INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.83

(03/93)

TELEPHONE TRANSMISSION QUALITY

SUBJECTIVE OPINION TESTS

SUBJECTIVE PERFORMANCE ASSESSMENT OF TELEPHONE-BAND AND WIDEBAND DIGITAL CODECS

ITU-T Recommendation P.83

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation P.83 was prepared by the ITU-T Study Group XII (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

2 In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

CONTENTS

SUBJECTIVE PERFORMANCE ASSESSMENT OF TELEPHONE-BAND AND WIDEBAND DIGITAL CODECS

(Helsinki, 1993)

1 Introduction

Digital telephony is now being introduced in the Public Switched Telephone Network (PSTN) at an increasing speed. Since the 1960s, digital transmission in the PSTN has been synonymous to 64 kbit/s, A-law PCM (see Recommendation G.711). Now the combination of better knowledge in signal processing techniques and advances in technology, most notably LSI/VLSI techniques, has lead to an increased interest in more efficient coding methods than PCM. One example of this development is the introduction of 32 kbit/s ADPCM (see Recommendation G.726). More recently a 64 kbit/s, 7 kHz wideband codec has been standardized by the CCITT (see Recommendation G.722). Considerable interest is now focused on Low Rate Encoding (LRE) i.e. at bit rates of 16 kbit/s and/or below.

When the transmission path is digital and/or non-linear, simple objective measurements such as Recommendation G.712 are insufficient to ensure adequate transmission performance. The aim of a subjective testing methodology is to measure the degradation contributed by the non-linear part of the transmission path, so that the performance of the complete system is satisfactory. To be suitable for this purpose, the measurements have to be:

a) reliable; and

b) carried out in a way that takes account of major interactions between the non-linear part and the other parts of the transmission system.

This implies both the ability to uniquely assign a numerical contribution to each digital process and the ability to use this assigned contribution in conjunction with other impairments to estimate telephone connection performance.

This Recommendation, based on Annex B/P.80 and experience gained from several international experiments [2, 3, 4, 5, 6], defines a specific testing method for evaluating digital processes in a manner such that the quantization distortion effects of these processes on transmission performance can be taken into account in the evolving international network.

A variety of methods are possible to characterize the performance of digital processes. At the present time these comprise listening-only tests involving:

1) opinion (category) ratings;

2) pair or multiple comparisons; and

3) articulation tests.

For most applications the CCITT recommends the use of the Absolute Category Rating (ACR) method using the Listening Quality scale. However, there are times when other scales and rating methods are more suitable and appropriate and these are used as well in this Recommendation. Only where there is a deviation from the use of the ACR method using the Listening Quality scale will it be stated.

NOTE – The Degradation Category Rating (DCR) method [1] is described in detail in Annex D/P.80. This method is purported to be suited when digital impairments are small. It may therefore be particularly useful to serve for system optimization. It should also be noted that a threshold method for direct comparison, described in detail in Annex E/P.80, is again applicable for system optimization.

This Recommendation contains advice on how to access the performance of digital codecs. This Recommendation must be read in conjunction with Recommendation P.80.

Furthermore, there is the need for supplementing listening-only tests with conversation tests but as yet no method can be agreed for inclusion in this Recommendation.

2 Source recordings

This clause is the same as B.1/P.80 with the following exceptions.

2.1 Sending system

For narrow-band systems (300-3400 Hz) it is recommended that if comparison of results is required between laboratories and administrations that the sending end of the Intermediate Reference System (IRS) conforming to Recommendation P.48 and calibrated to Recommendation P.64 be used. However, it has been found that the performance of low-bit rate speech codecs may depend significantly on the frequency characteristic applied to the input speech signal.

NOTE –The IRS is representative of analogue telephone sets (1970s) and Study Group 12 is currently studying a "Reference Telephone" that is representative of modern telephone sets (see Figure 1/P.31).

If a wideband system (100-7000 Hz) is to be used for audio-conferencing, then the sending end should conform to IEC Publication 581.7.

2.2 Speech samples

It is recommended that a minimum of two sentences per sample is used. This allows, for example, for bit-error-rates (BERs) of 0, 1:10000, 1:1000 and 1:100 to be assessed (see 3.1.4). BERs between 0 and 1:10000 will require more sentences/sample.

3 Selection of experiment parameters

3.1 Codec conditions

3.1.1 Speech input levels

When assessing codecs it is usual to set the input levels to be relative to the overload point of the codec. It is recommended that input levels of 14, 26 and 38 dB below the overload point of the codec (equivalent to –8, –20, –32 dBm0) should be used. This is approximately equal to the mean value measured at the International Switching Point ± two standard deviations. Other applications may require different input levels. Figure 1 demonstrates the effect of change in speech input level.

