PAPER Special Section on Fundamentals of Multi-dimensional Mobile Information Network

Media Synchronization Control Based on Buffer Occupancy for Stored Media Transmission in PHS*

Masami KATO†, Noriyoshi USUI†, *and* Shuji TASAKA††, *Members*

SUMMARY This paper proposes a scheme for synchronization of stored video and audio streams in PHS. A video stream of H.263 is transmitted over a PHS channel with ARQ control, while an audio stream of 32 kbit/s ADPCM is sent on another channel without any control. In order to preserve the temporal constraints within the video stream as well as the relationship between the video and audio streams, we adopt a new control scheme which modifies the target output time according to the amount of video data in the receive-buffer. Through simulation we assess the characteristics of this scheme in both random and burst error environments and confirm the effectiveness of the scheme.

key words: *media synchronization, PHS, stored media, inter-stream synchronization, intra-stream synchronization, simulation*

1. Introduction

There is a great demand for mobile video communications. PHS (Personal Handy phone System) [1] is one of the second generation mobile telecommunication systems. It offers both speech and bearer services, whose transmission rate is 32 kbit/s. The H.263-based video coding [2] is suitable for mobile video communications over PHS because of its coding rate.

This paper proposes a scheme for synchronization of stored H.263 video and audio streams in PHS. These streams are transmitted over separate PHS channels. An H.263 coded video stream requires some error control for transmission over a PHS channel because of its inherent error sensitivity. We employ an ARQ (Automatic Repeat reQuest) technique specified by the PHS Internet Access Forum (PIAF) [3]. We also assume that an audio stream of 32 kbit/s ADPCM (Adaptive Differential Pulse Code Modulation) is sent on another channel without any control so that the speech service already offered can be used.

In order to obtain excellent quality of the media at the destination, we need to keep the temporal constraints of these streams. That is, we must achieve two types of media synchronization: *intra-stream synchronization* and *inter-stream synchronization* [4]. The former is for preservation of temporal constraints within each media stream, while the latter is for keeping the temporal relationships among plural media streams.

In this paper, we consider a media synchronization problem in the transmission scheme mentioned earlier. The problem arises since retransmissions of video packets disturb temporal constraints within the video stream; the H.263 coding rate we have selected is nearly equal to the PHS transmission rate. Moreover, delayed video frames distort the temporal relation between the video and audio streams.

In [5], the authors have proposed two kinds of control schemes for the media synchronization: *B-block (Bidirectionally predictive block) discarding control* and *slide control*. We have reported that the B-block discarding control is effective in the intra-stream synchronization. It discards approximately the same amount of B-blocks as that of retransmitted video data. We have also confirmed that the slide control is effective in the inter-stream synchronization. It modifies the *target output time*, which is a calculated time at the destination to keep the intra-stream synchronization; however, the modification of the target output time has a side effect that the intra-stream synchronization of an audio stream is temporarily disturbed.

As a criterion of the slide control in [5], we have adopted the intra-stream synchronization error of the video stream. This scheme modifies the target output time, if a video frame arrives a specified amount of time later than its target output time. In this situation, the quantity of video data in the receive-buffer is small. That is, the low buffer occupancy causes large error of media synchronization. Therefore, we can expect the improvement of the media synchronization, if the slide control is executed before the receive-buffer occupancy becomes low. Note that the receive-buffer occupancy increases during the term of the modification.

On the basis of the above observation, we propose a new version of the slide control scheme in which the quantity of video data in the receive-buffer is also used to judge the modification of the target output time. Through simulation we assess the characteristics of this scheme in random error environments. In addition, we also consider burst error environments which are caused by terminal mobility at a walking speed. We then confirm the effectiveness of the scheme in both environments.

The rest of the paper is organized as follows. Section 2 specifies media synchronization algorithms adopted in this paper. Section 3 describes our simulation methodology, and Section 4 discusses numerical results.

Manuscript received October 24, 1997.
Manuscript revised February 9, 1998.

† The authors are with SANYO Electric Co., Ltd., Gifu-ken, 503-0195 Japan.
†† The author is with the Dept. of Electrical and Computer Engineering, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan.

* This paper was presented in part at PIMRC'97, Helsinki, Finland, September 1997.

2. Media Synchronization Algorithm

2.1 Assumptions and Notations

We suppose that a PHS mobile terminal receives and plays a video stream of H.263 as well as an audio stream of 32 kbit/s ADPCM; they are transmitted from a server station via a PHS base station (see Fig. 1). As mentioned earlier, the video stream is transmitted over a PHS channel with ARQ control, while the audio stream is sent on another channel without any control.

We assume in this paper that the server station has a video server and an audio server, which store video and audio media units (MUs), respectively. The MU is a unit of data for the output of the media. In this paper, a video MU is defined as a video frame, and a constant number of audio samples constitutes an audio MU. Therefore, the size of an audio MU is constant, although the size of a video MU is variable. Each stream is recorded at a constant MU rate.

Each MU has a timestamp in order to preserve the temporal relationship among these MUs. The *TR* (Temporal Reference for I- and P-pictures) and TR_B (TR for B-picture) in an H.263 video stream can be utilized as timestamps. An audio MU does not always need a timestamp because of its constant MU rate and size.

In order to specify the media synchronization algorithm, we consider a stream (say stream *j*), which can be either the video stream (j=1) or the audio stream (j=2). For stream *j*, we define the following notations [6]:

$T_n^{(j)}$: the timestamp of the n-th MU in stream j at the source.
$\sigma_{n,n+1}^{(j)}$: the difference between timestamps of the n-th and (n+1)-st MUs in stream j; i.e., $\sigma_{n,n+1}^{(j)}=T_{n+1}^{(j)}-T_n^{(j)}$.
$t_n^{(j)}$: the target output time of the n-th MU in stream j at the destination.
$x_n^{(j)}$: the ideal target output time of the n-th MU in stream j at the destination.
$A_n^{(j)}$: the arrival time of the n-th MU in stream j at the destination.
$D_n^{(j)}$: the output time of the n-th MU in stream j at the destination.
$\tau_n^{(j)}$: the output waiting time of the n-th MU in stream j at the destination; i.e., $\tau_n^{(j)}=D_n^{(j)}-A_n^{(j)}$.
$M_R^{(j)}$: the generating rate of MUs in stream j.

where n = 1,2,⋯. Note that $x_n^{(j)}$ denotes the ideal target output time, which we would have if there were no network delay jitter.

