



INTERNATIONAL TELECOMMUNICATION UNION

**ITU-T**

**P.830**

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(02/96)

**TELEPHONE TRANSMISSION QUALITY  
METHODS FOR OBJECTIVE AND SUBJECTIVE  
ASSESSMENT OF QUALITY**

---

**SUBJECTIVE PERFORMANCE ASSESSMENT  
OF TELEPHONE-BAND AND WIDEBAND  
DIGITAL CODECS**

**ITU-T Recommendation P.830**

(Previously "CCITT Recommendation")

---

## FOREWORD

The ITU-T (Telecommunication Standardization Sector) is a permanent organ of the International Telecommunication Union (ITU). 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, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

ITU-T Recommendation P.830 was revised by ITU-T Study Group 12 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 6th of February 1996.

---

## NOTE

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

© ITU 1996

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

## CONTENTS

	<i>Page</i>
1 Scope .....	1
2 References .....	1
3 Definitions.....	2
4 Abbreviations .....	2
5 Conventions.....	3
6 Summary of subjective testing procedures.....	3
7 Source recordings .....	3
7.1 Recording system.....	3
7.2 Speech samples.....	4
8 Selection of experiment parameters .....	5
8.1 Codec conditions .....	5
8.2 Reference conditions .....	11
9 Experiment design.....	14
10 Listening test procedure .....	14
10.1 Receiving system.....	14
10.2 Opinion scales.....	14
10.3 Electrical noise.....	15
11 Analysis of results .....	15
Annex A – Comparison of different SNR definitions .....	17
Annex B – Sets of parameters to determine codec performance.....	17
B.1 Narrow-band systems (300-3400 Hz).....	18
B.2 Wideband systems (100-7000 Hz).....	18
Annex C – PCM codec corresponding to one qdu .....	19
C.1 Common analogue interface .....	19
C.2 Digital interface .....	19
C.3 Analogue interface.....	19
Annex D – Modified IRS send and receive characteristics .....	20
Bibliography .....	21

## SUMMARY

This revised Recommendation describes methods and procedures for conducting subjective performance evaluations of digital speech codecs.

Revisions encompassed by this version of the Recommendation are to include new information that reflects current practices in subjective evaluation of digital codecs, including an expanded section on creating source recordings and addition of two annexes. One annex describes an implementation of a PCM codec (A-/ $\mu$ -law) that generates one quantization distortion unit (qdu) of distortion to input signals. The other new annex describes the Modified IRS transmit and receive characteristics. These characteristics are recommended as the transmit and receive responses to be used in situations where the codec being tested is intended for use in fully digital circuits.

## INTRODUCTION

Digital telephony in the Public Switched Telephone Network (PSTN) is now commonplace. Since the 1960s, digital transmission in the PSTN has been synonymous with 64 kbit/s A-law or  $\mu$ -law PCM (see Recommendation G.711). The combination of better knowledge in signal processing techniques and advances in technology, most notably LSI/VLSI techniques, has led to an increased interest in more efficient coding methods than PCM. Examples of this development include the introduction of 32 kbit/s ADPCM (Recommendation G.726), 16 kbit/s LDCELP (Recommendation G.728), and a 64 kbit/s, 7 kHz wideband codec (Recommendation G.722). Considerable interest is now focused on coding of speech at bit rates below 16 kbit/s.

Subjective testing is the most widely used method of assessing the performance of digital codecs. When the transmission path is digital and/or non-linear, simple objective measurements, such as those specified in 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, and hence to ensure that the performance of the complete system is satisfactory. To be suitable for this purpose, the measurements must 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 assign a unique numerical contribution to each digital process and the ability to use this assigned contribution in conjunction with other impairments to estimate telephone connection performance.

## KEYWORDS

Subjective testing, subjective evaluation.

## SUBJECTIVE PERFORMANCE ASSESSMENT OF TELEPHONE-BAND AND WIDEBAND DIGITAL CODECS

(Helsinki, 1993; revised in 1996)

### 1 Scope

This Recommendation is based on Annex B/P.80 and on experience gained from several international experiments [5], [6], [9], [10], [12], [13], [15] and [17]. Further, this Recommendation defines a specific testing method for evaluating digital processes. This method allows the subjective effects of distortion introduced by the speech coding process to be taken into account when assessing the performance of the evolving international network. This Recommendation contains advice on how to assess the performance of digital codecs. This Recommendation must be read in conjunction with Recommendation P.80. Additional information is available in sources listed in the Bibliography.

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 ITU-T 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 (e.g. the Listening Effort scale has been found useful when interest centers on good correlation with measures of conversational performance), 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 [8] is described in detail in Annex D/P.80. This method is purported to be suited when the absolute quality of the source material is low (e.g. speech corrupted by background noise) or 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.

Furthermore, there may be a need for supplementing listening-only tests with conversation tests planned and conducted in accordance with Recommendation P.80. However, until there has been further consideration of the circumstances in which conversational tests are necessary, and further investigation is made regarding the choice of conditions that should be covered in such experiments, no detailed plans can be included in this Recommendation.

### 2 References

The following Recommendations contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated are valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations listed below. A list of the currently valid ITU-T Recommendations is regularly published.

- IEC Publication 581.7 (1986), *High fidelity audio equipment and systems; Minimum performance requirements – Part 7: Loudspeakers.*
- ITU-T Recommendation G.113 (1996), *Transmission impairments.*
- ITU-T Recommendation G.191 (1993), *Software tools for speech and audio coding standardization.*

---

<sup>1)</sup> Previously P.83.

- ITU-T Recommendation G.192 (1996), *A common digital parallel interface for speech standardization activities.*
- CCITT Recommendation G.711 (1988), *Pulse Code Modulation (PCM) of voice frequencies.*
- CCITT Recommendation G.712 (1992), *Transmission performance characteristics of pulse code modulation.*
- CCITT Recommendation G.722 (1988), *7 kHz audio-coding within 64 kbit/s.*
- CCITT Recommendation G.726 (1990), *40, 32, 24 and 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).*
- CCITT Recommendation G.728 (1992), *Coding of speech at 16 kbit/s using low-delay code excited linear prediction.*
- ITU-T Recommendation O.41 (1994), *Psophometer for use on telephone-type circuits.*
- CCITT Recommendation O.131 (1988), *Quantizing distortion measuring equipment using a pseudo-random noise test signal.*
- CCITT Recommendation O.132 (1988), *Quantizing distortion measuring equipment using a sinusoidal test signal.*
- ITU-T Recommendation O.133 (1993), *Equipment for measuring the performance of PCM encoders and decoders.*
- ITU-T Recommendation P.10 (1993), *Vocabulary of terms on telephone transmission quality and telephone sets.*
- CCITT Recommendation P.48 (1988), *Specification for an intermediate reference system.*
- ITU-T Recommendation P.50 (1993), *Artificial voices.*
- ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level.*
- ITU-T Recommendation P.80 (1993), *Methods for subjective determination of transmission quality.*
- ITU-T Recommendation P.810 (1996), *Modulated Noise Reference Unit (MNRU).*
- ITU-T Recommendation P.84 (1993), *Subjective listening test method for evaluating digital circuit multiplication and packetized voice systems.*
- CCITT Recommendation Q.35 (1988), *Technical characteristics of tones for the telephone service.*

