INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Appendix I

Rec. G.764

(11/95)

GENERAL ASPECTS OF DIGITAL TRANSMISSION SYSTEMS

PACKETIZATION GUIDE

Appendix I to

ITU-T Recommendation G.764

(Previously "CCITT Recommendation")

FOREWORD

The ITU-T (Telecommunication Standardization Sector) is a permanent organ of the International Telecommunication Union (ITU). The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

Appendix I to ITU-T Recommendation G.764 was prepared by ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 13th of November 1995.

NOTE

In this Appendix, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

CONTENTS

Packetization Guide

(Geneva, 1995)

(This appendix does not form an integral part of this Recommendation)

I.1 Introduction

This appendix summarizes the current views on speech packetization in Study Group 15 of the ITU-T during the study period 1992-1996 and in CCITT SG XV during the study period 1988-1992. These views may change in the future.

This appendix does not address the following topics:

1) the different treatments within the network that may be accorded to packets depending on their assignment to priority classes, in general; however, it is agreed that the network should give priority to speech over digital data to reduce delay and delay variability, and to decouple bursty traffic from real-time traffic;

2) quality of service for different traffic classes.

The purpose of this appendix is to:

- explain the various issues that affect the packetization of speech;

- provide an overview of the techniques and considerations in the transport of packetized speech, that are used in Recommendation G.764;

- disseminate information on the various topics of concern to designers, implementors of packetized speech equipment and to the service providers that use them.

I.2 Historical background

Traditionally voice services have been implemented in the Public Switched Telephone Network (PSTN) (also denoted as Wide Area Network or WAN) using a circuit-oriented approach. The growth of packet transport techniques [e.g. Recommendation X.25/X.75, Internet, Wideband Packet Technology, Frame Relay and the Asynchronous Transfer Mode (ATM)] has stimulated research in new techniques for the transport of speech.

Packetized systems can exploit the bursty nature of traffic to multiplex different types of traffic (e.g. voice, data, video) of many users so that they can share transmission bandwidth and switching resources dynamically. Packetization facilitates the integration of the different types of traffic to allow more efficient utilization of the available bandwidth and switching resources. Packetization offers more flexibility than circuit-oriented approaches because the packet header contains the necessary control information that identifies, for example, the type of traffic and, where appropriate, the coding scheme.

Work in speech packetization began in the CCITT since the middle of the 1984-1988 study period in Working Party XVIII/8 and continued in Working Party XV/2 during the 1989-1992 study period. It is now continuing, in the ITU-T for wideband packet and ATM networks. The goal is to provide a uniform basis for speech packetization, with or without speech compression and speech interpolation, to facilitate the interworking of equipment from various vendors in telecommunications applications.

The work in the CCITT has lead to the Voice Packetization Protocol of Recommendation G.764 and its extensions in Recommendation G.765. These protocols are compatible at the link layer with the ISDN protocols LAPD and LAPF specified in Recommendations Q.921 and Q.922, respectively.

I.3 Summary of design issues

Using the OSI protocol stack as a reference, the major design considerations in devising a speech packetization protocol are:

1) Layer I (physical layer) – The issue is whether the physical interface will conform to that of public telephone networks (Recommendation G.703/G.704) or to other local area networks, such as IEEE 802.2, 802.3 or 802.9, etc.

2) Layer 2 (link layer) – Some of the issues are:

   a) whether the logical layer will be compatible with ISDN (LAPD/LAPF) protocols or will have the same structure as those for LANs;

   b) how to deal with the loss of frames;

   c) robustness to errors.

3) Layer 3 (procedures to deal with digitized voice and voiceband data traffic) – Issues are:

   a) the delay variability for speech packets;

   b) the transport of channel-associated signalling.

4) Higher-layer issues involve the speech coder and the type of compression used.

A speech packetization protocol has the following requirements:

- The speech must be reconstructed at the receiving end from packets arriving at irregular intervals (or, in some architectures, out of order).
- The protocol must be robust against line errors.
- It must offer an easy method for congestion control in the network.
- It must specify procedures at the terminating end to recover from packet loss or excessive delay.
- It must carry channel-associated signalling.
- If digital speech interpolation is used to eliminate silence intervals, it shall specify the level at which noise is re-injected at the terminating end.

I.4 Reconstitution of speech signals

To achieve good speech quality, the terminating end must reconstitute a continuous speech stream and play it out at regular intervals despite varying packet arrival times. This involves two aspects:

1) preserving the relative timing of information within one speech burst; and

2) delay equalization.

In Recommendation G.764, a packet sequence number is used to encode the relative timing of the information within one speech burst. The first packet of a speech burst always has the sequence number of 0: subsequent packets in the same burst have the numbers from 1 to 15, rolling back to 1. The terminating endpoints use the packet sequence number to:

1) determine the first packet of a speech burst; and

2) to detect packet loss. The determination of the first packet is useful for delay equalization and may be needed for some speech coding algorithms such as Recommendation G.728. Delay equalization is discussed in the next subclause.

I.5 Delay equalization

Delays in packet communication consist of two components: a fixed delay and a variable delay [1]. The fixed delay arises from signal propagation on the transmission links, and from fixed processing delays in the originating and terminating endpoints and within the network. The effects of variations in the propagation delay for a given path are assumed to be negligible.

For speech packetization, the fixed processing delays consist of the following components:

1) packetization delay during which the speech samples are buffered for further processing;

2) if digital speech interpolation is used to remove silent intervals, the hang-over time of the speech detector [2];

3) the end-to-end algorithmic delay due to the encoding and decoding of speech – this delay depends on the coding scheme; for example, it is 125 μs for PCM, 250 μs for the Adaptive Differential Pulse Coded Modulation (ADPCM) algorithms of Recommendations G.726 and G.727 while it is less than 2 ms for the Low-Delay Code Excited Linear Predictor (LD-CELP) algorithm of Recomendation G.728;

4) any added delay at the terminating end to mask the timing jitter resulting from the variability in the delay – this added delay is denoted as build-out.

Variable delays result primarily from the queueing and processing of packets. They depend on the characteristics of the route of each packet: the number of hops (nodes), the type and speed of each link and the traffic intensity.

Speech traffic requires low and uniform delay. Recommendation G. 114 (1993) discusses the effect of end-to-end delays on the quality of a conversation. Delay variations that may, be acceptable for digital data transmission, usually affect the gaps between words and syllables and are troublesome for conversational speech. Available data suggest that variation in the interval between speech bursts should be less than 200 ms to avoid subjective degradation of speech quality. In applications such as videotelephony where both audio and video information are transmitted and must remain synchronized, the effect of the variable delay on this synchronization must also be taken into account.

There are several methods to mask the variability of the delay in the network. These techniques include:

1) blind delay;

2) absolute time stamp; and

3) relative time stamp.

The effect of all these methods is to increase the effective end-to-end buffering delay and, therefore, the total end-to-end delay.