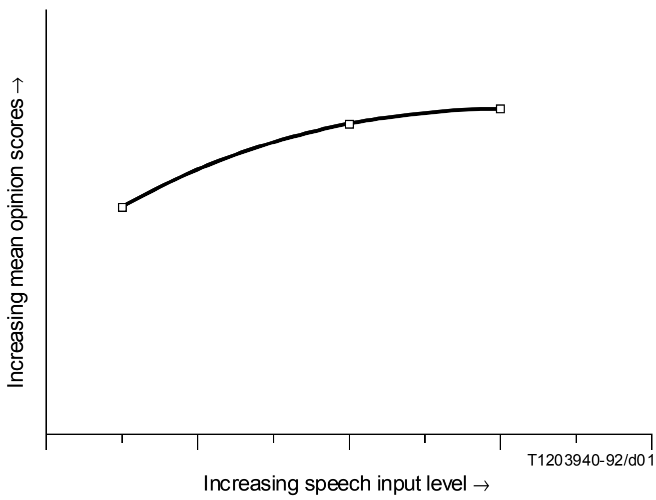

FIGURE 1/P.83

Mean opinion scores for speech input levels

3.1.2 Listening levels

It is recommended that at least three listening levels should be used. These levels should lie 10 dB either side of the preferred listening level (see B.2.1/P.80 for information on preferred listening level). Figure 2 demonstrates the effect of change in listening level.

3.1.3 Talkers

3.1.3.1 Different talkers

It is recommended that a minimum of two male and two female talkers be used. Figure 2 demonstrates the effect of different talkers. However, if talker dependency is to be tested, it is recommended to use the following talkers:

- 8 male;
- 8 female;
- 8 children.

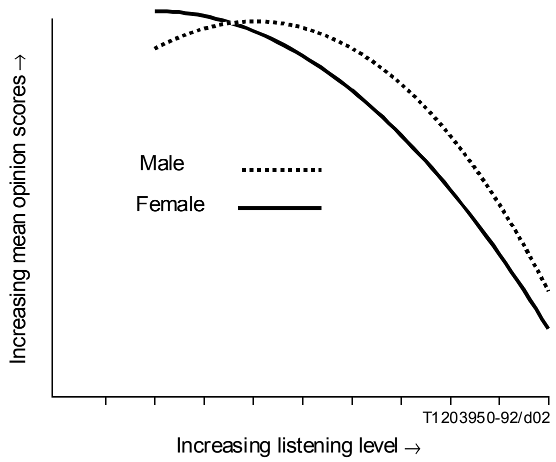

FIGURE 2/P.83

Mean opinion scores for listening levels

3.1.3.2 Multiple talkers

It is important to know the robustness of the codec to multiple voice input signals and to find out whether any adverse effects occur such as "break-up" of transmission, spurious signals, etc.

If only handset use is to be considered, then the mixing of two different talkers with input speech levels 20 dB apart is normally sufficient.

However, if it is to be used in a conference mode, i.e. hands-free, where more than one talker can speak at the same time, then one must ensure that the encoding algorithm can deal with multiple talkers where the difference in speech levels could be zero.

Either the Degradation Category Rating (DCR) method using the 5-point scale of 5.2.3 (see also Annex D/P.80) or the Quantal-Response detectability method using the 3-point scale of 5.2.4 (see also Annex C/P.80) are recommended for assessing the effects of multiple talkers.

3.1.4 Errors

This will depend on the application, i.e. whether it is used in the normal ordinary PSTN (i.e. line systems) or in say, a radio environment such as mobile radio.

It is recommended that for line systems randomly distributed errors should be used with bit-error-rates (BERs) in the range 0 to 1:1000. Of course, this will be dependent on the number of sentences used in a sample (see 2.2). In certain circumstances it may be necessary to test down to 1:100.

For other applications such as mobile radio the errors may well be of the burst-error type and therefore it would be appropriate to use errors of this type.

It is usual to use the ACR method using the Listening Quality scale of 5.2.1. If, however, the condition is expected to be of poor quality then the ACR method using the Listening Effort scale of 5.2.2 may be more appropriate.

3.1.5 Bit rates

The codec must be tested at all the bit rates that it is capable of working at e.g. for G.722 this is 48 kbit/s, 56 kbit/s and 64 kbit/s and for G.726 this is 16 kbit/s, 24 kbit/s, 32 kbit/s and 40 kbit/s. However, operation at certain bit rates may be dependent on operational conditions and system loading may be applicable (see Recommendation P.84). Figure 3 demonstrates the effect of change in bit rate.

