PAPER *Special Issue on the Latest Development of Telecommunication Research*

Performance Evaluation of Media Synchronization in PHS with the H.223 Annex Multiplexing Protocol*

Masami KATO†, **Yoshihito KAWAI**††**, *and* **Shuji TASAKA**††, *Members*

SUMMARY This paper studies the application of a media synchronization mechanism to the interleaved transmission of video and audio specified by the H.223 Annex in PHS. The media synchronization problem due to network delay jitters in the interleaved transmission has not been discussed in either the Annex or any related standards. The slide control scheme, which has been proposed by the authors, is applied to live media. We also propose a QOS control scheme to control both quality of the media synchronization and that of the transmission delay. Through simulation we confirm the effectiveness of the slide control scheme and the QOS control scheme in the interleaved transmission.

key words: *PHS, multiplexing protocol, media synchronization, transmission delay, QOS control, live media*

1. Introduction

PHS (Personal Handy phone System) [1] is one of the promising wireless infrastructures for mobile multimedia communications. It offers both speech and 32 kbit/s bearer services. Moreover, a 64 kbit/s bearer service will be started in the near future.

The application of the H.223 Annex multiplexing protocol [2]-[4] to a PHS transmission channel is an approach to mobile visual communications. The Annex specifies a multiplexing protocol for low bit rate multimedia communication over error-prone channels. It allows the transfer of any combination of audio, video and data streams over a single communication link. Consequently, an H.263 coded video stream [5], which is suitable for PHS because of its coding rate, can be transmitted with ARQ (Automatic Repeat reQuest) control or FEC (Forward Error Correction) over a logical channel, while a G.723.1 coded audio stream [6] can be sent without any control in another logical channel of the same link. The paper treats this system.

In this transmission scheme, network delay jitters due to retransmissions of video packets disturb the temporal constraints of these streams. In order to obtain the excellent quality of the media at the destination, a media synchronization mechanism should be applied. Media synchronization is the preservation of the media structure, in particular, the temporal relationship between media units (MUs) such as video frames in a single stream (i.e., *intra-stream synchronization*) and also that among plural media streams (i.e., *inter-stream synchronization*) [7]. It is typified by lip-synch, which adjusts the output timing between spoken voice and the movement of the speaker's lips. It should be noted that the media synchronization problem due to network delay jitters has not been discussed in either the Annex or any related standards.

For live media synchronization in PHS, the authors have already proposed the slide control scheme as a media synchronization mechanism [8]; the idea is based on the virtual-time rendering (VTR) algorithm [9], [10]. It modifies the target output time backward or forward. The target output time is a calculated time at the destination to preserve the temporal constraints. In addition, the maximum allowable modification time κ restricts the amount of the modification, so that the transmission delay is not too large.

In [8], we have confirmed the effectiveness of the slide control scheme in another transmission scheme, where audio and video streams are transmitted over separate PHS channels so that speech and 32 kbit/s bearer services, both of which are already offered, can be used. However, it is not clear how effective the slide control is in the interleaved transmission with the Annex.

Also, we have confirmed in [8] that the slide control improves the quality of the media synchronization at the expense of the transmission delay. That is, there is a trade-off between the quality of the inter-stream synchronization and that of the transmission delay. Moreover, which quality should be guaranteed depends on the kind of a given application. For example, the lip-synch should be preserved in the case of lessons in English conversation, whereas the transmission delay should be small in a heated discussion. Therefore, some QOS (Quality of Service) control scheme is necessary so as to guarantee the quality of the media according to the requirements.

In this paper, we study the application of the slide control scheme to the interleaved transmission specified by the H.223 Annex in PHS. Especially, we consider the case where the 64 kbit/s bearer service is used. Moreover, we propose a QOS control scheme in which the maximum allowable modification time κ is utilized to control not only the quality of the transmission delay but also that of the media synchronization. Through simulation we confirm the effectiveness of the slide control scheme and the QOS control scheme.

Manuscript received April 6, 1998.
Manuscript revised June 26, 1998.

† The author is with SANYO Electric Co., Ltd., Gifu-ken, 503-0195 Japan.

†† The authors are with the Department of Electrical and Computer Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan.

* This paper was presented at ICUPC'98, Florence, Italy, October 1998. An early version of this paper was presented at TJCOM'98, Hsinchu, Taiwan.

** Presently with Matsushita Research Institute Tokyo, Inc.

The rest of the paper is organized as follows. Section 2 specifies the slide control scheme and the QOS control scheme adopted in this paper. Section 3 describes simulation methodology, and Sect. 4 discusses numerical results.

2. Media Synchronization Mechanisms

2.1 H.223 Annex Multiplexing Protocol

The Annex specifies a multiplexing protocol for low bit rate multimedia communication over error-prone channels. The Annex A, B and C are different from one another in the robustness and the complexity. Especially, the Annex C, which is adopted in this paper, deals with a highly error-prone environment. Figure 1 shows the protocol stack for the Annex C. It describes two different layers; an *Adaptation Layer* (AL) and a *Multiplex layer* (MUX).

The AL adapts AL-SDUs (Service Data Units) given by the AL user to AL-PDUs (Protocol Data Units) by adding some octets for error controls. Three kinds of ALs are defined: *AL1M*, *AL2M* , and *AL3M*, which are designed for the transfer of digital data, audio, and video streams, respectively, in mobile applications. AL1M and AL3M support the use of error detection, FEC based on RCPC (Rate Compatible Punctured Convolutional codes) algorithm and ARQ, while AL2M does not provide any error control. An AL-PDU is conveyed to the MUX layer as a MUX-SDU.

The MUX layer interleaves some kinds of MUX-SDUs into a single MUX-PDU. The MUX-PDU consists of a synchronization flag, a header and some information slots, which are assigned to different kinds of MUX-SDUs. That is, digital data, audio, and video streams are transmitted over individual logical channels.

2.2 Slide Control Scheme

For live media synchronization in PHS with the Annex, we adopt the same mechanism as that of the slide control which has been proposed in [8]. In this paper, however, we change the definition of streams so as to consider an interleaved transmission. That is, we define a stream of MUX-PDUs as a *composite stream*. Also, each information stream given by the AL user is defined as a *substream* (say substream *j*), which can be either the video substream (*j*=1) or the audio substream (*j*=2).

We assume that each substream is composed of MUs; an MU is the unit of data for the output of the media. The *n*-th MU in substream *j* has a *timestamp* $T_n^{(j)}$ in order to preserve the temporal constraints (see Fig. 2) . The *Temporal Reference* (*TR*) in an H.263 video substream can be utilized as timestamps. An audio MU does not need a timestamp because of its constant MU rate and size.

