプロット)した場合、BERが $2.0 \times 10^{-4}$ 以下ではBブロック廃棄制御により $\Delta_n^{(1)}$ が0となるため、スライド制御も発生しない。

つぎに、総スライド時間 $S_n$ とビデオの再生増加時間とを比較して考える。再生増加時間(■,●でプロット)が総スライド時間 $S_n$ (□,○でプロット)を超えることは、スライド制御の能力を超えたメディア内同期ずれの発生を意味する。従って、スライド制御の有効適用範囲は、Bブロック廃棄制御とスライド制御を適用した場合(□,■でプロット)、BERが $2.0 \times 10^{-4} \sim 5.0 \times 10^{-4}$ と言える。

更に、どのスライド条件が適用されたかについて調べる。例えば、Bブロック廃棄制御とスライド制御を適用した場合(□でプロット)、BER= $3.0 \times 10^{-4}$ における総スライド時間4.4sの内訳は、すべてしきい値 $T_{h1}$ =100 ms、スライド時間 $\theta_1$ =100 msのスライドである。すなわち、1秒間に100 msの割合で音声を巻き戻すことが44回発生している。この音声を主観評価すると、エコーがかかったように聞こえるものの大きな違和感はない。従って、 $\theta_1$ =100 msは十分適切な値と言える。

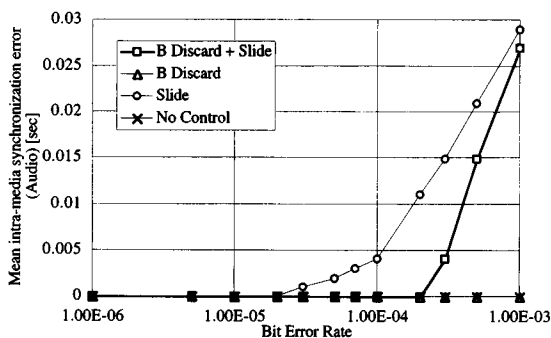


図12 音声の平均メディア内同期ずれ時間(ランダム誤り環境)  
Fig. 12 Mean of  $\Delta_n^{(2)}$  in random error environment.

また、BER= $5.0 \times 10^{-4}$ における総スライド時間18.9 sのうち、しきい値 $T_{h2}$ =2 s、スライド時間 $\theta_2$ =4 sのスライドが2回発生している。すなわち、4秒間だけ音声を巻き戻すことが2回発生している。この音声を主観評価すると、音声が予期せず巻き戻されるために違和感を感じるものの、音声の中断がなく良好である。

ここで、スライド制御が音声に及ぼす影響を、音声のメディア内同期ずれ時間 $\Delta_n^{(2)}$ を用いて調べる。図12に $\Delta_n^{(2)}$ の平均値とBERの関係を示す。これによれば、スライド制御を適用した場合(□,○でプロット)、音声のメディア内同期ずれが発生する。例えば、Bブロック廃棄制御とスライド制御を適用した場合(□でプロット)、BER= $5.0 \times 10^{-4}$ における $\Delta_n^{(2)}$ は平均15 msである。しかしこの値は、同じ通信環境におけるビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ (平均1.32 s)と比較して十分小さな値である。従って、音声とビデオのメディア間同期ずれ $\Delta_n^{(1-2)}$ の平均値は、 $\Delta_n^{(1)}$ の平均値と極めて近くなる。図13に $\Delta_n^{(1-2)}$ の平均値とBERの関係を示す。

文献[11]によれば、 $\Delta_n^{(1-2)}$ が80 ms以内であればメディア間同期は最適な状態であり、160 ms以上であれば同期はずれの状態である。この値を用いてメディア間同期を評価すると、Bブロック廃棄制御とスライド制御を適用すれば(□でプロット)、BERが $3.0 \times 10^{-4}$ 以下の通信環境で、 $\Delta_n^{(1-2)}$ が平均139 ms以下のほぼ良好なメディア間同期を維持できる。すなわち、スライド制御は音声のメディア内同期を乱すものの、ビデオと音声のメディア間同期を維持するのに非常に有効であると言える。なお、このときの $\Delta_n^{(1)}$ は平均140 ms以下、 $\Delta_n^{(2)}$ は平均4 ms以下である。