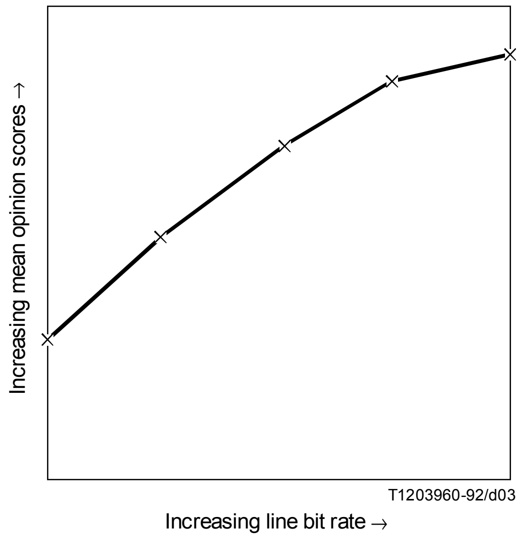

FIGURE 3/P.83

Mean opinion scores at different line bit rates

3.1.6 Transcodings

It is usual to assess a codec in association with an A- or μ-law companding process. Hence a single codec going analogue to analogue will have the configuration shown in Figure 4.

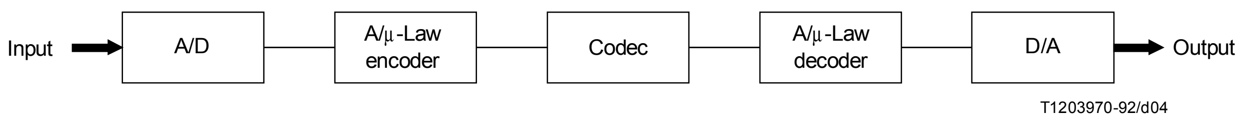

FIGURE 4/P.83

Transcoding scheme

3.1.6.1 Asynchronous tandeming

This method reflects the use of the codec in the network, based on a 64 kbit/s architecture. Therefore each transcoding includes additional quantization distortion produced by the linear to A- or μ-law companding process, additional attentuation distortion from tolerances in the anti-aliasing and reconstruction filters, plus accumulated idle-channel noise. This configuration is demonstrated in Figure 5.

The CCITT recommends that 1, 2 and at least 3 codecs are tested in tandem.

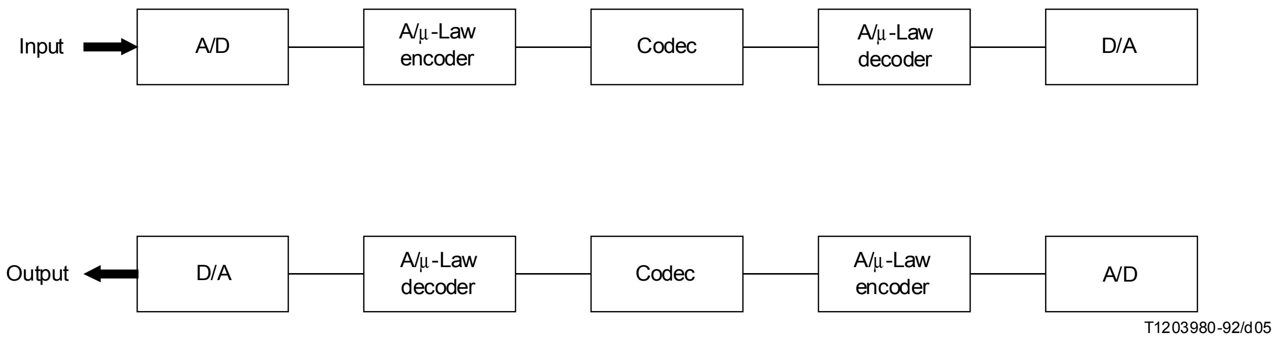

FIGURE 5/P.83

Asynchronous tandeming

3.1.6.2 Synchronous tandeming

The circuit configuration is shown in Figure 6 and has the advantage that there is only one A- or μ-law companding process and therefore has reduced quantization distortion when compared to 3.1.6.1.

The CCITT recommends that 1, 2 and at least 3 codecs are tested in tandem.

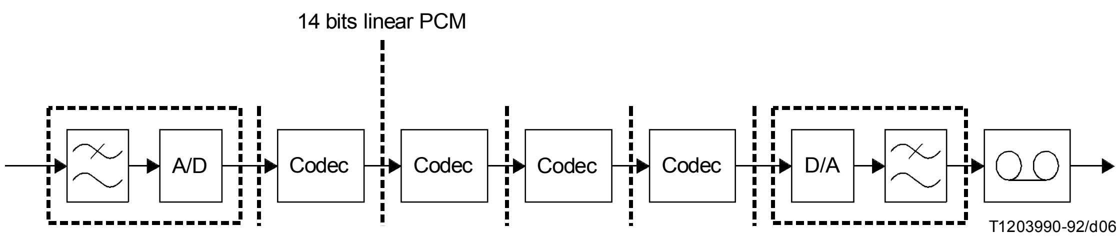

FIGURE 6/P.83

Synchronous tandeming

3.1.6.3 Mixed tandeming

It is important to establish the effects of tandeming systems that utilize encoding at different bit rates and a configuration is shown in Figure 7. It is essential that combinations of the most probable combinations are tested.

