5i’

Supplement No. 10

CONSIDERATIONS RELATING TO TRANSMISSION CHARACTERISTICS FOR ANALOGUE HANDSET TELEPHONES

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

1 Introduction

This Supplement based on reference [9] summarizes available information on how some characteristics for handset telephones can be optimized.

It contains information about sending and receiving sensitivities, frequency responses, sidetone characteristics, influence of impedance and handset dimensions. It must be remembered that there are different ways to make an optimization. For instance the number of degrees of freedom are essential. As there are different opinions in different countries (for instance, the different assumptions made) the results of the optimization will be different. This Supplement touches some of these aspects.

2 Receiving frequency response

Most Administrations seem to prefer a fairly flat frequency response between 300 Hz and 3400 Hz. This probably derives from the early days of telephone networks, when it was determined that possible pre-emphasis at higher frequencies should be located at the sending end to obtain the best possible overall signal-to-noise performance. If we consider free-field, two-ear listening as a reference (face to face conversation) and assume a frequency-independent (flat) response, we should in principle simulate these conditions also at one-ear telephone listening

Then, at the earphone listening, we should have a frequency response of the earphone as in Figure 1 to simulate the diffraction effect we have at free-field two-ear listening [1]. However, most Administrations seem to prefer a flat response and to put the corresponding correction at the sending end. It may also be easier to construct a receiver with high efficiency if the goal is a flat response. Reference [2] has suggested a response as in Figure 2 optimized for a mean local line. Where mains noise may cause problems, a response with greater loss at lower frequencies, e.g. at 200 Hz and lower frequencies, may be appropriate.

Figure 1 Sup.10, p.1

Figure 2 Sup.10, p.2

3 Receiving sensitivity

Receiving sensitivity today often is represented by values between an RLR of —4 dB and —12 dB respectively.

A further increase of the sensitivity by the use of amplifiers might technically be possible. However, the probability for the audibility of crosstalk will increase with increased sensitivity. Therefore, the information gathered in Recommendation P.16 must be considered and it is doubtful if it can be recommended to increase the sensitivity further beyond an RLR of —12 dB.

Increasing the receiving sensitivity also decreases the margins against the effects of speech-off noise on the connection, e.g. unwanted modulation products from PCM systems. The stability against singing will also be affected.

4 Sending frequency response

Having chosen the receiving response to be flat, the sending frequency response can be optimized to give the proper overall characteristic. Reference [3] suggests an optimization achieved by asking the listeners for the ‘‘preferred’’ response. The result is shown in Figure 3. Reference [4] suggests a 2 to 3 dB increase per octave with increasing frequency. This result was obtained in tests regarding ‘‘naturalness’’. Reference [2] suggests a steeper curve (Figure 4) as a result of an optimization where maximum loudness, minimum listening effort and lowest output level are combined. The degree of freedom used by [2] is of course less than in [3] and [4]. Here we may have a difference in opinion concerning which assumptions we must include in the optimization. If the signal-to-noise ratio is a problem, some decibels could be gained (without overloading) in the way shown by [2]. If there are no signal-to-noise ratio problems, an optimization for best naturalness as in [3] and [4] can be used. Thus, the result will depend on the assumptions.

Different opinions may also exist about the local cable length for which the frequency response should be optimized and if the high frequency loss at long lines should be compensated. Reference [2] suggests optimization of the mean local line which will be optimum to the highest number of subscribers (because of the statistical distribution of cable lengths).

Figure 3 Sup.10, p.3

Figure 4 Sup.10, p.4

The curves according to Figure 4 and [4] give with a flat receiving frequency response an overall characteristic close to what is obtained by the diffraction effect at free-field listening this is probably not the whole explanation to the preferred curves. Even if the receiving responses were flat during sealed measuring conditions, hardly anyone keeps the earphone tight to the ear during conversation. Therefore, the actual responses during conversation probably give some additional low frequency cut-off that certainly has an influence on the results (see also reference [5]).

5 Sending sensitivity

When we want to choose the sending sensitivity we have one degree of freedom less that at the receiving end. We must consider both the probability of crosstalk and the probability of overloading other parts of the telephone system. Actual output levels from the telephone must be considered. As shown in [6] different output levels for the same SRE-value have been found in different countries. However, the different results show one important feature in common: output levels during normal conversation are generally lower than during reference equivalent measurements. Hopefully we will get better agreement on this point in the future if we use the measuring distance defined in Recommendation P.76, Annex A for loudness rating measurements.