In order to specify the slide control, we define the slide time $\Delta S_n^{(j)}$, which is the difference between the original target output time $t_n^{(j)}$ and the modified one. Also, the total slide time $S_n^{(j)}$ is defined as $S_n^{(j)}=\Sigma_{i=1}^{n}\Delta S_i^{(j)}$. Then, $t_n^{(j)}$ and $D_n^{(j)}$ are calculated as follows:

$$t_1^{(j)}=A_1^{(j)}+\tau_1^{(j)},\qquad x_1^{(j)}=t_1^{(j)}\qquad (n=1)$$

$$t_n^{(j)}=x_n^{(j)}+S_{n-1}^{(j)},\qquad x_n^{(j)}=x_{n-1}^{(j)}+\sigma_{n-1,n}^{(j)}\qquad (n\geqq 2)$$

$$D_1^{(j)}=t_1^{(j)}\qquad (n=1)$$

$$D_n^{(j)}=t_n^{(j)}+\Delta S_n^{(j)}\qquad (n\geqq 2)\ \text{if}\ A_n^{(j)}\leqq t_n^{(j)}+\Delta S_n^{(j)}$$

$$D_n^{(j)}=A_n^{(j)}\qquad (n\geqq 2)\ \text{if}\ t_n^{(j)}+\Delta S_n^{(j)}<A_n^{(j)}.$$

It should be noted that $t_n^{(j)}$ means the original target output time before the modification due to the slide control; the modified one is calculated as $t_n^{(j)}+\Delta S_n^{(j)}$.

In order to evaluate the quality of the media synchronization, we define the intra-stream synchronization error $\Delta_n^{(j)}$ in stream j and the inter-stream synchronization error $\Delta_n^{(1\text{-}2)}$ between stream 1 and stream 2. When $\Delta_n^{(1\text{-}2)}$ is calculated, 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. Thus, $\Delta_n^{(j)}$ and $\Delta_n^{(1\text{-}2)}$ are defined as follows:

$$\Delta_n^{(j)}=D_n^{(j)}-t_n^{(j)}\qquad (n\geqq 1)$$

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

where $m=\mathrm{int}\{(n-1)\times M_R^{(2)}/M_R^{(1)}\}+1$.

2.2 Criteria for the Slide Control

To improve the slide control in [5], we now consider the slide control using two kinds of criteria: the intra-stream synchronization error of the video stream (say the criterion $\Delta_n^{(1)}$) and the quantity of video data in the receive-buffer, which is defined as $Q^{(1)}$, (say the criterion $Q^{(1)}$). The criterion $\Delta_n^{(1)}$, which is the same criterion as that in [5], modifies the target output time, when $\Delta_n^{(1)}$ exceeds a threshold value. The new criterion $Q^{(1)}$ modifies the target output time, when $Q^{(1)}$ is less than a threshold value.

It should be noted that the intra-stream synchronization of an audio stream is disturbed only by $\Delta S_n^{(2)}$, when the slide control is executed by these criteria. Taking the influence of the slide control on the audio stream into consideration, we introduce two types of modification schemes: *gradual recovery* and *fast recovery*†. The gradual recovery scheme modifies the target

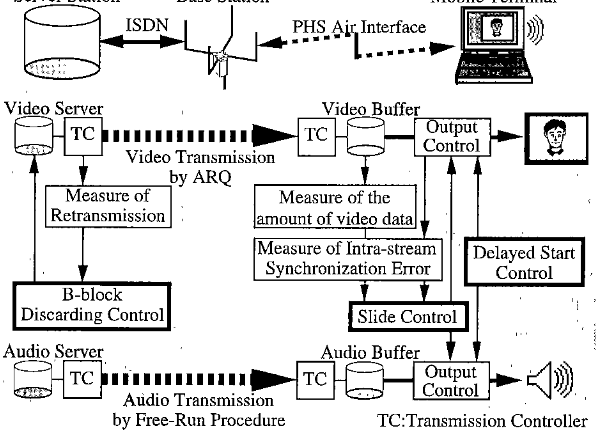


Fig. 1 Block diagram of the stored media transmission system .

† 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 [7] are alternative control policies on the modification.

Table 1 Patterns of the slide control.

pattern	description	
(a)	if $\Delta_n^{(1)} \geq T_{h2}$,	then $\Delta S_n^{(i)} = \theta_2$(Quick recovery).
(b)	if $T_{h1} \leq \Delta_n^{(1)} < T_{h2}$,	then $\Delta S_n^{(i)} = \theta_1$(Gradual recovery).
(c)	if $\Delta_n^{(1)} < T_{h1}$ and $Q^{(1)} < T_{hQ}$,	then $\Delta S_n^{(i)} = \theta_1$(Gradual recovery).
(d)	if $\Delta_n^{(1)} < T_{h1}$ and $Q^{(1)} \geq T_{hQ}$,	then $\Delta S_n^{(i)} = 0$.

output time step by step so that the disturbance of an audio stream is limited to a small one. The fast recovery scheme modifies the target output time largely at once in order to recover from the inter-stream asynchrony quickly, while the disturbance of an audio stream is large.

Then, we denote a threshold for the new criterion $Q^{(1)}$ by T_{hQ}, which is a threshold for the gradual recovery. We also define two kinds of thresholds for the criterion $\Delta_n^{(1)}$: T_{h1} and T_{h2}. The former is a threshold for the gradual recovery, while the latter is a threshold for the fast recovery. Additionally, in order to specify the two types of recovery schemes, we define both the modification size of the output time and a minimum allowable interval between two successive control. These parameters are denoted by θ_i and ω_i, respectively, where i indicates the kind of the recovery scheme: either the gradual recovery (i=1) or the fast recovery (i=2).