### 3 Definitions

For the purposes of this Recommendation, the following definitions apply:

- 3.1 **dBov**: dB relative to the overload point of a digital system.
- 3.2 **Q**: The ratio, in dB, of speech power to modulated noise power in the Modulated Noise Reference Unit, as described in Recommendation P.810.
- 3.3 **Q<sub>N</sub>**: Q for a narrow-band Modulated Noise Reference Unit.
- 3.4 **Q<sub>W</sub>**: Q for a wideband Modulated Noise Reference Unit.

### 4 Abbreviations

For the purposes of this Recommendation, the following abbreviations are used:

ACR	Absolute Category Rating
ADPCM	Adaptive Differential Pulse Code Modulation
BER	Bit Error Rate
DCR	Degradation Category Rating
eif	equipment impairment factor

IRS	Intermediate Reference System (Recommendation P.48)
ITU-T STL	ITU-T Software Tools Library
LDCELP	Low-Delay Code Excited Linear Prediction
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
PSTN	Public Switched Telephone Network
qdu	quantization distortion unit
RMS	Root Mean Square
SNR	Signal-to-Noise Ratio

## 5 Conventions

Subjective evaluation of digital speech codecs may, in principle, be conducted using listening-only or conversational methods of subjective testing. As a practical matter, listening-only tests are the only feasible method of subjective testing during the development of speech codecs when a real-time implementation of the codec is not available. This Recommendation discusses listening-only methods of subjective evaluation of speech codecs.

## 6 Summary of subjective testing procedures

When performing a subjective evaluation of a speech codec, consideration must be given to the intended application of the codec under test. The system used for recording speech materials, the selection of experimental test conditions, and the system used for delivery of processed speech materials to the listeners must be chosen so that they capture the important end-to-end characteristics of the telephone connection. In particular, the characteristics of the recording and receiving systems should reflect the characteristics of the transmitter and receiver that will be used in the intended application of the codec.

Performing subjective evaluations of digital codecs proceeds via a number of steps:

- 1) Preparation of source speech materials, including recording of talkers;
- 2) Selection of experimental parameters to exercise the features of the codec that are of interest;
- 3) Design of the experiment;
- 4) Selection of a test procedure and conduct of the experiment;
- 5) Analysis of results.

Each of these steps is described below.

## 7 Source recordings

### 7.1 Recording system

As noted above, it is important to use a representative system for the transmitting end of the system under test. When testing narrow-band (300-3400 Hz) digital codecs for network applications, the ITU-T recommends the use of a modified Intermediate Reference System (IRS) transmit characteristic. Such a device is constructed by removing the SRAEN filter from the IRS transmitter. The transmit characteristic of the modified IRS is found in Annex D. A software implementation of the modified IRS transmit characteristic is available in the ITU-T Software Tools Library (ITU-T STL).

NOTE – The ITU-T Software Tools Library contains a number of tools that are useful for the development of speech and audio coding standards. These tools are also useful when preparing materials for use in subjective evaluations of speech codecs. The tools that are available in the ITU-T STL are described in Recommendation G.191.

In some circumstances (e.g. when it is expected that low frequencies will be sharply limited), it will be necessary or desirable to assess the performance of the codec using the sending end of the Intermediate Reference System (IRS) conforming to Recommendation P.48 and calibrated to Recommendation P.64. 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. In particular, the IRS is representative of analogue telephone connections (1970s). Hence, the modified IRS transmit characteristic should be used under most circumstances.

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.

## 7.2 Speech samples

Speech material should consist of simple, short, meaningful sentences. These sentences should be chosen so as to be easy to understand (from the current non-technical literature or newspapers, for example). Further, the sentences should be made into sets of two or three in such a way that there is no obvious connection of meaning between the sentences in a set. Very short and very long sentences should be avoided, the aim being that each sentence when spoken should have a duration of 2-3 seconds.

NOTE – When considering bit errors, this allows, for example, for Bit Error Rates (BERs) of 0, 1:10000, 1:1000 and 1:100 to be assessed (see 8.1.4). BERs below 1:10000 will require more sentences/sample.

### 7.2.1 Production of source material

The talkers should be seated (one at a time) in a room with reverberation time less than 500 milliseconds, and room noise level below 30 dBA with no dominant peaks in the spectrum. The recordings should be made on a high quality recording system.

Talkers should pronounce the sentences fluently, but not dramatically, and maintain a constant level that they find comfortable, avoiding noise from sources such as the rustling of paper. The sentences may be presented to the talker one by one on a visual display controlled by a computer, or on a list at a distance suitable for easy reading.

The active speech level should be observed during recording. Care should be taken during the recording process that the active speech level for the recording is between 20 and 30 dB below the peak overload point of the recording system for each sentence measured separately. Any sentence for which this does not hold should be re-recorded.

NOTE – Care must be taken while recording, that no spurious noises (e.g. mains hum or switching spikes) are added.

### 7.2.2 Equalization of speech level

The measurement of signal levels in the digital part of the international network is normally expressed by telecommunications engineers as  $y$  dBm0, i.e. the level relative to 1 mW in 600  $\Omega$ . However, speech materials recorded digitally by computer are typically stored in integer format. It is more convenient to specify the levels of such material relative to the maximum signal power that can be stored on a computer, e.g.  $z$  dBov (where the characters 'ov' arbitrarily mean digital overload signal level). It is necessary to understand the relationship of these terms so that the levels in dBm0 can be expressed in dBov. Such a relationship has the form:

$$z \text{ dBov} = y \text{ dBm0} + C \quad (7-1)$$

For an A-law encoded system, a sine wave that excites the maximum level has a power given by  $T_{\text{max}} = 3.14$  dBm0 (Recommendation G.711). Similarly, for a  $\mu$ -law encoded system,  $T_{\text{max}} = 3.17$  dBm0.

The reference power for dBov is the maximum integer value for the system in use. A sine wave of maximum amplitude (e.g. 32 767 for a 16-bit system) will have power  $-3.01$  dBov. Thus,  $C = -6.15$  dB for A-law systems and  $C = -6.18$  dB for  $\mu$ -law systems.