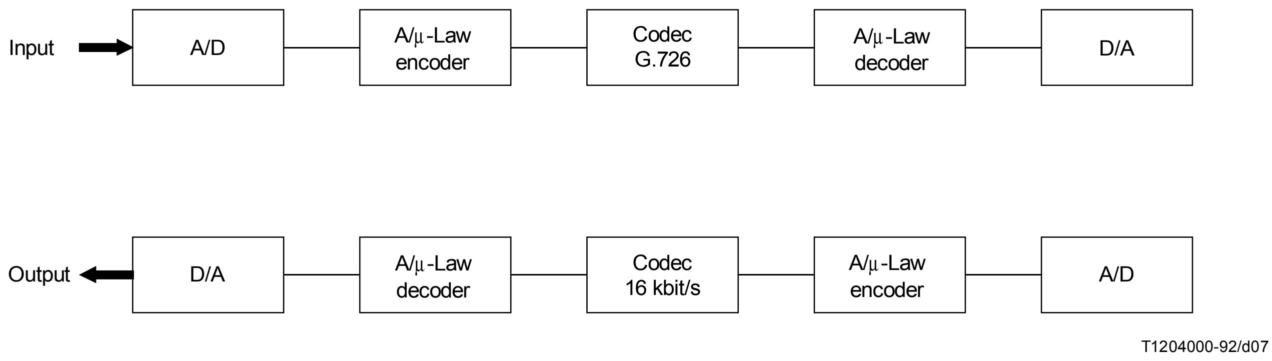

FIGURE 7/P.83

Mixed tandeming

The following example illustrates possible combinations:

8-16 kbit/s ⟶ mobile radio, aeronautical and recorded announcement applications	64 kbit/s ⟶ trunk cct.	32 kbit/s ⟶ variable bit rate DCME	64 kbit/s ⟶ trunk cct.	16-32 kbit/s ⟶ cordless telephone

For wideband systems (e.g. G.722) it is necessary to test for a wideband system in tandem with narrow-band system e.g. G.722 in tandem with G.726 and vice-versa.

3.1.7 Mismatch

It is recommended for wideband systems (G.722) that effects due to a mismatch problem, i.e. different mode of operation between the transmitter and receiver to be investigated for the following conditions:

Transmitter bit rate		*Receiver bit rate*
56 kbit/s	⟶	64 kbit/s
48 kbit/s	⟶	56 kbit/s
48 kbit/s	⟶	64 kbit/s

3.1.8 Environmental noise (sending)

Like the multiple talker considerations in 3.1.3.2, the interaction effect of environmental noise is an important factor for the following reasons:

a) low bit rate codecs can use speech synthesis processes as well as conventional waveform coding; and

b) the application of a codec might be that it is used in a noisy environment, e.g. a moving vehicle if used for mobile radio or a noisy office.

Sufficient testing should be made, with the appropriate noise (see A.1.1.2.2/P.80), to check this effect and the following signal-to-noise ratios (see 2.2.3 for definitions) are recommended:

– 30 dB for room noise;

– 10 dB and 20 dB for vehicular noise.

As the perceived quality is expected to be somewhat lower than normal, the CCITT recommends the use of the ACR method using the Listening Effort scale (see 5.2.2) which is based on the ability to understand the meaning of sentences.

NOTES

1 The noise can be electrically combined with the source tapes so that the level of the noise and speech are accurately known. It is not recommended that source recordings should be made in a noisy environment.

2 It should not be automatically assumed that the mixing of separately recorded speech and high level environmental noise will give the same effects as a subject talking in a noisy environment. A reason for this is that talkers will adapt their vocal characteristics (both level and spectrum) as well as their talking behaviour, in a noisy environment. Since some codecs may process different parts of the audio spectrum in different ways this change in vocal characteristics may affect the performance.

3.1.9 Network information signals

In any national network there are many information signals or tones transmitted for the benefit of the customer, and some for network equipment instruction. These signals may originate from within the PSTN or from private networks attached thereto. It is important that when passing through any speech processing device, that degradation of these signals does not cause them to become unrecognizable to the customer, nor to the equipment designed to take some action upon reception. It is possible that the latter situation, for example DTMF tones, will be less tolerant of degradation than the former, but it has the advantage that simple objective tests should be sufficient to detect an allowable limit to such degradation.

It is recommended that network originated signalling tones, conforming to Recommendation Q.35, are tested and the minimum should be:

- proceed to dial tone;
- called subscriber ringing tone;
- called subscriber engaged tone;
- equipment engaged tone;
- number unobtainable tone.