Table 1 describes the algorithm that determines which kind of slide control can be executed. On receiving the n-th video MU, the destination calculates $\Delta_n^{(1)}$ and compares it with both T_{h1} and T_{h2}. If $\Delta_n^{(1)}$ is larger than or equal to the threshold value T_{h2}, and if the interval between the last execution of the fast recovery and the current time is over the value of ω_2, the fast recovery is executed. Also, if $\Delta_n^{(1)}$ is larger than or equal to T_{h1} and less than T_{h2}, and if the interval between the last execution of the gradual recovery and the current time is over the value of ω_1, then the gradual recovery is carried out. When $\Delta_n^{(1)}$ is less than T_{h1}, the destination calculates $Q^{(1)}$. If $Q^{(1)}$ is less than the threshold value T_{hQ}, and if the interval is over the value of ω_1, the gradual recovery is made. These modifications can be executed only after the output of the first video MU.

Note that when the slide control is performed, the output of the audio pauses during $\Delta S_n^{(2)}$; this interval is called the *pause time*. However, if the destination can repeat the previous audio MUs, there is no pause of the audio stream. Especially, in the case of the gradual recovery, the repetition sounds like an echo, if we choose appropriate values of the parameters θ_1 and ω_1. This technique improves the quality of the output from a subjective point of view.

3. Simulation Methodology

3.1 Assumptions for Simulation

We make the following assumptions for simulation:

A1: An H.263 video stream is transmitted with SR (Selective Repeat) -ARQ control between the server and a PHS mobile terminal. Figure 2 shows the structure of the ARQ frame [3]. Each ARQ frame is composed of ARQ control fields and a data payload. The size of the data payload is 584 bits, while that of an ARQ frame is 640 bits.

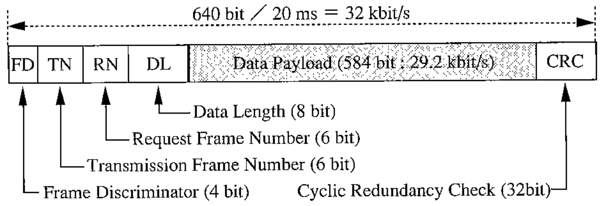


Fig. 2 Structure of the ARQ frame.

A2: Bit errors occur during the wireless transmission of the data payload over PHS, while there is no error on sending video and audio streams over ISDN.

A3: In the random error environment, bit errors occur independently with probability B_e for each bit.

A4: For the burst error environment, we use the Gilbert model (i.e., two state Markov chain model) [8]. The transition probabilities of the model are calculated from B_e and the mean of burst length; their values are measured actually in a fading environment with a fading frequency of 8 Hz, which corresponds to a walking speed [9], [10].

A5: The sum of the processing delay due to the ARQ frame structure and propagation delay in ISDN is 40 ms. The propagation delay in the PHS channel is negligible.

3.2 Characteristics of Stored Media Source

We have created a video stream by the H.263 compression scheme with only "PB-frame mode" of coding options [11]. We adopted a scene of the movie "*Labyrinth*" as the media source. We choose 128 × 96 pixels as the size of a picture (i.e., sub-QCIF: Quarter Common Intermediate Format) in order to obtain the good quality of pictures and a larger value of the video MU rate even in the low bit rate communication. We set $M_R^{(1)}$=15 MU/s, $M_R^{(2)}$=10 MU/s, and the size of an audio MU is 3,200 bits. The H.263 coding bit rate is 29.2 kbit/s, which is managed by a spatial resolution control scheme.

Characteristics of the stored media source are as follows:

C1: The original recording time of the stream is 122.333 sec.

C2: The total number of I-, P- and PB- frames is 1,102, while the number of PB- frames is 734.

C3: The amount of video data is 3,583,656 bits.

C4: The occupancy ratio of B-blocks is 13.1%.

C5: The mean of the quantizer information, which is defined as half the step size of the quantizer in each video frame, is 11.2.

3.3 Evaluation System

The structure of the software for performance evaluation is illustrated in Fig. 3. Not only original streams but also an information file of a video stream is necessary for simulation. This file contains the information on the reference number, picture type, frame size and B-block size for 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 stream.

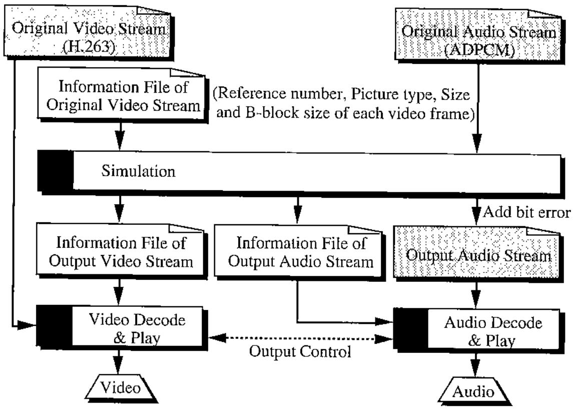


Fig. 3 Block diagram of the evaluation system.

We can watch the video output of the simulation with a display monitor and can listen to the audio output from a speaker.

3.4 Performance Measures

As performance measures in this paper, we adopt the mean of inter-stream synchronization error $\Delta_n^{(1\text{-}2)}$, the mean of intra-stream synchronization error $\Delta_n^{(j)}$, the coefficient of variation $C_v^{(j)}$ [12] and the total slide time $S_N^{(1)}$, where N is the reference number of the last MU. $C_v^{(j)}$ is defined as the coefficient of variation of MU output interval in stream j. It is calculated as follows:

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

$$E^{(j)} \triangleq (D_N^{(j)} - D_1^{(j)})/(N-1).$$

Note that $C_v^{(j)}$ indicates how smooth the output of MUs in stream j is, while the mean of $\Delta_n^{(j)}$ expresses how exactly the temporal constraints within stream j is preserved. We consider that it is important to obtain the smooth output of MUs from a subjective point of view. Therefore, in order to assess the quality of the intra-stream synchronization, we use not only the mean of $\Delta_n^{(j)}$ but also $C_v^{(j)}$.

4. Numerical Results

In this section, we first evaluate the performance of the slide control scheme using the new criterion $Q^{(1)}$. We next examine the pause time caused by the criterion $Q^{(1)}$ and make subjective assessment of the media quality. We then discuss appropriate values of the threshold T_{hQ}.