(3) 出力開始遅延制御

スライド制御と出力開始遅延制御の実行時にバッ



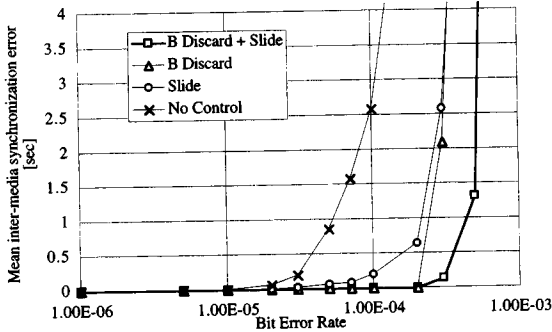


図13 平均メディア間同期ずれ時間 (ランダム誤り環境)  
Fig. 13 Mean of  $\Delta_n^{(1)}$  in random error environment.

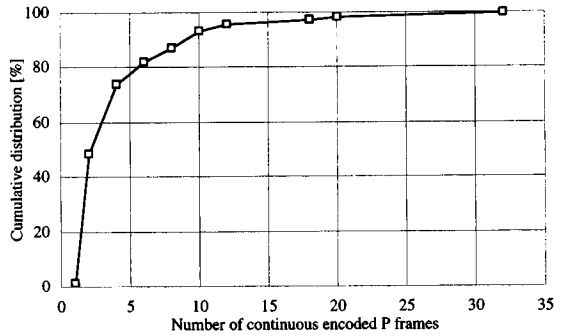


図15 Pフレーム連続発生回数の累積分布  
Fig. 15 Cumulative distribution of continuous encoded P frames.

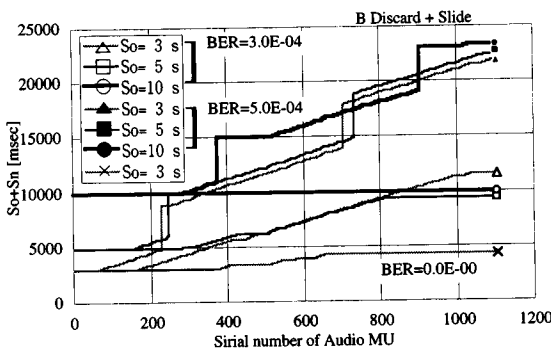


図14 出力開始遅延時間 $S_0$ +総スライド時間 $S_n$ の経過  
Fig. 14 Progress of  $S_0 + S_n$ .

ファリングされたビデオ情報は、ビデオのメディア内同期の乱れを吸収するのに役立つ。従って、出力開始遅延時間 $S_0$ と総スライド時間 $S_n$ の和( $S_0 + S_n$ )は、ビデオのメディア内同期の状態を示す評価尺度として利用できる。なお $S_0 + S_n$ は、[6]における $\hat{\Delta}_{max}$ に相当する。

図14にBブロック廃棄制御とスライド制御を適用した場合における $S_0 + S_n$ の時間的変化を示す。例えば $BER = 5.0 \times 10^{-4}$ 、 $S_0 = 10$  s (●でプロット)において、 $S_0 + S_n$ が変化しない場合は、ビデオ情報のバッファリングとBブロック廃棄制御により、メディア内同期の乱れを吸収している状態である。また、 $S_0 + S_n$ のなだらかな増加は、しきい値 $T_{h1}$ (=100 ms)を超えるメディア内同期ずれが発生し、スライド時間 $\theta_1$ (=100 ms)のスライドが発生している状態である。更に $S_0 + S_n$ の急激な増加は、メディア内同期ずれ時間がしきい値 $T_{h2}$ (=2 s)を超え、スライド時間 $\theta_2$ (=4 s)のスライドが発生した状態である。

