INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.177

(09/99)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

International telephone connections and circuits – Transmission plan aspects of special circuits and connections using the international telephone connection network

---

Transmission planning for voiceband services over hybrid Internet/PSTN connections

ITU-T Recommendation G.177

(Previously CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
General definitions	G.100–G.109
General Recommendations on the transmission quality for an entire international telephone connection	G.110–G.119
General characteristics of national systems forming part of international connections	G.120–G.129
General characteristics of the 4-wire chain formed by the international circuits and national extension circuits	G.130–G.139
General characteristics of the 4-wire chain of international circuits; international transit	G.140–G.149
General characteristics of international telephone circuits and national extension circuits	G.150–G.159
Apparatus associated with long-distance telephone circuits	G.160–G.169
Transmission plan aspects of special circuits and connections using the international telephone connection network	**G.170–G.179**
Protection and restoration of transmission systems	G.180–G.189
Software tools for transmission systems	G.190–G.199
INTERNATIONAL ANALOGUE CARRIER SYSTEM	
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TESTING EQUIPMENTS	
TRANSMISSION MEDIA CHARACTERISTICS	G.600–G.699
DIGITAL TRANSMISSION SYSTEMS	
TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION G.177

TRANSMISSION PLANNING FOR VOICEBAND SERVICES OVER HYBRID INTERNET/PSTN CONNECTIONS

Summary

This Recommendation provides transmission performance guidelines for voiceband services over networks that are an interconnection of an IP network (e.g. "the Internet") and the PSTN. As modern telecommunications systems continue to evolve, hybrid networks of various kinds will become more prevalent. This tendency, together with the reality that voiceband communications will be an important source of traffic on such networks, has led ITU-T to develop this Recommendation. The intent is to provide to a wide audience a set of principles on transmission performance aspects of hybrid Internet/PSTN connections.

Source

ITU-T Recommendation G.177 was prepared by ITU-T Study Group 12 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on 30 September 1999.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the 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.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation the term *recognized operating agency (ROA)* includes any individual, company, corporation or governmental organization that operates a public correspondence service. The terms *Administration, ROA* and *public correspondence* are defined in the *Constitution of the ITU (Geneva, 1992)*.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Recommendation G.177

TRANSMISSION PLANNING FOR VOICEBAND SERVICES OVER HYBRID INTERNET/PSTN CONNECTIONS

(Geneva, 1999)

1 Scope, purpose and application

1.1 Scope

This Recommendation provides transmission planning guidelines for voiceband services over connections that include Internet and PSTN segments, i.e. hybrid Internet/PSTN connections. In particular, "voice over the Internet" or "Voice over Internet Protocol" (VoIP) connections are considered, where at least part of the connection is over an Internet Protocol (IP) network. PSTN-only connections (i.e. connections that begin, end, and are carried only on the PSTN) are considered as the reference against which the performance of other connections is compared.

This Recommendation provides end-to-end transmission performance guidance for hybrid Internet/PSTN connections carrying voiceband services. When one of the terminals is attached to an IP network, it is assumed that this terminal has functionality defined in Recommendation H.323. (The terminal need not be an H.323 terminal, but is assumed to have that functionality.) To support these end-to-end guidelines, guidance is given for interworking functions between Internet and PSTN. The functionality provided by Internet/PSTN gateways presents special performance issues that must be addressed to ensure high quality service. Protocol details of IP and higher level protocols (e.g. TCP, UDP, RTP, RSVP) are not considered in this Recommendation. Rather, the impact on performance of voiceband services of the packetized aspects of IP networks, such as delay associated with packet assembly, lost or discarded packets, is of primary interest here.

The importance of user expectations for voiceband services on hybrid Internet/PSTN networks is recognized. However, this Recommendation does not address this subject, nor does it offer detailed suggestions of how to assess the impact of user expectations on the acceptability of these services. Additional guidance on many of these issues is available in ITU-T P-series and G.100-series Recommendations.

This Recommendation also includes discussion of issues associated with the interworking functions provided by Internet/PSTN gateways.

1.2 Purpose

This Recommendation provides guidance on basic telephony issues involved in providing voiceband services over hybrid Internet/PSTN connections. This information will also be useful to equipment manufacturers, VoIP service providers, Internet Service Providers (ISPs), software developers, and PSTN providers where service goes over a gateway to IP networks. Traditional PSTN engineers will be familiar with the telephony aspects of this Recommendation, but will find useful the material on Internet-specific features of voiceband services over connections that incorporate an Internet segment.

1.3 Application

It is intended that the guidelines provided in this Recommendation be applied by equipment manufacturers, VoIP service providers, ISPs, software developers, and PSTN providers. Those who provide services where the PSTN interconnects with IP networks through a gateway should find this information useful as they design equipment, software and services for voiceband services over hybrid Internet/PSTN connections. Application of the guidelines herein by equipment manufacturers, VoIP service providers, Internet Service Providers, software developers, and PSTN providers should enhance the quality of voiceband services delivered over hybrid Internet/PSTN connections. Furthermore, if widespread customer acceptance of VoIP service is desired, the resulting voiceband service quality should approach that of the PSTN.

The guidance in this Recommendation applies to situations/connections where at least part of the connection is on an IP network. Special attention focussed on the "Gateway/Inter-Working Function" (which provides the essential interworking capabilities of these hybrid connections). A major application of this Recommendation will be Internet Telephony where the terminal (e.g. a PC) enjoys H.323 stack functionality. As noted above, this Recommendation will use PSTN-only connections as the reference against which the performance of other connections will be compared.