I.5.1 Blind delay

In the blind delay method, a fixed buffer delay is always added at the terminating end at the first packet of a speech burst. This delay corresponds to the maximum variable delay expected.

The advantage of the blind delay scheme is its simplicity which makes it a good candidate when the transmission speech is such that the variable delays are of the order of a fraction of a millisecond (e.g. local area networks or broadband networks at 150 Mbit/s). In these cases, a fixed build-out delay on the order of 10 ms will be adequate to eliminate end-to-end delay jitter [3].

Over long haul connections in the PSTN, the scheme may require too large a delay that the total end-to-end delay would exceed the performance limits for network delays specified in Recommendation G. 114. For example, if the first packet has already experienced the worst case delay variation, the total variable delay to be added will be twice the worst case value [1]. Because this approach does not mask the delay variability completely, the gaps between the words may be varying which may degrade the subjective quality of speech. Voiceband traffic, included demodulated facsimile, may be perturbed; total delays larger than 500 ms cause premature disconnections of facsimile calls, especially when echo is present [4].

I.5.2 Absolute time stamp

This is the method used in datagram networks. The packet header includes a field for a time stamp that represents real time. The time stamp has a resolution sufficient to allow accurate detection of packet jitter, and to cover the worst case transit time of a packet across the network. Thus, packets that arrive out of sequence may be correctly ordered and buffered at the receiver using time stamp information [5]. This datagram scheme also requires clock synchronization between the transmitter and receiver so that the delay of each incoming packet can be compared to the previous ones, assuming that the network delay is fixed.

I.5.3 Relative time stamp

In the relative time stamp method, an estimate of the play-out time is obtained for the first packet of a speech burst, for all signalling packets and for the first packet after a missing packet. This time is then used to adjust the delay of all remaining packets of this burst conveyed on that virtual circuit.

The accumulated variable delay experienced by a packet is recorded in the time stamp field of the packet header [1]. Each network node adds to the time stamp the amount of time it took to serve a packet before sending it, using its local clock as reference.

The maximum allowable variable delay for a virtual circuit is specified as the build-out. The build-out is defined for a given virtual circuit. Once an estimate of the play-out time has been made, subsequent packets are placed in the order of their sequence number in the play-out buffer and then held for the following duration:

time before play-out = build-out delay – time stamp value

The terminating endpoint must, therefore, store the speech packets that arrive before their scheduled play-out time and then play them at regular intervals. Packets whose time stamp field exceeds the build-out delay are considered late and are dropped.

The relative time stamp method is less complicated than the absolute time stamp method of I.5.2 when virtual circuits are used to guarantee that the packets remain in sequence. The scheme does not rely on clock synchronization between the endpoints and places the timing function in layer 3. Furthermore, the delay measurement in each packet can be used to detect network congestion and then to invoke overload management strategies.

In a multisite conference, the introduction of build-out delay improves the conversational dynamics because it ensures that the playback of the speech to all parties is synchronized.

One disadvantage of the relative time stamp method is the added complexity for network nodes. Each network node must distinguish between voice and data packets, update the time stamp field and recalculate the CRC. Another disadvantage is that the end-to-end delay is always increased by the build-out even in the absence of network congestion. Finally, compared to the fixed delay method, the time stamp field adds some protocol overhead.

In general, the selection of the value for build-out is a trade-off between accepting excessively large delays and dropping many packets. Because a virtual circuit emulates the physical circuit connection, the build-out delay is typically higher for digital data traffic than for voiceband traffic. In a point-to-point connection, simulation studies have shown that, at the primary rate, a build-out delay of 10 ms will ensure that less than one tenth of one percent of speech packets are dropped in a given node for excessive delay even for data loads reaching 60 to 70 percent of the offered bandwidth [6]. To allow some margin of safety, the build-out delay can be set around 20-30 ms per hop. More conservative figures (e.g. 70 ms) may be used, especially when facsimile demodulation/remodulation is in the path to account for the processing time for facsimile demodulation and for the timing constraints imposed by the T.30 protocol.

The build-out delay is added only once at the terminating endpoint; thus, if two G.764 networks are connected back-to-back, the delay is added at the last node only. This property assists the network planner/operator in maintaining the overall one-way delay within the limit of 400 ms of Recommendation G.114.

Recommendation G.764 sets the maximum value of the build-out at 200 ms to accommodate both voiceband traffic and digital data traffic. Satellite links introduce a one-way delay of 250 ms, while terrestrial extensions add delays of about 50 ms. Therefore, when a single-hop satellite link is used, a build-out around 50 ms can give a total one-way delay of about 350 ms with a very low percentage of packet loss. The build-out for speech is usually limited to about 70 ms for a path of three or more nodes or when facsimile demodulation of Recommendation G.765 is used.

I.6 Robustness to errors

Robustness to errors involve two aspects:

1) robustness of the speech coder; and

2) robustness of the protocol.

In this subclause, the issue of how to make the protocol more robust to errors on the line is discussed. In particular, it is explained why it is better to calculate the CRC over a part of the frame rather than on the whole frame. In this way, the frame would not be dropped if a bit error occurs in the unprotected field. It should be noted that new coding algorithms may be more robust to burst errors (i.e. frame erasure or packet loss) than PCM or ADPCM.

The use of unrestricted 64 kbit/s packetized channels has to be cautiously addressed for a class of narrow-band ISDN services such as videotelephones in the presence of line errors. Packet system transform random errors into burst of errors [7].

I.6.1 Sensitivity of speech to bit errors

Line errors are a source of packet loss (also called *frame erasure* in wireless communication). Speech traffic is more sensitive to packet loss than to bit errors [8], [9]. Therefore, it is better to deliver a speech packet with some corrupted bits than to discard it, or attempt to recover the original bits by retransmission. This applies to the case of PCM speech as well as compressed speech if a bit error in a speech sample has low impact on speech quality. This fact is now well established for speech carried in HDLC [10] and for TCP/IP networks [11].

I.6.2 CRC calculation over part of the frame

Recommendation G.764 has introduced a new HDLC frame, called “Unnumbered Information” with a Header check or UIH, to restrict calculation of the CRC to the address field (to ensure correct delivery), the control field (to guarantee the validity of the frame type), and the layer 3 header. ISO has generalized the scope of applications to all cases where the timely delivery of the information is more important than the integrity of the information being transferred (e.g. packetized speech or the transfer of periodically updated information or video applications) [12]. The length of the protected bits is determined by negotiation as a system parameter through the exchange of XID frames at link establishment or by Administration.

I.6.3 CRC calculation over whole frames

Some proprietary speech packetization implementations have preferred the use of the traditional method of calculating the CRC over the whole frame. This provides minimal impact on existing HDLC devices that do not have the capability of performing the CRC over part of the frame.

