Zeitschrift: Comtec : Informations- und Telekommunikationstechnologie = information and telecommunication technology

Herausgeber: Swisscom

Band: 74 (1996)

Heft: 9-10

Artikel: Real-time voice and sound services over ATM

Autor: Vörös, Paul

DOI: https://doi.org/10.5169/seals-876794

REAL-TIME VOICE AND SOUND SERVICES OVER ATM

While the Asynchronous Transfer Mode (ATM) was not expressly invented with speech and sound or any other real-time, synchronous, low bit rate services in mind, most people agree that ATM must also carry such traffic. Since voice still represents anywhere from 50 to 80 % of the total traffic volume, the inability of ATM to satisfactorily carry this service may jeopardize or delay the wide-spread introduction of ATM. This and similar considerations have prompted ITU-T and the ATM Forum lately to step up their efforts in the area of voice over ATM.

The reasons for the earlier neglect may be that data traffic increases at a much higher rate and real-time services over ATM are more difficult to handle. In ITU-T, ATM has been proposed first of all as an universal multi-

PAUL VÖRÖS, BERN

plex and switching scheme for services well in excess of 64 kbit/s, able to support flexible bit rates over asymmetrical connections particularly well suited for data and multimedia communication. A general thematic review of the ATM technology may be found in [1] and a more organic treatment of this subject in [4].
The main concern of this contribution will be the effects of ATM on real-time voice and sound services and to examine how signal processing might deal with some of the particularities and incongruities of ATM networks to carry such services and help to achieve the required level of Quality of Service (QoS) combined with an efficient use of the transmission medium.

Basic characteristics of ATM influencing the quality of signal transmission

Compared with the classical Synchronous Transfer Mode (STM), ATM has the additional particularity that cells may be lost or inserted into other connections and that, unless the cell rate of the source is an exact submultiple of the cell transfer rate of the system, there is jitter (min. 2.7 µs at 155 Mbit/s link rate) on the delay even in otherwise ideal systems. Similarly, system loading has a bearing not only on connection admission, but also on cell jitter and cell loss. In ATM, moreover, a single channel monopolizes the total transmission capacity for a much longer period at a time.
For real-time audio and video, the effects of ATM transmission in the presence of network impairments can be very different from those observed in STM networks; e.g. burst errors are fairly well distributed among PDH (bit-wise interlacing of bit streams of lower hierarchical levels) and SDH channels (byte-wise interlacing of the payload) and are often reduced to single bit/byte errors, while for ATM a whole cell (376 bits of the same channel) or, for high speed sources, several consecutive cells may be affected.
The role of the ATM Adaptation Layer (AAL) is to map the various types of traffic onto the underlying ATM cell layer. Of the currently defined AAL, AAL types 1 and 2 [2] are reserved for real-time, connection-oriented applications, where the *recovery of the source clock frequency* is important. AAL1 supports real-time traffic (audio, video) generated at a Constant Bit Rate (CBR) and circuit emulation. AAL2, which is still only sketchily defined, supports Variable Bit Rate (VBR) traffic. The reasons for the relative neglect of AAL2 may be the increased *difficulty to maintain the flow of timing information across a call and the handling of lost and misinserted cells.* The timely and successful development of AAL2 is, however, crucial for the full potentiality of the ATM technology to unfold, since the information content of signal sources destined for immediate human appreciation usually varies with time.
The AAL is actually split into two parts, viz., Convergence Sublayer (CS), Segmentation and Reassembly (SAR) Sublayer. For the AAL1 and AAL2, the 48-octet SAR Service Data Unit (SAR-SDU) is made up of the 1-octet SAR Protocol Data Unit (SAR-PDU) header and the 47-octet SAR-PDU payload issuing from the signal source. The SAR generates a three-bit Sequence Number

(SN) at the transmitting end and checks it at the receiving end to detect missing or misinserted cells. The AAL handles most of the information responsible for the QoS, viz., AAL user information, Cell Delay (CD), Cell Delay Variation (CDV – a one-point CDV is implied), lost and misinserted cells, timing relation, correction of bit errors and lost cells, etc.

Solutions for CBR coding

With some rare exceptions, all existing ITU-T source coding algorithms have been developed for synchronous networks and produce CBR output for transmission over fixed capacity channels. The transfer rate is dictated by the desired quality: In broadcasting, it covers the range of contribution quality to commentary grade of service. If the average transfer rate is smaller than the peak rate, then the quality must suffer each time the demand for instantaneous bit rate exceeds the transfer rate, while a channel with capacity equal to the peak rate will be very inefficiently used most of the time. For this reason, AAL1 can support a variable QoS only. On the other hand, AAL2 adapts itself to the natural rate of any source and does thus most efficiently guarantee a *constant* QoS. Coding for VBR will be treated under 'VBR coding for ATM'.

Delay in ATM networks

Real-time communication means the *timely* delivery of information. The subjective degradation produced by a given average cell delay is application-dependent. For instance, in the case of 'off-the-air' monitoring, even an experienced reporter may find a delay greater than 13 to 18 ms disturbing. For telephony, the excess delay of 6 ms due to packetizing 64-kbit/s speech may be enough to push the one-way delay above the 24-ms limit and make the use of Echo Cancellers (EC) necessary, even in national networks (e.g. for voice circuits crossing an ATM island involving at least one PSTN terminal). Longer delays may restrict the choice of suitable voice terminals to four-wire-connected equipment, headsets or special telephones with high acoustic loss between microphone and earphone. Conversation can become difficult even in the absence of echo, and round-trip delays on the order of 400 to 500 ms or greater are undesirable for normal verbal interaction. Reference [3] contains a short but informative treatment of the effects of delay on voice communication in ATM networks.