This Recommendation applies to connections where the data transfer phase occurs over hybrid Internet/PSTN connections. Issues of call set-up, number translation (between PSTN and IP numbering), etc., are not considered in this Recommendation.

This Recommendation applies to IP networks that provide voice telephony in accordance with any of the scenarios described in clause 5, Performance of hybrid Internet/PSTN connections. This Recommendation contains general information on end-to-end quality and the way in which quality is affected by various components in the VoIP system. A description of the relationship of the performance of terminals and of the network is also included.

2 Terminology

This Recommendation defines the following terms:

2.1 hybrid PSTN/Internet: A connection that includes at least one segment where traffic is carried on the PSTN and at least one segment where traffic is carried on a network that uses the Internet Protocol suite.

2.2 gateway/IWF: A connection element that interconnects different networks and performs the necessary translation between the protocols used on those networks.

2.3 interwork: The ability of two networks to be connected and transfer traffic from one to the other.

2.4 H.323-based terminal: A terminal that is either dedicated (e.g. a telephone set) or general purpose (e.g. a computer running an application that performs the terminal function) and that:

– is intended for connection to an IP network;

– provides the functionality defined in Recommendation H.323.

3 Acronyms and Abbreviations

This Recommendation uses the following abbreviations:

%GoB	Per cent Good-or-Better
%PoW	Per cent Poor-or-Worse
ACELP	Algebraic-Code-Excited Linear-Prediction
ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
ADSL	Asymmetric Digital Subscriber Line
ATM	Asynchronous Transfer Mode

BER	Bit Error Ratio
CNG	Comfort Noise Generator
CS-ACELP	Conjugate Structure Algebraic-Code-Excited Linear-Prediction
dB(A)	dB SPL, A-weighted
dBm	decibels referred to 1 milliwatt
dBm0p	dBm as measured at the zero dBr point, weighted psophometrically
dBmp	dBm, psophometrically weighted
DTMF	Dual-Tone Multi-Frequency
EC	Echo Canceller
eif	E-Model equipment impairment factor
ERL	Echo Return Loss
GSM	Global System for Mobile communications
GSM EFR	GSM Enhanced Full Rate Speech Coder
GSM FR	GSM Full Rate Speech Coder
GSM HR	GSM Half Rate Speech Coder
GW	Gateway
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
IWF	Interworking function
LAN	Local Area Network
LD-CELP	Low-Delay Code-Excited Linear Prediction
PCME	Packetized Circuit Multiplication Equipment
PPP	Point-to-Point Protocol
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RLR	Receive Loudness Rating
RSVP	Resource Reservation Set-Up Protocol
RTP	Real-Time Transport Protocol
SCN	Switched Communications Network
SLR	Send Loudness Rating
SPL	Sound Pressure Level
STU	Secure Telephone Unit
TCL_W	Weighted Terminal Coupling Loss
TCP	Transmission Control Protocol
TDD	Telecommunications Terminal for the Deaf
TTY	Teletype

UDP	User Datagram Protocol
VBD	Voiceband Data
VDSL	Very High Speed Digital Subscriber Line
VoIP	Voice over Internet Protocol
VTOA	Voice and Telephony over ATM
xDSL	ADSL, VDSL and other Digital Subscriber Line Techniques

4 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

– ITU-T Recommendation G.101 (1996), *The transmission plan*.

– ITU-T Recommendation G.107 (1998), *The E-model, a computational model for use in transmission planning.*

– ITU-T Recommendation G.108 (1999), *Application of the E-model – A planning guide*.

– ITU-T Recommendation G.109 (1999), *Definition of categories of speech transmission quality*.

– ITU-T Recommendation G.113 (1996), *Transmission impairments*.

– ITU-T Recommendation G.114 (1996), *One-way transmission time*.

– ITU-T Recommendation G.116 (1999), *Transmission performance objectives applicable to end-to-end international connections*.

– ITU-T Recommendation G.131 (1996), *Control of talker echo*.

– ITU-T Recommendation G.168 (1997), *Digital network echo cancellers*.

– ITU-T Recommendation G.173 (1993), *Transmission planning aspects of the speech service in digital public land mobile networks*.

– ITU-T Recommendation G.174 (1994), *Transmission performance objectives for terrestrial digital wireless systems using portable terminals to access the PSTN*.

– ITU-T Recommendation G.175 (1997*), Transmission planning for private/public network interconnection of voice traffic*.

– ITU-T Recommendation G.176 (1997), *Planning guidelines for the integration of ATM technology into networks supporting voiceband services*.

– CCITT Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

– ITU-T Recommendation G.723.1 (1996), *Speech coders: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*.

– CCITT Recommendation G.726 (1990), *40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)*.

– CCITT Recommendation G.728 (1992), *Coding of speech at 16 kbit/s using low-delay code excited linear prediction.*

– ITU-T Recommendation G.729 (1996), *Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP).*

– ITU-T Recommendation G.729 Annex A (1996), *Reduced complexity 8 kbit/s CS-ACELP speech codec.*

– CCITT Recommendation G.764 (1990), *Voice packetization – Packetized voice protocols*.

– ITU-T Recommendation H.225.0 (1998), *Call signalling protocols and media stream packetization for packet-based multimedia communication systems*.

– ITU-T Recommendation H.245 (1998), *Control protocol for multimedia communication*.

– ITU-T Recommendation H.323 (1998), *Packet-based multimedia communications systems.*

– ITU-T Recommendation P.11 (1993), *Effect of transmission impairments*.

– ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level*.

– ITU-T Recommendation P.79 (1993), *Calculation of loudness ratings for telephone sets*.

