INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.161

(03/2001)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

IPCablecom

Audio codec requirements for the provision of bidirectional audio service over cable television networks using cable modems

ITU-T Recommendation J.161

(Formerly CCITT Recommendation)

ITU-T J-SERIES RECOMMENDATIONS

CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

General Recommendations	J.1–J.9
General specifications for analogue sound-programme transmission	J.10–J.19
Performance characteristics of analogue sound-programme circuits	J.20–J.29
Equipment and lines used for analogue sound-programme circuits	J.30–J.39
Digital encoders for analogue sound-programme signals	J.40–J.49
Digital transmission of sound-programme signals	J.50–J.59
Circuits for analogue television transmission	J.60–J.69
Analogue television transmission over metallic lines and interconnection with radio-relay links	J.70–J.79
Digital transmission of television signals	J.80–J.89
Ancillary digital services for television transmission	J.90–J.99
Operational requirements and methods for television transmission	J.100–J.109
Interactive systems for digital television distribution	J.110–J.129
Transport of MPEG-2 signals on packetised networks	J.130–J.139
Measurement of the quality of service	J.140–J.149
Digital television distribution through local subscriber networks	J.150–J.159
IPCablecom	**J.160–J.179**
Miscellaneous	J.180–J.199
Application for Interactive Digital Television	J.200–J.209

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

ITU-T Recommendation J.161

Audio codec requirements for the provision of bidirectional audio service over cable television networks using cable modems

Summary

Many cable television operators are upgrading their facilities to provide two-way capability, and are using this capability to provide high-speed IP data services per ITU-T J.83 and ITU-T J.112. These operators now want to expand the capability of this delivery platform to include bidirectional voice communication and other time-critical services. This Recommendation is one of a series of Recommendations required to achieve this goal. It provides guidance on audio (voice) codec selection that will provide for a high-quality, interoperable service.

Source

ITU-T Recommendation J.161 was prepared by ITU-T Study Group 9 (2001-2004) and approved under the WTSA Resolution 1 procedure on 9 March 2001.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. 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 Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 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 expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. 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, 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

ITU-T Recommendation J.161

Audio codec requirements for the provision of bidirectional audio service over cable television networks using cable modems

1 Scope

This Recommendation specifies the audio (voice) codecs that are to be used in the provisioning of bidirectional audio services over cable television distribution networks using IP technology (i.e., IPCablecom service). The Recommendation also addresses codec options and packetization issues.

2 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 are valid. All Recommendations and other references are subject to revisions; 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 published regularly.

– ITU-T G.165 (1993), *Echo cancellers*.

– ITU-T G.168 (2000), *Digital network echo cancellers*.

– ITU-T G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

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

– ITU-T G.729 Annex E (1998), *11.8 kbit/s CS-ACELP speech coding algorithm*.

– ITU-T J.83 (1997), *Digital multi-programme systems for television, sound and data services for cable distribution*.

– ITU-T J.112 Annex A (2001), *Digital video broadcasting: DVB interaction channel for cable TV distribution systems*.

– ITU-T J.112 Annex B (2001), *Data-over-cable service interface specifications: Radio frequency interface specification*.

– ITU-T V.18 (2000), *Operational and interworking requirements for DCEs operating in the text telephone mode*.

– IETF RFC 1890 (1996), *RTP Profile for Audio and Video Conferences with Minimal Control*.

– IETF RFC 2327 (1998), *SDP: Session Description Protocol*.

NOTE – The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

3 Terms and definitions

This Recommendation defines the following terms:

3.1 access node: As used in this Recommendation, an Access Node is a layer two termination device that terminates the network end of the J.112 connection. It is technology specific. In Annex A/J.112 it is called the INA (interactive network adapter) while in Annex B it is called the CMTS (cable modem termination system).

3.2 cable modem: A cable modem is a layer two termination device that terminates the customer end of the J.112 connection.

3.3 IPCablecom: An ITU-T project that includes an architecture and a series of Recommendations that enable the delivery of real-time services over the cable television networks using cable modems.

3.4 MUST: The term **MUST** or **MUST NOT** is used as a convention in the present Recommendation to denote an absolutely mandatory aspect of the specification.

4 Abbreviations

This Recommendation uses the following abbreviations:

AN	Access Node
CPE	Customer Premises Equipment
DTMF	Dual-Tone Multi-Frequency
IP	Internet Protocol
IVR	Interactive Voice Response
MTA	Media Terminal Adaptor
PSTN	Public Switched Telephone Network
TDD	Telecommunications Device for the Deaf
VAD	Voice Activity Detection

5 Audio codec requirements

Offering a competitive and/or superior product requires support for more than high-quality delivery of audio. In addition to features and signalling capabilities, which are beyond the scope of this Recommendation, the audio codec application must provide transparent support for certain audio features. These include general detection mechanisms, DTMF, fax, analog modem, echo cancellation, and hearing-impaired support.

