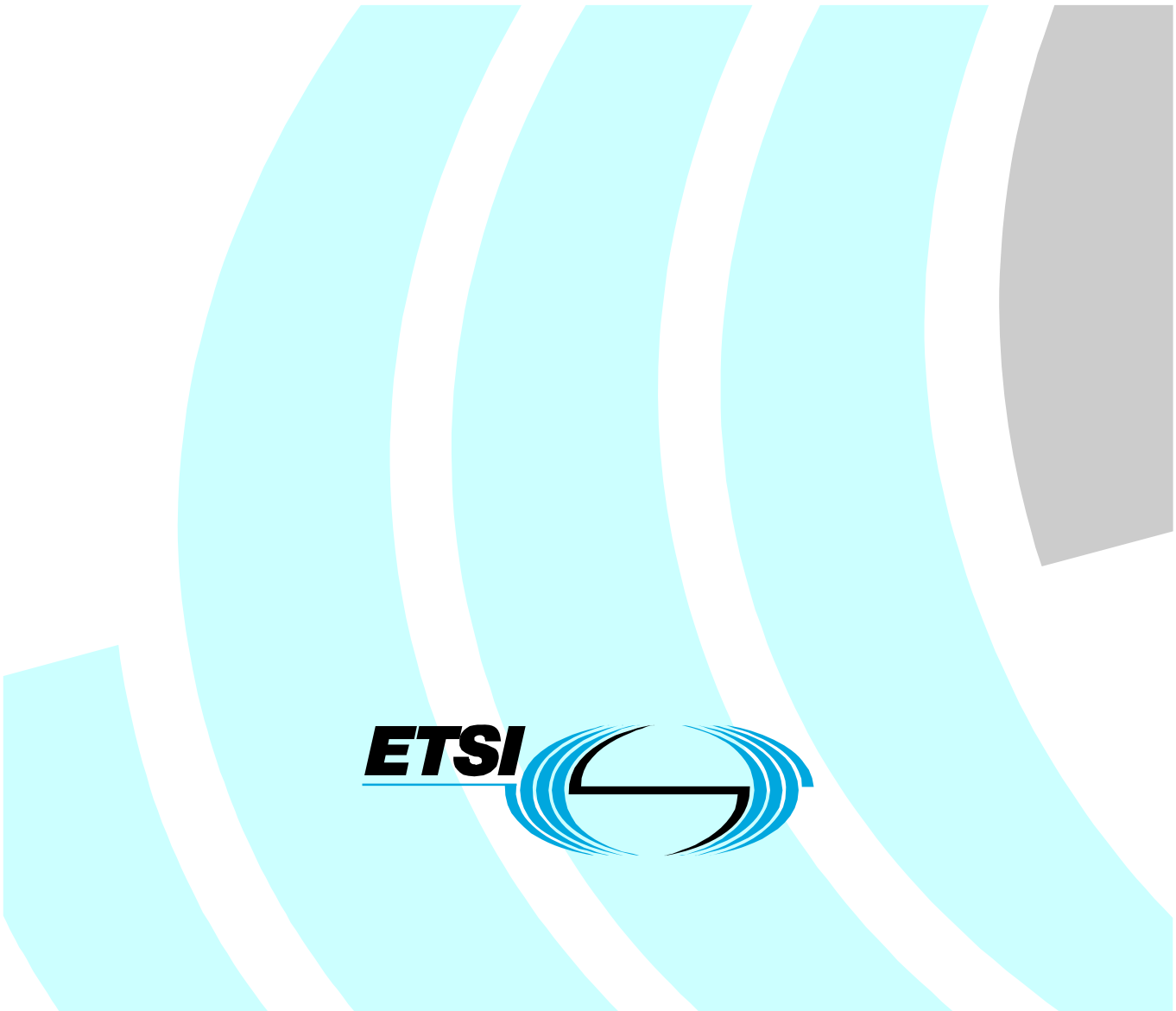
ETSI ES 202 667 V1.1.1 (2009-04)

ETSI Standard

Speech and multimedia Transmission Quality (STQ); Audiovisual QoS for communication over IP networks



Reference
DES/STQ-00097

Keywords
multimedia, QoS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Standard (ES) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

1 Scope

The present document addresses combination network performance parameters and user perceived media (audio and video) quality parameters for audiovisual communications on IP networks.

The access technologies covered include both wired (e.g. xDSL) and wireless (e.g. UMTS, WLAN) technologies.

The display size range covered is from those of small mobile terminals (e.g. 2") up to large TV sets (e.g. 40" or more).

It is applicable to:

- Broadcasting and streaming applications such as IPTV and VoD.
- Interactive point-to-pint applications such as videotelephony and videoconferencing.

Where the media coding standards define two or more profiles, the baseline profile is addressed in the normative part of the standard.

Informative annexes present an overview of network QoS mechanisms and the effects on connection performance as well as guidance on terminal parameters that may influence the user perceived media performance.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

[1] ITU-T Recommendation Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[2] ITU-T Recommendation Y.1541: "Network performance objectives for IP-based services".

[3] ITU-T Recommendation G.711: "Pulse Code Modulation (PCM) of voice frequencies".

[4] ITU-T Recommendation G.722: "7 kHz audio-coding within 64 kbit/s".

[5] ITU-T Recommendation G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5,3 and 6,3 kbit/s".

[6] ITU-T Recommendation G.726: "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".

[7] ITU-T Recommendation G.728: "Coding of speech at 16 kbit/s using low-delay code exited linear prediction".

[8] ITU-T Recommendation G.729: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)".

[9] ITU-T Recommendation G.729.1: "G.729 Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729".

[10] ETSI TS 126 071 (V6.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech Codec; General description (3GPP TS 26.071, version 6.0.0 Release 6)".

[11] 3GPP2 C.S0014-C v1.0 Enhanced Variable Rate Codec, Speech Service Option 3, 68 and 70 for Wideband Spread Spectrum Digital Systems, January 2007.

[12] ITU-T Recommendation G.722.1: "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss".

[13] ITU-T Recommendation G.711.1: "Wideband embedded extension for G.711 pulse code modulation".

[14] ETSI TS 126 171 (V6.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; General description (3GPP TS 26.171, version 6.0.0 Release 6)".

NOTE: TS 126 171 is identical to ITU-T Recommendation G.722.2.

[15] IETF RFC 3351: "RTP Profile for Audio and Video Conferences with Minimal Control".

[16] ISO/IEC 11172: "Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s (MPEG 1, 5 parts)".

[17] ISO/IEC 13818: "Information technology -- Generic coding of moving pictures and associated audio information (MPEG 2, 9 parts)".

[18] ISO/IEC 14496: "Information technology -- Coding of audio-visual objects (MPEG 4; currently in 11 parts)".

[19] ETSI TS 126 290 (V6.3.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS) Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions (3GPP TS 26.290, version 6.0.0 Release 6)".

[20] ITU-T Recommendation G.719: "Low-complexity full-band audio coding for high-quality conversational applications".

[21] ETSI TS 102 366: "Digital Audio Compression (AC-3, Enhanced AC-3) Standard".

[22] ITU-T Recommendation H.261: "Video codec for audiovisual services at p × 64 kbit/s".

[23] ITU-T Recommendation H.262: "Information technology - Generic coding of moving pictures and associated audio information: Video".

[24] ITU-T Recommendation H.263: "Video coding for low bit rate communication".

[25] ITU-T Recommendation H.264: "Advanced video coding for generic audiovisual services".

NOTE: This recommendation is identical to MPEG 4 Annex 10.

[26] SMPTE 421M (2006): "Television - VC-1 Compressed Video Bitstream Format and Decoding Process".

[27] ITU-R Recommendation BT.1359-1: "Relative timing of sound and vision for broadcasting".

[28] IETF RFC 768: "User Datagram Protocol".

[29] ETSI TS 122 146 (V7.1.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Multimedia Broadcast/Multicast Service (MBMS); Stage 1 (3GPP TS 22.146, version 7.1.0 Release 7)".

[30] IETF RFC 793: "Transmission Control Protocol".

[31] ITU-T Recommendation G.995.1: "Overview of digital subscriber line (DSL) Recommendations".

[32] IEEE 802.3 (2005): "IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks; Specific requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications".

[33] ITU-T Recommendation G.992: Parts 1 to 5.

[34] ITU-T Recommendation G.993: Parts 1 and 2.

[35] ITU-T Recommendation G.991.1: "High bit rate Digital Subscriber Line (HDSL) transceivers".

[36] ITU-T Recommendation G.991.2: "Single-pair high-speed digital subscriber line (SHDSL) transceivers".

[37] ETSI TS 101 113 (V7.5.0): "Digital cellular telecommunications system (Phase 2+) (GSM); General Packet Radio Service (GPRS); Service description; Stage 1 (GSM 02.60, version 7.5.0 Release 1998)".

[38] ETSI TS 122 228 (V8.5.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Service requirements for the Internet Protocol (IP) multimedia core network subsystem (IMS); Stage 1 (3GPP TS 22.228, version 8.5.0 Release 8)".

[39] ETSI TS 122 173 (V7.5.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1 (3GPP TS 22.173, version 7.5.0 Release 7)".

[40] ETSI TS 125 308 (V7.7.0): "Universal Mobile Telecommunications System (UMTS); High Speed Downlink Packet Access (HSDPA); Overall description; Stage 2 (3GPP TS 25.308, version 7.7.0 Release 7)".

[41] ETSI TS 125 319 (V7.6.0): "Universal Mobile Telecommunications System (UMTS); Enhanced uplink; Overall description; Stage 2 (3GPP TS 25.319, version 7.6.0 Release 7)".

[42] ETSI TS 123 107 (V7.1.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Quality of Service (QoS) concept and architecture (3GPP TS 23.107, version 7.1.0 Release 7)".

[43] ETSI TS 123 207 (V7.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); End-to-end Quality of Service (QoS) concept and architecture (3GPP TS 23.207, version 7.0.0 Release 7)".

[44] ETSI EN 300 175-2: "Digital Enhanced Cordless Telecommunications (DECT); Common Interface (CI); Part 2: Physical Layer (PHL)".

[45] IEEE 802.16 (2004): "Standard for Local and metropolitan area networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems".

[46] IEEE 802.11 (2007): "Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications".

[47] ETSI EN 300 401: "Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers".

[48] ETSI TS 102 428: "Digital Audio Broadcasting (DAB); DMB video service; User Application Specification".

[49] ETSI EN 302 307: "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications".

[50] ETSI EN 300 744: "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television".

[51] ETSI EN 300 419: "Access and Terminals (AT); 2 048 kbit/s digital structured leased lines (D2048S); Connection characteristics".

[52] ETSI EN 302 304: "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)".

[53] ETSI EN 302 583: "Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz".

[54] ETSI TS 101 154: "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream".

[55] ETSI TS 102 005: "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols".

[56] ITU-T Recommendation G.114: "One-way transmission time".

[57] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".

[58] IETF RFC 3095: "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed".

[59] ETSI TS 181 005: "Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); Service and Capability Requirements".

[60] ITU-R Recommendation BS.1534-1: "Method for the subjective assessment of intermediate quality levels of coding systems".

[61] ITU-T Recommendation G.107: "The E-model, a computational model for use in transmission planning".

[62] ITU-T Recommendation G.1010: "End-user Multimedia QoS Categories".

[63] ETSI ES 202 737: "Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for narrowband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

[64] ETSI ES 202 738: "Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for narrowband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[65] ETSI ES 202 739: "Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user".

[66] ETSI ES 202 740: "Speech Processing, Transmission and Quality Aspects (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user".

[67] ETSI TS 126 235 (V7.4.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Packet switched conversational multimedia applications; Default codecs (3GPP TS 26.235, version 7.4.0 Release 7)".

[68] ITU-T Recommendation J.247: "Objective perceptual multimedia video quality measurement in the presence of a full reference".

[69] ITU-T Recommendation P.911: "Subjective audiovisual quality assessment methods for multimedia applications".

[70] ETSI TS 181 018: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Requirements for QoS in a NGN".

[71] ETSI TS 122 105 (V8.4.0): "Universal Mobile Telecommunications System (UMTS); Services and service capabilities (3GPP TS 22.105, version 8.4.0 Release 8)".

[72] ETSI TS 126 234 (V7.5.0): "Universal Mobile Telecommunications System (UMTS); Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (3GPP TS 26.234, version 7.5.0 Release 7)".

[73] ETSI TS 126 346 (V7.8.0): "Universal Mobile Telecommunications System (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (3GPP TS 26.346, version 7.8.0 Release 7)".

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area**.** For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI ETR 310: "Digital Enhanced Cordless Telecommunications (DECT); Traffic capacity and spectrum requirements for multi-system and multi-service DECT applications co-existing in a common frequency band".

[i.2] ETSI TS 126 091 (V7.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech Codec; Error concealment of lost frames (3GPP TS 26.091, version 7.0.0 Release 7)".

[i.3] ETSI TS 126 191 (V7.0.0): "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Error concealment of erroneous or lost frames (3GPP TS 26.191, version 7.0.0 Release 7)".

[i.4] ETSI TR 102 479: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Review of available material on QoS requirements of Multimedia Services".

[i.5] IETF RFC 1633: "Integrated services in the Internet architecture: An overview".

[i.6] IETF RFC 2205: "Resource ReSerVation Protocol (RSVP) Version 1 Functional Specification".

[i.7] IETF RFC 2475: "An Architecture for Differentiated Services".

[i.8] Cicconetti, C., Lezini, L., Mingozzi, E. and Eklund, C.: "Quality of Service support in 802.16 networks. IEEE Network, vol. 29", March/April 2006.

[i.9] Pedersen, K., Mogensen, P. and Kolding, T.: "Overview of QoS options for HSDPA. IEEE Communications Magazine vol. 44", July 2006.

[i.10] ETSI ES 282 003: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-System (RACS): Functional Architecture".

[i.11] Perkins, C, Hodson, O. and Hardman, V.: "A Survey of Packet Loss Recovery Techniques for Streaming Audio. IEEE Network vol. 12, No. 5", 1998.

[i.12] Wah, B., Su, X. and Lin, D.: "A Survey of Error-Concealment Schemes for Real-Time Audio and Video Transmissions over the Internet. IEEE International Symposium on Multimedia Software Engineering 2000 (MSE 2000)"; Taipei, Taiwan, 11-13 December 2000.

[i.13] ITU-T Recommendation G.711 (Appendix I).: "Pulse code modulation (PCM) of voice frequencies; A high quality low-complexity algorithm for packet loss concealment with G.711".

[i.14] ITU-T Recommendation G.722 (Appendix III): "7 kHz audio-coding within 64 kbit/s; A high-quality packet loss concealment algorithm for G.722".

[i.15] ITU-T Recommendation G.722 (Appendix IV): "7 kHz audio-coding within 64 kbit/s; A low-complexity algorithm for packet loss concealment with G.722".

[i.16] Wenger, S.: "H.264/AVC over IP. IEEE Transactions on circuits and systems for video technology, vol. 13, No. 7", 2003.

[i.17] ETSI TR 101 329-6: "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 6: Actual measurements of network and terminal characteristics and performance parameters in TIPHON networks and their influence on voice quality".

[i.18] Kövesi, B. and Ragot, S.: "A low complexity packet loss concealment algorithm for ITU-T Recommendation G.722. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)". Las Vegas, USA, 30th March - 4th April, 2008.

[i.19] ITU-T Recommendation I.113: "Vocabulary of terms for broadband aspects of ISDN".

[i.20] ITU-T Recommendation G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[i.21] ITU-T Recommendation SG.12: "Temporary Documents".

[i.22] Layer 1 specifications.

NOTE: Available at http://3GPP specification series: 05series.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

audio: all signals that are audible to human beings, including speech and music

broadcasting: communication capability which denotes unidirectional distribution from a single source to all users connected to the network

multipoint: value of the service attribute "communication configuration", which denotes that the communication involves more than two network terminations

NOTE: Source: ITU-T Recommendation I.113 [i.19].

narrowband speech: speech restricted to the frequency band from 300 Hz to 3 400 Hz

speech: oral production of information by a human being

streaming: mechanism whereby media content can be rendered at the same time that it is being transmitted to the client over the network

video: signal that contains timing/synchronization information as well as luminance (intensity) and chrominance (colour) information that when displayed on an appropriate device gives a visual representation of the original image sequence

videoconferencing: service providing interactive, bi-directional and real time audio-visual communication

NOTE: Normally intended for multiple users at each end.

videotelephony: service providing an interactive, bi-directional, real time audio-visual communication between users

wideband speech: speech restricted to the frequency band from 50 Hz to 7 000 Hz

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	3rd Generation Partnership Project
3GPP2	3rd Generation Partnership Project 2

NOTE: A 3G project comprising North American and Asian interests.

AAC	Advanced Audio Coding
ADPCM	Adaptive Differential Pulse Code Modulation
AMR	Adaptive Multi Rate
AMR-WB	Adaptive Multi Rate Wide Band
AMR-WB+	Adaptive Multi Rate extended Wide Band
AP	Access Point (IEEE 802.11 WLAN [46])
ATM	Asynchronous Transfer Mode
AVC	Advanced Video Coding
CCIR	Comité Consultatif International pour la Radio; Now ITU-R
CELP	Code-Excited Linear Predictive
CIF	Common Intermediate Format
CPCFC	Custom Picture Clock Frequency Code
CPFMT	Custom Picture ForMaT
DECT	Digital Enhanced Cordless Telecommunications
DPCM	Differential Pulse Code Modulation
EUL	Enhanced UpLink
FER	Frame Error Rate
FP	Fixed Part (DECT)
HDTV	High Definition TV
HE-AAC	High Efficiency AAC
HSPA	High-Speed Packet Access
HSDPA	High-Speed Downstream Packet Access
HSUPA	High-Speed Upstream Packet Access
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPDV	IP Packet Delay Variation
IPER	IP Packet Error Ratio
IPLR	IP Packet Loss Ratio
IPTD	IP Packet Transfer Delay
ITU-R	International Telecommunication Union - Radiocommunication sector
ITU-T	International Telecommunication Union - Telecommunication standardization sector
LPC	Linear Predictive Coding
MAC	Medium Access Control
MBMS	Mobile Broadcast/Multicast Service
MDCT	Modified Discrete Cosine Transform
MCU	Multipoint Control Unit
MPE	Multi-Pulse Excited
MPEG 2 TS	MPEG 2 Transport Stream
MPEG	Moving Picture Experts Group
MUSHRA	MUlti Stimulus with Hidden Reference and Anchors
NTSC	National Television System Committee

NOTE: Used to identify an analogue TV standard used outside Europe.

PAL	Phase-Alternating Line

NOTE: Colour-encoding system used in television systems.

PBX	Private Branch eXchange
PCM	Pulse Code Modulation
PP	Portable Part (DECT)
QCIF	Quart CIF
QVGA	Quart VGA
RTP	Real-time Transport Protocol
RTT	Round Trip Time
SDTV	Standard Definition TV
SVC	Scalable Video Coding
TCP	Transport Control Protocol
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
VGA	Video Graphics Array
W-CDMA	Wideband-Code Division Multiple Access
WLAN	Wireless Local Area Network

NOTE: IPER, IPDV, IPLR and IPTD are defined in ITU-T Recommendations Y.1540 [1] and Y.1541 [2].

4 Parameters affecting audiovisual user perceived quality

4.1 Audiovisual user perceived quality model

The characteristics affecting audiovisual user perceived quality and their interactions are illustrated in figure 1.