– CCITT Recommendation P.82 (1988), *Method for evaluation of service from the standpoint of speech transmission quality*.

– ITU-T Recommendation P.310 (1996), *Transmission characteristics for telephone-band (300-3400 Hz) digital telephones*.

– ITU-T Recommendation P.561 (1996), *In-service, Non-intrusive measurement device – Voice service measurements*.

– ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

– ITU-T Recommendation P.861 (1998), *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs*.

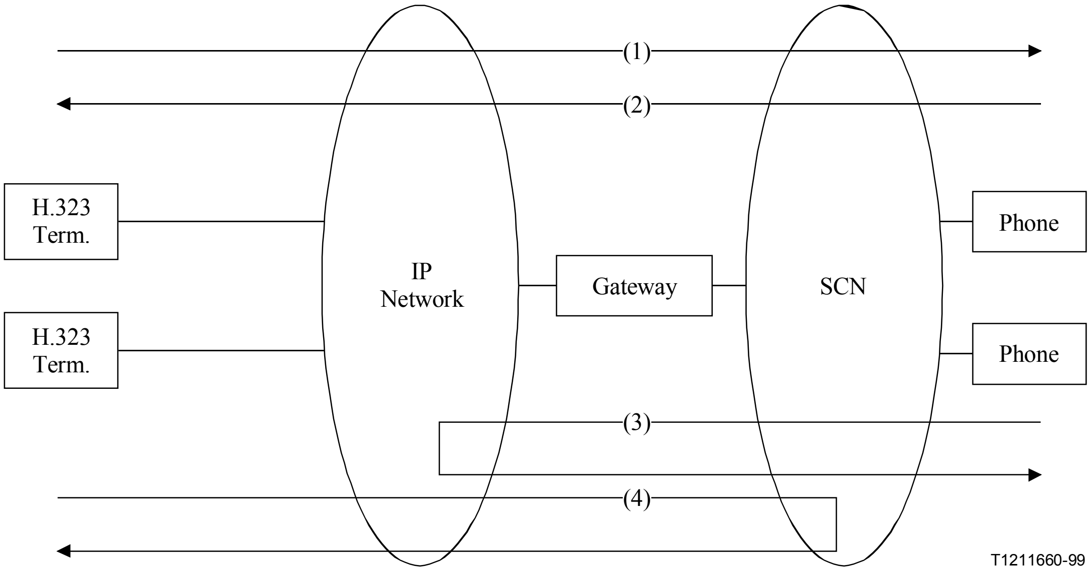
– RFC 1889, *RTP: A Transport Protocol for Real-Time Applications*.

– RFC 2205, *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*.

– RFC 2212, *Specification of Guaranteed Quality of Service*.

5 Performance of hybrid Internet/PSTN connections

Figure 1 shows specific Internet/PSTN connection arrangements that are of interest in this Recommendation. The terminals attached to the Internet are assumed to have H.323 functionality from the point of view of speech transmission. These terminals may be connected to the Internet via a direct connection (e.g. Ethernet, Token Ring, etc.) or a dial-up connection (e.g. modem and PPP link). The Internet and PSTN sections are connected through a gateway. For convenience, this gateway is designated with a single box in Figure 1. In practice, the gateway may be composed of multiple pieces of equipment, each with specialized functions. Performance aspects of the gateway are of particular interest here.



NOTE – Four connection types are shown. See text for details.

Figure 1/G.177 – Specific connection types of interest in this Recommendation

The specific functions of the gateway will depend on whether the direction of transmission is from the Internet to the PSTN or vice versa. In particular, the functions in the gateway include (but are not limited to):

- Internet → PSTN
  - Packet disassembly (including "IP stack")
  - Speech decoder (including error concealment, comfort noise, silence insertion, etc.)
  - Management or regulation of delay variation
  - Echo cancellation.
- PSTN → Internet
  - Speech encoder (including silence removal, comfort noise decisions, etc.)
  - Packet assembly (including "IP stack")
  - Management or regulation of delay variation.

Four connection arrangements are considered in this Recommendation, each of which is shown in Figure 1. They are:

1) H.323 → Phone (H.323 → Internet → PSTN → Phone)
2) Phone → H.323 (Phone → PSTN → Internet → H.323)
3) Phone → Phone (Phone → PSTN → Internet → PSTN → Phone)
4) H.323 → H.323 (H.323 → Internet → PSTN → Internet → H.323)

Each of these connection arrangements requires at least one use of the gateway. Hence, connections that are purely PSTN-based or are strictly H.323-to-H.323 using only the Internet are not the subject of this Recommendation.

- This Recommendation describes factors that play a role in determining end-to-end QoS and the parameters by which QoS is characterized.

6 Transmission channel characteristics of IP networks

In many respects, transmission in IP networks is similar to transmission of voiceband services over packetized systems and networks. Appendix I/G.764 gives an introduction to packetized systems and the performance issues associated with them.

Transmission in IP networks is accomplished by assembling multiple bytes into packets. These packets include headers with essential information such as source and destination for the packet, which is added at the transport and network layers. The size of the payload and of the header may vary, depending on the application and on the nature of the protocols being used. Communication between two endpoints is via a "data stream" (somewhat analogous to a connection in a connection-oriented network such as the PSTN), which usually consists of multiple packets. Each of the packets from a given source may take different routes to a given destination. As a consequence, packets from a given data stream may arrive at the destination in an order that is different from the order in which they were transmitted (i.e. the packets arrive out of sequence). Protocol features such as sequence numbers or time stamps allow the packets to be reassembled in the proper order at the destination.