A typical nominal value for mean active speech level (measured according to Recommendation P.56) is  $-20$  dBm0, corresponding to approximately  $-26$  dBov. Thus, the active speech level of each sentence should be measured, and the recorded level adjusted to be  $-27 \pm 1$  dBov for the digital system, i.e. in digital format. For languages or voices where the peak-to-average level exceeds 23 dB, the average level of the speech material should be reduced appropriately. The software implementation of Recommendation P.56, as found in the ITU-T STL, should be used to determine the average speech level while active.

Finally, groups of sentences are combined (with a suitable silence period between sentences) to produce the final source material for use in the subjective test.

## 8 Selection of experiment parameters

Experimental conditions are conveniently divided into those that involve the codec under test and those that serve as references. Codec conditions provide essential information about the performance of the codec, while reference conditions allow sensible comparisons to be made when the same codec is tested in different laboratories or in the same laboratory at different times. Figures 1 to 3 and 8 to 9 are examples to demonstrate typical effects on MOS of experiment parameters.

### 8.1 Codec conditions

Test conditions described in this subclause will serve to indicate how the performance of the codec is affected by changes in input and listening levels, different talkers, presence of background noise, etc.

#### 8.1.1 Speech input levels

It is recognized that digital speech codecs are sensitive to the level of the input speech and that the performance of the codec may depend on this level. When assessing the performance of 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 (approximately equivalent to  $-8$ ,  $-20$ ,  $-32$  dBm0) should be used for assessing the impact of input level variations on the performance of the codec. This is approximately equal to the mean value measured at the International Switching Point  $\pm$  two standard deviations. Other applications may require different input levels. Figure 1 illustrates the effect on Mean Opinion Score (MOS) of change in speech input level.

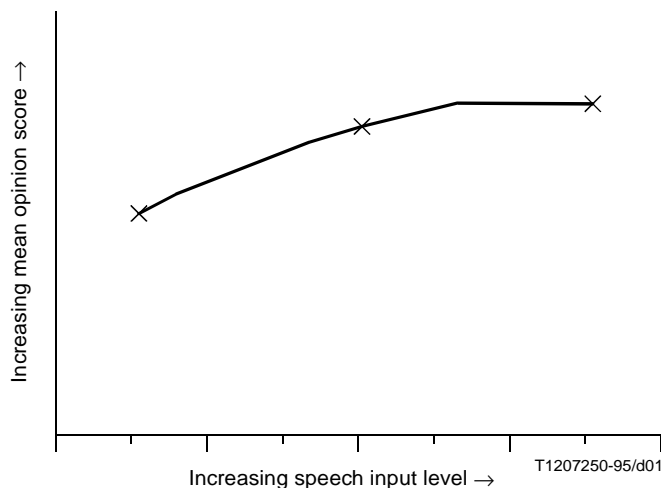


FIGURE 1/P.830

Mean opinion scores for speech input levels

### 8.1.2 Listening levels

It is recommended that at least three listening levels should be used when the effect of receive level is to be assessed. These levels should lie 10 dB to either side of the preferred listening level (for this purpose taken as 79 dB SPL at the ear reference plane). It should be noted that the preferred listening level is generally below the “optimum listening level”, i.e. the listening level corresponding to the maximum MOS value, as defined in Recommendation P.10.

Figure 2 illustrates the effect on MOS of changes in listening level (as well as the interaction between listening level with male and female talkers).

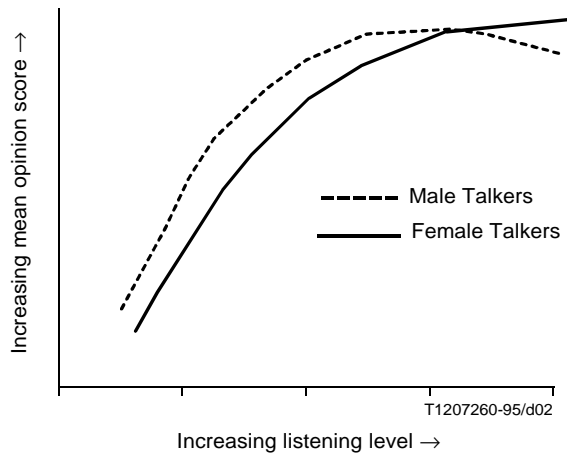


FIGURE 2/P.830  
Mean opinion scores for listening levels

### 8.1.3 Talkers

#### 8.1.3.1 Different talkers

It is recommended that a minimum of two male and two female talkers should be used. Figure 2 illustrates the effect on MOS of different talkers. However, if talker dependency is to be tested as a factor in its own right, it is recommended to use more talkers, as follows:

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

#### 8.1.3.2 Multiple talkers

It is important to know how robust the codec is to the presence of multiple simultaneous voice input signals and to understand 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 the codec 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 presence of multiple talkers, where the difference in speech levels could be zero, does not adversely affect the performance of the codec.

Either the Degradation Category Rating (DCR) method using the 5-point scale of 10.2.3 (see also Annex D/P.80) or the Quantal-Response detectability method using the 3-point scale of 10.2.4 (see also Annex C/P.80) is recommended for assessing the effects of multiple talkers.

#### 8.1.4 Errors

The selection of error conditions must be done in the context of the intended applications of the codec. If the codec is to be used in the PSTN (i.e. line systems), then errors should be randomly distributed bit errors. If, on the other hand, the intended application is a radio environment, such as mobile radio, then errors will occur either in isolation or in bursts, and the effect of these errors will be to corrupt entire frames of the coded speech, rather than single bits.

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 7.2). In certain circumstances it may be necessary to test at BERs up to 1:100. The Error Insertion Device (EID) in the ITU-T STL may be used for this purpose.

For other applications, such as mobile radio, the errors may be of the burst-error type and therefore it would be appropriate to use errors of this type. Again, the EID in the ITU-T STL may be used for this purpose.

It is usual to use the ACR method using the Listening Quality scale of 10.2.1. If, however, the condition is expected to be of poor quality then the ACR method using the Listening Effort scale of 10.2.2 may be more appropriate.

#### 8.1.5 Bit rates

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

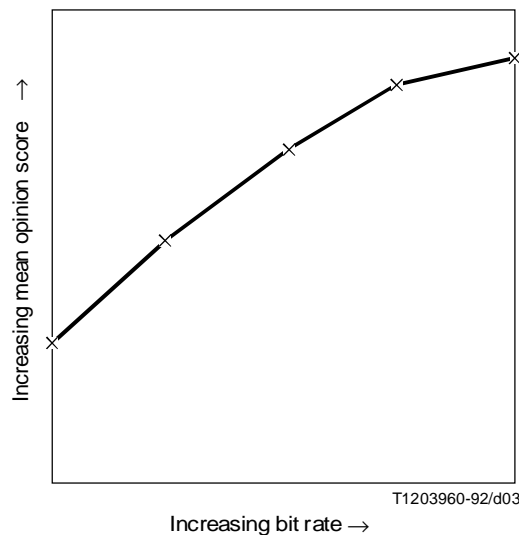


FIGURE 3/P.830

Mean opinion scores at different bit rates

### 8.1.6 Transcodings

Since codecs standardized by the ITU-T are likely to be used in the PSTN (and the associated 64 kbit/s architecture), speech codecs should be assessed in combination with an A- or  $\mu$ -law companding process. Hence, a single codec in a connection going from an analogue transmitter to an analogue receiver will have the configuration shown in Figure 4.