In the simulation, we use the following parameter values: T_{h1}=100 ms, θ_1=100 ms, ω_1=1 sec, T_{h2}=2 sec, θ_2=4 sec, and ω_2=60 sec. In subsections 4.1 and 4.2, we set T_{hQ}=29.2 kbit. We have chosen the values of θ_1 and ω_1 so that the repetition of audio MUs can sound like an echo. On the other hand, the output waiting time of the first video MU (i.e., $\tau_1^{(1)}$) is effective in the media synchronization, since the receive-buffer can absorb the temporary delay jitters of video MUs. In [5], we show that the quality of the media synchronization is insensitive to $\tau_1^{(1)}$ when the slide control is adopted. In this paper, we denote the sum of $\tau_1^{(1)}$ and the necessary time for transmission of the first video

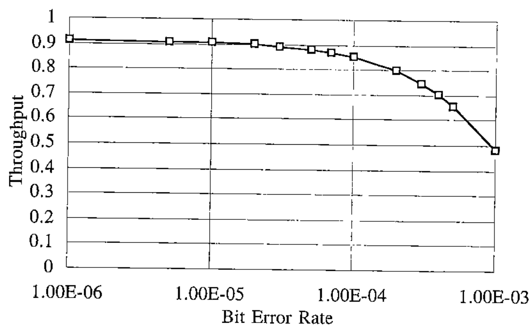


Fig. 4 Throughput versus B_e in random error environment.

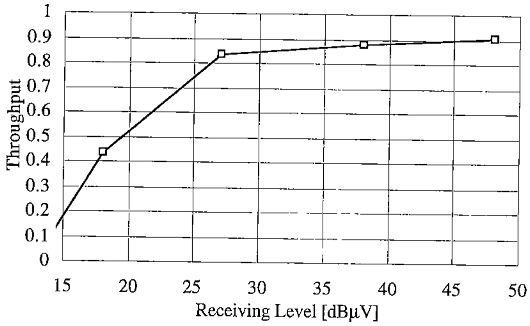


Fig. 5 Throughput versus R_e in burst error environment.

MU as S, and we set S=5 sec.

4.1 Performance of the Control Scheme with Criterion $Q^{(1)}$

We first examine the throughput of the video transmission channel versus the bit error rate B_e in the random error environment and the throughput versus the receiving level R_e in the burst error environment in Figs. 4 and 5, respectively. The throughput is defined as the ratio of the rate of successfully transmitted video data to the transmission capacity of PHS (32 kbit/s). We see in Fig. 4 that the throughput is 0.85 when $B_e = 1.0 \times 10^{-4}$. We also notice in Fig. 5 that the throughput is 0.84 when R_e=27dBµV.

Next, we examine the inter-stream synchronization error, the intra-stream synchronization error and the coefficient of variation in order to evaluate the performance of the slide control scheme with the criterion $Q^{(1)}$.

4.1.1 The Inter-Stream Synchronization Error

Figures 6 and 7 show the mean of $\Delta_n^{(1\text{-}2)}$ as a function of B_e in the random error environment and as a function of R_e in the burst error environment, respectively. The notation "Δ+Q" in the figures means that we adopt both the criterion $\Delta_n^{(1)}$ and the criterion $Q^{(1)}$, and "Δ" implies that the only criterion $\Delta_n^{(1)}$ is applied. Also, "BD+Δ+Q" and "BD+Δ" indicate that the B-block discarding control is adopted together with the same criteria as those of "Δ+Q" and "Δ," respectively. In the case of "NC," neither the slide control nor the B-block discarding control is exerted.

Comparing "Δ+Q" with "Δ" in Fig. 6, we see that the slide control scheme with the criterion $Q^{(1)}$ is effective in the inter-stream synchronization. For example, the mean of $\Delta_n^{(1\text{-}2)}$ of "Δ+Q" is smaller than that of "Δ" when $B_e \geqq 2.0 \times 10^{-5}$.

Now, in order to assess the quality of the inter-stream synchronization, we use Steinmetz's results [13], which indicate that $\Delta_n^{(1\text{-}2)}$ within 80 ms implies high quality of synchronization, whereas $\Delta_n^{(1\text{-}2)}$ over 160 ms is out of synchronization; we have confirmed that these criteria can be applied to a sub-QCIF based video communication by means of subjective assessment. Then, we find that "Δ+Q" can achieve excellent inter-stream synchronization when $B_e = 1.0 \times 10^{-4}$, for instance, while "Δ" cannot at the same value of B_e. On the other hand, the mean of $\Delta_n^{(1\text{-}2)}$ of "Δ+Q" is larger than that of "Δ" when $B_e < 2.0 \times 10^{-5}$. This is because the slide control due to the criterion $Q^{(1)}$ is often executed, even if B_e is small. Additionally, when $t_n^{(1)}$ is modified, not $t_m^{(2)}$ but $t_{m+1}^{(2)}$ is changed, while $\Delta_n^{(1\text{-}2)}$ is calculated using $t_m^{(2)}$. However, the difference of performance makes no sense since each value is much smaller than 80 ms.

Next let us consider the cases of applying the B-block discarding control along with the slide control (i.e., "BD+Δ+Q" and "BD+Δ"). Figure 6 shows that the slide control scheme with the criterion $Q^{(1)}$ also results in the improvement of the inter-stream synchronization. Comparing "BD+Δ+Q" with "Δ+Q," we notice that the B-block discarding control is effective in the inter-stream synchronization. However, it should be noted that the degree of the performance improvement by the B-block discarding control depends on the characteristics of the video stream such as the size of B-blocks, while the effectiveness of the slide control is independent of the characteristics of the video stream. Therefore, it is important to adopt the slide control scheme with the criterion $Q^{(1)}$ in order to keep the excellent media synchronization.

Furthermore, from Fig. 7, we see that the slide control scheme with the criterion $Q^{(1)}$ is effective in the inter-stream synchronization in the burst error environment as well.