The disadvantage of having the CRC over the whole frame is that any bit error will cause the entire frame to drop. In this case, the application of packetized speech becomes restricted to cases where the error rate is very low. This restriction would exclude the application of the protocol from many public and private networks worldwide, unless some degradation is accepted.

I.7 Congestion control

Congestion may be caused by the statistical nature of the traffic or from actions that the network has initiated. The most commonly used method of speech flow control is that of “call blocking”, i.e. preventing the initiation of new calls during busy periods. Data-oriented flow control mechanisms do not suit voice networks, because the speakers cannot be forced to slow down their speech. Also, long gaps or variable delays due to data-oriented flow control actions may decrease speech intelligibility [13], [14].

In integrated packet networks, there may be two types of congestion control techniques:

- Global control – The congestion is abated by system-wide procedures.

- Local control – The congested node relieves the congestion locally by reducing the offered load either by shortening the length of the packets or by reducing their numbers. The nature of the treatment depends on whether the packet length is variable or fixed (such as in ATM systems) and whether the speech coder can sustain block losses or packet losses.

I.7.1 Global control

This method prevents the establishment of new calls. In circuit multiplication equipment, irrespective of the technology used, this may be achieved through communication with the gateway switch according to the rules of "dynamic load control" defined in Annex A/Q.50 or Annex B/Q.50. In private networks, the interface between a PBX and the packetization equipment can be used to affect that call control.

I.7.2 Local control

Local congestion control can be achieved either by dropping elements of a packet, called "blocks", or alternatively whole packets. The effect of dropping blocks from a packet is to reduce the packet length while the effect of dropping packets is to decrease the number of packets to be transmitted. The resultant shortening of the speech buffers frees some bandwidth for non-compressible traffic through a controlled degradation of speech quality [6]. Thus, local control depends on the ability of the speech coder to operate with less bits and/or to sustain packet loss.

Local control allows each node to function independently, so that each node does not have to coordinate with other nodes in the backward direction. This allows an instantaneous response to traffic congestion at any node that is adapted to the characteristics of the traffic type without explicit exchanges of control messages among all the nodes in the path of the virtual circuits.

I.7.2.1 Bit-droppable and embedded algorithms

Bit-droppable algorithms are those algorithms whose code words consist of a core bits and several enhancement bits. The core bits are used both in the encoder and the decoder, while the enhancement bits are used to reduce the quantization noise in the reconstructed signal. Core bits must reach the decoder to avoid mistracking, but the enhancement bits can be dropped thereby shortening the packets to alleviate congestion.

Bit-droppable schemes provide a flexible way to alleviate congestion at any node of the network without the need for exchanging control messages in the backward path of the connection (towards the originating end). There are currently two standard bit-droppable algorithms, PCM of Recommendation G.711 and ADPCM of Recommendation G.727. Recommendation G.727 is also called an embedded algorithm because there is a feedback loop tying the encoder with the decoder. In this case, the coder is constructed such that the levels of the quantizer at the higher rates include the levels of the quantizer at the lower rates. Therefore, enhancement bits can be dropped when there is local congestion, without changing the quantization table [15].

Several recent proposals have shown how embedded coding can be achieved for CELP [16], [17], [18], [19], [20], [21]. An initial version of an embedded LD-CELP coder derived from the G.728 operates at 8 kbit/s, 14 kbit/s and 20 kbit/s and can achieve toll quality at its higher operating rate [22]. This coder has a total processing delay of approximately 12-18 ms and is robust to channel errors at an average bit-error rate of 10^{-4}.

The drawback of embedded coding schemes is that they introduce some degradation in speech quality compared to the non-embedded algorithm operating at the same bit rate. However, with good coder design, the difference in subjective quality between an embedded coder and the corresponding fixed rate coder can be made imperceptible (e.g. Recommendations G.727 and G.726 both operating with 4 bits/sample). Another drawback is that, as the bit rate decreases, the efficiency of transmission diminishes, because the overhead remains constant while some user's bits are discarded.

I.7.2.1.1 Block dropping schemes

The technique of block dropping is ideally suited for use with "embedded" speech coders because they assign different priorities to the various bits of the encoded bitstream. In the process of packetization, the bits of various priorities can then be assembled in separate "blocks" so that network nodes can shorten packets by deleting certain predefined blocks to alleviate congestion (with minimal impact to speech quality).

With the ADPCM embedded coding algorithm of Recommendation G.727, Recommendation G.764 specifies an arrangement of the samples such that the most significant bits of all samples fall in the same block, followed by the next most significant bits and so on. Thus, each block contains bits of the same significance for the 128 samples. Block 1 contains all the most significant bits, block 2 the next significant bits, etc. The core block contains sufficient information for decoding the speech, although full quality is achieved only when both the core and enhancement blocks are used.

A Block Dropping Indicator in the packet header indicates the maximum number of blocks that can be discarded and the number of blocks that are still droppable. Block dropping can occur at any node which allows intermediate nodes to reduce the bit rate and alleviate congestion by dropping bits from the speech packets *without* having to communicate with the originating node.

Simulation and field results have demonstrated that, with ADPCM coding, the block scheme offers a more graceful degradation under overload than with packet dropping [23], [24], [25], [26], [27], [28], [7]. When the mean bit per sample decreases to 3.97 bits/sample, the MOS value drops by about 0.36 from its value for PCM. For the same compression ratio, the effect of packet dropping is more drastic; the MOS value drops by about 1.36, with 0.54 percent packet loss. At a packet loss rate of 2.2 percent or 5.3 percent the difference between the block dropping and packet dropping schemes is 1.4 and about 2 MOS respectively.

The advantages of the Block Dropping Scheme are that it allows fast and decentralized decision. Also, the reduction in speech quality is gradual.

The disadvantages of the Block Dropping Scheme are:

1) increase in the complexity of the network nodes;

2) there is only one standard embedded algorithm;

3) the method might not be effective in at low bit rates because the packet overhead may occupy a significant portion of the bandwidth relative to the payload (unless several calls or several types of traffic are multiplexed in the same frame).

I.7.2.1.2 Selective packet dropping schemes

The scheme relies on the capability of dropping marked frames (packets) or cells. The idea is to divide frames (packets) into high priority frames (packets) and low priority frames (packets). High priority frames (packets) contain the most significant bits of the code word (or core bits) while low priority frames (packets) contain the less significant bits (or enhancement bits) [13], [29], [30], [31]. The enhancement blocks are put in the low priority frames (packets) while the core blocks are in the high priority frames (packets) so that the enhancement blocks can be dropped when congestion occurs.

Proposals have been made for the transport of ADPCM coded speech over ATM or over data-only frame relay networks. For example, for data-only frame relay, the “DE” bit is used for tagging the droppable frames [32].

The main advantages of the selective packet dropping scheme are:

1) Simplicity and effectiveness in handling congestion in the network.

2) Compatibility with the methods used in the data-only frame relay networks.

3) Relative ease of implementation.