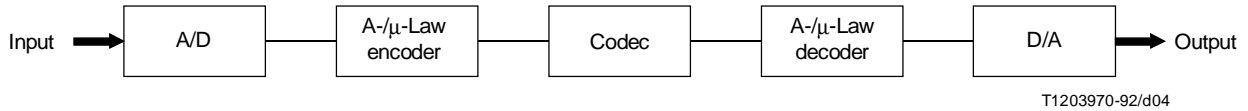


FIGURE 4/P.830  
Transcoding scheme

#### 8.1.6.1 Asynchronous tandeming

When the codec under test is to be used in the mixed analogue and digital PSTN, situations may occur where speech signals are encoded and decoded by the codec two or more times, i.e. various digital segments are interspersed with analogue segments. In such situations, the distortion introduced by the codec accumulates with each successive conversion to analogue, then back to digital with re-encoding by the codec. Further, when the codec is deployed in the network each transcoding includes additional quantization distortion produced by the linear to A- or  $\mu$ -law companding process and additional attenuation distortion from tolerances in the anti-aliasing and reconstruction filters, plus accumulated idle-channel noise. This configuration is demonstrated in Figure 5.

The ITU-T recommends that at least 1, 2 and 3 codecs should be tested in tandem. The recommended method to achieve asynchronous tandeming (to simulate the D/A and A/D process) is to use the up-and-down-sampling filter in the ITU-T STL.

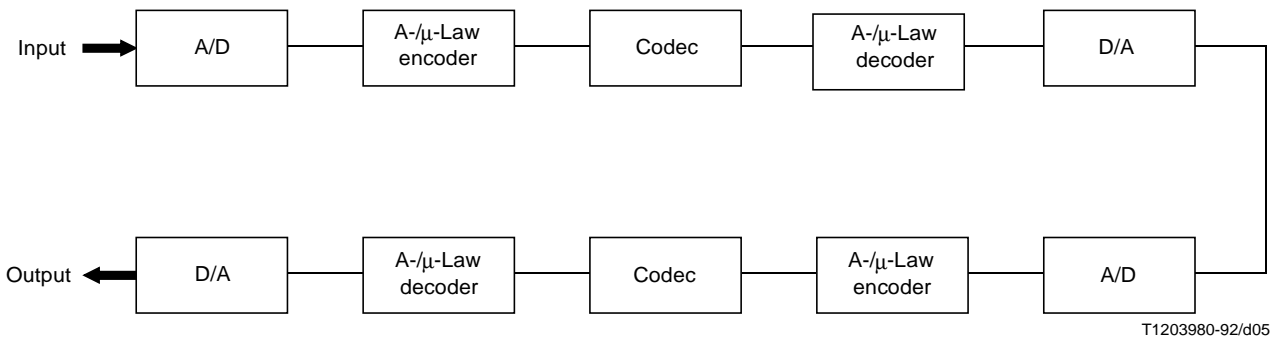


FIGURE 5/P.830  
Asynchronous tandeming

### 8.1.6.2 Synchronous tandeming

The circuit configuration shown in Figure 6 demonstrates situations similar to those of 8.1.6.1. In these situations, speech signals are encoded and decoded multiple times but without D/A and A- or  $\mu$ -law conversions (i.e. in an all-digital network). Thus, this situation has the advantage that there is only one A- or  $\mu$ -law companding process and therefore has reduced quantization distortion when compared to 8.1.6.1. However, the accumulated distortion produced by the test codec continues to occur and must be assessed.

The ITU-T recommends that at least 1, 2 and 3 successive encodings and decodings by the codec should be tested in tandem.

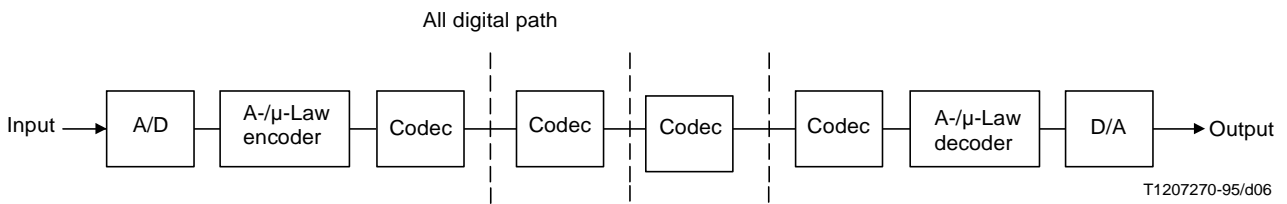


FIGURE 6/P.830  
Example of synchronous tandeming

### 8.1.6.3 Interoperability with other speech coding standards

It is important to establish the effects of tandeming systems that use encoding at different bit rates or different methods of coding. An example configuration is shown in Figure 7. Further examples of interoperability configurations can be found in [14]. It is essential that the most probable combinations of codecs are tested.

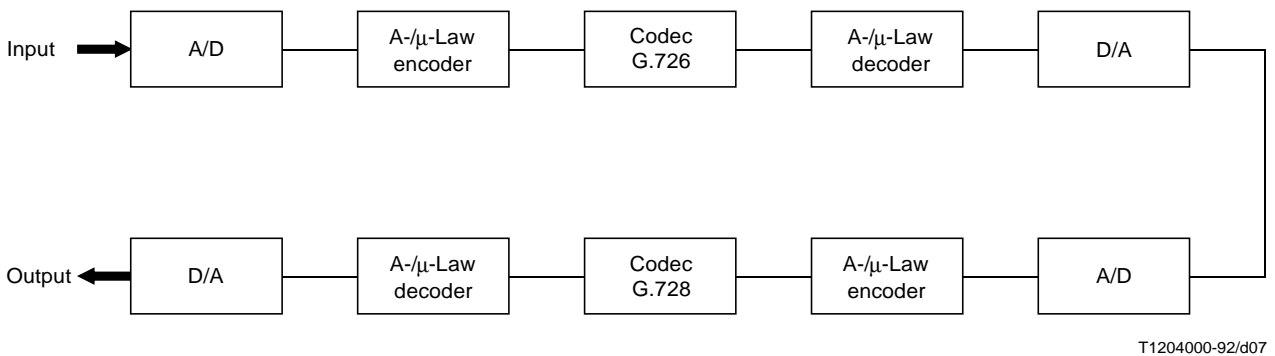


FIGURE 7/P.830  
Interoperability of speech coding standards

The following example illustrates possible combinations:

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

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

### 8.1.7 Mismatch

Some speech codecs (e.g. Recommendation G.722) are capable of operating at a variety of bit rates. In such systems, it is possible that the transmitter and receiver are working in different modes (i.e. different bit rates). When this situation is possible, it is necessary to investigate the performance of appropriate combinations of bit rate at the two ends of the connection. For example, a codec meeting the specifications in Recommendation G.722 should be tested for the following conditions where the encoder and decoder are operating at different bit rates:

<i>Transmitter bit rate</i>		<i>Receiver bit rate</i>
56 kbit/s	————→	64 kbit/s
48 kbit/s	————→	56 kbit/s
48 kbit/s	————→	64 kbit/s

### 8.1.8 Environmental noise (sending)

As with the multiple talker considerations in 8.1.3.2, the interaction of environmental noise and the voice of the talker is an important factor for the following reasons:

- a) low-bit-rate codecs may use coding techniques other than 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 carried out, with the appropriate noise (see A.1.1.2.2/P.80), to examine this effect. The following signal-to-noise ratios (see 8.2.3 for definitions) are recommended:

- 30 dB for room noise;
- 10 dB and 20 dB for vehicular noise.

As the perceived quality of speech corrupted by environmental noise is expected to be somewhat lower than that of speech in a quiet background, some laboratories have found the ACR method using the Listening Effort scale (see 10.2.2) to be especially useful since it is based on the ability to understand the meaning of sentences. These laboratories have found that the Listening Effort scale yields results that are better correlated with conversation opinion scores than are Listening Quality scores.

Modern speech coding techniques may attempt to suppress or eliminate background noise. Thus, it is possible for speech material that has been corrupted by background noise, and subsequently processed by a codec, to be of better quality than the original input (i.e. speech corrupted by noise). In these situations, the Comparison scale (see 10.2.5) has been found to be quite useful.

Environmental noises recorded with studio microphones must be filtered to compensate for the difference between diffuse (far field) sources and point (near field) sources. A software implementation of such a filter is available in the ITU-T STL. SNRs defined above should be measured after this filtering.

## NOTES

1 The noise can be electrically combined with the source recordings so that the level of the noise and speech are known accurately. It is not recommended that source recordings 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 Some laboratories have used a technique for recording speech samples that exploit the fact that talkers will adjust their speaking levels in noisy environments, but avoids the problems associated with recording in noisy environments. Talkers wear headphones over which background noise is played at a known level. The talker's speech is then recorded in the normal way, with the talker speaking as if he or she were in a noisy environment. Noise at a desired level relative to the speech is then added electronically to the recorded speech.

### 8.1.9 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 degradation of these signals, when passing through any speech-processing device, should 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 will be less tolerant of degradation than the former when certain signals such as DTMF tones are in question, 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, should be tested subjectively, 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 10.2.3 (see also Annex D/P.80) is recommended for use in evaluating the suitability (recognition) of information signals.

### 8.1.10 Music

Some equipment used in the PSTN, especially PBXs, has the facility to play music to the customer when “on hold”. The ITU-T recommends that only simple testing be used in order to ensure that the music is of reasonable quality.

## 8.2 Reference conditions

Reference conditions provide a convenient means for making meaningful comparisons of subjective test results from different laboratories or from the same laboratory at different times. These conditions include a *best possible* condition, as well as conditions where known, controlled degradations have been added to the speech materials.

### 8.2.1 Direct

The “direct” condition gives the very best condition that is attainable in the experiment and is theoretically the same as  $Q_N$  or  $Q_W$  of infinity (see 8.2.2). In particular, this condition has no coding of the input speech, having only the same input filtering, input level, and output filtering as speech that has been processed by the codec under test.

### 8.2.2 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.810, 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 processed 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 between experiments at the same laboratory but conducted at different times; and
- c) considerable experience and information has been accumulated with the MNRU.

A range of  $Q_N$  or  $Q_W$  should be included in the subjective test. 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 dB (also 5 to 7 different values).

NOTE – The distortion introduced by the MNRU is subjectively similar to quantization distortion. Such distortions are not necessarily similar, subjectively, to the distortion produced by modern speech coding techniques. Reference degradations that produce distortions more similar to modern coding impairments are currently under study.

Figure 8 illustrates the effect on MOS of change of  $Q_N$  or  $Q_W$ .

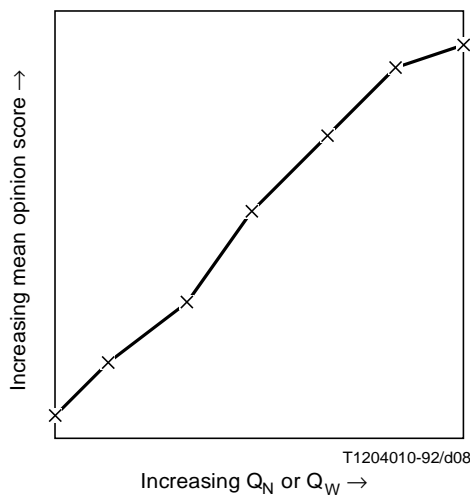


FIGURE 8/P.830  
Mean opinion scores for  $Q_N$  or  $Q_W$

### 8.2.3 Signal-to-Noise Ratio (SNR)

Administrations and Operating Companies 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 ITU-T for use with steady state noise:

- *Telephone (narrow-band) measurements – psophometric* (see Recommendation O.41)  
SNR(p) = Active speech level (see Recommendation P.56)/psophometrically 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 kHz) = Active speech level (see Recommendation P.56)/unweighted noise measurement.

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

Comparisons of SNRs, using the definitions for different types of noise spectra and different sending sensitivity/frequency characteristics, are 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 of SNRs should be 15 to 45 dB.

Figure 9 illustrates the typical effect on MOS of changes in SNR.

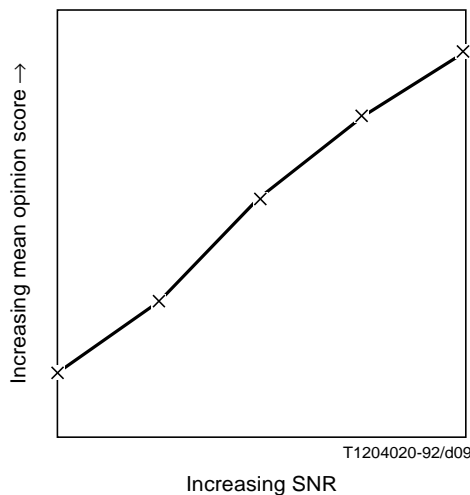


FIGURE 9/P.830  
Mean opinion scores for SNR

## 8.2.4 Reference codecs

Reference codecs serve two purposes:

- 1) they can be used to determine planning rules, either in the form of quantization distortion units (q dus) or equipment impairment factors (eifs) (see Recommendation G.113 for discussion of the q du and eif); and
- 2) they can be used as the standard to judge the overall performance in terms of parameters (see 8.1).