Proper sequencing of the packets in the data stream is the responsibility of the higher layers in the protocol stack and is not part of the IP protocol. When an application requires the arriving packets to be used in proper sequence, sufficient delay must be included to allow integration of late packets. For applications such as speech communications, where end-to-end delays as short as possible must be maintained, it may be necessary to declare "very late" packets as lost in order to achieve acceptable delay. The tradeoff between long delay (which may result in higher speech transmission quality, but will increase the difficulty of conversation) and dropped packets (which will result in lower speech transmission quality, but will ease the ability to have an interactive conversation) must be given careful consideration when designing VoIP services.

7 Speech performance (end-to-end)

Many factors influence the overall end-to-end speech quality of VoIP services. These factors include choice of speech coder, delay (and any associated echo), overall level (loudness), etc. Important factors are considered, in turn, in this clause. The widespread use of digital wireless terminals (and associated speech coding) in the PSTN, and the interaction of wireless systems with speech processing elements in VoIP systems is an area for further study.

A useful tool for assessing the relative impact of transmission planning decisions on speech performance is the E-Model. The E-Model has been included in various ITU-T Recommendations on transmission planning. In particular, the E-Model is the subject of Recommendation G.107.

The E-Model introduces the notion of "equipment impairment factor" (eif), which has shown great utility in capturing the effects on transmission performance of new speech signal processing devices. Recommendation G.113 outlines the eif methodology.

Recommendation G.109 provides useful guidance on categories of voice quality, using E-Model R-values, and their use in speech communications.

Overall assessment of the speech transmission quality of a hybrid Internet/PSTN service is advisable. While it is important to assess the speech quality of various system components (e.g. speech coders, echo cancellers), the cumulative effects of impairments from multiple speech signal processing devices will be the limiting factor in determining user acceptability of a new service. Recommendation P.800 provides general guidance on conducting subjective evaluations of speech transmission quality. Recommendation P.830 provides detailed guidance for evaluation of speech codecs.

Objective measures of speech quality may also be appropriate. Recommendation P.861 describes one such objective measurement algorithm for evaluation of speech codecs. Application of

Recommendation P.861 for other purposes should be made with caution. The scope of Recommendation P.861 must be consulted for guidance.

NOTE – ITU-T is studying the issue of what new or modified objective techniques can be applied to situations relevant to VoIP. In particular, Appendix I/P.861 provides a method that is quite robust with respect to frame erasure.

7.1 Speech coder performance

Transmission in IP networks is characterized by periods of relatively error-free transmission, punctuated by occasional bursts of lost frames. These lost frames occur when packets are lost in transit or are discarded due to late arrival as part of a delay management strategy at the destination. Thus, speech coders for VoIP applications must be evaluated in terms of their performance when multiple sequential frames are lost or discarded. Some VoIP applications are also likely to encounter high ambient acoustic noise at the sending end of the connection (e.g. when one of the endpoints is a personal computer). Hence, it is also desirable that the speech coder shows robust performance in the presence of acoustic background noise.

The performance effect caused by lost or missing packets at the destination, and the associated effects on low bit rate speech coders, is of particular concern. In many applications, it will be necessary to include multiple frames of coded speech in a single IP packet. Thus, speech coders for VoIP applications should be robust with respect to loss of multiple sequential frames. Accordingly, it is recommended that one of the G.700-series codecs be used. These coders have been tested extensively and have been shown to have good quality and to be robust in their performance under a wide range of conditions. In particular, G.723.1, G.728, and G.729 have been evaluated for their performance under conditions that lead to random frame loss as well as "bursty" frame loss. However, they have not been evaluated thoroughly for one effect that may be peculiar to IP transmission, viz., the effect of loss of multiple sequential frames, either in isolation or in "bursty" situations. H.323 makes specific provisions for G.711, G.723.1, G.728, and G.729 (including Annexes A and B).

Opinion tests of the coder that is selected for VoIP applications should indicate that (individually) its overall quality under error-free conditions should be no worse than that of 32 kbit/s ADPCM (Recommendation G.726) coder. The 8 kbit/s ITU-T CS-ACELP coder (G.729 and Annex A/G.729) satisfies this requirement, although its interworking performance with other coding technologies is different from that of Recommendation G.726. For this reason, the provisional eif assigned to G.729 and Annex A/G.729 is higher than the established eif for G.726 at 32 kbit/s.

Under conditions of bursty 3% missing frames, it is recommended that the proportion of Poor-or-Worse judgements should not increase by more than 0.1 relative to that of 32-kbit/s ADPCM for the error-free case. Similar requirements for loss of multiple sequential frames, as noted above, have not been determined.

The use of speech coders that are already in wide use in digital wireless systems (or newer coders that are under consideration for such applications) should be considered for use in VoIP systems. While the transmission characteristics of wireless channels and those of IP systems may differ, speech coders for wireless applications are designed to be robust with respect to coded frames that have been corrupted, declared unusable, and discarded. Hence, they exhibit properties that are desirable in a speech coder for VoIP applications. Moreover, since it can be expected that an increasing number of voice calls will include an IP/PSTN gateway and a wireless terminal, it is desirable that both systems should be using the same speech coder. In particular, it is recommended that, when possible, only one encode/decode of speech be performed. To achieve this, it will be necessary to establish appropriate methods for communicating the type of codec in use so that the IP/PSTN gateway does not do speech encoding or decoding in these situations. The objective is to eliminate tandem processing by speech coders whenever possible.