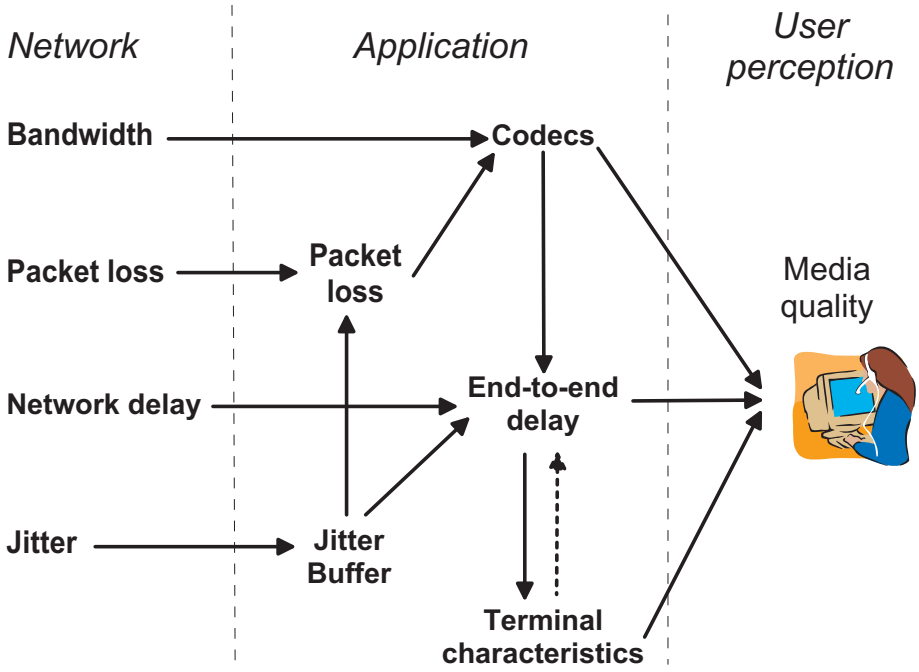


Figure 1: Characteristics affecting audiovisual user perceived quality

4.2 Coding algorithms

4.2.1 Speech

There are two groups of speech codecs:

- narrowband codecs, transmitting the frequency band from 300 Hz to 3 400 Hz;
- wideband codecs, transmitting the frequency band from 50 Hz to 7 000 Hz.

NOTE 1: The transmission bandwidth indicated is the bandwidth supported by the coding algorithm. The actual bandwidth supported may be restricted due to handset or terminal characteristics.

Speech codecs are often classified into three types:

- waveform codecs;
- source codecs;
- hybrid codecs.

Waveform codecs attempt, without using any knowledge of how the signal to be coded was generated, to produce a reconstructed signal whose waveform is as close as possible to the original. This means that in theory they should be signal independent and work well with non-speech signals. An example is the codec standardized in ITU-T Recommendation G.711 [74].

To reduce the required bit rate, the difference compared with the previous sample may be transmitted instead of the actual sample. This technique is call delta modulation or differential PCM (DPCM). This technique may be further enhanced by predicting the value of the next sample from the previous samples and transmit the difference between the predicted value and the actual sampled value (ADPCM).

The input speech signal may also be split into a number of frequency bands, or sub-bands, and each is coded independently. This is called Sub-band coding. An example is the codec defined in ITU-T Recommendation G.722 [4] where the 7 kHz frequency band is divided into two sub-bands, which are coded independent of each other.

Source coders operate using a model of how the source is generated, and attempt to extract, from the signal being coded, the parameters of the model. Coders using this technique require very low bit rate, but the quality is usually not good enough for public telecommunication applications.

Hybrid codecs attempt to fill the gap between waveform and source codecs. Although other forms of hybrid codecs exist, the most successful and commonly used are time domain Analysis-by-Synthesis (AbS) codecs. Such coders use the same linear prediction filter model of the vocal tract as found in LPC vocoders. However instead of applying a simple two-state, voiced/unvoiced, model to find the necessary input to this filter, the excitation signal is chosen by attempting to match the reconstructed speech waveform as closely as possible to the original speech waveform. Examples are Multi-Pulse Excited (MPE) codecs and Code-Excited Linear Predictive (CELP) codecs.

The narrowband speech codecs standardized by ITU-T and ETSI are listed in table 1.

Table 1: Narrowband speech codecs standardized by ITU-T and ETSI

Codec ID	Bit rate (kbit/s)	Frame size (ms)	Look ahead (ms)
ITU-T Recommendation G.711 [75]	64	n/a	
ITU-T Recommendation G.723.1 [5]	6,3 5,3	30	7,5
ITU-T Recommendation G.726 [6]	16, 24, 32 and 40	n/a	n/a
ITU-T Recommendation G.728 [7]	16	n/a	n/a
ITU-T Recommendation G.729 [8]	6,4, 8 and 11,8	10	5
ITU-T Recommendation G.729.1 [9] (see note 1)	8, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32	20 (see note 2)	25
TS 126 071 AMR [10]	4,75, 5,15, 5,90, 6,70, 7,40, 7,95, 10,20 and12,2	20	n/a
NOTE 1: ITU-T Recommendation G.729.1 [9] can be used as either a narrowband codec (interoperable with G.729 [8]) or a wideband codec. NOTE 2: When the encoder/decoder pair is working in Low delay/Narrowband mode, the algorithmic delay is the same as that of ITU-T Recommendation G.729 [8].			

3GPP2 has standardized a variable bit rate codec, Enhanced Variable Rate Codec (EVRC) [11]. One of four channel rates corresponding to the 9 600 bit/s, 4 800 bit/s, 2 400 bit/s and 1 200 bit/s frame 5 rates can be used. The speech coder source bit rates corresponding to the above mentioned rates are 8 550 bit/s, 4 000 bit/s, 2 000 bit/s and 800 bit/s. Service option 3 identifies the EVRC codec, service option 68 identifies the EVRC-B codec.

NOTE 2: EVRC Option 3 does not support the 2 400 bit/s rate.

The wideband speech codecs standardized by ITU-T and ETSI are listed in table 2.

Table 2: Wideband speech codecs standardized by ITU-T and ETSI

Codec ID	Bit rate (kbit/s)	Frame size (ms)	Look ahead (ms)
ITU-T Recommendation G.711.1 [13]			
ITU-T Recommendation G.722 [4]	48, 56 and 64 (see note 1)	n/a	n/a
ITU-T Recommendation G.722.1 [12]	24, 32 and 48	20	20
ITU-T Recommendation G.729.1 [9]	8, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32	20 (see note 2)	25 (see note 3)
TS 126 171 [14] (see note 4)	6,60, 8,75, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 and 23,85	20	n/a
NOTE 1: The ITU-T Recommendation defines three operating modes; 48 kbit/s, 56 kbit/s and 64 kbit/s. However, The IETF RTP profile for this codec defined in RFC 3351 [15] only addresses the 64 kbit/s mode. NOTE 2: For narrowband mode see note 2 to table 1. NOTE 3: 20 ms for the MCDT analysis and 5 ms for the narrowband LPC analysis. NOTE 4: TS 126 171 [14] is identical to ITU-T Recommendation G.722.2 [i.20].			

The 3GPP2 standardized EVRC codec [11] includes a wideband option (Service option 70).

4.2.2 Audio

Standardized audio coding algorithms have been developed by the ISO/IEC Moving Picture Experts Group (MPEG), a group working on coding standards for audio and video.

The first audio coding generation is defined in part 3 of MPEG 1 [16]. Three operating modes, called layers with increasing complexity and performance, are defined. Layer 3 is the highest complexity mode, optimized to provide the acceptable quality at bit rates around 128 kbit/s for stereo signals. This coding algorithm is often referred to as MP3. MPEG 1 [16] Part 3 defines three sampling frequencies; 32 kHz, 44,1 kHz and 48 kHz.

MPEG 2 [17] extends the sampling frequencies to 16 kHz, 22,05 kHz and 24 kHz allowing for lower transmission bit rates. There is no modification to the coding algorithm.

A second generation audio coding, called Advanced Audio Coding (AAC) is defined in MPEG 2 [17] Part 7. This is a completely new algorithm that is not backward compatible with MP3. The algorithm was extended in MPEG 4 [18] Part 3. The MPEG 4 AAC algorithm supports a wide range of sampling rates (8 kHz to 96 kHz). The range of bit rates supported is from 16 kbit/s up to 288 kbit/s per channel and up to 48 audio channels can be supported. This algorithm is often referred to as Low Complexity AAC (LC AAC).

There is also standardized a High Efficiency AAC (HE AAC) extension. HE AAC can deliver coding gains of more than 40 % compared to LC AAC. There are two versions of this extension; HE AAC version 1 (HE AAC v1) and HE AAC version 2 (HE AAC v2). HE AAC v2 enhance the compression efficiency of stereo signals by including a parametric stereo tool. Frequently used trade names are aacPlus v1 and aacPlus v2. HE AAC v2 (see note 1) is a superset of HE AAC v1 which is a superset of AAC. A HE AAC v2 can decode both HE AAC v1 and AAC and HE AAC v1 can decode AAC.

NOTE 1: The acronym Eaac+ is also used to identify this codec.

There is also a low delay AAC option (LD AAC). The algorithmic delay of this coding algorithm is typically around 20 ms depending on the sampling frequency.

3GPP has developed an extended AMR wideband codec that is designed for a wider range of applications including music. The AMR-WB+ codec [19] supports signal bandwidth up to 20 kHz. The sampling rates supported are from 16 kHz up to 48 kHz. Both mono and stereo are supported. The bit rates supported is from 6 kbit/s up to 36 kbit/s in mono mode, and from 8 kbit/s up to 48 kbit/s in stereo mode.

NOTE 2: The AMR-WB+ [19] decoder is able to decode AMR-WB [14] encoded content.

ITU-T Recommendation G.719 [20] specifies a low complexity fullband coding algorithm for conversational speech and audio and operating from 32 kbit/s up to 128 kbit/s. The encoder input and decoder output are sampled at 48 kHz. The codec enables full bandwidth, from 20 Hz to 20 kHz, encoding of speech, music and general audio content. The codec operates on 20 ms frames and has an algorithmic delay of 40 ms.

The AC-3/Enhanched AC-3 audio coding algorithm is standardized in TS 102 366 [21]. The algorithm can encode from 1 to 5,1 channels of source audio from a PCM representation into a serial bit stream at data rates ranging from 32 kbit/s to 640 kbit/s. Bit rates supported are: 32 to 80 (steps of 8), 96, 112, 128, 150 to 256 (steps of 32), 300 to 640 (steps of 64).

The optional enhanced AC-3 is an evolution of the AC-3 coding system. The addition of a number of low data rate coding tools enables use of Enhanced AC-3 at a lower bit rate than AC-3 for high quality, and use at much lower bit rates than AC-3 for medium quality. Enhanced AC-3 bit streams are similar in nature to standard AC-3 bit streams, but are not backwards compatible (i.e. they are not decodable by standard AC-3 decoders).

4.2.3 Video

There is standardization work on audio and video coding in both ISO and ITU-T. The work of ITU-T address mainly interactive communication such as videotelephony and videoconferencing, while the ISO work has a broader scope including streaming. There are joint ISO/ITU-T activities on video coding.

Most standardized video coding algorithms apply several encoding principles:

- Intraframe coding where each video frame is encoded separately. These frames are often called ***I*** frames.
- Interframe coding where the difference between successive video frames is encoded. These frames are often called ***B*** frames.

The interframe coding works well with a stationary background and a moving foreground. However, to cope with scenes where the camera is moving, zooming or panning, this technique is not working very well. To solve this problem motion compensation is introduced. These frames are coded using a predicted (motion compensated) frame as reference. These frames are often called ***P*** (predicted) frames.

Another problem related to interframe coding is the effect of packet loss. If a frame being reference for following frames is lost, the loss will propagate to these. To cope with this problem error resilience technologies have been introduced. These technologies are addressed by the standardization bodies.

There are four ITU-T standardized video coding algorithms:

1) ITU-T Recommendation H.261 [22] was considered as a break-through in the art of video coding enabling video communication over 64 kbit/s channels;

2) ITU-T Recommendation H.262 [23] is equivalent to the video part of MPEG 2 [17];

3) ITU-T Recommendation H.263 [24];

4) ITU-T Recommendation H.264 (AVC) [25].

The ISO video and audio coding standards are developed by the Moving Picture Experts Group (MPEG). There are currently three groups of standards defining audio and video coding:

1) MPEG 1 [16] was initially designed for storage and transmission of media at up to 1,5 Mbit/s. The quality of the MPEG 1 video coding is similar to the quality of a VHS recorder.

2) MPEG 2 [17] allows for coding of studio quality video for digital TV including High Definition TV. MPEG 2 is capable of coding of standard television pictures at bit rates in the range from 3 Mbit/s to 15 Mbit/s. High definition TV pictures can be coded at bit rates between 15 Mbit/s and 30 Mbit/s.

3) The MPEG 4 [18] standard supports a wider range of applications than MPEG 1 and MPEG 2 including communication over mobile links.
The video part of MPEG 4 includes both coding of live video and synthetic video. Several profiles are defined. The video coding of the simple profile can interwork with the video coding defined in ITU-T Recommendation H.263 [24]. As a rule of thumb AVC delivers equal quality to base MPEG 4 video at half the bit rate.

The video coding standards describes a large number of parameters that can be selected. The provide interoperability between implementations profiles and profile levels are defined. Profiles and levels specify restrictions on bitstreams and hence limits on the capabilities needed to decode the bitstreams. Profiles and levels may also be used to indicate interoperability points between individual decoder implementations.

Each profile specifies a subset of parameters, algorithmic features and limits that shall be supported by all decoders conforming to that profile.

Each level specifies a set of limits on the values that may be taken by the syntax elements of the standard. Individual implementations may support a different level for each supported profile. For any given profile, levels generally correspond to decoder processing load and memory capability.

The profiles defined in ITU-T Recommendation H.263 [24] are:

- Baseline Profile (Profile 0). The profile defines the minimal "baseline" capability of the recommendation.
- H.320 Coding Efficiency Version 2 Backward-Compatibility Profile (Profile 1).The profile defines a feature set adopted into the H.242 capability exchange mechanism for use by H.320 circuit-switched terminal systems.
- Version 1 Backward-Compatibility Profile (Profile 2). This profile provides enhanced coding efficiency performance within the feature set available in the first version of the recommendation.
- Version 2 Interactive and Streaming Wireless Profile (Profile 3). This profile provides enhanced coding efficiency performance and enhanced error resilience for delivery to wireless devices within the feature set available in the second version of this recommendation.
- Version 3 Interactive and Streaming Wireless Profile (Profile 4). This profile provides enhanced coding efficiency performance and enhanced error resilience for delivery to wireless devices within the feature set available in the third version of this recommendation.
- Conversational High Compression Profile (Profile 5). This is a low delay profile which defines enhanced coding efficiency performance without use of B pictures and error resilience features.
- Conversational Internet Profile (Profile 6). This profile adds to Profile 5 features error resilience mechanisms suitable for use on IP networks.
- Conversational Interlace Profile (Profile 7). This profile adds support of interlaced video sources to Profile 5.

- High Latency Profile (Profile 8). This profile provides enhanced coding efficiency performance for applications without critical delay constraints.

Eight levels of performance capability are defined for decoder implementation. Table 3 defines the detailed performance parameters of each of these levels.

Table 3: H.263 Profile levels

Level number	Max picture format	Max bit rate	Min picture interval
10	QCIF	64 kbit/s	2 002/30 000 s
20	CIF	128 kbit/s	2 002/30 000 s for CIF 1 001/30 000 s for QCIF and sub-QCIF
30	CIF	384 kbit/s	1 001/30 000 s
40	CIF	2 048 kbit/s	1 001/30 000 s
45	QCIF Support of CPFMT in profiles other than 0 and 2	128 kbit/s	2 002/30 000 s support of CPCFC in profiles other than 0 and 2
50	CIF support of CPFMT	4 096 kbit/s	1/50 s 1 001/(60 000) s at 352 × 240 or smaller support of CPCFC
60	CPFMT: 720 × 288 support of CPFMT	8 192 kbit/s	1/50 s at 720 × 288 or lower 1 001/(60 000) s at 720 × 240 or smaller support of CPCFC
70	CPFMT: 720 × 576 support of CPFMT	16 384 kbit/s	1/50 s at 720 × 576 or lower 1 001/(60 000) s at 720 × 480 or smaller support of CPCFC

The profiles defined in ITU-T Recommendation H.264 (AVC) [25] are:

- Baseline Profile. This is the simplest profile targeting applications with low complexity and low delay requirements.
- Main Profile. This profile allows best quality at the cost of higher complexity and delay.
- Extended Profile. Intended as streaming video profile.
- High Profile. The primary profile for broadcast and disk storage applications.
- High 10 Profile. Adding 10 bits per sample precision to the High Profile.
- High 4:2:2 Profile. Adding support for 4:2:2 chroma precision to the High 10 Profile.
- High 4:4:4 Predictive Profile. Adding support for 4:4:4 chroma precision and 14 bits per sample to the High 4:2:2 Profile.

For each of the profiles 16 levels are defined. The max. video bit rate for each profile/level combination can be found in table 4.

The MPEG 4 [18] video profiles and levels can be found in table 5.

Table 4: Max video bit rate of H.264 AVC profile levels

Level Number	Baseline, Main and Extended Profiles	High Profile	High 10 Profile	High 4:2:2 and High 4:4:4 Predictive Profiles
1	64 kbit/s	80 kbit/s	192 kbit/s	256 kbit/s
1b	128 kbit/s	160 kbit/s	384 kbit/s	512 kbit/s
1,1	192 kbit/s	240 kbit/s	576 kbit/s	768 kbit/s
1,2	384 kbit/s	480 kbit/s	1 152 kbit/s	1 536 kbit/s
1,3	768 kbit/s	960 kbit/s	2 304 kbit/s	3 072 kbit/s
2	2 Mbit/s	2,5 Mbit/s	6 Mbit/s	8 Mbit/s
2,1	4 Mbit/s	5 Mbit/s	12 Mbit/s	16 Mbit/s
2,2	4 Mbit/s	5 Mbit/s	12 Mbit/s	16 Mbit/s
3	10 Mbit/s	12,5 Mbit/s	30 Mbit/s	40 Mbit/s
3,1	14 Mbit/s	17,5 Mbit/s	42 Mbit/s	56 Mbit/s
3,2	20 Mbit/s	25 Mbit/s	60 Mbit/s	80 Mbit/s
4	20 Mbit/s	25 Mbit/s	60 Mbit/s	80 Mbit/s
4,1	50 Mbit/s	62,5 Mbit/s	150 Mbit/s	200 Mbit/s
4,2	50 Mbit/s	62,5 Mbit/s	150 Mbit/s	200 Mbit/s
5	135 Mbit/s	168,75 Mbit/s	405 Mbit/s	540 Mbit/s
5,1	240 Mbit/s	300 Mbit/s	720 Mbit/s	960 Mbit/s

Table 5: MPEG 4 video profiles and levels

Visual Profile	Level	Max number of objects	Picture size	Max bit rate
Simple	L0	1	QCIF	64 kbit/s
Simple	L1	4	QCIF	64 kbit/s
Simple	L2	4	CIF	128 kbit/s
Simple	L3	4	CIF	384 kbit/s
Advanced Real-Time Simple	L1	4	QCIF	64 kbit/s
Advanced Real-Time Simple	L2	4	CIF	128 kbit/s
Advanced Real-Time Simple	L3	4	CIF	384 kbit/s
Advanced Real-Time Simple	L4	16	CIF	2 000 kbit/s
Simple Scalable	L1	4	CIF	128 kbit/s
Simple Scalable	L2	4	CIF	256 kbit/s
Core	L1	4	QCIF	384 kbit/s
Core	L2	16	CIF	2 000 kbit/s
Advanced Core	L1	4	QCIF	384 kbit/s
Advanced Core	L2	16	CIF	2 000 kbit/s
Core Scalable	L1	4	CIF	768 kbit/s
Core Scalable	L2	8	CIF	1 500 kbit/s
Core Scalable	L3	16	CCIR601	4 000 kbit/s
Main	L2	16	CIF	2 000 kbit/s
Main	L3	32	CCIR601	15 000 kbit/s
Main	L4	32	1 920 × 1 088	38 400 kbit/s
Advanced Coding Efficiency	L1	4	CIF	384 kbit/s
Advanced Coding Efficiency	L2	16	CIF	2 000 kbit/s
Advanced Coding Efficiency	L3	32	CCIR601	15 000 kbit/s
Advanced Coding Efficiency	L4	32	1 920 × 1 088	38 400 kbit/s
N-Bit	L2	16	CIF	2 000 kbit/s

Scalable Video Coding (SVC) is a principle that can be used to adapt the video transmission to transmission channel characteristics such as bandwidth restrictions and/or transmission channel errors. It is also possible to adapt the video stream to a variety of receiving terminal characteristics. This technology is included in MPEG 2 [17]/H.262 [23], MPEG 4 [18]/H.263 [24] and H.264 (AVC) [25], and makes it possible to define two or more profiles.

VC-1 [26] is a video coding algorithm standardized by SMPTE, an organization developing industry standards for the imaging industry. The performance of VC-1 is similar to the performance of H.264 (AVC) [25].

The profiles defined for VC-1 are described in table 6.

