INTERNATIONAL INTERCONNECTION FORUM
FOR SERVICES OVER IP

(i3 FORUM)

(www.i3forum.org)

Workstream "Technical Aspects"

White Paper

Optimal Codec Selection in International IP based Voice Networks

(Release 1.0) May 2009

Executive Summary

This White Paper assists in correct codec 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.

Codec engineering (the practical application of codecs) 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 basics, 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, and may necessitate different network configurations being sought.

This white paper complements the content of the i3 Forum document "*Technical Interconnection Model for International Voice Services, Release 2 (May 2009)*" with regard to the media information flow management / treatment.

Table of Contents

1 Scope and Objective

The scope of this paper is the voice quality of the media path as affected by codecs as used in interconnected IP based voice networks.

The objective of this paper is to provide background to and to support the codec sections in "*i3 Forum, Technical Interconnection Model for International Voice Services, Release 2 (May 2009)*" as well as to draw attention to the adverse voice quality voice 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
ACELP	Algebraic-Code-Excited Linear Prediction
ADPCM	Adaptive Differential Pulse Code Modulation
ADSL	Asymetrical Digital Subscriber Line [equipment]
A-law	Companding profile (volume compression) used by all countries except for USA 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
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]
DTX	Discontinuous Transmission
EF	Expedited Forwarding
EV-CELP	Enbedded Variable bit rate – Code-Excited Linear Prediction
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
ISDN	Integrated Services Digital Network
ITU-T	International Telecommunications Union – Telecommunications
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
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
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	Algebraic-Code-Excited Linear Prediction
MR-ACELP	Multi-Rate Algebraic Code Excited Linear Prediction
ms	millisecond
µ-law	Companding profile (volume compression) used in USA and Japan
NB	Narrow Band (with respect to voice frequency signal band width)
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
pp	packetisation period
Ppl	Packet loss ratio (used for E-model calculations)
PSTN	Public Switched Telephone Netwok
qdu	quantisation distortion unit
QOS	Quality of Service
RAM	Random Access Memory
RCELP	Residual Code Excited Linear Prediction
Rec.	Recommendation
ROM	Read Only Memory
RPE-LTP	Residual Pulse Excited-Long Term Prediction
RTP	Real-Time Transport Protocol
SB-ADPCM	Sub-Band Adaptive Differential Pulse Code Modulation
SDP	Session Description Protocol
SIP	Session Initiation Protocol
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
VMR	Variable Multi Rate
VoIP	Voice over IP
VSELP	Vector Sum Excited Linear Predictive
WB	Wide Band (with respect to voice frequency signal band width)
WMOPS	Weighted Million Operations Per Second

3 References

[1] "i3 Forum "Technical Interconnection Model for International Voice Services", Release 2, May 2009

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

[3] 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)

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

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

[6] "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.

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

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

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

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

[11] ITU-T Recommendation G.107, "The E-model, A Computational Model for use in Transmission Planning" (03/2005)

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

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

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

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

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

[18] 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).

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

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

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

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

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

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

(http://books.google.co.nz/books?id=WLRBLsaVapwC&printsec=frontcover&dq=DCR+score+inaudible+annoying#PPR5,M1)

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

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

[27] 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)

[28] ITU-T Recommendation G.1020, "Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks", (01/2005)

[29] "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

[30] 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)

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

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

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

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

[35] 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.

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

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

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

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

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

4 Codec Engineering in IP networks

All TDM switching and interconnections used G.711 PCM coded 64kbit/s voice signals, and any transcoding to lower bit rates to lower transmission costs was undertaken by Digital Circuit Multiplication Equipment (DCME) within a carriers network. Consequently codec engineering expertise resides mainly with DCME equipment vendors. Codecs used in DCME were predominately 16kbit/s / 32kbit/s and speech quality reduction was low-to-moderate. In most international connections (i.e. all submarine cable links with global reach) "*All Users Satisfied*" quality levels were readily achieved in bilateral networks, and to a lesser extent, when a chain of networks is involved. Effectively, voice engineering had a low profile.

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 fidelity 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, 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 (several times higher), presenting a difficult commercial trade-off.

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/packetisation period 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 have lower voice quality (fidelity) 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.

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

5 General Reference Architecture

The general reference configuration for 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.

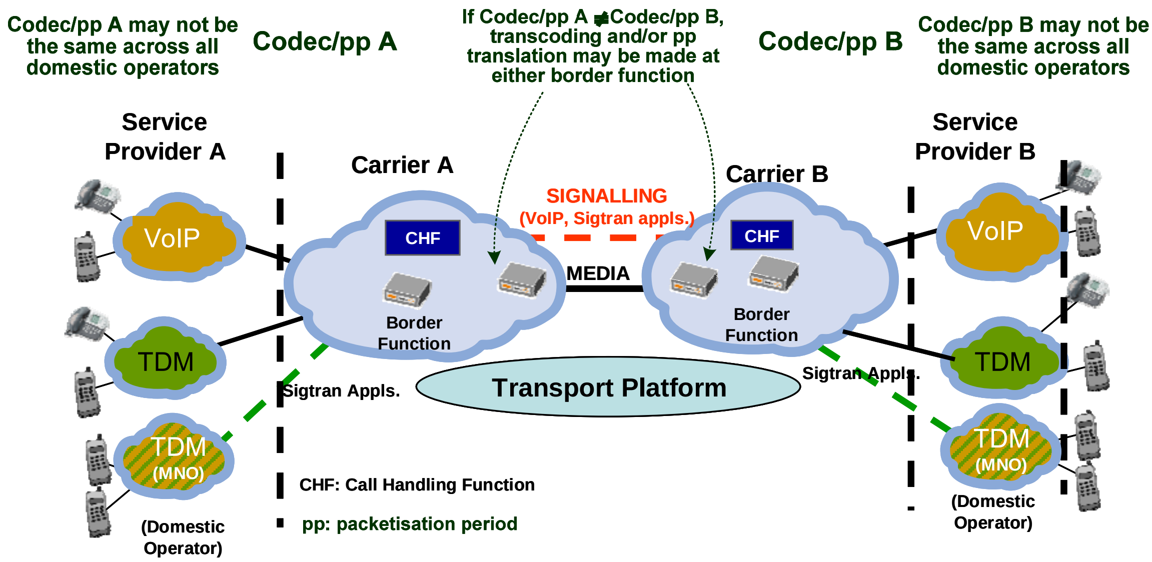


Figure 1 General Reference Configuration with codec annotation

6 Codec Basic Features

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 300-3400 Hz; this bandwidth (“narrow band”) ensured sufficient intelligibility and was the basis of the design of 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 (“wideband” up to 7 kHz or even higher) speech bandwidth [3]. 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 or wideband, complexity (CPU time to compute the code, static/dynamic RAM and ROM memory), delay and fidelity are typical trade-offs in codec design. It is predominantly the trade-offs in codec design that distinguish them[1].

6.1.1 Waveform codecs

Waveform codecs simply process the speech waveform as it arrives, sample by sample, e.g G.711 PCM[2] 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) [4]. These are non-waveform codecs and use speech synthesis techniques ([5], A.1.8).

In non-waveform codecs many G.711 0.125 ms 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. 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 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[3] becomes an important design trade-off. Sinusoidal coders, with frame independent coding which synthesise the output from slowly varying parameters (long term prediction) can tolerate up to one in 3 frames lost (~30% packet loss at 1 frame per packet) [6]. Latency generally increases which is an acceptable tradeoff when used for internet telephony where the IP transmission channel cannot be of guaranteed quality.