Transit time is made up of coding delay (sum of algorithmic and processing delay), ATM packetizing/depacketizing delay, queuing time in terminals and switches and propagation delay of the transmission medium. The total delay for a 1000-km, 64-kbit/s voice circuit (STM-ATM-STM chain) having the parameters

- number of ATM exchanges = 8 (50 consecutive queues)
- number of synchronous exchanges = 2
- link load = 80 %
- Cell Loss Ratio (CLR) = 10^{-10}

amounts to 12.53 ms [4] using the 'Fast SN Algorithm' [2]. This delay is mainly determined by the transmission and packetizing delay, and it is largely insensitive to transmission speed.

With more complex speech coding schemes, the removal of redundance and irrelevance (in a wide sense) from the source output allows the reduction of the bit rate to about 4 kbit/s while still maintaining 'toll quality'. In STM systems, the synchronization of the sink with the source is straightforward, as long as the bit rate is simply related to the 8-kHz network timing or if the frame size is compatible with e.g. ITU-T Rec. H.221 (i.e., the size of the frame or block of samples acquired by the encoder should be a sub-multiple of 10 ms). This also explains why most ITU-T speech codecs developed for synchronous networks do not have (to have) internal (algorithmic) framing. Consequently, the mapping of the output of such codecs onto ATM cells only rarely yields a 100-% cell-fill. *Table 1* summarizes the basic delay parameters of some speech and sound codecs when using ATM transport [5].

The delay values in Table 1 are valid for a single encoding-decoding operation only. Partial cell-fill to reduce mapping delay is always possible, if frames are not split up. In a connection involving conferencing bridges, at the MCU site, all the audio signals are decoded (returned to their uniformly quantized PCM-state), summed, and the composite signal is reencoded. This and similar tandeming operations (e.g. transcodings) further degrade the audio quality and double the delay. (If all parties use the same codec based on subband decomposition, then some delay reduction may be realized by summing the partial signals in each subband.)

The AAL and higher-layer protocols can correct almost all network imperfections, but they are powerless as far as the per-VC delay is concerned (the use of the 'Fast SN Algorithm' [2] may shorten the delay in the receiver by one cell period, if, for a given application, it is not required to distinguish between lost and misinserted cells), and only careful network planning can help. CS also supports any 8-kHz-based structure used in circuit-mode services of ITU-T Rec. I.231; in particular, ATM can transport a 30-channel, 2048-kbit/s primary multiplex signal (E1) with an additional delay of only about 0.5 ms.

For compressed speech used in low-rate, delay-sensitive applications, ITU-T, SG13, proposes a new bandwidth-efficient AAL for composite user (AAL-CU) information. The problem of audio delay presents itself quite differently for multimedia services (videoconferencing, videotelephony, etc.): In this case, the video coding delay (typically 150 ms) is usually the longer one, and, in order to synchronize the audio with the video signal, an additional delay must be added in the audio path.

Measures against transmission errors

For modern optical transmission systems, under normal operating conditions, very low Bit Error Ratios (BER) may be realized. In addition, over 99 % of the errors are single-bit errors that seldom cause cell losses, and cell loss probabilities between 10^{-8} and 10^{-12} are achievable for ATM systems. Under maintenance conditions such as protection switching, about 10 % of the errored seconds have bursts with more than 100 errors [4]. Packet-oriented networks are normally dimensioned so that switching and transmission systems have a comparable quality.

If the network BER and CLR are considered excessive for some application, error control methods may be used: Draft Rec. I.363.1 describes several Forward Error Correction (FEC) techniques using the Reed-Solomon (RS) codes. When used for bit error correc-

Codec description	Bit rate(s)	Frame size	Codec delay	# Frames per ATM cell	ATM mapping delay	Notes
G.711 PCM	64 kbit/s	0.125 ms	1.625 ms	47	5.875 ms	Partial cell-fill is possible
G.726 and G.727 ADPCM	40, 32, 24 and 16 kbit/s	10 ms (DECT)	10 ms (DECT)	1	0	40 octets per ATM cell at the rate of 32 kbit/s. They succeed ITU-T Rec. G.721
G.722 (7 kHz speech)	64, 56 and 48 kbit/s	0.125 ms	1.615 ms	47	5.875 ms	Based on the 64-kbit/s rate 2 subbands (SB): 0–4/4–8 kHz Partial cell-fill is possible
G.728 LD-CELP	16 kbit/s	2.5 ms	5 ms	9	20 ms	45 octets per ATM cell (N-ISDN videophone, H.320 series)
GSM full-rate PRE-LTP1)	13 kbits/s	20 ms	90 ms (total system + FEC)	1	0	260 bits per ATM cell
G.729 CS-ACELP	8 kbit/s	10 ms	25 ms	4	30 ms	40 octets per ATM cell; for wireless and cellular applications; may be used with VAD (see G.723.1 below)
GSM half-rate VSELP[1]	5.6 kbit/s	20 ms	90 ms (total system + FEC)	1	0	14 octets per ATM cell. Up to 3 frames may be packed
G.723.1 A-CELP (interoperable with G.729.1)	6.3 and 5.3 kbit/s	30 ms	67.5 ms	2	30 ms	PSTN videophone appl., H.324 series. Silent frames are coded with 4 octets only (irregular cell-fill)
Proposed 4 kbit/s	4 kbit/s	20 ms	45 ms	4	60 ms	40 octets per ATM cell. Standard available in 1998
Proposed 7-kHz algorithm (G.722-quality at half-rate or less)	32, 24 and 16 kbit/s	10, 20 or 30 ms	> 25 ms	1 (10 ms frame)	0	Only the 10-ms coder output would fit into a cell at all three bit rates filling 20, 30 and 40 octets, respectively. Standard available in 1998
SB15 [6] for 2 × 64-kbit/s ISDN channels	64–128 kbit/s sampling freq. f_s = 32 kHz	0.25 ms	10 ms	11	2.5 ms	44 octets per ATM cell (internal framing is used only to sync. 2 ISDN B-channels)
J.52 (ISO/IEC 11172-3 and 13818 for H.221)[2]	32–384 kbit/s, sampling frequency f_s= 32, 44.1 and 48 kHz	Layer I: 8 ms Layer II: 24 ms Layer III: 24 ms at all bit rates	30 ms 45–57 ms 59–157 ms	0 (min. 48 bytes are generated per frame at f_s = 32 kHz)	0	The algorithm has an internal frame structure – its output may be continuously mapped onto consecutive ATM cells