4.1.2 The Intra-Stream Synchronization Error

In addition, we examine the intra-stream synchronization. Figures 8 and 9 plot the mean of $\Delta_n^{(2)}$ versus B_e and the mean of $\Delta_n^{(1)}$ versus B_e, respectively, in the random error environment. Also, Figs. 10 and 11 show the mean of $\Delta_n^{(2)}$ versus R_e and the mean of $\Delta_n^{(1)}$ versus R_e, respectively, in the burst error environment.

As for the audio stream in the random error environment, we observe in Fig. 8 that the slide control with the criterion $Q^{(1)}$ produces larger errors than that without the criterion and that the difference in the error between the two increases as B_e becomes smaller. This is because the target output time of audio MUs is modified, even if B_e is small. However, each value is not too large. Also, note that the intra-stream synchronization error of the audio for "NC" is always zero. On the other hand, we notice in Fig. 9 that the slide control caused by the criterion $Q^{(1)}$ is also effective in the intra-stream synchronization of the video stream.

In the burst error environment (see Figs. 10 and 11), we again see that the slide control scheme with the criterion $Q^{(1)}$ is effective in the intra-stream synchronization of the video stream except for the case of applying the B-block

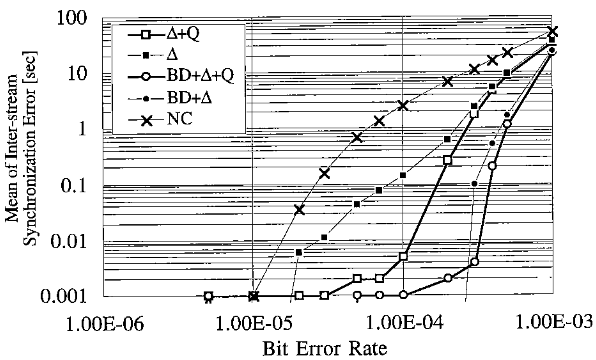


Fig. 6 Mean of $\Delta_n^{(1\text{-}2)}$ versus B_e in random error environment.

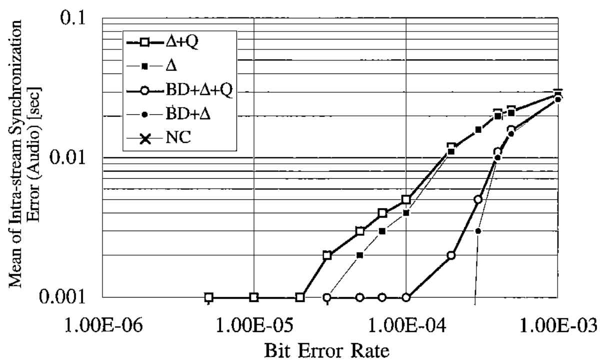


Fig. 8 Mean of $\Delta_n^{(2)}$ versus B_e in random error environment.

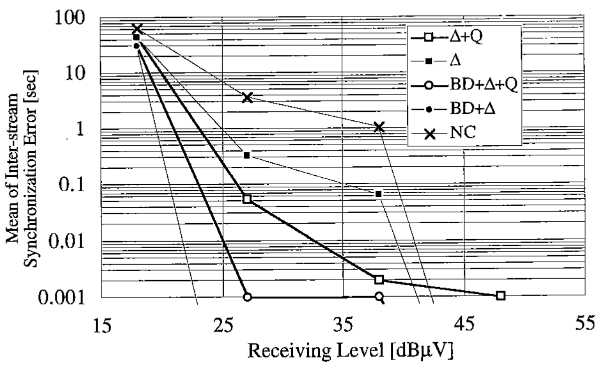


Fig. 7 Mean of $\Delta_n^{(1\text{-}2)}$ versus R_e in burst error environment.

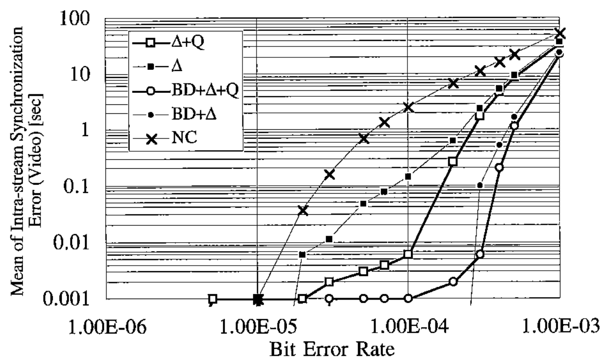


Fig. 9 Mean of $\Delta_n^{(1)}$ versus B_e in random error environment.

discarding control (i.e., "BD+Δ+Q" and "BD+Δ"). Note that although the mean of $\Delta_n^{(1)}$ of "BD+Δ+Q" is larger than that of "BD+Δ," the values for both "BD+Δ+Q" and "BD+Δ" are very small when $R_e \geqq 27$ dBμV. On the contrary, the intra-stream synchronization of the audio stream is disturbed to a lower degree because of the slide control scheme with the criterion $Q^{(1)}$.

4.1.3 The Coefficient of Variation $C_v^{(i)}$

We also examine the quality of the intra-stream synchronization in terms of the coefficient of variation $C_v^{(i)}$ for each stream. Figures 12 and 13 show $C_v^{(1)}$ versus B_e and $C_v^{(2)}$ versus B_e, respectively, in the random error environment. Also, Figs. 14 and 15 illustrate $C_v^{(1)}$ versus R_e and $C_v^{(2)}$ versus R_e, respectively, in the burst error environment; a smaller value of this means that we can get MUs more smoothly. Note that the cases of the B-block discarding control along with the slide control are not plotted in these figures, since the degree of the performance improvement by the B-block discarding control depends on the characteristics of the video stream.