Note that the slide control in this paper is the *distinctive* scheme [11], which distinguishes between audio and video MUs in the composite stream and executes the inter-stream synchronization control as well as the intra-stream synchronization control.

Let $A_n^{(j)}$ and $\tau_n^{(j)}$ denote the arrival time and the output waiting time, respectively, of the *n*-th MU in substream *j*. Then, the target output time $t_n^{(j)}$ and the *output time* $D_n^{(j)}$ are calculated as follows:

$$\begin{array}{lll} t_1^{(j)}=A_1^{(j)}+\tau_1^{(j)}, & x_1^{(j)}=t_1^{(j)} & (n=1) \\ t_n^{(j)}=x_n^{(j)}+S_{n-1}^{(j)}, & x_n^{(j)}=x_{n-1}^{(j)}+\sigma_{n-1,n}^{(j)} & (n \geqq 2) \\ D_n^{(j)}=t_n^{(j)}+\Delta S_n^{(j)} & (n \geqq 1) & \text{if } A_n^{(j)} \leqq t_n^{(j)}+\Delta S_n^{(j)} \\ D_n^{(j)}=A_n^{(j)} & (n \geqq 1) & \text{if } t_n^{(j)}+\Delta S_n^{(j)} < A_n^{(j)} \\ \sigma_{n-1,n}^{(j)}=T_n^{(j)}-T_{n-1}^{(j)}, & S_n^{(j)}=\Sigma_{k=1}^{n}\Delta S_k^{(j)}, & \end{array}$$

where the *slide time* $\Delta S_n^{(j)}$ is defined as the interval between the original target output time $t_n^{(j)}$ and the modified one, and $x_n^{(j)}$ denotes the ideal target output time, which we would have if there were no network delay jitter. Also, $S_n^{(j)}$ is the *total slide time*.

The slide control has two kinds of modification of the target output time; the *backward control* and the *forward control* (see Fig. 3). The former delays the target output time so as to preserve the inter-stream synchronization between the delayed video MU and the audio MU; this leads to increase in the transmission delay. The latter advances the target output time so that the transmission delay due to the backward control can be reduced, even though the inter-stream synchronization may be disturbed. Thus, we see that there is a trade-off between the quality of the inter-stream synchro-

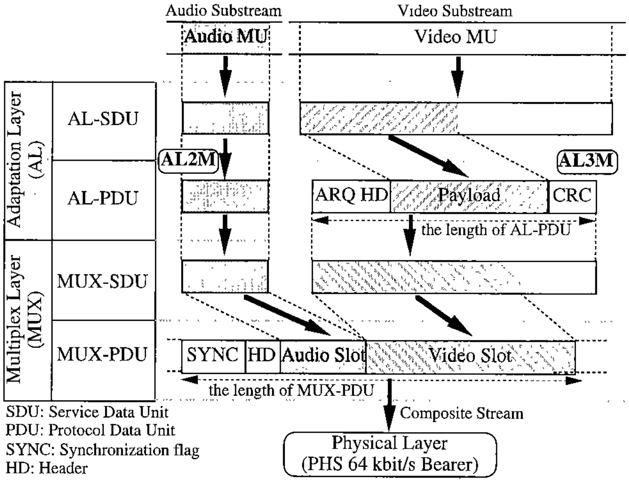


Fig. 1 Protocol stack for H.223 Annex C.

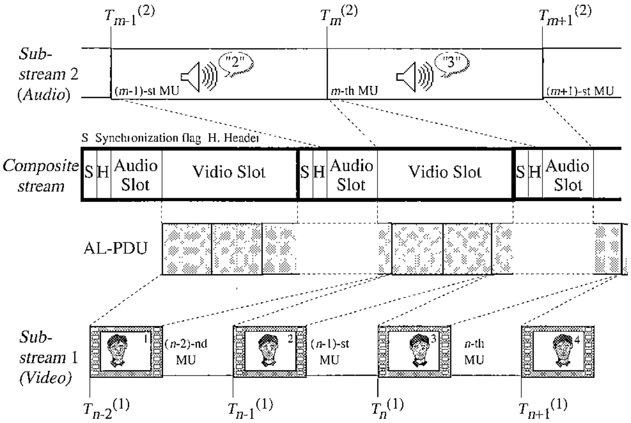


Fig. 2 Temporal relationship between video and audio MUs.

nization and that of the transmission delay. In order to control this relation, we introduce the maximum allowable modification time κ, which restricts the amount of the modification due to the backward control, so that the transmission delay is not too large.

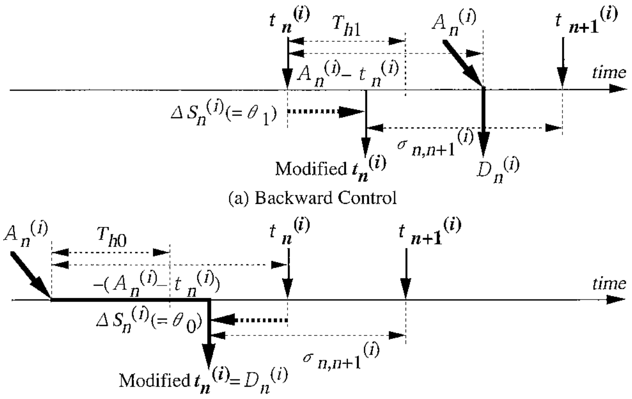
In this paper, we adopt the *gradual recovery*, not the *fast recovery* [12]†; the gradual recovery scheme modifies the target output time step by step so that the disturbance of an audio substream is limited to a small one. It is specified by both modification size of the output time θ_i and a minimum allowable interval between two successive control ω_i, where i indicates the kind of the modification: i=1 for the backward control and i=0 for the forward control. Note that $\theta_1 > 0$, while $\theta_0 < 0$.

Table 1 describes the algorithm that determines which kind of modification can be executed. The criterion of the slide control is $A_n^{(1)} - t_n^{(1)}$ ††. On receiving the n-th video MU, the destination calculates $A_n^{(1)} - t_n^{(1)}$. If the difference exceeds a threshold value T_{h1}, which is a positive constant, and if the total slide time after the modification (i.e., $S_{n-1}^{(1)} + \Delta S_n^{(1)}$) is not larger than the value of κ, the backward control is carried out (i.e., $\Delta S_n^{(1)} = \Delta S_m^{(2)} = \theta_1$, where m is the sequence number of audio MU which corresponds to the n-th video MU). Also, if $A_n^{(1)} - t_n^{(1)}$ is less than or equal to another threshold value $-T_{h0}$ ($T_{h0} > 0$), the forward control can be executed. (i.e., $\Delta S_n^{(1)} = \Delta S_m^{(2)} = \theta_0$). These modifications can be made only after the output of the first video MU.