If reference codecs are used to determine planning rules, the ITU-T recommends that asynchronous tandeming (see 8.1.6.1) of 1, 2, 4, 8 and 16 A- or  $\mu$ -law codecs, conforming to Recommendations G.711 and G.712, should be used. The specification for a reference G.711 codec is found in Annex C.

If used to determine the relative performance, e.g. relative to Recommendation G.726 or G.722, then the considerations of 8.1.6.1 and 8.1.6.2 should be applied.

## 9 Experiment design

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

To use every combination of the parameters described in clause 7 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 they would not cover every combination, would result in sufficient data to make sensible decisions. Annex B gives sets of experiments found suitable by the ITU-T in studies that led to Recommendations in the G.700-Series of narrow-band and wideband codecs.

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

## 10 Listening test procedure

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

### 10.1 Receiving system

When testing narrow-band (300-3400 Hz) digital codecs for network applications, the ITU-T recommends the use of a modified IRS receive characteristic. As with the modified IRS transmit characteristic, this device is constructed by removing the SRAEN filter from the IRS receiver. The receive characteristic of the modified IRS is found in Annex D. A software implementation of the modified IRS receive characteristic is available in the ITU-T STL. In some circumstances (e.g. when it is expected that low frequencies will be sharply limited), it will be necessary or desirable to assess the performance of the codec using the receiving end of the IRS, conforming to Recommendation P.48 and calibrated according to Recommendation P.64.

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.

NOTE – Some laboratories use headphones for presentation of speech materials to listeners during subjective evaluations. When using headphones for presentation of speech materials, care must be taken to ensure that presentation provides proper filtering to simulate the effects of real receivers, including the effects of acoustic leakage when necessary. See [11] for more information on the use of headphones as opposed to telephone handsets.

### 10.2 Opinion scales

The following opinion scales are recommended for assessing digital processes.

#### 10.2.1 Listening Quality scale

Quality of the speech:

- 5 Excellent
- 4 Good

- 3 Fair
- 2 Poor
- 1 Bad

### 10.2.2 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.

### 10.2.3 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.

### 10.2.4 Detectability Opinion scale

- 3 Objectionable
- 2 Detectable
- 1 Not detectable

### 10.2.5 Comparison scale

The second compared to the first is:

- 3 Much better
- 2 Better
- 1 Slightly better
- 0 About the same
- 1 Slightly worse
- 2 Worse
- 3 Much worse

## 10.3 Electrical noise

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

## 11 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 equivalent Q value for the codec and L is the line bit rate. One simple method for determining this function uses the MOSs 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 is determined. This MOS is then used to enter the right-hand graph to find the value of Q, in this case  $Q_2$ , corresponding to this MOS. 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 – The above operations are more reliable when carried out on fitted curves than on straight-line segments joining observed points. Methods for calculating appropriate fits for MOS as a function of Q are under study. As a provisional procedure, while the results of this study are awaited, attention is drawn to a sigmoid-fitting method that takes account of the “saturation effect” at high values of Q, and has been used with some success in ETSI work on assessing candidates for the Half-Rate GSM system [2].

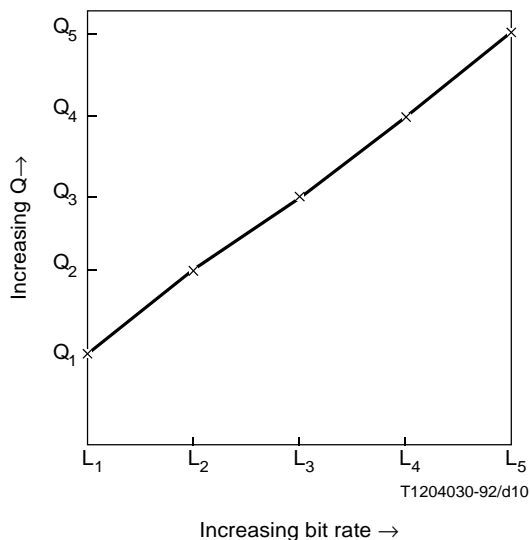


FIGURE 10/P.830  
Q as a function of bit rate

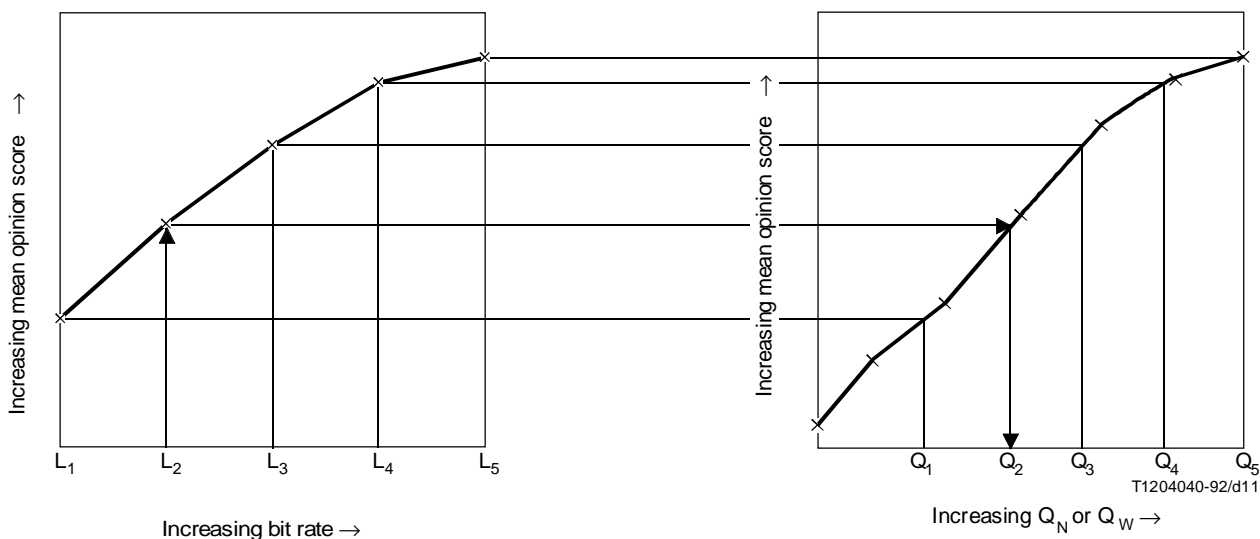


FIGURE 11/P.830  
Graphical method of deriving Figure 10 from Figures 3 and 8

## 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.830

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

#### NOTES

- 1 All values are differences in dB with respect to unweighted.
- 2 Negative sign means quieter.
- 3 The IRS response used here refers to Recommendation P.48.
- 4 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.
- 5 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.
- 6 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.
- 7 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)

This annex gives examples of experiments that should be conducted to evaluate the subjective performance of digital codecs. As noted in clause 6, it is important to augment these experiments with conditions and/or experiments that test the codec in special conditions of its intended application.