With the existing algorithm, the scheme suffers from several disadvantages such as:

- Two levels of congestion reduction are considered, while there are at least three levels with Recommendation G.764. In other words, the coding rate may fluctuate between two extremes, for example between 32 kbit/s and 16 kbit/s for the (4,2) embedded ADPCM coding of Recommendation G.727, i.e. with 4 bits for the code, divided into 2 core bits and 2 enhancement bits. In contrast, Recommendation G.764 allows more graceful degradation with the coding rate dropping from 32 kbit/s to 24 kbit/s to 16 kbit/s. Field data have shown that operation until an average bit rate of 3.0 bit/sample is very satisfactory [7].

- More delay and processing may be introduced because the speech samples are sorted in different packets.
- More transmission bandwidth is used for the packet overhead.
- More importantly, there is no guarantee with this scheme that the core blocks and enhancement blocks remain synchronized. For example, if the core blocks are dropped due to a CRC failure, the enhancement blocks may still arrive at the decoder. For ADPCM, this may cause the decoder to be out of step with the encoder since the ADPCM prediction relies on the core blocks; and the speech quality will degrade substantially until resynchronization occurs. Recommendation G.764 ensures that no enhancement blocks arrive without their core blocks.
- Finally, delayed packets are discarded in Recommendation G.764 while in the approach of Yin, *et all.*, [29], high priority packets are transmitted irrespective of the congestion condition while low priority packets are discarded.

I.7.2.2 Class-oriented schemes

In these proposals, [33] and [34], speech traffic is classified according to models of speech production (plosives, fricatives, voiced speech, etc.). Each class is encoded with a different algorithm and is assigned a different delivery priority. The objective is to adjust the optimal voice treatment with the characteristics of each segment. In contrast, Recommendation G.764 always encodes the speech at the maximum rate that is specified in the packet header and leaves the packet priority as an internal mechanism of the network.

I.8 Packet loss

As explained before, there are several causes for packet loss (frame erasures):

1) transmission impairments;

2) excessive delay; and

3) congestion.

The effects of packet loss depend on the traffic type as explained in the following subclauses.

I.8.1 Case of speech

Packet loss is perceived as gaps may cause a severe degradation in performance. In the past, CCITT algorithms have not been explicitly designed for robustness to packet loss. For example, with PCM and ADPCM, if the percentage of dropped frames exceeds one percent, speech quality can be significantly impaired [9], [10], [35]. In these cases, packet dropping does not produce significant impairments provided that the percentage packet loss is less than one percent.

Some non-standard ADPCM algorithms have been designed to increase their robustness for indoor wireless transmission [36]. Also, one of the requirements of the ITU coder at 8 kbit/s is to sustain about three percent frame erasures (or packet loss) for the radio channel.

Furthermore, in March 1994, AT&T presented modifications to the G.718 decoder to make it suitable for handling frame erasures in the received bitstream [37]. The idea is to extrapolate some parameters when a “frame is erased” (i.e. a packet is lost). Special and simplified decoding rules take effect only when there is packet loss, thereby assuring total compatibility with Recommendation G.728 during normal operation. While the complexity of the decoder is not increased, robustness to packet loss (frame erasures) is substantially improved, as shown by simulated results using bursty errors obtained with the tools of the CCITT/ITU software tools library. Some of these results are shown in Table I.1 for packets corresponding to 10 ms.

TABLE I.1/G.764

MOS of the robust LD-CELP

Coder	Packet loss rate	MOS	
		IRS speech	Non-IRS speech
32 kbit/s ADPCM	0%	3.90	3.78
G.728	0%	3.90	3.79
Extended G.728	1% 3%	3.95 3.77	3.82 3.59

These results demonstrate that there are ways of successfully concealing the effects of packet loss that would give no MOS degradation at one percent packet loss and a degradation of about 0.2 on the MOS scale for three percent packet loss. This is in opposition to the properties of ADPCM where packet loss had to be avoided at any cost. It seems, therefore, that Recommendation G.728 can be extended to tolerate more packet dropping (although uncontrolled packet dropping is to be avoided). The question of how to take advantage of these characteristics to simplify the design of the speech packetization protocols is an item for future study.

I.8.2 Case or voiceband data

Voiceband data are more susceptible to random bit errors than speech [7]. In the case of voiceband data, when a packet is dropped, the receiving modem may go through a recovery procedure if it detects a loss of carrier. The carrier-loss-disconnect timer in a typical modem ranges from 100 ms to about 20 s. If this timer expires, the modem usually disconnects and drops the call.

If the modem is connected to a network, the result will also depend on the internal procedures that the network implements to recover from carrier loss. For example, if the network reacts prematurely to a loss-of-carrier indication from the modem, it may disconnect the virtual circuit while the modem is undergoing the recovery procedure.

Depending on the modem characteristics, even a single bit error in the header may start this chain reaction. This is clearly one disadvantage of packet mode transmission compared to circuit mode transmission when random errors are significant. In contrast, for bursty errors, the packet mode is superior to the circuit mode.

I.8.3 Fill-in strategies (speech)

A number of fill-in strategies have been proposed to deal with gaps in the received speech due to dropped or lost packets:

1) Replace the discarded packet with samples with zero-amplitude values [35]. This is denoted in the literature as *silence substitution.*

2) Fill the gaps with injected noise called *comfort noise.* The method is called *noise interpolation k*.

3) Compensate for dropped packets through interpolation – this method can be used with the extension of Recommendation G.728 that AT&T has proposed [37].

4) Repeat the last packet received – if the speech is classified in several classes, and each class is sent in a different packet, the repetition algorithm may vary with the class of the previous packet and with the sequence number [33], [34].

5) Extract packet-length segments from the last few received packets (or from those preceding and following the lost packet) that correspond to the few milliseconds before the detection of packet loss – these segments are then used to replace the missing segment *(pattern matching)* [38], [39].

6) Repeat the last pitch waveform for voiced segments, otherwise use the previous packet [38].

Silence substitution or noise interpolation are simple techniques. Packet repetition requires less memory and signal processing than pitch waveform replication and pattern matching. These latter methods have the advantage of maintaining phase continuity at the boundaries between packets.

I.9 Choice or packet size

Several factors control the choice of the packet size: packetization delay, robustness to errors, end-to-end delay, jitter, overhead, etc. When several traffic types are considered, the requirements for each type of traffic may be in opposition to the others.

I.9.1 Speech considerations

Packets with no more than 50 ms of speech provide robustness to packet loss [1], [5] (the optimal length is in the range of 16-32 ms) [8], [35], [40] and the larger the packet the smaller the overhead. In Recommendation G.764, the input speech samples are collected over a period of 16 ms which corresponds to 128 samples at a sampling rate of 8 kHz.

For a given packetization interval, the actual packet size in octets will depend on the bit rate of the coding algorithm; with embedded coding, the size will adapt to time-varying network conditions. In ATM systems, a larger overhead is allowed and the speech samples correspond to 6 ms stored in 47 or 48 octets.