As for the video stream in the random error environment (see Fig. 12), we find that the slide control scheme with the criterion $Q^{(1)}$ is effective in decreasing the coefficient of variation; $C_v^{(1)}$ of "Δ+Q" is smaller than that of "Δ," when $B_e \geqq 2.0 \times 10^{-5}$. For example, when $B_e = 1.0 \times 10^{-4}$, the coefficient of "Δ+Q" is 0.35, which is about half of the value of "Δ." Note that, when $B_e < 2.0 \times 10^{-5}$, $C_v^{(1)}$ of "Δ+Q" is larger than that of "Δ." This is because the slide control due

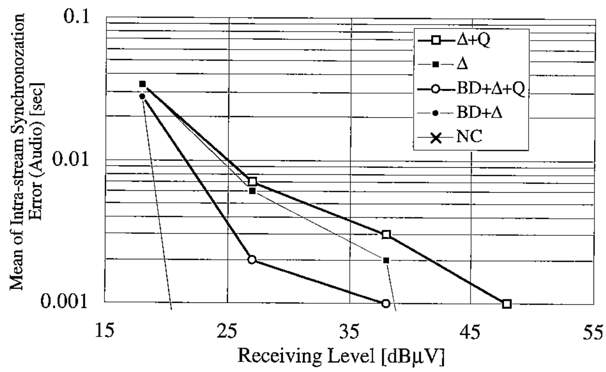


Fig. 10 Mean of $\Delta_n^{(2)}$ versus R_e in burst error environment.

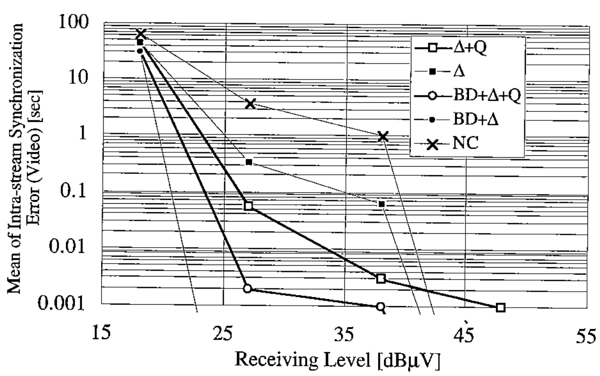


Fig. 11 Mean of $\Delta_n^{(1)}$ versus R_e in burst error environment.

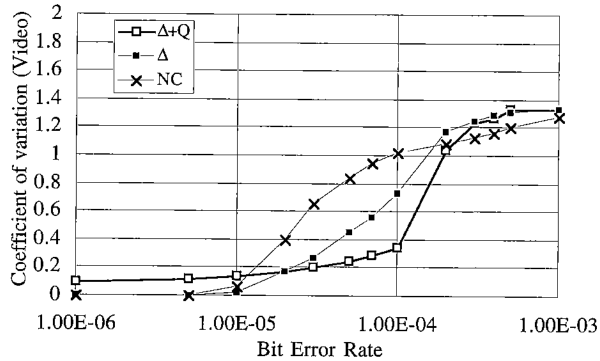


Fig. 12 $C_v^{(1)}$ versus B_e in random error environment.

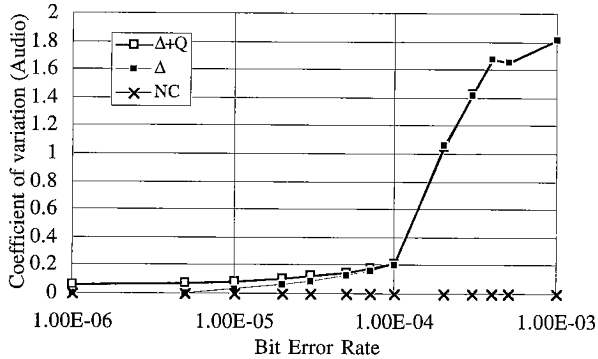


Fig. 13 $C_v^{(2)}$ versus B_e in random error environment.

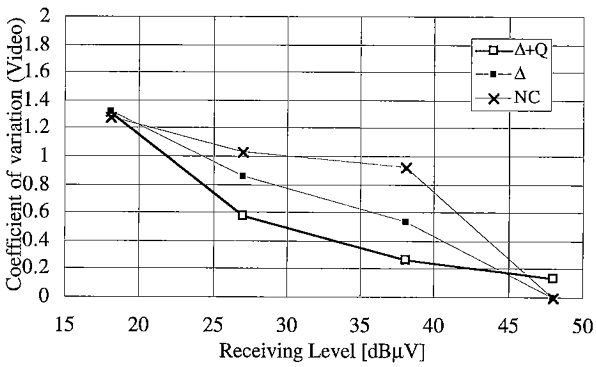


Fig. 14 $C_v^{(1)}$ versus R_e in burst error environment.

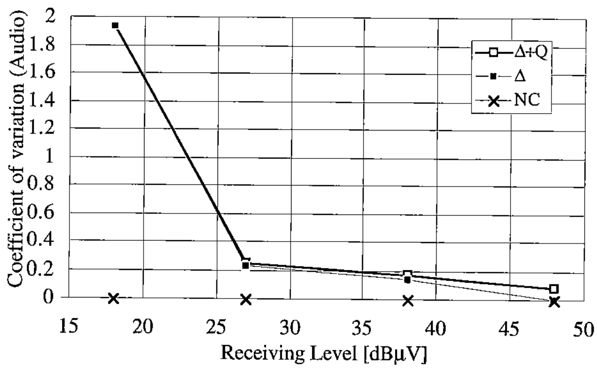


Fig. 15 $C_v^{(2)}$ versus R_e in burst error environment.

to the criterion $Q^{(1)}$ is often executed, even if B_e is small. However, each value is not large.

We also notice that $C_v^{(1)}$ of "Δ+Q" is less sensitive to B_e than that of "Δ," when $B_e \leqq 1.0 \times 10^{-4}$. This result means that "Δ+Q" can achieve the smooth output of video MUs even if B_e changes. This is because the slide control scheme with the criterion $Q^{(1)}$ is executed so as not to avoid low buffer occupancy. Actually, we have watched the pictures more smoothly, when $B_e \leqq 1.0 \times 10^{-4}$.

Note that, when $B_e \geqq 2.0 \times 10^{-4}$, both $C_v^{(1)}$ of "Δ+Q" and that of "Δ" are very large, since the fast recovery is made. Moreover, these values are larger than that of "NC."