The Degradation Category Rating (DCR) method using the 5-point scale of 5.2.3 (see also Annex D/P.80) is recommended for use in evaluating the suitability (recognition) of information signals.

3.1.10 Music

Some equipment used in the PSTN, especially PBXs have the facility to play music to the customer when "on hold".

The CCITT recommends that the only simple testing be used in order to ensure that the music is of reasonable quality.

3.2 Reference conditions

3.2.1 Direct

This gives the very best condition that is attainable in the experiment and is the same as Q_N or Q_W of infinity (see 3.2.2).

3.2.2 Modulated Noise Reference Unit (MNRU)

Random noise with amplitude proportional to the instantaneous signal amplitude in terms of Q_N or Q_W ratio, according to the MNRU as specified in Recommendation P.81, should be used as the reference system in terms of which subjective performance of digital processes should be expressed for the following reasons:

a) for waveform codecs and possibly some non-waveform codecs the signal processed through the MNRU is perceptually very similar to the process signal, thus resulting, in principle, in an easier assessment by test subjects;

b) experience has shown that the MNRU is a useful transfer standard and affords sensible comparisons to be made between different laboratories; and

c) considerable experience and information has been accumulated with the MNRU.

This range for narrow-band systems should be from Q_N = 5 to 35 dB (preferably 5 to 7 different values) and for wideband systems should be from Q_W = 10 to 45.

Figure 8 demonstrates the effect of change of Q_N or Q_W.

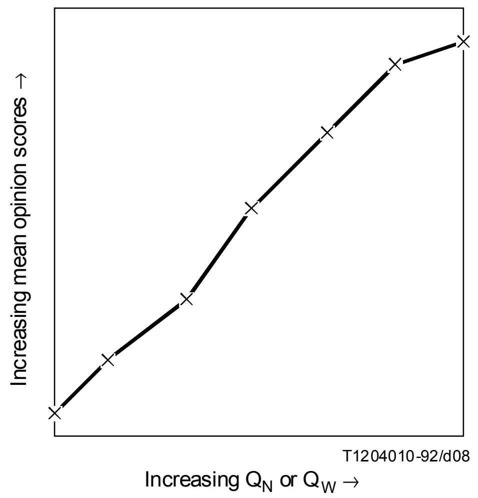

FIGURE 8/P.83

Mean opinion scores for Q_N or Q_W

3.2.3 Signal-to-Noise Ratio (SNR)

Administrations and Operating Companies in the past have found it useful to relate the effects of degradations in terms of SNR.

The following definitions of signal-to-noise ratio, measured on connections, are recommended by the CCITT for use with steady state noise:

– *Telephone (narrow-band) measurements – psophometric* (see Recommendation O.41)

  SNR(p) = Active speech level (see P.56)/psophometric weighted noise measurement

– *Wideband measurements – A-weighted* (see Recommendation P.54)

  SNR(A) = Active speech level (see Recommendation P.56)/A-weighted noise measurement

If definitions, other than those given above, are used then the following system of notations shall be adopted:

– *Narrow-band 300-3400 Hz – unweighted*

  SNR(N) = Active speech level (see Recommendation P.56)/unweighted noise measurement

– *Wideband 100-7000 Hz – unweighted*

  SNR(W) = Active speech level (see Recommendation P.56)/unweighted noise measurement

– *C-message weighting* (see Recommendation O.41)

  SNR(C) = Active speech level (see Recommendation P.56)/C-message weighted noise measurement

– *Bandwidth 100-5000 Hz*

  SNR(0.1-5) = Active speech level (see Recommendation P.56)/unweighted noise measurement

NOTE – If other bandwidths are used then the 0.1-5 should be replaced by the measurement bandwidth.

Comparison of SNR using the definitions for different types of noise spectra and different sending sensitivity/frequency characteristics, can be found in Annex A.

If gaussian noise is used in the test then it is suggested that for both narrow-band and wideband systems the range should be 15 to 45 dB.

Figure 9 demonstrates the effect of change of SNR ratio.

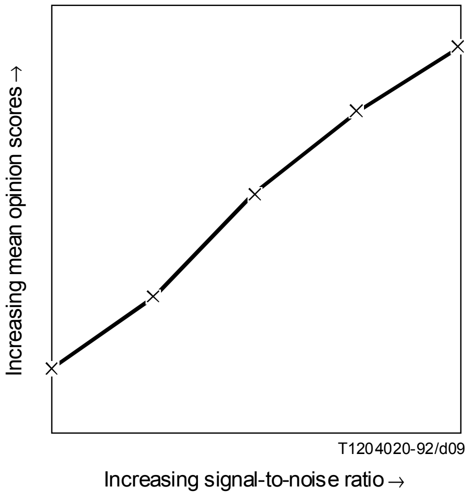

FIGURE 9/P.83

Mean opinion scores for SNR

3.2.4 Reference codecs

Reference codecs can serve two purposes:

1) they can be used to determine planning rules (see Recommendation G.113); and

2) they can be used as the standard to judge the overall performance in terms of parameters (see 3.1).