7.1.1 E-Model equipment impairment factors (eifs)

The subjective performance of a speech coder is conveniently captured by the Equipment Impairment Factor (eif) of the E-Model (see Recommendation G.113 for a summary of these eifs and methods for extracting eifs from subjective data). E-Model eifs are empirically (subjectively) determined and provide a useful metric for comparing speech coders on a scale of subjective performance (i.e. speech quality). The eif for a given codec captures the subjective effect of speech impairments that are introduced by the speech codec. Performance of multiple codecs (whether of the same or of different type) arranged in tandem is an important consideration for the overall speech quality of a connection. The eif also captures the effects of such arrangements. All other things being equal, a lower eif is better than a higher eif. For speech coders with equal eifs, other considerations (such as delay or bit rate) may play decisive roles in choice of a speech coder. A current list of eif assignments for many speech codecs, including those that are available as options in Recommendation H.323, is found in Appendix I/G.113.

The processing delay associated with a codec is also captured in the E-Model, but is not part of the eif.

7.1.2 Tandeming performance

In some calling scenarios, the speech signal will be subjected to encodings by more than one speech codec. An obvious example is the H.323 → H.323 example illustrated in Figure 1. The transmission path passes through two gateways. Hence, speech will be encoded and decoded two times by a speech codec. Other important cases will be calls where one of the voice terminals is a wireless terminal, digital answering machine, or digital voice mail system. In these cases, the ability of the VoIP codec to operate well (from a subjective standpoint) in tandem with the other codec must be given proper consideration.

In general, tandeming of speech codecs leads to an overall degradation in speech quality. If equipment impairment factors are available, the E-Model (Recommendation G.107) may be used to assess the potential impact of multiple speech codecs in the transmission path.

Use of a G.711 codec in the VoIP terminal will lead to toll quality results on the PSTN and ISDN and normal digital wireless system performance when terminated in a digital wireless network.

7.1.3 Other considerations for speech coders

As noted above, use of speech codecs has a number of possible negative effects on voice services. Designers of VoIP services using speech coders should consider the requirements of such services as they build their networks. In particular, the effects of speech coding on the following applications should be considered:

- Speaker Recognition – Speech codecs should not have an adverse affect on the ability of humans or machines to recognize and identify a talker.
- Speech Recognition – Speech codecs should not have an adverse affect on the ability of humans or machines to recognize speech content.
- Text-to-Speech (TTS) – The artificial speech generated by speech production systems may not have the same characteristics as real speech. Since modern speech codecs exploit such characteristics, it is possible that coding of artificial speech may generate unexpected impairments.

7.2 Effects of transmission errors and packet loss

For VoIP applications, two primary types of transmission errors may occur:

1) bit errors on the transmission facility; and
2) lost or discarded IP packets.

7.2.1 Bit errors

Bit errors may occur on transmission facilities. As a result, IP packets arriving at a destination may be corrupted. The IP layer will detect corrupted IP headers, but will not detect errors that occur in the payload of the IP packet, i.e. the coded speech. Higher level protocols must detect corrupted speech frames. Whether such detection and correction is handled at the Transport Layer or at the Application Layer, and how detection and correction of errors is achieved, may ultimately affect the overall quality of speech transmission.

7.2.2 Packet loss

A potentially more severe performance issue is that of lost or discarded IP packets. An IP packet may be lost due to congestion in the IP network. An IP packet may also be discarded at the destination. This would occur, for example, when a packet is sufficiently late that the destination declares the packet lost. Discarding packets that are very late is preferable to having the destination increase delay and delay variation by potentially waiting for long time periods (see 7.4, Delay and echo).

Loss of a single IP packet will result in loss of one or more coded speech frames, depending on the speech coder in use and the number of frames per packet. As noted above, the speech coder should be robust with respect to loss of coded frames. In particular, if multiple frames are assembled into an IP packet, the performance of the speech coder must be assessed under frame loss conditions that reflect those of the network in use. The impact of large percentages of frame loss (as much as 20-30%) should be assessed.

7.3 Loudness ratings and terminal characteristics

Two important characteristics of telephone sets are the Send Loudness Rating (SLR) and the Receive Loudness Rating (RLR), as defined in Recommendation P.79. SLR and RLR define, respectively, the acoustic-to-electric efficiency of the transmitter and the electric-to-acoustic efficiency of the receiver. Transmitters and receivers for VoIP applications may not be typical telephone voice terminals. Other possibilities include headsets with microphone, stand-alone microphones together with separate loudspeakers, and built-in microphone and speakers (as available in laptop computers).

Handsets and headsets provide specified means to control input and output levels. Usually the frequency characteristics are also well suited for telephony. Acoustic echo is also less of a problem since the acoustic coupling loss is generally over 50 dB. Handsets and headsets typically provide significantly greater background noise rejection than stand-alone microphones. When stand-alone microphones and speakers are used in hands-free situations, the performance is highly dependent on several factors, including the linearity of the equipment and their positioning. The acoustic coupling also need proper echo-control in the form of half-duplex switching solutions or full-duplex echo cancellation solutions. The echo canceller must cope with background noise (e.g. office environment) and double-talk conditions (when users speak at the same time), and cancel the echo in single-talk (normal working) conditions. Poor echo performance mainly affects the user at the other end of the connection.

The sending and receiving frequency response of microphones, loudspeakers, ear-pieces and headsets should be matched to the audio bandwidth used. For narrow-band telephony the bandwidth should be [300-3400] Hz with a flat frequency response (within ±3 dB).

While Loudness Ratings of non-traditional voice terminals are, in principle, easy to specify, the measurement procedures to verify compliance have not been defined. General guidance can be provided. Active speech levels in the network have an average of about −20 dBm, with a standard deviation of about 5 dBm. These values usually produce acoustic levels at the output of a "typical" telephone receiver that are centred at about 80 dB SPL. A "typical" talker should produce an active speech level (as measured according to Recommendation P.56) of about –20 dBm at the 0 dBr point, which should produce an acoustic level of about 80 dB SPL at the output of the receiver.