ここで、異なる $S_0$ (=3, 5, 10 s)における $S_0 + S_n$ の時間

的变化を比較してみる。例えば、しきい値 $T_{h1}$ のスライドしか発生しない $BER = 3.0 \times 10^{-4}$ ( $\Delta$ ,  $\square$ ,  $\circ$ でプロット)において、更に、しきい値 $T_{h2}$ のスライドも発生する $BER = 5.0 \times 10^{-4}$ ( $\blacktriangle$ ,  $\blacksquare$ ,  $\bullet$ でプロット)においても、 $S_0$ の値にかかわらず、 $S_0 + S_n$ はほぼ同じ値に収束する。すなわち、 $S_0$ の値を小さく設定しても、 $S_0$ の値を大きくした場合と同様の効果のスライド制御によって期待できる。従って、スライド制御の適用を許容するならば、 $S_0$ は大きめに設定しておけば十分である。なお、[6]においても同様の結論を得ている。

一方、 $BER = 0.0$ ( $\times$ でプロット)であってもスライド制御が発生している。これは、ビデオのMUサイズの2乗平均値が $2.993$  (kbit)<sup>2</sup>と大きく、またMUサイズの大きなMUが連続して発生したために、しきい値 $T_{h1}$ を超えるビデオのメディア内同期ずれが発生したものと考える。従って、 $BER = 0.0$ においてスライド制御が実行されないようにするには、 $S_0$ は4.6 s以上に設定する必要がある。

更に、Pフレームの連続発生回数の累積分布を(図15)を見てみると、連続して6回(0.4 s期間)以上のPフレームが発生する割合は約20%に達する。特に、最大32回(2.133 s期間)連続してPフレームが発生する場合もある。これは、変化の大きな動きのあるビデオでは、二つのPフレームとして符号化するほうが符号量を小さくでき、PBフレームが生成されないためである。従って、Bブロック廃棄制御を適用しても、最大2.133 sの間Bブロックを廃棄できない場合があり、これを補うのに十分な $S_0$ を設定する必要がある。

このようにメディア同期の観点からは、出力開始遅延時間 $S_0$ を大きくしたほうが望ましい。しかし、 $S_0$ は通信開始時の待ち時間となるため、ユーザにとって小

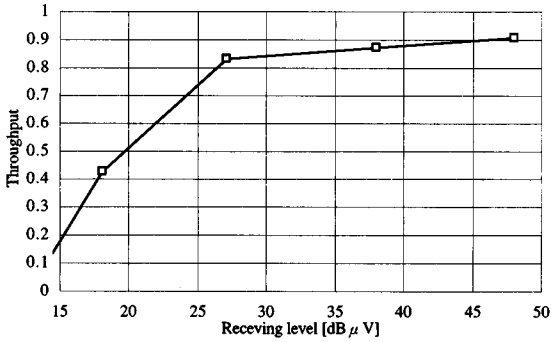


図16 スループット (バースト誤り環境)  
Fig. 16 Throughput in burst error environment.

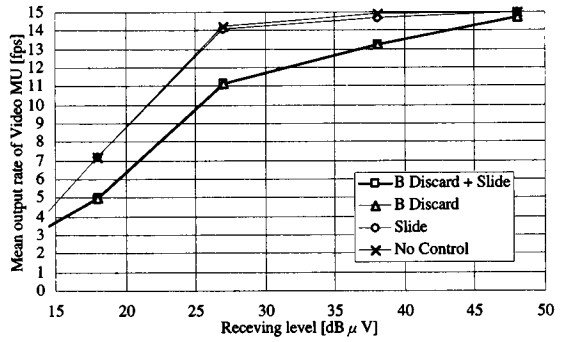


図18 ビデオの平均MU出力レート (バースト誤り環境)  
Fig. 18 Mean output rate of video MU in burst error environment.

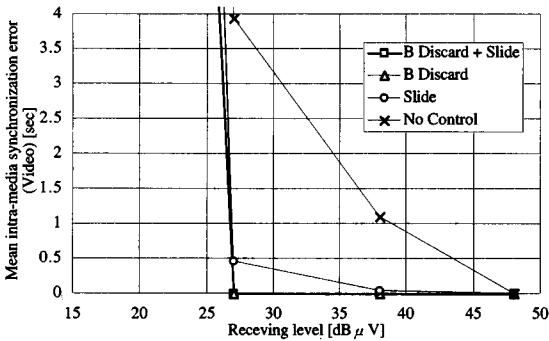


図17 ビデオの平均メディア内同期ずれ時間 (バースト誤り環境)  
Fig. 17 Mean of  $\Delta_n^{(1)}$  in burst error environment.