---

1 Examples are: G.729 was designed for lower complexity than G.728, and has higher delay (called algorithmic delay) for similar speech fidelity; 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), G.729a was designed for lower complexity than G.729, at expense of slightly lower voice fidelity [4].

2 The G.711 codec, as well as 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 with the µ-law countries (generally the international carrier taking the responsibility at the international/domestic interface). In codec engineering of IP-based voice, care should be taken not to overlook this requirement as it may have to be specifically included in some possible network configurations.

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

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, increased latency (see also 6.4.3).

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 (from 300 to 3400 Hz using a 8 kHz sampling frequency) to wideband (WB) quality (from 50 to 7000 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

Use of digitised voice 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.

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, i.e. divided into equal length sections which comprise the IP packet payloads. The length of each section is a multiple of the codec frame length.

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 and decreases voice quality. Examples of transmission bandwidth at the link layer are given in Section 6.4 of the i3 forum Tech Spec. [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 [5], 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 delay

A de-jitter buffer is required at the receive end to counter jitter introduced by queuing delay variations in the packet network. This enables a continuous playout of the de-packetised, coded digital signal into the decoder. This buffer is typically set equal to the packetisation period.

6.4.6 Combined effect of Delay factors

Minimum codec speech processing delay is

(frame length + look ahead) + frame length

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.

Increasing the packetisation period increases latency, thus reducing speech quality for end-to-end calls >150 ms (see section 7.5). Loading the frames of coded voice into IP packets is practically instantaneous, [7] Annex A, so that the additional latency is 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 + packetisation period) 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, the delay may increase over that given above. The maximum speech processing time is

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

The average delay would be obtained by using a multiplier of 1.5 prior to the N for a combination of the low and high speed access links.

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

6.5 Speech Fidelity

Preserving natural speech fidelity as much as possible is essential to satisfy users. Generally low bit rate codecs have an increased complexity (resulting in latency increase and more computation) to retain fidelity. In addition, they become optimised for speech (the "codebook" parameters are optimised for speech, see section 6.1.2) to maintain fidelity at lower bit rates[4] [5].

---

[4] 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.

[5] 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.

Codec	Frame size (ms)	Look-ahead (ms)	Typically used packetisation periods (ms)	Mean 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

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.

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

Voice codec basic features for the codecs cited in the *i3 Forum Technical Interconnection Model for International Voice Networks, Release 2 (May 2009)* [1] are presented in Table 2.

As codecs are developed it is common to conduct subjective tests (see section 7.1) according to ITU-T Rec. P.800 [8] on that codec, effectively an end-end connection where the only impairment is the codec. The values resulting from such tests (generally expressed as Mean Opinion Scores – MOS) 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 indicative values are included, where known, in Table 2.

6.6 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. Earlier generations of mobile codecs typically have lower voice quality than “wire-line” or fixed network codecs, and quality varied during a call. However most mobile codecs brought into service over the last few years have very good speech fidelity, 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.

A Narrowband codecs

Codec	Technology	Sampling Frequency	Audio Band	Bit Rate		Frame length	Packet length (a)	Look ahead	Min. IP Delay (b)	Mean one way delay ©	Max. one way delay ©	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+PLC	PCM	8	0,3 – 3,4	Fix	64	0,125	10	0	0,25	10,125	20,125	Yes	0,01	App II	0	25 (q)	Y			App I	92,3
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	0	4(q)				N	92,3
G.711+PLC	PCM	8	0,3 – 3,4	Fix	64	0,125	20	0	0,25	20,125	40,125	Yes	0,01	App II	0	25 (q)	Y			App I	92,3
G.711+PLC	PCM	8	0,3 – 3,4	Fix	64	0,125	20	0	0,25	20,125	40,125	Yes	0,01	App II	0		5,91	6/1.5	7	(p)	85,3
G.711+PLC	PCM	8	0,3 – 3,4	Fix	64	0,125	20	0	0,25	20,125	40,125	Yes	0,01	App II	0		7,84	8/2	10	(p)	82,3
G.711	PCM	8	0,3 – 3,4	Fix	64	0,125	20	0	0,25	20,125	40,125	Yes	0,01	App II	0	5 (r)	4			N	92,3
G.729	CS-ACELP	8	0,3 – 3,4	Fix	8	10	10	5	25	25	35	no	18	Ann B	10					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
G.729d	CS-ACELP	8	0,3 – 3,4	Fix	6,4	10	10	5	25	25	35	no	20	Ann B/F						Y	
G.729e	CS-ACELP	8	0,3 – 3,4	Fix	11,8	10	10	5	25	25	35	no	18	Ann B/G	4	8 (r)	4			Y	88,3
G.729	CS-ACELP	8	0,3 – 3,4	Fix	8	10	20	5	25	35	55	no	18	Ann B	10					Y	82,3
G.729a+VAD	CS-ACELP	8	0,3 – 3,4	Fix	8	10	20	5	25	35	55	no	10.5	Ann B	11	19 (r)	Y			Y	81,3
G.729d	CS-ACELP	8	0,3 – 3,4	Fix	6,4	10	20	5	25	35	55	no	20	Ann B/F						Y	
G.729e	CS-ACELP	8	0,3 – 3,4	Fix	11,8	10	20	5	25	35	55	no	18	Ann B/G	4					Y	88,3
G.729e	CS-ACELP	8	0,3 – 3,4	Fix	11,8	10	20	5	25	35	55	no	18	Ann B/G	4	8 (r)	5,91	6/1.5	9	native	83,3
G.729e	CS-ACELP	8	0,3 – 3,4	Fix	11,8	10	20	5	25	35	55	no	18	Ann B/G	4	8 (r)	7,84	8/2	11	native	81,3

G.723.1+VAD	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
G.723.1	MP-MLQ	8	0,3 – 3,4	Fix	6,3	30	30	7,5	67,5	67,5	97,5	no	18-20	Ann A	15	16 (r)				Y	77,3
AMR	MR-ACELP	8	0,3 – 3,4	Var	4,75-12,22	20	20	5	45	45	65	no	16,7 WMOP	Y						Y	
G.726	ADPCM	8	0,3 – 3,4	Fix	16	0,125	10	0	0,25	10,125	20,125	Yes	~ 30	N	50					N	42,3
G.726	ADPCM	8	0,3 – 3,4	Fix	24	0,125	10	0	0,25	10,125	20,125	Yes	~ 30	N	25					N	67,3
G.726	ADPCM	8	0,3 – 3,4	Fix	32	0,125	10	0	0,25	10,125	20,125	Yes	~ 30	N	7					N	85,3
G.726	ADPCM	8	0,3 – 3,4	Fix	40	0,125	10	0	0,25	10,125	20,125	Yes	~ 30	N	2					N	90,3

B_Wideband Codecs

Codec	Technology	Sampling Frequency	**Audio Band**	**Bit Rate**		**Frame length**	**Packet length (a)**	**Look ahead**	**Min. IP Delay (b)**	**Mean one way delay ©**	**Max. one way delay ©**	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 Narrow band low delay mode (8, 12 kbit/s)	EV-CELP +TDBWE+ MDCT	8	0,3 – 3.4	Var	8 12	20	20	5	25	25	45	no	14,48 WMOPS (8 kbit/s) 17.30 WMOPS (12 kbit/s)	Y							
G.729.1 Wideband low delay mode (14kbit/s)	EV-CELP +TDBWE+ MDCT	8 or 16	0,3 – 3.4 or 0,5 – 7,0	Var	14	20	20	5 (low delay mode)	28.9375	28.9375	48.9375	no	24 WMOPS	Y							
G.729.1 (I)	EV-CELP +TDBWE+ MDCT	8 or 16	0,3 – 3.4 or 0,5 – 7,0	Var	14, 16, 18, 20, 22, 24, 26, 28, 30, 32	20	20	28,9375	68,9375	68,9375	88,9375	no	36 WMOPS	Y							