When traditional telephone sets or handsets are used, they should meet the requirements in Recommendation P.310. New work is under way in ITU-T to define performance characteristics for a digital voice terminal intended for use in packetized systems.

7.4 Delay and echo

Extremely long transmission times will result in difficulty in conducting interactive conversations. As delay increases, interaction between users will degenerate to "push to talk" or "over and out" style of conversation. If widespread acceptance and use of IP-based voice services is desired, these situations should be minimized.

7.4.1 Delay

In modern digital telecommunication networks, delay is a key performance parameter whose increase should be minimized. Although the delay of IP networks may exceed the typical delay of the PSTN, the degradation caused by additional delay might be compensated for by benefits provided by new network and service capabilities. These tradeoffs need to be quantified.

Delay can have two effects on voice performance. Firstly, it increases the subjective effect of any echo impairment. Secondly, as indicated in Recommendation G.114, even when echo is controlled, one-way delays above 150 ms can interfere with the dynamics of voice conversation, depending upon the type of conversation and degree of interaction. Recommendations G.114 and G.131 and Annex A/G.173 give additional information regarding effects of delay and echo.

In addition, delay can impair the performance of particular voiceband data applications, some applications being even more sensitive to the delay than voice applications. Total delay of hybrid Internet/PSTN networks should be limited, even with the use of echo control. Recommendation G.114 should be consulted for additional information.

7.4.1.1 Codec delay

Modern speech codecs operate on collections of speech samples known as frames. Each input frame is processed into a compressed frame. The coded speech frame is not generated until all speech samples in the input frame have been collected by the encoder. Thus, there is a delay of one frame before processing can begin. In addition many coders also look into the succeeding frame to improve compression efficiency. The length of this advance look is known as the look-ahead time of the coder. The time required to process an input frame is assumed to be the same as the frame length since efficient use of processor resources will be accomplished when an encoder/decoder pair (or multiple encoder/decoder pairs operating in parallel on multiple input streams) fully uses the available processing power (evenly distributed in the time domain). Thus, the delay through an encoder/decoder pair is normally assumed to be:

$$2 \times \text{frame size} + \text{look-ahead}$$

If the output facility is running at the same rate as the speech codec (e.g. an 8 kbit/s facility for G.729), then an additional frame of delay is incurred when clocking the compressed frame to the facility. Thus, the maximum delay attributable to codec-related processing in conventional systems (i.e. the PSTN) is:

$$3 \times \text{frame size} + \text{look-ahead}$$

If the output facility is an IP network, then the frame output by the encoder will instantaneously be dropped into an IP packet. The additional delay required for IP packet assembly and presentation to the underlying link layer will depend on the link layer. When the link layer is a LAN (e.g. Ethernet), this additional time may be quite small.

If multiple speech frames are grouped together into a single IP packet, further delay is added to the speech signal. This delay will be the duration of one extra speech frame for each additional speech frame added to the IP packet:

$$(N + 1) \times \text{frame size} + \text{look-ahead}$$

where N is the number of frames in each packet.

7.4.1.2 IP terminal buffering delay

Audio cards and telephone cards in PCs usually include large internal buffers, in order to provide a fixed rate interface to A/D and D/A converter and an asynchronous interface to the application layer.

Additionally, modems and network adapters use internal buffers to increase network access efficiency. They have been optimized for data transmission where delay is not a problem, but this optimization may not be appropriate for voice transmission where delay is a critical issue.

There are also software buffering delays. Application or device drivers can store large amounts of data in order to process them easily and efficiently or to manage the delay jitter in received packets.

7.4.1.3 H.323 packetization/buffering delays

Packetization delay may be introduced while packets are being constructed. Buffering delay may be introduced when they are being disassembled.

Packetization delay is the time required to collect enough information to fill a packet. When fixed length packets are used with a frame-oriented codec, packetization can introduce an additional delay if the packet length differs from the frame length of the codec.

Buffering delay is due to queuing in the receiver and is usually used to compensate for network jitter. Voice playback requires equally spaced (in time) packets but network delays are variable, thus the receiver must delay packets that arrive early to synchronize them with those arriving later.

7.4.1.4 Network transmission delays

Transmission delay is the time spent by packets to reach their destination during transmission through the network. Components of network delay include:

– the transmission delay, introduced by sending a packet over a link (e.g. sending a 256 byte packet over a 64 kbit/s link takes 32 ms);
– the propagation delay, due to signal propagation over physical link. This delay is usually negligible if links are shorter than 1000 km;
– the node delay, due to router queuing and processing of packets;
– the protocol delay, due to packet retransmissions (if used, e.g. for TCP) or network access (e.g. CSMA-CD for Ethernet);
– gateway delay, introduced by interconnection between networks (e.g. packet disassembly/assembly and speech coding/decoding).

Network transmission delays may be negligible in fixed SCNs. However, significant transmission delays may be encountered in data networks (e.g. modem links or IP networks).

7.4.2 Delay variation

Packetized transmission systems exhibit variable delay in packet delivery times. Delay variation may have a negative impact on speech transmission quality. Depending on the nature of delay variations, the result may be experienced as time warping in speech or as impairments associated with lost speech packets.

Delay variations especially affect the performance of modems with auto-ranging echo cancellers. Accordingly, a guideline for limiting delay variation is desirable and is for further study.