As for the audio stream, we see in Fig. 13 that $C_v^{(2)}$ of "Δ+Q" is slightly larger than that of "Δ," when $B_e \leqq 1.0 \times 10^{-4}$. This is because the slide control due to the criterion $Q^{(1)}$ is often made, though B_e is small. However, each value is sufficiently small. When $B_e \geqq 2.0 \times 10^{-4}$, we notice that both $C_v^{(2)}$ of "Δ+Q" and that of "Δ" increase largely because of the fast recovery scheme. Note that the coefficient of "NC" is always zero.

In the burst error environment (see Figs. 14 and 15), we also observe that the slide control scheme with the criterion $Q^{(1)}$ is effective in decreasing the coefficient of variation for the video stream, while the coefficient of variation for the audio stream increases a little.

4.2 The Pause Time and its Effect on the Media Quality

We next examine the pause time caused by the slide control scheme with the criterion $Q^{(1)}$ and make subjective assessment of the media quality. Figure 16 shows the total slide time $S_N^{(1)}$; a large value of this obviously means that not only the total play time of the streams but also the amount of the pause time for MUs increase.

We find that the slide control caused by the criterion $Q^{(1)}$ is often executed, even if B_e is low; $S_N^{(1)}$ of "Δ+Q" is larger than that of "Δ." Especially, the difference between the two increases, as B_e becomes smaller. This result means that the criterion $Q^{(1)}$ causes the increase of both the total play time and the pause time for low values of B_e. However, the increment of the time is less than one second, which is sufficiently small compared to the original recording time, 122.333 sec. Note that large values of $S_N^{(1)}$ for $B_e \geqq 2.0 \times 10^{-4}$ are due to the fast recovery scheme.

Then, we examine the influence of the pause time on both the video stream and the audio stream by means of the subjective assessment. When the gradual recovery due to the criterion $Q^{(1)}$ is made, the output of a video MU and an audio MU pauses during θ_1, which is 100 ms. As for the video MU, we can see the previous video MU during the pause time, which is not too large. Therefore we did not notice the pause of the video MU when we watched the video stream actually.

On the other hand, when the output of an audio MU paused without repeating the previous audio MUs, we recognized the pause of the audio stream, since there was no sound during 100 ms. However, if the destination repeated the output of the previous audio MU, we did not notice the pause of the audio MU; we heard the audio stream with an echo. Therefore repeating the previous audio MU during the pause time is a good strategy from a subjective point of view, since the influence of the slide control upon the audio stream is reduced.

4.3 Appropriate Values of the Threshold T_{hQ}

We now consider how much the threshold T_{hQ} should be. For that purpose, we assess both the total slide time $S_N^{(1)}$ and the mean of $\Delta_n^{(1\text{-}2)}$.

We illustrate $S_N^{(1)}$ of "Δ+Q" versus α for various values of B_e in Fig. 17 and $S_N^{(1)}$ of "Δ+Q" for various values of R_e in Fig. 18, where α is defined as the minimum time to receive

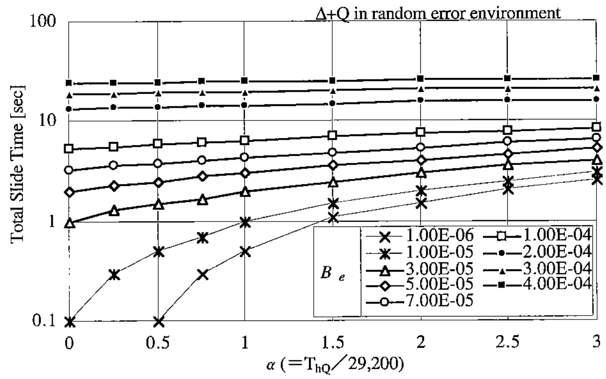


Fig. 17 Total slide time versus α in random error environment.

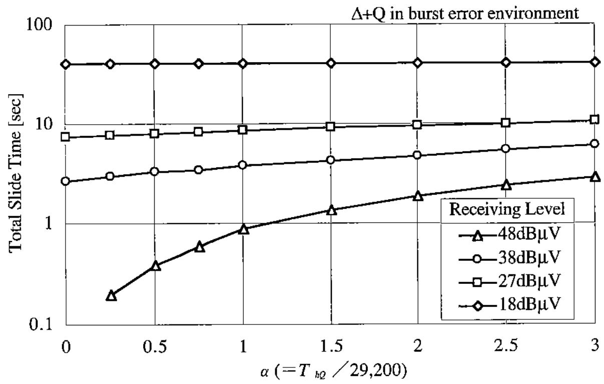


Fig. 18 Total slide time versus α in burst error environment.

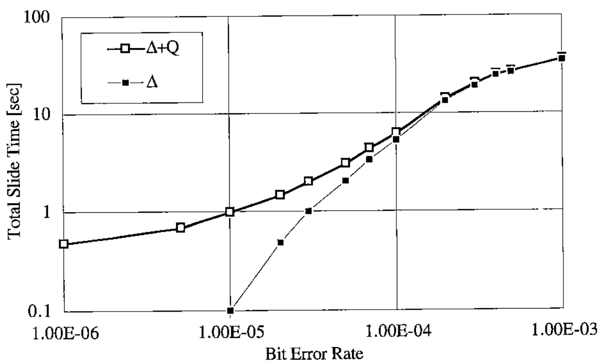


Fig. 16 Total slide time versus B_e.

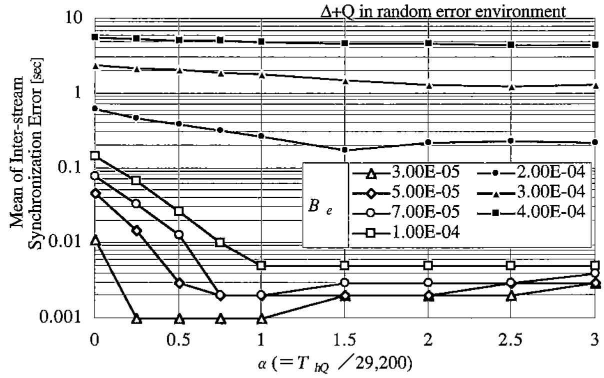


Fig. 19 Mean of $\Delta_n^{(1\text{-}2)}$ versus α in random error environment.