G.729.1 (l)	EV-CELP +TDBWE+ MDCT	8 or 16	0,3 – 3.4 or 0,5 – 7,0	Var	24	20	20	28,9375	68,9375	68,9375	88,9375	no	36 WMOPS	Y	16	3					113
G.729.1 (l)	EV-CELP +TDBWE+ MDCT	8 or 16	0,3 – 3.4 or 0,5 – 7,0	Var	32	20	20	28,9375	68,9375	68,9375	88,9375	no	36 WMOPS	Y	7	6					122
G.722	SB-ADPCM	16	0,5 – 7,0	Fix	48	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	31					N	98
G.722	SB-ADPCM	16	0,5 – 7,0	Fix	56	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	20						109
G.722	SB-ADPCM	16	0,5 - 7,0	Fix	64	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	13	7				App III	116
G.722	SB-ADPCM	16	0,5 - 7,0	Fix	64	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	13	5				App IV	116
G.722	SB-ADPCM	16	0,5 - 7,0	Fix	64	0,125	20	0	0,25	20,125	40,125		10 MIPS	N	13					N	116
G.722.1	MLT	16	0,5 - 7,0	Fix	24	20	20	20	60	60	80	no	< 5.5 WMOPS	N	19					N	110
G.722.1	MLT	16	0,5 - 7,0	Fix	32	20	20	20	60	60	80	no	< 5.5 WMOPS	N	13						116
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	6.6 – 23.85	20	20	5	45	45	65	no	39 WMOPS (s)	Y							
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	6,6	20	20	5	45	45	65	no	(t)	Y	41						88
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	8,85	20	20	5	45	45	65	no	(t)	Y	26						103
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	12,65	20	20	5	45	45	65	no	(t)	Y	13	4					116
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	15,85	20	20	5	45	45	65	no	(t)	Y	7						122
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	23,05	20	20	5	45	45	65	no	(t)	Y	1	5					128
G.722.2 (k)	ACELP	16	0,5 - 7,0	Var	23,85	20	20	5	45	45	65	no	(t) (u)	Y	8	5					121

Terms used in the table

1. Var — means Dynamically Variable bit-rate during a call. Some fixed codecs have different Bits 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.
4. Min Latency — is the Algorithmic Delay, the lowest time for a bit input to the encoder to emerge from the decoder algorithm. This is implementation independent.

Notes:

(a) Typically encountered packetisation rates only.
(b) Minimum IP Delay refers to the processing delay requirements of the encoder (send side) for a single frame in a packet as per G.114 A.2.3 [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 [9] 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 [10] 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 [11] par 3.5 & G.113 [10] Appendix I for specific limitations and conditions. For WB codecs, Bpl assumed in diotic listening context.
(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 [3].
(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 comparitive purposes.
(k) G.722.2 [12] is also known as AMR WB for mobile and is backward compatible with the AMR codec.
(l) G.729.1 [14] is also known as G.729EV and supports backward compatibility modes to G.729/a/b and a new Narrowband bit rate of 12Kbps. *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 [15] par 5.2 states Wideband algorithmic delay to be 42.875ms.
(n) G.711.1 [16] par 6.5 states Wideband algorithmic delay to be 11.875ms.
(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 [11].
(p) PLC type of "Repeat 1/Silence". Refer G.113 [10] 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 excpet 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.

Table 2 Basic voice codec features

7 Voice Quality Evaluation

7.1 Mean Opinion Score (MOS)

A commonly used scale is the Mean Opinion Score (MOS). MOS is a subjective value defined in ITU-T Rec. P.10/G.100 [17], 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 [8])
- 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 [18])
- parametric models, taking benefit of protocol information ; for IP Based voice, they must comply with ITU-T Rec. P.564 [19]

The audio bandwidth must also be taken into account:. Three contexts must be distinguished:

- narrow band only (e.g. MOS-LQON with P.862.1 [20])
- wide-band only (e.g. MOS-LQSW with P.800 [8] in wide-band context)
- mixed-band (e.g. MOS-LQOM with P.862.2 [21]).

ITU-T Rec. P.10/G.100 [17] gives all details about different MOS scales.

During subjective listening tests, listeners participate in a well balanced, subjective experiment [22], 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 [8]:

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 [23] 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 [24]. 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 on the stimulus set [3].

With properly designed experiments, wideband voice scores 0.5 to 1 MOS greater than narrow band voice (the G.711 PCM codec used as reference gets a MOS-LQSM score of 3.7 in mixed wideband and narrowband codec subjective test experiments). This is particularly important in view of the transcoding impairments presented in section 10.2.

7.2 E-Model

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 [11]. 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 [11]), 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[6] to estimate end-to-end voice quality.

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

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

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 [10] Appendix II and section 8.5.

Table 4. Impairments contributing to R-Factor.

6 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.

The term *Ro* and the *Is, Ie* and *Id* values may be subdivided into further specific impairment values. Further detail is in [11], in [5] and in section 8.

7.3 E-Model Relationship to MOS

The R-Factor can be transformed into estimates of customer opinion factors, such as MOS. When estimated from the E-model using the following formula it is called MOS Communication Quality Estimated or MOS_{CQE} [8], [25]:

For $R < 0$	$MOS_{CQE} = 1$
For $0 < R < 100$	$MOS_{CQE} = 1 + 0.035R + R(R - 60)(100 - R)\ 7\text{x}10^{-6}$
For $R > 100$	$MOS_{CQE} = 4.5$

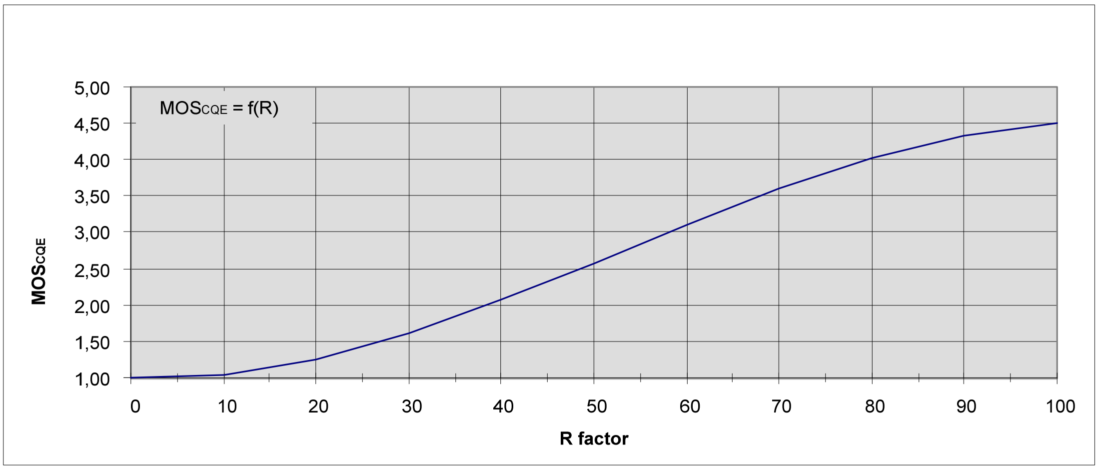


Figure 2 MOS_{CQE} = f(R)

7.4 Transmission Quality Category in the E-model

The R-Factor is related to User Satisfaction and to Speech Quality Transmission Category as shown in Figure 3 [26]. Customer opinion score estimates, MOS_{CQE}, are also indicated.