7.4.3 Echo control

Current assessment of delay in hybrid Internet/PSTN connections indicates that echo control is required for all types of calls. Control of echo from the PSTN should be provided in the gateway between the IP network and the PSTN.

The H.323-based terminal on the IP network should control echo from that terminal. Echo from a four-wire H.323-based terminal will be primarily acoustic.

7.4.3.1 Echo from H.323-based terminals

Figure 2 illustrates the echo path that may arise at the H.323-based terminal. When the terminal is using a microphone and loudspeaker as the transmitter and receiver, the echo will be due to acoustic coupling between transmitter and receiver. The existing PSTN infrastructure probably will not provide adequate echo protection if the acoustic coupling loss in the terminal is too low and the delay is too high. Recommendation H.225.0 indicates that control of acoustic echo from the H.323 terminal is the responsibility of the terminal. In order to provide echo protection all H.323-based terminals should meet the Weighted Terminal Coupling Loss (TCL_W) objective of 45 dB, as is specified for digital wireline terminals in Recommendation P.310. Such acoustic isolation may be achieved relatively easily in standard handset terminals by careful design. However, in hands-free operation (e.g. microphone and speaker), other more complex techniques may have to be used. For example, introduction of advanced echo control technology capable of increasing acoustic isolation in hands-free terminals may be needed (standard echo cancellers may not be capable of providing sufficient isolation in a non-linear acoustic environment).

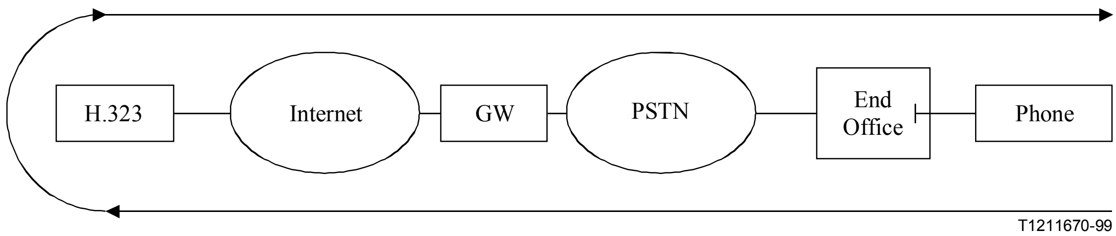


Figure 2/G.177 – Acoustic echo generated at the 4-wire H.323 terminal

7.4.3.2 Echo from the PSTN

Figure 3 illustrates the echo path that arises at the PSTN end of the connection due to a poor impedance match at the 4-to-2 wire conversion point. In this configuration an echo canceller is applied in the Interworking Function to control echo. The echo cancelling function may, in practice, be implemented at any location within the system. However, practical considerations (i.e. capabilities of existing echo cancellers) indicate that the appropriate location is in the gateway.

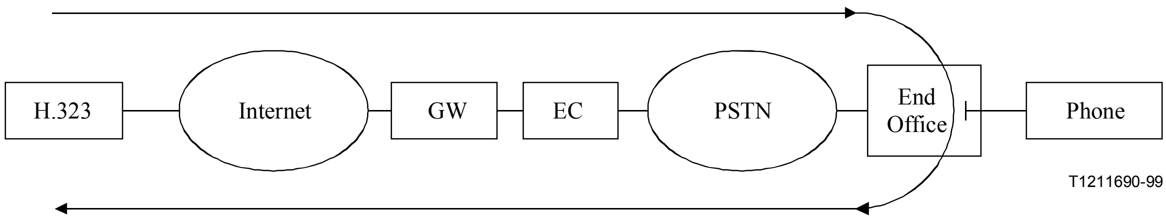


Figure 3/G.177 – Example of echo canceller deployment to control echo from the PSTN

Based on the existing PSTN infrastructure, the gateway should provide echo cancellation. It is likely that such echo cancellers will, in some configurations, be working in tandem with the PSTN echo control devices. This should not degrade the overall echo control function in the connection. In addition, effects of interaction of echo cancellers in the gateway with signal processing devices in the PSTN (e.g. PCME or conference bridges) is under study in ITU-T.

Recommendation G.168 provides specifications for digital network echo cancellers. At a minimum, echo cancellers deployed in the gateway should meet these requirements.

7.4.3.3 Voiceband data considerations

The issue of whether network echo control devices, such as echo cancellers, significantly degrade certain types of voiceband data transmission is unresolved. Additional information on this subject is available in Appendix I/G.168.

7.5 Temporal (syllable) clipping

Temporal speech clipping is the loss of speech signal at any time, and can occur when, for example, voice activity detection is used, when low bit rate coders change rate, or during protection switching and uncontrolled slips. The subjective impact of clipping depends on four factors: duration of clip; percentage of speech clipped; frequency of clipping; and overall speech activity. Based on the results of detailed subjective tests, two guidelines (specified in Recommendation G.116) to maintain good speech quality are:

– clipping of speech segments ≥ 64 ms should always be avoided; and

– clipped segments < 64 ms should be kept below 0.2 per cent of active speech.

7.6 Environmental (acoustic) noise

Acoustic noise at the transmitting end of a connection will have a negative impact on the performance of speech coders. The speech coders found in the G.72x-series have been tested for the effects of environmental noise at the sending end of the connection. However, some codecs have been tested more extensively than others. All have been shown to be fairly robust under conditions that included addition of circuit noise or speech babble to the input speech. If a particular type of background noise will be dominant in a given application, it is advisable to verify that performance of the speech codec is satisfactory under those conditions.

Pick-up of environmental noise with non-standard handsets may present special problems for the quality of speech when low bit rate codecs are used. In these situations, selection of a speech codec that is robust in the presence of acoustic background noise is especially important.