[1] GSM – Global System for Mobile Communications, an ETSI standard for digital cellular system
[2] Commonly known as MPEG-1; ISO/IEC 13818, also called MPEG-2, has a 1/2fs-extension for audio with a minimum bit rate of 16 kbit/s

Table 1. Impact of using ATM transport on delay for some speech and sound codecs.

tion only, the RS(128,124) code can correct up to two errored octets out of 128 at the cost of a 3.1 % overhead and a delay of about three cells at the receiver. When protection against both bit errors and cell losses is required, the same RS code with octet interleaving can correct (reconstruct), among others, any combination of up to four lost cells at the expense of a 128-cell delay in both the sender and the receiver. With the shortest RS code capable of recovering one lost cell and using a 16-cell interleaver, the delay for a 64-kbit/s voice channel would amount to 88.2 ms. For lower bit rate speech codecs, this delay would be higher still.

Such FEC techniques are obviously not very helpful for voice communication: They were primarily intended for high-speed video services (although there are some doubts as to the effectiveness of some RS codes in the presence of burst errors), where the resulting delay would be considerably less. For services that are characterized by low average cell arrival rate and short holding time, the probability that any one connection will be disturbed by cell losses is small; e.g., for telephony with CLR = 10^{-7} and typical call duration of 90 s, on the average, only one out of 653 calls will experience a cell loss. Such optimism notwithstanding, the problem of minimizing the subjective impairment of speech due to cell losses has received a lot of attention [7–9].

Techniques for PCM (G.711)

- zero-amplitude or white noise stuffing: replacing all the amplitudes of a missing cell by zero values or white noise of the same energy as the previous cell
- cell repetition: repeating last received cell
- assigning the higher half-byte and the lower half-byte of each PCM sample to alternate cells
- assigning each PCM sample to alternate cells based on an even/odd criterion

With the first two procedures, there is no significant difference between the subjective quality of the reconstructed signals; on the average, zero-substitution or white noise injection may be preferable, since possible phase hits at the cell boundary of repeated cells – unless special precautions such as pattern matching are taken – may be more objectionable for voiced signal segments. For CLR < 1 %, cell substitution and sample splitting yield almost identical Mean Opinion Scores (MOS) [8]. The missing odd or even cell is reconstructed by interpolation. In this case, one must recall that this technique produces spectral folding due to a 2 : 1 undersampling (aliasing), and no amount of postprocessing can fundamentally change this fact. The effects of a missing cell depend on the spectral content of the neigbouring cells (including possible subjective masking effects). It is implicit in the third method that, under overload conditions, the network drops the lower half-byte cells with low Cell Loss Priority (CLP)=1; for the other three methods, there is no such assumption to be made about the CLP bit. In addition, the last two methods also double the mapping delay (Table 1). Nevertheless, all of the above methods contain some basic ideas that may be used to develop robust coding techniques for ATM.

Techniques for ADPCM (G.726, G.727)

Rec. G.721 (32-kbit/s ADPCM algorithm) is the prototype for the two very similar Rec. G.726 and G.727. From an ATM view-point, the most important difference between the two recommendations is that the latter specifies an *embedded codec,* i.e., the prediction is performed with a reduced number of bits (so-called core bits) obtained by masking some of the LSB of the error signal. The masked bits serve only to reduce the quantization noise and may be lost during transmission without compromising the stability of the predictor in the receiver. (This feature is exploited in Digital Circuit Multiplication Equipment [DCME] under overload conditions to trade front-end clipping for the less objectionable quantization noise.) ATM networks may also take advantage of this possibility by packing embedded bits into separate low-CLP cells. If only low-CLP cells are lost, the transmission quality suffers only graceful degradation; e.g., the effect of a CLR of up to 50 % on a source coded with 32 kbit/s (2 core bits + 2 embedded bits per error sample) would not be much worse than that produced by encoding at 16 kbit/s without cell loss. Otherwise, if packed in the natural order, ADPCM is very sensitive to cell losses due to predictor mistracking in the receiver.

Abbreviations

AAL	ATM Adaptation Layer
AAL-PDU	AAL Protocol Data Unit
AAL-SDU	AAL Service Data Unit
ACELP	Algebraic Code Excited Linear Prediction
ADPCM	Adaptive Differential Pulse Code Modulation
ATM	Asynchronous Transfer Mode
AUT	Aperture Uncertainty Time
BER	Bit Error Ratio
CBR	Constant Bit Rate
CD	Cell Delay
CDV	Cell Delay Variation (1-point CDV)
CELP	Code-Excited Linear Prediction
CLP	Cell Loss Priority
CLR	Cell Loss Ratio
CS	Convergence Sublayer
CS-ACELP	Conjugate Structure ACELP
CS-PDU	CS Protocol Data Unit
DECT	Digital European Cordless Telecommunication
EC	Echo Canceller
ETSI	European Telecommunications Standards Institute
FBR	Fixed Bit Rate
FEC	Forward Error Correction
FIFO	First In/First Out
GSM	Global System for Mobile Communications
IPv6	Internet Protocol, version 6
ISDN	Integrated Services Digital Network
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
LD-CELP	Low-Delay Code Excited Linear Prediction
LSB	Least Significant Bit
MCU	Multipoint Control Unit
MOS	Mean Opinion Score
MPEG	Moving Pictures Expert Group
PCM	Pulse Code Modulation
PDH	Plesiochronous Digital Hierarchy
PLL	Phase-Locked Loop
PSTN	Public Switched Telecommunications Network
QMF	Quadrature Mirror Filter
QoS	Quality of Service
RS	Reed-Solomon
RSVP	Resource Reservation Protocol in Internet
RTS	Residual Time Stamp
SAR	Segmentation and Reassembly Sublayer
SAR-PDU	SAR Protocol Data Unit
SAR-SDU	SAR Service Data Unit
SB-ADPCM	Subband Adaptive Differential Pulse Code Modulation
SDH	Synchronous Digital Hierarchy
SDT	Structure Data Transfer
SN	Sequence Number
SRTS	Synchronous Residual Time Stamp
STM	Synchronous Transfer Mode
UPC	Usage Parameter Control
VBR	Variable Bit Rate
VC	Virtual Channel
VCXO	Voltage Controlled Crystal Oscillator