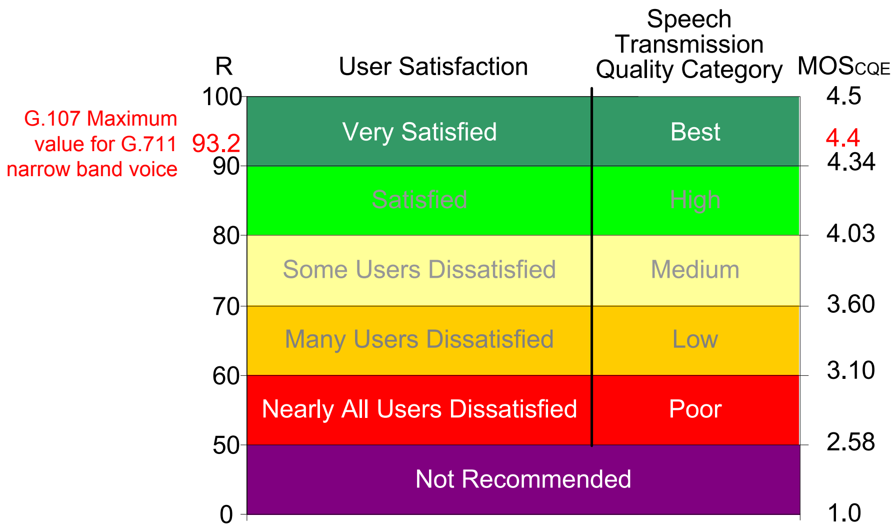


Figure 3 Classification of speech quality for different R-Factors

Note that the classifications in Figure 3 are for convenience only; the range of speech quality is actually a continuum; ref [26] 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 speech (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 13) gives data to construct this graph.

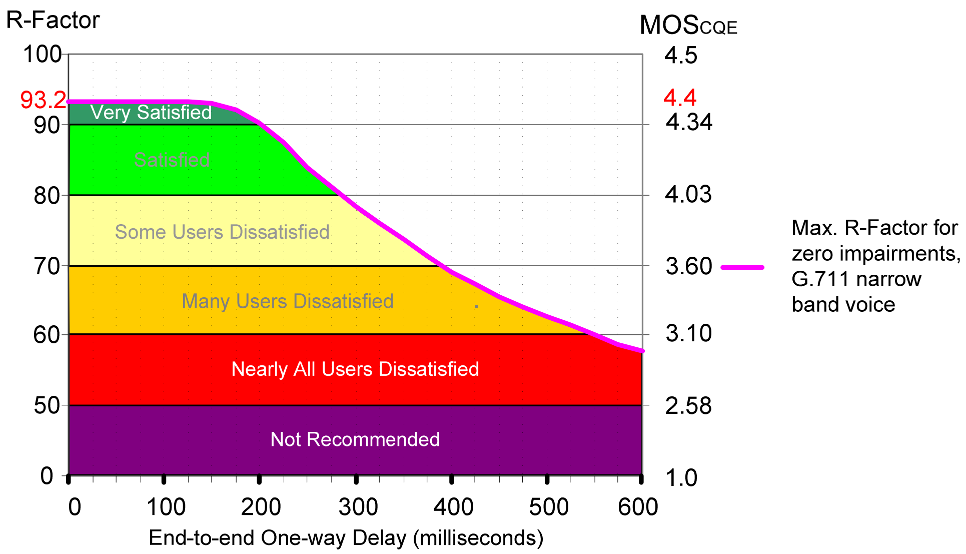


Figure 4 Maximum R-Factor vs absolute one-way delay for narrow band speech

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 [11] 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; [8] 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 [19] 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 this White Paper accounts for narrow band (NB) voice transmission only. The R-factor scale is now extended to support wideband (WB) voice (7kHz audio bandwidth) by extending the R-scale to R=129, [3] in a way which leaves the narrow-band use of the scale unaffected, including the position of the reference connection[7]. This scale extension occurs because, for wideband transmission (defined as 50-7000Hz) the quality is generally[8] judged better than for a narrow band channel [3].

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 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*." [27].

The Ie,wb values are derived from subjective test results and from objective measurements (according to methods described in 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).

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: [27] 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 material. The remainder of this White Paper uses the NB model to illustrate the impact of transcoding.

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)
- associated packetisation period – pp
- packet loss
- international propagation delay (latency)
- domestic/access (Service Provider) network latency
- transcoding.

---

7 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.

8 E.g there can be occasions where some noise is more objectionable when presented in a wideband channel.

Of the parameters in the E-model (formula in section 7.2), ITU-T Rec. G.108 [5] 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 [5] table 1 and following list, ITU-T Rec. G.108 [5] 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.[9]

Ro is set to the 93.2 default value when evaluating codec impairments.

8.2 E-Model Parameter *Is*

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

If *Ie* is used (as in this paper), the qdu impairment is not to be used [5].

The loudness ratings [11], [5] 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 conditions to evaluate codecs (section 11) 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, [11], 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 11.2.5.

8.3 E-Model Parameter *Id*

Id represents all mouth-to-ear delay 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 were typically well within 10 ms 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 [28], interleaving in wireless access networks etc. Further delay can also be introduced where multiple services share the access, and

[9] The year 2000 revision of G.107 [11] 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.

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	Typically a frame length, commonly 20 ms
Customer	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 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 a 20 ms. de-jitter buffering). Codec latency is additional and is added to the send 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, as the routes average ~14% longer than great circle distance to ensure a safe seabed path for the cable. Geostationary satellite links contribute 260 ms to one-way propagation time.

Table 6 provides typical propagation delays for four 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
1. Intra Region, e.g Europe	40 ms
2. Inter Region, e.g Europe –USA	80 ms
3. Global, e.g. Europe - Pacific	160 ms
4. Global Worst case e.g global including satellite such as Europe to many Pacific Islands	420 ms

Table 6 Representative One-way Propagation delays in international networks

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

8.4.1 Codec Equipment

Ie allows for the impairments of codec distortion[10]. *Ie* is by its very definition independent of all the other impairment factors: it is only dependent on the digital process[11] whose perceptual characteristics it aims to model [23]. 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 [23]. 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[12]. 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.

8.4.2 Packet Loss

Packet loss removes speech samples or frames, increasing *Ie,eff*.[13] 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[14] limits the achievable robustness), and because generally (with G.711 say, with typical packetisation periods) many speech samples are lost with one missing packet[15] [29], [30].

Application layer techniques called Packet Loss Concealment (PLC) are commonly used to mitigate the effect of packet loss; these use information on the speech signal from either side of the "gap" to interpolate a representation of the missing signal [29], [30].

Packet loss impairment is different for each codec, [5], Table 2b, and varies with network load and packetisation period[16] (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[17], see Figure 5.

---

10 Codecs distort speech, the impairment is a measure of the user perception of its effect.

11 For non-waveform codecs the encoding process is non-linear.

12 An example is the *Ie* for G.729, for which the initial provisional value was *Ie* = 15 [31], which then became provisionally *Ie*=12 in the 1998 version of G.107 [32] and was changed to the current value of *Ie* =10 as that data was removed from G.107 to G.113 late in 1998 [12].

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

14 The effect of the loss of one frame can propagate over several consecutive frames.

15 E.g. For G.711/20ms, 160 consecutive samples are lost.

16 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.

17 Impairments are generally slight below 0.5% packet loss for low bit rate codecs, [5] table 2b, [30] Table 2 and Figure 5 of this White Paper.