Table 6: VC-1 profiles

Profile	Level	Max. bit rate	Representative Resolutions by Frame Rate
Simple	Low	96 kbit/s	176 × 144 @ 15 Hz (QCIF)
Simple	Medium	384 kbit/s	240 × 176 @ 30 Hz 352 × 288 @ 15 Hz (CIF)
Main	Low	2 Mbit/s	320 × 240 @ 24 Hz (QVGA)
Main	Medium	10 Mbit/s	720 × 480 @ 30 Hz (480 p) 720 × 576 @ 25 Hz (576 p)
Main	High	20 Mbit/s	1 920 × 1 080 @ 30 Hz (1 080 p)
Advanced	L0	2 Mbit/s	352 × 288 @ 30 Hz (CIF)
Advanced	L1	10 Mbit/s	720 × 480 @ 30 Hz (NTSC-SD) 720 × 576 @ 25 Hz (PAL-SD)
Advanced	L2	20 Mbit/s	720 × 480 @ 60 Hz (480 p) 1 280 × 720 @ 30 Hz (720 p)
Advanced	L3	45 Mbit/s	1 920 × 1 080 @ 24 Hz (1 080 p) 1 920 × 1 080 @ 30 Hz (1 080 i) 1 280 × 720 @ 60 Hz (720 p)
Advanced	L4	135 Mbit/s	1 920 × 1 080 @ 60 Hz (1 080 p) 2 048 × 1 536 @ 24 Hz

4.2.4 Lip synchronization

The video coding introduces larger delay than the speech coding. Most of the video packets are larger than the speech packets, which may results in a slightly larger network delay, particularly in the access network. The result is than the speaking motion of a person is not synchronized with the speech.

There are several documents specifying the acceptable difference between speech and video. There are standards specifying symmetric thresholds. However, the general conclusion is that audio arriving ahead of video is more annoying than audio arriving after the video.

For broadcasting purposes ITU-R Recommendation BT.1359-1 [27] defines detectability and acceptability threshold for lip synchronization. Figure 2 describes these thresholds. The detectability thresholds are 125 ms when sound is delayed with respect to the video, and 45 ms when sound is advanced with respect to the video. The acceptability thresholds are 185 ms when sound is delayed with respect to the video, and 90 ms when sound is advanced with respect to the video.

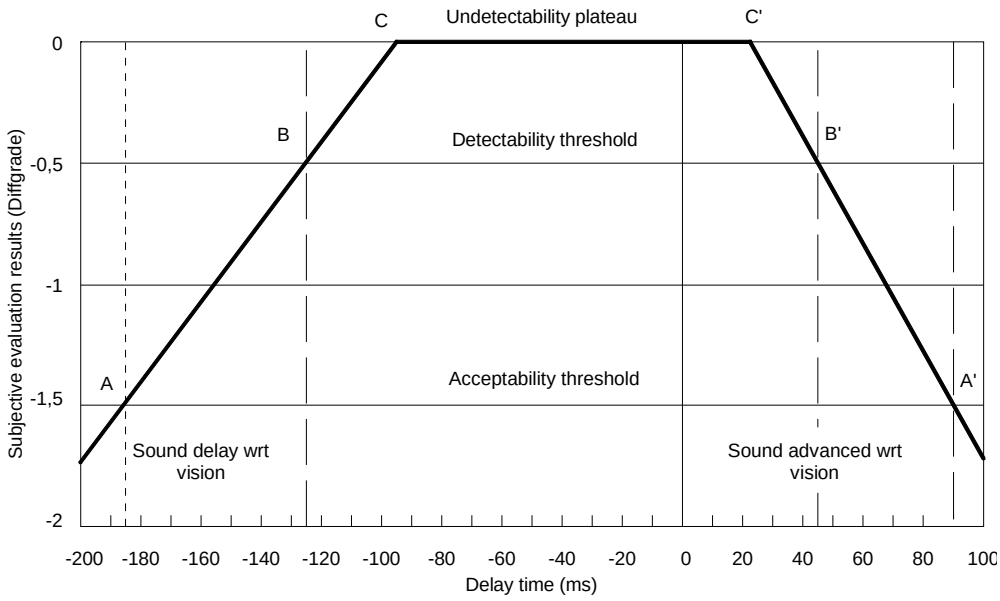


Figure 2: Detectability and acceptability thresholds for lip synchronization

4.3 Network transmission protocols and principles

4.3.1 Unreliable communication

The unreliable transport protocol standardized by IETF is UDP [28]. UDP does not guarantee delivery of packets and that they are delivered in correct order. Applications using UDP should not be sensitive to packet loss. When ordering of packets is important, the application protocol shall be able to reorder packets.

4.3.2 Multicast

Multicast is a communication capability where the same information is transmitted from a single source to a group of destinations simultaneously. The principle is depicted in figure 3. The information is transported using the UDP [28] protocol described above.

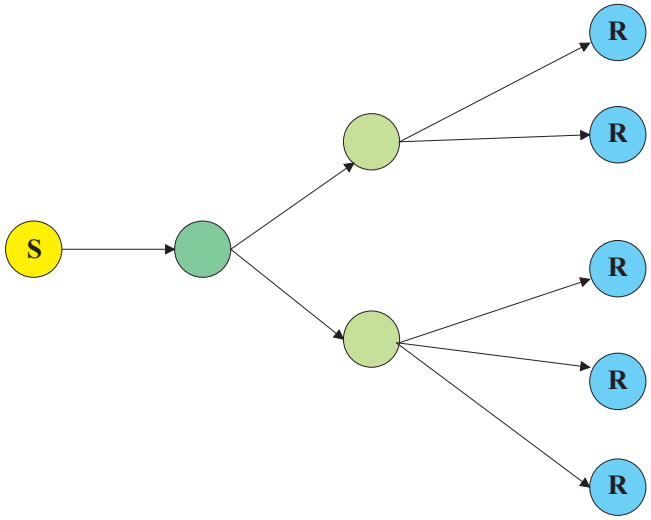


Figure 3: Multicast principle

3GPP has defined a unidirectional point to multipoint bearer service in which data is transmitted from a single source entity to multiple recipients (The Mobile Broadcast/Multicast Service - MBMS) [29] as depicted in figure 4.

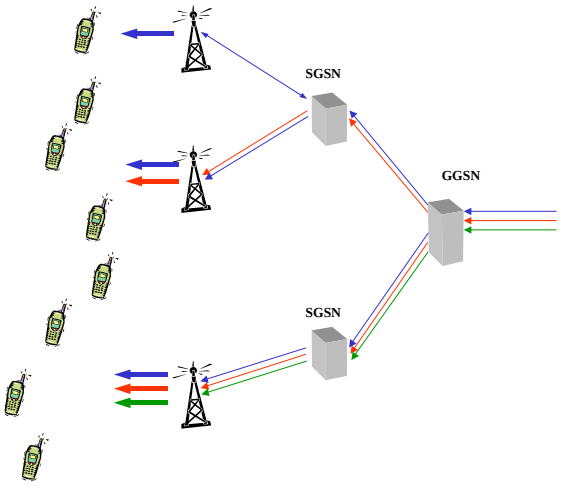


Figure 4: MBMS principle

4.3.3 Reliable communication

The most common reliable transport protocol used for communication over an IP network, is TCP described in RFC 793 [30]. TCP allows devices to establish and manage connections and send data reliably, and takes care of handling potential problems that can occur during transmission. It is stream-oriented; it is designed to have applications send data to it as a stream of bytes, rather than requiring fixed-size messages to be used. This provides maximum flexibility for a wide variety of uses, because applications does not need to worry about data packaging, and can send files or messages of any size. TCP takes care of packaging these bytes into messages called segments. Furthermore, TCP is a reliable protocol, when a packet is lost, it is retransmitted. TCP assigns a sequence number to each byte transmitted, and expects a positive acknowledgment (ACK) from the receiving TCP. If the ACK is not received within a timeout interval, the data is retransmitted. The receiving TCP uses the sequence numbers to rearrange the segments when they arrive out of order, and to eliminate duplicate segments Each byte has a sequence number. The transmission time is usually determined by the Round Trip Time (RTT) of the connection. Retransmissions or other events may also influence this period. One of the drawbacks of this protocol is increased transmission delay when packets are lost. It is therefore not suitable for delay sensitive applications such as interactive voice, but is an alternative for delay unsensitive applications such as media streaming.

4.3.4 Multipoint

Most multipoint applications are based on a central server (MCU, Focus). The connection between each participant and the centralized server is a point-to-point connection. The central server performs audio mixing and video switching. In this scenario the protocols described above are used for each of the point-to-point connections.

4.3.5 DSL Access

DSL is a group of technologies that provides high speed data transmission over telephone lines.

Two different groups of technologies are available:

- asymmetrical DSL where the downstream bandwidth (from the network to the user) are larger than the upstream bandwidth;
- symmetrical DSL where the downstream and upstream bandwidths are equal.

The DSL Access principle is depicted in figure 5. The telephone connection uses the low frequency band, and the data is transmitted in the high frequency part of the spectrum. A splitter is used to separate the two parts of the spectrum. The data connection from the end user DSL Modem is terminated in a Digital Subscriber Line Access Multiplexer (DSLAM) which concentrates a number of individual DSL connections.

NOTE 1: Systems without splitter are available. These systems do not support circuit-switched telephone transmission.

NOTE 2: Customer premises DSL Modems may include a router and a telephone interface to support VoIP.

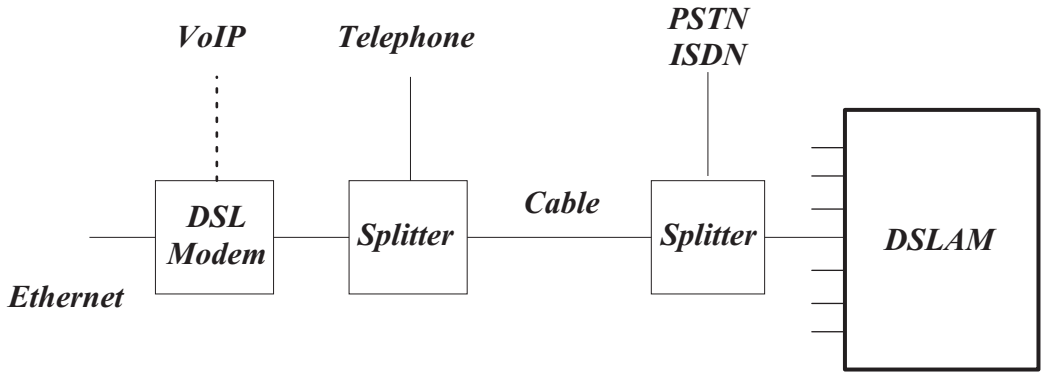


Figure 5: DSL Access principle

An overview of the ITU-T family of DSL recommendations are given in ITU-T Recommendation G.995.1 [31].

The most common solution for DSL communication is to use ATM as layer 2 carrier. The ATM cell size is fixed: 48 bytes payload and a 5 bytes header; i.e. more than 10 % overhead. For speech packets two cells or more, where the last cell is partly filled, might be required. The ATM related overhead might therefore be significant. IEEE 802.3ah [32] defines a solution for Ethernet applications without underlying ATM reducing this overhead.

The achievable bit rate depends on the copper line quality and the distance from the central office to the customer installation.

DSL communication may use interleaving reduce the degradation caused by noise bursts on a line. The cost of interleaving is increased connection delay. However, some systems support a dual latency option where delay sensitive data are transmitted on a low delay path, and less delay sensitive data are transmitted on an interleaved path. For further information see the following clauses.

4.3.5.1 Asymmetrical DSL

The asymmetrical DSL technologies have larger transmission capacity downstream than upstream. The justification for the asymmetry is partly commercial; more information is downloaded from the network than uploaded to the network. A technical justification is the near-end cross-talk effect on the transmission because all of the downstream signals can be of the same amplitude thus eliminating crosstalk between downstream channels.

There are two standardized asymmetrical DSL technologies:

- ADSL [33]
  There are three ADSL versions standardized; ADSL, ADSL 2, ADSL 2+.

- VDSL [34] (see note)
  There are two VDSL versions standardized, VDSL and VDSL2.

NOTE: VDSL supports both asymmetrical and symmetrical operations.

The maximum capacity of the first version, ADSL, was 8 Mbit/s downstream and 640 kbit/s upstream. The next generation, ADSL2, increased the upstream capacity up to approximately 1 Mbit/s. The downstream capacity is slightly increased, as is the supported distance between the central office and the customer installation. Annexes to the recommendation describe enhancements making it possible to increase the upstream data rate to 3 Mbit/s and to further extend the distance between the central office and the customer installation.

The ADSL2+ maximum capacity has been extended to approximately 16 Mbit/s downstream. As an option downstream bit rates above 16 Mbit/s and upstream rates above 800 kbit/s might be supported.

There are significant differences between VDSL and VDSL2.

The first version, VDSL, offers bit rates up to 70 Mbit/s downstream and up to 30 Mbit/s upstream. The reach is low, approximately 400 m. An optional synchronous mode is possible.

VDSL2 offers higher bit rates and longer reach than VDSL. Downstream bit rates above 200 Mbit/s on short lines has been indicated as is symmetrical bit rates around 100 Mbit/s. The reach can be extended to up to 2 500 m of 0,4 mm cable.

The VDSL2 standard defines eight different profiles for different application areas. At least one profile must be supported by a VDSL2 compliant device.

The VDSL2 payload may be divided into two equal parts; fast channel for latency sensitive services, and slow channel (mandatory). The slow and fast channel alternate.

4.3.5.2 Symmetrical DSL

There are two standardized symmetrical DSL technologies:

- HDSL (High bit rate Digital Subscriber Line) [35].
- SHDSL (Single-pair High-Bitrate Digital Subscriber Line) [36].

HDSL is a four-wire system first developed in North America. The system transmission capacity was 776 kbit/s. Later a European system providing 776 kbit/s, 1 160 kbit/s or 2 312 kbit/s on 1, 2 or 3 pairs was standardized.

SHDSL is a flexible multi-rate system that supports bit rates from 192 kbit/s to 2 312 kbit/s on a single pair (2 wire). SHDSL uses the entire frequency band; no PSTN or ISDN transmission is possible.

4.3.6 Wireless access

4.3.6.1 2G Mobile access

2G mobile access (GSM) is basically a circuit-switched voice access with an additional data channel offering a limited circuit-switched data transmission capacity. In GSM phase 2+ General Packet Radio Services (GPRS) [37] provides higher data rates. The available GPRS data rates increase up to 160 kbit/s downstream, depending on implementation. The upstream capacity is lower. It is possible to use this data channels for multimedia applications, but due to long delay it is not recommended to use the GPRS channel for two-way interactive real-time communication (e.g. telephony or videotelephony).

A further evolution, EDGE (Enhanced Data rate for GSM Evolution) offer higher data transmission capacity than GPRS. EDGE is realized via modifications of the existing layer 1 specifications [i.22] rather than by separate, stand-alone specifications. Like GPRS the throughput depends on the radio transmission conditions. At worst radio transmission conditions the EDGE throughput is marginally better than GPRS, at the best radio transmission conditions the throughput might be up to 384 kbit/s downstream. Like GPRS, the EDGE data channel delay does not make EDGE suitable for two-way real-time multimedia applications.

4.3.6.2 3G Mobile access

ITU has defined five 3G technologies. In Europe the term 3G is associated with W-CDMA (Wideband-Code Division Multiple Access) access. The maximum data rate was 384 kbit/s downstream and 64 kbit/s upstream when the system was introduced. Later enhancements have increased the capacity, maximum data rate may be 2 Mbit/s downstream and 384 kbit/s upstream. However, the effective bit rate is lower and implementation dependant.

The first 3G releases developed by 3GPP defined Circuit Switched voice and audiovisual applications. These are outside the scope of the present document. From 3GPP release 5 an IP multimedia core network subsystem (IMS) [38] is defined. An IMS telephony service is defined in TS 122 173 [39].

HSPA (High Speed Packet Access) is an extension to 3G offering higher data transmission capacity. The standards have been developed in two steps; HSDPA (High Speed Downlink Packet Access) [40] is providing downlink data rates up to 14 Mbit/s. Later HSUPA (High Speed Uplink Packet Access) [41] (see note) providing uplink data rates up to 5,8 Mbit/s are standardized.

NOTE: Another acronym used is EUL (Enhanched Uplink).

The 3G QoS framework and architecture is specified in TS 123 107 [42]. The scope of the document includes the UMTS Bearer Service which consists of two parts, the Radio Access Bearer Service and the Core Network Bearer Service.

The document defines four QoS classes:

- conversational class;
- streaming class;
- interactive class;
- background class.

The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: Conversational class is meant for traffic which is very delay sensitive while Background class is the most delay insensitive traffic class. Conversational and Streaming classes are mainly intended to be used to carry real-time traffic flows.

TS 123 207 [43] provides the framework for end-to-end Quality of Service involving GPRS and complements TS 123 107 [42] which describes the framework for Quality of Service within UMTS. The document is only applicable to GPRS packet switched access services.

4.3.6.3 DECT

Digital Enhanced Cordless Telecommunications (DECT) is based on a micro-cellular radio communication system that provides low-power radio (cordless) access between PPs and (DECT) FPs at ranges up to a few hundred metres (up to several kms for fixed access systems). The basic technical characteristics are as follows:

- frequency band: 1 880 MHz to 1 980 MHz and 2 010 MHz to 2 025 MHz (see note 1);
- number of carriers: typical 10 (see note 1);
- carrier spacing: 1 728 MHz (see note 1);
- maximum peak transmit power: 250 mW (see note 1);
- carrier multiplex: TDMA; 12 double slots/24 full slots/48 half slots per frame;
- frame length: 10 ms;
- basic duplexing: TDD using 2 slots on same RF carrier;
- gross bit rate: 1 152 kbit/s, 2 304 kbit/s, 3 456 kbit/s, 4 608 kbit/s or 6 912 kbit/s for 2-, 4-, 8-, 16- or 64-level modulation respectively (see note 2);
- net channel rates: 6,4 kbit/s A-field (control/signalling) per slot.

B-field (traffic) rates per slot are described in table 7.

Table 7: DECT B-field (traffic) rate per slot

Type of modulation	Maximum B-field (traffic) rate per slot					Maximum asymmetric B-field (traffic) data rate (11 double slots)
	half slot (j = 80)	long slot (j = 640)	long slot (j = 672)	full slot	double slot	
2-level modulation	8 kbit/s	64 kbit/s	67,2 kbit/s	32 kbit/s	80 kbit/s	880 kbit/s
4-level modulation	16 kbit/s	128 kbit/s	134,4 kbit/s	64 kbit/s	160 kbit/s	1 760 kbit/s
8-level modulation	24 kbit/s	192 kbit/s	201,6 kbit/s	96 kbit/s	240 kbit/s	2 640 kbit/s
16-level modulation	32 kbit/s	256 kbit/s	268,8 kbit/s	128 kbit/s	320 kbit/s	3 520 kbit/s
64-level modulation	48 kbit/s	384 kbit/s	403,2 kbit/s	192 kbit/s	480 kbit/s	5 280 kbit/s

NOTE 1: The complete definition of frequency bands and carrier positions for DECT are found in EN 300 175-2 [44]. DECT is a member of the IMT-2000 family, the only member that provides for uncoordinated installations on an unlicensed spectrum. The most common spectrum allocation is 1 880 MHz to 1 900 MHz, but outside Europe spectrum is also available in 1 900 MHz to 1 920 MHz and in 1 910 MHz to 1 930 MHz (several countries). Carrier positions in the 902 MHz to 928 MHz and 2 400 MHz to 2 483,5 MHz ISM bands have been defined for the US market. New or modified carrier positions and/or frequency bands can be defined when needed. The number of carriers depends on the frequency spectrum in use and the carrier spacing. Maximum peak transmit power also depends on local regulations or environment requirements.

NOTE 2: Depending on radio capabilities or number of radios used a DECT system can provide higher data rate. The indicated here values are relevant for radius operating on 10 carriers and non overlapping slots.

A connection is provided by transmitting bursts of data in the defined time slots. These may be used to provide simplex or duplex communications. Duplex operation uses one or several pairs of evenly (5 ms) spaced slots. Of the paired slots one is for transmit and one for receive.