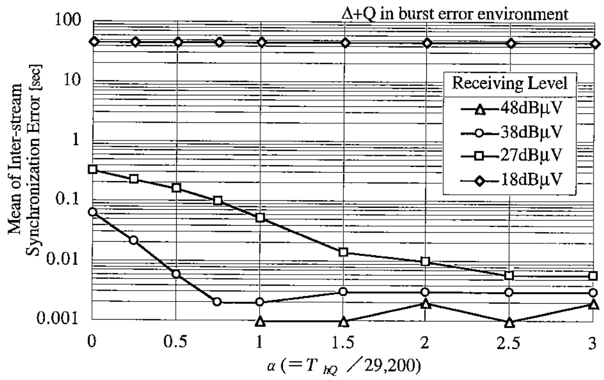


Fig. 20 Mean of $\Delta_n^{(1\text{-}2)}$ versus α in burst error environment.

the same amount of video data as the value of T_{hQ}. That is, we define $\alpha = T_{hQ}$ / 29,200, since the maximum video throughput is 29,200 bit/s. Note that the condition of α=0 is equivalent to "Δ." We notice in Fig. 17 that $S_N^{(1)}$ increases, as α becomes larger in the random error environment. Also, we obtain the same result in the burst error environment (see Fig. 18). Thus, it is desirable to select a smaller value of α in order to reduce $S_N^{(1)}$.

Figures 19 and 20 plot the mean of $\Delta_n^{(1\text{-}2)}$ of "Δ+Q" versus α in the random error and the burst error environments, respectively. In the random error environment (see Fig. 19), the mean of $\Delta_n^{(1\text{-}2)}$ increases, as α becomes smaller, while it is insensitive to α when $\alpha \geqq 1.0$. To keep the mean of $\Delta_n^{(1\text{-}2)}$ less than 80 ms, which is one of Steinmetz's results, we have to set $\alpha \geqq 0.25$ when $B_e \leqq 1.0 \times 10^{-4}$, for instance. To satisfy the same condition of $\Delta_n^{(1\text{-}2)}$ in the burst error environment (see Fig. 20), we must set $\alpha \geqq 1.0$ when the receiving level is 27 dBμV or more. From the above observations, we see that a value of α around 1.0 (i.e., $T_{hQ} \approx 29.2$ kbit) is an appropriate one under the condition that the state of the PHS channel is not so bad.

5. Conclusions

We proposed a new slide control scheme in which the quantity of video data $Q^{(1)}$ in the receive-buffer is used in addition to the intra-stream synchronization error of video to judge the modification of the target output time. We have evaluated the performance of the slide control scheme with the criterion $Q^{(1)}$ for stored video and audio streams which are transmitted over separate PHS channels.

We first observed that the slide control caused by the proposed criterion $Q^{(1)}$ is effective in the media synchronization. Again, we find that the B-block discarding control is effective. However, it is very important that the effectiveness due to the slide control is independent of the characteristics of a video stream unlike the B-block discarding control. Also, we confirmed that the quality of the media output under the control is often satisfactory to us from a subjective point of view. We further found that the threshold T_{hQ} should be approximately 29.2 kbit (i.e., $\alpha \approx 1.0$) for the PHS channel whose state is not so bad.

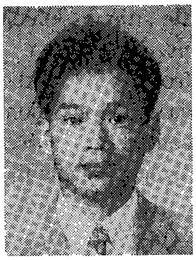
Our future work includes systematic studies about the criterion for the quality of the media synchronization in a small sized video communication. Also, we are planning to make an experiment of media synchronization schemes for PHS including the one proposed in this paper.

Acknowledgment

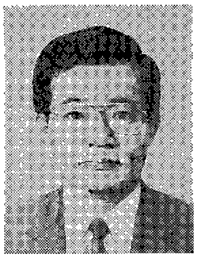
The work of Tasaka was supported by *Telecommunications Advancement Organization of Japan*.

References

[1] Association of Radio Industries and Businesses (ARIB), "Personal handy phone system ARIB standard version 2 (Rev.-2) RCR STD-28," June 1996.
[2] ITU-T Recommendation H.263, "Video coding for low bit rate communication, " March 1996.
[3] PHS Internet Access Forum (PIAF), "PHS internet access forum standard (PIAFS)," March 1997.
[4] G.Blakowski and R.Steinmetz, "A media synchronization survey: Reference model, specification, and case studies," IEEE J. Select. Areas Commun., vol.14, no.1, pp.5-35, Jan. 1996.
[5] M.Kato, N.Usui, and S.Tasaka, "Performance evaluation of stored media synchronization in PHS," IEICE Trans., vol.J80-B-II, no.9, pp.749-759, Sept. 1997.
[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] Y.Ishibashi, E.Minami, and S.Tasaka, "Performance measurement of a stored media synchronization mechanism: Graceful recovery scheme," IEICE Trans. Commun., vol.E79-B, no.3, pp.399-411, March 1996.
[8] R.W.Lucky, J.Salz, and E.J.Weldon, "Principles of data communication," McGraw-Hill, 1968.
[9] M.Kato, N.Usui, and S.Tasaka, "Performance evaluation of a video transmission scheme for PHS," IEICE Trans., vol.J79-B-II, no.10, pp.646-656, Oct. 1996.
[10] T.Fujimoto, J.Yamaguchi, T.Kobayashi, and T.Hamaki, "A study of a data transmission method for personal handy phone system," IEICE Technical Report, CS94-77, Aug. 1994.
[11] Source program code of "TMN H.263 Encoder ver.1.6c & Decoder ver.1.6c," Telenor R&D, March 1996.
[12] A.Tsuji, Y.Ishibashi, and S.Tasaka, "An inter-destination synchronization mechanism for live media in multicast communications," IEICE Technical Report, IN96-130, Feb. 1997.
[13] R.Steinmetz, "Human perception of jitter and media synchronization," IEEE J. Select. 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.



Noriyoshi Usui received the B.E. degree in electrical engineering from Gifu University, Gifu, Japan, in 1971. He joined Sanyo Electric Co., Ltd. in 1971, and engaged in the development of telecommunication systems including digital private automatic branch exchanges and public base stations of PHS. He is presently a general manager in Information & Telecommunications Development Center of Sanyo.

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.