6 Regulation

A possibility to increase the sending sensitivity on long lines exists if we use sending regulation dependent on line length. The probability for overloading and the probability for far end crosstalk will not increase if the mean power is kept to the same value as today. See also [2]. The probability of near end crosstalk in the local cable will of course increase and has to be considered.

If regulation is introduced both at sending and receiving, more subscribers may experience an overall loudness rating close to a preferred optimum, i.e. less calls will be rated poor and unsatisfactory. Another reason to introduce regulation is to obtain a better sidetone performance on short and long lines at the same time.

7 Impedance presented to the line

Some considerations concerning this topic are as follows:

— a conjugate match with the line maximizes the power transferred but creates sidetone problems on short lines and also stability/echo problems on long-distance calls;

— an image match to the line reduces the range of impedance presented to the exchange and eases the sidetone problem except for short subscriber-lines connected to resistive junction plant (e.g. PCM circuits);

— an impedance approximating the reference resistance (e.g. 600 ohms) eases standardization problems particularly in respect of alternative uses of the local line for non-speech services, but the optimum in respect of sidetone cannot be attained over the whole range of local line lengths.

References [2], [7] and [11] touch upon this subject.

8 Sidetone balance impedance

The degree of sidetone suppression is governed by the following parameters:

— microphone sensitivity;

— earphone sensitivity;

— sidetone balancing arrangement within the telephone instrument circuit;

— the impedance of the line to which the telephone is connected.

The microphone and earphone sensitivities and the instrument circuit are in part controlled by the required sending and receiving sensitivities. The impedance of the line to which the telephone is connected is not usually within the control of the telephone instrument designer. The only parameter freely available to the telephone designer to control the sidetone level is $Z_{S\backslash \mathrm{d}O}$, the sidetone balance impedance [7], [8], the impedance which when connected to the telephone completely suppresses sidetone (see also ref. [12]). If a transformer hybrid is used in the telephone then the internal balance network impedance is equal to the sidetone-balance impedance $Z_{S\backslash \mathrm{d}O}$ modified by the turns ratio of the transformer. However, the concept $Z_{S\backslash \mathrm{d}O}$ is not affected if the circuit uses any other form of balancing arrangement instead of a transformer.

9 Interworking with the existing network

The design of new handset telephones to be introduced into the telephone network must take account of the need to give satisfactory transmission on connections to existing local telephone circuits either directly or via the long-distance network. Reference [7] contains information touching upon this aspect.

Reference [10] is an example of a specification used in North America. Guidance for desirable sending and receiving levels are given as well as characteristics to be minimally acceptable for connection to the public switched network. It should be noted that this specification uses IEEE terminology, which is different from that found in CCITT Recommendations.

References

[1] CCITT Recommendation *Description of the ARAEN* , Green Book, Vol. V, Rec. P.41, Fig. 4, ITU, Geneva, 1972.

[2] CCITT Contribution COM XII-No. 32 (U.K. Post Office), Study Period 1973-1976.

[3] CCITT Contribution COM XII-No. 22 (Australia), Study Period 1973-1976.

[4] GLEISS (N.): Sound transmission quality, Tele. No. 1, 1972, pp. 44-53.

[5] CCITT Contribution COM XII-No. 229 (Sweden), Study Period 1985-1988.

[6] CCITT Recommendation *Subjective effects of direct crosstalk; thresholds of audibility and intelligibility* , Yellow Book, Vol. V, Rec. P.16, ITU, Geneva, 1981.

[7] CCITT *Manual Transmission planning of switched telephone networks* , Chapter V, Annex 1, ITU, Geneva, 1976.

[8] RICHARDS (D. | .): Telecommunications by speech, Chapter 5, *Butterworths* , London, 1973.

[9] CCITT Contribution COM XII-No. 105 (LME), Study Period 1973-1976.

[10] EIA Specification RS 470.

[11] CCITT Contribution COM XII-No. 144 (British Telecom), Study Period 1981-1984.

[12] CCITT Handbook on *Telephonometry* , Geneva, 1987.