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

- a) *Experiment 1 – Effect of errors, input level and listening level*
- |                                       |                                     |
|---------------------------------------|-------------------------------------|
| Speech input levels:                  | 3 (14, 26 and 38 dB below overload) |
| Listening levels:                     | 3 (P + 10, P, P – 10)               |
| Error rates:                          | 3 (0, 1:10000, 1:1000)              |
| Transcoding:                          | 1 transcoding                       |
| Environmental noise levels (sending): | 1 (< 30 dBA)                        |
- b) *Experiment 2 – Effect of transcodings, input level and listening level*
- |                                       |                                     |
|---------------------------------------|-------------------------------------|
| Speech input levels:                  | 3 (14, 26 and 38 dB below overload) |
| Listening levels:                     | 3 (P + 10, P, P – 10)               |
| Transcodings:                         | “x”                                 |
| Error rates:                          | 1 (1:1000)                          |
| Environmental noise levels (sending): | 1 (< 30 dBA)                        |
- c) *Experiment 3 – Effect of environment noise, room noise, input level and listening level*
- |                                       |                                     |
|---------------------------------------|-------------------------------------|
| Speech input levels:                  | 3 (14, 26 and 38 dB below overload) |
| Listening levels:                     | 3 (P + 10, P, P – 10)               |
| Transcodings:                         | 1 transcoding                       |
| Error rates:                          | 1 (1:1000)                          |
| Environmental noise levels (sending): | 2 (< 30 dBA and “y”)                |
| Room noise levels:                    | “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 5 to 7 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 levels:                  | 2 (20 and 38 dB below overload)                          |
| Listening levels:                     | 3 (P + 10, P, P – 10)                                    |
| Transcoding:                          | 1 transcoding  |
| Environmental noise levels (sending): | 1 (< 30 dBA)   |
| Bit rates:                            | “r” (e.g. 48, 56 and 64 kbit/s for Recommendation G.722) |
| BERs:                                 | 3 (0, 1:10000 and 1:1000)                                |
| Room noise levels:                    | 1 (< 30 dBA)   |

- b) *Experiment 2 – Effect of transcodings, input level and listening level*
- |                                       |  |
|---------------------------------------|--|
| Speech input levels:                  | 2 (20 and 38 dB below overload)                          |
| Listening levels:                     | 3 (P + 10, P, P – 10)                                    |
| Transcodings:                         | “x”, these include synchronous and asynchronous          |
| Environmental noise levels (sending): | 1 (< 30 dBA)   |
| Bit rates:                            | “r” (e.g. 48, 56 and 64 kbit/s for Recommendation G.722) |
| BERs:                                 | 3 (0, 1:10000 and 1:1000)                                |
| Room noise levels:                    | 1 (< 30 dBA)   |
- c) *Experiment 3 – Effect of mismatch, input level and listening level*
- |                                       |  |
|---------------------------------------|--|
| Speech input levels:                  | 2 (20 and 38 dB below overload)        |
| Listening levels:                     | 3 (P + 10, P, P – 10)                  |
| Transcodings:                         | 1 transcoding                          |
| Environmental noise levels (sending): | 1 (< 30 dBA)                           |
| Bit rates:                            | “m” (56 → 64, 48 → 56, 48 → 64 kbit/s) |
| BERs:                                 | 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;  
           “r”    is the number of bit rates at which the codec operates;  
           “m”    is the number of mismatch bit rate combinations at which the codec will operate.

All three experiments must also include the wideband MNRU conditions.

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

## Annex C

### PCM codec corresponding to one qdu

(This annex forms an integral part of this Recommendation)

#### C.1 Common analogue interface

The common analogue interface (Recommendation G.192) is essential to ensure identical input and output conditions for testing candidate algorithms for speech coding and speech processing. It is also expected to provide the exact 1-qdu for estimating network quality of the target system.

#### C.2 Digital interface

The digital interface connects the candidate speech coding or speech processing equipment to the common analogue interface. This digital interface conforms to the configuration in the new Recommendation on hardware tools. This interface has time sampling of 16 kHz and quantization precision of up to 16 bits.

#### C.3 Analogue interface

The analogue interface consists of A/D and D/A converters, a filter that provides attenuation distortion of PCM terminal equipment, G.711 log-PCM encoding and decoding, and the MNRU of Recommendation P.810 to make the exact 1-qdu distortion.

##### C.3.1 A/D and D/A converters

The A/D converter converts an analogue signal to uniform PCM data with a minimum precision of 13 bits and maximum precision of 16 bits at a sampling rate of 16 kHz. The input to the A/D must include anti-alias filtering, and the output of the D/A must be passed through a reconstruction filter. This filtering has a flat frequency response within the pass-band

and sufficient loss in the stop-band. The distortion assumed for this part is no more than the theoretical quantizing distortion of uniform PCM. The D/A converter should be balanced with the A/D converter.

### C.3.2 Attenuation-frequency distortion

The attenuation-frequency distortion of the sending or receiving side of the interface should conform to the mask of Figure 3/P.66. However, to limit any attenuation distortion in the pass-band due to tandeming, the amplitude of pass-band ripple should be less than 0.1 dB.

### C.3.3 MNRU

Total distortion, corresponding to one qdu, is provided by a PCM codec having a 35 dB signal-to-distortion ratio (S/D) when the input signal is in the range of 0 to -30 dBm0 and measured by the method shown in Recommendation O.132. The S/D ratio for the common analogue interface is chosen as 2 dB above the mask of Recommendation G.712 and is thought to be equivalent to a typical commercial codec (see Recommendation G.712).

Since such a typical PCM codec is a hypothetical device, it is difficult to design a codec in hardware that provides the exact S/D ratio. If the A/D and D/A converters are assumed to be ideal and the log-PCM coding is achieved by a table lookup, this codec performs significantly better than typical commercial PCM equipment. To provide practical emulation of a typical PCM codec, an MNRU (conforming to Recommendation P.810) with  $Q = 36$  dB should be embedded between the A/D converter and log-PCM encoder, and between the log-PCM decoder and D/A converter, as shown in Figure C.1.