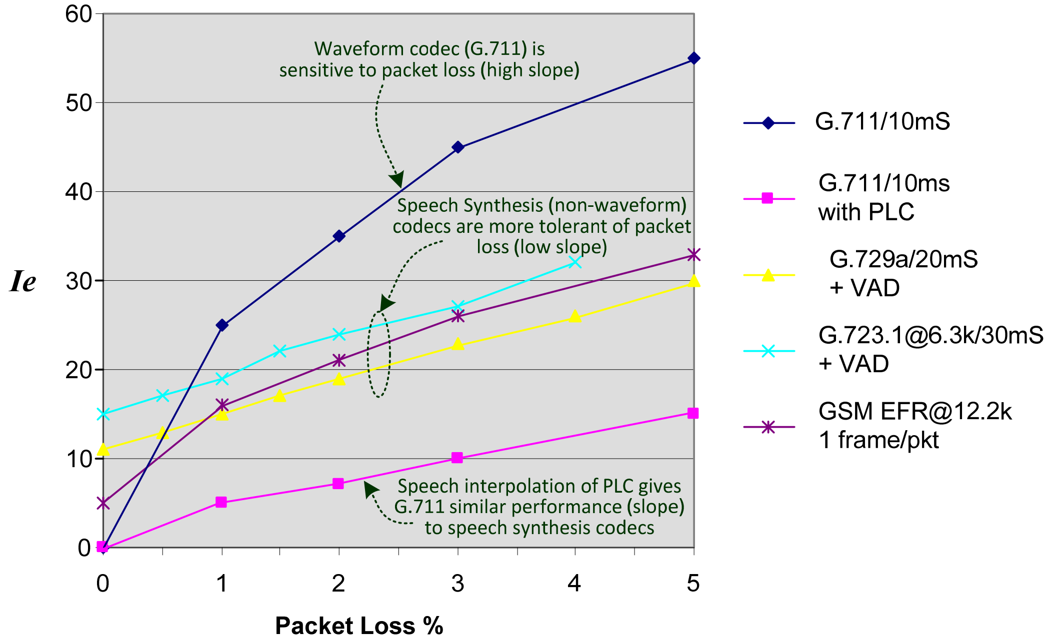


Figure 5 Distortion impairment as a function of packet loss for several codecs

Recent codecs designed for lossy packet networks (often called frame erasure channels) are more tolerant to packet loss. These have specific application in internet 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 [5] 7.8, and ITU-T Rec. G.107 [11] 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 [5]

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 mainstream[18], A tends to decrease, [12], Appendix II. Table 7 gives absolute upper limits, see [11], 3.6.

9 Transcoding in the E - model

Transcoding is defined [30], 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 linear PCM segments. 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 [33] - 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 process.

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.729a is a commonly used codec with good balance between bandwidth, speech fidelity, and latency, and is favoured by i3 Forum members for fixed networks [1].

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 [4].

In contrast, G.726 is a simple transcoder (ADPCM) which when decoded to G.711[19], 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.711 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 degrade with successive transcoding.

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 [23], 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 coders), 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[20], 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.

---

18 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.

19 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 coding standard that are identical if synchronous transcoding is invoked.

20 No material seen so far indicates it could be lower.

9.2 Codec Transcoding Issues – G.729

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

Some data is available on the G.729 codecs transcoding performance. P.833 [23] 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 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, which is listed as

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 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 [32] and was changed to *Ie* =10 as that data was moved from G.107 to G.113 [12]. Prior to 1998 the provisional value was *Ie*=15 [31].

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[21], and the inability to signal codec policy end-to-end[22] when multiple carriers are involved in a call, means that multiple transcodings can readily occur. Thus network planners should be vigilant given the significant impairments that tandemed codecs can cause.

9.3 Packetisation during Transcoding

When transcoding of an IP signal occurs, the digital IP signal must first be de-packetised to reconstruct the continuous coded digital signal, which introduces buffering latency. The recovered continuous signal is then transcoded (decoded to G.711 or linear PCM, 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

[21] Extensive testing in tandem configurations to ITU Rec. P.834 [34] 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.

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

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 translation.

9.4 Mobile Transcoding

Mobile SP's often transcode mobile-mobile calls within their network as the dynamically variable codecs may not match for Caller A and Caller B due to differing radio path conditions to/from their respective base stations. If a match is possible, Tandem Free Operation (TFO) is often invoked, which means the particular codec is "tunneled" through the 64K channel of the TDM switches[23] 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.

Mobile-fixed calls are transcoded. 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 = 30, [5] Table 2c, (although such high values usually exist for a short time only).

Mobile SP's currently want G.711 interconnections to extend TFO through intermediate networks, which requires a higher bandwidth international interconnection. It will be unlikely to find G.711 extensively used in future 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 [35] and RFC4348 [36] for AMR WB and VMR WB codecs respectively.

10 Impact of Transcoding using E-model

10.1 Single codec

An example of a high level estimate of the R-Factor derived by considering the contribution of the above factors of international call quality is given in Figure 6. The methodology used 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
- Nil transcoding
- Latency of four typical network distances from Table 6

[23] The means the codec signal is made to look like G.711 so the intermediate switches handle in the usual way, except transcoding impairments do not occur.

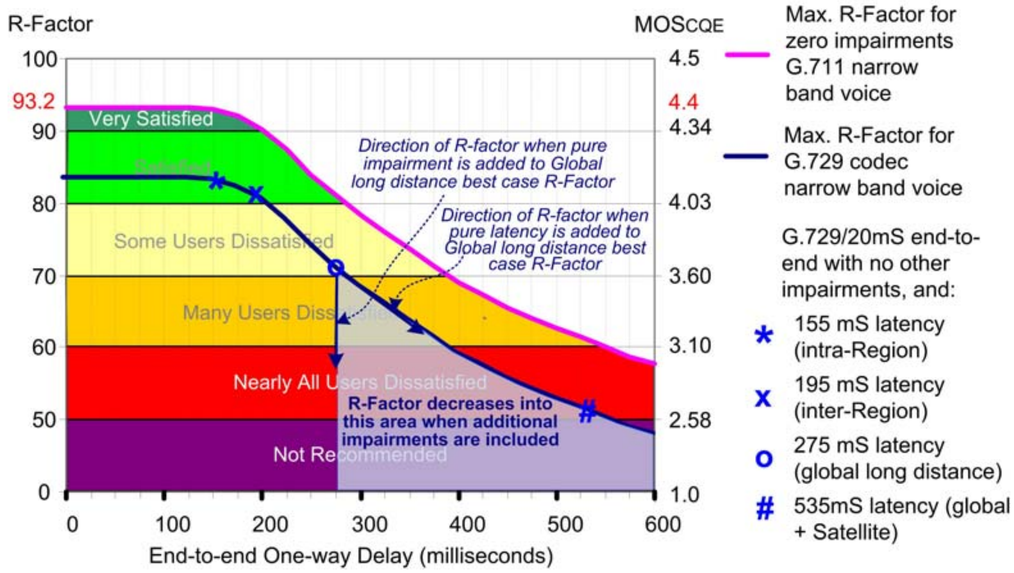


Figure 6 Best Case R-factor for international voice call with G.729 and several call distances

It is important to note that when other impairments are added to this best case estimate, <u>the R-Factor can only decrease</u> and (for the example of a global long distance call) moves into the shaded zone on Figure 6, i.e. the quality can be <u>no higher</u> than indicated 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 by customers.

Since G.729 is coded from linear PCM[24], 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.

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.
2. Service Providers assumed to use same codec as domestic network operators.
3. The international operator may have to undertake A-law to/from µ-law conversion if Service Providers have differing companding standards.