Techniques for CELP and ACELP codecs (G.728, G.729, G.723.1)

Some of these codecs have been especially developed for mobile applications, and, in most cases, an error concealment strategy for frame erasures has been implemented in the decoder. If a frame erasure occurs (signalled by a mechanism external to the decoder), the decoder switches to the frame erasure concealment mode; e.g., for the G.729 codec, the error free-MOS score of 3.49 was reduced to 3.13 over a channel with 3 % burst frame erasures. Cell losses should have similar effects, if the ATM packetization respects algorithmic frame boundaries indicated in Table 1.

Techniques for Subband Coding (G.722, SB15 [6], MPEG-1/2)

The coding quality at low bit rates depends on a subtle mixture of bandwidth and quantization noise. This consideration, among others, has prompted MPEG to introduce the MPEG-2 (ISO/IEC-13818) family of codecs at half the sampling rate of the corresponding MPEG-1 (ISO/IEC-11172-3) codecs. From the subjective quality point of view, the lower frequency bands are more important, and the higher bands serve only to enhance the basic quality by providing a better time resolution.
The ITU-T Rec. G.722 specifies a two-band codec with an independent ADPCM coder in each subband. The low-band predictor uses 4 bits and the high-band 2 bits per sample period. The remaining two embedded bits are assigned to the low-band and used either to enhance speech quality or to carry ancillary data. Thus, for each block of 4×47 octets, one cell containing enhancement bits only could be created, albeit at the cost of quadrupling the minimum packetization delay, which is quite acceptable, e.g. for videoconferencing. Furthermore, dropping the prediction bits in the highband would not be nearly as objectionable as dropping those in the low-band or losing the even/odd cells for G.711, since now the aliasing distortion is limited to a narrow band around the QMF crossover frequency of 4 kHz. For the four-subband SB15 [6], two octets per frame are assigned to the three lower frequency bands covering the 0-to 8-kHz portion of the spectrum and one octet to the 8-to 16-kHz subband. The remaining 8 bits may be distributed *on demand* among the four subbands.
A test of the 2×64- vs.1×64-kbit/s mode yielded a score of +0.71 in favour of the 2×64-kbit/s mode on the seven-grade comparison scale (+1.0 means 'slightly better'). This suggests the possibility of creating three different cell hierarchies: Out of every four cells generated, two contain basic information guaranteeing a certain minimum quality, one carries information pertaining to the 8- to 16-kHz portion of the spectrum, and one cell contains the enhancement bits to further improve the quality by reducing the quantization noise.
Even for these relatively simple codecs, at least three different hierarchies (layers) of information can be identified. Unfortunately, ATM recognizes only two CLP classes (primarily intended to implement the 'tagging' option to convert CLP=0 cells identified by the UPC function as 'nonconforming' into CLP=1 cells to be merged with the user-submitted CLP=1-traffic), which may be *completely inadequate to satisfy future (multimedia) communication needs.* There is, however, an essential difference between the network-generated and the user-submitted CLP=1-cells: While the network generates these cells in bursts at random intervals during periods of traffic violation, the user-generated CLP=1-cells are almost uniformly and deterministicly distributed in time. With a good traffic mix it can be assured that even in switches with the most efficient central queuing [4], there are always a few cells with CLP=1 in the queue. Consequently, *significantly higher CLR could be tolerated and bigger link loads attained without measurably degrading the subjective quality.*
While in ATM the time-sensitivity issue has been largely solved, coders could certainly use more than one CLP bit (the next generation Internet protocol IPv6 will have four bits reserved for time-sensitivity and drop-priority classes) to support *scaleable* speech/audio coding. Scaleability is the ability of the terminal or network to select the bit rate, bandwidth, decoder complexity, delay, etc., compatible with terminal capabilities and/or network environment. Thus, *scaleability of the source coding algorithm becomes one of the important side conditions for the interoperability of disparate terminals.* The codec that comes close to being all of the above is the ISO/MPEG layer 3 [10], where the bit rate varies from 128 kbit/s, ≥16 kHz bandwidth for CD quality stereo to 16 kbit/s, 4.5 kHz bandwidth for 'better than short wave' mono, and at least *five drop-priority classes can be identified.* If only two CLP classes are available, however, discarding one low-priority cell may render all the others in the same coding block useless.
An alternative way to transport layered information may be to set up a distinct VC for each subflow (at least as many as terminal modes of operation). *This would also permit the receiver to select its own QoS* (just as with RSVP of IPv6). The drawbacks of this technique are that all the received subflows must be synchronized, and it does not really solve the problem of multilayered information transfer in ATM.