2.3 QOS Control Scheme

In this paper, we propose a QOS control scheme in which the maximum allowable modification time κ is used to control both quality of the media synchronization and that of the transmission delay. For example, if we choose a small value of κ, we can expect that the transmission delay is small, though the inter-stream synchronization may be disturbed. On the other hand, if we select a large value of κ, we can anticipate that the inter-stream synchronization quality is improved, while the transmission delay may be large.

However, it is critical who selects the value of κ. Of course, the users can directly set the value of κ according to their requirements by a simple method such as pushing a button or clicking a control bar; the user is responsible for choosing the value. However, it is troublesome. Therefore, it is desirable that the value of κ will be chosen adaptively by means of discriminating the contents of the communication, while it is not easy. The application of the adaptive QOS control is one of our future studies.



(a) Backward Control

(b) Forward Control

Fig. 3 Backward control and forward control.

Table 1 Patterns of the slide control.

pattern	description
(a)	if $T_{h1} \leqq A_n^{(1)} - t_n^{(1)}$, and $S_{n-1}^{(1)} + \Delta S_n^{(1)} \leqq \kappa$ then $\Delta S_n^{(j)} = \theta_1$ (Backward).
(b)	if $A_n^{(1)} - t_n^{(1)} \leqq -T_{h0}$, then $\Delta S_n^{(j)} = \theta_0$ (Forward).
(c)	otherwise $\Delta S_n^{(j)} = 0$.

3. Simulation Methodology

3.1 Assumptions for Simulation

We suppose that a PHS mobile terminal receives a composite stream, which is transmitted from an ISDN terminal via a PHS base station over a single communication link (see Fig. 4†††). An H.263 video substream and a G.723.1 audio substream are multiplexed into a composite stream. We make the following assumptions for simulation:

A1: The 64 kbit/s bearer service is used.

A2: In the AL3M, the AL-PDU for the video substream is generated with an ARQ header of 10 octets, a CRC of 4 octets, and a data payload. We assume that the size of the data payload is constant; it is set to 90 octets, which was determined by experiments as explained later. Thus, the length

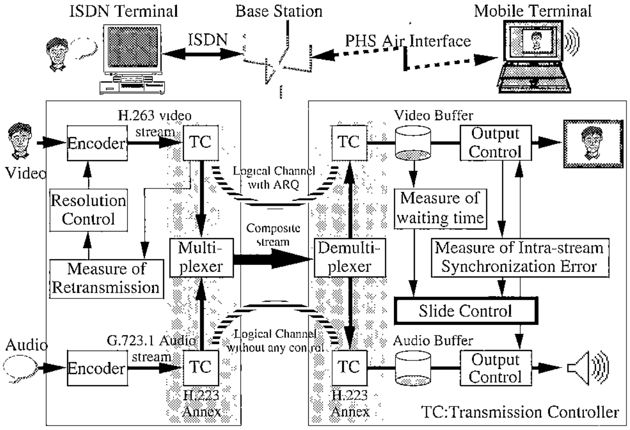


Fig. 4 Block diagram of the live media transmission system .

† It should be noted that these recovery schemes are concerned with the degree of the modification in a single control policy, while *graceful recovery* and *quick recovery* proposed in [10] are alternative control policies on the modification.

†† The criterion $A_n^{(1)} - t_n^{(1)}$ is equivalent to the combination of $\Delta_n^{(1)}$ before the modification (i.e., $A_n^{(1)} - t_n^{(1)}$) and $\tau_n^{(1)} (= t_n^{(1)} - A_n^{(1)})$ used in [8].

††† The maximum size of video and audio buffer in a mobile terminal is given by the product of the maximum allowable modification time κ and the transmission rate.

of AL-PDU is 104 octets except for in Figs.15 and 16.

A3: The function of FEC based on RCPC is not utilized; that is, we set the code rate of RCPC to 1.0. Also, we select ARQI control scheme, which uses the same code rate of RCPC in the case of the retransmission.

A4: The MUX layer creates the MUX-PDU with a synchronization flag of 4 octets, a header of 4 octets, an audio slot and a video slot. We assume that the length of MUX-PDU is constant; it is set to 400 octets except for in Figs. 13 and 14 so that it takes 50 ms to transmit an MUX-PDU with the 64 kbit/s bearer service. Then, the size of the audio slot is 40 octets because of the coding rate (6.3 kbit/s) and the size of the video slot is 352 octets. In Figs. 13 and 14, we use different values from these.

A5: In the AL2M for the transfer of the audio substream, an audio MU is treated as an AL-SDU, which is equal to an AL-PDU. Then, the size of audio MU is set to 315 bits except for in Figs. 13 and 14.

A6: A video frame is defined as a video MU.

A7: Bit errors occur only during the transmission of the data payload. Neither the headers of the AL-PDU nor that of the MUX-PDU has any error, because FEC is applied to these headers.

A8: Bit errors occur independently with probability B_e for each bit (i.e., a random error environment[†]).

A9: The sum of the processing delay and propagation delay in ISDN is 20 ms. The propagation delay in the PHS channel is negligible.

A10: Each audio MU is generated just before the beginning of the audio slot.

A11: The processing time to encode a video sample is 33.3 ms, which is the same value of the sampling interval.

3.2 Characteristics of Live Media Source

The video substream used in the simulation was created by the H.263 compression scheme without any coding options [13]. We adopted the head view of a speaker and his voice as the media source. The size of a picture is 128 × 96 pixels (i.e., sub-QCIF: Quarter Common Intermediate Format). The H.263 coding bit rate is 47 kbit/s, and its target MU rate is 15 MU/s. These values are managed by two kinds of resolution control schemes: the *temporal resolution control* and the *spatial resolution control* [8]. Note that the former can skip some video samples according to the size of each video frame as well as the number of retransmission.

The audio substream was generated by the G.723.1 compression scheme, whose coding bit rate is 6.3 kbit/s. The size of an audio MU is 315 bits. Then, its MU rate is 20 MU/s. Other values of these parameters are used in Figs. 13 and 14.

3.3 Evaluation System

The structure of the software for performance evaluation is illustrated in Fig. 5. An H.263 video substream is encoded with the number of retransmission which is calculated by the simulator. Also, the video encoder notifies the simulator of the size of each video frame. In simulating the protocol and the error environment, we calculate the output time of each MU and add bit errors to the audio substream. We can watch the video output of the simulation with a display monitor and can listen to the audio output from a speaker.