Table 9 Network situations applicable to Figure 6

[24] G.711 without companding.

10.2 Transcoding

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 [11] but the effect is known to be small, and is presently disregarded to preserve the simple additive nature of the E-model [11].

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 [37], 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.

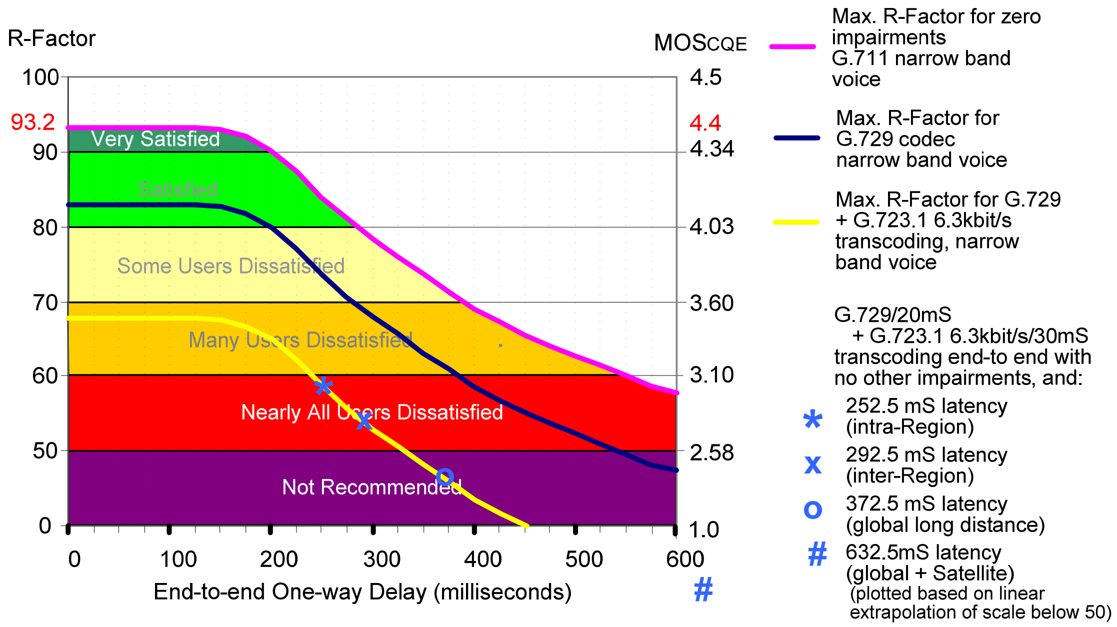
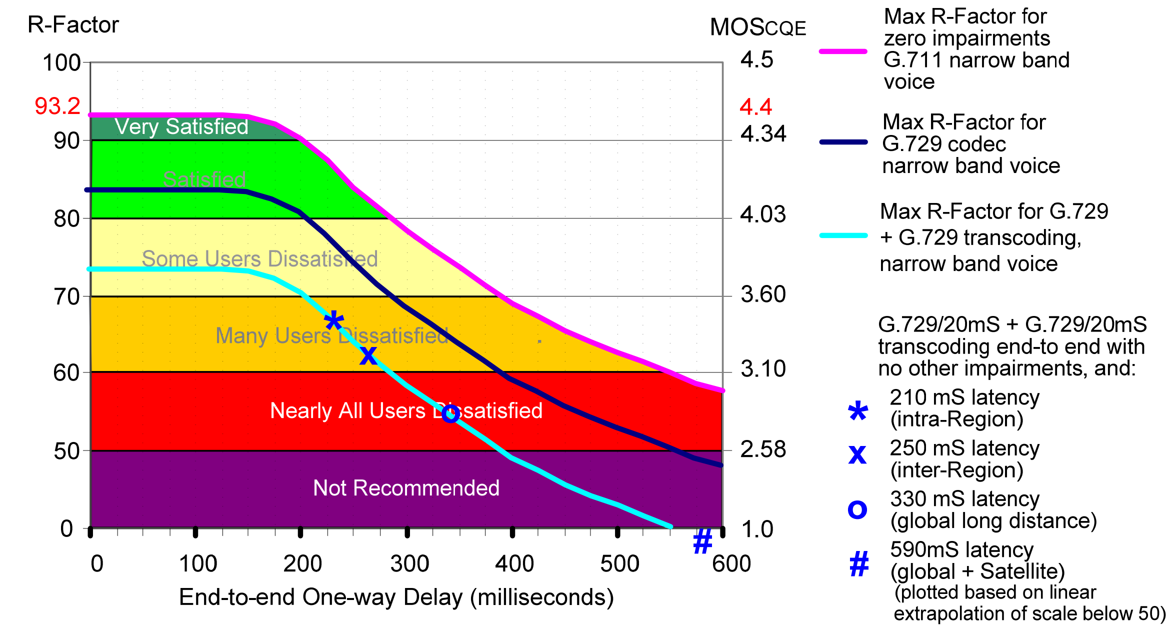


Figure 7 Illustrating the high impact on voice quality of adding a transcoding to G.723.1 to the Best Case R-factor for international voice call of Figure 6.

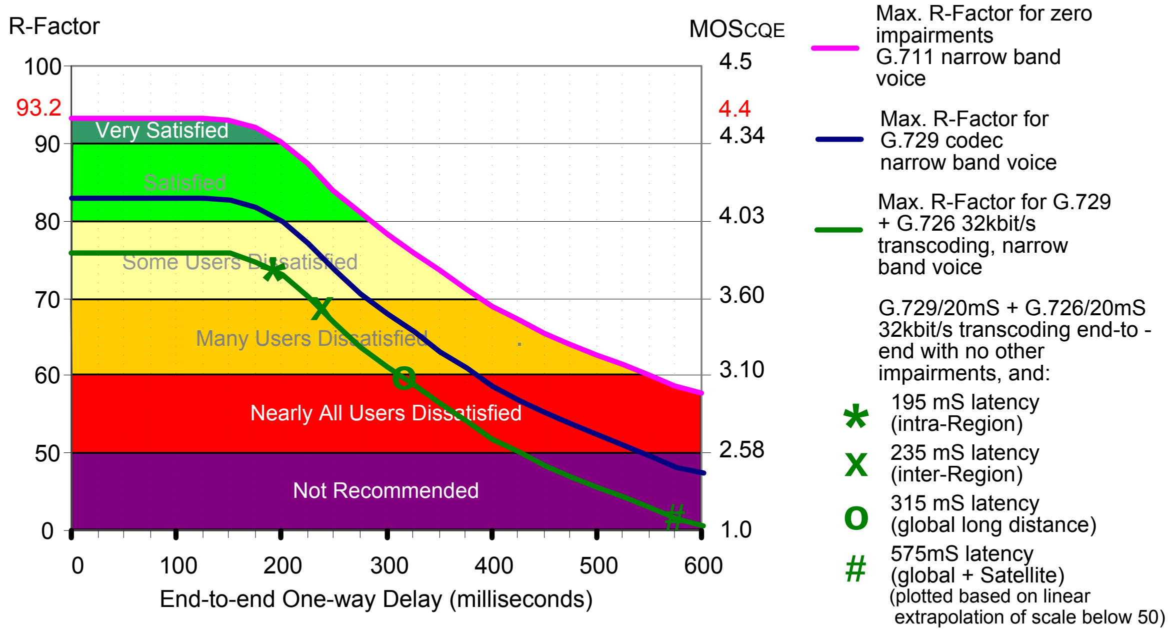
This illustration applies to a direct bilateral configuration where Domestic Operator A and Domestic Operator B do not use the same codec (specifically A uses G.729 and B uses G.723.1), or when an intermediate carrier transcodes to G.723.1 to save bandwidth in an otherwise all G.729 configuration (and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

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.