Synchronization of CBR connections

There are several source clock recovery methods detailed in the ITU-T Rec. I.363.1 [2]. The size of the buffer at the receiver must be sufficient to ensure that for the worst-case CDV no cell under-/overflows occur. The CDV may be considered as a random variable with a Gaussian probability density function [4] which exceeds a given threshold CDV_t with probability p_t. For this case, the standard deviation σ may be calculated as

$$\sigma = CDV_t / \mathrm{erfc}^{-1}(p_t) \qquad (1)$$

Synchronous CBR services

This is the simplest case where the 8-kHz clock is available from the network. ITU-T Rec. G.823 specifies the peak-to-peak jitter for a 2048-kbit/s primary multiplex to be smaller than 1.5 UI = 732 ns. Most of this phase noise power lies in the 20-Hz to 18-kHz band, and the standard clock extraction circuit does not substantially modify this value. If CDV_t = 732 ns and $p_t=10^{-10}$, then Eq. 1 yields $\sigma = 159 \times 10^{-9}$ s.
If the jitter, like the signal, is white Gaussian noise, then its contribution to the base-band signal to noise ratio (S/N) in the *transmit direction* is given by

$$S/N = 20 \cdot \log(\sqrt{3}/[\pi \cdot f_s \cdot \sigma]) \qquad (2)$$

Eq. 2 may be easily derived for an A/D converter, where s is the effective val-

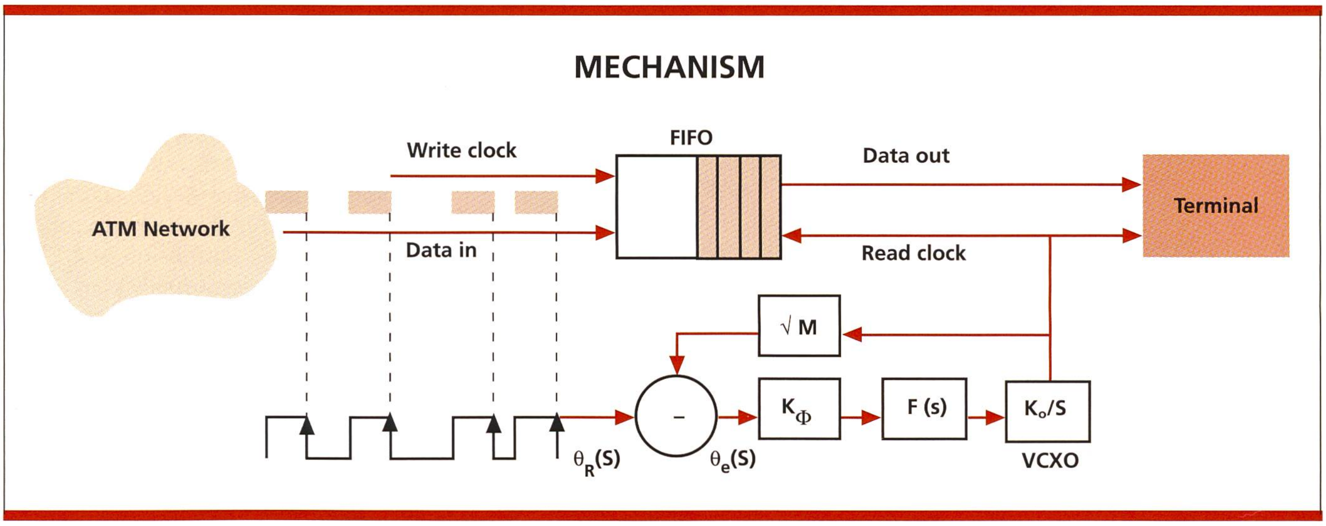

Fig. 1. Adaptive clock recovery mechanism.

ue of the uncertainty of the sampling instants, and it is also valid, when the transmit clock is extracted from the incoming data stream (clock loop-back). For PCM speech sampled at 8000 Hz, Eq. 2 yields 54 dB, which does not significantly reduce the required signal to total distortion ratio of 34 dB (cf. ITU-T Rec. G.712).

For music signals uniformly quantized with 16 bits, a maximal signal to quantization noise ratio of 96 dB is obtained. In a well-dimensioned system, the contribution of the jitter to the S/N should be similar, i.e., at a sampling frequency of 32 kHz, the phase jitter should be filtered to have s – 0.273 ns. To achieve this, the PLL would have to have a bandwidth of 0.053 Hz and a time constant of ca. 1.5 s for a proportional control loop.

For the *receive direction,* the results obtained from Eq. 2 are accurate within a few decibels. In any case, the S/N gives little indication of subjective impairments (with some perceptual codecs, near CD quality can be obtained at a S/N of ~25 dB), and nothing can replace subjective testing.

Adaptive clock method

Here the assumption is that there is a known relationship between the sampling frequency f_s at the source and the average cell-arrival rate, which can be locally modified by CDV. The CDV for a 1000 km long connection described in Sec. 'Delay in ATM networks' amounts to 600 μs [4] with a standard deviation σ_c = 130 μs for p_t=10^{-10}. The cells are read into a FIFO buffer, which is emptied at the local clock rate. The control voltage for the VCXO in *Figure 1* is derived from the filtered (smoothed) fill-level of the FIFO buffer.

The rising edges of the reference input signal are shown to coincide with the cell arrival times. In this way the phase detector output reflects exactly the state of the FIFO.

For an active loop filter F(s), the phase transfer function may be written as

$$H(s) = (2\xi\omega_n s + \omega^2_n)/ (s^2 + 2\xi\omega_n s + \omega^2_n) \qquad (3)$$

where $\omega^2_n = K_\Phi K_o / M\tau_1$ is the resonance frequency and $\xi = \tau_2\omega_n/2$ is the damping factor; τ_1 and τ_2 are time constants. The damping factor ξ is constrained by Rec. Q.551, and the parameters of the VCXO are critical for the feasibility of the PLL.