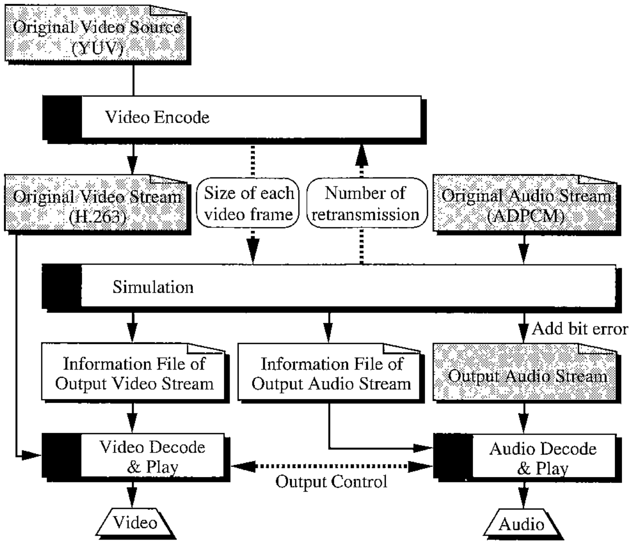


Fig. 5 Block diagram of the evaluation system.

3.4 Performance Measures

As performance measures in this paper, we adopt the root mean square (RMS) of *inter-stream synchronization error* (say the RMS of $\Delta_n^{(1-2)}$), the average MU delay, and the coefficient of variation $C_v^{(j)}$.

The *inter-stream synchronization error* $\Delta_n^{(1-2)}$ between substream 1 and substream 2 is expressed as follows:

$$\Delta_n^{(1-2)}=(D_n^{(1)}-D_m^{(2)})-(T_n^{(1)}-T_m^{(2)}) \qquad (n \geqq 1, m \geqq 1),$$

where the n-th video MU is compared with the m-th audio MU which is the last audio MU before the n-th video MU. The average MU delay of substream j is the average of the difference between $T_n^{(j)}$ and $D_n^{(j)}$. $C_v^{(j)}$ is defined as the coefficient of variation of MU output interval in stream j. It is calculated as follows:

$$C_v^{(j)} = [\ \{\ \Sigma_{n=1}^{N-1} (D_{n+1}^{(j)}-D_n^{(j)}-E^{(j)})^2\ \}\ /\ (N-1)\ ]^{1/2} / E^{(j)}$$

$$E^{(j)} = (D_N^{(j)}-D_1^{(j)})/(N-1),$$

where N is the reference number of the last MU. Note that $C_v^{(j)}$ indicates how smooth the output of MUs in stream j is. Then, we use $C_v^{(j)}$ in order to assess the quality of the intra-stream synchronization.

4. Numerical Results

In this section, we first evaluate the performance of the slide control scheme in the interleaved transmission. We next ex-

[†] The performance evaluation in a burst error environment is one of further studies.

amine the influence of the MUX-PDU size and that of the AL-PDU size. We then assess the effectiveness of the QOS control scheme with the maximum allowable modification time κ.

In the simulation, we use the following parameter values: T_{h0}=100 ms, θ_0=−50 ms, ω_0=1 sec, T_{h1}=100 ms, θ_1=50 ms, ω_1=1 sec. We choose these values in order that the increment of the transmission delay and the disturbance of the audio stream due to the slide control are not too large [8]. In Sects. 4.1 and 4.2, we choose κ=200 ms. We set $\tau_1^{(2)}$=0 sec except for in Fig. 8 so that the transmission delay of the first audio MU is minimized at the start of the communications (i.e., $t_1^{(2)}=A_1^{(2)}$). The target output time of the first video MU is calculated as $t_1^{(1)}=t_1^{(2)}+(T_1^{(1)}-T_1^{(2)})$. The duration of each simulation run in the following figures is 120 sec.

4.1 Performance of the Slide Control Scheme in the Interleaved Transmission

We first examine the throughput of the video transmission channel versus the bit error rate B_e in Fig. 6. The throughput is defined as the rate of successfully transmitted video data. We see that the throughput is 43.2 kbit/s when B_e=1.0 × 10^{-4}, for instance. However, it becomes a small value of 28.7 kbit/s when B_e=6.0 × 10^{-4}.

(1) Inter-stream Synchronization Error

Next, we assess the inter-stream synchronization error in order to evaluate how effective the slide control scheme is. In the following figures, the notation "SC" means that we adopt the slide control scheme, and "NC" implies that it is not exerted (i.e., $D_n^{(i)}=A_n^{(i)}$).

Figure 7 shows the RMS of $\Delta_n^{(1-2)}$ as a function of B_e. We see that the slide control scheme is effective in the inter-stream synchronization in the interleaved transmission with the Annex. In order to assess the quality of the inter-stream synchronization, we use Steinmetz's results [14], which indicate that $\Delta_n^{(1-2)}$ within 80 ms implies high quality of synchronization, whereas $\Delta_n^{(1-2)}$ over 160 ms is out of synchronization. Then, we find that "SC" can achieve excellent inter-stream synchronization even when B_e=6.0 × 10^{-4}, for instance, while "NC" cannot at the same value of B_e.

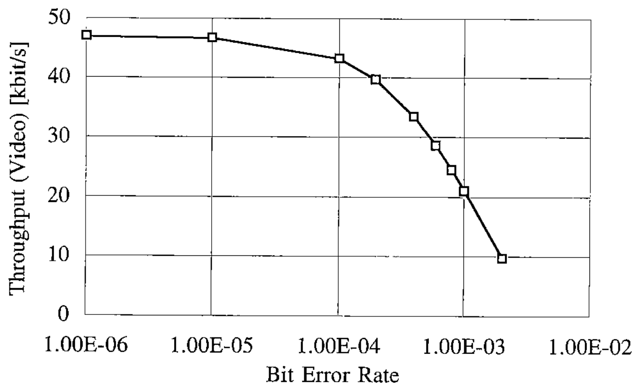


Fig. 6 Throughput vs. B_e.

We also observe that the error of "NC" is about 52 ms for small values of B_e, e.g. B_e=1.0 × 10^{-6}. This is because the difference in the MU size between audio and video causes a difference in the MU delay between the two and therefore an error of the inter-stream synchronization. This is one of the characteristics of the interleaved transmission.