If used to determine planning rules the CCITT recommends that asynchronous tandeming (see 3.1.6.1) of 1, 2, 4, 8 and 16 A- or µ-law codecs, conforming to Recommendations G.711/G.712, are used.

If used to determine the relative performance e.g. relative to G.726 or G.722, then the considerations of 3.1.6.1 and 3.1.6.2 should be applied.

NOTE – The specification for a reference G.711 codec is under study.

4 Experiment design

The considerations detailed in B.3/P.80 apply.

To use every combination of the parameters described in clause 2 would result in a single experiment that would be logistically too large. It is recommended that a minimum set of experiments be conducted which although would not cover every combination would result in sufficient data to make sensible decisions. Annex B gives sets of experiments found suitable by the CCITT in studies that lead to Recommendations in the G.700-Series of narrow-band and wideband codecs.

Extreme caution should be used when comparing within the same test systems with widely differing degradations, e.g. digital codecs, frequency division multiplex systems, vocoders, etc.

5 Listening test procedure

This clause is the same as B.4/P.80 with the following exceptions.

5.1 Receiving system

For narrow-band systems (300-3400 Hz) it is recommended that the IRS receiving end conforming to Recommendation P.48 and calibrated to Recommendation P.64 be used.

If a wideband system (100-7000 Hz) is to be used for audio-conferencing, then the receiving end should conform to IEC Publication 581.7.

5.2 Opinion scales

The method to be used is the absolute category rating (ACR) type test using a 5-point scale as described in B.4.5/P.80.

The following opinion scales are recommended for assessing digital processes.

5.2.1 Listening Quality scale

Quality of the speech:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

5.2.2 Listening Effort scale

Effort required to understand the meaning of sentences:

5 Complete relaxation possible; no effort required.

4 Attention necessary; no appreciable effort required.

3 Moderate effort required.

2 Considerable effort required.

1 No meaning understood with any feasible effort.

5.2.3 Degradation Category scale

5 Degradation is inaudible.

4 Degradation is audible, but not annoying.

3 Degradation is slightly annoying.

2 Degradation is annoying.

1 Degradation is very annoying.

5.2.4 Detectibility Opinion scale

3 Objectionable

2 Detectable

1 Not detectable

5.3 Electrical noise

Gaussian noise equivalent to –68 dBmp at the input to the receiving system should be added to reduce noise contrasts effects at the onset of speech utterances.

6 Analysis of results

The considerations detailed in B.4.7/P.80 apply.

An objective of the analysis is to determine $Q_2 = F(L)$ where Q_2 is the Q value for the codec and L is the line bit rate. One simple method for determining this function uses the MOS values shown in Figures 3 and 8 and can produce a graph of this function as shown in Figure 10. This method is shown in Figure 11, wherein a value of line bit rate is chosen, say L_2, and its corresponding MOS value is determined. This MOS value is then used to enter the right hand graph to find the value of Q, in this case Q_2, corresponding to this MOS value. Q values for all other L values are obtained in a similar way and the resulting set of (L_i, Q_i) gains are plotted as in Figure 10.

NOTE – Curve fitting applied to the results of MOS versus Q is currently under study.