With low bit rate voice coders, 8 kbit/s for example, the size should be larger than for Recommendation G.764 or in ATM systems to keep the amount of overhead within reasonable bounds.

I.9.2 Bit error considerations

As per Recommendation G.826 and ANSI/IEEE C37.1-1979 [41], bit error ratios may exceed 10^{-4} for short intervals. By taking into account this factor in the computation of the residual error probability and considering the transparency mechanism of HDLC through bit insertion and flag synchronization, it is possible to show that the frame length should not exceed 500 octets if the residual error probability is less than 10^{-6} with bit error rates of 10^{-4} [42].

I.9.3 Integrated traffic

Long frames take a longer time to transmit than short frames. Therefore, when long and short frames are mixed, the short frames will experience a longer average delay (for a given link utilization) than if no long frames were present [43], [44]. This happens because long frames take a longer time to be transmitted than short frames. This is illustrated in the following simulation that presents two different scenarios for mixing frames of differing sizes in various proportions [43].

In the first case, half of the frames have a length of 74 octets and the length of the remaining half is 256 octets. For a link utilization of 90 percent, the average waiting queue length is 742.5 octets and the 0.1 percent percentile delay limit is 7226.536 octets.

If the traffic mix changes to the following profile:

Frame length (octets)	Proportion
74	40%
256	20%
512	20%
1024	10%
1500	5%
2048	5%

then, for the same link utilization, the average waiting queue length becomes 2083 octets (i.e. about three times larger) and the 0.1 percent delay limit becomes 36171.875 (i.e about five times larger). Clearly, the smaller frames are severely delayed in the presence of very large frames. While large delays may be tolerable for some types of traffic, speech (and to a lesser degree facsimile) has more stringent requirements on the variability of the delay.

I.9.4 Recommendation G.764

Recommendacion G.764 specifies that the maximum information field be restricted to 490 octets. Therefore, larger frames such as those allowed in Recommendation Q.922 have to be segmented at the originating endpoint and reassembled at the terminating endpoint.

I.10 Compression issues

A typical digital transmission system codes the speech twice. The first encoding converts the source speech from the analog into the digital domain using a special coding scheme that may involve compression by removing redundancy. The second encoding is used to ensure that the coded binary information satisfies the constraints that the transmission scheme imposes, for example ensuring ones density. Thus, there are two general approaches for efficient use of the available transmission bandwidth:

1) to compress the source signal using an appropriate speech coding algorithm by removing redundancy in the source material; and

2) to remove idle periods, i.e. to remove the silence intervals from the speech or at least code them at a lower bit rate than the bit rate used for speech.

I.10.1 Speech coding algorithms

The speech traffic can be coded by one of various standard coding algorithms. There are several CCITT/ITU recommended coding schemes for narrow-band telephony that provide toll quality: PCM of Recommendation G.711 (64 kbit/s), ADPCM of Recommendation G.726 (at 40 kbit/s and 32 kbit/s), 4-bit encoding algorithms of Recommendation G.727 (i.e. at 32 kbit/s) and Recommendation G.728 (16 kbit/s). Each one of these algorithms introduce a different end-to-end algorithmic delay. As mentioned earlier, the end-to-end algorithmic delay of the G.726 and G.727 algorithms is 250 μs while that of G.728 is less than 2 ms.

Recommendation G.764 accommodates PCM and ADPCM algorithms, while the initial implementation of ATM is likely to favor PCM systems (without digital speech interpolation).

I.10.2 Digital speech interpolation

Further reduction in the bit rate needed for speech transmission can be accomplished by removing idle periods, i.e. removing silent intervals from speech or by coding the silence at a lower bit rate than that used for speech. Time Assignment Speech Interpolation (TASI) was originally used for undersea analog cable systems and then applied to digital satellite transmission systems. Later on, digital speech interpolation was combined with variable rate ADPCM coding [45] to increase the efficient use of the transmission path.

To recognize that speech is being transmitted, a highly sensitive speech detector is required [2], [46]. Speech detection is based on several measurements such as short-time energy, zero-crossing rates and sign-bit sequences. Depending on the design, a hang-over time is used to extend the speech interval and avoid speech clipping. This extension obviously reduces the gain in bandwidth; therefore, it is usually recommended that the speech detector operates such than the speech activity it declares does not exceed the actual speech activity by more that five percent. For example, if the actual speech activity is 38 percent, then the speech activity measured at the packet interface shall be less than 43 percent.

It is to be noted that the quality of the speech detector at the originating end is one of the important factors that determines the overall quality of the speech. If the speech detector does not detect speech correctly, it could clip the beginning of the speech bursts thereby causing severe degradation of the speech quality. In contrast, if the speech detector is too sensitive, then more silence intervals will be passed and the compression gain will be reduced.

At the terminating end, a comfort noise or noise fill may be played out instead of silence to minimize the discontinuities between the background noise for speech and silence. The level of the noise fill may be specified in a field in the packet header. Careful selection of the noise power is necessary to avoid the problem of “noise pumping”, an annoying contrast between the background noise during the silence period and the background noise during speech bursts [47].

Note that in the PSTN, echo control is needed whenever digital speech interpolation is used because the speech detector may erroneously classify echoed signals as speech and reduce the compression gain [48], [49]. This is independent of the need for echo cancellation to compensate for the delay due to packetization or to avoid speech clipping due to "freeze-out" [50], [51].

A protocol must allow a method for distinguishing between gaps due to silence and gaps due to missing/discarded packets. This distinction is necessary if a recovery action is used to replace the missing packet. In Recommendation G.764, this distinction is through the *More-Bit* or M-bit which allows a transmitter to signal whether a packet is part of a burst or is the last packet in a burst. The M-bit is set to 1 for all packets except for the last of a burst; in such a case it is set to 0.

With the M-bit mechanism, the terminating endpoint can distinguish the absence of packets at playout time due to silence at the source from the absence due to packet loss in transit. When it finishes playing out a packet and the next packet is not available, the terminating end checks the M-bit of the last packet. If this value is 0, then the absence of a packet indicates a real pause in the transmission at the originating endpoint of the transmitting node. If the value of the M-bit of the last packet is 1, the absence indicates a late or lost packet that had not arrived. This procedure will allow the terminating endpoint to use appropriate fill procedures for speech packet.

For a single speaker, such as in cellular/wireless telephony, adaptive coding techniques may be used to achieve gradual reduction in average bit rate by allocating variable bits to different parts of the speech. In this way, the lowest coding rate is assigned to speech pauses while the highest rate is for active speech.

If the background noise is coded by the lowest rate of a variable bit rate coder, there is less chance of clipping the initial parts of the words. As a consequence, the requirements on the performance of the speech detector may be less stringent in this case. This could be advantageous in more difficult applications where speech interpolation becomes less effective, such as when the ambient noise is high.