From the above observation, we can expect that $\Delta_n^{(1-2)}$ of "NC" can be reduced if $\tau_m^{(2)}$ is set to a constant value which gives the delay of audio MUs nearly equal to the average MU delay of video. Then, we examine the influence of $\tau_m^{(2)}$ in the case of "NC". Figure 8 shows the RMS of $\Delta_n^{(1-2)}$ for four values of $\tau_m^{(2)}$ in "NC" as a function of B_e. Each symbol " × ," " △ ," " ○ ," and " ◇ " implies the case of "NC," where $\tau_m^{(2)}$=0, 25, 50, and 75 ms, respectively. For comparison, " □ " shows the case of "SC"; $\tau_1^{(2)}$=0 ms. We can see that the error of " ○ ," where $\tau_m^{(2)}$ is nearly equal to the difference in the average MU delay between the two (see Figs. 9 and 10), is the smallest among the four cases of "NC" for small values of B_e; the value of "○" is close to that of "SC." However, the RMS in every case of "NC" is much larger than that of "SC" for large values of B_e, since the retransmission disturbs the temporal constraints. Therefore, the slide control scheme is indispensable to preserve the inter-stream synchronization for large values of B_e.

(2) Average MU delay

For a live media transmission, the transmission delay is an

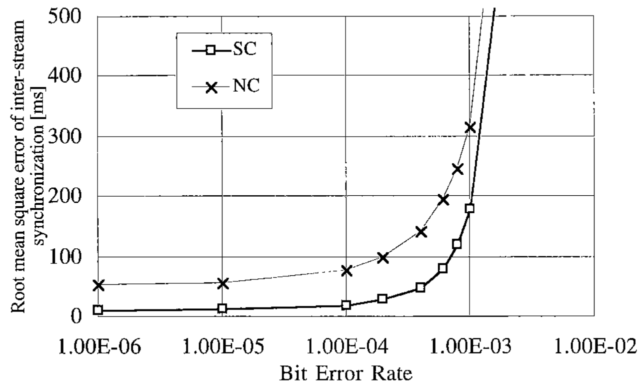


Fig. 7 Root mean square of $\Delta_n^{(1-2)}$ vs. B_e.

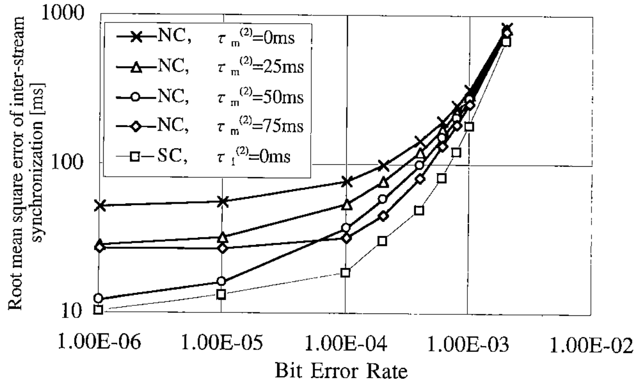


Fig. 8 Influence of $\tau_m^{(2)}$ on root mean square of $\Delta_n^{(1-2)}$.

important performance measure as well as the inter-stream synchronization. Then, we examine the average MU delay. Figures 9 and 10 plot the average MU delay of audio and that of video, respectively, as a function of B_e. We find that the slide control scheme increases the average MU delay of audio as well as that of video. In the case of audio substream (see Fig. 9), as B_e becomes larger, the average MU delay of "SC" increases, while "NC" is constant. However, it should be noted that the value of "SC" is less than 270 ms for large values of B_e. This is because the maximum allowable modification time κ restricts the amount of the modification due to the backward control, so that the transmission delay is not too large.

We also notice that the average MU delay of audio even in the case of "NC" is large (i.e., 75 ms). This is because it takes 50 ms to create an audio MU; the capturing time of an audio MU depends on the length of the MUX-PDU and the multiplexing pattern of information slots in the MUX-PDU. We will examine the influence of the MUX-PDU size in Sect. 4.2.

On the other hand, we see in Fig. 10 that the average MU delay of video is about 125 ms when B_e=1.0 × 10^{-6}. This is due to the encoding time (i.e., 33.3 ms), the propagation delay in ISDN (i.e., 20 ms), and the transmission time of a video MU, whose average is 66.6 ms, since the target MU rate is 15 MU/s.

(3) Intra-stream Synchronization Error

We examine the quality of the intra-stream synchronization by means of the coefficient of variation $C_v^{(i)}$ for each substream. Figures 11 and 12 illustrate $C_v^{(1)}$ versus B_e and $C_v^{(2)}$ versus B_e, respectively; a smaller value of this measure means that we can get MUs more smoothly.

As for the video substream, we notice in Fig. 11 that the slide control scheme is effective in decreasing the coefficient of variation; the $C_v^{(1)}$ of "SC" is smaller than that of "NC." On the other hand, we see in Fig. 12 that $C_v^{(2)}$ of "SC" is somewhat larger than that of "NC," which is always zero. However, each value of "SC" is sufficiently small.

Note that $C_v^{(2)}$ of "SC" has a peak; when B_e is large, the coefficient of "SC" decrease with increment of B_e. This is because a large bit error rate causes successive backward controls, which makes the total slide time reach the value of κ in a short time. However, the forward control scarcely occurs. Therefore, the frequency of slide control execution is low.

4.2 Influence of the MUX-PDU Size and the AL-PDU Size

We next examine the influence of the MUX-PDU size and that of the AL-PDU size for the video substream.

(1) Influence of the MUX-PDU size

Figures 13 and 14 illustrate the average MU delay of audio and that of video, respectively, as a function of B_e for three values of the MUX-PDU size, namely, 400, 800 and 1200 octets. Empty symbols represent the case of "SC," while solid

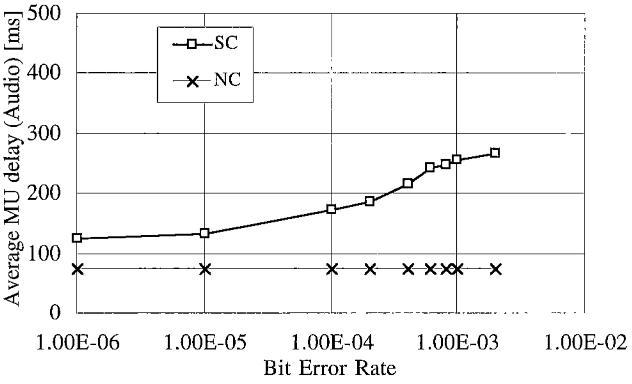


Fig. 9 Average MU delay of audio vs. B_e.