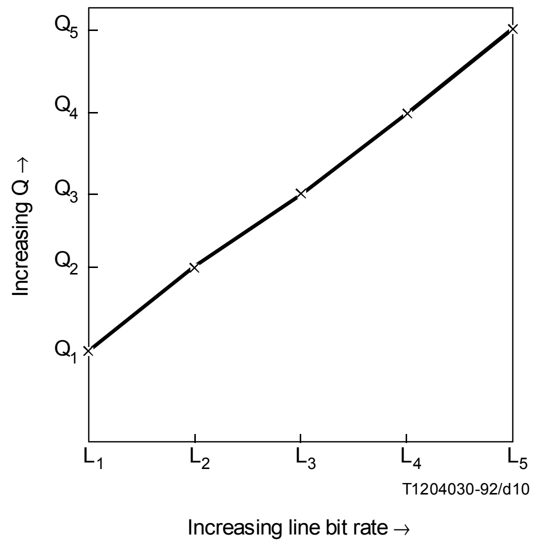

FIGURE 10/P.83

Q as a function of line bit rate

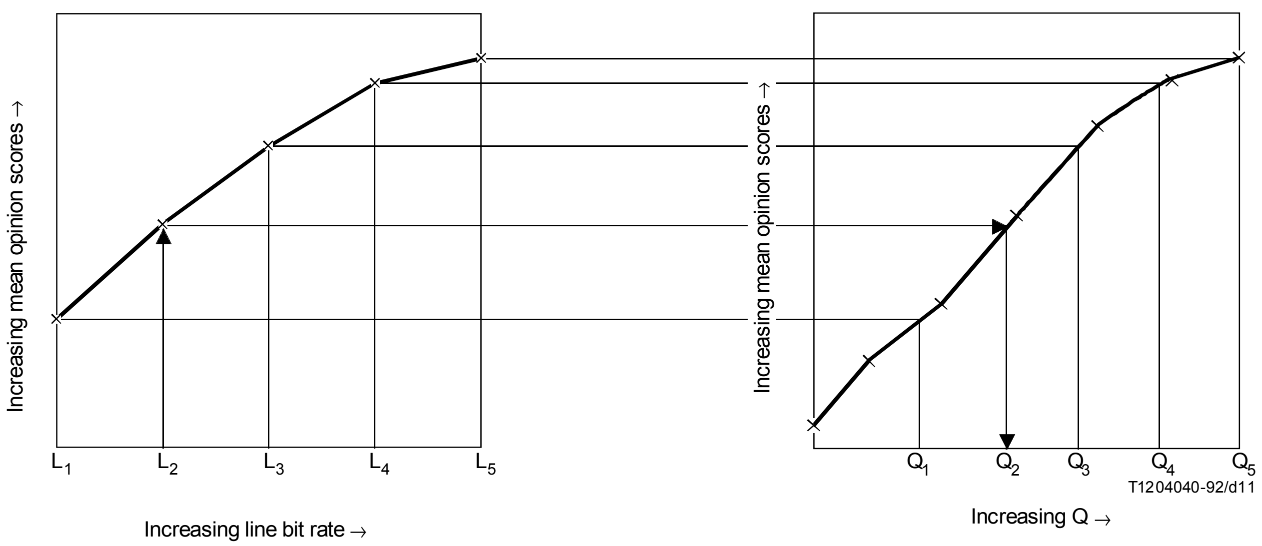

FIGURE 11/P.83

Graphical method of deriving Figure 10 from Figures 3 and 8

Annex A

Comparison of different SNR definitions

(This annex forms an integral part of this Recommendation)

Table A.1 shows some calculations using the weightings described in Recommendation O.41, on known spectra using two different frequency responses. These spectra were chosen to represent those most commonly used in subjective testing.

– White (gaussian) noise consistent with circuit noise-psophometric and C-message designed to measure the effects of this type of noise.

– Hoth (room) noise and vehicle noise are used as environmental noise in subjective tests (see A.1.1.2.2/P.80) and represent the noise(s), picked up by the microphone, measured at the telephone line terminals.

TABLE A.1/P.83

		White noise	Hoth noise	Moving vehicle	Stationary vehicle
Flat response (narrow-band)	Unweighted C-message Psophometric	0 –1.9 –2.6	0 –3.6 –2.6	0 –6.4 –4.1	0 –4.2 –2.8
IRS response (narrow-band)	Unweighted C-message Psophometric	0 –1.9 –3.6	0 –2.0 –2.3	0 –3.7 –2.4	0 –2.3 –2.3
Flat response (wideband)	Unweighted C-message Psophometric	0 –5.2 –5.9	0 –7.3 –6.3	0 –16.5 –14.0	0 –12.7 –11.3
IRS response (wideband)	Unweighted C-message Psophometric	0 –1.9 –3.6	0 –2.1 –2.4	0 –4.4 –3.1	0 –2.5 –2.5

NOTES

– All values are differences in dB with respect to unweighted.

– Negative sign means quieter.

– Calculation of narrow-band unweighted used the 1/3rd octave bands, centred at the preferred frequencies as defined in ISO R.266, from 315 Hz to 3150 Hz inclusive.

– Calculation of wideband unweighted used the 1/3rd octave bands, centred at the preferred frequencies as defined in ISO R.266, from 100 Hz to 6300 Hz inclusive.

– Calculation of C-message and psophometric weightings used 1/3rd octave bands, centred at the preferred frequencies as defined in ISO R.266, from 100 Hz to 5000 Hz inclusive.

– Care must be taken when assuming that C-message weighting is equivalent to psophometric weighting – this is only true for gaussian type noise (see Recommendation O.41).