I.11 Channel-oriented signalling

The packetization protocol must also relay in a transparent manner channel associated signalling (ABCD signalling).

The information to be passed consists of two types: "signalling transition" and "refresh". A signalling transition occurs whenever the signalling state changes including the presence of alarms. Refresh information is sent on a periodic basis to indicate that the link is still active and to recover from the effects of lost signalling transition packets.

For each voiceband connection with channel-associated signalling, Recommendation G.764 specifies one layer 2 address for voiceband traffic and another for signalling traffic. This out-of-band control is consistent with the ISDN approach to separate the control and user planes (defined in Recommendation I.320). The integration of control and signalling procedures for all sources is now possible while the transfer of control and signalling functions are kept distinct from the transfer of user speech or data frames.

I.12 Extensions

This subclause considers existing and future extensions of the basic speech packetization protocol.

I.12.1 Facsimile

The framework of Recommendation G.764 has been extended to facsimile traffic in Recommendation G.765. Tests have been performed to evaluate the transmission quality of Group III facsimile on Intelsat's Intermediate Data Rate (IDR)[1)] satellite links when Recommendation G.764 is adapted for facsimile transmission as per Recommendation G.765. Under the test conditions (i.e. no tail impairments in the national extensions), packetized facsimile traffic according to the demodulation/remodulation procedure of Recommendation G.765 provides significant performance improvement over

[1)] IDR is the Intermediate Data Rate digital service from Intelsat that allows the transmission of speech, voiceband data, digital data, or video at information rates from 64 kbit/s to 45 Mbit/s. IDR channels use V.35 scramblers. rate 3/4 Forward Error Correction (FEC) with soft decision decoding, and QPSK modulation [52].

circuit-oriented transmission on IDR links in terms of error-free pages, if service-specific Forward Error Correction (FEC) is not used [53], [54]. Furthermore, the results for severely-errored pages show that packetized remodulation is superior even up to a BER of 10^{-5}. The concentration of the effects of IDR error bursts on a single packet together with the distributed control structure of packet systems reduces the frequency of exposure to errors and explains the dramatic improvement achieved with packetized transmission.

I.12.2 Speech/video synchronization

When transmitting a video signal in conjunction with speech, and perhaps data, across the PSTN, the packetization protocol may need to consider the synchronization between the audio and video information. A recent result indicates that a speech channel delay of ± 80 ms relative to the video channel may have little or no effect on intelligibility. Speech intelligibility appears reduced with desynchronization of more than +280 ms or –160 ms relative to the video channel [55].

I.12.3 Interface between the PSTN and LANs

LAN traffic can be integrated through one of several interfaces to the PSTN in emerging enterprise networks. To maintain high quality, it is important to appreciate the differences between the conditions in the PSTN and LANs that may affect speech packetization. For example:

1) Public telecommunications networks have synchronization plans to avoid buffer overflow or underflow [56]. In a LAN, there is no need for such a tight synchronization and other mechanisms are available to monitor buffers at the various stations. Any multimedia protocol design (voice/digital data/video) should specify how the receiver can synchronize with the transmitter for the various types of media.

2) In a LAN, bit error ratios of 10^{-9} or better are typical. In a general PSTN connection spanning several countries, such a bit error ratio cannot always be achieved. Therefore, when transmitting a voiceband service from a LAN to a wide area network, the LAN protocol should take into account this difference in bit error rates.

3) Transmission of LAN traffic across the PSTN may potentially cause large delays and jitter at the receiver (or packet loss) during congestion and high traffic periods. The effects of congestion in a LAN, e.g. packet loss or delay, are often compensated by higher layer protocols such as TCP/IP for digital data traffic. TCP/IP, for example, can invoke retransmission strategies that are valid for data but are not suitable for interactive speech (from a subjective quality viewpoint).

4) Finally, if the LAN traffic is ATM (i.e. 64 kbit/s PCM without digital speech interpolation), the bandwidth required on the PSTN may be excessive.

In summary, the differences between the PSTN traffic and LAN traffic arise from:

1) the conditions in the wide area network are more diverse and less predictable than in a LAN; and

2) it is relatively easier to add capacity to a LAN if congestion is a problem on that LAN.

Accordingly, the constraints on the speech packetization within a LAN may be less stringent than those for the wide area network. Although direct encapsulation of the LAN traffic is the easiest proposition, this approach does not take into account the above differences in the nature of wide area networks and LANs. It is likely that an interface for protocol conversion will be needed between the various LANs and the PSTN.

I.12.4 Extension to new algorithms

Based on the available information, it can be stated that an embedded LD-CELP algorithm can be designed to be used in compression equipment both for mobile (wireless/cellular) and fixed networks. With a single "universal" embedded algorithm, a "seamless" integration of many services can be achieved and many of the operational problems that have arisen because of the usage of non-embedded and/or incompatible encodings can be avoided.

For the end-user, this would optimize bandwidth utilization, while retaining service quality by achieving rapid deployment of access lines and a minimum number of encodings/decodings on an end-to-end basis. Other advantages would be streamlined and centralized administration of the whole network.

Because the network of the future will be more dynamic and its boundaries (from an engineering view-point) will be less clear, the rules of access control may have to be changed without affecting the service quality. For example, a mobile user may want to communicate at a given time with a stationary user through the PSTN. A few moments later, the same user may want to call another wireless/cellular user at a different location or even on a different continent.

Clearly, because of the dynamic nature of the complexities, it would be easier to minimize the number of encodings/decodings along a call path. In addition, compression may be needed to increase the number of calls that can be transmitted over scarce transmission media. If the user remains in the packet domain, the tandem encoding problem can be avoided and graceful degradation under overload can be achieved by using the same embedded algorithm at every link [57].

During the final study period of the CCITT (1988-1992), there was significant effort to investigate how to avoid tandem encodings. An exchange of liaisons took place between SGs XV and II regarding the implications of using one DCME equipped circuit in a connection. It was recognized that the "situation causes major problems, not only in the area of routing, but for Quality of Service, network management, network design, planning and operations" [58]. These problems can be alleviated, if not totally avoided, in the packet domain. Thus, the implications of potential tandemings on networking operations and rules cannot be overlooked when activities for new coders are launched.

Should there be a universal agreement on a low rate embedded coder, it would greatly improve the quality of voiceband service in a decentralized telecommunication network.

I.13 Summary

This appendix has summarized and reviewed the standardization activities for speech packetization in the CCITT (now ITU-T). Topics that are left for further study are:

1) speech packetization in ATM systems; and

2) extensions of Recommendation G.764 to new toll-quality standard algorithms.

References

[1] MONTGOMERY (W.A.): Techniques for packet voice synchronization. *IEEE J. Select. Areas Commun.*, Vol. SAC-1, No. 6: 1022-1028, December 1983.