The simplest duplex service uses a single pair of time slots to provide e.g. a 32 kbit/s (2-level modulation) digital information channel capable of carrying coded speech or other low rate digital data. Higher data rates are achieved by using more time slots in the TDMA structure, and a lower data rate may be achieved by using half-slot data bursts. Different uplink and downlink bit rates are realized by using asymmetric connections, where a different number of time slots is used for the uplink and downlink. For efficient transmission of packet data the radio connection can be suspended after the data has been sent and as soon as new data arrives, the radio connection is resumed again.

DECT is able to support a number of alternative system configurations ranging from single cell equipment (e.g. domestic FPs) to large multiple cell installations (e.g. large business cordless PBXs or converged IP based business systems), public pedestrian systems and fixed wireless access (radio local loop) systems.

The protocols are designed to support uncoordinated system installation, even where the systems co-exist in the same physical location. Efficient sharing of the radio spectrum (of the physical channels) is achieved using a careful mechanism for selection of channels prior to their use. This is called dynamic channel selection (see ETR 310 [i.1]).

In addition, the DECT protocols provide two internal mechanisms to support rapid handover of calls in progress (both intracell and intercell handover are supported). These handover mechanisms allows a high quality of service to be maintained where the mobility of the PP requires transparent re-connection to another FP or where a new physical channel is required in response to a disturbances in the radio environment.

Wireless Relay Stations (WRSs) for wireless coverage enhancements, direct communication from PT to PT and wireless communication between FTs is also supported.

DECT is an access technology providing sufficient flexibility for access to various communication networks, e.g. IP, PSTN, ISDN, LAN, GSM, UMTS, etc.

4.3.6.4 Broadband Wireless Access

WiMax is a Broadband Wireless Access (BWA) technology specified in IEEE 802.16 [45]. IEEE 802.16 [45] is a family of standards covering different modes of operation:

- fixed point-to-point;
- fixed point to multipoint;
- mobile WiMAX.

Two transmission techniques can be used. Frequency-division duplex (FDD) is where downlink and uplink subframes occur simultaneously on separate frequencies, and time-division duplex (TDD) is where downlink and uplink subframes occur at different times and usually share the same frequency. Subscriber Stations (SS) can be either full duplex (i.e. they can transmit and receive simultaneously) or half-duplex (i.e. they can transmit and receive at nonoverlapping time intervals).

With point to multipoint (PMP), the BS serves a set of SSs within the same antenna sector in a broadcast manner, with all SSs receiving the same transmission from the BS. Transmissions from SSs are directed to and centrally coordinated by the BS. On the other hand, in mesh mode, traffic can be routed through other SSs and can occur directly among SSs.

Access coordination is distributed among the SSs. The PMP operational mode fits a typical fixed BWA scenario, where multiple service subscribers are served by one centralized service provider so that they can access external networks.

Typical WiMax bit rate is 70 Mbit/s, maximum bit rate is 268 Mbit/s.

4.3.6.5 WLAN

WLAN, also referred to as WiFi, can be seen as a wireless alternative to Local Area Networks (LAN). The most frequently used systems are based on IEEE 802.11 [46].

The IEEE 802.11 [46] standard specifies use of two frequency bands:

- 2,4 GHz specified in IEEE 802.11b (11 Mbit/s), IEEE 802.11g (54 Mbit/s) and IEEE 802.11n (see note 1). These can interwork, but when 802.11g co-exists with 802.11b, the capacity is reduced. This band is license free (see note 2).
- 5 GHz specified in IEEE 802.11a (54 Mbit/s). There is a European version specified in IEEE 802.11h and a Japanese version specified in IEEE 802.11j.
  There are restrictions on the use of the 5 GHz band, it cannot be used outdoors.

NOTE 1: IEEE 802.11n is a new planned standard that will work in both the 2,4 GHz and the 5 GHz frequency bands. Thee maximum transmission capacity is 300 Mbit/s. It is expected published in 2009.

NOTE 2: For unlicensed usage of the 2,4 GHz frequency band there are radio power emission restrictions.

It is important to note that the capacity indicated is the theoretical capacity. In a WLAN network the actual throughput will depend on:

- The radio transmission environment.
- The characteristics of the information to be transmitted (i.e. packet size).
- The number of transmitters active.

The actual throughput will therefore be lower than the maximum rate indicated.

It is possible to use the technology to communicate directly between WLAN terminals or via an Access Point (AP). In the context of the present document, only communication via an AP is considered.

4.3.7 Broadcasting

There are several packet-based broadcasting systems standardized; Digital Audio Broadcasting (DAB), Digital Multimedia Broadcasting (DMB) and Digital Video Broadcasting (DVB).

DAB is a technology for audio broadcasting using digital radio transmission. DAB is specified in EN 300 401 [47]. The standard defines a broadcasting system designed for delivery of high-quality digital audio programme and data services for mobile, portable and fixed reception from terrestrial or satellite transmitters in the Very High Frequency (VHF)/Ultra High Frequency (UHF) frequency bands as well as for distribution through cable networks.

The first version of DAB, published in 1997 specifies use of a MPEG Layer II audio encoding algorithm. A new version of the standard, often referred to as DAB+, was published in 2007. This version specifies use of the MPEG AAC+ (Eaac+) audio codec, is not backward compatible with the 1997 version.

DMB [48] is an extension of the DAB system. DMB can be broadcast from both terrestrial and satellite based transmitters.

Digital Video Broadcasting (DVB) is a group of standards specifying systems for transmitting multimedia information over:

- Satellite networks (DVB-S2) [49].
- Terrestrial networks (DVB-T) [50].
- Cable networks (DVB-C) [51].

- To handheld terminals (mobiles) over terrestrial networks (DVB-H) [52].
- To handheld terminals (mobiles) over a combination of satellite and terrestrial networks (DVB-SH) [53].

Initially the audio and video transport were based in the MPEG 2 transport stream (MPEG 2 TS) [17]. Implementation guidelines for this solution is described in TS 101 154 [54]. MPEG-1 [16] or MPEG-2 [17] layer 2 backward compatible audio is specified.

Implementation guidelines for the use of H.264/AVC and High Efficiency AAC for DVB compliant delivery in RTP packets over IP networks is described in TS 102 005 [55].

4.4 Network performance parameters

4.4.1 Transmission bandwidth

4.4.1.1 General considerations

When transmitting speech (or other media) over a packet network, headers are added to each packet. The Real-time Transport Protocol (RTP) defined in RFC 3550 [57] adds, as a minimum, 40 bytes header to each packet in IP version 4 (see note). For 20 ms packet intervals this corresponds to 16 kbit/s transmission requirement increase.

NOTE: IP headeer 20 bytes, UDP header 8 bytes aand RTP header 12 bytes.

Where IP version 6 is used the header added to each packet is minimum 60 bytes.

The required transmission bandwidth can be found by adding the headers identified above plus technology dependent link layer headers (e.g. 14 bytes for an Ethernet link) and any tails (e.g. 4 CRC bytes for an Ethernet link) to the encoded media bit rate.

To reduce the overhead, header compression algorithms can be used. 3GPP has specified an optional use of an algorithm described in RFC 3095 [58] on the wireless link of the 3GPP system. Using this algorithms the IP/UDP/RTP headers might be compressed to 4 bytes or less.

4.4.1.2 Speech transmission

For some of the standardized speech coding algorithms listed in tables 1 and 2, several modes of operation having different transmission bandwidth requirements are defined. These codecs can be considered as Constant Bit Rate (CBR), i.e. when they operate in a defined mode the bit rate is constant. The required transmission bandwidth can be changed by changing mode.

The bit rates given in tables 1 and 2 are describing the encoded speech bitstream. When transmitting over an IP network, the bitstream is split into separate packets. Each packet consist of a header as described above and the payload. The payload size depends on the packet interval chosen. TS 181 005 [59] specifies 10 ms packet intervals when the G.711 coding algorithm is used. There are frame based codecs where the frame size is larger, e.g. 20 ms for the AMR and AMR-WB, and 30 ms for the G.723.1 codec.

When using the G.711 and 10 ms packet interval, the transmission bandwidth requirements at IP level is 96 kbit/s. To find the transmission system bandwidth requirement, lower layer headers needs to be added. These depends on the transmission technology used.

4.4.1.3 Audio transmission

There are a number of profiles and operating modes defined for audio coding algorithms standardized by ISO/IEC. The bit rates varies from 16 kbit/s up to 288 kbit/s per channel. The quality ratings as a function of transmission bit rate for three different AAC modes (stereo) are depicted in figure 6. The input sampling rate was 44,1 kHz.

There is a difference between the AAC algorithm and the HE-AAC extensions for lower bit transmission rates (i.e. below 100 kbit/s). At 64 kbit/s the difference is 18 points on the MUSHRA [60] scale (i.e. approximately one grade).

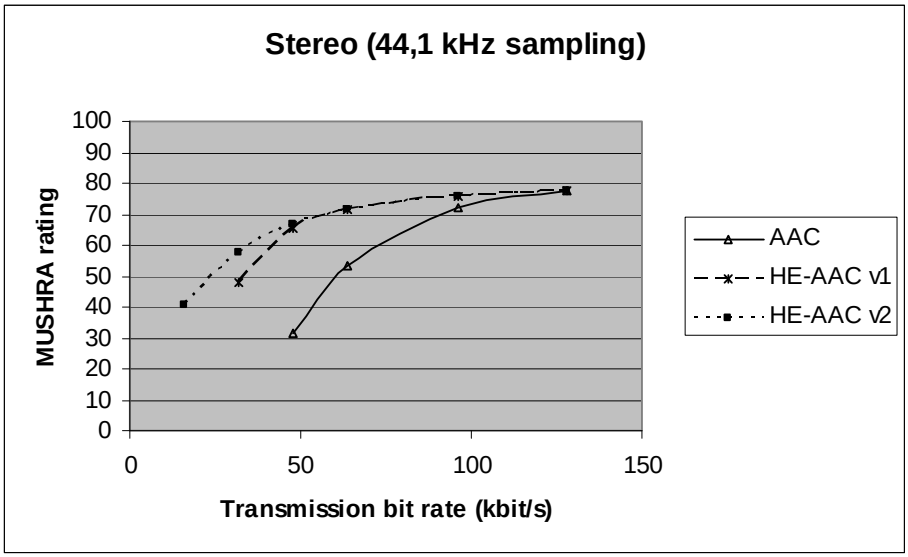


Figure 6: MUSHRA rating as a function of transmission bit rate for three AAC codec versions defined in MPEG-4

The AMR-WB+ codec transmits information at bit rates between 5,2 kbit/s and 36 kbit/s in mono, and between 6,2 kbit/s and 36 kbit/s in stereo. The codec is scalable and backward compatible with AMR-WB. A comparison between the AMR-WB+ codec and the HE-AAC v2 (Eaac+), mono signals, is depicted in figure 7. The input sampling rate was 48 kHz.

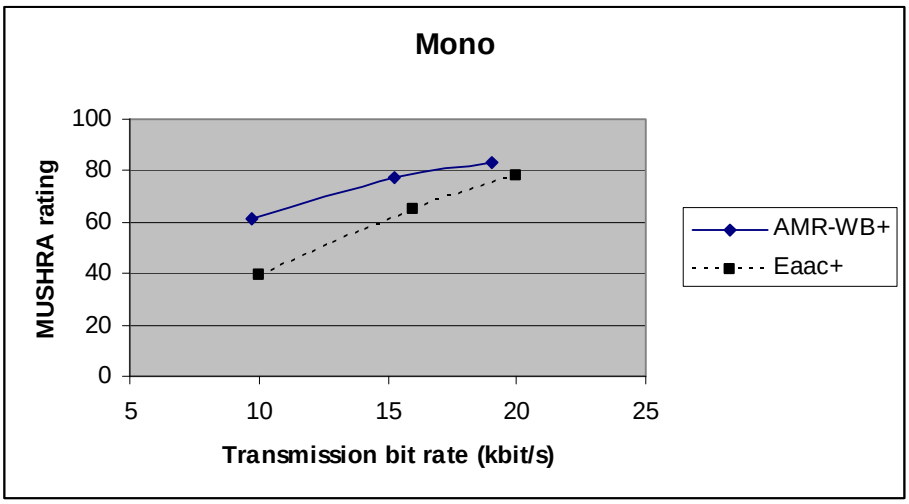


Figure 7: MUSHRA rating as a function of transmission bit rate for AMR-WB+ and HE-AAC v2 (Eaac+), mono

Figure 8 depicts the comparison for stereo signals.

While the AMR-WB+ algorithm [19] performs better than the HE-AAC v2 [18] algorithm for mono signals at low transmission bit rates, the ratings are similar for stereo signals. It should be noted that the information presented is based on average ratings for different types of signals. There are variations depending on type of signal.

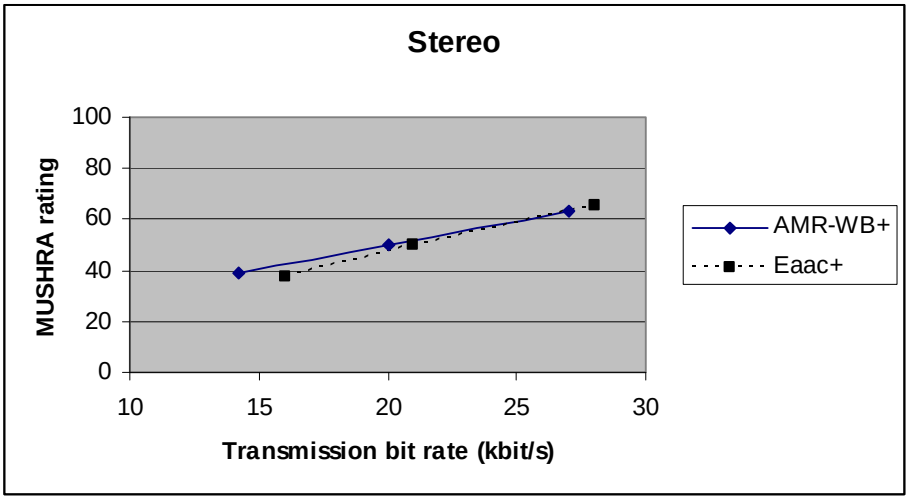


Figure 8: MUSHRA rating as a function of transmission bit rate for AMR-WB+ and HE-AAC v2 (Eaac+), stereo

Like speech the Real-time Transport Protocol is used for transport of audio information over the network. The headers identified in clause 4.4.1.1 need to be added for each packet sent. Most audio applications are less delay sensitive than conversational speech. The packet intervals may therefore be larger, thus reducing the packet overhead.

4.4.1.4 Video transmission

The video coding algorithms are Variable Bit rate (VBR). The actual bit rate is depending on a number of factors:

- the image content (e.g. amount of details) and changes;
- the video coding principles;
- the video frame rate;
- the video spatial resolution;
- the codec quantization parameter setting.

The video codec profile levels defined in tables 3, 4, 5 and 6 specify the max bit rate that needs to be supported. The actual bit rate might be lower.

Like speech and audio the Real-time Transport Protocol is used for transport of video information over the network. The headers identified in clause 4.4.1.1 need to be added for each packet sent.

4.4.2 Packet loss

4.4.2.1 General considerations

Media (e.g. speech, audio or video) are usually transported over IP networks using an unreliable protocol which does not guarantee that packets are delivered or delivered in order. Packets may be dropped under peak loads and during periods of congestion (caused, for example, by link failures or inadequate capacity). Packets may also be dropped at the receiving device due to jitter buffer overflow.

4.4.2.2 Speech transmission

The codecs identified in tables 1 and 2 are initially designed for use on circuit-switched networks. For some error recovery mechanism might be included. The effect of packet loss on user perceived speech quality in terms of MOS rating without use of error recovery is indicated in figure 9 for two of the narrowband codecs listed in table 1. The slope of the MOS rating vs. packet loss curve is different for different codecs, which means that the effect of packet loss depends on the codec algorithm used. To estimate this effect the algorithm described in ITU-T Recommendation G.107 [61] (E model) can be used.

NOTE: Provisional parameter values required to estimate the effect of packet loss on wideband codec perceived quality are available in the 2008 version of the E-model documentation.

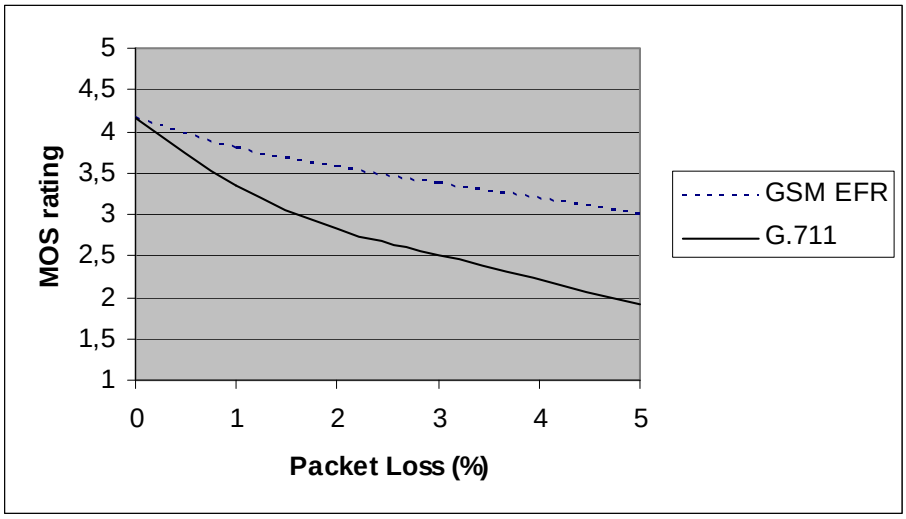


Figure 9: Effect of packet loss on speech MOS rating

To limit the degradations caused by packet loss, optional error recovery mechanisms are specified for some of these codecs.

4.4.2.3 Audio transmission

In ITU-T Recommendation G.1010 [62] performance target for high quality streaming audio is indicated to be less than 1 % PLR. A note in table I.1/G.1010 points out that exact values depend on specific codec, but assumes use of a packet loss concealment algorithm to minimize effect of packet loss.

- The amount of degradation caused by packet loss depends on several factors.
- The codec standard and the codec implementation.
- The choice of codec parameters (e.g. mono coding or stereo coding) and transmission bit rate.
- The audio material.
- The packet loss characteristics.

The effect of random packet loss on user perceived audio quality in terms of MUSHRA [60] rating is indicated in figure 10 for mono presentation. It can be seen that that there is a non-linear relation between the MUSHRA [60] rating and the packet loss ratio. The trend is similar for stereo presentation, but the degradation effect is larger; tests carried out have indicated that the degradation at 10 % packet loss is between 10 and 20 point larger than the corresponding mono presentation degradation, corresponding to between 0,5 and 1 point on a MOS scale.

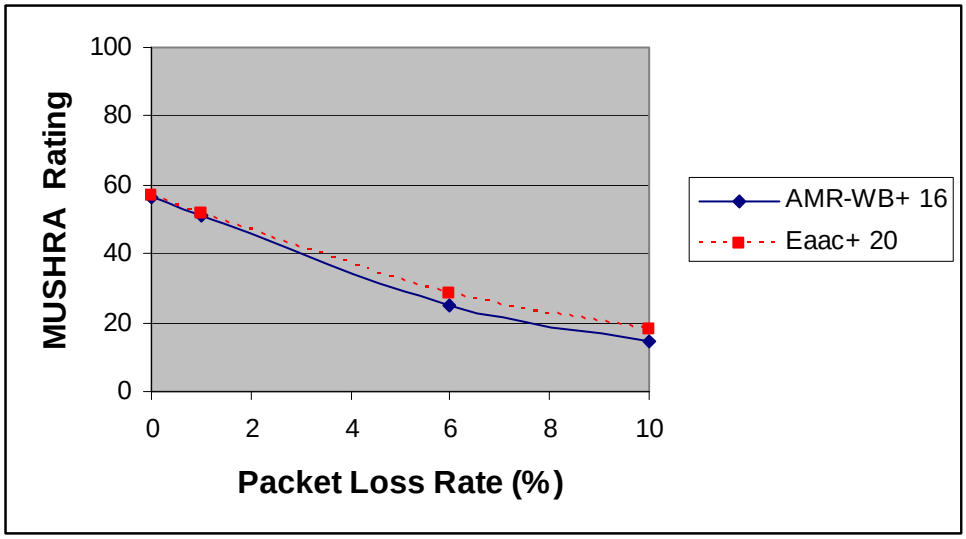


Figure 10: Effect of random packet loss on user perceived audio quality

4.4.2.4 Video transmission

Video streams are sensitive to packet loss. In ITU-T Recommendation G.1010 [62] performance target for video is indicated to be less than 1 % PLR. A note in table I.1/G.1010 points out that exact values depend on specific codec, but assumes use of a packet loss concealment algorithm to minimize effect of packet loss.

The amount of degradation caused by packet loss depends on several factors:

- the codec standard and the codec implementation;
- the choice of codec parameters;
- the picture resolution and picture quality;
- the packet loss characteristics.

The video degradation caused by a lost packet differs depending on the actual frame (B frame, I frame or P frame) the lost packet contains. When an I frame or a P frame is lost, the effects propagates to subsequent frames that uses the lost frame as a reference.

The effect of random packet loss on user perceived video quality in terms of MOS rating is indicated in figure 11 for two transmission bit rates (and quality level) using the same codec. It can be seen that the degradation is largest for high quality, and that there is a non-linear relation between the MOS rating and the packet loss ratio.

The packet loss frequency (i.e. the interval between each packet lost) affects the video quality. Burst packet loss might therefore be less annoying than random packet loss at the same packet loss rate.

ITU-T Recommendation G.1010 [76] provides guidelines on the selection of appropriate objective perceptual video quality measurement methods when a full reference signal is available. The following are example applications that can use ITU-T Recommendation G.1010 [77]:

1) Internet multimedia streaming.
2) Video telephony and conferencing over cable and other networks.
3) Progressive video television streams viewed on LCD monitors over cable networks including those transmitted over the Internet using Internet Protocol. (VGA was the maximum resolution in the validation test).
4) Mobile video streaming over telecommunications networks.
5) Some forms of IPTV video payloads (VGA was the maximum resolution in this validation test).

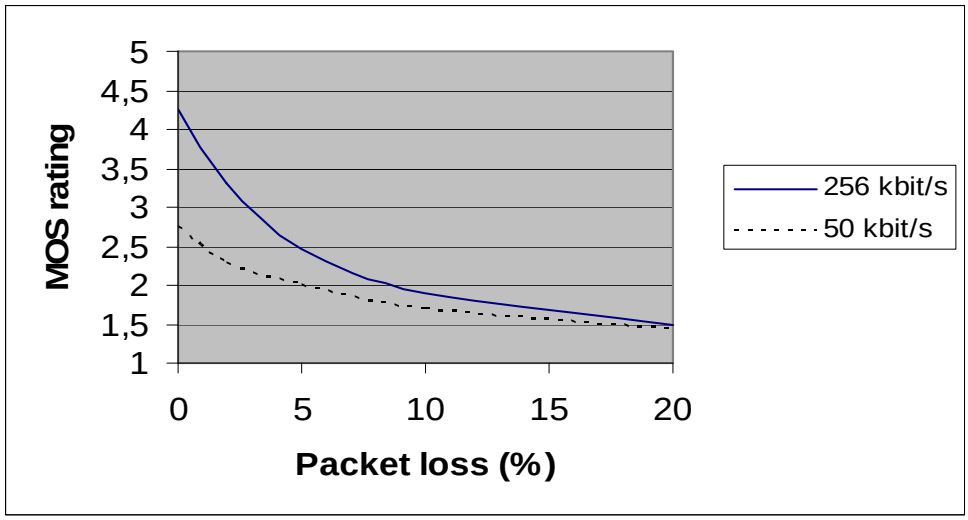


Figure 11: Effects of packet loss on video MOS rating

4.4.3 Transmission delay

4.4.3.1 General

The delay sources of an IP network connection are:

- transmitting terminal delay;
- access network delay;
- core network delay;
- receiving terminal delay.

The transmission delay of interactive voice and audiovisual communication shall meet the requirements specified in ITU-T Recommendation G.114 [56]. For non-interactive applications such as media streaming, delay is less disturbing, but not insignificant.

4.4.3.2 Transmitting terminal delay

The algorithmic delay of speech codecs is addressed in ITU-T Recommendation G.114 [56].

The minimum speech codec algorithmic delay of a sample based coded is equal to the packetizing interval.

The minimum speech codec algorithmic delay of a frame-based codec used in an IP system can be estimated using the formula:

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

where N is the number of frames included in a single packet.

In default operation the ITU-T Recommendation G.729.1 [9] coder has an algorithmic delay of 48,9375 ms.

The algorithmic delay of AAC codecs may depend on the sampling rate as depicted in figure 12. However, the algorithmic delay of the AAC-LD is almost constant (approximately 20 ms).

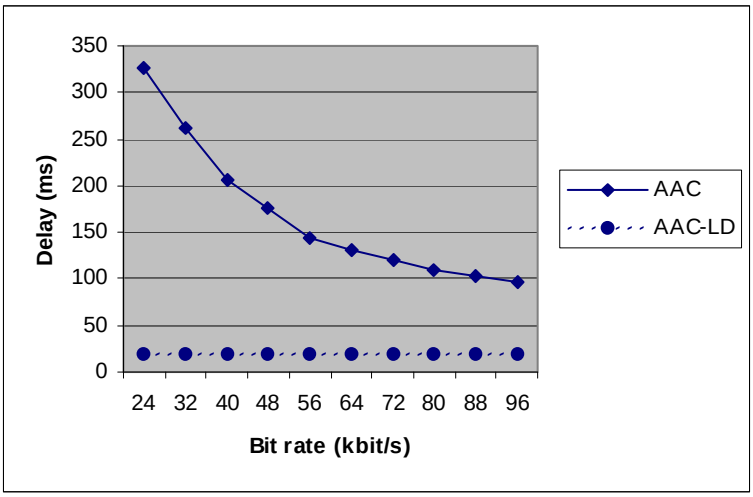


Figure 12: MPEG AAC algorithm delay

The delay of the AMR-WB+ algorithm [19] is shown in table 8.

Table 8: AMR-WB+ codec delay

	Internal Sampling Frequency = 12,8 kHz	Internal Sampling Frequency = 25,6 kHz	Internal Sampling Frequency = 38,4 kHz
Mono	227,5625 ms	113,7813 ms	77,5208 ms
Stereo	325,6785 ms	162,8438 ms	108,5625 ms

Video coding is more challenging than speech and audio coding, and introduces more delay than speech/audio coding. A frequently used estimate for the video codec delay is:

$$Video\,coding\,delay\,(ms) = 3\frac{1000}{Video\ \ frame\,rate}$$

To achieve acceptable audio/video synchrony, the speech and audio signals need to be delayed.

4.4.3.3 Access network delay

The access network delay depends on the actual technology used and the performance parameters (e.g. transmission speed). Examples are presented in table 9.

NOTE: The values presented in table 9 should be taken as an indication. The actual delay depends on implementation, traffic conditions, any QoS mechanisms and packet size.

Table 9: Access network delay examples

Access technology	Delay (ms)
ADSL (see note 1)	≤ 2 (see note 2)
VDSL (see note 1)	≤ 2 (see note 3)
SHDSL (see note 1)	≤ 2
HDSL	≤ 2
GPRS	600 to 700 (see note 4)
EDGE	400 to 500
Enhanced EDGE	150 (see note 5)
3G	80 (see note 6)
HSPA	Downstream (HSDPA): 50 Upstream (HSUPA): 50
DECT Enterprise network	14
WLAN	< 5
WiMAX	Downstream: 10 Upstream: 25
NOTE 1: Interleaving may be applied to improve the noise robustness. The interleaving depth has to be added to these numbers (valid for all DSL technologies). NOTE 2: The actual delay depends on the transmission speed and packet size. Downstream delay is normally lower than upstream delay. NOTE 3: Low delay mode. NOTE 4: RTT Terminal - Server - Terminal. Upstream delay is larger than downstream delay. Some papers indicate lower delays. NOTE 5: Some suppliers indicate a lower delay. NOTE 6: For 200 bytes packets. For smaller packets the delay is lower (e.g. 60 ms for 32 bytes packets).	

4.4.3.4 Core network delay

The core network delay sources are the delay caused at each router of the network connection and the propagation delay.

The router delay consists of queuing delay and processing delay. The queuing delay is a function of the router load while the processing delay is a function of the router processing capacity.

The propagation delay depends on the technology used. Table A.1/G.114 of ITU-T Recommendation G.114 [56] presents planning values for calculating propagation delay for various transmission technologies.

4.4.3.5 Receiving terminal delay

The main receiving delay source is the playout (jitter) buffer, see clause 4.5.2.

4.4.4 Transmission delay variations (jitter)

Transmission delay variations (jitter) are caused by queuing in network elements or by routing the packets along different network paths. A significant multimedia delay variation source is the simultaneous transmission of packets containing information for two or more media; e.g. voice and video. In serial access networks such as xDSL systems only one packet can be transmitted at an instant, other packets has to be delayed until the transmission link is free.

In wireless systems (e.g. WLAN) several connections share the wireless link. To get access to the wireless link a queuing mechanism is implemented. This mechanism increases the delay variation. The number of participants sharing the wireless link influences the amount of delay variation.

4.5 Terminal characteristics

4.5.1 Packet loss recovery

Several packet loss recovery approaches are possible. Some speech coding algorithms have defined error concealment mechanisms as an integral part of the algorithm. Packet loss recovery mechanisms have been standardized as an option for other algorithms.

NOTE: Annex C provides information about packet loss recovery mechanisms.

4.5.2 Playout buffer

The objective of a playout (jitter) buffer is to ensure continuous playout of audio and video information to the user. The incoming media packet are stored in a buffer to allow the packets to arrive in time to be played in correct sequence. Packets that arrive later is rejected (lost). Jitter buffers cause additional connection delay. To minimize the delay, the jitter buffer should be as short as possible. On the other hand, if the jitter buffer is too short, the packet loss increases, causing user perceived media quality degradation. The amount of jitter may vary for each connection set up. It may also vary during a communication session.

There are two jitter buffer design options:

- fixed jitter buffer;
- adaptive jitter buffer.

The fixed jitter buffer adds a constant delay to a connecting, while the delay added by an adaptive jitter buffer adds a delay that depends on the connection delay variation (jitter).

4.5.3 Audio characteristics

Audio terminal principles used are:

- handset;
- headset;
- microphone/loudspeaker (Handsfree or loudspeaking).

Telephony applications might be narrowband (300 Hz to 3 400 Hz) or wideband (50 Hz to 7 000 Hz).

Four ETSI Standards specify performance requirements for VoIP terminals. ES 202 737 [63] addresses narrowband handset and headset terminals; ES 202 738 [64] addresses narrowband loudspeaking and handsfree terminals; ES 202 739 [65] addressees wideband handset and headset terminals and ES 202 740 [66] addresses wideband loudspeaking and handsfree terminals.

NOTE: There is no ETSI documents specifying the audio transmission characteristics of VoIP terminals or terminals for other audiovisual applications over IP networks.

4.5.4 Video display

There are three video display technologies used:

- CRT.
- LCD.
- Plasma technology.

The display size may vary from less than 2" on a mobile handset to more than 70". Issues such as type of screen (CRT, LCD, plasma, etc.), size of screen, and image resolution can all affect perceived video quality. Some impairments, are generally considered tolerable, if noticeable at all, on a small/medium to low resolution screen, but to be more pronounced and objectionable when viewed on a large/high resolution screen.

There are indications that users rate video on a large display higher than video on a smaller display. It is however difficult to quantify the difference.

Another quality element is the characteristics of the Human Visual System (HVS).

There are a number of display resolution formats supported by one or several video coding algorithms:

- Sub-QCIF (128 × 96) pixels;
- QCIF (176 × 144) pixels;
- CIF (352 × 288) pixels;
- 4CIF (704 × 576) pixels;
- SIF (384 × 288) pixels;
- QVGA (320 × 240) pixels;
- VGA (640 × 480) pixels;
- CCIR 601 (720 × 576) pixels.

NOTE 1: State of the art signal processor are not capable of encoding 4CIF and CCIR 601 real-time.

NOTE 2: There is no ETSI documents specifying the video transmission characteristics of terminals for audiovisual applications over IP networks.

4.6 Audio-video interaction

Prediction of audiovisual quality from audio only and video only quality ratings shall use the following models:

1) Delay insensitive applications presenting low motion video

$$MOS_{AV} = \alpha (MOS_A * MOS_V) + C$$

2) Delay insensitive applications presenting high motion video

$$MOS_{AV} = \alpha MOS_A + \beta MOS_V + \gamma (MOS_A * MOS_V) + C$$

3) Delay sensitive (conversational) applications

$$MOS_{AV} = \alpha MOS_V - \theta GD + C$$

The parameter *GD* is the Global Delay expressed in seconds.

Information about parameter values are presented in annex A.

5 Audiovisual applications classification

The audiovisual applications can be classified into two groups:

- delay sensitive applications;
- delay insensitive applications.

Further considerations are made in the following clauses.

5.1 Delay sensitive applications

The delay sensitive audiovisual applications consist of the interactive real-time applications telephony, audio conferencing, video telephony and video conferencing. Low end-to-end delay is important to achieve acceptable user perceived quality.

5.2 Delay insensitive applications

The delay insensitive audiovisual applications consist of applications where no conversational element is involved: examples are downloading of audio/video files, streaming audio/video and audio/video broadcasting.

6 Delay sensitive audiovisual applications requirements

6.1 Coding algorithms

6.1.1 Narrowband speech

TS 181 005 [59] specifies narrowband speech codecs and codec characteristics to be used by network elements in TISPAN defined NGNs.

In order to enable interworking between the NGN and other networks (including the PSTN, mobile networks and other NGNs) the NGN must be capable of receiving and presenting G.711 [78] coded speech when interconnected with another network. When a packetization size is not selected by codec negotiation between terminals and/or network elements or agreed by bilateral arrangement, a speech packetization size of 10 ms samples should be used for G.711 [79] coded speech; this is recommended as an optimum value balancing end-to-end delay with network utilization. It is recognized that there may be network constraints which require that a higher value is agreed by bilateral arrangement; in such cases a value of 20 ms is recommended.

NOTE 1: Where a packetization size is selected by codec negotiation between terminals and/or network elements the present document places no requirements on the value to be selected.

NOTE 2: The above does not put any requirement about the codecs to be supported by terminals nor does it mandate that NGN networks shall support speech transcoding between any arbitrary codec to G.711 [80].

In addition, support for the following speech codecs is recommended:

- AMR [10] in order to support 3GPP terminals and to facilitate the interwork with 3GPP network.
- G.729A [8] in order to facilitate the interwork with existing VoIP networks and support existing VoIP terminals.
- EVRC/EVRC-B [11] in order to support 3GPP2 terminals and to facilitate interworking with 3GPP2 networks.

TS 126 235 [67] specifies the set of default codecs for packet switched conversational multimedia applications within 3GPP IP Multimedia Subsystem. The narrowband AMR codec [10] is mandatory.

Transcoding between different coding algorithms should be avoided.

To minimize the end to end delay 10 ms or 20 ms packet intervals are recommended for voice only applications. For audiovisual applications the audio/video synchronization requirement and longer video encoding times make it possible to increase the speech packet intervals without further performance degradation. The packet intervals should not exceed the video encoding delay.

The speech transmission performance can be estimated using the E-model [61].

6.1.2 Wideband speech

TS 181 005 [59] specifies narrowband speech codecs and codec characteristics to be used by network elements in TISPAN defined NGNs.

Terminals and network elements originating and terminating end to end NGN IP media flows, supporting wideband speech should provide one or more of the following wideband speech codecs:

- G.722 [4].
- G.729.1 [9].
- AMR-WB [14].
- EVRC+WB [11].

The wideband AMR codec [14] shall be used when wideband speech is provided on 3GPP networks.

Transcoding between different coding algorithms should be avoided.

To minimize the end to end delay 10 ms or 20 ms packet intervals are recommended for voice only applications. For audiovisual applications the audio/video synchronization requirement and longer video encoding times make it possible to increase the speech packet intervals without further performance degradation. The packet intervals should not exceed the video encoding delay.

The speech transmission performance can be estimated using the E-model [61].

NOTE: The Wideband parameter values presented in the 2008 version of the E-model [61] are provisional.

6.1.3 Video

TS 181 005 [59] recommends that ITU-T Recommendations H.263 [24] profile 0 and H.264 (AVC) [25] baseline profile are supported by TISPAN NGNs.

NOTE: The above does not put any requirement about the codecs to be supported by terminals nor does it mandate that NGN networks shall support video transcoding between any arbitrary codec and ITU-T Recommendations H.263 [24] or H.264 (AVC) [25].

TS 126 235 [67] specifies that 3G PS multimedia terminals offering video communication shall support ITU-T Recommendation H.263 [24] profile 0, level 45.

The following video algorithms/profiles should be supported:

- H.263 [24] version 2 Interactive and Streaming Wireless Profile (Profile 3) Level 45;
- ISO/IEC 14496-2 [18] (MPEG-4 Visual) Simple Profile at Level 0b;
- H.264 (AVC) [25] Baseline Profile at Level 1b without requirements on output timing conformance.

ITU-T Recommendation J.247 [68] provides video quality estimations for video classes TV3 to MM5B, as defined in ITU-T Recommendation P.911 [69], annex B. The maximum resolution is VGA and the maximum bit rate covered well in the test was 4 Mb/s. The model is a full reference model; i.e. a copy of the source is required to estimate the quality.

Within the limitations indicated above, the video performance can be estimated using one of the algorithms specified in ITU-T Recommendation J.247 [68].

6.2 Network performance requirements

ITU-T Recommendation Y.1541 [2] defines classes of network Quality of Service (QoS) with objectives for Internet Protocol network performance parameters measured between user network interfaces (UNI). Two classes are defined for real-time, jitter sensitive, high interaction applications (VoIP, VTC). Class 0 applies for constrained routing and distance while class 1 apply for less constrained routings and distances. The values are presented in table 10.

Table 10: Network performance objectives defined in ITU-T Recommendation Y.1541 for delay sensitive audiovisual applications

QoS class	IPTD	IPDV	IPLR	IPER
Class 0	100 ms	50 ms	1×10^{-3}	1×10^{-4}
Class 1	400 ms	50 ms	1×10^{-3}	1×10^{-4}

NOTE 1: A set of Provisional QoS Classes; class 6 and Class 7 are defined in ITU-T Recommendation Y.1541 [2]. The distinction between these classes and those in table 10, is that the values of all objectives are provisional and they need not be met by networks until they are revised (up or down) based on real operational experience. These classes are intended to support the performance requirements of high bit rate user applications that have more stringent loss/error requirements than those described in table 10. Annex D presents the definitions and network performance objectives of these classes.

NOTE 2: The network performance and bearer service classes described in this clause are intended to enable the support of a wide variety of services and applications over the Internet protocol. There may be some service or applications that are highly sensitive to IP packet delay, jitter, loss or errors and which require more stringent performance than is described here.

TS 122 105 [71] outlines the Bearer service QoS requirements that shall be provided by 3GPP networks to the end user/applications. Table 11 presents the end-user performance expectations for conversational/real-time services.

Table 11: 3GPP specified end-user performance expectations - conversational/real-time services

Medium	Application	Degree of symmetry	Data rate	Key performance parameters and target values		
				End-to-end One-way Delay	Delay Variation within a call	Information loss
Audio	Conversational voice	Two-way	4 to 25 kb/s	< 150 ms preferred < 400 ms limit (see note 1)	< 1 ms	< 3 % FER
Video	Videophone	Two-way	32 to 384 kb/s	< 150 ms preferred < 400 ms limit Lip-synch: < 100 ms		< 1 % FER
Data	Telemetry - two-way control	Two-way	< 28,8 kb/s	< 250 ms	N.A	Zero
Data	realtime games	Two-way	< 60 kb/s (see note 2)	< 75 ms preferred	N.A	< 3 % FER preferred, < 5 % FER limit (see note 2)
Data	Telnet	Two-way (asymmetric)	< 1 KB	< 250 m	N.A	Zero
NOTE 1: The overall one way delay in the mobile network (from UE to PLMN border) is approximately 100 ms. NOTE 2: These values are considered the most demanding ones with respect to delay requirements (e.g. supporting First Person Shooter games). Other types of games may require higher or lower data rates and more or less information loss but can tolerate longer end-to-end delay.						

Both TS 181 018 [70] (TISPAN NGN) and TS 123 107 [42] (3G networks) specify implementation of QoS mechanisms.

6.3 Terminal characteristics

6.3.1 Narrowband speech

The characteristics specified in ES 202 737 [63] and/or ES 202 738 [64] shall be met.

NOTE 1: These ETSI Standards specify the characteristics of voice only terminals. When working in an audiovisual mode, packet intervals might be increased.

Error concealment techniques (Packet Loss Concealment - PLC) should be implemented on all terminals and media gateways.

NOTE 2: Packet Loss Concealment algorithms are available for most speech coding algorithms, either as a part of the standard or as an appendix to the standard. TS 126 091 [i.2] defines a mandatory error concealment procedure which shall be used by the AMR speech codec.

6.3.2 Wideband speech

The characteristics specified in ES 202 739 [65] and/or ES 202 740 [66] shall be met.

NOTE 1: These ETSI Standards specify the characteristics of voice only terminals. When working in an audiovisual mode, packet intervals might be increased.

Error concealment techniques (Packet Loss Concealment - PLC) should be implemented on all terminals and media gateways.

NOTE 2: Packet Loss Concealment algorithms are available for most speech coding algorithms, either as a part of the standard or as an appendix to the standard. TS 126 191 [i.3] defines a mandatory error concealment procedure which shall be used by the AMR-WB speech codec.

6.3.3 Video

The human eye spatial resolution is approximately 0,3 arc minute. The standard definition of normal visual acuity (20-20 vision) (see note) is the ability to resolve a spatial pattern separated by a visual angle of one minute of arc. The video display minimum resolution shall be adapted to the display size fulfilling the minimum resolution requirement of normal vision acuity.

NOTE: The term 20-20 vision refers to the distance in feet that objects separated by an angle of 1 arc minute can be distinguished as separate objects.

All video decoder implementations should include basic error concealment techniques. These techniques may include replacing erroneous parts of the decoded video frame with interpolated picture material from previous decoded frames or from spatially different locations of the erroneous frame. The decoder should aim to prevent the display of substantially corrupted parts of the picture. In any case, it is recommended that the terminal should tolerate *every* possible bitstream without catastrophic behaviour (such as the need for a user-initiated reset of the terminal).

7 Delay insensitive audiovisual applications requirements

7.1 Coding algorithms

7.1.1 Audio

TS 126 234 [72] specifies the protocols and codecs for the PSS within the 3GPP system. TS 126 346 [73] specifies the set of media decoders that are supported by the MBMS Client. The decoders specified below are relevant for both MBMS Download and Streaming delivery.

The audio codecs that should be supported by the 3GPP system are:

- HE AAC v2 [18] (see note).
- AMR-WB+ [19].

NOTE: In 3CPP documentation this codec is called enhanced aacPlus.

The ISO/IEC 14496 [18] and TS 126 290 [19] contain a table describing which codecs performs best for bit rates of 18 kbit/s to 48 kbit/s (stereo) for the following types of content:

- music;
- speech over music;
- speech between music;
- speech.

If speech is supported, the AMR [10] decoder shall be supported for narrow-band speech. The wideband AMR [14] decoder shall be supported when wideband speech is supported.

TS 102 005 [55] specifies use of the following audio codecs for the use in DVB IP-based systems:

- High Efficiency AAC Profile Level 2 (mono or 2-channel-stereo) or the High Efficiency AAC Profile Level 4 (multi-channel);
- AMR-WB+ [19];
- AC3/Enhanched AC-3 [81].

Conformance to the present document requires use of the HE AAC v2 codec.

7.1.2 Video

TS 126 234 [72] specifies the protocols and codecs for the PSS within the 3GPP system. The audio codecs that should be supported by the 3GPP system are:

- H.263 [24] version 2 Interactive and Streaming Wireless Profile (Profile 3) Level 45;
- ISO/IEC 14496-2 [18] (MPEG-4 Visual) Simple Profile at Level 3;
- H.264 (AVC) [25] Baseline Profile at Level 1b without requirements on output timing conformance.

TS 126 346 [73] specifies use H.264 (AVC) [25] Baseline Level 1.2 decoder for the MBMS service.

NOTE: MBMS does not offer dynamic negotiation of media codecs.

TS 102 005 [55] defines five ITU-T Recommendation H.264 (AVC) [25] video coding capabilities and five VC-1 [26] video coding capabilities that can be used The H.264 (AVC) [25] video coding capabilities are listed in table 12 and the VC-1 [26] capabilities are listed in table 13.

Table 12: H.264 video encoding capabilities defined for DVB services delivered directly over IP

Capability	Profile and Level
A	H.264 [25] Baseline Profile at Level 1b
B	H.264 [25] Baseline Profile at Level 12
C	H.264 [25] Baseline Profile at Level 2
D	H.264 [25] Main Profile at Level 3
E	H.264 [25] Main Profile at Level 4

Table 13: VC-1 encoding capabilities defined for DVB services delivered directly over IP

Capability	Profile and Level
A	VC-1 [26] Simple Profile at Level LL
B	VC-1 [26] Simple Profile at Level ML
C	VC-1 [26] Advanced Profile at Level L0
D	VC-1 [26] Advanced Profile at Level L1
E	VC-1 [26] Advanced Profile at Level L3

ITU-T Recommendation J.247 [68] provides video quality estimations for video classes TV3 to MM5B, as defined in ITU-T Recommendation P.911 [69], annex B. The maximum resolution is VGA and the maximum bit rate covered well in the test was 4 Mb/s. The model is a full reference model; i.e. a copy of the source is required to estimate the quality.

Within the limitations indicated above, the video performance can be estimated using one of the algorithms specified in ITU-T Recommendation J.247 [68].

7.2 Network performance requirements

ITU-T Recommendation Y.1541 [2] defines classes of network Quality of Service (QoS) with objectives for Internet Protocol network performance parameters measured between user network interfaces (UNI). Two classes are defined for interactive applications, and a third class for video streaming and similar applications. Class 2 applies for constrained routing and distance highly interactive (signalling) applications while class 3 apply for less constrained routings and distances interactive applications. Class 4 applies for low loss only applications (e.g. video streaming). The values are presented in table 14.

Table 14: Network performance objectives defined in ITU-T Recommendation Y.1541 for delay insensitive audiovisual applications

QoS class	IPTD	IPDV	IPLR	IPER
Class 2	100 ms	Unspecified	1×10^{-3}	1×10^{-4}
Class 3	400 ms	Unspecified	1×10^{-3}	1×10^{-4}
Class4	1 000 ms	Unspecified	1×10^{-3}	1×10^{-4}

NOTE 1: A set of Provisional QoS Classes; class 6 and Class 7 are defined in ITU-T Recommendation Y.1541 [2]. The distinction between these classes and those in table 14 is that the values of all objectives are provisional and they need not be met by networks until they are revised (up or down) based on real operational experience. These classes are intended to support the performance requirements of high bit rate user applications that have more stringent loss/error requirements than those described in table 14. Informative Annex D presents the definitions and network performance objectives of these classes.

NOTE 2: The network performance and bearer service classes described in this clause are intended to enable the support of a wide variety of services and applications over the Internet protocol. There may be some service or applications that are highly sensitive to IP packet delay, jitter, loss or errors and which require more stringent performance than is described here.

TS 122 105 [71] outlines the Bearer service QoS requirements that shall be provided by 3GPP networks to the end user/applications. Table 15 presents the end-user performance expectations for streaming services.

Table 15: 3GPP specified end-user performance expectations - streaming services

Medium	Application	Degree of symmetry	Data rate	Key performance parameters and target values		
				Start-up Delay	Transport delay Variation	Packet loss at session layer
Audio	Speech, mixed speech and music, medium and high quality music	Primarily one-way	5 kb/s to 128 kb/s	< 10 s	< 2 s	< 1 % Packet loss ratio
Video	Movie clips, surveillance, real-time video	Primarily one-way	20 kb/s to 384 kb/s	< 10 s	< 2 s	< 2 % Packet loss ratio
Data	Bulk data transfer/retrieval, playout and synchronization information	Primarily one-way	< 384 kb/s	< 10 s	N.A.	Zero
Data	Still image	Primarily one-way		< 10 s	N.A.	Zero

7.3 Terminal characteristics

7.3.1 Audio

There is no audio terminal characteristics specified.

All audio decoder implementations should include basic error concealment techniques.

7.3.2 Video

The human eye spatial resolution is approximately 0,3 arc minute. The standard definition of normal visual acuity (20-20 vision) (see note) is the ability to resolve a spatial pattern separated by a visual angle of one minute of arc. The video display minimum resolution shall be adapted to the display size fulfilling the minimum resolution requirement of normal vision acuity.

NOTE: The term 20-20 vision refers to the distance in feet that objects separated by an angle of 1 arc minute can be distinguished as separate objects.

All video decoder implementations should include basic error concealment techniques. These techniques may include replacing erroneous parts of the decoded video frame with interpolated picture material from previous decoded frames or from spatially different locations of the erroneous frame. The decoder should aim to prevent the display of substantially corrupted parts of the picture. In any case, it is recommended that the terminal should tolerate *every* possible bitstream without catastrophic behaviour (such as the need for a user-initiated reset of the terminal).

Annex A (informative): Audio-video quality interaction

A.1 Introduction

The intention of this annex is to present some background information and to present possible parameter values that can be used in the formulas given in clause 4.6 of the present document.

A.2 Available information

A.2.1 TR 102 479

TR 102 479 [i.4] provides and overview of factors the influence user perceived in systems supporting multimedia applications. One of the topics addressed is Media quality interaction. The following results are considered:

a) Results published by Beerends and de Caluwe. These results are also presented to ITU-T Recommendation SG 12 [i.21] in COM 12-19 (1997-2000 Study period). A nine point ACR scale has been used. The model for predicting audiovisual quality from audio and video quality rating developed by Beerends and de Caluwe is:

$$MOS_{AV} = 0{,}007 MOS_A + 0{,}24 MOS_V + 0{,}088 (MOS_A * MOS_V) + 1{,}12$$

An alternative model using a single multiplicative term is:

$$MOS_{AV} = 0{,}11 (MOS_A * MOS_V) + 1{,}45$$

The alternative model gave a slightly lower correlation between predicted and measured ratings, 0,98 for the first model and 0,97 for the second model. This model is visualized in figure A.1.

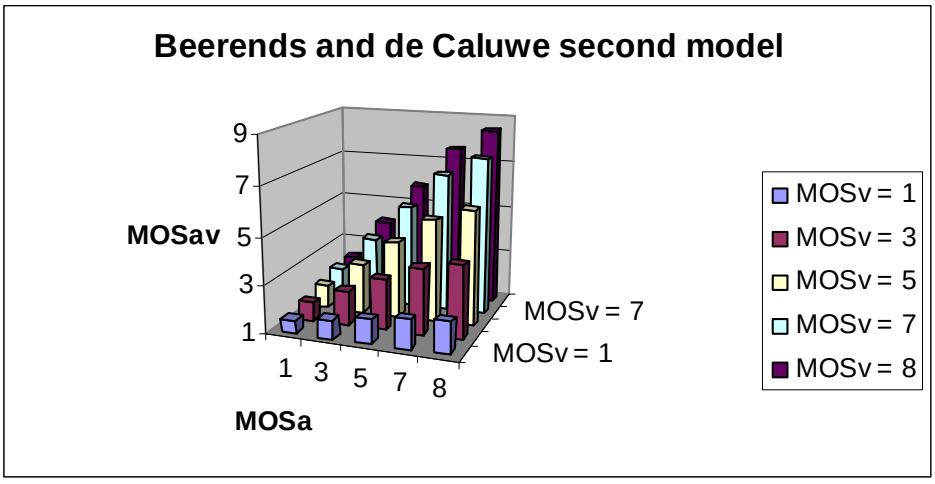


Figure A.1: AV interaction model developed by Beerends and de Caluwe

b) Results published by Hands. Two experiments were carried out.
The first experiment compared the test sequence with a reference (un-degraded) sequence; i.e. a double stimulus continuous quality scale (DSCQS) using a five point scale description. The test material consists of short (5 s) head and shoulder sequences. The model developed from these results is.

$$MOS_{AV} = 0{,}85 MOS_A + 0{,}76 MOS_V - 0{,}01 (MOS_A * MOS_V) - 3{,}34$$

The second experiment consists of two sets of test material; head and shoulder sequences and high motion sequences. The single stimulus quality scale (SSQS) methodology (5 points) was used. Two models were developed.

Head and shoulder model:

$$MOS_{AV} = 0{,}17\,(MOS_A * MOS_V) + 1{,}15$$

High motion model:

$$MOS_{AV} = 0{,}25\,MOS_V + 0{,}15\,(MOS_A * MOS_V) + 0{,}95$$

The head and shoulder model is visualized in figure A.2, and the high motion is visualized in figure A.3.

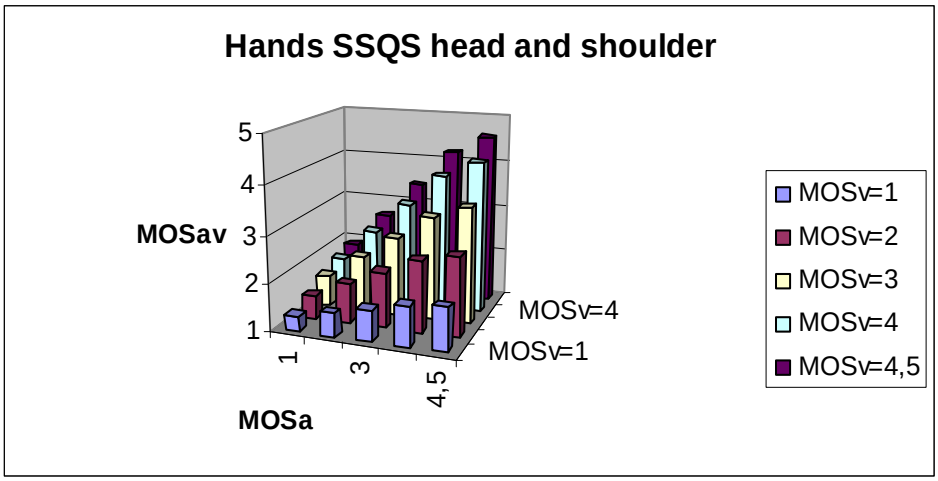


Figure A.2: AV head and shoulder interaction model developed by Hands

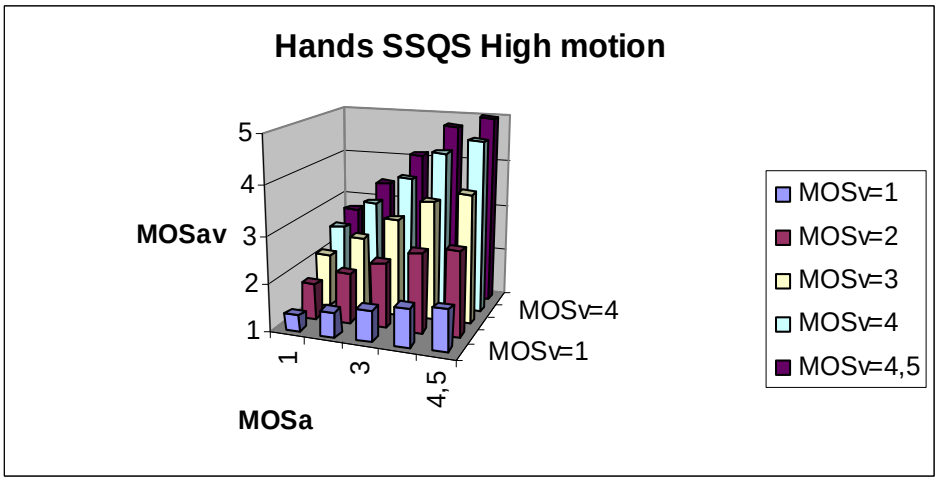


Figure A.3: AV high motion interaction model developed by Hands

c) Results published by Winkler and Faller. These tests reflect mobile applications use. The video was downsampled QCIF. A five point ACR scale was used. The test scenes used included head and shoulder scenes, high motion scenes, movie trailers and football. A multiplicative term model based on these results is:

$$MOS_{AV} = 0{,}103\,(MOS_A * MOS_V) + 1{,}98$$

For these results a linear mode gave a slightly better fit, but to easily compare the Winkler and Faller results with the other results reviewed in this contribution, the multiplicative term model is used. This model is visualized in figure A.4.

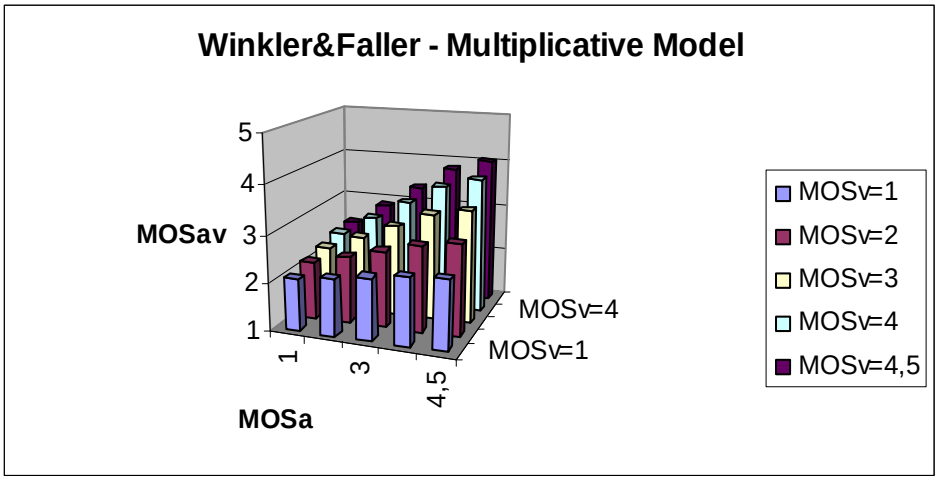


Figure A.4: AV multiplicative model developed by Winkler and Faller

d) Results published by Kitawaki et al. A slightly different approach where the *Mutual interactivity quality* is introduced. The model takes into account the influence of audio when accessing video and vice versa. The ACR 5 point scale is used. The model is:

$$MOS_{AV} = 0{,}188\,V_q(A_q) + 0{,}211\,A_q(V_q) + 0{,}112\,V_q(A_q) * A_q(V_q) + 0{,}618$$

To compare this model with the models described above, the transform described in TR 102 479 [i.4] has been made. The model is visualized in figure A.5.

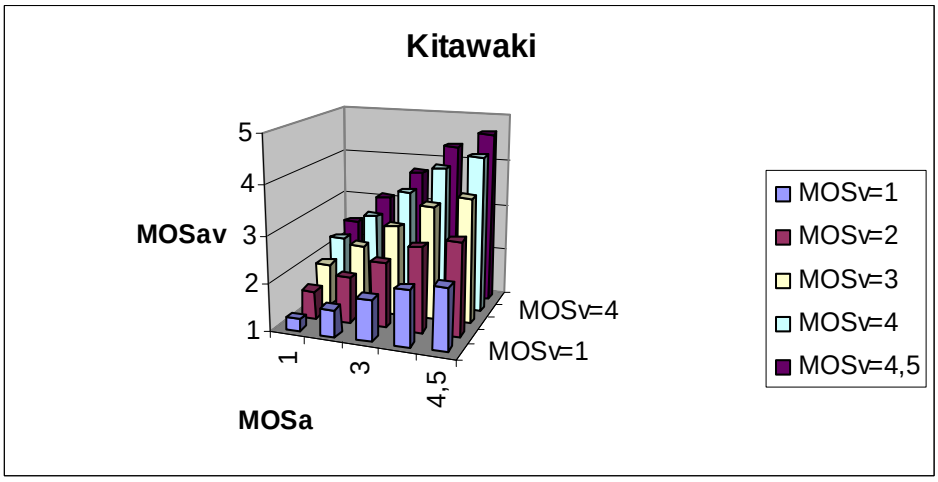


Figure A.5: AV multiplicative model developed by Kitawaki

A.2.2 Results presented to ITU-T Recommendation SG 12

A France Telecom contribution to ITU-T Recommendation SG.12 [i.21] (COM 12-61, 1997-2000 Study period) presents results of two psychophysical experiments investigating the relationships between audio, video and audiovisual qualities in passive and conversational contexts. A passive context application example is distribution applications such as IPTV, while a conversational context application example is videoconferencing.

Four audio and four video objective qualities were used. The audio characteristics were:

- G.728 codec at 16 kbit/s;
- G.711 codec at 56 kbit/s;
- G.722 codec at 56 kbit/s;

- no coding (Full Band - FB).

The video characteristics were:

- QCIF at 456 kbit/s, 25 images/s;
- CIF at 72 kbit/s (CIF1), 12 images/s;
- CIF at 456 kbit/s (CIF5), 12 images/s;
- no coding (PAL), 25 images/s.

The combination of these audio and video configurations gave 16 objective audiovisual qualities.

The passive context experiments were using videoconference samples involving a male and a female speaker, i.e. similar to head and shoulder sequences used in some of the experiments described above.

One of the observations that was highlighted in the contribution is that both audio and video contributes to the audiovisual quality. The influence of video is however found to be larger than the influence of audio.

The best audiovisual quality prediction model in passive context was based on a logarithmic transformation of MOS_{AV}. However, the be able to easily compare the results with the models presented to ITU-T Recommendation SG.12 [i.21] by KPN (Beerends and de Caluwe, see above), Bellcore and ITS, a multiplicative term model was used. The five point ACR scale results of France Telecom were also transformed to a nine point scale that was used by the experiments reported in the other contributions.

The model found by France Telecom is:

$$MOS_{AV} = 0{,}10\,(MOS_A * MOS_V) + 1{,}76$$

The model is visualized in figure A.6.

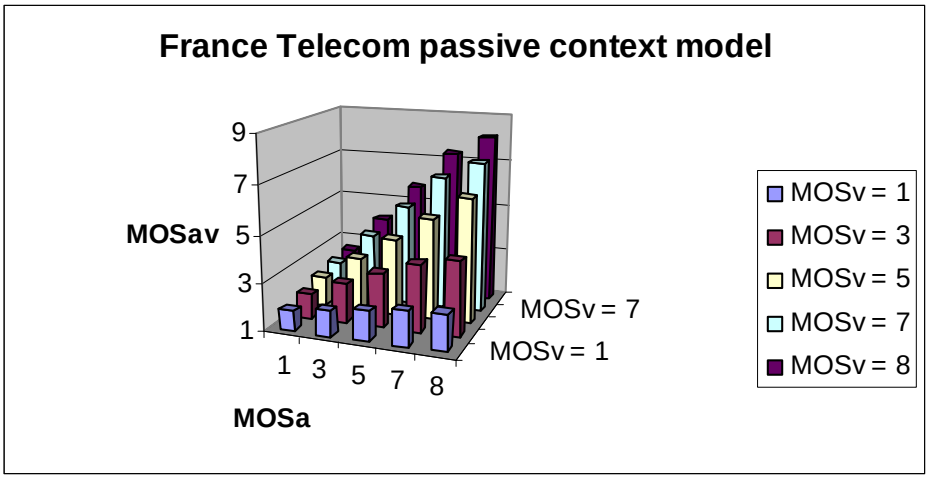


Figure A.6: AV passive context model developed by France Telecom

The model found by Bellcore is:

$$MOS_{AV} = 0{,}11\,(MOS_A * MOS_V) + 1{,}07$$

The model is visualized in figure A.7.

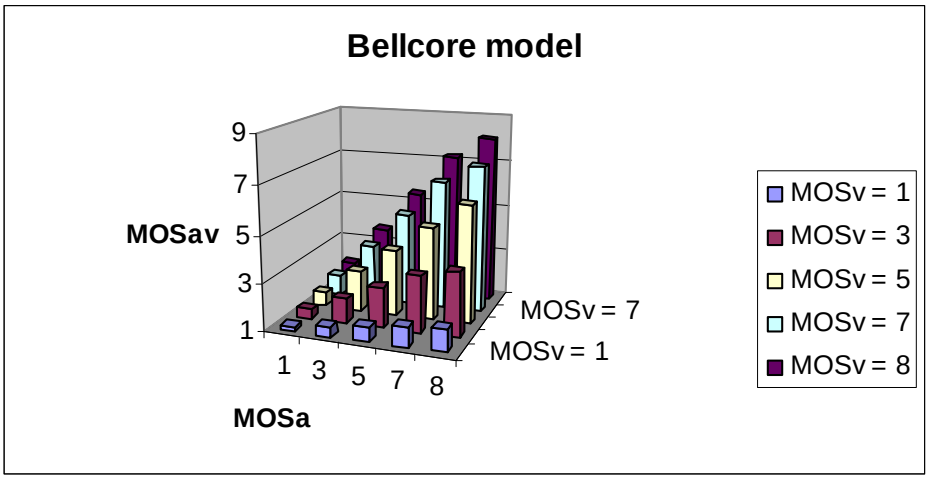


Figure A.7: AV model developed by Bellcore

The model found by ITS is:

$$MOS_{AV} = 0{,}121(MOS_A * MOS_V) + 1{,}54$$

The model is visualized in figure A.8.

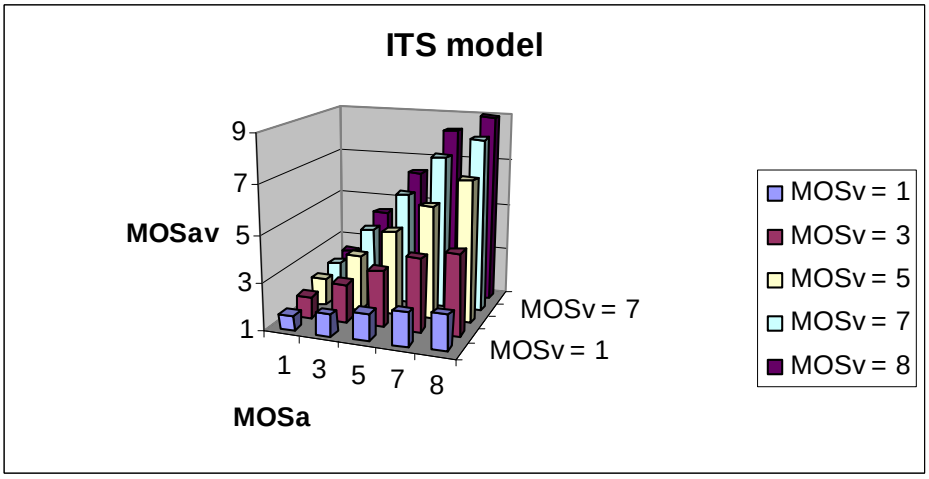


Figure A.8: AV model developed by ITS

The conversational context part of the France Telecom experiments indicated clearly that audio did not influence the audiovisual quality. A model to predict the conversational audio quality is:

$$MOS_{AV} = 0{,}657\,MOS_V + 0{,}573$$

It is also found that delay influences the perceived audiovisual quality in a conversational context. A prediction model where delay is included is:

$$MOS_{AV} = 0{,}569\,MOS_V - 0{,}586\,GD + 1{,}125$$

The parameter *GD* is the Global Delay expressed in seconds.

Annex B (informative): QoS mechanisms and the effects on connection performance

B.1 Introduction

IP network without a QoS mechanism is a **Best effort** network. In a best effort network all users obtain unspecified variable bit rate and delivery time, depending on the current traffic load.

Transmission of real-time information over a best effort network may cause performance degradation. Although high transmission capacity may limit the performance degradation, there is a need for QoS mechanisms to offer an acceptable QoS to audiovisual real-time applications.

This annex presents an overview of some QoS mechanisms available. In addition to these mechanisms, some of the transmission technologies available include Forward Error Correction (FEC) in order to keep the transmission error rate as low as possible.

B.2 QoS mechanisms

Two different internet QoS models have been proposed by the Internet Engineering Task Force (IETF), namely, integrated services (IntServ) and differentiated services (DiffServ). There are also QoS mechanisms included in transmission technology standards; e.g. WLAN, WiMax and HS(D)PA.

B.2.1 Integrated services (IntServ)

The integrated services architecture is defined in RFC 1633 [i.5]. The document defines an extended service model, called the Integrated Services (IS) model. The model includes two sorts of services targeted towards real-time traffic; guaranteed and predictive service. The model relies on the Resource Reservation Protocol (RSVP) [i.6] to signal and reserve the desired QoS for each flow in the network. The guaranteed service provides for firm bounds on end-to-end delay and assured bandwidth for traffic that conforms to the reserved specifications. The predictive service is a controlled load service that provides for a better than best effort and low delay service under light to moderate network loads. It is possible (at least theoretically) to provide the requisite QoS for every flow in the network, provided it is signalled using RSVP and the resources are available.

IntServ is not frequently used because there are scalability challenges and RSVP support need to be implemented in all network elements from the source to the destination.

B.2.2 Differentiated services (DiffServ)

The differentiated services architecture is defined in RFC 2475 [i.7]. In DiffServ, instead of explicit reservation, traffic is differentiated into a set of classes for scalability, and network nodes provide priority-based treatment according to these classes. Within the core of the network, packets are forwarded according to the per-hop behaviour associated with defined codepoints.

Differentiated services enhancements to IP protocols are intended to enable scalable service discrimination without the need for per-flow state and signalling at every hop. The services may be either end-to-end or intra-domain; they include both those that can satisfy quantitative performance requirements (e.g. peak bandwidth) and those based on relative performance (e.g. "class" differentiation). Services can be constructed by a combination of:

- setting bits in an IP header field at network boundaries (autonomous system boundaries, internal administrative boundaries, or hosts);
- using those bits to determine how packets are forwarded by the nodes inside the network;

- conditioning the marked packets at network boundaries in accordance with the requirements or rules of each service.

DiffServ networks classify packets into one of a small number of aggregated flows or "classes", based on the DiffServ codepoint (DSCP) in the packet's IP header. This is known as behaviour aggregate (BA) classification. At each DiffServ router, packets are subjected to a "Per-Hop Behaviour" (PHB), which is invoked by the DSCP. Using packet markings (code points) and queuing policies it results in some traffic to be better treated or given priority over other (use more bandwidth, experience less loss, etc.). It is important to note that providing QoS on a per-hop basis cannot guarantee an end-to-end QoS.

To simplify router design the number of PHB groups should be low. Currently the two most frequently PHBs are:

- **Assured Forwarding (AF)**
  This PHB provides *independently* forwarded traffic classes. Each class is assigned some partition of bandwidth and buffer capacity.

- **Expedited Forwarding (EF)**
  This PHB can be used to build a low loss, low latency, and low jitter assured bandwidth, end-to-end service.

B.2.3 WLAN QoS mechanism

IEEE 802.11 [46] specifes mechanisms that provides QoS on the 802.11 MAC layer (see note). The basis of 802.11 QoS mechanisms is the *Hybrid Coordination Function* (HCF), which is the enabler for QoS support. HCF has two medium access mechanisms; *Enhanced Distributed Channel Access* (EDCA) and *HCF Controlled Channel Access* (HCCA). EDCA is used for contention access, while HCCA is for contention-free access controlled by polling from the access point. These relationships are shown in figure B.1.

NOTE: QoS support was initially published in IEEE 802.11e. It is now included in the 2007 version of IEEE 802.11 [46].

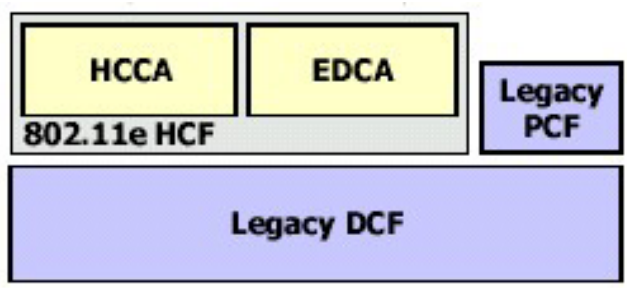


Figure B.1: Relation between HCCA and EDCA of 802.11 HCF and relation to PCF and DCF of legacy 802.11

To understand the principal differences between EDCA and HCCA, one may draw the analogy to IP level DiffServ and IntServ. EDCA, like DiffServ, divides traffic into traffic classes and provides prioritized, qualitative, and relative differentiation between the traffic classes without providing any hard guarantees. On the other hand, HCCA (like IntServ) can handle traffic on a per-application level and provides parameterized, quantitative and "guaranteed" differentiation. Although the "guarantees" of HCCA are harder than those of EDCA, it is always difficult to talk about guarantees when dealing with the inherently unreliable 802.11 medium.

Four QoS classes, also known as Access Categories (ACs) are defined. These are:

- voice (Highest priority);
- video;
- best effort;
- background (Lowest priority).

B.2.4 WiMax QoS mechanism

The IEEE 802.16 [45] has three main methods for QoS provisioning:

- service Flow Classification;
- dynamic Service Establishment;
- two-Phase Activation Model.

The main feature of the WiMax QoS mechanism is that each packet is associated with a service flow.

The WiMax MAC contains a queuing and traffic-shaping engine that is ultimately responsible for the reception and transmission of 802.16a packets according to the enforced QoS parameters. These parameters can be different from service flow to service flow. An overview of QoS support in 802.16 networks can be found in a paper by Cicconetti et. al. [i.8].

The WiMax service classes are listed in table B.1.

Table B.1: WiMax QoS service classes

WiMax QoS Service Class	Description
Unsolicited Grant Service (UGS)	Supports CBR services such as T1/E1 emulation and VoIP without silence suppression.
Real-Time Polling Service (rtPS)	Supports real-time services with variable size data packets that is issued at periodic interval (video and VoIP with silence suppression).
Non-Real-Time Polling Service (nrtPS)	Supports delay tolerant services consisting of variable size data packets for which a minimum data rate is required.
Best Effort (BE)	For applications that do not require QoS.

Each network application has to register with the network, where it will be assigned one of these service flow classifications with a Service Flow ID (SFID). When the application wants to send data packets, the service flow is mapped to a connection using a unique CID.

IEEE 802.16 [45] provides a dynamic service establishment function that allows an application to acquire more resources when required. Multiple service flows can be allocated to the same application, so more service flows can be added if needed to provide good QoS.

The four service classes, described in table B.1, are implemented by using the QoS mechanisms built into the grant-based MAC. table B.2 describes how each service class is implemented.

Table B.2: WiMax QoS Service Class implementation

WiMax QoS Service Class	Implementation
Unsolicited Grant Service (UGS)	Base Station (BS) provides fixed-size data grant bursts periodically.
Real-Time Polling Service (rtPS)	Base Station (BS) provides the Subscriber Station (SS) the opportunity to request bandwidth on a regular basis.
Non-Real-Time Polling Service (nrtPS)	Base Station (BS) provides the Subscriber Station (SS) the opportunity to request bandwidth using unicast and contention methods, etc.
Best Effort (BE)	Base Station (BS) allows the Subscriber Station (SS) to use all available mechanisms for transmission requests.

B.2.5 HSPA packet scheduling

An overview of QoS options for HSDPA is presented in a paper by Pedersen et. al [i.9]. There are basically three mechanisms that can be used to implement QoS control for HS(D)PA:

- dynamic allocation of transmission resources for HS(D)PA;
- quality-based HS(D)PA access control and congestion control algorithms;
- qoS-aware MAC-hs packet scheduling.

Dynamic allocation of transmission resources for HSDPA includes the power and channelization codes that the MAC-hs is allowed to use for transmission. The latter is primarily relevant for cells where the available power and channelization codes for each cell are shared between Release 99 dedicated channels and HSDPA. Quality-based HSDPA access control algorithms decide whether a new HSDPA user should be granted or denied access, depending on the user's QoS requirements and the current load in the cell. Finally, the QoS-aware MAC-hs packet scheduler is responsible for scheduling the allocated users on the shared data channel so their QoS requirements are fulfilled.

Scheduling downstream is performed by the logical node responsible for radio transmission/reception in one or more cells to/from the User Equipment (Node B) It determines which device to send data to in the next 2 ms time frame thus making the most efficient use of the bandwidth available.

There are some differences between the uplink and the downlink. While the downlink channel is shared between users, the uplink channel is dedicated to a single user. It is possible to initiate a scheduling mechanism similar to downlink scheduling by the device. However, for delay sensitive applications such as VoIP, there is another method where the device initiates the transmission. With Scheduled request-grant activity, the node B determines the power level of the device transmission and is controlled dynamically to ensure maximum efficiency for all devices on that cell.

The QoS requirements depend on the application. Hence, the algorithms and parameters of the mechanisms indicated above need to be optimized to the distribution of real-time and non real-time applications. Investigations have indicated that different downlink scheduling algorithms for different media (i.e. voice, video and/or data) might be beneficial. It is up to the equipment manufacturers and the network operators to choose algorithms and parameters that give best possible performance for the actual traffic conditions.

B.2.6 RACS

RACS is the ETSI TISPAN NGN Sub-System responsible for the implementation of policy-based transport control features, by using procedures and mechanisms that handle resource reservation and admission control for both unicast and multicast traffic in access and core networks.

Besides acting as a Resource Control Framework, RACS also includes support for controlling Network Address Translation (NAT) at the edge of networks and for assisting in remote NAT traversal.

RACS also covers aspects related to the derivation, modification, and installation of traffic policies, end to end quality of service, transport-level charging and overload control.

The RACS functional architecture is defined in ES 282 003 [i.10].

RACS essentially provides policy based transport control services to applications. This enables applications to request and reserve transport resources from the transport networks within the scope of RACS.

RACS scope extends to the access and core networks, as well as to points of interconnection between them in order to support e2e QoS. The work of ETSI TISPAN is organized in Releases. In Release 3 of ES 282 003 [i.10] the e2e QoS handling is limited to scenarios involving only wholesale and roaming between two domains.

By offering a set of generic policy based transport control services to applications, RACS ensures that any existing or future application shall be able to request transport resources appropriate to that service as long as the application supports the interface to RACS defined in this architecture specification.

Moreover, by offering a level of hidden interaction between applications and the transport resources themselves, RACS also ensures that applications do not need to be aware of the underlying transport networks. RACS allows for real-time multimedia services (VoIP, Videoconferencing, Video on Demand, on-line gaming) to request some particular bandwidth and/or address mediation capabilities for the service from the network. As the network element responsible for policy based transport control, RACS evaluates these requests in the context of predefined policy rules, which the network operator has provisioned. RACS reserves the appropriate resources and admits the request provided the request passes the policy tests and the appropriate resources are available in the transport network. Therefore, RACS offers the means for an operator to enforce admission control and set the respective bearer service policies.

In addition, RACS can also provide the means for value-added services to obtain network resources that are necessary to offer services to the end-user.

RACS is resource-reservation session aware but application session agnostic, i.e. it can support transport resource reservations for both session based and non-session based applications.

RACS also provides access to services provided by the Border Gateway Function. Examples of those services are gate control, NAT and hosted NAT transversal.

Basically, RACS offers to applications the following set of functionalities on a one per RACS resource reservation session request basis:

- Admission Control: RACS implements Admission Control to the access and aggregation segment of the network. One can imagine various types of admission control going from a strict admission control where any overbooking is to be prevented, to admission control that allows for a certain degree of over subscription or even a trivial admission control where the authorization step is considered sufficient to grant access to the service.

- Resource reservation: RACS implements a resource reservation mechanism that permits applications to request bearer resources in the access, aggregation, and core networks.

NOTE: Resource reservation mechanisms in the core network are not standardized in the Release 3 version of ES 282 003 [i.10].

- Policy Control: RACS uses service based local policies to determine how to support requests from applications for transport resources. Based on available information about resource availability and on other policy rules, e.g. priority of the application, RACS determines if a request can be supported and (if successful) RACS authorizes appropriate transport resources and derives L2/L3 traffic polices to be enforced by the bearer service network elements.

- NAT transversal: RACS controls the transversal of far end (remote) NAT.

- NAT/Gate control: RACS controls near-end NAT at the borders of the NGN core network and at the border between a core network and an access network.

RACS offers services to applications that may reside in different administrative domains.

The RACS architecture supports both guaranteed and relative QoS control - allowing the access provider to select the most suitable QoS architecture for their needs.

When relative QoS is used, the QoS differentiation shall be performed at the IP_Edge, e.g. compliant with the DiffServ Edge functionality defined in IETF specifications for Differentiated Services (see RFC 2475 [i.7]). Moreover, RACS takes into account the ability of some CPN to provide QoS differentiation, e.g. by applying DiffServ marking, and take steps to allow this to have effect only where it is required by operator defined RACS local policies.

For guaranteed QoS control, enforcement of QoS admission control decisions (throughput control and traffic policing) shall be performed in the IP_Edge and may also be performed in the CPN and/or Access Node.

The RACS shall support the "proxy QoS reservation request with policy-push" as a QoS Push resource reservation mechanism, e.g. among others, the one shown in figure 3. In this case, the CPN does not itself support native application independent QoS signalling procedures. When a CPN invokes a specific service of an AF using the NGN signalling (e.g. SIP), the AF will issue a request to the RACS for QoS authorization (policy control) and resource reservation. The AF may extract implicit user requested QoS class from Service Request, e.g. by SIP SDP, based on operator's policy, and send the appropriate QoS class information to RACS.