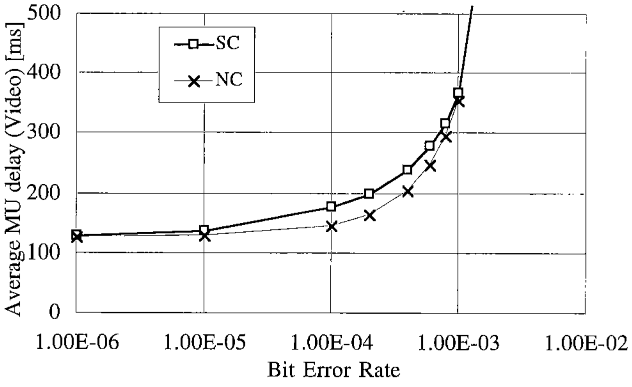


Fig. 10 Average MU delay of video vs. B_e.

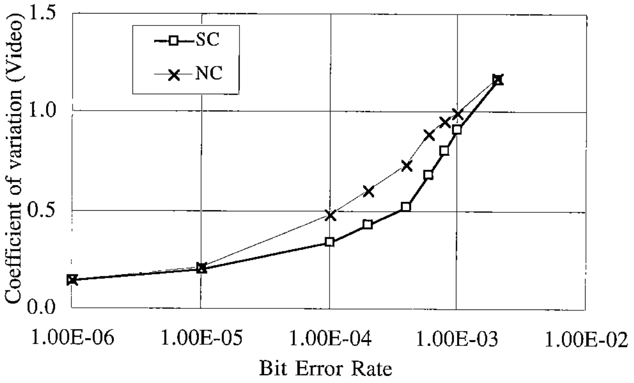


Fig. 11 $C_v^{(1)}$ vs. B_e.

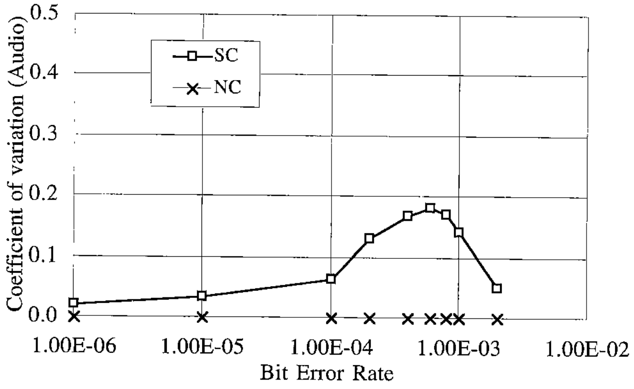


Fig. 12 $C_v^{(2)}$ vs. B_e.

symbols correspond to "NC." The figure associated with each symbol is the size of the MUX-PDU in octets. When the MUX-PDU size is 400, 800, and 1200 octets, the size of the audio slot is set to 40, 80 and 120 octets, respectively, because of the coding rate (i.e., 6.3 kbit/s). Since we assume that each audio MU is mapped onto an audio slot, the size of the audio MU is set to 315, 630 and 945 bits, respectively.

We see in the case of "NC" that as the size of the MUX-PDU becomes larger, the average delay of audio MUs increases, while the average MU delay of video is independent of the MUX-PDU size. That is, the audio MU delay depends on the MUX-PDU size.

However, the average MU delay of audio can be reduced by means of the multiplexing pattern, even though the MUX-PDU size is large. For example, when the MUX-PDU size is 800 octets, the audio slot is divided in half and the latter part is mapped into the middle of the video slot; an audio slot of 40 octets followed by a video slot of 356 octets is repeated twice in the MUX-PDU. Since the multiplexing pattern of this case is nearly equal to that of the case where the MUX-PDU size is 400 octets, we can expect that the audio MU delay becomes small. Thus, we see that the interval of the audio slot is a dominant factor of the audio MU delay. This result is another characteristic of the interleaved transmission with the Annex.

We also notice in the case of "SC" that the average MU delay of audio and that of video increases, as the MUX-PDU size becomes larger. Especially, the difference in the average MU delay of audio among the MUX-PDU sizes is large for large values of B_e, even though the same value of κ restricts the amount of the modification due to the backward control. This is because the target output time of the first audio MU for a larger MU size is behind that for a smaller size.

(2) Influence of the video AL-PDU size

Figures 15 and 16 plot the average MU delay of video and the RMS of $\Delta_n^{(1-2)}$, respectively, as a function of the AL-PDU size in the case of "SC." In these figures, each symbol of " ○ ," " ◇ ," " □ ," " △ ," " ▲ ," and " × " shows the case where $B_e = 1.0 \times 10^{-6}$, 1.0×10^{-4}, 4.0×10^{-4}, 6.0×10^{-4}, 8.0×10^{-4}, and 1.0×10^{-3}, respectively.

We find in Fig. 15 that the average MU delay of video has the bottom for each value of B_e; the value of the bottom increases, as B_e becomes larger. When the AL-PDU size is not too small, the MU delay increases gradually with the increment of the size; the average MU delay for large values of B_e is more sensitive to the AL-PDU size than that for small values of B_e. This result means that a large size of the AL-PDU causes the large delay of video MUs for large values of B_e. On the other hand, when the AL-PDU size is too small, the average MU delay of video decreases steeply with an increase in the size, since the ratio of the header size to the AL-PDU size is very large. Thus, the AL-PDU size of 104

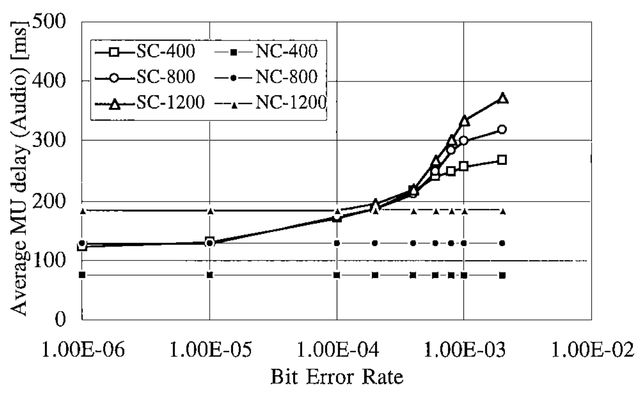


Fig. 13 Influence of MUX-PDU size on average MU delay of audio.