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.



(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

Using codecs with lower delay and lower *Ie* impairment value is seen to increase end-to-end quality compared to Figure 7. Note also that quality would increase further if the packetisation period was lowered to 10 ms, the saving of 20 ms ≈3 R "points".

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 [38], 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 [5]). Other predominant DCME types used the G.726 @ 32kbit/s ADPCM codec [9] (also Ie=7), which, when the DCME is not highly loaded so that temporary bit robbing reduces the codec bit rate, 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.

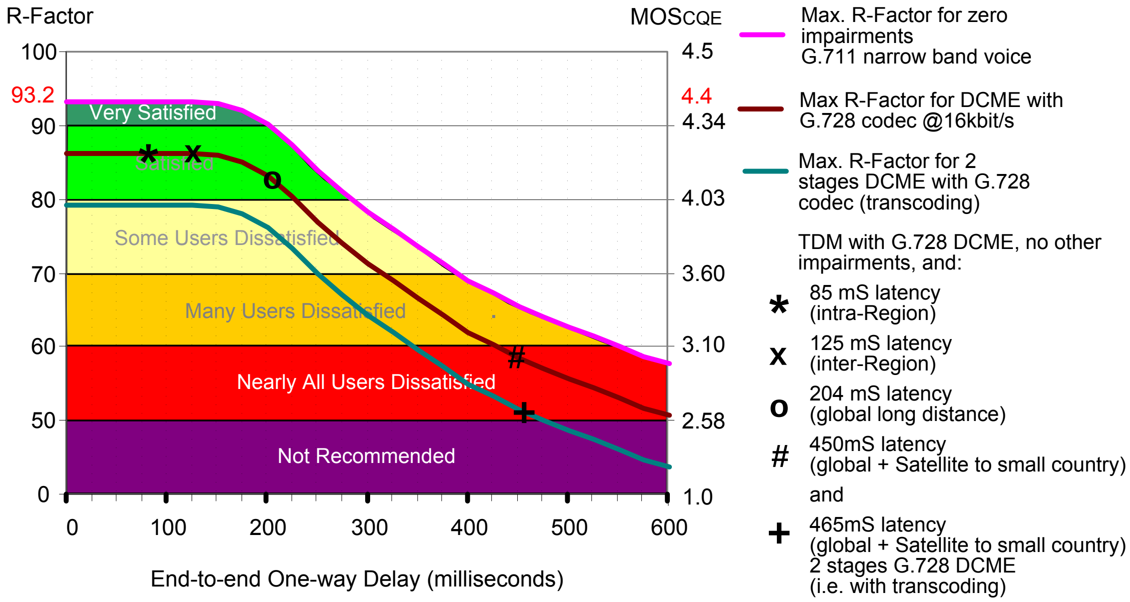


Figure 9 Best Case R-factor for international voice call on TDM network with DCME using G.728 codec with VAD, and several call distances, plus one example of two stages of DCME (transcoding)

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

Given the popularity of G.729, it would not normally be expected that multiple transcodings of G.729 would occur as all carriers can be expected to support it (it is a mandatory codec for compliance with i3 Forum recommendations [1]). However for "codec" negotiation to succeed the packetisation period also has to match and when, say, there is incorrect soft switch setup, the fall-back codec G.711 could be invoked in an intermediate network. Under such conditions, a multiple G.729 coding, with adverse call quality, could 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 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 interconnection with domestic fixed operators (at a transcoded interface) and the international call is passed through an inappropriately transcoded international configuration. 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 (Ie =7, air-path delay = 14 ms) is used in current generation DECT handsets. 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. The quality level customers are willing to accept to preserve mobility in their homes with terminals they have already used satisfactorily with the TDM PSTN will be interesting.

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 11.2.4. 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 [29] 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. 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[25], thus it is possible that quality improvement over time could be achieved if carriers support the use of developing codecs in that family.

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 carriers to consider supporting mobile codecs, particularly AMR-WB or G.722.2 [39], in fixed Next Generation Networks to maximise call quality outcomes.

---

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

10.5 Unsuitability of G.723.1 Codec in International Carrier Networks

The G.723.1 codec [37], for a small (~3kbit/s) bandwidth (transmission cost) reduction, has such high fidelity impairment 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 i3 Forum carriers take every opportunity to eradicate this codec from general use in IP based voice international networks.

11 Evaluation of Codec Choice in International IP Interconnections

11.1 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 to save capacity costs regardless of ingress and egress carrier codec primary offers[26], 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).

Generally, configurations with a number of carries 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.

---

[26] In the SDP part of SIP signalling

11.1.1 Bilateral Interconnection Configuration

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

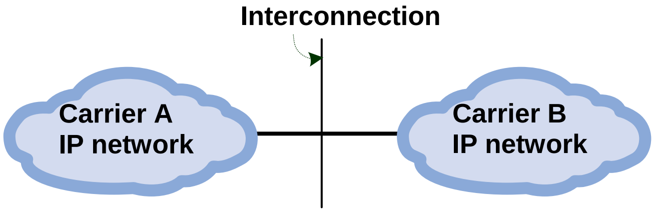


Figure 10 Bilateral interconnection configuration

11.1.2 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 providing the 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, low frame rate) to minimise compounding impairments that may already have occurred “downstream”.

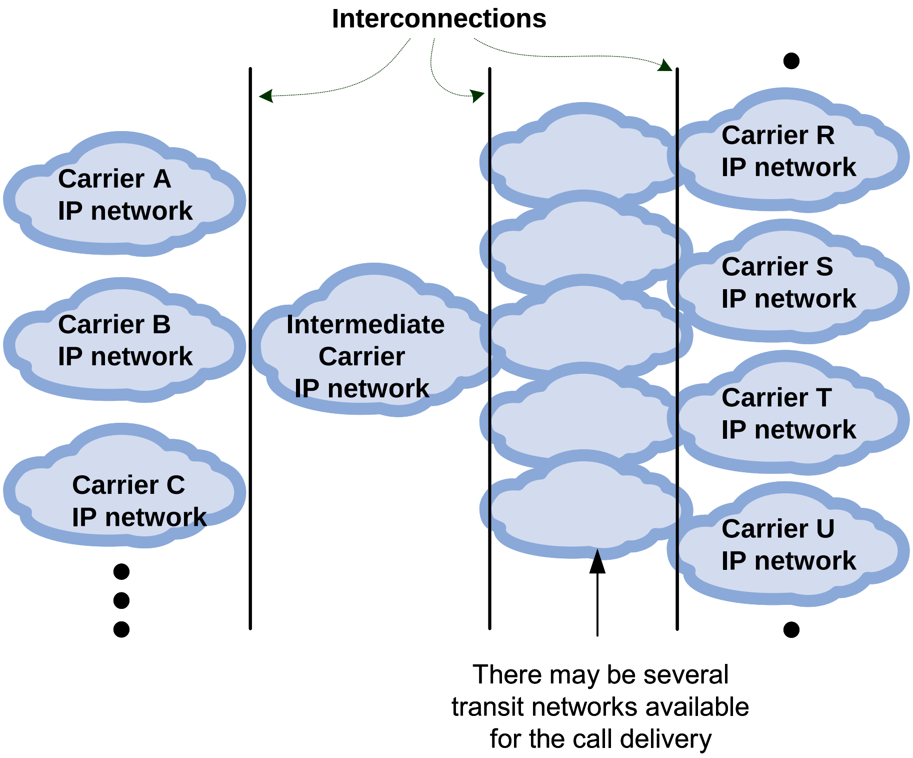


Figure 11 Configuration with Multiple Carriers end-to-end.