[2] MIEDEMA (H.), SCHACHTMAN (M.G.): TASI quality – Effect of speech detectors and interpolation. *The Bell System Technical Journal*, Vol. XLI, No. 4, pp. 1455-1473, July 1962.

[3] SRIRAM (K.), MCKINNEY (R.S.), SHERIF (M.H.): Voice packetization and compression in broadband ATM networks. *IEEE Journal Selected Areas in Communication*, special issue on Teletraffic Analysis of ATM Systems, Vol. 9 (3): 294-305, April 1991. Also in part in the *Proc. of the ITC Seminar*, Morristown, NJ, October 9-11, 1990.

[4] ITU Telecommunications Standardization Sector: Impact of transmission delay on Group 3 facsimile. *USA Contribution,* Questions: 4, 8, 9, 13/15, COM 15/D.26, September 1993.

[5] WEINSTEIN (C.J.), FORGIE (J.W.): Experience with speech communication in packet networks. *IEEE J. Select. Areas Commun.*, Vol. SAC-1, No. 6: 963-980, December 1983.

[6] SRIRAM (K.): Dynamic bandwidth allocation and congestion control schemes for voice and data multiplexing in wideband packet technology, *ICC'90*, Vol. 3, p. 1003-1009, Atlanta, GA, April 15-19, 1990.

[7] MONTEILLET (A.), SHERIF (M.H.), RUBIN-JAFFE (K.), BAILLY (R.), ERDREICH (M.): Expérimentation sur site d'un système de multiplication de circuits par mise en paquets, *Annales des Télécommunications*, Vol. 47: (5-6), 175-179, 1992.

[8] GRUBER (J.), STRAWCZYNSKI (L.): Subjective effects of variable delay and speech loss in dynamically managed voice systems. *GLOBECOM'82*, Vol. 2: F.7.3.1-F.7.3.5, Miami, FL, November-December 1982.

[9] GRUBER (J.G.), LE (N.H.): Performance requirements for integrated voice/data networks, *IEEE J. Select. Areas Commun.*, Vol. SAC-1, No. 6: 981-1005, December 1983.

[10] SHERIF (M.H.): Overview of wideband packet protocols. *Proceedings of the 9th International Conference on Digital Satellite Communications ICDSC 9*, Session A4, pp. 135-143, Copenhagen, 18-22 May, 1992.

[11] CASNER (S.), DEERING: First IETF Internet Audiocast, *Connexions Newsletter*, pp. 10-17, 1992.

[12] ISO/CEI JTC 1/SC6/WG1: The Evolution of HDLC: Past! Present! Future!, *WG1 Convener (R.I.G. Prince), Project JTC 1.06.16, Doc. No. 337r*, US Contact: American National Standards Institute, 11 West 42nd Street, NY, NY 10036, 27 January 1993.

[13] BIALLY (T.), GOLD (B.), SENEFF (S.): A technique for adaptive voice flow control in integrated packet networks, *IEEE Trans. Commun.*, Vol. COM-28: 325-333, March 1980.

[14] BIALLY (T.), MCLAUGHLIN (A.J.), WEINSTEIN (C.J.): Voice communications in integrated digital voice and data networks, *IEEE Trans. Commun.*, Vol. COM-28: 1478-1490, September 1980.

[15] SHERIF (M.H.), BOWKER (D.O.), BERTOCCI (G.), ORFORD (B.A.), MARIANO (G.A.): Overview and performance of CCITT/ANSI embedded ADPCM algorithms, *IEEE Trans. Comm.*, Vol. 41, No. 2, 391-399, February 1993.

[16] De IACOVO (R.D.), SERENO (D.): Embedded CELP coding for variable bit-rate between 6.4 kbit/s and 9.6 kbit/s, *ICASSP 91 International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 681-684, Toronto, Ontario, Canada, May 14-17, 1991.

[17] DAVIDSON (G.), GERSHO (A.): Multiple stage vector excitation coding of speech waveforms, *ICASSP 88 International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 163-166, NY, April 11-14, 1988.

[18] LE GUYADER (A.), LOZACH (B.), MOREAU (N.): Embedded algebraic CELP coders for wideband speech coding, *Signal Processing VI Theories and Applications, Proceedings of EUSIPCO-92 Sixth European Signal Processing Conference*, Vol. 1, pp. 527-530, Brussels, Belgium, August 24-27, 1992. J. Vandewable, R. Boch, M. Moonen and A. Osterlinch (eds.), Elsevier, Amsterdam, 1992.

[19] LE GUYADER (A.), BOURSICAUT (E.): Embedded wideband VSELP speech coding with optimized codebooks, *Proceedings of IEEE workshop on speech coding for Telecommunications: Speech coding for the network of the future*, pp. 15-16, Sainte-Adèle, Québec, Canada, October 13-15, 1993.

[20] TANIGUCHI (T.), OTHA (Y.), YAO (J.): Structure codebook for variable-rate CELP coding, *Proceedings of IEEE workshop on speech coding for Telecommunications: Speech coding for the network of the future*, pp. 87-88, Sainte-Adèle, Québec, Canada, October 13-15, 1993.

[21] LEUNG (T.W.), LE BLANC (W.P.), MAHMOUD (S.A.): Speech coding over frame relay networks, *Proceedings of IEEE workshop on speech coding for Telecommunications: Speech coding for the network of the future*, pp. 75-76, Sainte-Adèle, Québec, Canada, October 13-15, 1993.

[22] USA: Embedded LD-CELP, UIT-T SG 15 Contribution D.225, Geneva, Switzerland, 16-27 May, 1994.

[23] BOWKER (D.O.), ARMITAGE (C.B.): Performance issues for packetized voice communication, *Proc. Nat. Comm. Forum*, 41 (3): 1087-1092, 1987.

[24] BOWKER (D.O.), DVORAK (C.A.): Speech transmission quality of Wideband Packet Technology, *GLOBECOM'87*, Vol. 3: 1887-1889, Tokyo, Japan, November 15-18, 1987.

[25] DRAVIDA (S.), SRIRAM (K.): End-to-end performance models for variable bit rate voice over tandem links in packet networks, *IEEE INFOCOM'89*, Vol. 3: 1617-1622, Ottawa, Canada, April 23-24, 1989.

[26] KARANAM (V.R.), SRIRAM (K.), BOWKER (D.O.): Performance modeling of variable bit rate voice in packet switched networks, *GLOBECOM'88*, Vol. 3: 1617-1622, Hollywood, FL, November 28-1 December, 1988.

[27] SRIRAM (K.), LUCANTONI (D.M.): Traffic smoothing effects of bit dropping in a packet voice multiplexer, *IEEE INFOCOM'88*, pp. 759-770, New Orleans, LA, March 1988.

[28] SHERIF (M.H.): Overview of wideband packet protocols, *Int. J. Sat. Comm.*, Vol. 10, 261-268, 1992.