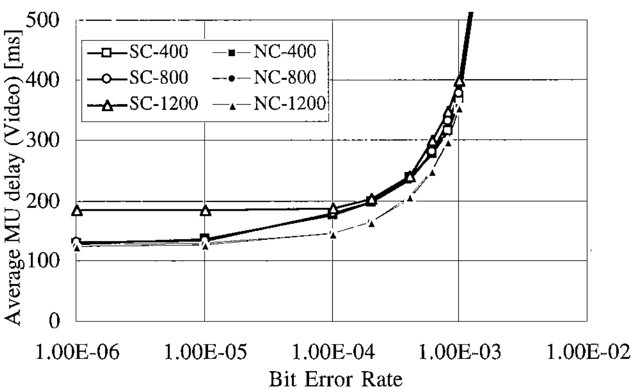


Fig. 14 Influence of MUX-PDU size on average MU delay of video.

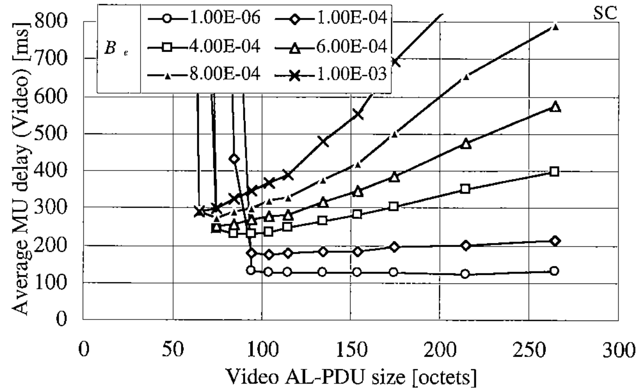


Fig. 15 Average MU delay of video vs. video AL-PDU size.

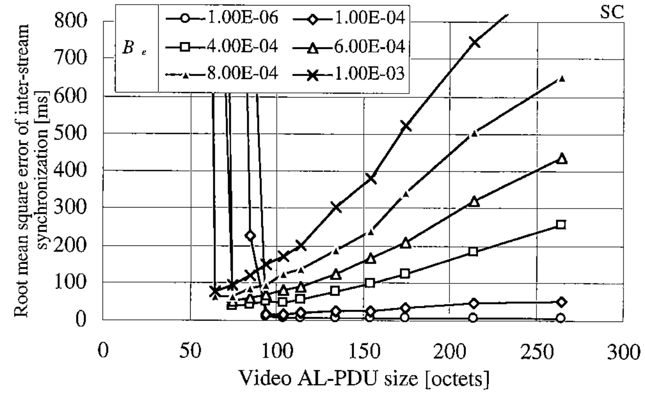


Fig. 16 Root mean square of $\Delta_n^{(1-2)}$ vs. video AL-PDU size.

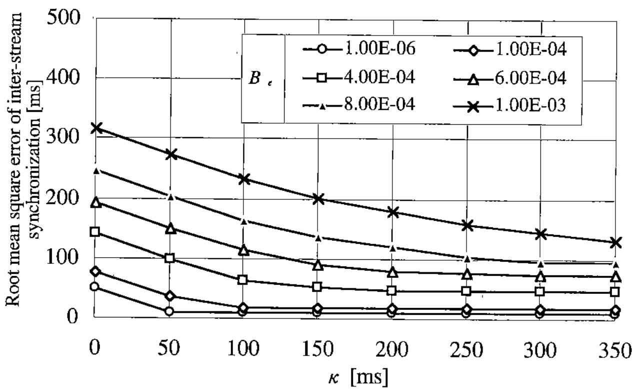


Fig. 17 Root mean square of $\Delta_n^{(1-2)}$ vs. κ.

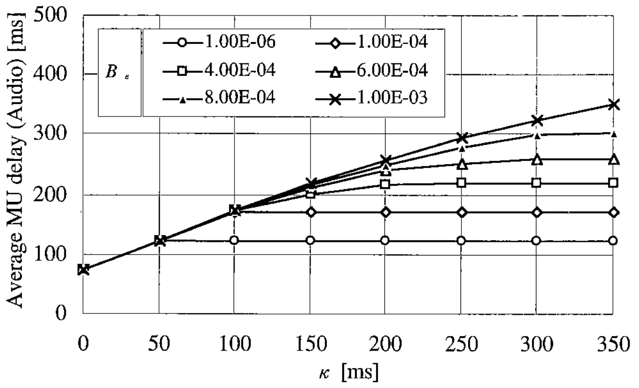


Fig. 18 Average MU delay of audio vs. κ.

octets, which we have adopted in assumption A2, is a good selection, since the size provides small MU delays for all the values of B_e. This implies the size of the data payload is 90 octets.

As for the inter-stream synchronization (see Fig. 16), we can get the same results as those of the average MU delay of video. That is, the AL-PDU size is a crucial parameter for the lip-synch.

4.3 Effectiveness of the QOS Control Scheme

We examine the effectiveness of the QOS control scheme with the maximum allowable modification time κ. Figure 17 shows the RMS of $\Delta_n^{(1-2)}$ as a function of κ. Figures 18 and 19 plot the average MU delay of audio and that of video, respectively, as a function of κ. In these figures, each symbol shows the same case as that of Figs. 15 and 16.

We see in Figs. 17 and 18 that there is a trade-off between the quality of the inter-stream synchronization and that of the transmission delay of audio; as the value of κ becomes larger, the RMS of $\Delta_n^{(1-2)}$ decreases, while the average MU delay of audio increases. This result means that a large value of κ allows the frequent execution of the backward control in order to keep the inter-stream synchronization at the expense of the transmission delay of audio. Also, a small value of κ restricts the execution of the backward control so as to reduce the transmission delay, even though the inter-stream synchronization is disturbed. Especially, these performance measures for large values of B_e are more sensitive to the value of κ than those for small values of B_e. Therefore, the QOS control scheme using the parameter κ is effective in managing both quality of the inter-stream synchronization and that of the transmission delay of audio for large values of B_e. However, the QOS control scheme is less effective in managing the quality of the transmission delay of video (see Fig. 19).

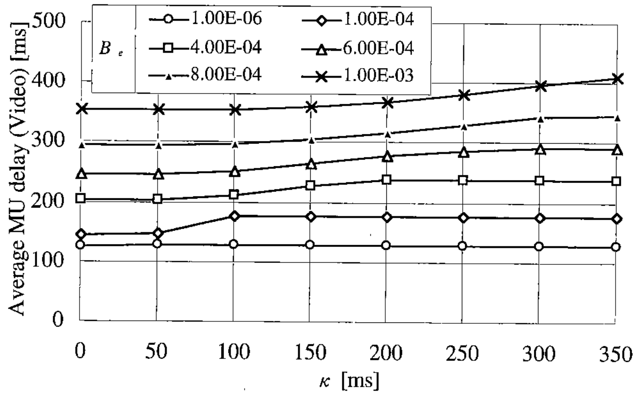


Fig. 19 Average MU delay of video vs. κ.