5.1 DTMF support

Dual-tone multi-frequency (DTMF) support allows employment of dual-tone multiple-frequency signals by either an autodialling system or through manual entry of tones. In order for DTMF tones to be captured correctly by the receiving device, tonal integrity (frequency accuracy and signal duration) must be maintained even through compression and transcoding.

MTA devices must successfully pass DTMF tone transmissions. The specified codecs must be capable of transparently passing these tones in-band.

5.2 Fax and modem support

IPCablecom needs to support analog fax and modem interfaces for two reasons. First, fax and modem equipment are common, and customers will continue to use these familiar devices for some years to come. Second, even with cable modem access, many users will continue to access their dial-up networks using a traditional modem.

In order to provide customers with access for analog fax and modems, Media Terminal Adapter (MTA) devices must be able to detect fax/modem signals and signal these detections using the appropriate protocol. The codec at each end is then switched to G.711 for the remainder of the session. Additionally, echo cancellation is disabled for the duration of the fax/modem session. After the fax/modem session has completed, echo cancellation is re-enabled.

A more robust solution for supporting fax is to employ fax relay. Fax relay involves demodulating the T.30 transmission and sending control and image data over the IP network. At the receiving end, the received data is remodulated and sent to the fax terminal using another T.30 session. This is described in ITU-T T.38. Client devices may employ fax relay.

5.3 Echo cancellation support

When end-to-end delay in an audio communication is more than 20 ms, an artifact called line echo can occur. This echo, if not removed, will be heard by the remote talker (thus it is also called talker echo) whenever he or she speaks.

Line echo cancellation must be provided in IPCablecom MTA and Gateway devices to mitigate the effects of line echo. This echo canceller must allow both parties to speak simultaneously (double-talk), so that one talker does not seize the line and block out the other user from being heard.

During periods when only the remote talker is speaking, the local echo canceller should either inject comfort noise or allow some noise to pass through to the remote talker, so that a "dead line" is not perceived. However, if local voice activity detection (VAD) is enabled, either the noise injection should be disabled, or the echo canceller should communicate its state with the VAD, in order for the VAD to not estimate the injected noise mistakenly as the true background noise.

In an application where the MTA is located in a home, the length of the echo canceller is typically short (eight ms or less). For PSTN gateway applications, the echo canceller length is typically much longer (32 ms or longer). Vendors may choose to differentiate their products by providing longer echo canceller lengths suitable for their application, or other programmable parameters.

In applications where a non-standard bidirectional audio interface is used (e.g., microphone and headset), echo cancellation may not be necessary. However, where a microphone and speakers are used, acoustic echo cancellation may be necessary, and vendors implementing these products should employ acoustic echo cancellation.

The performance of the line echo canceller should comply with either ITU-T G.165 or ITU-T G.168.

5.4 Asymmetrical services support

MTA devices should be capable of supporting employment of different codecs for upstream and downstream audio channels. This allows potential optimization of device resources, network bandwidth, and user service quality.

5.5 Hearing-impaired services support

For hearing-impaired people, TDD equipment can be the primary communication link to the outside world. This type of equipment has evolved lacking the type of standardization allowing broad interoperability among international manufacturers. The ITU recently adopted the V.18 standard to begin alleviating this problem. ITU-T V.18 outlines a procedure, which includes protocol negotiation for connecting these devices.

Since CPE for the hearing impaired consists of text input/output devices coupled with voiceband modems, any system designed to support them would need to be able to pass voiceband modem tones coherently. Of the list of proposed voice codecs, only G.711 would be able to achieve this, given that the other standards are not designed for passage of complex tones associated with modem communications. Typically, these devices will interface to the PSTN via an acoustical coupler to a phone or with a regular telephone jack.

MTA devices must support detection of ITU-T V.18 hearing-impaired tones, including V.18 Annexes A, B, and F. Support for Annex G/V.18 is optional. Upon detection of a V.18 signal, the codec at each end is then switched to G.711 for the remainder of the session. Additionally, echo cancellation is disabled for the duration of the V.18 call. It is optional to disable echo cancellation for Annex B/V.18 because it is DTMF-based. After the session has completed, echo cancellation is re-enabled.