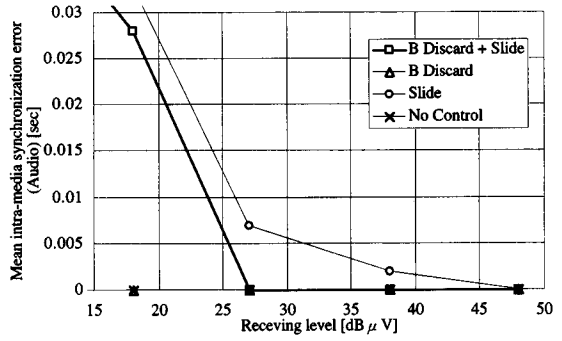


図19 音声の平均メディア内同期ずれ時間 (バースト誤り環境)  
Fig. 19 Mean of  $\Delta_n^{(2)}$  in burst error environment.

さいほうが望ましい。従って $S_0$ は、5~10 sが適当と考える。

#### 4.2 バースト誤り環境

つぎに、バースト誤り環境における評価結果を示す。まず、SR-ARQによるビデオ通信チャネルのスループット特性は図16のとおりである。

図17にビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ の平均値、図18にビデオの平均MU出力レート、図19に音声のメディア内同期ずれ時間 $\Delta_n^{(2)}$ の平均値、図20にビデオと音声のメディア間同期ずれ時間 $\Delta_n^{(1-2)}$ の平均値を、受信レベルとの関係で示す。

ビデオのメディア内同期ずれ(図17)は、Bブロック廃棄制御の適用により、バースト誤り環境においても小さく制御できる。例えば、No Control( $\times$ でプロット)の場合、受信レベルが27 dB $\mu$ Vにおいて $\Delta_n^{(1)}$ は平均3.9 sである。しかし、Bブロック廃棄制御を適用( $\square, \triangle$ でプロット)すれば、受信レベルが27 dB $\mu$ V以上において $\Delta_n^{(1)}$ を0にできる。またこのとき、Bブロックを廃棄し

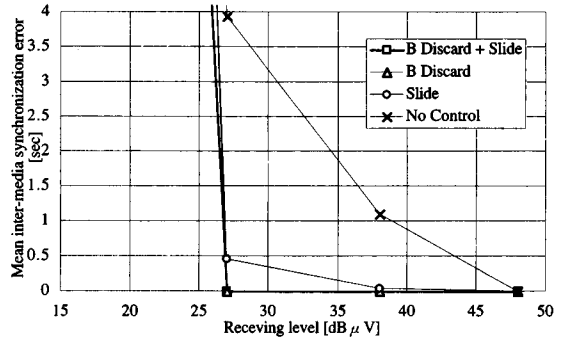


図20 平均メディア間同期ずれ時間 (バースト誤り環境)  
Fig. 20 Mean of  $\Delta_n^{(1-2)}$  in burst error environment.

ても11.2 fps以上の平均MU出力レート(図18:  $\square, \triangle$ でプロット)を確保できるため、主観評価においてもなめらかなビデオ表示に感じられた。

一方、音声のメディア内同期ずれは、Bブロック廃棄制御とスライド制御を適用(図19:  $\square$ でプロット)すれば、受信レベルが27 dB $\mu$ Vにおいてビデオのメディア

内同期ずれが発生しないため、スライド制御が実行されず、音声のメディア内同期ずれも発生しない。従って、このときメディア間同期ずれ(図20: □, △でプロット)も発生せず、良好なメディア同期を維持できる。すなわち、バースト誤り環境においても、特に、Bブロック廃棄制御を適用することがメディア同期の維持に非常に効果的であると言える。