Annex B

Sets of parameters to determine codec performance

(This annex forms an integral part of this Recommendation)

B.1 Narrow-band systems (300-3400 Hz)

a) *Experiment 1 – Effect of errors, input level and listening level*

Speech input level:	3 (14, 26 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Errors:	3 (0, 1:10000, 1:1000)
Transcoding:	1 transcoding
Environmental noise (sending):	1 (< 30 dBA)

b) *Experiment 2 – Effect of transcodings, input level and listening level*

Speech input level:	3 (14, 26 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Transcoding:	"x"
Error:	1 (1:1000)
Environmental noise (sending):	1 (< 30 dBA)

c) *Experiment 3 – Effect of environment noise, room noise, input level and listening level*

Speech input level:	3 (14, 26 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Transcoding:	1 transcoding
Error:	1 (1:1000)
Environmental noise (sending):	2 (< 30 dBA and "y")
Room noise:	"z"

where "P" is the preferred listening level;

"x" is the number of transcoding combinations to be tested;

"y" is the sending noise level to be tested;

"z" is the number of room noise conditions, typically 2.

All experiments must also include the narrow-band MNRU conditions.

These three experiments are not intended to be exhaustive and should be supplemented with other experiments to better characterize the codec.

B.2 Wideband systems (100-7000 Hz)

a) *Experiment 1 – Effect of bit rate, BER, input level and listening level*

Speech input level:	2 (20 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Transcoding:	1 transcoding
Environmental noise (sending):	1 (< 30 dBA)
Bit rates:	3 (48, 56 and 64 kbit/s)
BER:	3 (0, 1:10000 and 1:1000)
Room noise:	1 (< 30 dBA)

b) *Experiment 2 – Effect of transcodings, input level and listening level*

Speech input level:	2 (20 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Transcoding:	“x” these include synchronous and asynchronous
Environmental noise (sending):	1 (< 30 dBA)
Bit rates:	3 (48, 56 and 64 kbit/s)
BER:	3 (0, 1:10000 and 1:1000)
Room noise:	1 (< 30 dBA)

c) *Experiment 3 – Effect of mismatch, input level and listening level*

Speech input level:	2 (20 and 38 dB below overload)
Listening level:	3 (P + 10, P, P – 10)
Transcoding:	1 transcoding
Environmental noise (sending):	1 (< 30 dBA)
Bit rates:	3 (56 → 64, 48 → 56, 48 → 64 kbit/s)
BER:	2 (0 and 1:1000)
Room noise:	1 (< 30 dBA)

where “P” is the preferred listening level;

“x” is the number of transcoding combinations to be tested.

All three experiments must also include the wideband MNRU conditions.

References

[1] COMBESCURE (P.), *et al*: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.

[2] CCITT Report of the meeting of Working Party XVIII/2 (Speech processing), COM XVIII-R 28, Annex 1, pp. 13-39, December 1983.

[3] GOODMAN (D.J.), NASH (R.D.): Subjective quality of the same speech transmission conditions in seven different countries, Proc. ICASP 82 (International Conference on Acoustics, Speech and Signal Processing), Vol. 2, Paris, May 1982.

[4] MODENA (G.), COLEMAN (A.), USAI (P.) and COVERDALE (P.): Subjective performance evaluation of the 7 kHz audio coder, *IEEE Global Telecommunications Conference 1986 (Globecom ’86)* Houston, Texas, 1-4 December 1986.

[5] EUROPEAN JEG Subjective testing methodology for the evaluation of low-bit rate codecs for mobile radio, CCITT, COM XII-68, May 1986.

[6] SWEDISH TELECOMMUNICATION ADMINISTRATION Report on subjective test on candidate codecs for mobile radio, CCITT, COM XII-147, February 1987.

Bibliography

CCITT Recommendation G.726 *40, 32, 24 and 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*, Vol. III.

DAUMER (W.R.) and CAVANAUGH (J.R.): A subjective comparison of selected digital codecs for speech, *Bell Systems Technical Journal*, Vol. 57, No. 9, November 1978.

RICHARDS (D.L.) and BARNES (G.J.): Pay-off between quantizing distortion and injected circuit noise, *Proc. ICASP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.

BOYD (I.) and SOUTHCOTT (C.B.): A speech codec for the Skyphone service, *British Telecom Technology Journal*, Vol. 6, No. 2, April 1988.

COLEMAN (A.), GLEISS (N.) and USAI (P.): A Subjective Testing Methodology for Evaluating Medium Rate Codecs for Digital Mobile Applications, *Speech Communications*, Vol. 7, pp. 151-166, June 1988.

CROWE (D.P.): Selection of Voice Codec for the Aeronautical Satellite Service, *European Conference on Speech Communication and Technology*, Vol. 2, S37, pp. 320-323, September 1989.

COLEMAN (A.), GLEISS (N.), SOTSCHECK (J.), USAI (P.) and SCHEUERMANN (H.): Subjective performance evaluation of the RPE-LTP codec for the Pan-European cellular digital mobile radio system, *IEEE Globecom '89*, Dallas, Texas, 27-30 November 1989.

WHEDDON (C.) and LINGGARD (R.): Speech and Language Processing, *Chapman and Hall*, 1990.

Printed in Switzerland
Geneva, 1994