Supplement No. 11

SOME EFFECTS OF SIDETONE

(Malaga-Torremolinos, 1984; amended Melbourne, 1988

(referred to in Recommendations P.11 and P.79)

1 Introduction

Over a number of years sidetone has been studied in CCITT Study Group XII under Question 9/XII. Some important conclusions have been reached from the point of view of the subscriber in his role as both talker and listener. These conclusions relate to the effect of sidetone on a subscriber, as he hears his own voice, the way his talking level changes as a result and some effects of sidetone when the subscriber is listening in conditions of moderate to high-level room noise. These effects are summarized in Figures 1 and 3.

2 Talker sidetone

Figure 1 shows that there is a preferred range for sidetone when the subscriber is talking under quiet conditions, and that the difference between the sidetone being objectionable or too quiet is of the order of 20 dB. (These results were obtained from talking-only tests and need to be confirmed by conversation tests.) The preferred range lies between 7 and 12 dB, STMR (sidetone masking rating — Recommendation P.76) [1], [5].

Figure 1 Sup.11, p.

The acceptable range is wider and lies between an STMR of 1 dB and 17 dB, (although it must be stated that increasing STMR to a value greater than 17 dB is likely to affect only the talking level, and that only marginally). This range corresponds to the difference between the two curves at the 50% appraisals level. It is not proposed that the 17 dB figure should in any way be considered a maximum value. However, for an STMR above 20 dB, the connection sounds ‘‘dead’’.

For telephone connections where the OLR is in the preferred range, the STMR values may similarly be positioned in the preferred STMR range given above. However, on high loss connections the STMR value should be close to, or even exceed 12 dB. On low loss connections the STMR value may be sometimes permitted to become less than 7 dB, but only rarely should it become as low as 1 dB, e.g. telephone sets with receive volume control. Recommendation G.121 interprets those results for transmission planning purposes.

Figure 2 shows the way in which the talking level changes with sidetone level [1], [2], [3], [4]. These results were obtained by means of conversation tests [6], for a connection close to the preferred overall loss. The speech voltage will also be a function of room noise for the same connection conditions.

Figure 2 Sup.11, p.

3 Listener sidetone

High room noise in the subscriber’s environment disturbs the received speech in two ways:

i) noise being picked up by the handset microphone and transmitted to the handset receiver via the electric sidetone path,

ii) noise leaking past the earcap at the handset receiver.

Studies have shown that at low frequencies the earcap leakage path dominates over the electric sidetone path in much the same way as the human sidetone signal does in talker sidetone. The weightings applied in the STMR loudness calculation are therefore applicable and the listener sidetone rating (LSTR, Recommendation P.76) has been developed, which makes use of the room noise sidetone sensitivity (see Recommendation P.64, § 9) in the STMR rating method (Recommendation P.79).

Results of subjective tests from two Administrations [7], [8] (using in this case a mean opinion scale of 0-10) are given in Figure 3. In each case the LSTR was derived by making use of $\Delta_{S\backslash\mathrm{d}m}$(see Recommendations P.10, P.64, P.79 and the *Handbook on Telephonometry* , § 3.3.17c) to convert the sidetone sensitivities $S_{m\backslash\mathrm{d}e\backslash\mathrm{d}s\backslash\mathrm{d}t}$ to $S_{R\backslash\mathrm{d}N\backslash\mathrm{d}S\backslash\mathrm{d}T}$ before calculating LSTR (Australian results) or applied as a weighted correction to STMR (Swedish results) as described in Recommendation G.111, § A.4.3.3. Room noise levels were comparable at 55-59 dBA.

Based upon these results Recommendation G.121 recommends that a value of 13 dB LSTR should be striven for.

The value 13 dB is based on a 10 dB LSTR (which may be considered a minimum value), where no further improvement in mean opinion score was possible by increasing LSTR (Figure 3), plus an allowance of 3 dB reflecting the fact that room noise in some office locations can exceed the values used in these experiments. Other tests (Sweden) have also suggested that a higher figure might be more appropriate.

The value that is satisfactory in a given telephone connection will depend on such factors as the level of room noise, the OLR of the connection, the talking levels used, etc. This is still under study in Question 9/XII.

Figure 3 Sup.11, p.

References

[1] CCITT Contribution COM XII-No. 50, Study Period 1977-1980 (ITT).

[2] CCITT Contribution COM XII-No. 171, Study Period 1977-1980.