The spectral density of the output phase is given by

$$\Theta_o(f) = |H(j2\pi f)|^2\Theta_R(f) \qquad (4)$$

The designer is faced with the practical problem of calculating the equivalent noise bandwidth in order to dimension the PLL or simply to find out, if it can be built at all. The distribution function of the CDV, however, depends on the cell rate of the observed as well as of the background traffic. Moreover, the interarrival times are correlated, which makes the analytical modeling of the input spectral density function extremely complex. On the one hand, the PLL must have the necessary high capture ratio $\Delta\omega_c \approx 2\xi\omega_n$, which for ξ>2 equals to the 3_{dB}-bandwidth of the PLL, and, on the other hand, the worst case output jitter must be sufficiently attenuated to prevent any additional *subjective impairment* of the reconstructed speech. There are indications [11] that periodic jitter may represent the worst case. Two extreme cases will be examined:

CDV without background traffic

In this case, only the observed traffic is loading a perfectly synchronized system. It was mentioned in Sec. 'Basic characteristics of ATM in fluencing the quality of signal transmission' that even in such cases – unless the cell rate of the source is an exact submultiple of the cell transfer rate of the system – there is a minimal jitter of 2.7 μs at the 155-Mbit/s link rate (this may be likened to the 'waiting-time jitter' in PDH systems, only this time it is ~30 times bigger). The jitter frequency depends on the source to link cell rate ratio. An analogous situation exists with the SRTS method, when the same stable network clock is available to both the sender and the receiver: It was reported in [11], however, that the RTS oscillates between two stable values and *may cause audible low frequency wow and flutter* that characterize some old play-back equipment. Except for the phase detector, the PLL used with the SRTS method could be similar to the one shown in Figure 1.

Fig. 2. Bit rate of a 'lossless coder' (uncompressed bit rate: 705.6 kbit/s).

CDV with background traffic

Here the problem is similar, except that now the amplitude may reach hundreds of microseconds. If the jitter and wander requirements of ITU-T Rec. G.823 for a 64-kbit/s connection are to be met using the adaptive clock recovery method, a PLL with $\omega_n \approx 10^{-6}$ will have to be realized, and it is *highly unlikely* that such a PLL can be built. This would leave the SRTS as the only practicable source-clock recovery method in ATM, unless the CDV happens to be a zero mean uncorrelated process, in which case for speech a PLL with $\omega_n \approx 10^{-2}$ satisfying Rec. G.823 could be easily constructed. In any case, in order to attain a measure of certainty in this matter, CDV needs to be *much better characterized not only as a probability density function but also as a process.* As far as ATM is concerned, the requirements of Rec. G.823 could be relaxed. In order to be able to correctly dimension the PLL, subjective tests should be conducted to determine:

- the jitter amplitude threshold, where subjective impairments appear in function of the jitter frequency (e.g., a 500-μs_{p-p} sinusoidal jitter is certainly not audible, if it has a period of one day; $\Delta f/f = 5.9 \sim 10^{-9}$)
- the effects of isolated but large phase changes (the CDV passes from its minimum to its maximum value) as a function of the adaptation time

These results will very likely depend on the service. There is very little experimental data available in this area, and ITU-T, SG15/WP2 is planning some activities for the next study period to evaluate the effects of ATM on subjective speech and audio quality.

From a synchronization point of view, the most complex situation arises when a connection includes two PC-based terminals that use their own soundboard for A/D and D/A conversion. In this case, a buffer may be inserted to take into account the possible frequency deviation between the send and receive clocks as well as typical holding times for the service. If the resulting delay is not acceptable, then, to reduce it, the received signal may have to be resampled digitally.

VBR coding for ATM

The purpose of this chapter is not signal processing but the illustration of some of the difficulties associated with the transmission of VBR sources over ATM. For fixed-rate multimedia services, there are some possibilities to vary the transmission rate of each component (voice, video, data, etc.) separately through dynamic bit allocation to improve the overall QoS. The most recent ITU-T Rec. G.723.1 (Scaleable Channel Coder for Multimedia Communications) is a good example for doing this. The same thing may be done for the subflows of a single component, too (see e.g. [6]).

There are basically two (more or less royal) ways to VBR coding, both entailing greater complexity:

- silence or voice activity detection (VAD) for CBR sources
- entropy coding (removing the redundance and irrelevance from the source output; implicitly contains VAD)

The function of the VAD is to ensure that only useful information will be transmitted and silence (including background noise only) coded with a minimum number of bits. VAD in one form or another has been used for a long time in specialized equipment such as EC and DCME, but the terminal equipment based on G.723.1 is the first one to include VAD in order to reduce the per-source transmission rate.

Signal compression usually starts with uniformly quantized source samples. Most of the modern music codecs (e.g. those of the MPEG family) are perceptual coders exploiting the laws of psycho-acoustics: With this compression method, the just audible noise levels (noise masking thresholds) are calculated for each of the 26 so-called 'critical bands' of the useful frequency range ($f_s \leq 48$ kHz) to enable coding at a *constant quality* level. The associated bit-stream is necessarily VBR, and presently only AAL2 is designated to carry such *real-time* traffic. There are three main parameters that may be varied in order to obtain a scaleable codec and a variable QoS at the receiver:

- varying the number of 'critical bands'
- varying the noise margin in each 'critical band'
- varying the codec complexity

By combining the above possibilities, a large number of quality and information priority classes may be generated, in any case many more than ATM can handle on a cell basis only.

The codec output may be characterized by the average and the peak bit rate or the ratio between the peak and the average information rate called *burstiness.* The burstiness of the codec output depends on the information type, the signal content and on the codec.