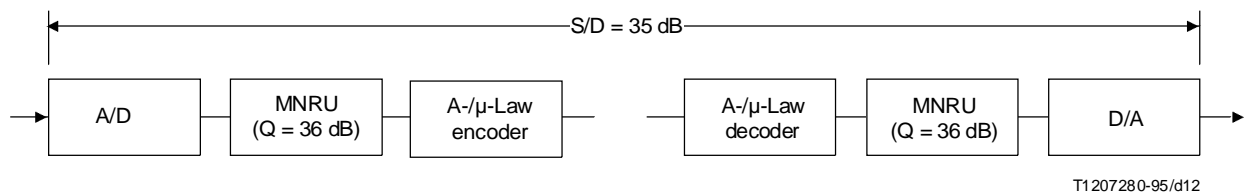


FIGURE C.1/P.830  
Configuration for one qdu

The system should be calibrated according to the following procedure, which assumes that a tolerance of  $\pm 0.1$  dB for the overall S/D ratio of 35 dB is acceptable:

- 1) Set both MNRUs to  $Q = 36$  dB;
- 2) Measure the S/D ratio of the complete system by the method of Recommendation O.132, letting the result be denoted by  $Q_0$ ;
- 3) If  $34.9 < Q_0 < 35.1$ , then calibration is complete, otherwise;
- 4) Add  $(Q_0 - 35)$  dB to the setting of each MNRU and return to step 2).

## Annex D

### Modified IRS send and receive characteristics

(This annex forms an integral part of this Recommendation)

The origin of the IRS send and receive frequency characteristics comes from an extensive series of measurements made on a number of contemporary analogue telephones in the early 1970s [3] and [4]. From these measurements, average send and receive frequency response characteristics were derived. However, for the loudness balance purposes for which the IRS is designed, it was also necessary to include a 300-3400 Hz bandpass filter, known as the SRAEN filter. The values of send and receive sensitivity currently given in Recommendation P.48 are therefore composed of the average send and receive responses for a number of telephones, as well as the response of the SRAEN filter.



While these responses may have been representative of connections over analogue long-distance transmission facilities, they are not representative of connections over modern digital facilities. In particular, the low frequency roll-off gives rise to unnecessary quality degradation. For the purposes of low bit-rate codec evaluation, especially where the codec is located in the telephone handset, a better characteristic can be obtained by modifying the IRS response by removing the SRAEN filter, as shown in Table D.1. For information purposes, the response of the SRAEN filter is also included in Table D.1. Tolerance on the nominal points shown in Table D.1 for modified IRS send and receive characteristics is  $\pm 2.5$  dB for frequencies between 200 Hz and 3400 Hz. Below 200 Hz, the roll-off of the modified IRS should be at least 15 dB/octave. Above 3400 Hz, appropriate low-pass filtering should be applied to satisfy anti-aliasing or reconstruction requirements. Implementations of these filters are available in the ITU-T STL.

TABLE D.1/P.830

Frequency (Hz)	Modified IRS Send (dBV/Pa)	Modified IRS Receive (dBPa/V)	SRAEN Filter Insertion Loss (dB)
100	-31.7	-13.4	14.1
125	-24.7	-7.4	11.4
160	-17.2	-2.4	8.4
200	-13.3	3.2	5.9
250	-10.3	6.7	4.0
300	-8.5	9.2	2.8
315	-8.3	9.7	2.5
400	-7.0	11.3	1.4
500	-6.3	11.9	10.6
600	-6.0	12.1	0.3
630	-5.9	12.1	0.2
800	-4.9	12.3	0.0
1000	-3.7	12.6	0.0
1250	-2.3	12.5	0.0
1600	-0.5	13.1	0.1
2000	0.1	12.9	-0.2
2500	1.3	12.6	-0.5
3000	2.0	13.0	0.5
3150	2.1	12.9	0.3
3500	-0.3	10.9	7.0
4000	-3.5	2.1	33.7
5000	-9.0	-11.7	43.2

## Bibliography

- [1] BOYD (I.), SOUTHCOTT (C.B.): A speech codec for the Skyphone service, *British Telecom Technology Journal*, Vol. 6, No. 2, April 1988.
- [2] BT Laboratories: Enhanced Equivalent-Q Rating Algorithm, *ETSI/TM/TM5/TCH-HS Document TD 93/126*, December 1993.

- [3] CCITT Contribution COM XII-79: *Specification for an Intermediate Reference System*, Study Period 1973-1976.
- [4] CCITT Contribution COM XII-104: *Recapitulation and analysis of the results of subjective and objective Loudness Rating measurements carried out with eleven telephone systems by the CCITT Laboratory*, Study Period 1973-1976.
- [5] CCITT Report of the meeting of Working Party XVIII/2 (Speech processing), *COM XVIII-R 28, Annex 1*, pp. 13-39, December 1983.
- [6] COLEMAN (A.), GLEISS (N.), SOTSHECK (J.), USAI (P.), 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.
- [7] COLEMAN (A.), GLEISS (N.), USAI (P.): A Subjective Testing Methodology for Evaluating Medium Rate Codecs for Digital Mobile Applications, *Speech Communications*, Vol. 7, pp. 151-166, June 1988.
- [8] COMBESURE (P.), *et al.*: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.
- [9] 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.
- [10] DAUMER (W.R.), CAVANAUGH (J.R.): A subjective comparison of selected digital codecs for speech, *Bell Systems Technical Journal*, Vol. 57, No. 9, November 1978.
- [11] DIMOLITSAS (S.), CORCORAN (F.), RAVISHANKAR (C.): Correlation between headphone and telephone handset listener opinion scores for single-stimulus voice coder performance assessments, *IEEE Signal Processing Letters*, Vol. 2, No. 3, March 1995.
- [12] European JEG: Subjective testing methodology for the evaluation of low-bit rate codecs for mobile radio, *CCITT, COM XII-68*, May 1986.
- [13] GOODMAN (D.J.), NASH (R.D.): Subjective quality of the same speech transmission conditions in seven different countries, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.
- [14] ITU-T Contribution COM 15-20: Transmission quality of interconnected PSTN-digital cellular networks, *COMSAT*, Study Period 1993-1996.
- [15] MODENA (G.), COLEMAN (A.), USAI (P.), COVERDALE (P.): Subjective performance evaluation of the 7 kHz audio coder, *IEEE Global Telecommunications Conference 1986 (Globecom '86)*, Houston, Texas, 1-4 December 1986.
- [16] RICHARDS (D.L.), BARNES (G.J.): Pay-off between quantizing distortion and injected circuit noise, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Vol. 2, Paris, May 1982.
- [17] Swedish Telecommunication Administration Report: Subjective test on candidate codecs for mobile radio, *CCITT, COM XII-147*, February 1987.
- [18] WHEDDON (C.), LINGGARD (R.): *Speech and Language Processing*, Chapman and Hall, 1990.