6 Mandatory codecs

G.711 must be supported in all MTAs. This codec provides high-quality service and is ubiquitous. It provides the "fallback" position for services such as fax, modem, and hearing-impaired services support, as well as common gateway transcoding support. In addition, G.711 is used as the fallback mode if there are not enough resources to establish a new connection using the requested codec (e.g., two channels of G.728 or G.729 Annex E are already in existence, and there are not enough resources for a third connection to use a compressed codec).

6.1 μ-law and A-law support

Both G.711 encoding modes (μ-law and A-law) must be supported. If the analog-to-digital interface at the encoder uses A-law encoding, then A-law encoding must be selected for G.711 encoding. Otherwise, if the analog-to-digital interface at the encoder uses μ-law encoding, then μ-law encoding must be selected for G.711 encoding.

However, if one end of a voice connection uses an A-law interface, while the other end uses μ-law, then the A-law decoder MUST transcode the incoming μ-law packets to A-law, while the μ-law decoder MUST transcode the incoming A-law packets to μ-law.

6.2 Packet loss concealment

All media gateways and multimedia terminal adaptors MUST detect audio packet loss and implement some method to conceal losses from end-users. Specifications for low bit rate codecs (e.g. G.728, G.729) include methods for concealment. For G.711 implementations, the method defined in Appendix I/G.711 is recommended.

7 Additional codecs

In addition to G.711, MTAs also should support at least one of the following codecs.

7.1 G.728

G.728 should be supported in all MTAs. IPCablecom has a need to provide best or high voice quality. G.728 is a mid-bitrate (16 kbit/s), high-quality solution. Supporting a codec in this range provides high-quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications, such as IVR systems. In addition, it provides superior background noise handling.

7.2 G.729 Annex E

G.729 Annex E should be supported in all MTAs. IPCablecom has a need to provide best or high voice quality. G.729E is a mid-bitrate (11.8 kbit/s), high-quality solution. Supporting a codec in this range provides high-quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications, such as IVR systems. In addition, it provides superior background noise handling.

8 Optional features

8.1 Wideband codecs

Given that the majority of early customers will be "black phone" users, support for a wideband (i.e., greater than circuit voice bandwidth) codecs is not necessary. However, some vendors optionally may choose to differentiate their product by selecting components that will support higher fidelity in the event a wideband codec is provisioned at some time in the future.

8.2 Optional codecs

A vendor may supply any codec not described herein.

8.3 Voice activity detection (VAD)

A vendor may employ VAD to reduce bandwidth consumption. If employed, this capability must be optional, allowing disabling.

9 Packetization

Packet size influences both delay and the impact of packet loss. The larger the packet size, the greater the delay and the greater the impact of a lost packet. This suggests that the optimal packet size for voice applications is fairly small. Individual packets should not contain more than 20 ms of voice frames and must not contain more than 30 ms of voice frames. In addition, individual packets must contain an integral number of frames of sampling data, and any one frame of sampling data must be totally contained within one packet.

ANNEX A

Session description of CODECs

Session descriptor protocol (SDP) messages are used to describe multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation. SDP descriptions are used in both network call signalling (NCS) [ITU-T J.162] and distributed call signalling (DCS) [for future study]. This annex describes the required specification of the codec in SDP, and the required mapping of the SDP description into RSVP flowspecs.

A typical SDP description contains many fields that contain information regarding the session description (protocol version, session name, session attribute lines, etc.), the time description (time the session is active, etc.), and media description (media name and transport, media title, connection information, media attribute lines, etc.). The two critical components for specifying a codec in an SDP description are the media name and transport address (m) and the media attribute lines (a).

The media name and transport address (m) are of the form:

m=<media> <port> <transport> <fmt list>

The media attribute line(s) (a) are of the form:

a=<token>:<value>

A typical IP-delivered voice communication would be of the form:

m=audio 3456 RTP/AVP 0

a=ptime:10

On the transport address line (m), the first term defines the media type, which in the case of an IP voice communications session is audio. The second term defines the UDP port to which the media is sent (port 3456). The third term indicates that this stream is an RTP Audio/Video profile. Finally, the last term is the media payload type as defined in the RTP Audio/Video Profile (reference RFC 1890). In this case, the 0 represents a static payload type of μ-law PCM coded single channel audio sampled at 8 kHz (A value of 8 would be used to represent A-law). On the media attribute line (a), the first term defines the packet formation time (10 ms).

Payload types other than those defined in RFC 1890 are dynamically bound by using a dynamic payload type from the range 96-127, as defined in RFC 2327, and a media attribute line. For example, a typical SDP message for G.726 would be composed as follows:

m=audio 3456 RTP/AVP 96

a=rtpmap:96 G726-32/8000

The payload type 96 indicates that the payload type is locally defined for the duration of this session, and the following line indicates that payload type 96 is bound to the encoding "G726-32" with a clock rate of 8000 samples/s.