[29] YIN (N.), LI (S.Q.), STERN (T.E.): Congestion control for packet voice by selective packet discards, *GLOBECOM'87*, Vol. 3: 1782-1786, Tokyo, Japan, 15-18 November, 1987.

[30] YIN (N.), STERN (T.), LI (S.Q.): Performance analysis of a priority-oriented system, *IEEE INFOCOM'87*, p. 856-862, San Francisco, March 1987.

[31] KASHORDA (M.), JONES (E.V.): Simultaneous voice/data transmission over a single 64 kbit/s ISDN bearer channel, *Proceedings of the Third IEE Conference on Telecommunications,* pp. 281-285, Heriot-Watt University, Edinburgh, UK, 1991.

[32] Dowty Communications Ltd.: Packetized voice over frame relay: Issues and considerations, *Contribution to the frame relay forum*, Mai 1992. Contact: Ted Hatala, Cray Communications, Mayze House, Westmead, Swindon, Wiltshire SNS 7UQ, England.

[33] Da SILVA (L.A.), PETR (D.W.), FROST (V.S.): A class-oriented replacement technique for lost speech packets, *IEEE INFOCOM'89*, Vol. 3: 1098-1105, Ottawa, Canada, April 23-24, 1989.

[34] PETR (D.W.), Da SILVA Jr. (L.A.), FROST (V.S.): Priority discarding of speech in integrated packet networks, *IEEE J. Select. Areas Commun.*, Vol. 7, No. 5: 644-656, June 1989.

[35] JAYANT (N.S.), CHRISTENSEN (S.W.): Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure, *IEEE Trans. Commun.*, Vol. COM-27, No. 2, 101-109, February 1981.

[36] GOULD (K.W.), COX (R.V.), JAYANT (N.S.), MELCHNER (M.J.): Robust speech coding for the indoor wireless channel, *AT&T Technical Journal*, Vol. 72, No. 4, pp. 64-73, July/August 1993.

[37] AT&T: G.728 decoder modifications for frame erasure concealment, document No. AH-5-10, ITU-T Rapporteur's meeting, Tokyo, Japan, March 1994. Contact: Mr. S. Hayashi, NTT, Human Interface Laboratories, Speech Laboratory, 9-11 Midori-cho 3 Chome, Musashino-shi, Tokyo 180, Japan.

[38] GOODMAN (D.J.), LOCKHART (G.B.), WASEM (O.J.), WONG (W.C.): Waveform substitution techniques for recovering missing speech segments in packet voice communications, *IEEE Trans. Accoust. Speech, Signal Processing*, Vol. ASSP-34, No. 6: 1440-1448, December 1986.

[39] WASEM (O.J.), GOODMAN (D.J.), DVORAK (C.A.), PAGE (H.G.): The effect of waveform substitution on the quality of PCM packet communications, *IEEE Trans. Accoust. Speech, Signal Processing*, Vol. ASSP-36, No. 3: 342-348, March 1988.

[40] MINOLI (D.): Optimal packet lenght for packet voice communication, *IEEE Trans. Commun.*, Vol. COM-27, No. 607-611, March 1979.

[41] American National Standard ANSI/IEEE C37.1-1979: Definition, specification, and analysis of manual automatic, and supervisory station control and data acquisition, Sponsored by the *Substations Committee of the IEEE Power Engineering Society,* Deccember 15, 1977.

[42] HOEFFELMAN (J.): Performance de la procédure de télécommunication HDLC et adaptation aux exigences d'une haute sécurité de transmission, *Onde Electrique,* Vol. 60, No. 11, 52-60, 1980.

[43] SRIRAM (K.), SHERIF (M.H.): Frame size requirements in frame relay networks, contribution No. T1S1.1/90-185 (T1S1.2/90-128), April 17, 1990.

[44] SHERIF (M.H.), BOSSE (M.P.): Les paquets de bande élargie: une nouvelle technique de transmission, *Annales des Télécommunications*, 46, No. 7-8, pp. 392-407, 1991.

[45] YATSUZUKA (Y.): High-gain digital speech interpolation with adaptive differential PCM encoding, *IEEE Trans. on Commun.*, Vol. COM-30, No. 4, 750-761, April 1982.

[46] YATSUZUKA (Y.): Highly sensitive speech detector and high-speed voiceband data discriminator in DSI-ADPCM systems, *IEEE Trans. on Commun.*, Vol. COM-30, No. 4, 739-750, April 1982.

[47] DVORAK (C.A.), ROSENBERGER (J.R.): Deriving a subjective testing methodology for digital circuit multiplication and packetized voice systems, *IEEE J. Select. Areas Commun.*, Vol. 6, No. 2: 235-241, February 1988.

[48] ANSI T1.508-1992: Loss Plan for Evolving Digital Networks.

[49] Committee T1 Report No. 27: A Technical report on echo cancelling, Prepared by *T1A1.6 Working Group on Specialized Signal Processing*, Washington, D.C., November 1993.

[50] KIHLSTRÖM (L.O.), VIKLUND (N.), BRIDGEWATER (D.): The development of an open-network digital circuit multiplication equipment system, *International Journal of Satellite Communications*, Vol. 8, 461-490, 1990.

[51] SHERIF (M.H.), CLARK (R.J.), FORCINA (G.P.): CCITT/ANSI Voice Packetization Protocol, *International Journal of Satellite Communications*, Vol. 8, 429-436, 1990.

[52] Performance Characteristics for Intermediate Data Rate (IDR) Digital Carriers, *Intelsat Earth Station Standards, IESS-308 (Revision 6B)*, December 1992.

[53] SHERIF (M.H.), CUEVAS (E.), LIEBERT (T.), DIMOLITSAS (S.): Improvement of the Transmission Quality of Group III Facsimile on IDR Satellite Links, Proceedings of the *First IEEE Symposium on Global Data Networking*, pp. 101-111, Cairo, Egypt, December 12-15, 1993.

[54] SHERIF (M.H.), CUEVAS (E.), LIEBERT (T.), DIMOLITSAS (S.): Transmission Quality of Group III Facsimile on IDR Satellite Links with Circuit Multiplication Equipment, *International Journal of Satellite Communications*, Vol. 12, 167-180, 1994.

[55] Royal PTT, The Netherlands: Effects of desynchronization of vision and speech on the perception of speech: Preliminary results, *Delayed Contribution D.81 to CCITT SG XII. Q.27/XII and 31/XII*, Brasilia, 6-13 September, 1991.

[56] REY (R.F.), Technical Editor: Engineering and operation in the Bell System, *AT&T Bell Laboratories*, 1983.

[57] AT&T: Cellular applications of PCME, *ITU-T White Contribution* COM 15-36-E, July 1993.

[58] Speech processing techniques in the international digital network, *CCITT Liaison statement from WP II/1, Q.6/II (Routing), COM II-R8*, 24.10-3.11.89, Geneva, Switzerland.