Note that both the RMS of $\Delta_n^{(1-2)}$ and the average MU delay of audio when B_e=1.0 × 10^{-6} or 1.0 × 10^{-4} (plotted by "○" or "◇") are not sensitive to the value of κ when it is not small. This is because that the backward control is sufficiently made with small values of κ to preserve the temporal constraints in these situations. Note that the slide control scheme keeps the transmission delay very small, even if we select a large value of κ for small values of B_e. The result is highly significant for a live media transmission.

Next, we assess the quality of the inter-stream synchronization by means of Steinmetz's results. For example, consider the case where B_e =6.0 × 10^{-4}. Then, if we set κ to 200 ms or more, we can achieve excellent inter-stream synchronization. However, the average MU delay of audio is about 242 ms. On the other hand, if we set κ to 0 ms, the average MU delay is 75 ms, whereas the temporal relationship between the video substream and the audio substream is out of synchronization.

5. Conclusions

We studied the application of the slide control scheme to the H.223 Annex in PHS. We also proposed a QOS control scheme to control both quality of the media synchronization and that of the transmission delay. We have evaluated the effectiveness of these control schemes for live media transmission over PHS.

We first observed that the slide control scheme is effective in the inter-stream synchronization in the interleaved transmission with the H.223 Annex at the expense of the transmission delay. We also found that the QOS control scheme using the parameter κ is effective in managing both quality of the inter-stream synchronization and that of the transmission delay of audio. In addition, we noticed several features of media synchronization quality due to the interleaved transmission. That is, the interval of the audio slot is a dominant factor of the audio MU delay. Also, the difference in the MU size between audio and video causes a slightly large error of the inter-stream synchronization unless the control is exerted.

Our future work includes the application of the adaptive QOS control scheme and performance comparisons between the interleaved transmission and non-interleaved one. In addition, it is desirable to devise new control schemes which utilize characteristics of media streams, such as silent periods in an audio stream.

Acknowledgment

The work was supported by *Telecommunications Advancement Organization of Japan* and by the *Grant-In-Aid for Scientific Research of Japan's Ministry of Education, Science and Culture.*

References

[1] Association of Radio Industries and Businesses (ARIB), "Personal Handy Phone System ARIB Standard Version 3, RCR STD-28," Nov. 1997.

[2] ITU-T Draft Recommendation H.223/Annex A (Draft), "Multiplexing Protocol for Low Bitrate Mobile Multimedia Communication over Low Error-Prone Channels," Sept. 1997.

[3] ITU-T Draft Recommendation H.223/Annex B (Draft), "Multiplexing Protocol for Low Bitrate Mobile Multimedia Communication over Moderate Error-Prone Channels," Dec. 1997.

[4] ITU-T Draft Recommendation H.223/Annex C (Draft), "Multiplexing Protocol for Low Bitrate Mobile Multimedia Communication over Highly Error-Prone Channels," Sept. 1997.

[5] ITU-T Recommendation H.263, "Video coding for low bit rate communication," March 1996.

[6] ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communication transmitting at 5.3 & 6.3 kbit/s," March 1996.

[7] G.Blakowski and R.Steinmetz, "A media synchronization survey: Reference model, specification, and case studies," IEEE J. Sel. Areas Commun., vol.14, no.1, pp.5-35, Jan. 1996.

[8] M.Kato, N.Usui, and S.Tasaka, "Performance evaluation of live media synchronization in PHS," IEICE Technical Report, CQ97-62, Dec. 1997.

[9] S.Tasaka, H.Nakanishi, and Y.Ishibashi, "Dynamic resolution control and media synchronization of MPEG in wireless LANs," Conf. Rec. IEEE GLOBECOM '97, pp.138-144, Nov. 1997.

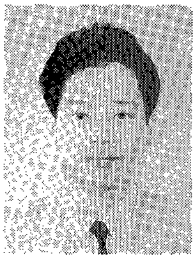
[10] Y.Ishibashi and S.Tasaka, "A synchronization mechanism for continuous media in multimedia communications," Proc. IEEE INFOCOM '95, pp.1010-1019, April 1995.

[11] S.Tasaka, Y.Ishibashi, and H.Imura, "Stored media synchronization in wireless LANs," Conf. Rec. IEEE GLOBECOM '96, pp.1904-1910, Nov. 1996.

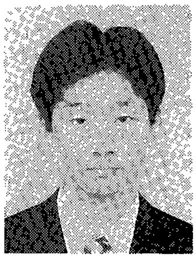
[12] M.Kato, N.Usui, and S.Tasaka, "Stored media synchronization based on buffer occupancy in PHS," Proc. IEEE PIMRC '97, pp.1049-1053, Sept. 1997.

[13] Source program code of "TMN H.263 Encoder ver.2.0 & Decoder ver.2.0," Telenor R&D, June 1996.

[14] R.Steinmetz, "Human perception of jitter and media synchronization," IEEE J. Sel. Areas Commun., vol.14, no.1, pp.61-72, Jan. 1996.



Masami Kato received the B.S. degree in physics from Nagoya University, Nagoya, Japan, in 1984. He joined Sanyo Electric Co., Ltd. in 1984, and engaged in the development of mobile multimedia communication systems in PHS and integrated wired and wireless network systems. He is presently a chief researcher in Hypermedia Research Center of Sanyo.



Yoshihito Kawai received the B.S. and the M.S. degrees in electrical and computer engineering from Nagoya Institute of Technology, Nagoya, Japan, in 1995 and 1998, respectively. At Nagoya Institute of Technology, he studied mobile multimedia communication systems in PHS. In April 1998, he joined Matsushita Electric Industrial Co., Ltd.

Shuji Tasaka received the B.S. degree in electrical engineering from Nagoya Institute of Technology, Nagoya, Japan, in 1971, and the M.S. and Ph. D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1973 and 1976, respectively. Since April 1976, he has been with Nagoya Institute of Technology, where he is now a Professor in the Department of Electrical and Computer Engineering. In the 1984-1985 academic year, he was a Visiting Scholar in the Department of Electrical Engineering at the University of California, Los Angeles. His current research interests include wireless networks, high-speed networks and multimedia communication protocols. He is the author of a book entitled *Performance Analysis of Multiple Access Protocols* (Cambridge, MA: The MIT Press, 1986). Dr. Tasaka is a member of the IEEE, ACM and Information Processing Society of Japan.