The behaviour of the so-called 'lossless' coder enabling the bit-exact delivery of the uniformly quantized PCM signals is typical for VBR sources, and it will be used for illustrative purposes only. This coder is based on LPC analysis and Huffman coding i.e., only the redundance is removed from the signal but not the irrelevance. *Figure 2* [12] shows the variation of the average bit rate per frame for glockenspiel taken from the EBU SQAM compact disk (published by the European Broadcasting Union, Cat. Nr. 422 204-2): The original signal is sampled at 44.1 kHz, and an analysis frame contains 576

Sequence	Average bit rate [kbit/s]	Min. bit rate [kbit/s]	Peak bit rate [kbit/s]	Standard deviation [kbit/s]	Burstiness
SQAM 35 01 glockenspiel	302.19	96.12	609.31	126.64	2.02
SQAM 61 01 soprano	329.96	232.93	471.92	96.54	1.43
SQAM 65 01 orchestra	462.43	387.36	681.15	75.00	1.47
SQAM 66 01 wind ensemble	424.99	322.43	564.46	96.72	1.33
SQAM 67 01 wind ensemble	353.00	310.76	439.31	56.73	1.24

Table 2. Statistics of the lossless coder output for various signals.

samples (48 octets/cell is assumed) to yield ~77 frames/s.
Table 2 gives the statistics of various types of music. A perceptual coder can further reduce the average bit rate well below 100 kbit/s for most signals. Of the parameters listed in Table 2, only the peak cell rate is used for admission control in the first generation ATM systems. If call charges are also based on this parameter exclusively, then there will be little incentive to develop VBR coding for real-time applications and even less to use it. Presently, Rec. I.363.1 contains only a very sketchy definition of AAL2, and the discussion in Sec. 'Synchronization of VBR connection' below is just an attempt to resolve some of the difficulties native to VBR transmission.

Synchronization of VBR connections

In most speech, sound and video codec, a fixed number of uniformly quantized samples are collected to form a *frame,* which is then analyzed and compressed. Although the number of ATM cells per frame varies, the sampling frequency of the source can still be recovered from the regular *frame arrival times,* albeit only at a layer above AAL. When both sender and receiver have access to the same network clock, the synchronization problem is the same as for CBR services (the buffer must hold the maximum number of bits that can be generated in a coding frame); otherwise, neither the adaptive nor the SRTS method recommended for AAL1 may be used directly for VBR, since they both rely on a fixed sampling to cell emission rate ratio.
For VBR transmission, *any synchronization scheme must be tied to the periodic algorithmic frame structure.* At protocol layers above AAL, flags delimiting coding frames may be used to regenerate the sampling frequency by using the *frame-fill level* of the buffer much the same way as for CBR cells in Figure 1. If the timing recovery is to be done in the AAL, the timing information may be transmitted by controlling the cell emission times and using time windows in the receiver:

- The emission time of the first cell of a frame (SN = 0, if used) coincides with the start of that frame.
- The remaining cells are transmitted with fixed intercell time $T_0 > CDV$.
- A window of width $W_1 = T_0 + CDV + \delta$ can be used to detect one missing cell, except when it is the last one to be transmitted for the current frame.
- A window of width $W_2 = 2 \cdot T_0 + CDV + \delta$ can be used to signal the last cell transmitted for the current frame or the start of the next one. The same window may also be used to detect the presence of misinserted cell(s), when more than two cell arrivals are detected.

The various time windows are illustrated in *Figure 3.* In fact, the theoretical celltimes and the time windows must be derived from the less than ideal PLL output containing residual jitter. In the above equations, the parameter δ is chosen to account for this jitter and to provide for safety margins as well.
If *sufficiently stable* PLL can be built, it may not be necessary to transmit any information at all during short 'silent' periods (e.g. all digital 0), since missing frames could be detected and substituted for the purposes of synchronization, too. The smaller the CDV compared to T_0, the more robust the time-

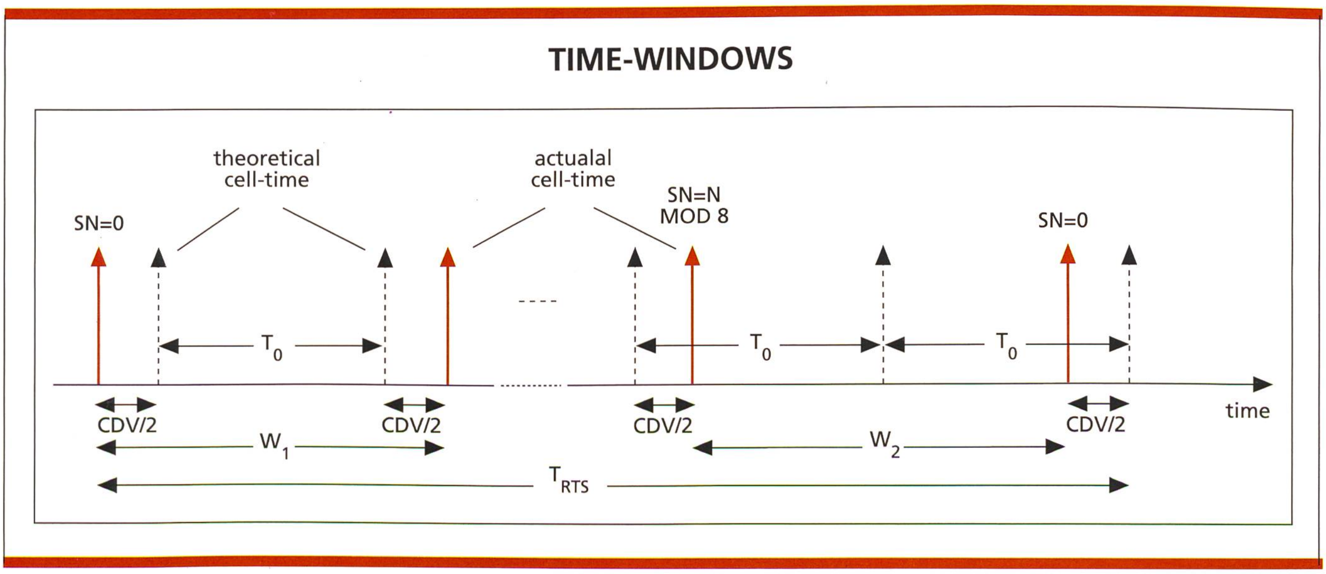

Fig. 3. Time windows for detecting cell errors and for synchronization.