## 5. む す び

本研究では、PHSの二つの通信チャネルを使用し、蓄積されたH.263圧縮ビデオ情報をARQで、蓄積されたADPCM音声情報を無手順で伝送しながら再生することを考えた。ビデオや音声のメディア内同期や両者のメディア間同期を維持するために、Bブロック廃棄制御、出力開始遅延制御、スライド制御を提案した。そしてランダム誤りとバースト誤りの各環境下において、実際のビデオ/音声情報を使用したシミュレーションを行い、メディア同期の性能評価、および主観評価を行った。その結果、次のような結論を得た。

Bブロック廃棄制御は、平均MU出力レートが低下するものの、ビデオのメディア内同期の維持に極めて効果的である。また、スライド制御と出力開始遅延制御によるビデオ情報のバッファリングは、ビデオの一時的なメディア内同期の乱れを吸収するため非常に有効である。なお、出力開始遅延時間 $S_0$ の大きさは、スライド制御を適用すれば、柔軟に設定できる。更に、スライド制御は、音声のメディア内同期を故意に乱すものの、音声とビデオのメディア間同期の維持に非常に効果的である。

また本提案方式を適用すれば、BERが $3.0 \times 10^{-4}$ 以下のランダム誤り環境において、ビデオのメディア内同期ずれ時間 $\Delta_n^{(1)}$ が平均140 ms以下、音声のメディア内同期ずれ時間 $\Delta_n^{(2)}$ が平均4 ms以下、ビデオと音声のメディア間同期ずれ時間 $\Delta_n^{(1-2)}$ が平均139 ms以下になる。また、このとき、ほぼ良好にメディア同期が維持されていることを主観評価によっても確認した。更に、バースト誤り環境においても、本提案方式がメディア同期の維持に効果的であることを確認した。

## 文 献

- [1] “パーソナル移動通信小特集,”信学誌, vol.78, no.2, pp.122-178, Feb. 1995.
- [2] 社団法人電波産業会, 第二世代コードレス電話システム標準規格 第二版(改訂-2) RCR STD-28, June 1996.
- [3] ITU-T Recommendation H.263, Video coding for low bit rate communication, March 1996.
- [4] 加藤正美, 臼井憲義, 田坂修二, “PHSにおけるビデオ伝送

方式の性能評価,” 信学論(B-II), vol. J79-B-II, no.10, pp.646-656, Oct. 1996.

- [5] 的場直人, 近藤 靖, 山階正樹, “ARQを用いた無線動画像通信の伝送遅延の改善法の検討,” 信学技報, RCS96-80, pp.95-101, Aug. 1996.
- [6] S. Tasaka, Y. Ishibashi, and H. Imura, “Stored media synchronization in wireless LANs,” Conf. Rec. IEEE GLOBECOM '96, pp.1904-1910, Nov. 1996.
- [7] マルチメディア通信研究会編, 最新MPEG教科書, アスキー出版局, Aug. 1994.
- [8] PHSインターネット・アクセス・フォーラム, PIAFS概要, June 1996.
- [9] R.W. Lucky, J. Salz, and E.J. Weldon, Principles of data communication, McGraw-Hill, Inc., 1968.
- [10] Source program code of “TMN H.263 Encoder ver.1.6c & Decoder ver.1.6c,” Telenor R&D, March 1996.
- [11] R. Steinmetz, “Human perception of jitter and media synchronization,” IEEE J. Sel. Areas Commun., vol. 14, no. 4, pp. 61-72, Jan. 1996.

(平成8年12月24日受付, 9年3月5日再受付)



加藤 正美 (正員)

昭59名古屋大・理・物理卒。同年三洋電機(株)入社。現在、同社ハイパーメディア研究所主任研究員。移动通信システム、無線・有線統合型マルチメディアネットワークの研究に従事。



臼井 憲義 (正員)

昭46岐阜大・工・電気卒。同年三洋電機(株)入社。現在、同社情報通信事業本部情報通信開発研究所 部長。PBXをはじめ、移动通信システム、特にPHS公衆基地局の研究・開発に従事。



田坂 修二 (正員)

昭46名工大・工・電気卒。昭51東大大学院博士課程了。同年名工大情報勤務。現在、同大・電気情報教授。昭59~60 UCLA客員研究員。マルチメディア通信、無線LANの研究に従事。工博。IEEE, ACM, 情報処理学会各会員。