11.2 Calculation Example

11.2.1 Assumptions

Evaluation of configurations may be made using the E Model as defined in ITU-T G.107 [11] (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.

11.2.2 Determination of Reference Configuration

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

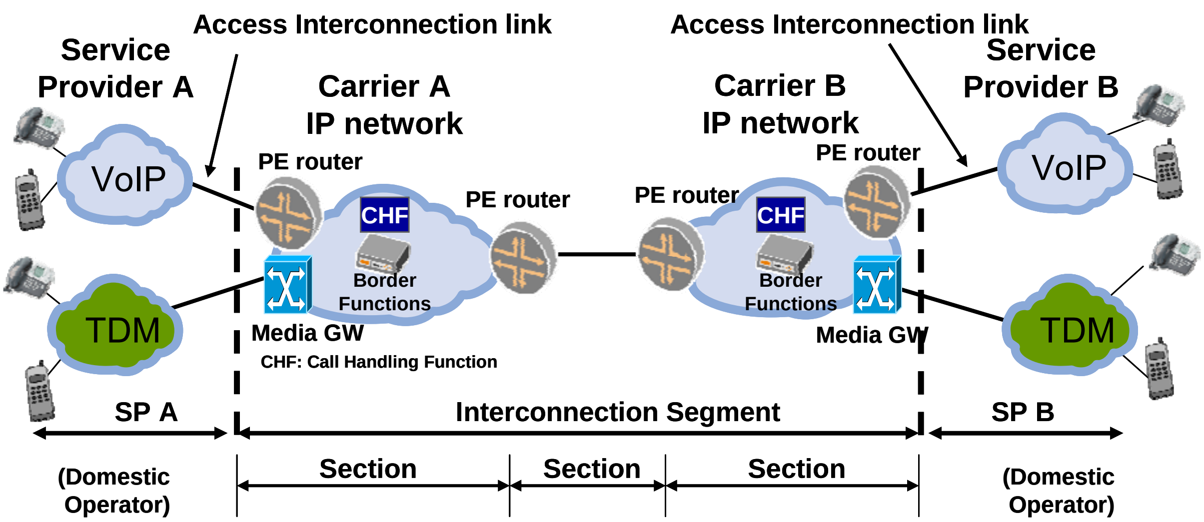


Figure 12 Reference configuration

11.2.3 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 digitalised 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 [10] (11/2007) Appendix 1 Tables I.1 thru I.5, noting that packet loss is characterised as "random" or "bursty".

11.2.4 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 [5].

Example of calculation:

Section 1 Service Provider A network (sending)						
Impairments			**Impairments**	**Delay**		
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				20	ms	
Others					ms	
Total Impairments & Delay, Carrier B			10	30	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				20	ms	
Others				0	ms	
Total Impairments & Delay, Service Provider B			0	40	ms	
Total Service Provider evaluation						
Total Impairments & Delay, Service Provider A & B			0	80.375	ms	
Total international and domestic evaluation						
Total Impairments & Delay, Carrier A & B + International			10	120	ms	
Total end to end evaluation						
Total Impairments & Delay, end-to-end			**10**	**200.375**	**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

11.2.5 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 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 [40]
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 13.

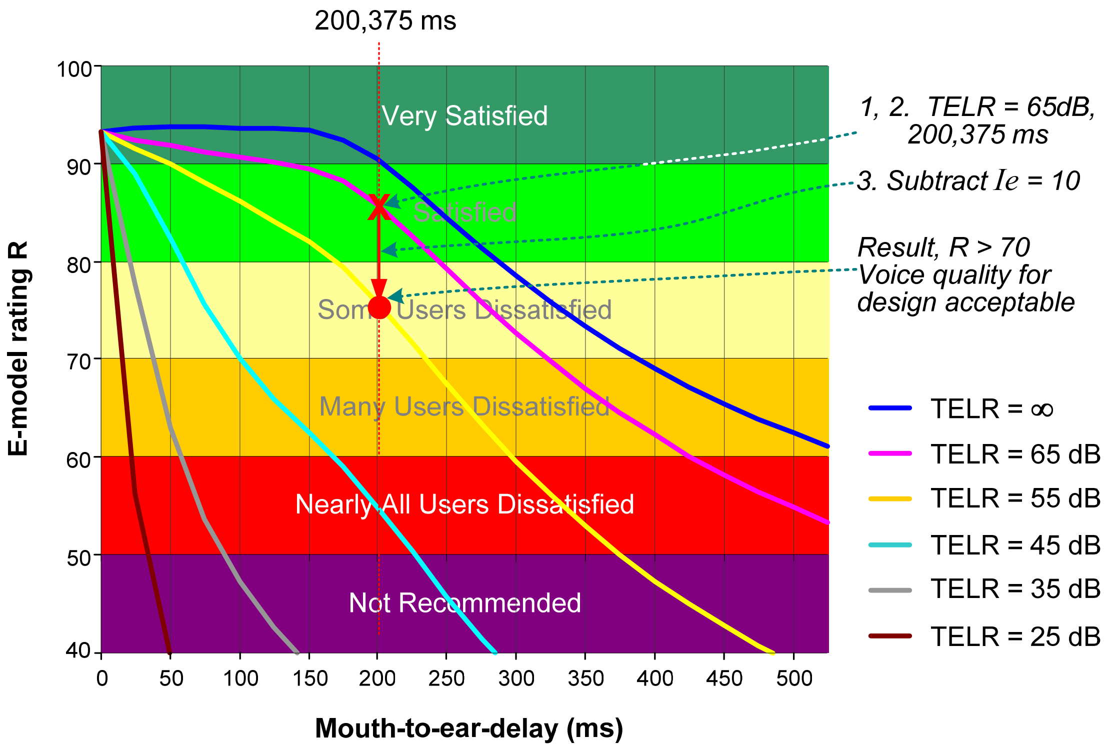


Figure 13. Calculation Example Result shown on R factor as a function of Total Delay and Talker Echo Loudness Rating (TELR) Graph.

Thus, in this example, using the "default curve TELR=65 dB" and total delay = 200.375 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, 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 (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 *http://www.itu.int/ITU-T/studygroups/com12/emodelv1/*.

12 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 "*Users Satisfied*" 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.
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 higher intrinsic fidelity, 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.

12.1 Recommendations on Codec Choice

8. 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.
9. Packet loss should be kept as low as possible (total packet loss < 0,1%) so that its influence on voice quality may be neglected. 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.
10. 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 G.711 in the remainder of the configuration).
11. The G.729 codec family offers a good balance of latency, bandwidth (cost) and voice fidelity.

12. Care is needed in an end-to-end IP based voice design, to ensure that the appropriate A-law to μ law conversion is included, where applicable.
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 higher intrinsic fidelity 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.

12.2 Transcoding

15. Transcoding of low bit rate codecs greatly decreases international call quality, especially on long connections, and <u>should be avoided unless absolutely necessary</u>.
16. 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.
17. 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. Favoring 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.

12.3 Call Setup

18. Order of codec/packetisation period preference is determined by the originating terminal and should be honoured where possible.
19. 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 interfacing international carrier doing the companding conversion.
20. 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.
21. In case of fixed-mobile interconnection, transcoding if necessary shall always be performed by mobile network.

13 Appendix 1 Table of maximum R-Factors for narrow band speech (G.711 PCM coded)

Absolute Delay (ms)	R-Factor	Absolute Delay (ms)	R-Factor	Absolute Delay (ms)	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				