window method becomes. The condition $T_0 > CDV$, however, may not be satisfied for channel rates above H_0.
Similarly, the SN can be used to detect cell errors and for synchronization (Fig. 3). When combined with the time-window technique, it renders the algorithm more robust, e.g., misinserted cells can be *identified.* A modified SRTS method may also be used for VBR, if the RTS period T_{RTS} coincides with the frame period and the RTS is transmitted – unlike for AAL1, where the 4-bit RTS is carried by the odd-SN cells – all at once. T_0 should be chosen as large as possible to reduce the peak cell rate: The condition $T_0 = T_{RTS}/(N_{max}+1)$ satisfies this requirement, while leaving the last theoretical cell position of the frame vacant, where N_{max} is the maximal number of cell emissions per frame, which is upperbounded by that of the uncompressed signal. FEC, as it was used in AAL1, cannot be used for VBR, since it would produce variable delay.
Although neither of these methods can detect whether the last cell of the frame is missing (this can be done on the algorithmic level), the synchronization of the terminal is ensured even in this case (this is far more important, since some terminals, such as modems, may take several seconds to resynchronize). If the bridging of the 'silent' intervals cannot be made sufficiently robust, then a minimum cell rate of 1 cell/frame (29.4 kbit/s in this example) must be maintained.
As a rule, the last cell of the frame is not completely filled, and an average free capacity of ~15 kbit/s is available to transmit ancillary data or, at the cost of an additional one-frame delay, to fill it with information bits from the following frame. Alternatively, depending on the coding margins, the last cell may also be either completed to increase quality or suppressed.

Summary and conclusions

There are some indications that, at least for real-time services, the transition from STM to ATM may not be as smooth as anticipated. Delay is one of the most critical parameters that must be controlled, since it may limit the choice of codecs and synchronization methods and make the introduction of EC mandatory even in the national network. The most generally applicable *adaptive* synchronization method is also the most problematic, due to the lack of reliable information about the true nature of CDV. Considerable amount of time and effort will have to be invested to ascertain that CDV in ATM does not unduly degrade the subjective quality available in STM networks.
It was argued that the role of signal processing (robust, scaleable codecs) is just as important as network capabilities – the two should ideally complement each other – to enable disparate terminals to interwork over heterogeneous networks and to increase network efficiency. Conversely, signal processing activity should be reoriented to produce codecs that are robust in an ATM- or, more generally, in a packet-based network environment. Layered VBR coding could play an important role in the development of statistical multiplexing techniques, but further research effort is needed in this area to develop low-delay codecs and packetization techniques for subflows that yield gradual QoS sensitivity to cell losses. To achieve these goals, however, ATM should ideally be able to handle more than just two CLP classes in order to get the most out of scaleable codecs. 8.4

References

[1] Technische Mitteilungen PTT 4/1993 (special edition on B-ISDN)
[2] ITU-T Draft Recommendation I.363.1, B-ISDN ATM Adaptation Layer (AAL) Specification, Types 1 and 2, SG13, 10–21 July 1995, TD.53(P)/COM 13-R 51
[3] E.-H. Göldner, ATM und Sprache, Siemens telecom report, 1/95 pp. 38–41
[4] M. de Prycker, Asynchronous Transfer Mode solution for Broadband ISDN, © Prentice Hall International (UK) Limited 1995
[5] ITU-T, SG15, Geneva, 13–24 November 1995, TD.96(2/15)
[6] R. Mury et al., Système de codage MICDA-SB pour signaux sonores de haute qualité, Technische Mitteilungen PTT 8/1991, pp. 320–327
[7] N. S. Jayant and S. W. Christensen, Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample Interpolation Procedure, IEEE Trans. Commun., vol. COM 29, pp. 101–109, February 1981
[8] J. Suzuki and M. Taka, Missing Packet Recovery Techniques for Low-Bitrate Coded Speech, IEEE J. Select. Areas in Commun., vol. 7, pp.707–717, June 1989
[9] N. Kitawaki et al., Speech Coding Technology for ATM Networks, IEEE Commun. Mag., January 1990, pp. 21–27
[10] M. Dietz et al., Audio Compression for Network Transmission, J. Audio Eng. Soc., vol. 44, pp. 58–72, January/February 1996
[11] D. Greaves and M. Taunton, ATM for Video and Audio on Demand, Audio Eng. Soc. UK Conference on Audio for New Media, 25–26 March 1996, pp. 52– 63
[12] Telecom Report FE 321.069 U, Report of Internship: VBR Lossless Audio Coding for Fixed- and Variable Rate (ATM) Transmission Channels, 28 August 1995

ZUSAMMENFASSUNG

Signalverarbeitung und -synchronisation für die Echtzeitübermittlung via ATM

Echtzeitdienste mit niedriger Bitrate wie Telefonie erzeugen immer noch 50 bis 80 % des gesamten Verkehrsaufkommens. Aus diesem Grund ist es wichtig, abschätzen zu können, wie gut diese Dienste unter Zeitdruck die ausserordentlich hohen Jitterwerte und die neuen Fehlermechanismen von ATM tolerieren können. Für den jeweiligen Dienst ist die Verzögerung massgeblich bei der Wahl von Codierung, Synchronisationsmethode, Fehlerüberwachung, Multiplexierung usw. Die Codierung und die Synchronisation für CBR- sowie VBR-Quellen werden besprochen, und die Rolle der Signalverarbeitung beim Erzielen der Interoperabilität für ungleiche Terminals und höhere Netzeffizienz wird betont.

Paul Vörös received a M. Sc. in Electrical Engineering from the University of Notre Dame, USA, in 1963. After graduating, he worked for nine years at Bell-Northern Research in Montreal as a member of the scientific staff on filters. Since 1972 he has been with the Swiss Telecom PTT Research Laboratories, where he is responsible for solving transmission problems and, during the last ten years, primarily for speech and audio coding.