Codecs defined in this Recommendation MUST be encoded with the following string names in the rtpmap parameter:

Table 1/J.161 – Codec RTPMap parameters

Codec	RTPMap parameter
G.726 at 16 kbit/s	G726-16/8000
G.726 at 24 kbit/s	G726-24/8000
G.726 at 32 kbit/s	G726-32/8000
G.726 at 40 kbit/s	G726-40/8000
G.728	G728/8000
G.729A	G729A/8000
G.729E	G729E/8000

In addition, G.711 MAY be encoded with a dynamic payload type code, in an rtpmap parameter using the name PCMU/8000.

For every defined codec (whether it is represented in SDP as a static or dynamic payload type), Table 2 describes the mapping that MUST be used from either the payload type or ASCII string representation to the bandwidth requirements for that codec.

The Mapping of RTP/AVP code to RSVP Flowspec (as used by Dynamic Quality of Service [ITU-T J.163]) must be according to Table 2. The implementation of the 15-ms time period for G.711 codec is mandatory. All other implementations are optional.

Table 2/J.161 – Mapping of Session Description Parameters to RSVP Flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	**Rtpmap**	**Ptime (ms)**	**Values b, m, M (bytes)**	**Values r, p (byte/s)**	
0	<none>	9	112	12 444	G.711 using the Payload Type defined by IETF
0	<none>	10	120	12 000	
0	<none>	15	160	10 666	
0	<none>	20	200	10 000	
0	<none>	30	280	9 333	
96-127	PCMU/8000	9	112	12 444	G.711 PCM, 64 kbit/s, default CODEC
96-127	PCMU/8000	10	120	12 000	
96-127	PCMU/8000	15	160	10 666	
96-127	PCMU/8000	20	200	10 000	
96-127	PCMU/8000	30	280	9 333	
96-127	G726-16/8000	10	60	6 000	
96-127	G726-16/8000	20	80	4 000	
96-127	G726-16/8000	30	100	3 333	
96-127	G726-24/8000	10	70	7 000	
96-127	G726-24/8000	20	100	5 000	
96-127	G726-24/8000	30	130	4 333	
2	<none>	10	80	8 000	G.726-32, identical to G.721, which is assigned Payload Type 2 by IETF
2	<none>	20	120	6 000	
2	<none>	30	160	5 333	
96-127	G726-32/8000	10	80	8 000	
96-127	G726-32/8000	20	120	6 000	
96-127	G726-32/8000	30	160	5 333	
96-127	G726-40/8000	10	90	9 000	
96-127	G726-40/8000	20	140	7 000	
96-127	G726-40/8000	30	190	6 333	
15	<none>	10	60	6 000	G.728, assigned Payload Type 15 by IETF
15	<none>	15	70	4 666	
15	<none>	20	80	4 000	
15	<none>	30	100	3 333	
96-127	G728/8000	10	60	6 000	G.728, LD-CELP, 16 kbit/s
96-127	G728/8000	15	70	4 666	
96-127	G728/8000	20	80	4 000	
96-127	G728/8000	30	100	3 333	

Table 2/J.161 – Mapping of Session Description Parameters to RSVP Flowspec *(concluded)*

Parameters from session description			**Flowspec parameters**		**Comments**
RTP/AVP code	**Rtpmap**	**Ptime (ms)**	**Values b, m, M (bytes)**	**Values r, p (byte/s)**	
18	<none>	10	50	5 000	G.729A, identical to G.729, assigned Payload Type 18 by IETF
18	<none>	20	60	3 000	
18	<none>	30	70	2 333	
96-127	G729A/8000	10	50	5 000	G.729A, CS-ACELP, 8 kbit/s, 10 ms frame size with 5 ms lookahead
96-127	G729A/8000	20	60	3 000	
96-127	G729A/8000	30	70	2 333	
96-127	G729E/8000	10	55	5 500	G.729E, CS-ACELP, 11.8 kbit/s, 10 ms frame size with 5 ms lookahead
96-127	G729E/8000	20	70	3 500	
96-127	G729E/8000	30	85	2 833	
b bucket depth *m* minimum policed unit *M* maximum datagram size *r* bucket rate *p* peak rate					

APPENDIX I

Bibliography

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

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

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

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

– ITU-T G.177 (1999), *Transmission planning for voiceband services over hybrid internet/PSTN connections*.

– ITU-T T.30 (1999), *Procedures for document facsimile transmission in the general switched telephone network*.

– ITU-T T.38 (1998), *Procedures for real-time Group 3 facsimile communication over IP networks*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of 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	**Cable networks and 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

Geneva, 2001