7.7 Idle channel noise

Idle channel noise in VoIP applications should be negligible. If present, however, background idle channel noise should be less than −68 dBm0p, a value consistent with Recommendation G.106.

7.8 Noise contrast and comfort noise

Noise contrast occurs when background noise is interrupted due to digital speech processing, such as echo cancellation using centre clippers, and voice activity detection (silence removal). Comfort noise is noise that can be introduced to mask the negative effects of noise contrast. Recommendations on noise contrast limits, and comfort noise values, are for future study.

For comfort noise insertion, some digital cellular systems (e.g. GSM) use an approach where noise parameters are extracted at the sending end and transmitted to the receiving end at a low bit rate. It is then possible to reconstruct (to good approximation) the background noise. This approach should provide superior subjective performance for voice users of circuits using voice activity detection and comfort noise insertion. The voice activity detectors and comfort noise generators described in Annex B/G.729 and Annex A/G.723.1 both operate in this fashion.

The best (subjective) performance will be realized when the noise inserted at the receiving end matches, as closely as possible, the background noise at the sending end. The following comments on CNGs can be made:

- the noise used should match the background noise, both in frequency content and level;
- level of the inserted noise should match that of the background noise; appropriate level measurements and adjustments should be done using dBm0p;
- the time course of changes in the level of the inserted noise should match, as closely as possible, the level changes that occur in the background noise.

7.9 Bandwidth

To maintain good speech quality and intelligibility, a minimum passband of 300-3400 Hz (3 dB points) should be delivered. For non-waveform coders, traditional measurement methods using single-frequency sine waves may not be adequate to evaluate effective bandwidth and level stability. At the date of this Recommendation, there are no industry-accepted methods to assess bandwidth of these non-linear systems.

7.10 Stability loss

For VoIP systems interfacing digitally to the PSTN, a minimum loss of 6 dB is recommended between the digital input and output paths of the VoIP system at the access port of the terminal. This guideline is provided to assure that singing does not occur when the handset terminal is used under conditions different from those to which the TCL_W measurement applies (e.g. placing a handset on a hard surface should not cause singing).

7.11 Distortion

Distortion in packetized systems such as Internet Telephony will be due primarily to the operation of speech codecs. It is essential that high quality speech codecs are used and that they have been thoroughly tested (subjectively) to ensure that there are no annoying effects.

8 Speech transmission planning in hybrid networks

The E-model (Recommendation G.107) is the recommended approach for use by network designers and transmission planners to describe and plan for handling the impairments that affect the quality of transmitted speech. This approach uses "the Equipment Impairment Factor Method" (Recommendation G.113), and is intended for use in a wide variety of transmission planning scenarios. Table 1 shows sample E-Model calculations for combinations of delay and equipment impairment factor. Recommendation G.108 provides general guidance for using the E-Model in transmission planning. Ie values in Table 1 are taken from Appendix I/G.113. Cells in the table are shaded according to the categories defined in Recommendation G.109.

Table 1/G.177 – R-values for indicated combinations of Ie and end-to-end mean one-way delay

Delay (ms)	Ie value								
	0	5	7	10	15	19	19	20	26
	G.711	GSM-EFR	G.726@32	G.729	G.723.1@6.3	G.729A+VAD w/ 2% loss	G.723.1@5.3	GSM-FR	G.729A +VAD w/ 4% loss
			G.728@16				G.723.1@6.3 +VAD w/ 1% loss	IS-54	
~0	94		87						
50	93		86	83		74			67
100	92	87	85	82	77	73	73	72	66
150	90	85	83	80	75	71	71	70	64
200	87	82	80	77	72	68	68	67	61
250	80	75	73	70	65	61	61	60	54
300	74	69	67	64	59	55	55	54	48
350	68	63	61	58	53	49	49	48	42
400	63	58	56	53	48	44	44	43	37
450	59	54	52	49	44	40	40	39	33

NOTE 1 – R-values in this table have been calculated using the indicated values for Ie and T (T=Ta=Tr/2) along with the default values from Table 3/G.107 for all other parameters.

NOTE 2 – Unless indicated otherwise, examples do not include packet loss or Voice Activity Detection (VAD).

NOTE 3 – Blackened cells indicate combinations of delay and codec that are impossible to realize.

9 Non-voice transmission (end-to-end)

There are a variety of user applications used on the PSTN that must continue to operate properly on hybrid connections. These include facsimile, encryption of voice and data (e.g. STU-III), ASCII file transfer, and use of special terminals. Packet loss and/or effects of low bit rate speech coding may limit the success of many of the more stringent applications. Difficulties associated with low bit rate coding require special care to ensure that these applications continue to operate at level that is satisfactory to end-users. These applications include:

- Fax – Proper fax transmission over hybrid Internet/PSTN connections will require special consideration. The role of the gateway will be central to this process. Each of the scenarios shown in Figure 1 must be considered.
- DTMF – DTMF signals from the VoIP terminals may be used to interact with DTMF-based services such as message retrieval. Thus, the VoIP system should support good end-to-end transmission of DTMF signals. Since some speech codecs will corrupt DTMF signals, special considerations may be necessary to ensure acceptable DTMF transmission. For an example of typical DTMF requirements, see ANSI/TIA/EIA/464-B-96.
- Call Progress Signals – It is also expected that call progress signals, such as audible ringback and busy, not be seriously degraded by the VoIP system. Detailed guidelines are for further study.
- TTY Devices and TDD Devices – The very low bit rate terminals are used on the PSTN and may find use on hybrid connections.

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems