INTERNATIONAL INTERCONNECTION FORUM
FOR SERVICES OVER IP

(i3 FORUM)

(www.i3forum.org)

Workstream "Technical Aspects"

White Paper

Voice Path Engineering in International IP-based Networks

(previously titled
"Optimal Codec Selection in International IP-based Voice Networks")

(Release 3.0) May 2011

Executive Summary

This White Paper assists in correct codec and voice path parameter selection in different IP-based voice interconnection configurations, as well as to predict IP-based voice interconnection configurations which will have unacceptable voice quality degradation.

Voice path engineering (the practical application of codecs and choice of associated IP parameters) in IP-based voice networks is more complex in comparison to existing TDM networks; this document deals with the factors and configurations indispensable in correct network configuration and interconnection agreement planning, which have to be considered in order to deliver voice quality levels satisfactory for Service Providers.

Having introduced codec and VoIP media basics, voice quality planning basics and the significance of proper codec choice, this White Paper provides a methodology, spreadsheets and a calculation template useful to evaluate codec choice(s) for a particular distance of network configuration, thus indicating if it will be possible to achieve the required speech quality. If this calculation shows that expected (customer) quality will be below a satisfactory level it is possible to go through the calculations step by step and try to change codec or other parameters to reach the desired quality level.

It is shown that transcoding significantly affects call quality, and should be avoided unless absolutely necessary. The impact of transcoding is likely to be much higher when a chain of downstream carriers is involved in the end-user to end-user communication, than for bilateral interconnections engineered directly between network operators in the end countries, and may necessitate different network configurations being sought.

Extending IP-based voice networks into remote or island nations often needs expensive satellite transmission. Low Bit Rate codec choices and bandwidth reducing transmission techniques are given to assist network planners with this voice quality/bandwidth tradeoff.

This paper discusses the voice quality of the media path as affected by codecs as used in interconnected IP-based voice networks, covering and addressing narrow band, wideband, and low bit rate codecs used in links where bandwidth is costly such as satellites.

This white paper complements the content of the i3 Forum document “*Technical Interconnection Model for International Voice Services*” [1] with regard to the media information flow management / treatment.

Table of Contents

1 Scope and Objective

This paper discusses the voice quality of the media path as affected by codecs as used in interconnected IP-based voice networks, covering and addressing narrow band, wideband, and low bit rate codecs used in links where bandwidth is costly such as satellites.

The objective of this paper is to provide background to and to support the media section in "*i3 Forum, Technical Interconnection Model for International Voice Services*" [1] as well as to draw attention to the adverse voice quality which will result from inappropriate transcoding of low-bit-rate codecs. The causes and degradation of voice quality are established, tools for voice transmission planning are provided, with particular attention being drawn to transcoding impairments which may result in voice quality reduction so severe that alternative network arrangements to get to the final destination may need to be explored.

2 Acronyms

A/D	Analogue to Digital Converter, Analogue to Digital
ACELP	Algebraic-Code-Excited Linear Prediction
ADPCM	Adaptive Differential Pulse Code Modulation
ADSL	Asymetrical Digital Subscriber Line [equipment]
A-law	Companding (volume compression) profile used by all countries except North America and Japan
ALOC	Average Length of Call
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
Ann	Annex
Bpl	Robustness factor against packet loss (used for E-model calculations)
BurstR	Packet loss burst ratio (used for E-model calculations)
CELP	Code Excited Linear Prediction
CLR	Circuit Loudness Rating
CNG	Comfort Noise Generation
COS	Class Of Service
CPU	Centralised Processing Unit
CRTP	Compressed RTP
CS-ACELP	Conjugate-Structure Algebraic-Code-Excited Linear Prediction
D/A	Digital to Analogue Converter
DCME	Digital Circuit Multiplication Equipment
DECT	Digital Enhanced Cordless Telecommunications
DSL	[Symmetrical] Digital Subscriber Line [equipment]
DSP	Digital Speech Processor
DTX	Discontinuous Transmission
E1	2Mbit/s TDM transmission bearer, comprising 30 x 64bkit/s channels.
EF	Expedited Forwarding
EV-CELP	Embedded Variable bit rate – Code-Excited Linear Prediction
FoIP	Fax over IP
FR-AMR	Full-Rate Adaptive MultiRate
GSM	Global System for Mobile Communications
GSM-EFR	Global System for Mobile Communications – Enhanced Full Rate
Hz	Hertz
IP	Internet Protocol
IPv4	Internet Protocol, version 4
IPv6	Internet Protocol, version 6

IPDV	IP packet Delay Variation
ISDN	Integrated Services Digital Network
ITU-T	International Telecommunications Union – Telecommunications standardization sector
LBR	Low Bit Rate
LD-CELP	Low Delay Code Excited Linear Prediction
LPAS	Linear Prediction Analysis-by-Synthesis
MDCT	Modified Discrete Cosine Transform
MIPS	Millions of Instructions per Second
MLT	Modulated Lapped Transform
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
MOS-CQ	Mean Opinion Score-Conversational Quality
MOSCQE	Mean Opinion Score, Communication Quality Estimated
MOS-LQ	Mean Opinion Score-Listening Quality
MOS-LQO	Mean Opinion Score-Listening Quality Objective (i.e. objectively assessed)
MOS-LQOM	Mean Opinion Score-Listening Quality Objective in Mixed band [wideband and narrowband] context
MOS-LQON	Mean Opinion Score-Listening Quality Objective in Narrow band context
MOS-LQS	Mean Opinion Score-Listening Quality Subjective (i.e. subjectively assessed)
MOS-LQSM	Mean Opinion Score- Listening Quality Subjective in a Mixed-band context
MOS-LQSW	Mean Opinion Score-Listening Quality Subjective in Wideband context
MOS-TQ	Mean Opinion Score-Talking Quality
MP3	MPEG-1 Audio Layer 3, more commonly referred to as MP3
MPEG-2	Moving Pictures Expert Group 2 (for generic coding of moving pictures and associated audio information)
MPEG-4	Moving Pictures Expert Group 4 (for generic coding of moving pictures and associated audio information)
MP-MLQ	Multi Pulse Maximum Likelihood Quantisation
MR-ACELP	Multi-Rate Algebraic Code Excited Linear Prediction
ms	millisecond
µ-law	Companding (volume compression) profile used in North America and Japan
NB	Narrow Band (with respect to voice frequency signal band width), 300Hz to 3,400Hz
PC	Personal Computer
PCM	Pulse Code Modulation
PDV	Packet Delay Variation (see also IPDV)
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
POS	Packet Over SONET
pp	packetisation period
Ppl	Packet loss ratio (used for E-model calculations)
PSTN	Public Switched Telephone Network
qdu	quantisation distortion unit
QOS	Quality of Service
RAM	Random Access Memory
RCELP	Residual Code Excited Linear Prediction
Rec.	Recommendation
ROHC	RObust Header Compression
ROM	Read Only Memory
RPE-LTP	Regular Pulse Excitation-Long Term Prediction
RTCP	Real-Time Transport Control Protocol
RTP	Real-Time Transport Protocol
SB	Super Wide Band Audio – sometimes Super-Band - (with respect to voice frequency signal band width), 50Hz to 14,000Hz
SB-ADPCM	Sub-Band Adaptive Differential Pulse Code Modulation
SDP	Session Description Protocol
SIP	Session Initiation Protocol

SONET	Synchronous Optical Networking
SP	Service Provider
TCP	Transmission Control Protocol
TDBWE	Time-Domain BandWidth Extension
TDM	Time Division Multiplex
TELR	Talker Echo Loudness Rating
TFO	Tandem Free Operation
TrFO	Transcoder Free Operation
VAD	Voice Activity Detection
Var	Dynamically Variable bit-rate
VBD	Voice Band Data
VMR	Variable Multi Rate
VoIP	Voice over IP
VoIPv4	Voice over IP specifically packetised using the IPv4 protocol
VoIPv6	Voice over IP specifically packetised using the IPv6 protocol
VSELP	Vector Sum Excited Linear Predictive
WB	Wide Band (with respect to voice frequency signal band width), 50Hz to 7,000Hz
WMOPS	Weighted Million Operations Per Second

3 References

[1] "i3 Forum "Technical Interconnection Model for International Voice Services", Release 4, May 2011

[2] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies" (11/88)

[3] ITU-T Recommendation Y.2201, "Requirements and Capabilities for ITU-T NGN", (09/2009)

[4] ITU-T Recommendation G.107, Amendment 1 "The E-model, A Computational Model for use in Transmission Planning, Amendment 1: New Appendix II – Provisional impairment Factor Framework for Wideband Speech Transmission", (06/2006)

[5] "Low Bit-Rate Speech Coders for Multimedia Communication", R. Cox and P. Kroon, IEEE Communications Magazine, December 1996, pp 34 – 40

[6] ITU-T Recommendation G.108, "Application of the E-Model: A planning guide", (09/99)

[7] ITU-T Recommendation G.114, "One-way transmission time", (05/2003)

[8] ITU-T Recommendation H.324 "Terminal for low bit-rate multimedia communication".. (04/09)

[9] "A Sinusoidal Voice Over Packet Coder Tailored for the Frame Erasure Channel", Jonas Lindblom, IEEE Transactions on Speech and Audi Processing, vol 13, n5, September 2005, pp 787-798.

[10] ITU-T Recommendation P.800, "Methods for subjective Determination of Transmission Quality" (08/96)

[11] "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems", 3GPP2 C.S0014-A, Version 1, April 2004.

[12] IETF RFC3551, "RTP Profile for Audio and Video Conferences with Minimal Control", July 2003.

[13] ITU-T Recommendation G.722, "7 kHz audio-coding within 64 kbit/s" (11/88)

[14] ITU-T Recommendation G.729.1, "G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729" (05/06)

[15] ITU-T Recommendation G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)" (07/03)

[16] ITU-T Recommendation G.718, "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s" (06/08 published)

[17] ITU-T Recommendation G.726, , « 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM) » (12/90).

[18] ITU-T Recommendation G.113, "Transmission impairments due to speech processing" (11/2007)

[19] ITU-T Recommendation G.107, "The E-model, A Computational Model for use in Transmission Planning" (08/2008)

[20] ITU-T Recommendation G.113, Appendix 1, "Provisional planning values for the equipment impairment factor Ie", (12/98)

[21] ITU-T Recommendation G.711.1, "Wideband embedded extension for G.711 pulse code modulation" 03/08).

[22] ITU-T Recommendation P.10/G.100, "Vocabulary for Performance and Quality of Service" (07/2006)

[23] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", (02/01).

[24] ITU-T Recommendation P.564, "Conformance testing for voice over IP transmission quality assessment models", (11/07)

[25] ITU-T Recommendation P.862.1, “Mapping function for transforming P.862 raw result scores to MOS-LQO”, (11/2003)

[26] ITU-T Recommendation P.862.2, “Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs”, (11/07)

[27] ITU-T Recommendation P.805, “Subjective Evaluation of Conversational Quality“ (04/2007)

[28] ITU-T Recommendation P.833, “Methodology for derivation of equipment impairment factors from subjective listening-only tests”, (02/2001

[29] “VoIP Handbook, Applications, Technologies, Reliability and Security” Edited by Syed Ahson, Mohammad Ilyas, CRC Press, section 9.2.7.3, p160

[30] ITU-T Recommendation P.800.1, “Mean Opinion Score (MOS) terminology”, (07/2006)

[31] ITU-T Recommendation G.109, “Definition of Categories of Speech Transmission Quality“, (09/99)

[32] ITU Study Group 12 – Delayed Contribution 151, “Towards a Wideband E-Model: R-Scale Extension and Impairment Factors for Wideband Speech Codecs” COM 12 – D 151 – E, Question 8/12, Geneva 5-13 June 2006 (available to ITU Members)

[33] ITU-T Recommendation G.1020, “Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks”, (07/2006)

[34] ITU-T Recommendation Y.1541, “Network performance objectives for IP-based services”, (02/2006)

[35] IEEE 802.16 Broadband WirelessAccess Working Group, “Downlink VoIP Packet Delay Jitter Model”, 2007-07-10.

[36] “An Inter-arrival Delay Jitter Model using Multi-Structure Network Delay Characteristics for Packet Networks”, E. J. Daniel, C. M. White and K. A. Teague, Proc. 37th Asilomar Conference of Signals, Systems, and Computers, vol. 2, November 2003, pp. 1738-1742.

[37] “Quality bounds for packetized voice transport”, D. De Vleeschauwer, J. Janssen, G. H. Petit, F. Poppe, Alcatel Telecommunications Review – 1st Quarter 2000, pp 19 – 24

[38] TIA/EIA Technical Services Bulletin PN-4689, “Telecommunications IP Telephony Equipment: Voice Quality Recommendations for IP Telephony”, (to be published as TIA/EIA/TSB116) (undated)

[39] ITU-T Recommendation G.711 Appendix I, “Pulse code modulation (PCM) of voice frequencies, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711”, (09/99).

[40] “The ETSI Computational Model: A Tool for Transmission Planning of Telephone Networks” IEEE Communications Magazine, January 1997, pp. 70-79

[41] ITU-T Recommendation G.107 “The E-model, A Computational Model for use in Transmission Planning”, (12/98)

[42] ITU-T Recommendation G.729 “Coding of Speech Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)”, (01/2007)

[43] ITU-T Recommendation P.834, “Methodology for the derivation of equipment impairment factors from instrumental models”, (07/2002).

[44] IETF RFC3267, “Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs”, June 2002.

[45] IETF RFC4348, “Real-Time Transport Protocol (RTP) Payload Format for the Variable-Rate Multimode Wideband (VMR-WB) Audio Codec”, January 2006.

[46] ITU-T Recommendation G.723.1, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s”, (05/06)

[47] ITU-T Recommendation G.728, “Coding of speech at 16 kbit/s using low-delay code excited linear prediction”, (09/92)

[48] ITU-T Recommendation G.722.2, “Wideband coding of speech at around 16k bit/s using Adaptive Multi-Rate Wideband (AMR-WB)”, (07/2003)

[49] “E-model supported switching between narrowband and wideband speech quality”,. International Workshop on Quality of Multimedia Experience, QoMEx 2009, Volume 29,

Issue 31, July 2009, Pages:98 – 103, Digital Object Identifier 10.1109/QOMEX.2009.5246970

[50] IETF RFC1661, “The Point-to-Point Protocol (PPP)”, July 1994.

[51] IETF RFC1662, “PPP in HDLC framing”, July 1994.

[52] IETF RFC2615, “PPP over SONET/SDH”, June 1999

[53] ITU-T Recommendation G.131, “Talker echo and its control”, (11/2003)

[54] 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 " (05/2005)

[55] “Quality Comparison of Wideband Coders Including Tandeming and Transcoding”, Catherine Quinquis, France Telecom R&D, ETSI Workshop on Speech and Noise in Wideband Communication, 22-23 May 2007, Sophia Antipolis, France.

[56] “Quantifying the Quality Difference between narrow-Band and Wideband Speech Codecs”, Sebastain Moller and Alexander Raake, Fortschritte Der Austik, 2005, Vol 31, Band 1, pages 157-158; also in English at http://www.isti.tu-berlin.de/fileadmin/fg41/download/Publikationen/Sebastian_Moeller/Extern/MR-DAGA-05.pdf

[57] ITU-T Recommendation G.711 – Appendix II, “Pulse code modulation (PCM) of voice frequencies, Appendix II: A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems” (02/2000)

[58] IETF RFC2508, “Compressing IP/UDP/RTP Headers for Low Speed Serial Links”, February 1999

[59] IETF RFC3095, “RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed:, July 2001 “.

[60] IETF RFC4995, “The Robust Header Compression (ROHC) Framework”, July 2007.

[61] ITU-T Recommendation Q.50 (03/93) (see also Implementors’ Guide (05/98)

[62] ITU-T Recommendation V.29, “9600Bits per second Modem Standardised for use on Point-to-Point 4-Wire Leased Telephone–type circuits”, ITU Blue Book, 1993.

4 Voice Path Engineering in IP networks

All TDM switching and network interconnections use G.711 PCM coded 64kbit/s voice signals and the TDM PSTN is engineered around this specific voice path design. Any transcoding to lower bit rates to lower transmission costs is undertaken by Digital Circuit Multiplication Equipment (DCME) either within a carriers network or on the bilateral interconnection link (interfaces to TDM switches are always 64kbit/s). Consequently voice path engineering expertise resided mainly with DCME equipment vendors. Codecs used in DCME are predominately 16kbit/s / 32kbit/s and speech quality reduction is low-to-moderate. In most international connections (i.e. all submarine cable links with global reach) an "*All Users Satisfied*" quality levels is readily achieved in bilateral TDM international networks, and to a lesser extent, when a chain of networks is involved. Effectively, voice path engineering had a low profile.

The freedom to use different codecs (combined with packet transmission parameters) now requires voice path engineering to be part of IP-based voice design. With IP-based voice, there are many changes which will have profound impacts:

1 transmission bandwidths increase because of packetisation overheads encouraging the use of low-bit-rate codecs to offset the bandwidth (cost) increases. These codecs generally have worse speech quality (both in voice distortion and codec related delay);

2 the delay of the IP packetisation processes throughout the call chain has significant additional impact on speech quality;

3 many more codecs have been developed, so that a significant diversity of codec types will be encountered in domestic networks (codecs are chosen predominantly for domestic market reasons; international carriers generally carry signals, and, if required, mediate technically mismatching voice signals);

4 interconnections are no longer to a common codec standard, but are according to the codecs used by the respective carriers being interconnected, thus codec and packetisation matters are now a required component of interconnection negotiations.

For the reasons given above, Service Provider (access) networks now introduce significant delay to the end-to-end delay budget formerly dominated (for intercontinental distances) by propagation delay, increasing the probability of lower user satisfaction.

Such increased delay, combined with low-bit-rate codec impairment (voice distortion), could reduce the best case estimate (with codec and delay impairments only accounted for) of customer opinion almost to the "*Many Users Dissatisfied*" level, so that when other impairments unavoidable in practical international connections are included, international call user quality can be demonstrably lower for IP-based voice (contrasted with current PSTN quality which typically meets customer "*Satisfied*" scores for similar calls).

In addition, the already mentioned diversity of codecs now available means that it would be unrealistic to expect all Service Providers to use the same codec. While it is firstly the responsibility of Service Providers to transcode if needed to ensure voice service interoperability (particularly relevant if that SP chooses a different codec from other carriers in their domestic interconnect environment), however, in case no common codec can be negotiated between end Service Providers, international carriers may provide transcoding for some calls simply to connect them.

Particularly hard hit will be calls of global reach (halfway around the World) and those necessitating satellite for completion (as is the case from Europe to many Pacific Islands). Clearly such degradation could be mitigated slightly by choosing higher bit rate codecs but this comes with a bandwidth cost (often several times higher), presenting a difficult commercial trade-off. There may be no practical alternative but to transcode to a low bit rate codec (as well as use bandwidth reducing VoIP transmission techniques) when using satellite links to access some geographic regions.

The involvement of a chain of international carriers poses a particular problem for planning quality in that there may be several intermediate carriers, and information about codec and packetisation downstream from the contracting first operator may be hard to obtain, thus frustrating call quality estimation.

If cost is the dominant criterion of an intermediate carrier, they may transcode within to save capacity costs, consequently profoundly impacting the end-to-end call they are involved with. Conversely, it may happen that the same codec is used throughout, with quality maintained.

It is concluded that for IP-based voice, bilaterally engineered interconnections will offer predictable quality better able to be matched to voice product requirements, and particularly will offer the lowest quality reductions vis-à-vis TDM because of more direct connections reducing impairments.

Mobile Service Providers use codecs designed for spectrum conservation, and dynamically change the codec parameters to compensate for radio signal strength variations during a call so that, taken together with packet loss on the radio path, generally mobile codecs, under practical use conditions, often have lower voice quality (higher distortion) than fixed codecs. Further mechanisms have been defined within IP centric mobile networks to allow end to end packet connections with no transcoding, but transcoding is likely to remain a feature of mobile-fixed network calls for the short term.[1]

As a result, carriers now require voice path engineering knowledge (the practical application of codecs and associated packet transmission parameters) to be able to engineer voice circuits in IP-based voice networks.

5 General Reference Architecture

The general reference configuration for an international voice interconnection based on the IP protocol given in [1] is reproduced here to include codec/transcoding functions which can be invoked at the Border Function.

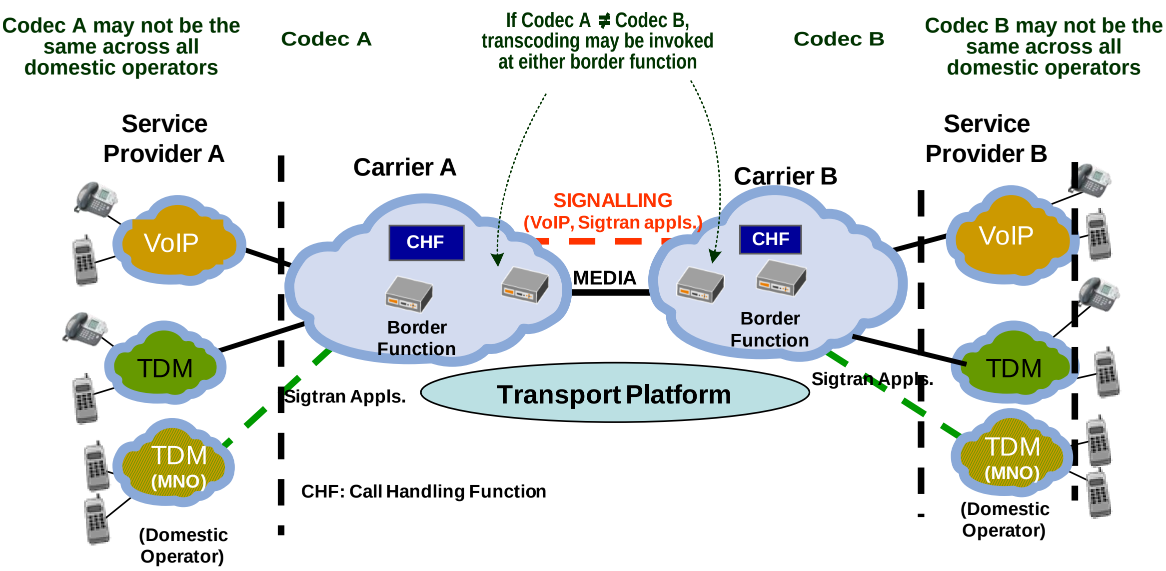


Figure 1 General Reference Configuration with codec annotation

1 Devices are expected to become increasingly convergent and so will share the same codec for mobile and fixed usage. With the likely evolution towards IP mobile, negotiation from end-to-end of the same codec is likely and thus communication will become increasingly transcoding free.

6 Codec and Voice Path Engineering Basics

The voice (user) signal is converted to a digital signal at (or near) the user end-point in the domestic/access network by a codec. A codec is a device for encoding and/or decoding a digital signal (coder/decoder), either from analogue (e.g. end user voice) or from a differently coded digital signal.

6.1 Coding Algorithm – Technology

Speech coding is the process of reducing the bit rate of digital speech representations while maintaining a quality acceptable for the application.

Most codecs are designed for the telephony speech bandwidth of 300-3400Hz; this bandwidth ("narrow band") ensured sufficient intelligibility and was the basis of the design of fixed TDM networks, which use the G.711 codec [2]. This bandwidth constriction does not apply to IP-based voice networks, and codecs are now being designed for higher speech bandwidth. Wideband codecs have a frequency range of 50Hz to 7,000Hz ([3], section 14.2.2, [4]), and higher bandwidth superwideband audio codecs are defined with frequency range 50Hz to 14,000Hz ([3], section 14.2.2). Narrow band codecs are still in very common use due to interworking with the PSTN.

Some speech coders are optimised for multi-media (where several applications signals will share the communications channel), and some for telephony. Bit rate, encoded bandwidth (narrow band, wideband or higher), complexity (CPU time to compute the code, static/dynamic RAM and ROM memory), delay and speech quality are typical trade-offs in codec design. It is predominantly the trade-offs in codec design that distinguish them[2].

6.1.1 Waveform codecs

Waveform codecs simply process the speech waveform as it arrives, sample by sample, e.g. narrow band codec's G.711 PCM[3] and G.726 ADPCM as 0.125 ms samples.

6.1.2 Non-Waveform Codecs

Many of the low bandwidth codecs used in IP-based voice telecommunications (commonly referred to as low bit rate codecs) are Linear Prediction Analysis-by-Synthesis (LPAS) codecs (e.g. G.729 and its annexes, G.728, G.723.1, GSM full rate, half rate and enhanced full rate etc) [5]. These are non-waveform codecs and use speech synthesis techniques ([6], A.1.8).

In non-waveform codecs many speech samples are grouped into a frame (see section 6.4.1), and processed (encoded) en-bloc into a new digital signal (code) with certain assumptions such as knowledge that the signal represents speech, so that certain fixed characteristics can be assumed. For narrow band codecs the speech samples input are provided by the G.711 PCM codec (or linear for some VoIP terminals) at 0.125ms intervals (8kHz sampling frequency). For wideband codecs a 16kHz sampling frequency is used, providing speech samples at 0.0625ms intervals.

Additional accuracy is obtained by including part of the next frame also in the calculation; this extra information is called "look-ahead" (see section 6.4.2) and improves the speech representation for a small increase in coding time. The encoding entails each frame of input signal being processed at

---

[2] Examples: G.729 was designed for lower complexity than G.728, and has higher delay (called algorithmic delay) for similar speech quality; G.723.1 was designed for low-bit-rate videophones of that era (where delay was increased to lower the frame rate to match videophones and encoded bandwidth was made as low as possible to fit alongside video in the relatively low bandwidth lines available); it has now been superseded by the AMR codec in low bit rate circuit switched videotelephony based on ITU-T H.324 [8]; G.729a was designed for lower complexity than G.729, at expense of slightly higher voice distortion [5].

[3] The G.711 codec, being predominantly the A/D function of converting from analogue to digital (linear) PCM, also contains a companding function, which follows the μ-law recommendation in USA and Japan, and the A-law recommendation in other countries. Companding conversion responsibility is covered in section 11.

the encoder to extract a set of parameters that are quantised (using codebooks for vector quantization or scalar quantiser) to be converted to a bit stream (the new coded signal) and transmitted to the decoder.

When decoding from a non-waveform codec, the frame information is computed along with characteristics of speech assumed at the encoder which is also stored in the decoder in what is called a ‘codebook’ (i.e. this information is not transmitted). Thus the speech is “synthesised” from the coded information sent plus the transmitted “codebook” index.

Recently codecs have also been designed specifically for use in packet networks, where packet loss[4] becomes an important design trade-off (e.g. [9]). Most of them include a Packet Loss Concealment (PLC) algorithm to generate at the decoder side the best possible synthesis signal even if all corresponding input frames have not been received. Concealment is achieved by using information from the previously received frames. Latency generally increases which is an acceptable tradeoff when used for internet telephony where the IP transmission channel cannot be of guaranteed quality.

6.2 Bit rate – necessary bandwidth

The bandwidth of IP-based voice signals is higher than that of equivalent TDM signals primarily because of packet overheads. This encourages the use of more bandwidth efficient codecs and more coded voice frames per IP packet to offset the increase in international transmission costs as TDM to IP-based voice migration occurs. The drawbacks of augmenting the size of IP packets for transporting voice are therefore important and must not be forgotten, namely increased sensitivity to packet loss and increased latency (see also section 6.4.3).

As the bit rate of the codec is reduced to seek bandwidth efficiency, speech distortion increases (section 6.5). Section 12.1 provides more in depth information on minimising the bandwidth of VoIP signals.

6.3 Encoded bandwidth: narrow band versus wideband codecs

IP-based voice gives the opportunity to improve encoded voice quality decisively by moving from the “historic” PSTN narrowband (NB) quality (300 to 3,400 Hz using a 8 kHz sampling frequency) to wideband (WB) quality (50 to 7,000 Hz using a 16 kHz sampling frequency). Wideband quality means voice better encoded on all its frequencies, with more natural sound and a greatly improved sensation of presence (in the voice sense), intelligibility and listening comfort.

6.4 Encoding and Packetisation Latency

Encoding/decoding digitised voice and loading/unloading packets for transmission in packet networks introduces several types of delay.

6.4.1 Frame length

The frame length is the length of the speech waveform that is generally processed at a time (see also “look-ahead” in section 6.4.2). A waveform sample is digitalised in the case of waveform codecs or speech parameters are computed in the case of speech synthesis (non-waveform) codecs for each frame and transmitted for every frame. The speech representation is reconstructed at the decoder.

4 A packet network with packet loss equates to a frame erasure channel.

6.4.2 Look-ahead

To analyse the speech properly, speech data beyond the frame boundary is commonly included in non-waveform codecs frame encoding calculations. This is called look-ahead. Thus it is necessary to buffer a frame plus look-ahead, and this is called algorithmic delay. It cannot be reduced in implementation (the subsequent CPU processing time to calculate the speech parameters may vary, and is assumed by the ITU-T to be optimum when equal to the frame length, [7] Annex A).

6.4.3 Packetisation

For IP transmission the continuous digital voice signal from the codec has to be packetised, requiring dividing the encoded signal into equal length sections which comprise the IP packet payloads. The length of each section is a multiple of the codec frame length.

Transmitted bandwidth can be reduced by increasing the size of the IP packet payload by loading multiple speech frames into each packet, however this increases the total latency thus reducing speech quality for end-to-end calls >150-200 ms (see section 7.5). Examples of transmission bandwidth at the link layer are given in the i3 forum Technical Interconnection Model [1].

Packetisation periods (pp) longer than 40 ms are not used in telecommunications networks due to additional latency and increased risk of voice clipping (an upper limit of 64 ms per IP packet is recommended by the ITU-T G.108 [6], Annex B, B.3).

6.4.4 Output Queuing Delay

This is the time taken at the send end to “clock” the packetised signal into an IP facility, and is generally low except for some Service Provider (Access) networks which have low bandwidth.

6.4.5 De-Jitter Buffer

A de-jitter buffer is required at the receive end of an IP transmission network to store the arriving packets to facilitate a continuous playout of the de-packetised, coded digital signal into the decoder. This buffer counters transmission timing variations in the packet network and clock asynchronism. The de-jitter buffer is described more fully in section 8.3.3.

6.4.6 Combined Effect of Speech Processing Delay Factors

The minimum codec speech processing delay is

(frame length + look ahead) + frame length = 2 x frame length + look-ahead

where the second frame length is the time to calculate the coded signal (CPU time), assumed optimised when calculation is finished just as the next frame is available for calculation.

Loading the frames of coded voice into IP packets is practically instantaneous, [7] Annex A. However for multiple frames per packet, additional latency results from the time the first frame is held until the final frame is calculated and available to concatenate and drop into the IP packet. Additional delay to clock the packets out into the link layer is low for a high speed link, thus speech processing time (codec processing + packetisation) is

(N + 1) x Frame length + look-ahead

where N is the number of frames per packet [7].

The codec is generally located in the Service Provider access network where, if the bandwidth is limited, or congestion occurs, the delay may increase over that given above. The maximum speech processing time permitted, [7] Annex A, is

(2N + 1) x Frame length + look-ahead

Common frame lengths and packetisation periods used for several codecs, together with one-way delays of coder and packetisation time processing in accordance with ITU-T G.114, [7] table I.4, are given in Table 1.

6.5 Speech Distortion

Preserving speech as naturally as possible is essential to satisfy users. Generally low bit rate codecs have an increased complexity (resulting in latency increase and more computation) to minimize distortion. In addition, they become optimised for speech (the "codebook" parameters are optimised for speech, see section 6.1.2) to minimize degradation at lower bit rates[5, 6]. When codecs are operated at very low bit rates speech tends to become metallic or robotic, losing its naturalness.

Codec	Frame size (ms)	Look-ahead (ms)	Typically used packetisation periods (pp) (ms)	One-way delay introduced by coder-related processing per G.114 (ms)	
				Min	Max
G.711	0.125	0	10	10.125	20.125
			20	20.125	40.125
			40	40.125	80.125
G.729	10	5	10	25	35
			20	35	55
			30	45	75
			40	55	95
G.723.1	30	7.5	30	67.5	97.5
AMR	20	5	20	45	65
			40	65	105
G.726	0.125	0	10	10.125	20.125
			20	20.125	40.125
			30	30.125	60.125
FR-AMR	20	5 (note 1)	20	45	65
G.722	0.125	0	10	10.125	20.125
G.722	0.125	0	20	20.125	40.125
G.722.2 / AMR-WB	20	5	20	45	65

Note 1. The 5mS look ahead is a dummy at the 12.2kbit/s Full Rate to allow seamless frame-wise mode switching with the rest of the FR-AMR rates.

Note 2. The higher pp values such as 40ms are less commonly used than the lower pp values (particularly 20ms) and are included here because of their relevance to reducing transmitted bandwidth on links with high bandwidth cost, see section 12.

Table 1 Common Codec Frame Sizes, Packetisation Periods and Encoding + Packetisation times

[5] This means that Low Bit Rate speech codecs generally cannot handle music, nor do they transmit tones or fax transmissions reliably, so that if tones must be transmitted, codecs such as G.711 must be used.

[6] It is common to optimise codecs for the application. Other codecs are optimised for music, such as MP3, and video, such as MPEG-2 and MPEG-4.

Voice codec basic features, including the codecs cited in the "*i3 Forum Technical Interconnection Model for International Voice Networks"* [1] are presented in Table 2.

As codecs are developed it is common to conduct subjective voice quality tests (see section 7.1) according to ITU-T Rec. P.800 [10] on that codec. The values (generally expressed as Mean Opinion Scores – MOS) resulting from such tests depend on the test configurations (see section 7.1) but are an accurate customer opinion rating when many listeners and many languages are admitted to the experiments. These values are known as the intrinsic MOS for that codec, and equivalent values expressed as the R-factor are included, where known, in Table 2 (see section 7.2 for the R-factor and its conversion to MOS in section 7.3).

6.6 Voice Activity Detection and Discontinuous Transmission

Conversational speech is generally punctuated by periods of talker "silence" as the "far-end" customer speaks. During such periods of "silence", the outgoing transmission rate may be discontinued (discontinuous transmission is referred to as DTX). Periods of active speech are detected by a Voice Activity Detector (VAD), with a fast attack time to avoid speech clipping, and a hangover time to ensure that speech has truly stopped. VAD functions may be built into the codec design, and the transmitted, packetised signal contains "instructions" to the decoder to decode the resulting signal correctly.

Digital transmission channels are never completely silent, containing a base level of quantising noise inherent from the A/D process plus added background noise of the speakers local environment. Thus discontinuous transmission would create an unnatural (uncomfortable) silence interpretable by the listener as a broken connection if artificial noise was not inserted to simulate a continuous channel. Such "comfort" noise is injected at the receiving end by a Comfort Noise Generator (CNG). Codecs with VAD/DTX/CNG also send, in the "silence" period, a description of the noise level and associated spectral information to allow the CNG to mimic and track the actual sending end noise level, so that the listener does not notice noise level changes between speaking and silent periods. This information occupies a small transmitted bandwidth, thus the use of VAD/DTX/CNG can considerably reduce the average voice packet transmission rate and hence improve bandwidth efficiency. Packet size may be reduced and transmission rate lowers during silence.

Some mobile codecs utilise dynamic control of the codec bit rate to achieve similar bandwidth efficiency to VAD/DTX. For example, EVRC [11] codes background noise at lower rates than active speech [7].

Codecs with VAD/DTX/CNG are included in Table 2. More information on VoIP signal bandwidth reduction by VAD/DTX/CNG is given in section 12.1.2. Information on RTP packetisation of voice signals with DTX (silence suppression) independent of the codec is given in [12], section 4.1.

VAD/DTX introduces some latency, typically 20-40ms, with the higher values associated with codecs which have longer frame lengths.

6.7 Coding of Wideband Speech

Wideband codecs often code information for different sub-bands separately to diminish complexity. For example the speech frequency band input to the G.722 [13] and G.729.1 [14] wideband codecs is split into a lower sub-band to 4KHz and a higher sub-band 4KHz to 8KHz, and each sub-band signal is separately ADPCM encoded, the technique being called sub-band ADPCM (SB-ADPCM). The AMR-WB (G.722.2 [13]) and G.718 [16] codecs encode separately the sub-bands 50Hz – 6.4KHz and 6.4 – 7KHz .

[7] EVRC [11] codes active speech at Rate 1 (171 bits per packet = 8.55kbit/ss) or Rate ½ (80 bits per packet = 4kbits/s) and background noise at Rate 1/8 (16 bits per packet = 0.8kbits/s).

6.8 Mobile Codecs

Generally, mobile codecs are designed for radio spectrum conservation and commonly have dynamically variable bit rates to compensate for radio signal strength variations during a call. For a call between two mobiles in a single SP area, the codec may operate at different bit rates on the A and B legs of the call, due to different radio paths conditions. This means a transcoding[8] takes place between the two “codecs” (except if TFO or TrFO are used)

Earlier generations of mobile codecs typically have lower voice quality than “wire-line” or fixed network codecs, and quality varied significantly during a call. However mobile codecs brought into service over the last few years (such as GSM-EFR, AMR at bit rates above 8 kbit/s) have very good speech quality, which under no packet loss conditions in fixed networks perform significantly better than G.729.

For mobile calls the codec impairment is mainly increased due to Frame Erasure caused by packet loss on imperfect radio paths. Some mobile codecs (such as G.722.2 [15]) also have the capability to compensate for lost frames if externally signaled when frames are lost or corrupted.

6.9 The Media Stream and Media Stream Conversion

The stream of packets containing the voice signal in a VoIP network is usually referred to as either the RTP stream (after the Layer 4 protocol header) or the media stream. Building a packetised media stream from the basic PCM continuous digital voice signal involves many signal processing steps, most of which are effectively standardised (such as choice of the UDP transport header rather than TCP[9]).

The three parameters of particular interest to the IP-based voice path engineer are the codec, the packetisation period (both covered earlier in this section) and the G.711 companding law (covered in section 11). Changing any of these parameters in the transmission path is called media stream conversion. The most commonly referred to media stream conversion is transcoding, which strictly means converting the encoding of the voice from one codec (such as G.729) to another codec (such as AMR-NB). This is covered more in section 9. Changing the packetisation period is sometimes called transrating, and some situations where transrating might be considered are given in sections 12.1.2 and 12.1.3. Both codec and packetisation period conversions require the VoIP signal to be de-packetised, introducing additional latency and interrupting the continuity of the RTCP stream, which potentially limits the usefulness of RTCP for QOS measurements.

The G.711 codec, being predominantly the A/D function of converting from analogue to digital (linear) PCM, also contains a companding function, which follows the µ-law recommendation in USA and Japan, and the A-law recommendation in other countries [2]. Companding conversion is also a media stream conversion and IP-based voice engineers working with G.711 must take care not to overlook this requirement as companding conversion may have to be specifically included in some possible network configurations (note that responsibility lies, by international agreement [2], with the µ-law countries, and generally the international carrier at the international/domestic interface has taken the responsibility for conversion in TDM networks). This is a particularly important media conversion give the prevalence of the G.711 codec in NB voice; more details are in section 11.

---

8 This type of transcoding is often called self tandeming.

9 This eliminates the latency that TCP retransmission requests can introduce, it is better to have some missing voice packets than delay the whole signal to wait for missing packets to be re-transmitted.

A Narrowband codecs

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	Ie-eef (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
G.711	PCM	8	0,3 – 3,4	Fix	64	0,125	10	0	0,25	10,125	20,125	Yes	0,01	App II (y)	0	4(q)				N	92,3
							20			20,125	40,125					5 (r)	4			N	
G.711+PLC							10			10,125	20,125					25 (q)	Y			App I	
							20			20,125	40,125					25 (q)	Y			App I	
							20			20,125	40,125						5,91	6/1.5	7	(p)	85,3
							20			20,125	40,125						7,84	8/2	10	(p)	82,3
G.711.0	LC PCM	8	0,3 – 3,4	Var	~32	5	10	0	10	10	15	Yes	1.667 WMPOS	G.711 AppII	0						92.3
G.729	CS-ACELP	8	0,3 – 3,4	Fi	8	10	10	5	25	25	35	no	18	Ann B	10					Y	82,3
							20			35	55									Y	82,3
G.729a+VAD	CS-ACELP	8	0,3 – 3,4	Fix	8	10	10	5	25	25	35	no	10.5	Ann B	11	19 (r)	Y			Y	81,3
							20			35	55						Y			Y	81,3
G.729d	CS-ACELP	8	0,3 – 3,4	Fix	6,4	10	10	5	25	25	35	no	20	Ann B/F						Y	
							20			35	55			Ann F						Y	
G.729e	CS-ACELP	8	0,3 – 3,4	Fix	11,8	10	10	5	25	25	35	no	25-30	Ann B/G	4	8 (r)	4			Y	88,3
							20			35	55			Ann G	4					Y	88,3
														Ann G	4	8 (r)	5,91	6/1.5	9	Y	83,3
														Ann G	4	8 (r)	7,84	8/2	11	Y	81,3

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	Ie-eef (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
G.729.1 Narrow band low delay mode (8, 12 kbit/s)	EV-CELP +TDBWE+ MDCT	8	0,05–4.0	Var	8 12	20	20	5	25	25	45	Yes	14,48 WMOPS (8 kbit/s) 17.30 WMOPS (12 kbit/s)	Ann C						Y	
G.723.1	ACELP	8	0,3 – 3,4	Fix	5,3	30	30	7,5	67,5	67,5	97,5	no	18-20	Ann A	19		Y			Y	73,3
	MP-MLQ				6,3										15	16 (r)				Y	77,3
AMR	MR-ACELP	8	0,3 – 3,4	Var	4,75-10.2	20	20	5	45	45	65	no	16,7 WMOP	Y						Y	
					12.22							GSM-EFR			5 (z)						87.3 (z)
G.726	ADPCM	8	0,3 – 3,4	Fix	16	0,125	10	0	0,25	10,125	20,125	Yes	~8	N	50					N	42,3
					24										25					N	67,3
					32										7					N	85,3
					40										2					N	90,3
ILBC (v)	BI-LPC	8	0,3 - 3,4	Fix	15.2	20	20	0	40	40	60	no	18	Y	10					Y	82.3
					13.3	30	30	0	60	60	90									Y	
GSM-HR	VSELP	8	0,3 - 3,4	Fix	5.6	20	20	0	40	40	60	no	15.2 WMOPS	Y	23					Y	69.3
GSM-FR	RPE-LTP	8	0,3 - 3,4	Fix	13	20	20	0	40	40	60	no	15.2 WMOPS	Y	20					Y	72.3
GSM-EFR	ACELP	8	0,3 - 3,4	Fix	12.2	20	20	0	40	40	60	no	15.2 WMOPS	Y	5	10				Y	87.3
G.718 (x)	CELP + MDCT	8, 16	0.3 - 3.4	Var	8, 12, 16, 24, 32	20	20	13.875 – 23.875	53.875 – 63.875	53.875 – 63.875	73.875 – 83.875	Yes	43.9 WMPOS	Y							
G.728	LD-CELP	8	0,3 - 3,4	Fix	16	0.625	10	0	1.25	10.625	20.625	Yes	35-40	N	7					Ann I	85.3
					12.8 Ann H								35-41	N	20					Ann I	72.3
					9.6 Ann H																

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	Ie-eef (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
IS-127 EVRC	RCELP/ACELP	8	0,3 - 3,4	Var	0.8 , 4, 8.55	20	20	5	45	45	65	no	25-30	Y						Y	
	ACELP				8.5										6					Y	86.3
SILK		8, 12,		Var	6 - 40	20	20	5	25					N							
SVOPC (w)		8		Fix or Var				30ms													

B Wideband Codecs

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie, wb* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	*Ie-eef* (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
G.711 (j)															36						93
G.711.1 (n)	PCM & MDCT	8 and 16	0,05 - 7,0	Var	64,80,96	5	5	6.875	16.875	16.875	21.875	Yes	8.7 WMOPS	N						Y	
G.718 (x)	CELP + MDCT	16	0.05-7.0	Var	8	20	20	22.875	62.875	62.875	82.875	Yes	42.8 WMOPS	Y							
					12								48 WMOPS								
					16								52.8 WMOPS								
					24								54.9 WMOPS								
					32								55.9 WMOPS								
					12.65								42.1 WMOPS		13	4					116

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie, wb* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	*Ie-eef* (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
G.729.1	EV-CELP +TDBWE	8 or 16	0,05–4.0 or 0,05– 7,0	Var	14	20	20	5 (low delay mode)	28.9375	28.9375	48.9375	Yes	24 WMOPS	Ann C						Y	
G.729.1 (l)	EV-CELP +TDBWE+ MDCT	8 or 16	0,05–4.0 or 0,05– 7,0	Var	14, 16, 18, 20, 22, 24, 26, 28, 30, 32	20	20	28,9375	68,9375	68,9375	88,9375	Yes	36 WMOPS	Ann C						Y	
					24										16	3				Y	113
					32										7	6				Y	122
G.722	SB-ADPCM	16	0,05 –7,0	Fix	48	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	31					N	98
					56	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	20						109
					64	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	13	7				App III	116
													10 MIPS	N	13	5				App IV	116
													10 MIPS	N	13					N	116
G.722.1	MLT	16	0,05 - 7,0	Fix	24	20	20	20	60	60	80	no	< 5.5 WMOPS	N	19					N	110
					32	20	20	20	60	60	80	no	< 5.5 WMOPS	N	13						116
AMR-WB / G.722.2 (k)	ACELP	16	0,05 - 7,0	Var	6.6 – 23.85	20	20	5	45	45	65	no	39 WMOPS (s)	Y						App I	
					6,6								(t)	Ann A&B	41						88
					8,85								(t)		26						103
					12,65								(t)		13	4					116
					15,85								(t)		7						122
					23,05								(t)		1	5					128
					23,85								(t) (u)		8	5					121
SILK		16		Var	8-30	20	20	5	25	25	45			N							

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Codec Processing Delay (b)	Min Codec + Packetisation one way delay (c)	Max. Codec + Packetisation one way delay (c)	Transcoding tolerance (e)	CPU Load	VAD / DTX / CNG (d)	*Ie, wb* (f)	B_{pl} (g)	Burst R (g)	n/ Ppl (o)	*Ie-eef* (o)	PLC (i)	R-factor (h)
		kHz	kHz		kbit/s	ms	ms	ms	ms	ms	ms		MIPS								
SILK		24		Var	12-40	20	20	5	25	25	45			N							
SVOPC (w)		16		Fix or Var		10	10	30ms	40	40	50			Y							

Terms used in the table

1. Var means Dynamically Variable bit-rate during a call. Some fixed codecs have different bit rates available, but the rate does not change during call.
2. MIPS Millions of Instructions per Second
3. WMOPS Weighted Million Operations Per Second, an ITU-T measure of computational complexity, similar to MIPS. WMOPS are roughly equivalent to MIPS for fixed-point processors used in commercial codecs.

Notes:

(a) Typically encountered packetisation rates only.

(b) Codec Processing Delay refers to the processing delay requirements of the encoder (send side) for a single frame as per G.114 A.2 [7] This is longer than quoted algorithmic delays for encoders.

(c) Min. and Max. delays refer to multiple frames per packet calculated as per G.114 A.2.4 [7]. The difference between Max. and Min. is determined by the serialisation delay to line of the codec frames wrapped in layer 2 and layer 3 protocol headers/trailers, and is therefore dependent on the link bit rate, application of QOS, and queuing delays for other session packets already transmitting to line (link congestion characterisation).

(d) Refers to Annexes/Amendments to base standards for operation, and "N" does not preclude use of proprietary forms. Impairments within the E-Model are not generally considered, but if so they are then explicitly included under Codec type.

(e) There is always transcoding tolerance within families where backward compatibility applies, for example AMR-NB to GSM-EFR would negotiate to operate as GSM-EFR. G.726 [17] is transcoding tolerant when synchronously tandemed.

(f) *Ie* for narrow band codecs and *Ie,wb* for wideband codecs (in a monotic listening context) as per ITU-T G.113 [18] Appendices I and IV respectively. For narrow band codecs, *Ie,wb* = *Ie* + 35,8.

(g) Burst Ratio is valid when Bpl≥16, denoted by "Y", and BurstR is stated the Bpl is valid for this value only. Refer ITU-T G.107 [19] par 3.5 & G.113 [18] Appendix I for specific limitations and conditions. For WB codecs, Bpl assumed in diotic listening context. G.107 is currently being updated to provide an improved treatment of packet loss robustness.

(h) Highest achievable R score within an optimal speech channel with no packet loss, and taking no account of the additional delay that is introduced (typically negligible impact for a domestic TDM baseline). Narrowband codecs have a maximum of 93.2, Wideband 129 [4].

(i) Packet Loss Concealment (PLC) improves performance under packet loss conditions and is incorporated in complex codecs by default. For others such as G.711 [2], impairment values may be available with and without PLC, either incorporating the performance into an effective *Ie* value (*Ie-eff*) or through the factors Bpl, Ppl and BurstR as defined in the E-model. See also notes (g),(o),&(p).

(j) G.711 [2] performance when compared to wideband codecs yields an *Ie* that allows use of an expanded R scale, and is shown here for comparative purposes.

(k) G.722.2 [15] is also known as AMR WB for mobile but is not backward compatible with the AMR codec.
(l) G.729.1 [14] is also known as G.729 Annex J and G.729EV and supports backward compatibility modes to G.729/a/b and a new Narrowband bit rate of 12Kbps. Low frequency range is extended to 50Hz. *Ie,wb* values are not ratified and are proposed for diotic (two ears or speakerphone) listening. Algorithmic delay is stated in G.729.1 par 5.6, as 48.9375ms.
(m) G.718 [16] par 5.2 states Wideband algorithmic delay to be 42.875ms.
(n) G.711.1 [21] par 6.5 states Wideband algorithmic delay to be 11.875ms. G.711.1 falls back to G.711 when used in a mixed wideband/narrowband call path.
(o) For some specific cases of number of lost packets "n", percentage Packet Loss "Ppl" and BurstR, an *Ie-eff* (effective *Ie* value) may be directly used in formula 3 of G.107 [19].
(p) PLC type of "Repeat 1/Silence". Refer G.113 [18] Table I.5.
(q) for 10 ms packets
(r) for 20 ms packets
(s) The complexity quoted for G.722.2 is for the highest complexity implementation.
(t) The complexity varies with bit rate, generally being higher for higher bit rates except for the 23.85kbit/s rate which is slightly less complex.
(u) The higher voice frequencies are handled differently in this highest bit rate, from directly transmitted information.
(v) iLBC is specified in the experimental RFC 3951.
(w) Sinusoidal Voice Over Packet Coder is based on quasi-harmonic modeling of the Linear Prediction (LP) residual [9]. Relies on floating point implementation such as provided by a PC rather than on a DSP.
(x) G.718 is designed to be highly robust to frame erasures, enhancing voice quality on IP transport applications. It also has an alternate coding mode, at 12.65kbit/s, which is bitstream interoperable with ITU-T Rec G.722.2, 3GPP AMR-WB, and 3GPP2 VMR-WB mobile wideband codecs. The lookahead numbers in the table include input and output resampling filters and an additional 10ms decoder delay, the overall delay can be reduced by 10ms if the output is limited to Layer 2 for NB input and output.
(y) G.711 App II only describes the CNG payload and does not specify VAD nor DTX.
(z) Estimated figure, same as GSM-EFR since the AMR codec is compatible with the GSM-EFR codec at this bit rate.

Table 2 Basic voice codec features

7 Voice Quality Evaluation

7.1 Mean Opinion Score (MOS)

A commonly used voice quality scale is the Mean Opinion Score (MOS). MOS is a subjective value defined in ITU-T Rec. P.10/G.100 [22], as follows: "*The mean of opinion scores, i.e. of the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material*."

There exist different MOS scales depending on the task undertaken. The most common and known is MOS-LQ for the listening-only context. MOS-TQ applies for talking-only situations. MOS-CQ applies for real conversational quality.

MOS scores can also have different origins:

- subjective tests (e.g.: MOS-LQS from P.800 tests [10]),
- measurement tools or methods (e.g. MOS-LQO with PESQ),
- planning and estimation tools (e.g. MOS-CQE with the E-model).

Measurement methods have to be divided into two families:

- psycho-acoustical models, signal-based ; the most commonly used model of this family is PESQ (ITU-T P.862 [23]),
- parametric models, taking benefit of protocol information ; for IP-based voice, they must comply with ITU-T Rec. P.564 [24].

The audio bandwidth must also be taken into account. Three contexts must be distinguished by adding "N", "W" or "M" to the MOS scale names mentioned above for respectively:

- narrow band only (e.g. MOS-LQON with P.862.1 [25]),
- wide-band only (e.g. MOS-LQSW with P.800 [10] in wide-band context),
- mixed-band (e.g. MOS-LQOM with P.862.2 [26]).

ITU-T Rec. P.10/G.100 [22] gives all details about different MOS scales.

During subjective listening tests, listeners participate in a well balanced, subjective experiment [27], listening to a pre-defined set of sentences, and score the results (their opinions on quality) on a scale of 1 to 5, which are then averaged [10]:

MOS	Classification
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 3 Scale of MOS values.

It is important to note that subjective test results exhibit a variability, ITU Recommendation P.833 [28] states "*Subjective tests, even if carefully designed and carried out under controlled conditions, cannot provide quality ratings which are 100% reproducible under the same conditions. The composition and experience of the test panel, choice of test conditions and stimulus material, test set-up and environment lead to an inherent variability. This variability can also be found in the mean ratings calculated over a large number of individual responses. As a consequence, equipment impairment factors derived from one test will vary to a certain extent if compared to other test data*".

Care is also needed when comparing MOS from different laboratories because MOS is also affected by language and culture, e.g. Japanese MOS tends to be less than that measured in other countries [29]. To minimise such effects, reference conditions (clean speech, MNRU'S) are used.

Subjective tests have historically only applied to narrow band voice, and there is a wealth of MOS data available for most narrow-band codecs. This remains highly relevant because there is a vast embedded base of narrow band telephony (contributed by the existing PSTN) which will co-exist and interwork with IP-based voice networks for many years.

However because IP-based voice networks are not specifically designed for narrow band voice, wideband voice codecs are now coming into use (see sections 6.1 and 6.3) and MOS measurements are also applied to those codecs. Care is needed in designing experiments to be meaningful to both narrow band and wideband codec's as these may be mixed within a network. For example MOS ratings differ between tests according to whether narrowband, mixed narrowband/wideband or only wideband stimuli are presented, as the use of the MOS scale is largely dependent of the stimulus set [4].

The subjective assessment of wideband codecs [19] without any comparative reference of narrowband coding leads to a range of MOS scores similar to narrowband codecs. However when subjective tests are conducted with a mix of wideband and narrowband codecs the narrow band codecs receive lower MOS scores. With such properly designed experiments, wideband voice scores 0.5 to 1 MOS greater than narrow band voice. In such wideband/narrow band voice comparisons, the G.711 PCM codec used as reference gets a MOS-LQSM score between 3.5 and 3.7 in mixed wideband and narrowband codec subjective test experiments, compared to a MOS-LQSN score of 4.4 - 4.5 when listeners are presented with only narrow band voice codecs (customers exposed to wideband speech rate narrow band speech as lower quality).

The additional quality perceived of wideband codecs is particularly important in view of the transcoding impairments presented in section 10.2.

7.2 E-Model – Narrowband Codecs

It is not practical to perform auditory tests during transmission planning. A widely used Transmission Rating Model for representing voice quality is the E-model as defined by the ITU Rec G.107 [19]. ITU Rec. P.834 adds "[*It*] *is the only one [method] recommended by ITU-T for describing the subjective effects of digital processes other than pure PCM on the integral quality for transmission planning purposes*". This model uses transmission impairment factors that represent the effects of modern signal processing devices (including codecs). All impairments modeled are additive (the E-model model being based on psychological factors which on a psychological scale are additive [19]), thus the impairments of transmission segments (e.g. Carrier A, the International Carrier, and Carrier B as well as Service Provider networks) can be added[10] to estimate end-to-end voice quality.

The E-model was developed for the PSTN, thus most development and experience is with narrow band codecs. Extension of the E-model to wideband codecs is given in section 7.7.

The primary output of the E-model is the Rating Factor or R (often called the R-Factor) which is composed of:

$$R = Ro - Is - Id - Ie + A$$

[10] Note however that some impairments such as echo and loudness ratings need to be calculated for the end to end call, while impairments such as delay etc are able to be added for each segment of the call but again are considered in a single calculation for the end to end call.

Ro	Represents in principle the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise.
Is	Is a combination of all impairments which occur more or less simultaneously with the voice signal.
Id	represents the impairments caused by delay
Ie	Effective equipment impairment factor: represents impairments caused by low bit-rate codecs. It also includes impairment due to packet-losses of random distribution
A	Advantage factor: allows for compensation of impairment factors when there are other advantages of access to the user. See [18] Appendix II and section 8.5.

Table 4. Impairments contributing to R-Factor.

The term *Ro* and the *Is, Ie* and *Id* values may be subdivided into further specific impairment values. Further detail is in [19], in [6] and in section 8.

7.3 E-Model Relationship to MOS for Narrow Band Codecs

The R-Factor can be transformed into estimates of customer opinion factors, such as MOS. When estimated from the E-model it is called MOS Communication Quality Estimated or MOS_{CQE} [10], [30]. The following formula for estimating MOS_{CQE} applies to narrow band voice only.

For $R < 0$ $\quad MOS_{CQE} = 1$

For $0 < R < 100$ $\quad MOS_{CQE} = 1 + 0.035R + R(R - 60)(100 - R)\, 7\text{x}10^{-6}$

For $R > 100$ $\quad MOS_{CQE} = 4.5$

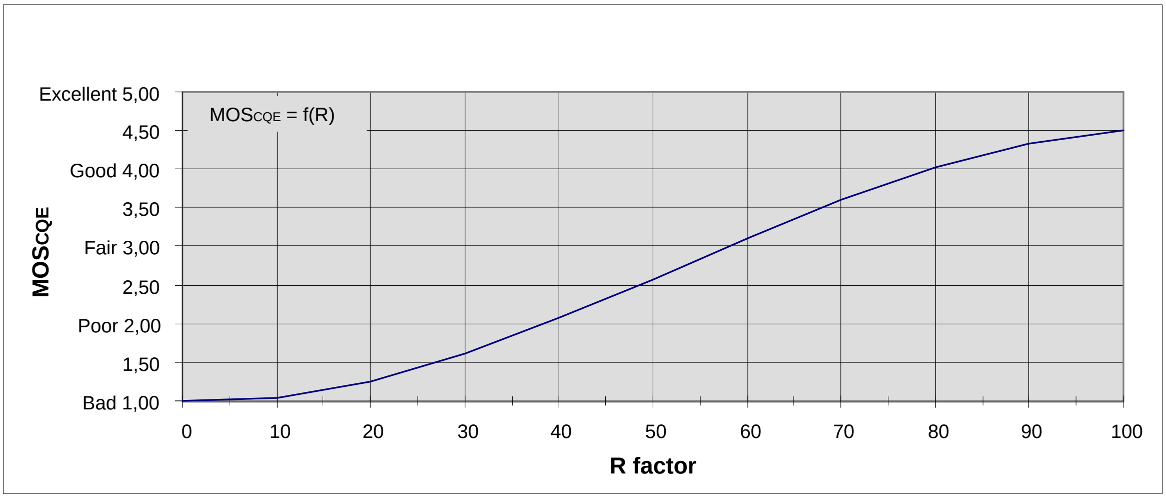


Figure 2 MOS_{CQE} = f(R)

7.4 Transmission Quality Category in the E-model – Narrow Band Codecs

The R-Factor is related to User Satisfaction and to Speech Quality Transmission Category as shown in Figure 3 for homogeneous voice paths containing only narrow band codecs [31]. Customer opinion score estimates, MOS_{CQE}, are also indicated.

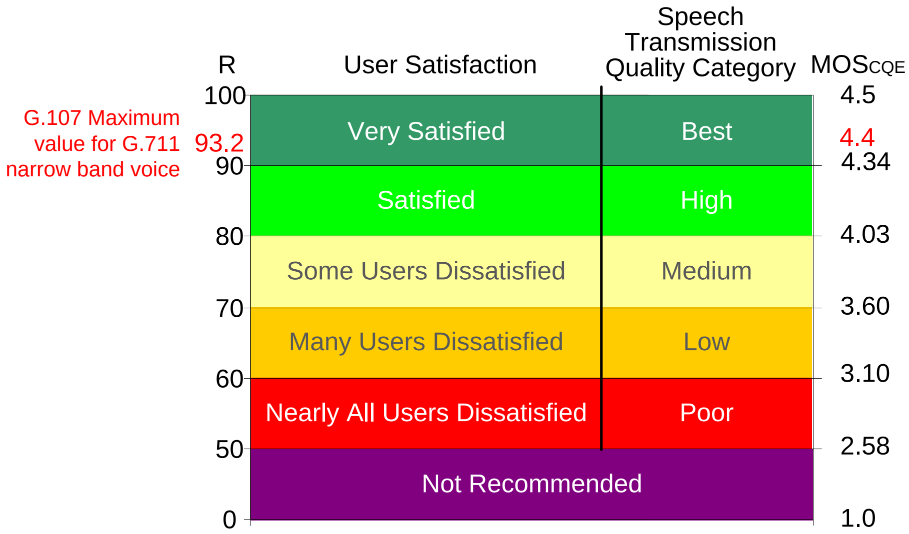


Figure 3 Classification of speech quality for different R-Factors

Note that the classifications in Figure 3 are for convenience only; the range of speech quality is actually a continuum; ref [31] stresses "*It is very important to fully understand the principle …. the R-value is a measure of a quality perception to be expected by the average user when communicating via the connection under consideration: quality is a subjective judgment such that assignments cannot be made to an exact boundary between different ranges of the whole quality scale. Rather, the quantitative terms should be viewed as a continuum of perceived speech transmission quality varying from high quality through medium values to a low quality as illustrated*".

7.5 The R-Factor and Delay – Introducing the E-model Graphical Representation

The delay impairment Id depends on total (end-to-end) latency and R-Factor is often represented on an R-Factor/delay graph. The maximum R-Factor for narrow band voice (G.711 PCM encoded including A/D conversion quantizing distortion) plotted against absolute one-way delay with no other impairments is at Figure 4. Significant latency (>150 ms) is perceived by users as an impairment. All delay, end-to-end (mouth-to-ear), must be included in any estimates of R-Factor. Appendix 1 (section 15) gives data to construct this graph.

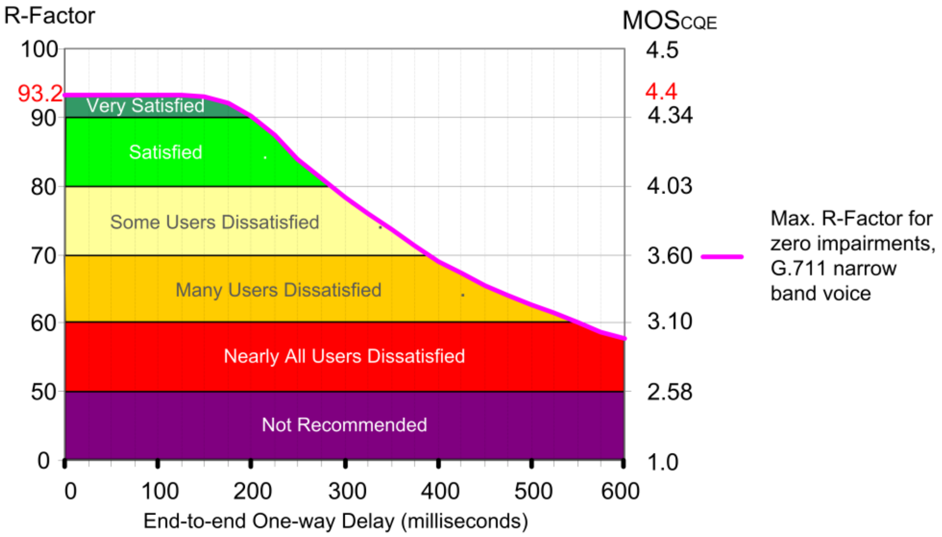


Figure 4 Maximum R-Factor vs absolute one-way delay for narrow band voice

7.6 E - model Limitations as an Estimator of Customer Opinion

Estimation of MOS from the R-factor should be made for transmission planning purposes only and not be fully relied upon for actual customer opinion prediction (akin to telling customers what they think). ITU-T Rec. G.107 [19] pointedly comments as follows: "*It must be emphasized that the primary output from the model is the "Rating Factor" R but this can be transformed to give estimates of customer opinion. Such estimates are only made for transmission planning purposes and not for actual customer opinion prediction (for which there is no agreed-upon model recommended by the ITU-T)*". In practical terms, such estimates nevertheless provide a useful indication of likely customer opinion.

It is also important to note that the E-model is a practical model and caution must be exercised in its use; [10] draws attention to some known conditions and combinations of certain types of impairments where caution should be exercised. Section 9.2 of this paper specifically proposes caution in determining voice quality in the case of transcoding.

In summary, the E-model is a transmission planning tool, and the R-factor a transmission planning rating, while MOS is a customer opinion measure, and derivation of MOS_{CQE} from the E-model R-factor gives an estimate, NOT a MOS customer opinion. Exact alignment of MOS_{CQE} and MOS (be it in listening or conversational context, from subjective tests or objective measures, in either a narrow-band or wide-band context) should not be expected.

For supervision purposes, methods compliant with ITU-T Rec. P.564 [24] must be preferred, even though many measurement tools implement MOS calculations based on the E-model.

7.7 E - model Extension for Wideband Codecs

The E-model described in section 7.2 - 7.6 of this White Paper accounts for narrow band (NB) voice transmission only. The R-factor scale has been extended to support wideband (WB) voice (7kHz audio bandwidth) by extending the R-scale to R=129, [4] in a way which leaves the narrow-band use of the scale unaffected, including the position of the reference connection[11]. This scale extension occurs because, for wideband transmission (50-7,000Hz) the quality is generally[12] judged better than for a narrow band channel [4].

While the extension to the scale is defined, and many provisional measurements made of Ie of wideband codecs for use on this scale (called Ie,wb, see Table 2), the full development of the E-model for wideband transmission is not considered sufficiently stable or complete to present in this White Paper (an ITU-T Study Period 2005-'08 document says "*In this contribution, we will present a new method for calculating impairment factors for WB speech codecs, on the basis of subjective quality judgments. The derived impairment factors are input parameters to a future wideband network planning model, e.g. to a WB version of the E-model which is currently under development in ITU-T SG 12*." [32]). For example, Ro has yet to be defined for wideband noise.

The mixed Wideband/Narrowband scale allows the comparison of end to end voice quality for all narrow band codecs (using the narrowband Ie values) or all wideband codecs (using the Ie,wb values). If transcoding between WB and NB is present in the voice path there is no way to calculate the resulting impairment value. Despite not yet being fully developed, the extended E-model scale is however useful to determine wideband codec impairments because of the Ie,wb, value data already available.

The Ie,wb values are derived from subjective test results and from objective measurements (according to methods described in ITU-T Recs. P.833.1 and P.834.1 respectively). A revised transformation rule between R-Factor for the newer scale, and MOS still needs to be derived (Figure 2 presents the NB transformation rule).

---

[11] The reference connection is the "direct" channel, usually associated with a standard ISDN connection, G.711 codec and other default parameters, resulting in the R-factor of 93.2.

[12] For example, there can be occasions where some noise is more objectionable when listened to in wideband quality.

Of particular relevance to this White Paper is codec behaviour when multiple codecs are used in tandem, so that transcoding occurs (see section 9). In case of NB/WB tandems, there is a strong dependency on the codec order: [32] concludes "*the additivity property for WB speech codec tandems requires further investigation.*"

Readers requiring further information on WB codecs are invited to research the ITU-T Study Groups 12 and 16 materials.

8 Major factors influencing Voice Quality in International Transmission

Major effects on international voice quality (following section 7.2) are

- codec choice (fidelity impairment and associated delay),
- voice bandwidth,
- associated packetisation period (pp),
- packet loss,
- international propagation delay (latency),
- domestic/access (Service Provider) network latency,
- transcoding.

Of the parameters in the E-model (formula in section 7.2), ITU-T Rec. G.108 [6] suggests only the most significant factors be included in normal E-model planning, with the remainder being set to default values (refer to ITU-T Rec. G.108 [6] table 1 and following list, ITU-T Rec. G.108 [6] p19).

8.1 E-Model Parameter Ro

This parameter represents the maximum achievable call quality with other quality degradation factors (Is, Id, Ie,) set to zero, thus representing the basic signal-to-noise ratio. For a call on a TDM network with near zero delay, optimum sender and receiver loudness levels, circuit noise and background noise this is 93.2.[13]

Ro is set to the 93.2 default value when evaluating codec impairments for homogeneous voice paths containing only narrow band codecs.

8.2 E-Model Parameter Is

Is includes factors such as talker loudness, network loudness ratings (speech level changes), side tone, quantisation distortion units (qdu) and echo.

If Ie is used (as in this paper), the qdu impairment is not to be used [6].

The loudness ratings [19], [6] are Service Provider network matters and are generally low/negligible impairment unless the network is set up incorrectly or where significantly different transmit and receive levels are standardised in different national networks. A call will seldom have optimal values for these, particularly when transiting international links and encountering different transmission plans, thus achieving lesser or greater than the optimal 10dB loudness rating. The international and domestic networks, being digital interconnections, do not change the speech level so that Circuit Loudness Rating (CLR) is 0dB. Loudness ratings are set to reference

13 The year 2000 revision of ITU-T G.107 [19] provides an enhanced version of the E-model algorithm (see Annex A). Due to this revision the resulting rating R with all parameter values default has slightly changed (from R = 94.2 to R = 93.2). For practical planning purposes, however, this slight deviation should be considered insignificant.

conditions to evaluate codecs (section 13) but may be included in specific detailed transmission planning.

Echo (as TELR) is also a Service Provider matter but may be a significant impairment if echo cancellation is not to the highest standard. It is set to the G.107 65 dB default value, [19], in this part of this White Paper to allow codec impairments to be gauged, but its influence on the R-factor is shown in section 13.2.5.

8.3 E-Model Parameter *Id*

Id represents all mouth-to-ear delay (latency) impairments. Delay is of utmost significance in international calls, both absolute one-way delay (mouth-to-ear) and the one-way delay of the echo path used in TELR assessment.

8.3.1 Domestic and Access (Service Provider) Network Latency

Domestic TDM access network latencies are typically well within 10 ms one-way and domestic network propagation time to the international gateway is, for most nations, <~10 ms.

Conversion of Service Provider access networks to IP-based voice increases access network latency due to serialisation delay, ADSL/DSL delay (where used) and associated packetisation processing delays including de-jitter buffers on receive [33], interleaving in wireless access networks etc. Further delay can also be introduced where multiple services share the access, and voice packets wait for the serialisation of a packet such as TCP to complete – these delays are limited to a maximum of a single non-voice packet transmission when prioritisation is applied to voice packets, and can be significant as a typical TCP packet is much larger than a voice packet.

Delay in Service Provider and domestic networks must also be obtained (or estimated) to obtain valid E-model codec results. Significant delay factors are indicated in Table 5:

Codec delay	From codec data, choose for appropriate packetisation period
Access network latency (serialization etc)	= 25 ms max. (e.g. DSL connection with interleaving on contributes 4 -16 ms)
Domestic network latency (propagation)	= 12 ms max. (except when >1200km great circle distance from international gateway, when additional allowance is permitted)
De-jitter buffer - receive only (there may be an additional de-jitter buffer if media stream conversion takes place at a domestic network border)	Typically half the maximum PDV, commonly 25 ms for carrier voice networks compliant with ITU-T Rec. Y.1541 [34]
Customer equipment	e.g. DECT (cordless telephones) = 14 ms
Mobile networks	Typically 35 ms

Note: The stated maximum delays are design objectives.

Table 5 Typical access and domestic one-way network latencies.

Note particularly that codec/pp delay occurs in the access network, so care must be exercised in end-to-end transmission planning to not count this twice.

Since the above figures are maxima and typical performance data (except propagation time) is not yet available to the authors, in this paper a planning figure of 30 ms (comprising 20 ms access packetisation / serialisation functions and 10 ms propagation time to the international gateway) is used at the send end, and 50 ms at the receive end (comprising the same factors plus 20 ms. de-jitter buffering). Codec latency is additional and is added to the sending end.

8.3.2 International and long distance network latencies

International network latency is dominated by the propagation time, typically 5μs/km for optical submarine cable systems, [7], Annex A. In estimating international propagation time it is recommended that actual latencies for the particular cable systems be obtained, or, failing that, that great circle distance +14% be used in a latency estimate (because actual cable lengths on the seabed average ~14% longer than great circle distance to ensure a safe seabed path). Geostationary satellite links contribute ~260 ms to one-way propagation time.

Table 6 provides typical propagation delays for international network distances used in the examples in this paper. Associated multiplex equipment delays are comparatively small and may be neglected.

Network Distance	Representative One-way Delay (terrestrial/submarine cable)	Representative One-way Delay (including added satellite hop)
1. Intra Region, e.g Europe	40 ms	300 ms
2. Inter Region, e.g Europe – USA	80 ms	340 ms
3. Global, e.g. Europe - Pacific	160 ms	420 ms

Note: Worst case international transmission latency occurs when satellite transmission is also required, such as might be used into Africa, or many island nations, particularly the Pacific Islands.

Table 6 Representative One-way Propagation Delays in International Networks

High latency causes impairment to conversation. The E-model itself references MOS measurements in subjective testing but is less specific about listening quality (one way) versus conversational quality (two way), although the modeling of delay and the impairment this causes as delay increases beyond approximately 200ms (see Figure 4) suggests conversational quality to be the primary quality measure.

The extended wideband scale appears to offer the promise of path delay mitigation as R>90 is readily achievable with a good quality wideband codec even with delay in the region of 400ms (see Appendix 1 at section 15). However this seems to be intuitively problematic as delays at this level challenge many aspects of a conversational model such as doubletalk and interruption.

Therefore, until a wideband E-model is fully developed and the impact of high levels of delay are better defined, it is advisable to exercise caution in the degree of call quality offset that wideband codecs seem to provide as a compensation for the additional delay introduced by VoIP.

8.3.3 De-Jitter Buffers and Latency

A de-jitter buffer is required at the receive end to remove variations in transmission delay in the packet network (called Packet Delay Variations - PDV or IP packet Delay Variation – IPDV [34]) and jitter from clock asynchronism[14]. This buffer is also called a play-out buffer or just a Jitter buffer.

Packet Delay Variations mainly occur from variations in the queuing delay of packet forwarding in routers, which vary from packet to packet. Voice packets have to wait behind other voice (variation sensitive) packets in the same forwarding queue which arrived earlier, and have also to wait behind data (variation-insensitive) packets in a lower forwarding queue which arrived earlier but have already commenced being forwarded (transmitted) when the voice packets arrive[15].

[14] The de-jitter buffer may also re-order packets which arrive out of sequence.

[15] This serialisation delay is higher on lower speed links.

IPDV is exacerbated by:

- Multiple router hops (more forwarding queues),
- Different packet sizes on shared transmission paths, particularly lower speed paths, (such as if data transmissions with longer data packets share the transmission path with voice sessions),
- Transmission link congestion (higher traffic).

IPDV increases as IP links congest so it is particularly important to adequately dimension interconnecting links according to [1] (as well as ensuring all packet networks involved in a end-to-end high quality voice call operate without undue congestion) to ensure IPDV is not excessive and is within the design size of the de-jitter buffer. Packets arriving when the buffer is full are discarded (buffer overflow). Packets that arrive too late to be played out at the correct time are also discarded (buffer underflow).

ITU Rec. Y.1541 [34] groups services into QoS classes defined according to desired QoS objectives. Classes 0 and 1 would be implemented with DiffServ Expedited Forwarding (EF) per hop behavior [1] which includes carrier IP-based voice services. The IPDV performance objective for IP-based networks supporting these service classes is less than 50ms average.

The de-jitter buffer should be set to at least a desired quantile of the maximum estimated IPDV to avoid excessive packet loss, consistent with achieving design voice quality[16]. This suggests for carrier voice network buffers approaching 50ms are likely to be needed. Jitter buffers should ideally re-order packets which may arrive out of order to avoid discarded packets affecting voice quality (or requiring PLC to be invoked in attempted remedy).

Should indirect public internet interconnections be used [1], voice traffic will be mixed with data traffic on a best endeavours link so that periodic congestion or even overload may be experienced. IPDV's are likely to be higher on calls using such public interconnecting links than in appropriately dimensioned all-private networks, so that increased de-jitter buffer lengths are likely to be needed.

It is also noted that shared radio paths such as will be used in packet mobile radio (LTE) and satellite transmission (see section 12) will require careful optimisation to achieve an adequate balance between radio spectrum cost and congestion leading to voice quality impairing high IPDV values as well as packet loss.

De-jitter buffers are required at VoIP receive ends, which will generally be in the SP network(s) but are also required where de-packetisation occurs for media stream conversion such as transcoding, packetisation period trans-rating (section 9.3) and G.711 companding conversion (section 11)[17]. In all cases the de-jitter buffer needs to be sized to accommodate PDV contributions from the international network(s).

The use of adaptive de-jitter buffers, which adapt the buffering delay to changing network characteristics by using IPDV estimates computed from the arrival characteristics of the voice packets, would alleviate design uncertainty and minimise latency. However discontinuities introduced to the voice play-out as the de-jitter buffer "resizes" may be objectionable to the listener, although the effect of these discontinuities is minimised when adaptive buffers are combined with VAD/DTX coded voice (section 12.2.3) as then the discontinuity can be arranged to change the length of a silence period rendering it inaudible to a listener.

The contribution E-Model parameter latency Id is the average time packets spend in the buffer, not the peak buffer size. If the exact IPDV distribution is not known then half the de-jitter buffer size should be used as the E-Model parameter Id ([33], section 7.2.1.3). If transcoding takes place, ensure that the Id value is included for every de-jitter buffer in the end-to-end call.

---

16 It is recommended [34] that this be based on the 99.9 percentile of the underlying IPDV distribution for the packet flow.

17 In the case of such media stream conversion(s) being implemented in an international/intermediate network, the VoIP call is effectively split into two concatenated VoIP calls interconnected by what is effectively a TDM stream. The media conversion occurs on this TDM stream. Thus the de-jitter buffer at the media conversion VoIP endpoint and at the receiving SP VoIP endpoint may operate on VoIP signals of different packetisation periods and different (codec) bit rates.

8.4 E-Model Parameter *Ie* and *Ie,wb* - Equipment and Codecs

8.4.1 Codec Equipment

8.4.1.1 Narrow Band Codecs

Ie allows for the impairments of codec distortion[18]. *Ie* is by its very definition independent of all the other impairment factors: it is only dependent on the digital process[19] whose perceptual characteristics it aims to model [28]. It also requires care in measurement, as it depends on listening only measurements (MOS) which have not been proven to have the same quantitative psychological degradation as conversational speech, but this is assumed for simplicity [28]. It also suffers from the variability of MOS measurements (see section 7.1).

Of all the impairment factors, *Ie* is the one most likely to deviate from the additivity rule, see section 7.2 (i.e. *Ie* when added together for tandemed codecs may not necessarily give results in exact agreement with listening tests).

Setting *Ie* values for codecs is not a precise science; depending on many MOS measurements plus judgment as to where the codec "fits" with respect to other codecs *Ie* values (resulting *Ie* values should be viewed as being "about right"). *Ie* values are generally assigned provisional values by the ITU-T, which are subsequently changed as modeling and measurement data accumulates and analysis develops[20]. This is important in analysing quality in transcoding configurations, as any residual "about right" *Ie* differences, plus possible non-additivity, compound when codecs are tandemed.

8.4.1.2 Wide Band Codecs

The corresponding impairment factor for wideband codecs for use on the extended E-model scale (with homogeneous wideband codec voice paths only) is called *Ie,wb,* (see section 7.7).

It is important to note that for planning purposes <u>additivity</u> cannot be directly applied to narrowband codecs when represented along with wideband codecs on a wideband scale, and that there are currently no agreed methods to allow a mixed wideband to narrow band transcoded call (see section 9) to be assessed. In practical terms this means that the extended wideband scale can display the results of an E-model calculation for a homogeneous narrowband call scenario, and a homogenous wideband call scenario, <u>but not a mixed wideband/narrowband call scenario</u>.

8.4.2 Packet Loss

Packet loss removes speech samples or frames, increasing *Ie,eff*.[21] Non-waveform codecs perform better than waveform codecs in that the speech synthesis techniques are more robust against missing frames (although the use of inter-frame coding[22] limits the achievable robustness), and because generally (with G.711 say, with typical multi-sample packetisation periods) many speech samples are lost with each missing packet[23] [37], [38].

---

[18] Codecs distort speech, the impairment is a measure of the user perception of its effect.

[19] For non-waveform codecs the encoding process is non-linear.

[20] An example is the *Ie* for G.729, for which the initial provisional value was *Ie* = 15 [40], which then became provisionally *Ie*=12 in the 1998 version of G.107 [41] and was changed to the current value of *Ie* =10 as that data was removed from G.107 to G.113 late in 1998 [20].

[21] *Ie* is a fixed value, depending on codec only. When impacted by packet loss it is called *Ie,eff*

[22] The effect of the loss of one frame can propagate over several consecutive frames.

[23] For example, for G.711/20ms, 160 consecutive samples are lost.

Application layer techniques called Packet Loss Concealment (PLC)[24] are commonly used to mitigate the effect of packet loss for voice media; these use information on the speech signal from either side of the “gap” to interpolate a representation of the missing signal [37], [38].

PLC algorithms also introduce latency and this “PLC Frame” depends on the “gap” size and the amount of information required to estimate and generate the missing packet. Latency added depends on implementation and the particular codec (it is at least 3.75ms for G.711 [39] and worked examples in [34] indicate 10ms allowance is appropriate for general design),

Packet loss impairment is different for each codec, [6] Table 2b, and varies with network load and packetisation period[25] (see Figure 5) thus evading practical planning approaches. As packet loss influence on Ie value is significant it is important to keep it as low as possible. For carrier networks dimensioned adequately and conditioned for voice transmission (e.g. with Expedited Forwarding – EF - Class of Service – COS - at the IP layer and interconnections dimensioned at +15% [1]), packet loss ≤ 0.1% is readily achievable so that packet loss impairment may be neglected[26], see Figure 5 for narrow band codec examples.

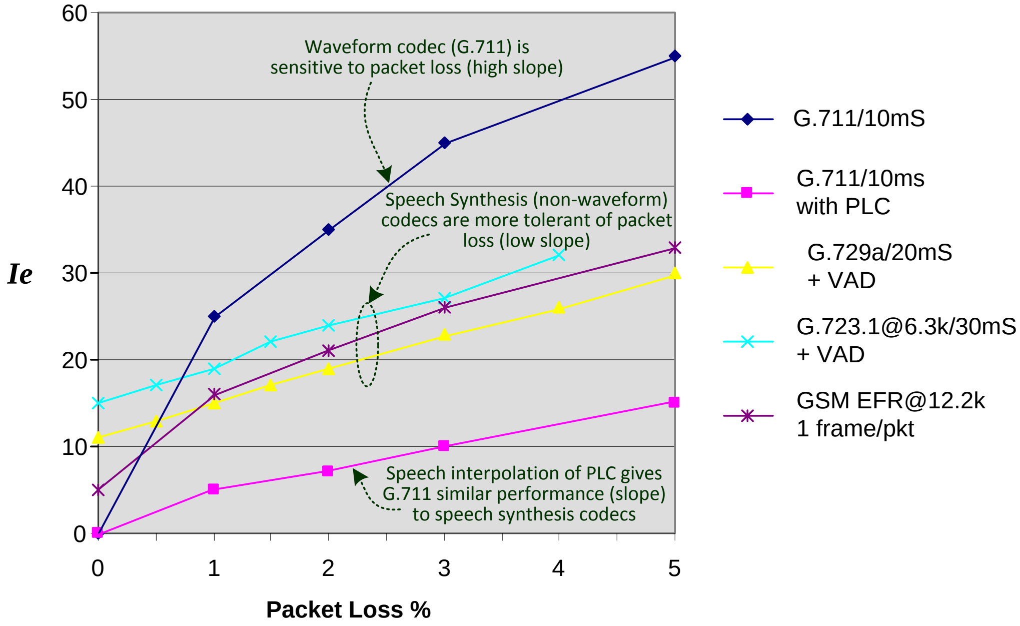


Figure 5 Distortion impairment as a function of packet loss for several narrow band codecs

It is recognized that bursty versus random packet loss also impacts the degree of impairment and these effects have been modeled in the E-model [19]. However there is limited published data and care needs to be exercised when using these parameters as exclusions may apply – refer to Table 2 for known published values.

Wideband codecs may also have associated packet loss Ppl and Bpl factors, however the only published values available are Bpl values for diotic sound presentation (ITU Rec. G.113 [18], Appendix IV).

Recent codecs designed for lossy packet networks (often called frame erasure channels) are more tolerant to packet loss (e.g. see SVOPC in Table 2). These have specific application in internet

24 Packet Loss Concealment (PLC) algorithms are also known as frame erasure concealment algorithms.

25 Increasing packetisation period (pp) means that when a packet is lost, more speech frames are lost, so that higher pp means less tolerance to packet loss.

26 Impairments are generally slight below 0.5% packet loss for low bit rate codecs, [6] table 2b, [38] Table 2 and Figure 5 of this White Paper.

telephony where the transmission channel is “best efforts” and cannot be engineered to the packet loss standards obtainable with carrier networks. These are not included in Figure 5.

8.5 E-Model Parameter - A =Advantage factor

A represents “Advantage of Access” whereby customers may tolerate some decrease in quality (over a “standard” system such as a wired connection) for access advantage e.g. mobility or just being able to talk to hard to get regions. A is very relevant when considering mobile call quality. Examples of A from ITU-T Rec. G.108 [6] section 7.8, and ITU-T Rec. G.107 [19] section 3.6 are in Table 7.

Communication system example	Maximum value of A
Wire-line	0
Mobile in a building	5
Mobile in moving vehicle	10
Hard to reach locations e.g. by several satellite hops	20

Table 7 Examples of Advantage Factor A from G.108 [6]

A = 0 for the IP-based voice fixed network interconnection work of the i3 Forum, but consideration of A = 5 or 10 may be given when serving mobile Service Providers, bearing in mind that, as mobile technology diffuses more into the mainstream[27], A tends to decrease, [20], Appendix II. Table 7 gives absolute upper limits, see [19], section 3.6.

A may also be relevant when low bit rate codecs are used to minimise satellite bandwidth charges (as network operators are likely to invoke such transcoding thus impacting voice quality), although the quality already experienced through long-standing experience with existing TDM satellite connections (coupled with general lack of customer knowledge about what transmission method carries their call) may limit customers tolerance to adverse change should that happen when migrating from TDM to VoIP transmission.

9 Transcoding and the E - model

The remainder of this White Paper predominantly uses NB codecs to illustrate the impact of transcoding due to the limited wideband codec data available. It is important to note however that the principles apply also to wideband codecs (see section 10.4.2 for specific comment on transcoding of wideband codecs)

Transcoding is defined [38], section 6.2.4, as two or more encodings of a signal through different types of non-G.711 codecs, separated by G.711 or 8kHz sample rate linear PCM segments, or in the case of wideband voice, 16kHz sample rate linear PCM. The series use of codecs is also called tandeming in the ITU-T (tandeming admits two or more encodings of a signal through the same type of non-G.711 codec - e.g. G.729 [42] - separated by G.711 or linear PCM segments). The terms are used interchangeably in this paper. Direct conversion between non-G.711 codecs does not occur (although it might be developed in future). When transcoding occurs, particularly for low bit-rate codecs, additional distortion and delay is introduced by each transcoding event.

[27] The “late majority” do not feel they are buying a new service and dilute the “early adopters” who are more accepting of a quality decrease.

In transcoding, voice quality degradations caused by the codecs are cumulative: the degradation is consequently more important for codecs of low intrinsic quality. Transcoding between wideband codecs of higher quality results in lower degradations with resulting WB quality still higher than NB quality.

9.1 Codec Transcoding Issues - General

Low bit rate codecs achieve their lower bit rates by using more complex algorithms that make certain assumptions, such as those about the media (voice, music etc). Other codecs may not make those same assumptions. G.729 is a commonly used fixed network narrow band voice codec with good balance between bandwidth, speech fidelity, and latency[28].

The design requirement of G.729 was that two tandem asynchronous transcodings had to produce a total distortion less that 4 tandem asynchronous transcodings of G.726 [5].

In contrast, G.726 is a simple transcoder (ADPCM) which when decoded to G.711[29], and again encoded to G.726, produces the exact digital signal of the original G.726. Thus if synchronous transcoding of G.726 is used as in a complete digital path with G.711 separating the G.726 instances, any number of transcoding stages to/from G.726 may be used without additional voice quality degradation. Asynchronous transcoding of G.726 would occur if the G.726 instances were separated by a codec other than G.711 (say G.729) and voice quality would then degrade with successive transcoding to/from the G.726 instances.

Transcoding (also known as tandeming) is one of the factors where caution should be exercised in the additivity of the E-model. In particular ITU Rec. P.833 [28], 4.2.1, says '*It is important to check the additivity of the newly derived equipment impairment factor in the framework of other equipment impairment factor values defined so far. If such an additivity check is not performed, the property of a simple summation of equipment impairment factors in order to cater for codec tandems should not be regarded as valid*". Thus, in determining the Ie for newly tested codecs in tandem (including the same codec tandemed, or with different codecs), unless experiments have shown that the summation of Ie values for that particular combination is valid, then it should not be taken as correct. No guidance is given as to whether Ie would be higher or lower than the summation[30], although to eliminate risk, users seeking to apply new codecs for which additivity data is not available, would be wise to examine the impact of the combined Ie being higher for the particular tandem configuration of interest.

These additivity comments also apply to wideband codecs.

For cascaded wideband and narrowband codecs in a voice path, there appears to be additional degradation beyond simple additivity, but more work is needed – this is expected to be part of developing a future mixed wideband/narrowband E-model.

9.2 Codec Transcoding Issues – G.729

The use of G.729 and G.729a is so widespread that this is an important codec family. The i3 Forum carriers, on a basis of a survey, have identified codecs G.729 and G.729a as currently the most popular wire-line (fixed network) low-bit-rate codecs.

Some data is available on the G.729 codecs transcoding performance. P.833 [28] also says "*When equipment impairment factors for non-waveform codecs disregarding transmission errors are determined, the set of 14 reference codec conditions given in Table 1 should be included in the*

---

[28] The lower complexity version G.729a [5] is also commonly used.

[29] Note that this G.711 signal will NOT be identical to the original G.711 because of the bit rate reduction in the G.726 encoding. This is what gives rise to the distortion represented by the Ie of 7, see narrowband codecs in Table 2. It is all subsequent signals to the G.711 encoding standard that are identical if synchronous transcoding is invoked.

[30] No material seen so far indicates it could be lower.

subjective test conditions. This list has been chosen from well-investigated codecs to cover the whole range of Ie values and degradation types." The list in the Table 1 referred to contains several tandemed codec combinations including G.729, whose tandemed Ie values are reproduced here in Table 8:

Codec combination	Ie value
G.729	10
G.729 x 2	20
G.729 x 3	30

Table 8 Ie values for G.729 codec in tandem, without transmission errors, from ITU-T Rec. P.833

This indicates that G.729 in tandem is Ie additive. It is noted that for G.729 Ie was provisionally 12 in the 1998 version of G.107 [41] and was changed to Ie =10 as that data was moved from G.107 to G.113 [20]. Prior to 1998 the provisional value was Ie=15 [40].

The wide gap between the provisional values and the current value may suggest difficulty in deciding what Ie value is "about right", and may suggest care should be exercised in transcoding this codec in marginal configurations.

Transcoding (and possible non-additive behaviour) would not be an issue if (as is possible in an end-to-end IP-based voice call) a single codec was utilised, or, at most, a single transcoding could be implemented (such as if domestic carriers A and B – and the respective Service Providers - use different codecs). However the proliferation of codecs in recent years, the relative absence of data on whether the newer codecs are Ie additive when used in tandem[31], and the inability to signal codec policy end-to-end[32] when multiple carriers are involved in a call, means that multiple transcodings can readily occur. Thus network planners should be vigilant given the significant voice quality impairments that tandemed codecs can cause.

9.3 Packetisation during Transcoding

When transcoding of an IP-based voice signal occurs, the digital IP signal must first be de-packetised to reconstruct the continuous coded digital signal (which introduces de-jitter buffering latency). The recovered continuous signal is then transcoded (decoded to G.711 or 8kHz sample rate linear PCM for narrow band codecs, and 16kHz sample rate linear PCM for wideband codecs, and re-encoded), then re-packetised, incurring an additional packetisation latency. Thus latency compounds if multiple transcodings occur.

> Note that the term transcoding strictly refers to the conversion of a continuous digital signal from one codec to another. Packetisation is an additional function. Sometimes the two are erroneously combined: this should be avoided as the two functions are separate, can be implemented separately and in different parts of the hardware. Changing the packetisation period only is sometimes called transrating.

[31] Extensive testing in tandem configurations to ITU Rec. P.834 [43] is expensive, and if done may not be completed for some time after codec release. In the case of codecs destined for internet telephony, this may never be done since tandeming is not contemplated in the intended use.

[32] This could tell intermediate networks what the codecs at each end are, so that transcoding could be minimised.

9.4 Mobile Transcoding

Mobile SP's routinely transcode mobile-mobile calls within their network as the dynamically variable bit rate codecs may not match for Caller A and Caller B due to differing radio path conditions to/from their respective base stations. When transcoding takes place between two bit rates of the same codec (such as within a mobile SP's coverage area), voice quality impairments are less that would occur to a different codec type, and in any event, voice quality becomes an optimisation of codec bit rate and other error correction mechanisms used on the radio path.

If the mobile call traverses an intermediate network, a standard interconnection bit rate of the mobile codec is used[33] and Tandem Free Operation (TFO) is often invoked, which means the particular codec digital signal is "tunneled" through the 64K channel of the TDM switches and intermediate network(s)[34] without transcoding impairment.

IP-based mobile networks will increasingly support Transcoder Free Operation (TrFO) where there is no tunneling involved and the radio access networks exchange IP packets – it can be expected that fixed networks will support this form of mobile connection in time.

Modern mobile codecs such as the AMR family are inherently high quality. However radio path variance leads both to packet loss and a trade off of codec bandwidth to error correction as radio carrier-to-interference varies. As a result the *effective* impairment attributed to mobile codecs is significantly worse than that of codecs within fixed networks, these *effective* impairments may rise to over Ie,eff = 30, [6] Table 2c, (although such high values usually exist for a short time only).

Mobile SP's currently use G.711 interconnections to extend TFO through intermediate networks, which requires a higher bandwidth international interconnection. It will be unlikely that G.711 will be extensively used in future IP-based international networks for cost reasons. When IP-based voice is introduced into mobile networks similar considerations of codec use and transcoding as discussed in this White Paper will occur, however a better target would be to utilise TrFO mechanisms for fixed IP network interworking with mobile networks as described in RFC3267 [44] and RFC4348 [45] for AMR WB and VMR WB codecs respectively.

Currently fixed network SP's and mobile network SP's use different families of codecs, thus mandating a transcoding event in every mobile – fixed call. Unifying the codecs across mobile and fixed SPs would raise the quality of fixed-mobile calls. For 3GPP technologies[35] this would require fixed SPs to support AMR-WB/G.722.2 in SIP phones, Residential Gateways and cordless devices[36] that today only mandate G.722 for wideband operation.

10 Impact of Transcoding using E-model – Illustrated for Narrow Band Codecs

10.1 Single codec

An example of a high level estimate of the R-Factor derived by considering the contribution of the above factors on narrow band international call quality is given in Figure 6. The methodology used

---

33 For the AMR codec and the GSM-EFR codec this is 12.22 kbit/s, see Table 2

34 The means the codec signal is made to look like G.711 (i.e. a 64kbit/s data signal containing the data that is the mobile codec coded signal instead of being a PCM voice signal) so the intermediate switches and network(s) handle in the usual way, except transcoding impairments do not occur.

35 3GPP is the most widely deployed mobile technology, with 3G and support for AMR-WB codecs beginning to ship in handsets. With handset turnover on average every 2 years it can be expected that there will be significant volumes of AMR-WB capable handsets deployed within the next 5 years.

36 DECT is one of the most common cordless technologies and the New Generation DECT standard that supports IP connectivity mandates for wideband operation G.722 (and G.729.1 as an optional codec) but does not include AMR-WB/G.722.2 in the optional codec list (see section 10.4 for comment on needing to account for transcoding in cordless handsets in an end-to-end connection).

is to derive the R-Factor vs latency curve for the codec, then the R-Factor sought is the intersection of this curve and the total end-to-end latency.

The parameters chosen are:

- Domestic/Access (Service Provider) Network latency of 30 ms send, 50 ms receive;
- Codec/pp G.729/20 ms, impairment Ie =10, latency = 35 ms ;
- No transcoding;
- Latency of four typical network distances from Table 6.

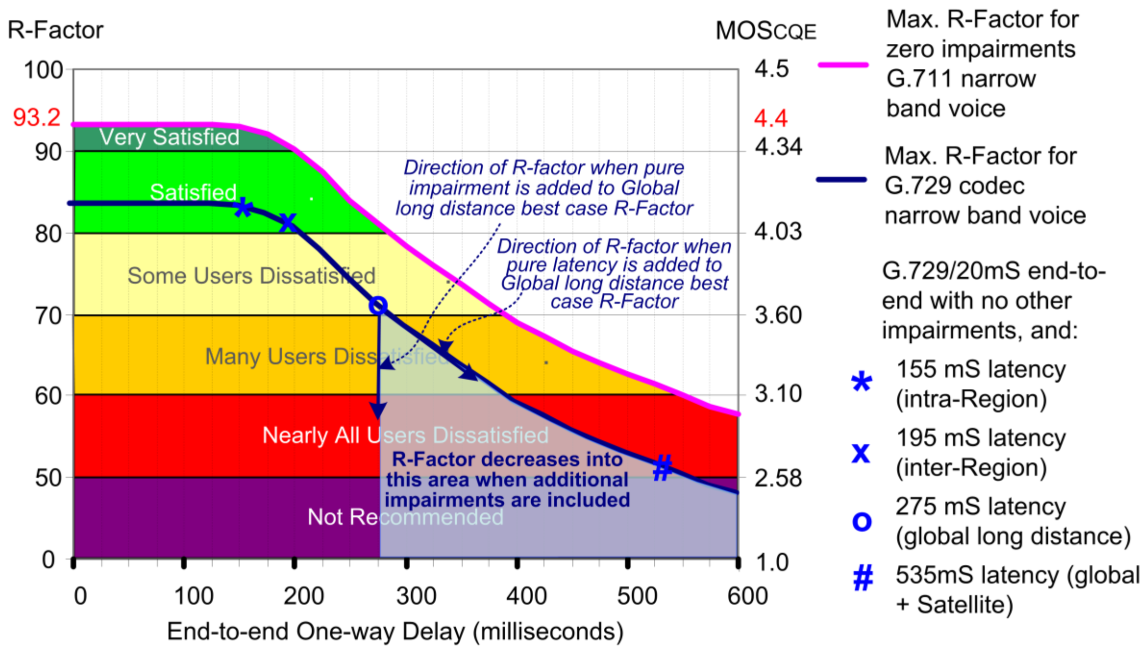


Figure 6 Best Case R-factor for international voice call with G.729 and several call distances

It is important to note that when other impairments are added to this best case estimate, the R-Factor can only decrease and (for the example of a global long distance call) moves into the shaded zone on Figure 6, i.e. the quality can be no higher than indicated by the shaded area after accounting for the speech processing effect of the codec(s) and the transmission delay.

Figure 6 indicates that codec impairments, IP-based voice latency, and international distance latency are important design parameters in international IP-based voice networks, particularly for calls requiring satellite. Appealing to the Advantage Factor is invalid as this is transfer of an existing fixed service to a replacement platform, not requested nor understood by customers.

Since G.729 is coded from PCM[37], this result applies to all network situations in Table 9. Although all but line 2 of the table does suggest that transcoding is taking place, because G.711 is the base for G.729 when coded initially, the G.711/G.729 conversion point in the examples is simply being shifted along the transmission path so that there are no additional impairments normally associated with transcoding. This illustrates that care should be taken in assessing the codec transitions along the entire call path to correctly determine the Ie values to apply.

[37] The G.711 signal (8 bit companded PCM) is converted to 16 bit linear PCM (sometimes called uniform PCM) for the input to the encoder. The reverse happens at the output of the decoder [42].

Service Provider A	Carrier A Domestic	International Network	Carrier B Domestic	Service Provider B
G.711	G.711	G.729	G.711	G.711
G.729	G.729	G.729	G.729	G.729
G.711	G.711	G.729	G.729	G.729
G.729	G.729	G.729	G.711	G.711

Notes:

1. Networks with G.711 internationally, although valid, have been eliminated from this table as they have a bandwidth (cost) disadvantage and would be unlikely to be used when G.729 is available in one of the domestic carriers networks.
2. Service Providers assumed to use same codec as domestic network operators.

Table 9 Network situations applicable to Figure 6

10.2 Transcoding – Illustrated with Narrow Band Codecs

The E-model represents transcoding by summing the Ie of the particular codecs concerned, without regard to order. Codec order is acknowledged to affect voice quality for low-bit-rate-codecs [19] but the effect is known to be small, and is disregarded to preserve the simple additive nature of the E-model [19].

To illustrate why transcoding should be avoided, best case estimates of the R-Factor for trancoded calls are modeled in Figure 7. The parameters chosen are:

- Domestic/Access (Service Provider) Network latency of 30 ms send, 50 ms receive;
- Codec/pp G.729/20 ms, impairment Ie =10, latency = 35 ms;
- Transcoded to G.723.1 @ 6.3kbit/s/30 ms [46], additional impairment Ie = 15, plus 30 ms de-jitter buffer plus 67.5 ms additional codec/pp processing ;
- Latency of four typical network distances from Table 6.

Codec impairments are taken as the E-model simple summation of Ie factors.

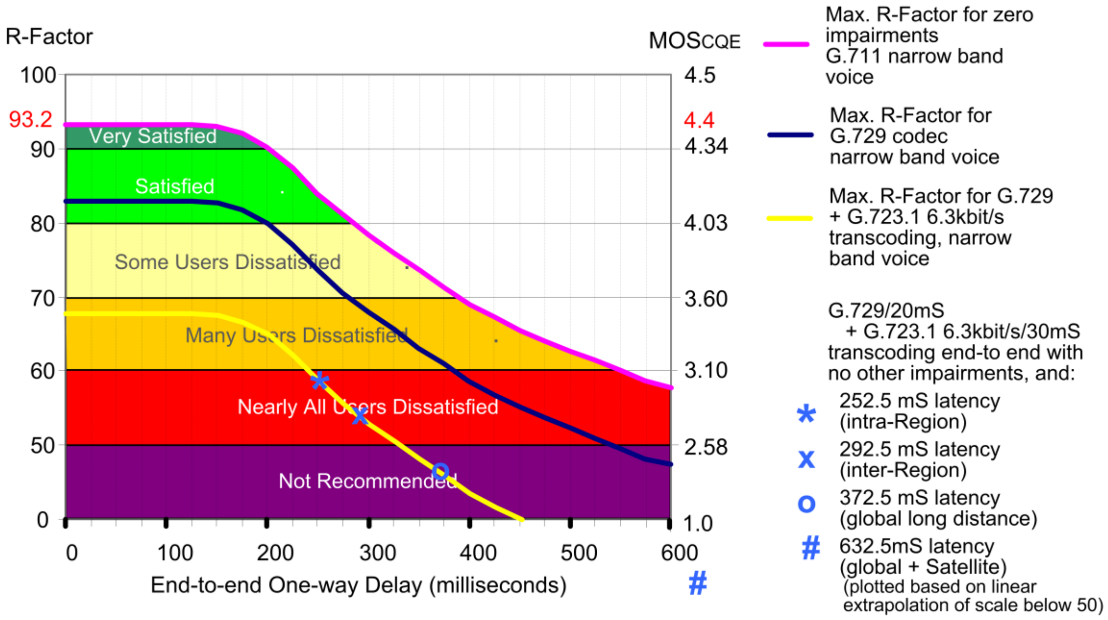
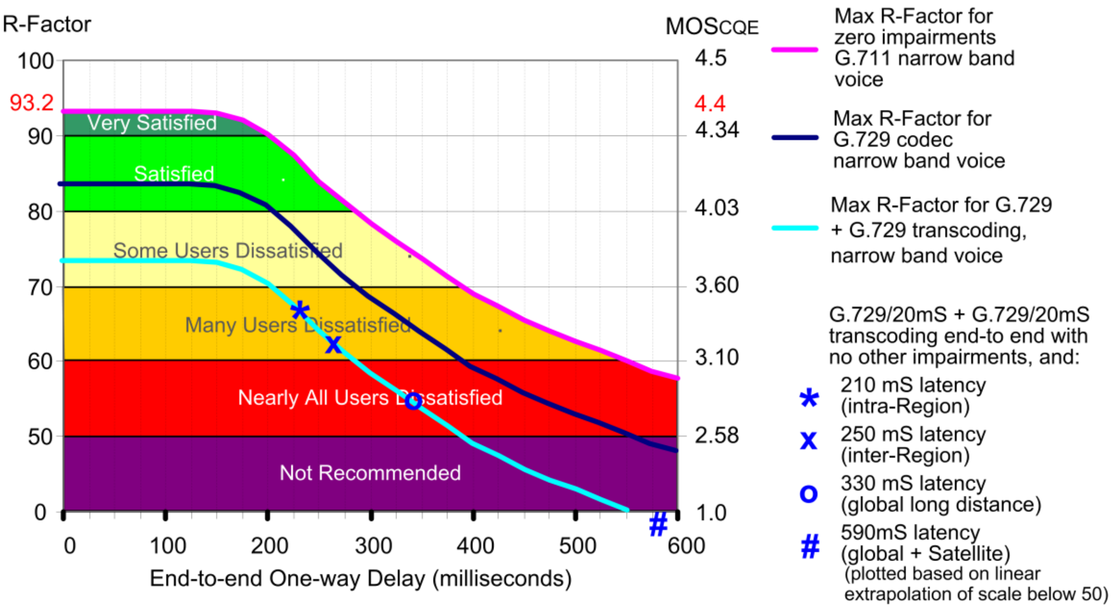


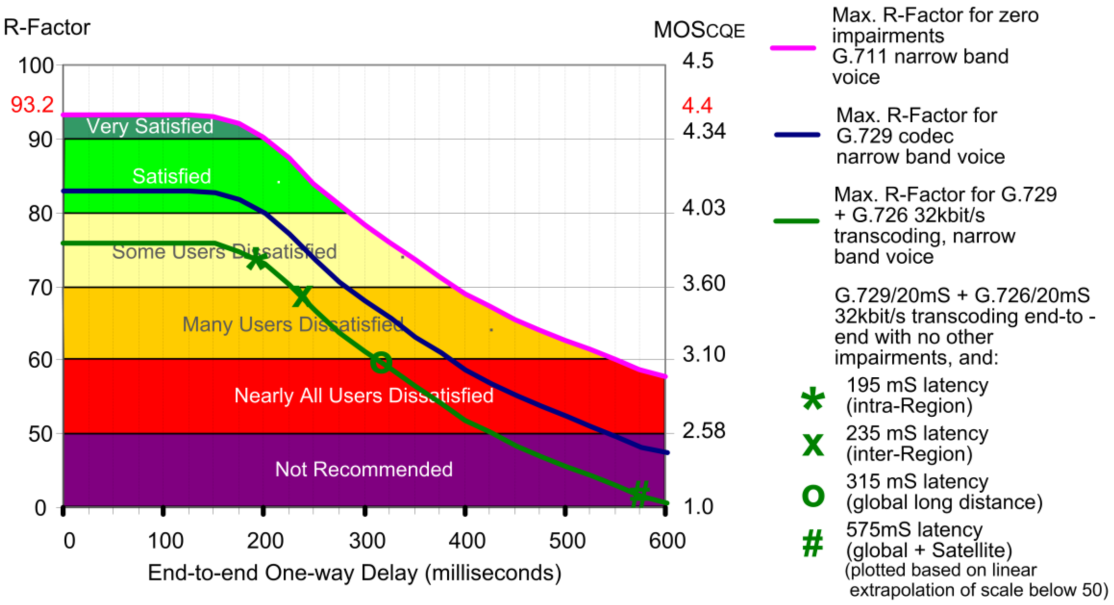
Figure 7 Illustrating the high impact on voice quality of adding a transcoding to G.723.1 to the Best Case R-factor for the international voice calls of Figure 6.

This illustration applies to a direct bilateral configuration where Domestic Operator A and Domestic Operator B do not use the same codec (specifically A uses G.729 and B uses G.723.1), or when an intermediate carrier transcodes to G.723.1 to save bandwidth in an otherwise all G.729 configuration (or any other configuration represented by the same end-to-end codec transitions).

The severe impact on call quality of this transcoding is plainly evident, and the high intrinsic impairment of the G.723.1 codec renders it unsuitable for use in international voice calls involving interconnected IP-based voice networks because voice quality is profoundly affected.



(a)



(b)

Figure 8 Different transcoding configurations, impact on Best Case R-Factor.
(a) G.729/20 ms x G.729/20 ms (b) G.729/20 ms x G.726@32kbit/s/20 ms

Two different codecs used for the same transcoding configuration are presented in Figure 8. Figure 8(a) shows a double transcoding from G.729/20 ms to G.729/20 ms and Figure 8(b) from G.729/20 ms to G.726 @ 32kbit/s/20 ms.

Using codecs with lower delay and lower Ie impairment value is seen to improve end-to-end quality compared to Figure 7. Note also that quality would decrease further if the packetisation period was lowered to 10 ms, the saving of up to 20 ms (Table 1), being equivalent to ≈ 2 R-factor units.

10.3 Comparison with TDM

To illustrate the importance of careful engineering of IP-based voice networks, and the severe impacts of transcoding low-bit rate codecs, the following example of R-factors in corresponding TDM networks is given.

The parameters chosen are:

- Domestic/Access (Service Provider) Network latency of 15 ms each end;
- Codec/pp G.711, impairment Ie =0, latency = 0.125 ms;
- Transcoding in Digital Circuit Multiplication Equipment (DCME), G.728 codec @ 16kbit/s with VAD [47], Ie =7, latency =15 ms (VAD dominates latency, codec contribution is 1.25 ms [7]);
- Latency of four typical network distances from Table 6 (with the global + satellite distance assumed to a small island country, domestic latency ~0 ms);
- Plus in addition, Global + satellite distance from Table 6 to small island country, domestic latency ~0 ms, with two stages of DCME (i.e. transcoding).

The DCME assumed here is the highest impairment type predominantly used in TDM networks (using the G.728 codec, Ie=7 [7]). Other predominant DCME types used the G.726 @ 32kbit/s ADPCM codec [17] (also Ie=7), which, when the DCME is not highly loaded so that temporary bit robbing reduces the codec bit rate[38], would have lower transcoding impairments because of the synchronous transcoding advantages of that codec. This example thus is not best case, but is realistic and typical of long distance international TDM interconnections.

[38] Bit robbing is a technique used in DCME whereby if another speech channel is needed by a new talker while other talkers are occupying all available talker transmit channels, several other talkers are temporarily assigned a lower bit rate (by dynamically invoking a lower bit rate version of the same codec, often developed for this purpose) and the released bits are then used to create an additional, temporary, talker transmit channel. Immediately the talker peak abates, all transmit channels are returned to full bit rate. This "graceful overload" mechanism avoids clipping speech until a transmit channel would otherwise become available.

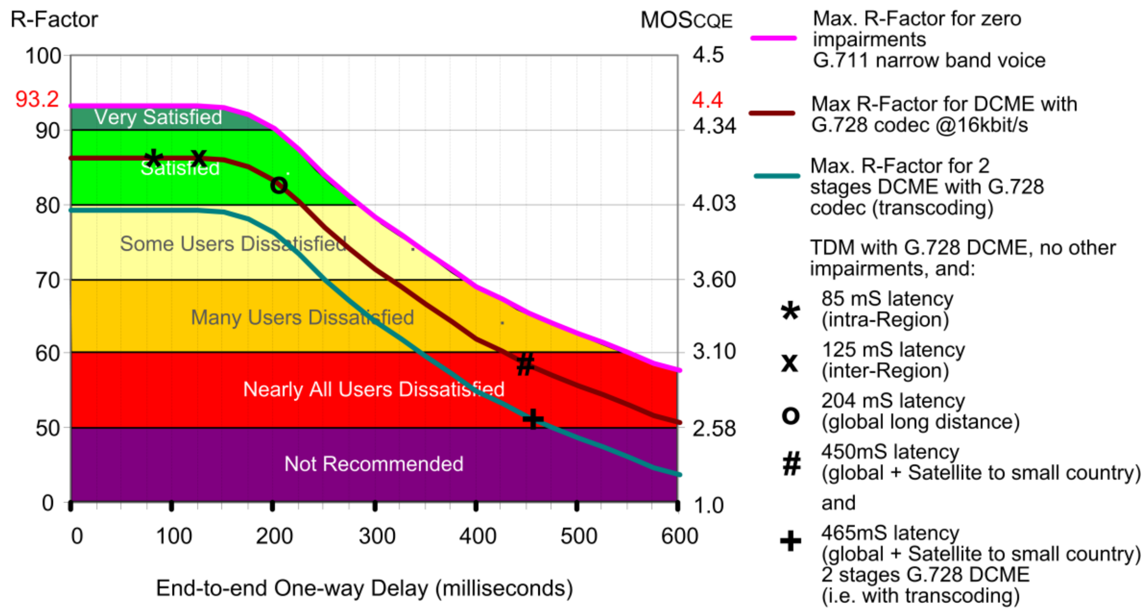


Figure 9 Best Case R-factor for international voice calls on TDM network with DCME using G.728 codec with VAD, and several call distances, plus one example of two stages of DCME (transcoding)

The impacts on voice call quality of migrating the PSTN to IP-based voice networks, particularly when transcoding is necessary is clearly evident when Figure 6 through Figure 8 are contrasted with Figure 9.

10.4 Transcoding – Observations

10.4.1 Narrow Band Codecs

Given the popularity of G.729, it would not normally be expected that multiple transcodings of G.729 would occur as many fixed carriers can be expected to support it, at least until wideband codecs supersede it (it is a mandatory codec for compliance with i3 Forum recommendations [1]). If any other codec (e.g. the fall-back codec G.711) is invoked in an intermediate network between two G.729 networks, a multiple G.729 coding, with adverse call quality, would occur.

It is important to note that the results above show that international call quality will be adversely affected even in the best possible configuration of a direct bilateral connection if the two countries domestic networks use different low bit rate codecs with the resulting voice quality being greatly dependent on the particular codecs. This illustrates the importance of end-to-end voice quality planning involving all carriers and Service Providers in the configuration, as is done for direct bilateral networks.

Further, it can be readily understood that any additional transcoding is likely to lead to unacceptable voice quality, such as if mobile Service Providers interconnect with domestic fixed operators (at a transcoded interface) and the international call is passed through an inappropriately transcoded international configuration. There may also be situations where digitised voice used in customer service platforms (e.g. voicemail) may be encoded so that additional transcoding is necessary to access them internationally. Such Service Provider situations should ideally be appropriately voice engineered in conjunction with the Domestic Network operator.

Another example of additional transcoding is cordless handsets. These are now commonly used by customers of fixed networks and may further complicate call quality by introducing an additional (asynchronous) transcoding into the mouth-to-ear call path. G.726 is used in current generation

DECT handsets (Ie =7, air-path delay = 14 ms). While not a current problem given the PSTN's impairment tolerance (see section 10.3), it is readily seen that introducing one (or two) additional such transcoding steps into IP-based voice networks could create an intolerable result for Service Providers customers. It will be interesting to monitor as VoIP migration occurs the quality level changes that customers are willing to accept to preserve mobility in their homes with terminals they have already used satisfactorily with the TDM PSTN.

Since there are many codecs available and these are generally chosen by Service Providers and Domestic Operators over which the International Carrier has limited influence, transcoding will not be completely avoidable. A calculation method for completing the analysis by adding other impairments is also given in section 13.2.4. It is recommended that Carriers undertake complete analysis for each situation as there may be other significant impairments to consider in individual cases: this paper generally focuses only those which are typical of all international connections with particular focus on codecs.

Reference [37] came to a strong conclusion: "*transcoding should be avoided at all cost*". No evidence has been found during researching this White Paper to indicate that this statement is any less correct than when it was written in 2000. It is stressed that, for interconnected IP-based voice networks, some instances of transcoding will be inevitable, so that wherever possible, compensating (low impairment) choices should be made in transcoded networks (domestic Network Operators cooperation in this would be needed during international bilateral negotiations).

If, in detailed analysis, transcoding impairments are indicated to be severe and unacceptable, it is recommended that different network arrangements be sought. This may necessitate different commercial and different carrier relationships be implemented.

The use of G.729 and G.729a is wide spread and compatibility within that codec family exists[39].

10.4.2 Wide Band Codecs

Wideband codecs, by virtue of the greater speech quality, offer more headroom above the level at which customers deem a call unacceptable, and thus their use (provided wideband codecs are used end-to-end) is likely to reduce the impact of transcoding.

As an example, take a typical transcode from fixed to mobile with a WB call, that is G.722/64Kbps (Ie,wb=13) in the fixed network, transcoded to G.722.2/12.65Kbps (Ie,wb=13) in the mobile network. Assuming no Frame Error Rate on the radio path, there is a total of 26 R units impairment incurred due to the transcode resulting in an R-Factor (extended scale) of 103, higher than the narrow band reference G.711 codec of 93.2 R units impairment on the extended scale.

A further observation is that mobile handsets supporting wideband codecs connecting over mobile networks operating TrFO will exceed the call quality of the best narrow band fixed line calls today, leading some domestic fixed line carriers to consider supporting mobile codecs, particularly AMR-WB or G.722.2 [48], in fixed Next Generation Networks to maximise call quality outcomes.

10.5 Unsuitability of G.723.1 Codec in International Carrier Networks

The G.723.1 codec [46], for a small (~3kbit/s) encoded bandwidth (transmission cost) reduction compared to G.729, has such high distortion and latency (see Table 2) due to the low frame rate and low encoded bandwidth that it should not be deployed in IP-based international telecommunications networks, and NEVER transcoded when other low bit rate codecs are also in the network configuration (see Figure 7). The only possible application this codec could have is if the bandwidth (cost) was an overwhelming factor for a special link, and then G.711 should be used as compensation in the remainder of the network.

It is suggested that the i3 Forum carriers take every opportunity to eradicate this codec from general use in IP-based voice international networks.

---

39 Such compatibility includes a reduction in, or elimination of, transcoding impairments.

10.6 Mixed Narrow Band and Wideband Codecs in a Voice Path

There are two scenarios in which mixed wideband and narrowband calls could occur:

- when a call changes codec type during the call or when the bit rate of a scalable codec is changed, for example from wideband to narrow band[40]or vice versa , and
- where there is transcoding between wideband and narrowband.

For the former case there is active research into the impacts but no publicly available published material is available. Frequent switching between NB and WB during a call could have negative impact on the perceived quality and should be limited (customers are likely to judge the resultant call as worse than if constant NB frequencies only were presented).

For the latter case the resulting quality can be seen as the mean of the quality of both legs of the call (however at the current state of research, this is based on very restricted amount of test results, in a narrow scope of application [49]).

11 A-Law/μ-Law Companding Conversion for G.711 PCM Codec

The G.711 codec comes in two versions, differing only by a companding function which follows the μ-law recommendation [2] in North America[41] and Japan and the A-law recommendation [2] in all other countries. Companding is utilised to allow the desired voice signal-to-noise ratio (which would require 13 or 14 bits in "uniform", "uncompressed" or "linear" PCM) to be accommodated in only 8 bits ([2], section 3.6).

Quality voice calls cannot be operated using the G.711 PCM codec without companding matching at both encoder and decoder. Companding conversion[42] responsibility lies, by international agreement [2], with the μ-law countries, and traditionally, in a TDM environment, the international carrier at the border of the μ-Law countries has taken the conversion responsibility. For IP-based voice however the responsibility may not be so obvious. ITU-T Rec. Y.1541 [34] states "*IP connectivity spans international boundaries, but does not follow circuit switched conventions (e.g., there may not be identifiable gateways at an international boundary if the same network section is used on both sides of the boundary)*", so that in voice path engineering of IP-based voice networks, care should be taken not to overlook this requirement as G.711 companding conversion has to be specifically included by a μ-Law country in G.711 encoded calls between A-Law and μ-Law countries.

The G.711 signal is packetised to construct the VoIP media stream. Thus the companding conversion for VoIP signals typically requires the same media stream conversion process as transcoding, i.e. de-packetising to a continuous digital signal, converting the companding law, then re-packetising. Note that this is NOT transcoding as no conversion to a different coding algorithm takes place: only the PCM amplitude levels are corrected to remove distortion that would otherwise occur at the receiving end if the digital representation of the companding law did not match that expected by the receiver. Thus no voice quality impairment results from a G.711 companding conversion (it is an $Ie = 0$ conversion). However the added latency of the de-packetisation and re-packetisation will contribute to voice quality impairment (e.g. this adds an additional de-jitter buffer of latency – section 8.3.3).

In IP-based voice network soft switches G.711 μ-Law and G.711 A-Law are separate codecs in negotiation lists and thus, unless media conversion is invoked, call negotiation between networks using differently companded G.711 VoIP signals will not succeed. Since G.711 is a very common codec, it is usually supported in all soft switch codec negotiation lists to ensure codec negotiation

[40] This might occur when recorded messages are accessed during a call session.

[41] This includes all the North American (E.164) Numbering Plan countries.

[42] The conversion tables are given in Table 3/G.711 and Table 4/G.711 of [2]).

will succeed. The requirement to undertake companding conversion for G.711 encoded calls to/from North America and Japan has the following implications:

1. media stream conversion capability is required, and
2. media stream conversion breaks the RTP and RTCP stream continuity, potentially limiting the usefulness of RTCP based QoS measurements between North America and Japan and other countries.

In respect of 2 above, since all 8 bit samples of the packetised PCM signal are independent, and companding conversion does not impair quality, an alternative method of conversion could be to:

- modify all PCM samples within the packet "on the fly" at wire speed;
- change the payload type in the RTP header, PT=8 (PCMA) to/from PT=0 (PCMU).

The RTCP header does not contain any information that requires changing as delay, jitter and loss do not depend on the codec, thus no adjustment is required to the VoIP metrics as Ie is identical for both A-Law and µ-Law at zero. If this method were qualified and adopted widely, the utility of RTCP based VoIP QoS metrics could be preserved in situations where transcoding is being avoided, but companding conversion is unavoidable.

12 Codec and VoIP Transmission Considerations for High Cost Bandwidth Links such as Satellite

Satellite links are often the only way to extend international communications to remote areas/countries. The transmission costs of satellite links are generally considerably higher than modern, high capacity submarine cable links, and voice compression using DCME is commonly used in TDM satellite links to minimise bandwidth costs[43]. Satellite bandwidth costs become even more important for VoIP transmission, driving the use of low bit rate codecs and other techniques to compensate for the packet overheads.

Satellite transmission costs are directly proportional to transmitted (occupied) bandwidth. Satellite bandwidth is still commonly sold as SDH, and the most common bearer size is E1[44, 45]. For VoIP transmission over SDH (or SONET), IP datagrams are mapped into SONET or SDH payloads using Packet Over Sonet (POS) technology. The IP datagram is encapsulated in Point-to-Point Protocol (PPP) packets [50] with framing information supplied by the High Level Data link Control (HDLC) protocol [51]. Gaps between frames are filled and are then mapped synchronously, octet by octet, into the SONET or SDH frame [52]. The IP/UDP/RTP headers of the (layer 3) IP datagram occupy 40 bytes for IPv4 and 60 bytes for IPv6, and the POS mapping into a layer 2 synchronous digital signal adds another 6 bytes.

---

[43] There is a well established tradition of transcoding to lower bit rates for satellite transmission. DCME is almost exclusively used on current TDM voice satellite links to conserve bandwidth, and, with the G.726/32kbit/s codec combined with proprietary VAD/DTX/CNG, average voice transmission bit rates of 16kbit/s are attained (called 4:1 voice compression gain). With the G.728/16kbit/s codec, average voice transmission rates of 8kbit/s are attained (i.e. 8:1 voice compression gain). DCME using the G.729/8kbit/s codec was introduced (average voice transmission rate ~4kbit/s –16:1 voice compression gain) but did not come into widespread use. Note that the VAD/DTX generally used in the most commonly used DCME type is not the codec associated VAD/DTX referred to in section 6.6, but a proprietary implementation (including CNG) separate from the codec (thus operating with any codec).

[44] This is a convenient size since most TDM satellite voice links are thin routes. Also, DCME has E1 interfaces.

[45] In some countries (e.g. USA) the similar SONET technology is used.

The main factors influencing the transmitted bandwidth of VoIP signals are[46]:

- codec bit rate (lower bit rates lower the speech information bandwidth),
- VAD/DTX (roughly halves the average speech information bandwidth by not transmitting silence),
- packetisation period (longer packetisation periods increase the size of the speech information payload relative to a given packet header size, meaning less header information is transmitted for a given speech information content (payload), and
- IP/UDP/RTP compression[47] (greatly reduces the size of the IP/UDP/RTP headers by exploiting constancy of differences between headers of a particular stream, therefore the header can, for most packets, be calculated and re-inserted at the receiving end).

One of these factors (codec bit rate) also has a strong effect on voice quality, per the Ie factor (sections 8.4 and 12.2).

12.1 Factors Predominantly Affecting Transmission Bandwidth

12.1.1 Codec Bit Rate

Using codecs of low bit rate decreases the payload size, thus decreasing transmitted bandwidth. There is however a tradeoff with reduced voice quality, see sections 12.2.1 and 12.3.

12.1.2 Packetisation Period and VAD/DTX

Figure 10 shows the occupied bandwidth on an SDH bearer for a G.729 coded VoIP signal for IPv4, four different packetisation periods and with/without Discontinuous Transmission (DTX)[48] (enabled by Voice Activity Detection - VAD), contrasted with the typical occupied bandwidth of the same voice channel in existing TDM networks[49]. The speech activity is taken as 50%. Note that:

1. the higher occupied bandwidth of G.729/10mS would encourage satellite network operators to translate pp to higher values, such as G.729/40ms,
2. the significant reduction in occupied bandwidth when VAD/DTX is also used, and
3. the occupied bandwidth is generally higher than in TDM networks equipped with DCME, despite the voice being coded at 8kbit/s rather than 16kbit/s or 32kbit/s as commonly used in DCME.

[46] A method of calculating the occupied bandwidth accounting for IP/UDP/RTP headers and DTX/CNG can be found in [57], section II.3.2. RTP compression (see later in section 12.1.4) may be accommodated by substituting 2 bytes for the 40 bytes of IP/UDP/RTP header, and POS by adding 6 bytes.

[47] Sometimes referred to just as RTP compression.

[48] Discontinuous transmission refers to sending speech information only when the talker is talking, the idle periods contain little information (e.g. comfort noise characterisation updates only). Thus the sending channels can accommodate more talkers on average.

[49] The comparison in the figures includes the use of G.729 equipped DCME which did not achieve the high penetration into TDM networks as did the those equipped with G.726 and G.728 (DCME growth was stifled by the advent of VoIP). It is however very useful for direct comparison here with G.729 in VoIP transmission, but readers should note that the transmission rates most commonly found in TDM networks are provided by the G.726 and G.728 equipped versions of DCME.

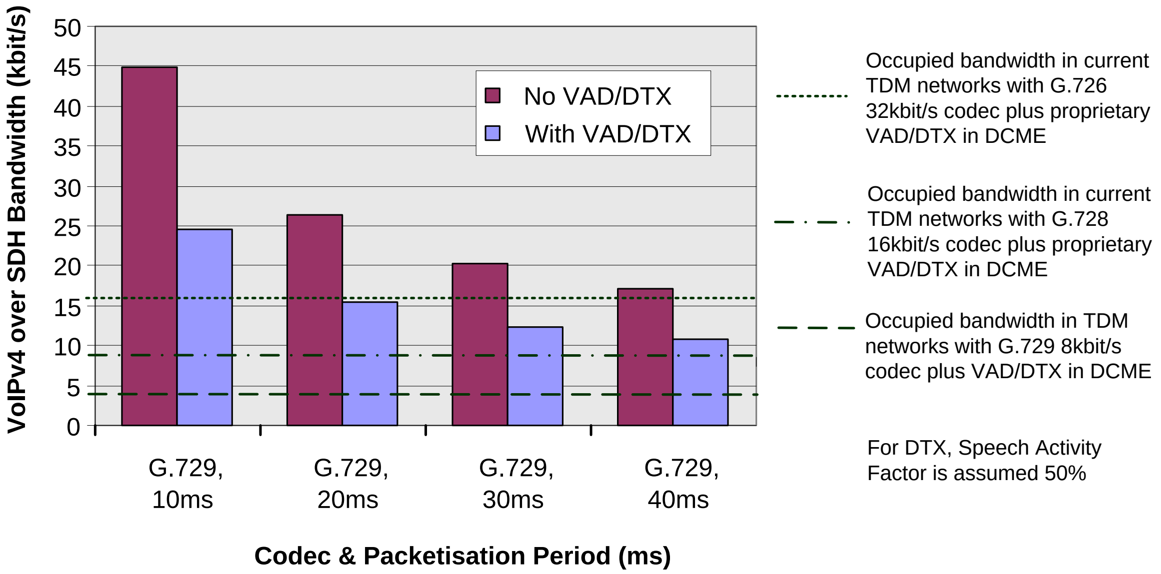


Figure 10 Showing the effect on transmitted bandwidth on SDH bearer of G.729 encoded VoIPv4 signal of different packetisation periods, with and without Voice Activity Detection/Discontinuous Transmission.

Further to point 3 above, Figure 10 illustrates very clearly that an IP-based voice signal bandwidth is higher than the corresponding TDM voice signal using the same codec (compare all G.729 IP-based voice signal bandwidths with the voice bandwidth of 4kbit/s attained by the TDM DCME using the G.729 codec). The additional occupied bandwidth is because of the added packet overheads required (TDM has no similar overheads).

12.1.3 Packetisation Process Latency and Satellite Latency

While reducing bandwidth, pp transrating introduces additional latency (section 6.9), and latency has an adverse impact on quality (see section 8). For satellite links, this needs to be kept in perspective.

Figure 11 shows the packetisation process latency of the VoIP signals used in the Figure 10 illustration, together with the satellite link latency. Observe that while significant latency is introduced by the higher pp values, when the satellite latency is also considered, the pp latency is dominated by the satellite latency, so that translating pp to higher values may be an attractive tradeoff.

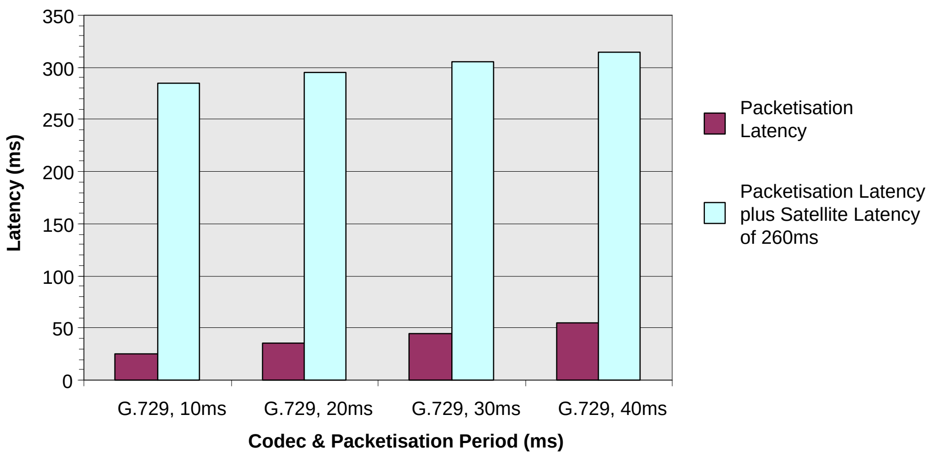


Figure 11 Illustrating VoIP packetisation latency for G.729 encoded voice signals, plus when compounding the latency of a satellite link.

12.1.4 IP/UDP/RTP Header Compression

The 46 bytes of POS/IPv4/UDP/RTP headers (66 bytes for POS/IPv6/UDP/RTP headers) compared to the relatively small voice information payload (10 bytes in the case of G.729,10ms) is the predominant reason for the increased voice bandwidth required in VoIP networks compared to TDM networks.

It is possible to compress (on a link-by-link basis) the 40 (or 60) bytes of IP/UDP/RTP headers to remove header redundancy for a particular stream using Compressed RTP (CRTP) to RFC2508 [58]. Although several header fields change in every packet, the difference from packet to packet is often constant[50]. All that has to be transmitted (after sending the initial packets of a new flow to establish what is called a context [58] at the receiving end) is an indicator that the differences are indeed constant and the header can be re-calculated at the receiving end from the transmitted compressed header information and the context information previously stored at the receiving end[51]. CRTP reduces the 40 bytes of IPv4/UDP/RTP packet overhead or the 60 bytes of IPv6/UDP/RTP packet overhead to 2 bytes [52].

The associated RTCP header is not compressed, but the IP/UDP part of that header may be compressed using the same technique. However since RTCP flows are a small fraction of the size of the associated RTP flows the compression efficiency impact of transmitting RTCP uncompressed is small in practice

IP/UDP/RTP compression is applied by processing below layer 3 and is thus limited to a single link. This compression is generally associated with radio equipment, e.g. satellite modem equipment contains packet processing IP/UDP/RTP compression modules. Considerable CPU power is needed to recalculate the headers at wire speed, generally discouraging the use of such compression unless the bandwidth savings warrants.

There are more recently developed compression schemes such as RObust Header Compression (ROHC), RFC3095 [59], being particularly targeted at radio links on mobile radio networks, and this has also been applied by satellite equipment vendors. ROHC can reduce the IP/UDP/RTP headers (IPv4 or IPv6) to a minimum of 1 byte, although in practice for packet flows containing many concurrent calls 3 bytes would be the typical compression achieved. In terms of the practical conclusions sought in this White Paper, this difference is insignificant.

Figure 12 compares the occupied bandwidth on an SDH satellite bearer of the G.729 coded VoIPv4 signals in the examples above, with the same signals with IPv4/UDP/RTP compression (CRTP) to 2 bytes applied.

---

50 RFC 2508 [58] describes the following mechanism: "*…although several fields change in every packet, the difference from packet to packet is often constant and therefore the second-order difference is zero. By maintaining both the uncompressed header and the first order differences in the session state shared between the compressor and the decompressor [across a satellite link, say], all that must be communicated is an indication that the second order difference was zero. In that case, the decompressor can construct the original header without any loss of information simply by adding the first order differences to the saved uncompressed header as each compressed packet is received*".

51 The context information maintained at the receiving end is also updated from the compressed headers, and if transmission errors cause context damage a resynchronising recovery mechanism is invoked.

52 If the UDP checksum is also generated and transmitted this increases to 4 bytes. This is required if the application requires end-to-end error detection. The PSTN voice application does not require end-to-end error detection.

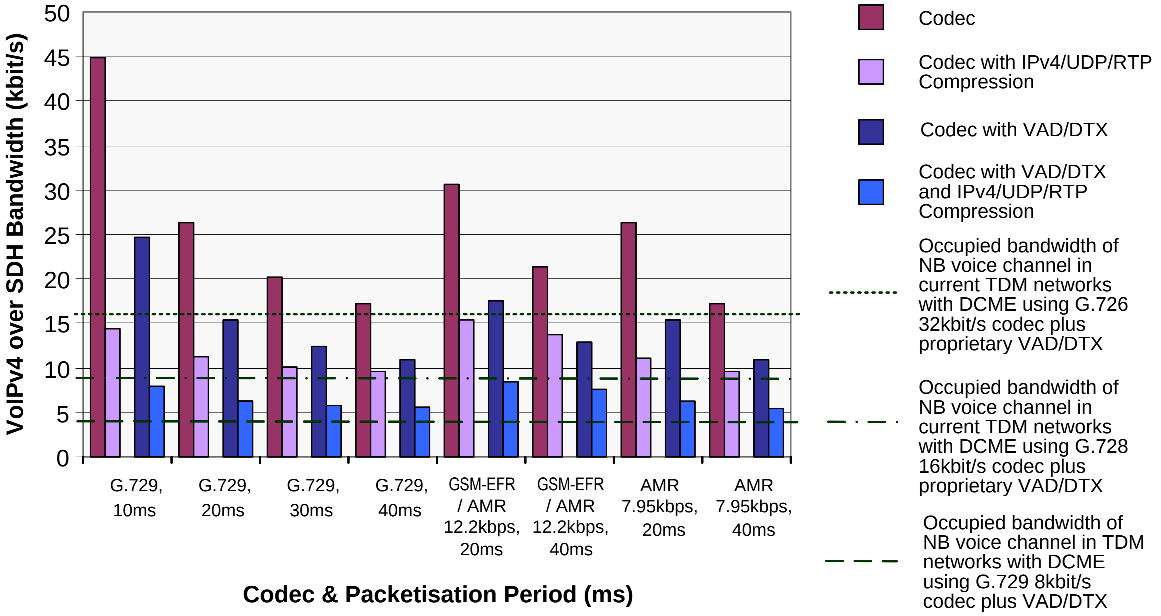


Figure 12 Demonstrating the bandwidth reduction of IPv4/UDP/RTP compression to 2 bytes on the transmitted bandwidth on SDH bearer of variously encoded VoIPv4 signals of different packetisation periods. Results are shown both with and without VAD/DTX. Speech Activity = 50%

The large reduction in bandwidth militates for the use of IP/UDP/RTP compression techniques to save transmission costs without lowering voice quality. IP/UDP/RTP compression would to need be applied at both ends of the link by bilateral negotiation.

Thus, is highly likely that carriers using satellite transmission will choose to:

- transcode to a Low Bit Rate codec (particularly one with VAD/DTX/CNG),
- transrate the VoIP signal to a higher packetisation period, and
- apply IP/UDP/RTP compression

to conserve bandwidth.

The case for using IP/UDP/RTP compression becomes even higher for IPv6 packetisation, as a G.729/20ms VoIPv4 signal over SDH occupying 26.4kbit/s will increase by 30% to 34.4kbit/s for VoIPv6.

12.2 Factors Affecting Voice Quality

12.2.1 Codec Bit Rate

Of the four factors listed in section 12.1, lowering the codec bit rate has the most effect on voice call quality. *Ie* generally increases as the codec bit rate lowers, thus creating a customer service counterbalance to the financial desire to use the lowest bit rate codec possible (although as the codec bit rate reduces the packet overheads occupy an increasing proportion of the transmitted signal so that reductions in codec bit rate do not reduce transmission costs "proportionally").

Because carriers using satellite transmission are likely to transcode to a Low Bit Rate codec, the transcoding considerations of the previous sections are particularly relevant. Table 2, sections 7-10, and section 13 give information to enable the voice quality of such call paths to be estimated.

It is further noted that transcoding to a Low Bit Rate codec will not admit mobile-mobile calls using TrFO and transcoding distortion will thus apply.

12.2.2 Packetisation Period Transrating

To a lesser extent packetisation period, through increased call latency, also reduces quality. This is covered in section 7.5.

12.2.3 Voice Activity Detection/Discontinuous Transmission

VAD/DTX is particularly effective in reducing the transmitted bandwidth and has no significant effect on voice quality[53], other than indirectly through increased latency (see section 6.6). VAD/DTX has been successfully used in DCME for two decades.

12.3 Voice Quality – Bandwidth Cost Tradeoff

The predominant tradeoff for network planners is that between occupied bandwidth and voice quality impairment, Ie. Section 12.1 highlights the bandwidth affecting parameters and section 12.2 that the predominant impairment is due to the codec bit rate, which is set by the codec type (and operating bit rate if user selectable). Figure 13 gives examples of these two critical factors in satellite VoIP transmission (occupied bandwidth and Ie) for several Low Bit Rate codecs and VoIP transmission parameters to provide the reader with a perspective of the signal bandwidth and voice quality attainable. Such calculations will need be undertaken for particular codecs of interest to network designer(s), so that appropriate choices are made for the particular network link under consideration. Note that the AMR codec is not included (other than the 12.2kbit/s rate at which it performs as GSM-EFR and hence Ie is known by association) because no Ie data can be found by the authors.

Points to note are

- the high impairment of the G.723.1 codec,
- the significantly lower impairment of the mobile codecs,
- that IP/UDP/RTP compression and VAD/DTX are both key to reducing the occupied bandwidth of VoIP signals to that approaching that of voice in TDM networks [54], and
- that voice occupied bandwidth equal or better to that achieved with TDM transmission using the predominantly high compression DCME type is attainable with VoIP with lower speech impairment (distortion).

The penultimate point drives consideration of lower bit rate codecs than are used in TDM networks, which brings the voice quality tradeoff into stark consideration.

Supporting the last point, Figure 13 shows that VoIP signal bandwidths on satellite equal to or less than is currently attained with TDM transmission and with lower distortion is attainable. Consider a TDM link equipped with the predominant high compression DCME (G.728 codec), which delivers 8:1 voice compression resulting in an average voice channel occupied bandwidth of 8kbit/s. The Ie is 7 for the G.728 codec. For the VoIPv4 signal, consider the AMR codec at 12.2kbit/s with VAD/DTX and IP/UDP/RTP compression, which achieves a similar occupied bandwidth (8.46kbits/s with pp=20ms, 7.66 kbits/s with pp=40ms) with Ie=5, less (better) than the DCME. Although lower bandwidth could be achieved with G.729 (~6.4kbit/s with pp=20ms) the voice distortion would be higher (*Ie*=11).

---

[53] In the particular case of G.729, the difference in voice quality between G.729 and G.729a with VAD/DTX in Table 2 and ITU-T Rec. G.113 [18] is primarily due to the lower complexity encoding used in G.729a. For G.729, the VAD in [42], Appendices II and III, which is optimised for VoIP, should be used.

[54] Figure 13 indicates that for any given codec Low Bit Rate codec bit rate, bandwidth costs of ~50% higher must be accepted for VoIP transmission compared to TDM transmission. This includes the use of VAD/DTX in both cases. E.g. G.729 with VAD/DTX achieves an average voice transmission rate of ~4kbit/s on TDM bearers equipped with DCME and achieves ~6.4kbit/s for VoIP (IPv4 or IPv6) with pp=20ms, VAD/DTX and IP/UDP/RTP compression.

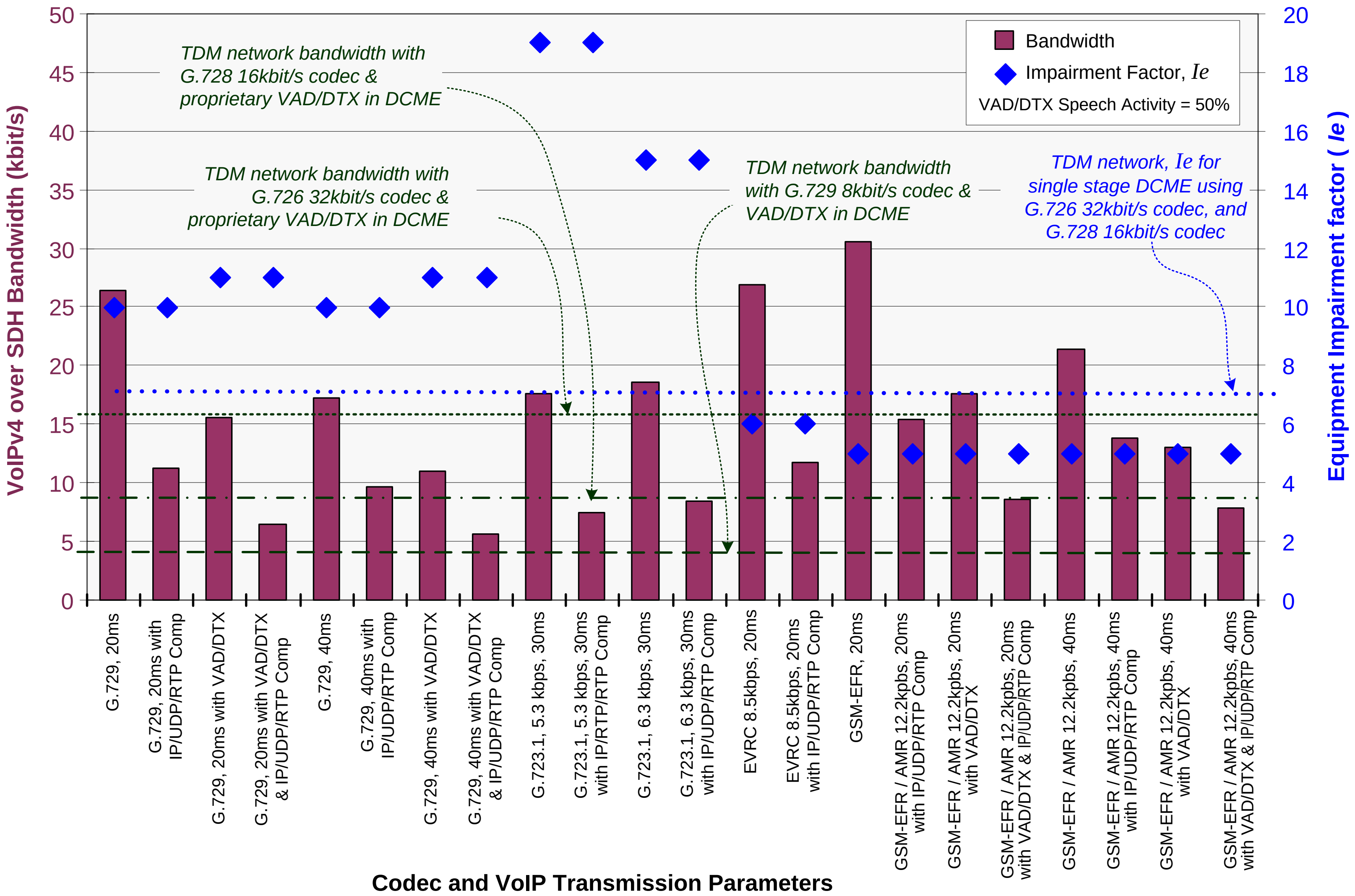


Figure 13 **Examples of occupied bandwidth and voice quality Impairment Factor *Ie* for several NB codecs and VoIP transmission parameters**

12.3.1 Other Network Link Considerations: Voice

Given that a low bit rate satellite link may contribute significant call quality impairment, it may be possible for special consideration to be given, by inter-carrier cooperation in end-to-end design, to minimizing quality impairments on other links in an end-to-end voice connection (e.g. in a narrow band context, links extending the satellite link by, say, submarine cable, to the ultimate destination could use the high quality reference codec, G.711). This would be a matter for the particular bilateral network designers.

In the case of satellite links serving communities of predominantly mobile SP's, it may be possible, through co-operative network design, to extend the use of the mobile codec through the satellite link, thus avoiding the generally high impairments of transcoding to a different codec just for the satellite link. Mobile codecs seem to have less Ie data available however, making such potential design choices problematic.

The introduction of wideband codecs by SP's will exacerbate the bandwidth cost considerations given in this section, although the use of newer wideband mobile codecs (such as AMR-WB) at the lowest bit rate supporting wide band audio (Table 2) may give a better cost/quality tradeoff than fixed WB codecs. However, the lowest bit-rates of AMR-WB were meant to be used as fallback modes under temporarily poor radio conditions; permanent use of those modes is not advisable.

12.4 General Transmission Considerations for IP Satellite Links used for Migrating PSTN Voice Services

DCME is used on TDM satellite links fundamentally to gain control over, and to limit the bandwidth on such links, thus making them economic. This section presents some additional factors for the designer(s) of VoIP satellite links to consider, particularly when the satellite link is used to migrate the PSTN to VoIP.

Voice networks have also traditionally handled fax services and some modem traffic – the latter as Voice Band Data (VBD). Since satellite links are expensive they are usually dimensioned with low headroom at traffic peaks. Thus the following practices evolved:

- calculating fairly accurately the bandwidths of all signal types,
- arranging for some traffic and bandwidth limiting devices to prevent link overload affecting all customers,
- utilising DCME on long international or high cost of bandwidth links.

Thus the network designer(s) will need to estimate fax and VBD traffic volumes and bandwidths, and add them to the expected voice traffic bandwidth, as well as to consider the encoding required to reliably digitise the signals for these VBD based services[55]. The use of T.38 Fax over IP (FoIP) transmission limits the high bandwidths associated with fax being sent as VBD over IP [56, 57].

[55] DCME generally accommodates VBD and fax signals by switching in a codec capable of reliably encoding/decoding such signals. This was commonly the G.726 codec operated at the 40kbit/s rate (G.726 at 32kbit/s and other low bit rate codecs, particularly G.729, are conditioned for voice and do not reliably encode/decode VBD signals, see sections 6.1 and 6.5).

[56] For fax to work reliably as a VBD transmission, in a narrow band context the G.711 codec must be used with pp=10ms or less. G.711/10ms would occupy ~70kbit/s on a satellite SDH link with IP/UDP/RTP compression. G.711 with pp=5mS increases the occupied bandwidth to 76kbit/s (with IP/UDP/RTP compression).

[57] Some DCME models demodulate the VBD fax signals to a digital signal for transmission at the ingress DCME terminal, and remodulate to VBD at the egress DCME terminal. For example, fax signals coded to the ITU-T V.29 standard [62] could thus be transmitted over the international circuit at, typically, 9.6kbit/s. T.38 may be viewed as the IP successor of that technology.

TDM PSTN links with DCME were statistically dimensioned, and it was possible for the DCME to signal the TDM switch that no more calls could be accepted until some talker channels were released [61]. It is not known if any current soft switches can limit traffic based on specific media stream size, so that this would represent an overload risk suggesting VoIP satellite links be dimensioned with additional headroom. This would increase link costs putting additional pressure on the voice transmission bandwidth.

Since network planners of TDM satellite links use DCME (with its transcoding to lower bit rate codecs) and call limiting features [61] to control the bandwidth of the link and to prevent overload affecting the voice quality of all active calls, the factors presented in this section also strongly point to satellite IP link designer(s) providing transcoding/pp transrating capability to gain control over the VoIP signal bandwidths on the satellite link, rather than being at the mercy of the VoIP traffic bandwidths sent by SP's. If satellite link overload were permitted to occur by simply accepting all VoIP signals delivered by SP's it would degrade the quality of all active calls on the link. Thus carriers using satellite links and wishing to gain control over bandwidths (and hence both link economics and voice quality) would need to utilise transcoding and pp transrating. Satellite transmission thus appears to be a situation justifying transcoding, as otherwise access to remote areas/countries could become uneconomic.

13 Evaluation of Codec Choice in International IP Interconnections

This section presents calculations for narrow band codecs. Where *Ie,wb* values are known for wideband codecs the same calculations may be made for WB scenarios as are made for NB scenarios, noting that:

- the examples given in this section for NB can normally be translated into the WB context by simply replacing Ie by Ie,wb values,
- mixed WB/NB scenarios are <u>explicitly excluded,</u>
- <u>calculations beyond R-factor, such as the effect of TELR impairment, cannot be done</u>, and
- the E-model is still under study in the ITU-T, so the application of such results should be treated with high caution.

13.1 Bilateral and Series Configurations

The transcoding examples analysed show that configurations with a series of carriers involved pose a particular problem for voice quality in that there may be several intermediate (transit) carriers in a particular international configuration, and information about codec and packetisation downstream from the contracting carrier may be hard to obtain, thus frustrating call quality estimation.

In addition, there is presently no way for any intermediate network to automatically determine (e.g. via signalling) what the codecs are in the end Service Providers networks, so that codec choice can only be based on the immediately adjacent carriers and the particular codec policies of those carriers (rather than on end-to-end call considerations). If cost is the dominant criterion of an intermediate carrier, they may transcode within their network to save capacity costs regardless of ingress and egress carrier codec primary offers[58], consequently profoundly impacting the end-to-end call they are involved with. Conversely, it may happen that the same codec/pp is used throughout, with quality maintained (this could occur if a carrier was able to enforce common code/pp parameters on all suppliers by agreement).

---

[58] In the SDP part of SIP signalling.

Generally, configurations with a number of carriers involved from end-user to end-user are highly likely to be non-optimum overall, with significant transcoding occurring, that transcoding being sufficient to lower call quality into the "*Not Recommended*" zone (Figure 3). This could correct in time as configurations evolve form in response to poor quality reports, but the effect on customer quality meantime may give carriers (and Service Providers) a bad name. This paper provides a methodology to predict such adverse outcomes and it is highly recommended that planning estimates be made, with scenario data if necessary.

It is concluded that for IP-based voice, bilaterally engineered direct interconnections, with full information available from Domestic and Service Provider Network Operators, will offer predictable quality better able to be matched to voice product requirements, and particularly will offer the lowest quality reductions vis-à-vis TDM because more direct connections reduce impairments.

13.1.1 Bilateral Interconnection Configuration

In the bilateral interconnection configuration the design is fully controlled, hence coding impairments are predictable and minimised because of direct connections.

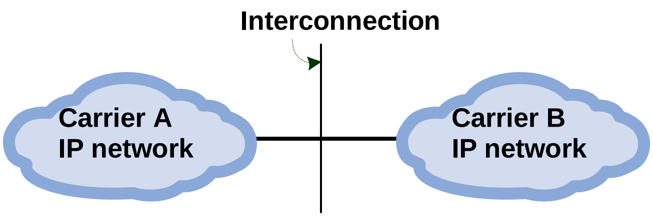


Figure 14 Bilateral interconnection configuration

13.1.2 Series Configuration

In this configuration a carrier receives voice traffic from multiple sources and offers voice traffic to multiple destinations regardless of the bilateral commercial relationship this carrier has with its own downstream carriers (i.e the traffic is not generated, in general, in the country where the carrier requesting the delivery is located and it is not terminated, in general, in the country where the carrier accepting the traffic for delivery is located).

In this configuration the design is thus not fully controllable, hence coding impairments may be higher, and design has to be carefully chosen (like low pp with a low frame length codec) to minimise compounding impairments that may already have occurred "downstream".

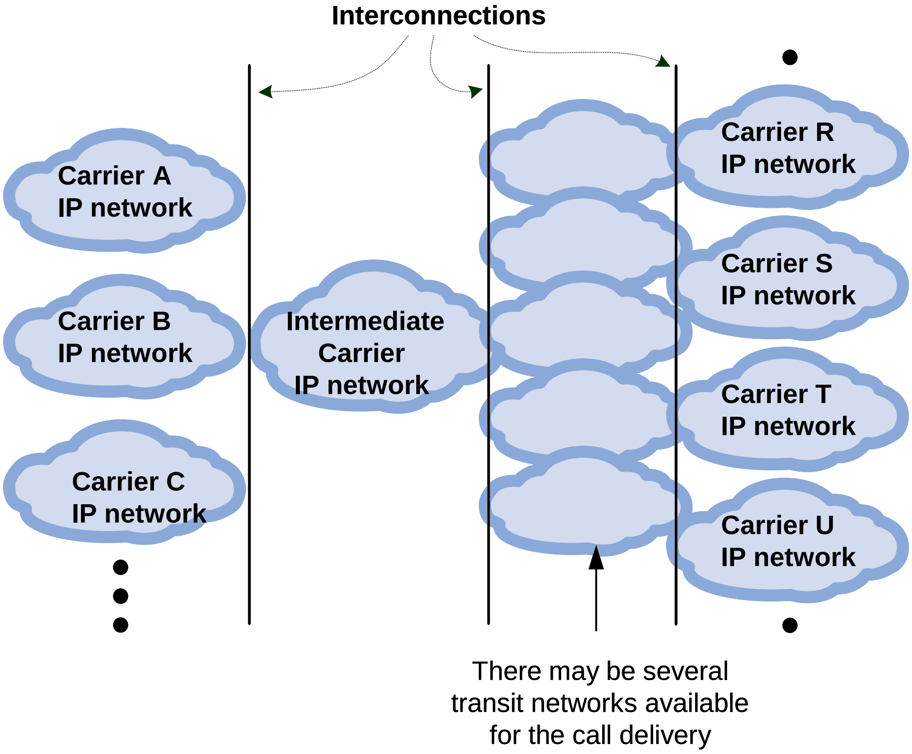


Figure 15 Configuration with Multiple Carriers end-to-end.

13.2 Calculation Example for Configurations with all Narrow Band Codecs

13.2.1 Assumptions

Evaluation of configurations may be made using the E-model as defined in ITU-T G.107 [19] (see also section 7.2), i.e. based on the R-factor calculated by the following formula:

$$R = Ro - Is - Id - Ie + A$$

where:

Ro = 93.2

$Is \approx 0$

A = 0 for fixed networks.

Id – delay impairment

Ie – equipment impairment

Packet Loss <0.1%

If the packet loss is kept below 0.1% then $Ie,eff \approx Ie$ which means that the influence of packet loss on the Ie impairment value may be neglected.

13.2.2 Determination of Reference Configuration

For evaluation purposes the following recommended i3 Forum configuration will be used:

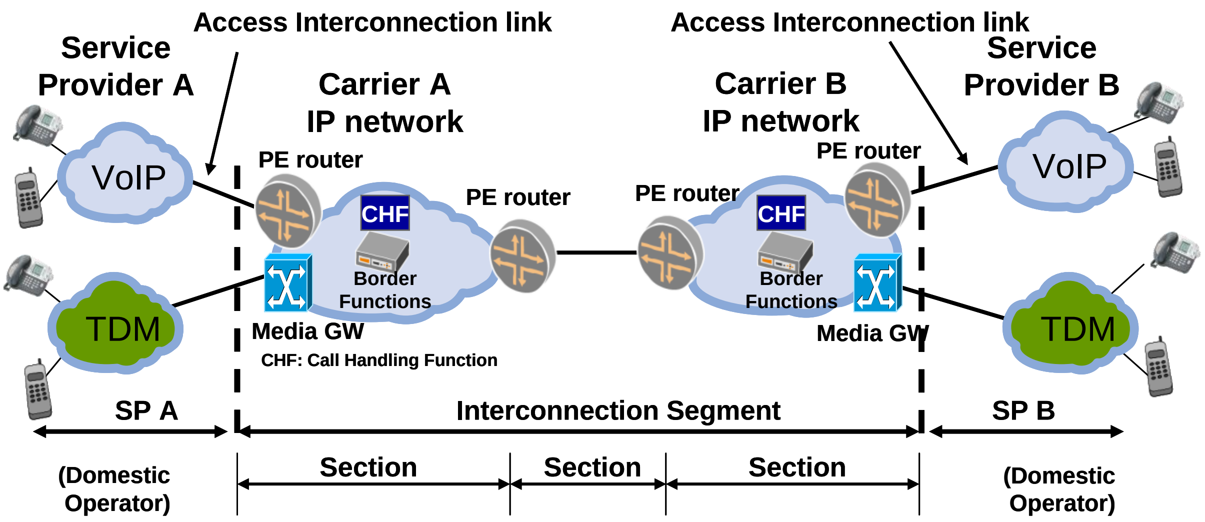


Figure 16 Reference configuration

13.2.3 Ascertainment of Actual Transmission Impairments in each Section

Following the above assumptions only two impairments are taken into consideration:

Id – delay impairment which results from transmission delay and codec delay.

- transmission delay can be evaluated basing on ITU-T G.114 [7] and other values such as measurement results or equipment technical data.
- codec delay is introduced mainly in an SP network where voice is digitised and encoded/decoded. In the international part of the network it is introduced only in the case of transcoding/tandeming. The values of delay for most popular codecs can be found in ITU-T G.114 [7] Annex 2 or in Table 2 of this document.

Ie – equipment impairment which results mainly from quantising distortion and codec algorithms. The *Ie* impairment value depends strongly on packet loss. The values of impairment introduced by most the frequently used codecs/bit rates and for packet loss = 0 can be found in ITU-T G.113 [18] (11/2007) Appendix 1 Tables I.1 thru I.5, noting that packet loss is characterised as "random" or "bursty".

13.2.4 Impairment Calculation and End-to-End Evaluation

To calculate total impairment it is necessary to evaluate the two domestic segments. As an international carrier cannot always know all Service Provider network parameters, it is recommended to take into consideration a reasonable safety margin.

The following calculation assumes that there is no packet loss in each segment. Practically if packet loss is kept below 0.1% its influence on the Ie value may be neglected. For other packet loss values Ie must be separately determined on the basis of the table 2B/G.108 in ITU-T G.108 [6].

Example of calculation:

			Impairments	Delay		
Section 1 Service Provider A network (sending)						
Impairments						
Codec/pp	G.711/20mS	Ie	0			
packet loss %	0		0			
Delay						
Access Network delay				20	ms	
Codec Delay incl pp (*associated with codec above*)				20.375	ms	
Propagation Delay (mS)				0	ms	Note 1
Others				0	ms	
Total Impairments & Delay, Service Provider A			0	40.375	ms	
Section 2 Carrier A IP-based voice network.						
Impairments						
Codec/pp	G.711/20mS	Ie	0			
packet loss %	0		0			
Others						
Delay						
Domestic Network delay				10	ms	Note 2
Transcoding Delay incl pp					ms	
Others					ms	
Total Impairments & Delay, Carrier A			0	10	ms	
Section 3 International Carrier Bilateral IP-based voice network.						
Impairments						
Codec/pp	G.711/20mS	Ie	0			
packet loss %	0		0			
Others						
Delay						
International Network delay				80	ms	Note 3
Transcoding Delay incl pp					ms	
Others					ms	
Total Impairments & Delay, International Network			0	80		
Section 4 Carrier B IP-based voice network						
Impairments						
Codec/pp	G.729/20mS	Ie	10			Note 4
packet loss %	0		0			
Others						
Delay						
Domestic Network delay				10	ms	Note 2
Transcoding Delay incl pp				25	ms	
Others					ms	
Total Impairments & Delay, Carrier B			10	35	ms	

Section 5 Service Provider B network (receiving)						
Impairments						
Codec/pp	G.729/20mS	*Ie*	0			
packet loss %	0		0			
Delay						
Access Network delay				20	ms	
Transcoding Delay (*associated with codec above*)				0	ms	
Propagation Delay (mS)				0	ms	Note 1
De-Jitter Buffer				25	ms	
Others				0	ms	
Total Impairments & Delay, Service Provider B			0	45	ms	
Total Service Provider evaluation						
Total Impairments & Delay, Service Provider A & B			0	85.375	ms	
Total international and domestic evaluation						
Total Impairments & Delay, Carrier A & B + International			10	125	ms	
Total end to end evaluation						
Total Impairments & Delay, end-to-end			**10**	**210.375**	**mS**	

Note 1 Add propagation delay if SP has significant domestic reach
Note 2 Delay for small country
Note 3 From Table 6 e.g. Europe - America
Note 4 The codec Impairment (*Ie*) is recorded when the codec is first encoded from G.711

Table 10 Calculation Example

13.2.5 Judgment of Results

Now it is possible to check the overall voice quality. In the E-model the voice quality is satisfactory if the end-to-end R-factor is equal to or greater than 70.

To determine an R value by considering the various impairments in turn as previously described the following steps should be performed:

1. Identify the echo (TELR) curve appropriate for the configuration under consideration on Figure III.1/G.131 p.10 in ITU-T G.131 [53],
2. Read the R value for the calculated above Total Delay,
3. Subtract from read R value the total *Ie* impairment calculated above. If R ≥ 70 voice quality should be acceptable.

This procedure is illustrated in Figure 17.

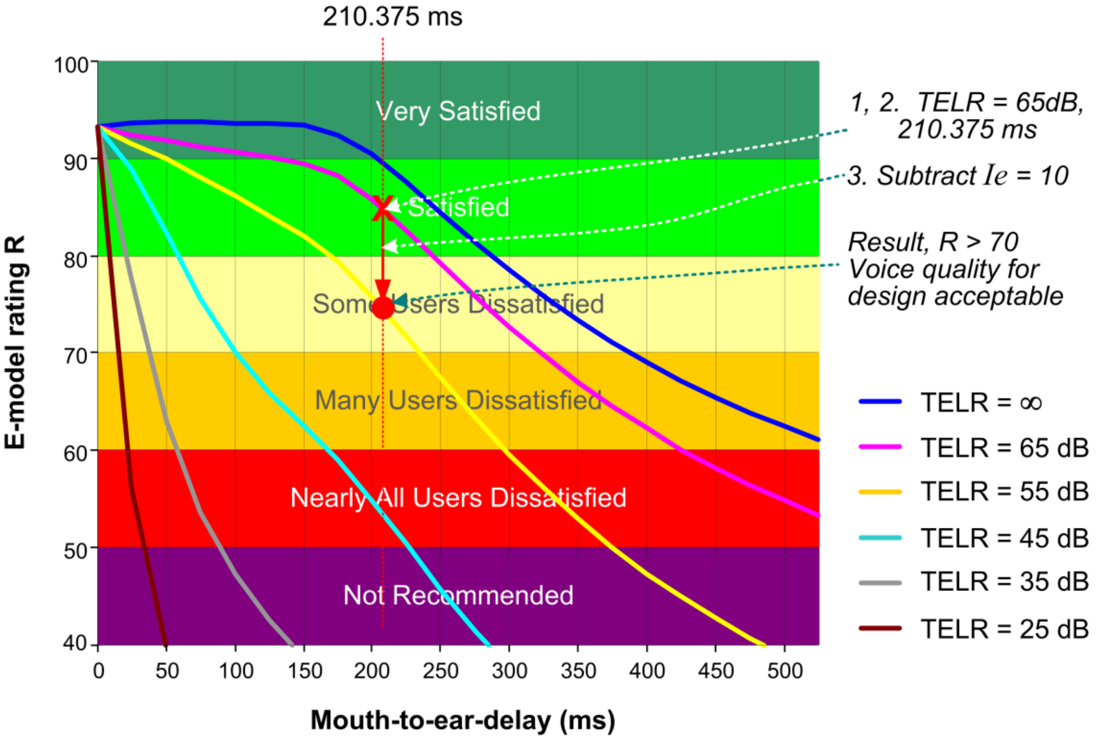


Figure 17. Calculation Example Result shown on R-factor as a function of Total Delay and Talker Echo Loudness Rating (TELR) Graph.

Thus, in this example, using the "default curve TELR=65 dB" and total delay = 210.375 ms, after subtracting the codec impairment for the design (Ie=10), R>70 which is still in the "acceptable" area.

In this case however it is necessary to consider once more the whole configuration. These calculations are not precise because of an assumption that packet loss = 0. We are close to the chosen design limit of R=70 and if packets were lost the R-factor would fall below an acceptable level for this design. If transcoding had been avoided then the voice quality would have been satisfactory for all users (although it is noted that this would also be the result for a design in which both Service Providers used the G.729 codec, i.e. G.729 end-end, in which case it should be noted that additional transcoding would result in unacceptable quality for the design target of R≥70 chosen here).

It is also possible to shift down the relevant TELR curve, subtracting the Ie value at each point and checking the delay margin remaining until the point where this implied curve intersects the R=70 level.

Another way to calculate R-factor is to use the ITU web based tool at *http://www.itu.int/ITU-T/studygroups/com12/emodelv1/* (free of any need for software copyright licences when used in accordance with the conditions and disclaimers noted in G.107 (08/2008) Appendix III) to calculate an R value, allowing default parameters to be changed as required.

14 Conclusions and Recommendations

1. IP-based voice networks using narrow band codecs provide lower quality international voice calls than the TDM networks they replace, with the quality of all-cable network calls falling from "*Users Satisfied*" levels regardless of international distance to**,** when a single codec is used end-to-end, a voice quality ranging from lower "*Users Satisfied*" scores within regions such as Europe to "*Some/Many Users Dissatisfied*" for long international calls such as New Zealand/Australia to UK/Europe.
2. A single codec cannot be guaranteed for calls between all countries (or Service Providers), and when transcoding is necessary, voice call quality will range from only "*Some/Many Users Dissatisfied*" for intra region calls to "*Nearly All Users Dissatisfied*" for long international calls.
3. Careful planning will be required to minimise voice quality degradation, and carriers are encouraged to apply transmission voice quality analysis to all interconnections.
4. The E-model R-Factor/delay graph is a convenient planning tool for carriers to assess voice quality of international interconnections and its usage is recommended:
    a. for scoping major voice quality impairments,
    b. for more detailed voice quality design, if sufficient information is available from domestic network operators and Service Providers,
    c. if intermediate carriers are involved in international calls, estimates of latency to the final destination could be used, as well as the best knowledge that can be obtained about intermediate network codec usage and possible packet loss,

    noting that the E-model applies only for homogeneous voice paths containing only narrow band codecs (using narrowband impairment vaues Ie), or, with the extended scale, for homogeneous voice paths containing only wideband codecs (using wideband impairment vaues Ie,wb). It <u>does not</u> apply to mixed wideband and narrow band codecs in a connection.
5. IP-based voice with direct bilateral interconnections, engineered with full information available from the corresponding carriers, will offer predictable quality, at levels fulfilling voice product requirements.
6. IP-based voice via multiple downstream networks will generally present more difficulty in engineering to direct bilateral standards because several intermediate international carriers are often involved.
7. Longer term, wideband codecs, which have lower impairments and lower intrinsic distortion, which is potentially a compensation for quality lost in transcoding of narrow band low bit rate codecs, will counteract the quality degradation if used widely and their introduction by SP's should be encouraged.
8. Adaptive de-jitter buffers introduce discontinuities in speech which may be objectionable to the listener, and are best associated with Voice Activity Detection and Discontinuous Transmission (VAD/DTX) whereupon the discontinuities can be hidden in the silences.

14.1 Recommendations on Codec Choice

9. In network configurations where total delay is a critical parameter (particularly important for trans-oceanic international calls) it is recommended to use codecs with low algorithmic latency. Total delay can also be decreased by choosing shorter packetisation periods.
10. Packet loss should be kept as low as possible (total packet loss < 0,1%) so that its influence on voice quality may be neglected, although it is recognised that such low packet loss may not always be achievable on some links such as satellite when link occupancy

exceeds ~80%. In this case network planners should incorporate the expected packet loss into their calculations. It is also recommended to use Packet Loss Concealment whenever possible and to take into consideration the “Packet Loss Robustness” parameter of the codec used in configuration planning.

11. The G.723.1 codec (because of long frame length and relatively high distortion) is unsuitable in general for international voice networks (it could have application only where bandwidth is the over-whelming consideration, and then only if compensated for by using, say, G.711, in the remainder of the configuration).
12. The G.729 codec family offers a good balance of latency, bandwidth (cost) and voice fidelity in fixed networks. In mobile networks the AMR codec offers similar attributes but with less distortion.
13. Mobile SP’s will, until mixed IP-based voice/TDM networks are eliminated, experience best interconnection call quality for mobile-fixed calls where G.711 coded transmission is applied. However this is unattractive for international interconnections that must preserve bandwidth. Over time the use of TrFO in mobile networks will allow end-to-end carriage of mobile voice packets with all transcoding and “mobile tandems” eliminated, and this technical solution needs to be actively promoted for interconnection to fixed networks and for international transit.
14. Wideband (voice) codecs have lower impairments and lower intrinsic distortion which is potentially a compensation for quality lost in transcoding of narrow band low bit rate codecs. The application of newer wideband codecs in the G.729 family may offer a migration path in time due to backwards compatibility although this would take time and benefits gained would be limited meantime due to the predominance of TDM PSTN’s for the foreseeable future in international configurations. An alternative would be the introduction of the AMR family of codecs into fixed networks, which would eliminate much of the transcoding impairments between fixed and mobile networks and especially AMR-WB for end-to-end high WB voice quality with WB transcoding or fall back to NB.
15. In network configurations where occupied bandwidth is a critical parameter (particularly important for satellite transmission) it is:
    a. considered acceptable to utilise transcoding (if necessary), and it is recommended to utilise packetisation period transrating and other IP transmission techniques to gain control of transmission bandwidth (and hence link economics):
        i. select codecs with low bit rate and low Ie (this balance between cost and voice quality needs due consideration of the end-to-end performance required),
        ii. apply Voice Activity Detection and Discontinuous Transmission (VAD/DTX),
        iii. consider transrating packetisation period to higher values, such as 40ms,
        iv. implement IP/UDP/RTP compression on the links with restricted or costly bandwidth,
        v. consider also, when migrating from the PSTN to VoIP, the bandwidth and encoding requirements of other traffic on the link such as fax and Voice Band Data.

    Note that transcoding events may be minimised e.g. if a satellite link serves mobile SP’s, use the mobile codec on the satellite link rather than transcoding to a different codec such as might be commonly used in a fixed network,

    b. where end-to-end performance is being bilaterally designed, inter-carrier cooperation in end-to-end design may allow other links in a connection to be engineered to minimise total quality impairments (such as by using a high quality codec in the remainder of the network).

14.2 Transcoding

16. Transcoding of low bit rate codecs greatly decreases international call quality, especially on long connections, and should be avoided unless absolutely necessary.

17. Generally, arrangements with a number of carriers involved in the end-user to end-user communication are likely to have significant transcoding, quite possibly sufficient to render call quality completely unacceptable (or even unintelligible under normal listening conditions) so that alternative network configurations may need to be sought.

18. If transcoding is necessary (or is known to happen in another part of the end-user to end-user communication), complete the international design by:

    a. Favouring codecs with low frame lengths and choosing low packetisation periods to minimise compounding latency,

    b. When multiple carriers have to be crossed, carriers should ascertain downstream codec information for transmission planning wherever possible,

    c. If not available, estimates of delay to destination plus "what-if" scenarios to assess possible quality degradation should be done as part of interconnection negotiation.

14.3 Companding Conversion for G.711 codec

19. If a G.711 encoded call is routed across the borders of either the North American Numbering Plan countries or Japan then G.711 A-law/µ-law companding conversion is necessary and this companding conversion will be done by the countries using the µ-law.

20. With IP connectivity (in contrast to circuit switched TDM which has identifiable gateways at the international boundaries), the same network section may span those international boundaries, so that in voice path engineering of IP-based voice networks, care should be taken not to overlook this requirement as G.711 companding conversion has to be specifically included by a µ-Law country for G.711 encoded calls between A-Law and µ-Law countries.

21. This companding media stream conversion is usually implemented using the same media stream conversion method as transcoding (i.e. de-packetising, conversion, and re-packetising) and thus introduces additional latency (impairing voice quality through the E-Model paramter *Id)*, although there is no voice quality *Ie* impairment.

22. Companding conversion by this method interrupts the RTP and RTCP streams thus limiting the utility of QOS measurements based on RTCP.

23. Companding conversion methods based on conversion of the G.711 PCM voice samples within the packet without de-packetising, with appropriate "adjustments" of the codec payload type in the RTP header would appear to offer promise to solve both the voice quality impairments caused by additional latency and the RTCP interruption issues presented by a depacketising/repacketising media conversion.

24. If a call is to be routed to a TDM network, appropriate G.711 A-law or µ-law shall be chosen with the µ-law country doing any necessary companding conversion.

14.4 Call Setup

25. Order of codec/packetisation period preference is determined by the originating terminal and should be honoured where possible.

26. If the call is to be routed to a TDM network and if the originating terminal does not support G.711 interconnection, the carrier interconnecting to the TDM network shall perform transcoding.

27. In case of fixed-mobile interconnection, transcoding if necessary shall always be performed by mobile network.

15 Appendix 1 Maximum R-Factors for Narrow Band Speech (G.711 PCM encoded) and Wide Band Speech (16kHz Sampling Frequency PCM encoded)

Absolute Delay (ms)	NB R-Factor	Absolute Delay (ms)	NB R-Factor	Absolute Delay (ms)	NB R-Factor
0	93.2	225	87.5	425	67.2
25	93.2	250	84.0	450	65.5
50	93.2	275	81.0	475	64.1
75	93.2	300	78.3	500	62.7
100	93.2	325	76.0	525	61.4
125	93.2	350	73.6	550	60.0
150	93.0	375	71.3	575	58.7
175	92.0	400	69.0	600	57.8
200	90.3				

For Wideband Codecs, the extended R-factor scale is obtained by adding 35.8 to the above figures.