RACS policy decisions are "pushed" to the policy enforcement point (IP_Edge) in the NGN access (e.g. xDSL).

The RACS also supports the QoS Pull resource reservation mechanism, e.g. the one depicted in figure 4. This "user requested QoS with policy-pull" mechanism requires that the UE is able to handle Layer 3 QoS signalling capability, and perform a QoS request for its own needs through the use of path-coupled signalling, e.g. RSVP. Similarly the UE may request a service which in turn may cause a QoS Request to be originated from a Transport Function resulting in a Policy Pull request.

B.3 Effects on connection performance

The connection performance metrics addressed in the present document are bandwidth, delay/delay jitter, and packet loss rate. Using these performance metrics, network performance guarantees can be specified in various forms, such as *absolute* (or *deterministic*), e.g. a network connection is guaranteed with e.g. a specified bandwidth all the time; *probabilistic* (or *stochastic*), e.g. connection performance metrics is guaranteed not to exceed a specified threshold. The *guarantee* feature of network QoS is what differentiates it from the "best-effort" network services. The exact form of performance would depend on the operator(s) policies and agreements with the customers as well as the different mechanisms described.

While WiMax and WLAN QoS mechanisms provide DiffServ-like QoS, the HSPA Scheduling mechanism is different. The main objective of this mechanism is to enable best possible use of resources available. Performance parameters such as delay and congestion/packet loss might be included in the scheduling algorithms.

By implementing QoS mechanisms such as DiffServ EF, reduced delay and delay variations (jitter) has been reported. The network related packet loss might also be lower. More accurate information is however required to quantify the user perceived performance improvement. Until this information is available, it is difficult to estimate the user perceived performance improvement.

Annex C (informative): Packet loss recovery

C.1 Introduction

Real-time media (e.g. speech, audio or video) transport over IP networks are using unreliable protocols which does not guarantee that packets are delivered or delivered in order. Packets may be dropped under unfavourable transport conditions, peak loads or periods of congestion (caused, for example, by link failures or inadequate capacity). Packets may also be dropped at the receiving device due to jitter buffer overflow.

Several technologies may be applied to eliminate or reduce the degradations caused by packet loss. This annex presents an overview of some of these technologies.

Some transmission systems (e.g. DVB-H/SH, WiMAX) apply FEC at the physical layer and/or link layer. For DSL access interleaving is a technique being used.

These transmission system mechanisms can be supplied by error recovery mechanisms at the application layer. The application layer error recovery mechanisms are the topic addressed in this annex.

C.2 Application layer packet loss recovery methods

Application layer methods to repair packet loss can be divided into sender based and receiver based methods. Sender-based methods, which introduce redundancy in the transmitted bit stream, are generally more powerful but require implementations in both encoder and decoder, whereas receiver-based methods require changes in the decoder only.

It is important to note that sender based methods and receiver based methods are complementary techniques that can (and should) be combined to obtain best possible performance.

C.2.1 Speech and audio recovery

A paper by Perkins et. al. [i.11] presents a survey of packet loss recovery techniques for streaming audio. The paper address both sender based methods and receiver based methods as described in the introduction.

Sender based methods can be split into two major classes; active retransmission and passive coding. A taxonomy of sender based audio related methods provided by Perkins et. al. [i.11] is depicted in figure C.1. Another classification approach is chosen by Wah et. al. [i.12] who divides the error concealment schemes into source coder-independent schemes and source code-dependent schemes.

Sender based methods are usually more efficient than the receiver methods because there are a combination of both encoder and decoder implementation, and could convey information about the lost packets to the receiver.

Retransmission is a simple approach, but increases the delay, and is not realistic for real-time audiovisual applications, but is a good (and simple) option for applications where the audiovisual information is downloaded. Retransmission can be achieved using the TCP protocol which is a reliable transport protocol; see clause 4.3.3.

The principle of Forward Error Correction (FEC) is to send redundant information, along with the primary information, in order to allow a receiver to reconstruct (exactly or approximately) the missing signal. FEC allows to significant improvement of the perceived quality, but it has the main drawback of increasing the end-to-end delay, since the destination has to wait for the redundant packets to be received in order to reconstruct the missing sample. The actual delay depends on the algorithm and algorithm parameters used. Another FEC related drawback is increased bit rate requirement.

Perkins et. al. [i.11] classifies FEC into media independent FEC and media specific FEC. Media specific FEC may be flexible where the FEC specific information is transmitted only when parameters of the codec are deemed to have changed sufficiently. Advantages might be reduced overhead and/or reduced delay. These approaches are usually codec dependent.

In addition to link layer error control interleaving can be used on the application layer. Unit (smaller than the packet) are re-sequenced so that originally adjacent units are separated by a defined distance (interleaving depth). Like FEC interleaving increases the end-to-end delay.

MPEG 4, Part 3 [18] defines a set of error resilience tools that are source coding related and thus codec specific. Designated algorithms are defined for optional functions such as Spectral Band Replication (SBR) and parametric Stereo. The core decoder employs signal-adaptive spectrally shaped noise generation for error concealment, in the SBR and Parametric Stereo decoders; error concealment is based on extrapolation of guidance, envelope, and stereo information.

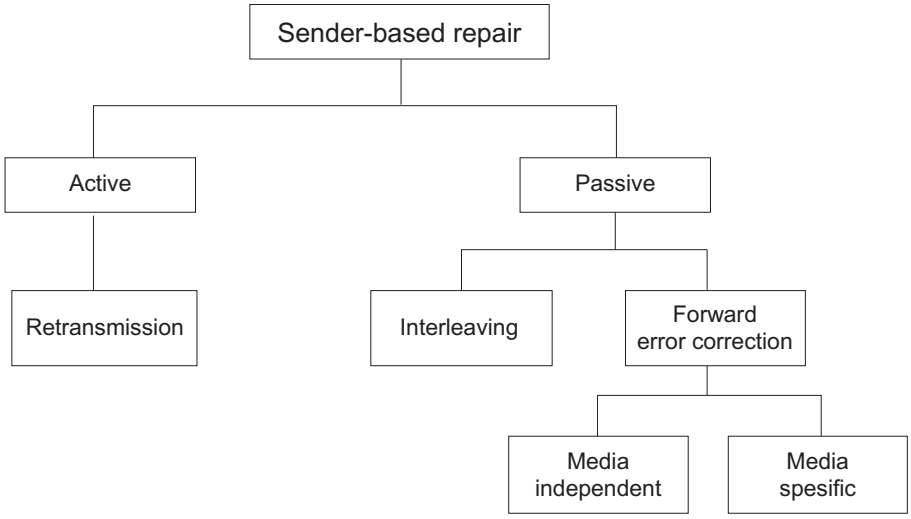


Figure C.1: A taxonomy of sender based audio packet loss repair methods

A taxonomy of audio related receiver based methods provided by Perkins et. al. [i.11] is depicted in figure C.2. These methods create a packet that substitutes the lost packet. The advantage is no requirement to sender implementation; the disadvantage is less powerful methods that work best for loss of single, small packets.

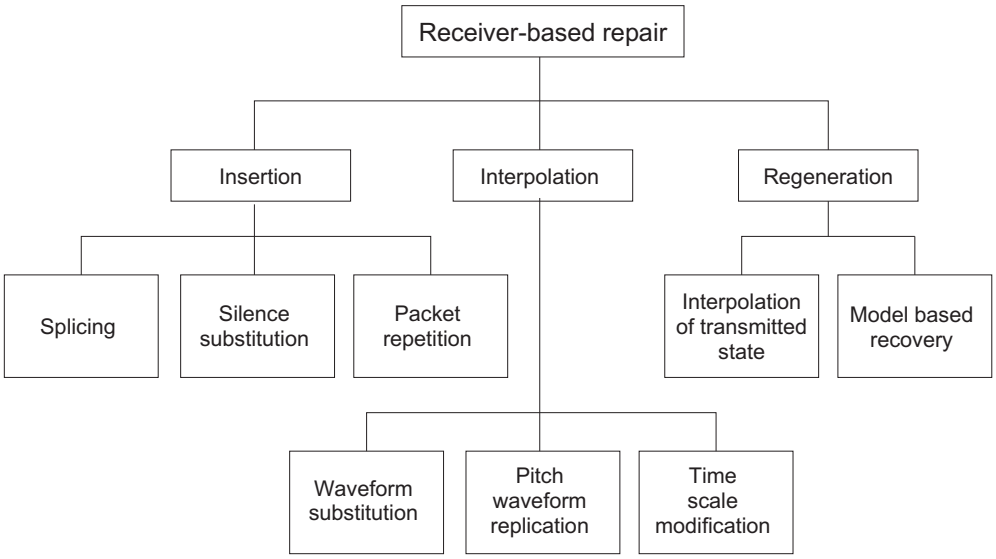


Figure C.2: A taxonomy of receiver based audio packet loss repair methods

Insertion based schemes repair the lost packet(s) by inserting packet(s) into the playout bit stream. The characteristics of the inserted packet(s) might be simple; silence, noise or repetition of the previous packet(s). These solutions may work when single short packets are lost, but the performance is not good for long packets or burst packet losses. In the case of burst packet losses, the performance is improved by gradually reducing the amplitude of the repeated packets.

This approach is chosen by 3GPP [i.2] and [i.3]. In order to improve the subjective quality, lost speech frames shall be substituted with either a repetition or an extrapolation of the previous good speech frame(s). This substitution is done so that it gradually will decrease the output level, resulting in silence at the output. Use of this procedure is mandatory for implementation in all networks and User Equipment (UE) capable of supporting the AMR [10] or AMR-WB [14] speech codecs. The narrowband specification [i.2] provides two example solutions; the wideband specification [i.3] provides a single example solution.

Another alternative is splicing where the audio of either side of the loss is spliced together. A disadvantage is timing disruption. Perkins et. al. [i.11] also highlights risk for interference with adaptive playout buffers, and states that splicing is not an acceptable repair technique.

A more advanced approach is interpolation where attempt is made to interpolate a replacement of the lost packet(s) from the packets surrounding a loss. The advantage of interpolation-based schemes over insertion-based techniques is that they account for the changing characteristics of the speech/audio signal transmitted. Perkins et. al. [i.11] identifies three different approaches:

- Waveform substitution.
  Waveform substitution uses audio before, and optionally after, the loss to find a suitable signal to cover the loss. Implementations that use information both before and after the loss performs better, but introduces increased delay.

- Pitch waveform replication.
  This is a refinement of waveform substitution by using a pitch detection algorithm either side of the loss. Losses during unvoiced speech segments are repaired using packet repetition and voiced losses repeat a waveform of appropriate pitch length. This approach is marginally better than waveform substitution.

- Time scale modification.
  Time scale modification allows the audio on either side of the loss to be stretched across the loss. It is more resource demanding, but performs better than the previous approaches.

The method standardized in ITU-T Recommendation G.711 Appendix I [i.23] is using pitch waveform replication. The signal is delayed 3,75 ms before it is sent to the output. This algorithm delay, used for an Overlap Add (OLA) at the start of an erasure, allows the PLC code to make a smooth transition between the real and synthesized signal.

The method standardized in ITU-T Recommendation G.722 Appendix III [i.14] uses periodic waveform extrapolation to fill in the waveform of lost packets, Additional processing takes place for each packet loss in order to provide a smooth transition from the extrapolated waveform to the waveform decoded from the received packets.

Regenerative repair techniques use knowledge of the audio compression algorithm to derive codec parameters, such that audio in a lost packet can be synthesized. These techniques are necessarily codec-dependent but perform well because of the large amount of state information used in the repair. Typically, they are also somewhat computationally intensive.

For codecs based on transform coding or linear prediction, it is possible that the decoder can interpolate between states. For example, the ITU-T Recommendation G.723.1 speech coder [82] interpolates the state of the linear predictor coefficients either side of short losses and uses either a periodic excitation the same as the previous frame, or gain matched random number generator, depending on whether the signal was voiced or unvoiced. For longer losses, the reproduced signal is gradually faded.

Another example is the error concealment specified in ITU-T Recommendation G.729 [8]. The concealment strategy is to reconstruct the current frame, based on previously received information. The method replaces the missing excitation signal with one of similar characteristics, while gradually decaying its energy. This is done by using a voicing classifier based on the long-term prediction gain, which is computed as part of the long-term post filter analysis.

The linear prediction technique is also used in the low complexity algorithm specified in ITU-T Recommendation G.722 Appendix IV [i.15]. In case of frame erasures, the decoder performs an analysis of the past lower-band reconstructed signal and extrapolates the missing signal using linear-predictive coding (LPC), pitch-synchronous period repetition and adaptive muting. In the higher band the decoder repeats the previous frame pitch-synchronously, with adaptive muting and high-pass post-processing.

- In model-based recovery the speech on one, or both, sides of the loss is fitted to a model that is used to generate speech to cover the period loss.

C.2.2 Video recovery

Video coding is more complex than speech and audio coding. It is therefore logical that video error recovery is more demanding than speech recovery.

The paper by Wah et. al. [i.12] uses the same classification for audio and video concealment; i.e. dividing the error concealment schemes into source coder-independent schemes and source code-dependent schemes.

The paper groups source coder-independent schemes into three classes:

- sender based schemes where intelligent packetization prevent loss of synchronization and propagation effects due to difference coding applied;
- receiver based schemes which could be carried out in either spatial domain, temporal domain or frequency domain;
- sender receiver based schemes where both sender and receiver are involved in the error concealment process. Examples are Forward Error Correction (FEC), retransmissions and interleaving.

The source coder dependent classes described are:

- robust entropy coding that decreases the error propagation when packet losses result in wrong decoding states;
- restricting prediction domain which may include different forms of picture segmentation and dynamic reference picture selection;
- layered coding where the video data is portioned into a base layer and enhancement layers;
- multiple description coding (MDC) where the video data is divided into equally important streams such that the decoding quality using any subset is acceptable.

The video coding standards specify the bitstream that shall be correctly decoded, and is offering more freedom to the implementers than the ETSI and ITU-T speech codec standards. A paper by Wenger [i.24] on H.264 over IP presents an overview of error resilience tools that are standardized in the ITU-T Recommendation H.264 (AVC) [25]. Some of these tools are defined in older video standards [16], [17], [22], [23] and [24]. These tools are based on selections made in the encoding process, and is considered as source code-dependent schemes.

C.3 Performance improvement

It is difficult to quantify the effects of generic resilience tools. No doubt there is performance improvement. Videoconferencing equipment vendors have demonstrated that for H.264-encoded video implementation of proprietary algorithms in combination with standard based tools make degradations caused by up to 10 % random packet loss almost invisible. This is a significant improvement compared with the performance target indicated in ITU-T Recommendation G.1010 [62] (1 % packet loss).

TR 101 329-6 [i.17] presents results of VoIP quality tests carried out among ETSI members. Among the information presented are results from subjective tests on the effects of packet loss. Figure C.3 depicts the effect of packet loss concealment where G.711 codec without any packet loss concealment is compared with two G.711 based PLC algorithms and G.729/G.7229A.

A paper by Kövesi and Ragot [i.18] presents results from selection tests of PLC algorithm for G:722 codec. Figure C.4 depicts results for both random packet loss and burst packet loss without and with PLC.

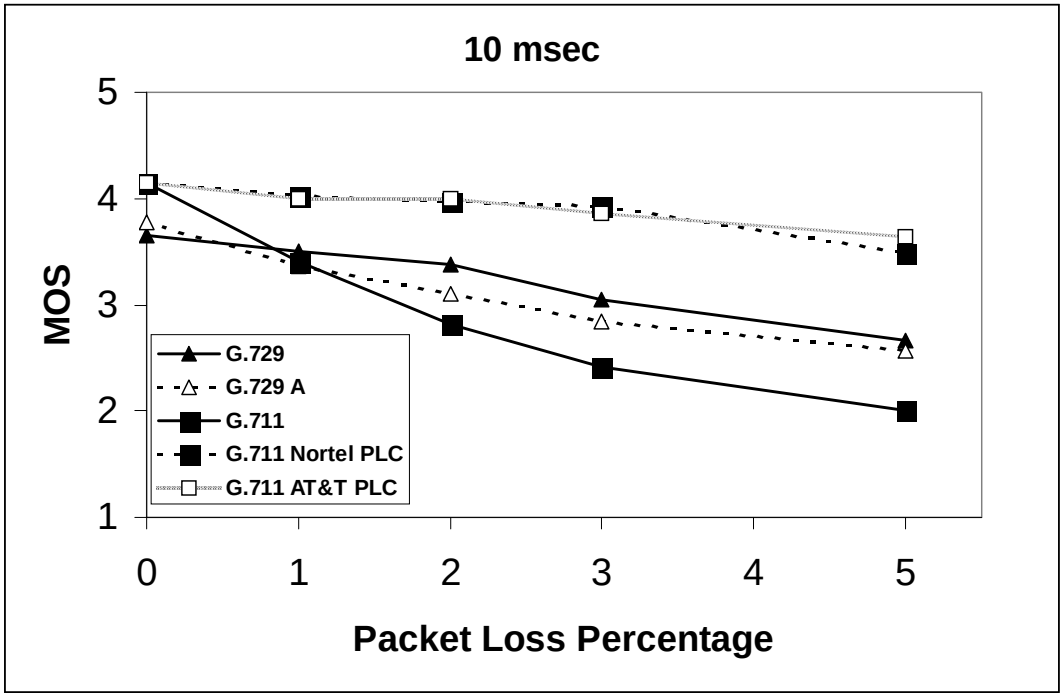


Figure C.3: Effects of packet loss on voice quality with 10 ms packets

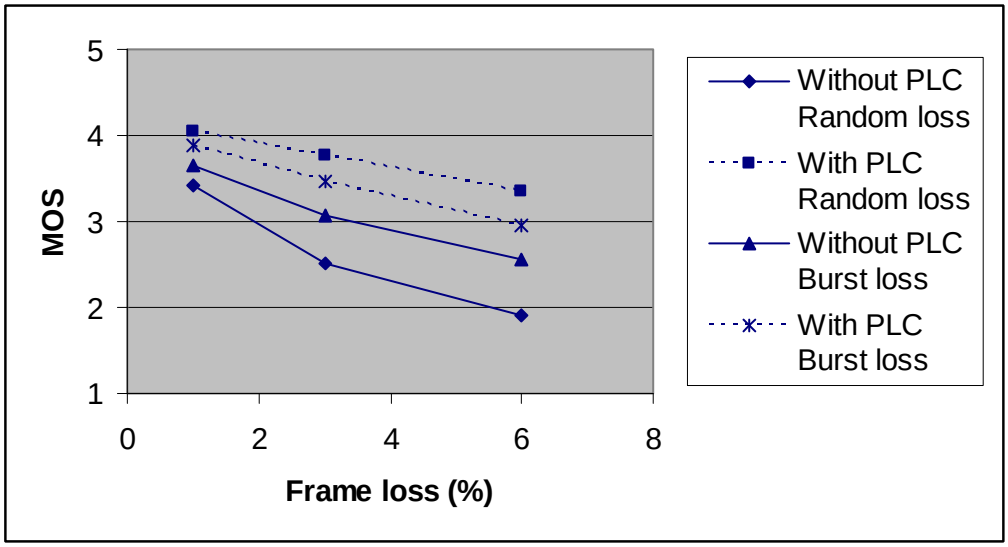


Figure C.4: Effects of packet loss on G.722 without and with packet loss concealment (PLC)

Annex D (informative): Provisional QoS Classes defined in ITU-T Recommendation Y.1541

This annex provides information about the provisional QoS classes described in ITU-T Recommendation Y.1541 [2]. For further information the recommendation should be consulted.

Table D.1: Provisional QoS classes defined in ITU-T Recommendation Y.1541

Network performance parameter	Nature of network performance objective	QoS Classes	
		Class 6	Class 7
IPTD	Upper bound on the mean IPTD	100 ms	400 ms
IPDV	Upper bound on the $1 - 10^{-5}$ quantile of IPTD minus the minimum IPTD	50 ms	
IPLR	Upper bound on the packet loss ratio	1×10^{-5}	
IPER	Upper bound	1×10^{-6}	
IPRR	Upper bound	1×10^{-6}	

History

Document history		
V1.1.1	February 2009	Membership Approval Procedure MV 20090410: 2009-02-10 to 2009-04-10
V1.1.1	April 2009	Publication