[3] CCITT Contribution COM XII-No. 199, Study Period 1977-1980 (Australia).

[4] CCITT Contribution COM XII-No. 116, Study Period 1977-1980 (Hungary).

[5] CCITT Contribution COM XII-No. 152, Study Period 1981-1984 (NTT).

[6] Results of conversation tests sent directly to Special Rapporteur for Question 9/XII, British Telecom, 1978.

[7] CCITT Contribution COM XII-No.151, Study Period 1981-1984 (Australia).

[8] CCITT Contribution COM XII-No.70, Study Period 1985-1988 (Sweden).

Supplement No. 13

NOISE SPECTRA

(Malaga-Torremolinos, 1984)

(quoted in Recommendations P.44 and P.45 (Orange Book, Volume V)

and Question 24/XII)

(Contribution from British Telecom)

1 Introduction

This Supplement gives the descriptions of noise spectra used in the evaluation of telephony transmission performance that are recommended by the CCITT or have been employed in studying questions assigned to Study Group XII.

Controlled environmental noise is used in subjective evaluations such as:

a) AEN determinations as described in Recommendations P.44 [1] and P.45 [2];

b) conversation and listening experiments as described, for example, in Supplement No. 2 [3].

Spectra for two different environments are described, one for room noise and two for internal vehicle noise.

2 Room noise

The room noise should have a power density spectrum corresponding to that published by Hoth [4]. Table 1 gives the spectrum density adjusted in level to produce a reading of 50 dBA on a sound level meter conforming to IEC Recommendation Publication 179 [5]. This is reproduced in Figure 1. This spectrum is independent of level, i.e. for 40 dBA the level in each band will be 10 dB less than that shown in Table 1. Additional information on the power in each 1/3rd octave band is also given in Table 1.

3 Internal vehicle noise

Two spectra representing internal vehicle noise [6], [7] have been recommended for use in the study of Question 24/XII [8] for evaluating mobile radio systems. They are adequately represented by simplified curves [9]; one spectrum for moving vehicles and the other for stationary vehicles. Table 2 gives the spectrum densities together with additional information on the power in each 1/3rd octave band. The spectrum density for moving vehicles is shown in Figure 2 | fIa) and for stationary vehicles in Figure 2 | fIb) . These spectra are independent of level.

Table 3 gives the computed values of the unweighted sound pressure levels for various speeds calculated over the ISO 1/3rd octave frequency bands centred on 63 Hz to 8000 Hz.

H.T. [T1.13]
TABLE 1
Room noise spectrum

Frequency (Hz) Bandwidth 10 log 1 0 Δ*f* (dB) } Total power in each 1/3rd octave band (dB SPL) }	Spectrum density (dB SPL/Hz) { Tolerance (dB)	{		
100	32.4	13.5	45.9	
125	30.9	14.7	45.5	
160	29.1	15.7	44.9	
200	27.6	16.5	44.1	
250	26.0	17.6	43.6	
315	24.4	18.7	43.1	
400	22.7	19.7	42.3	
500	21.1	20.6	41.7	
630	19.5	21.7	41.2	
800	17.8	22.7	40.4	± |
1000	16.2	23.5	39.7	
1250	14.6	24.7	39.3	
1600	12.9	25.7	38.7	
2000	11.3	26.5	37.8	
2500	9.6	27.6	37.2	
3150	7.8	28.7	36.5	
4000	5.4	29.7	34.8	
5000	2.6	30.6	33.2	
6300	—1.3	31.7	30.4	
8000	—6.6	32.7	26.0	

Note 1 — The electrical input signal, e.g. white noise, shall be band-limited to the 1/3rd octave bands centred on the ISO preferred frequencies (ISO 266) between 100 Hz and 8000 Hz with the band edges conforming to the filters described in IEC 225.

Note 2 — The acoustical room noise is difficult to control at low frequencies, especially in the unspecified region below 100 Hz because of the dimensions of typical test cabinets, poor attenuation of such cabinets and the influence of extraneous noises, e.g. air-conditioning plant. It is therefore desirable to select a test cabinet that keeps these unwanted low frequency sound pressure levels to a minimum.

Tableau 1 [T1.13], p.8

Figure 1 Sup.13, p.9

Figure 2 Sup.13, p.10

H.T. [T2.13]
TABLE 2
Internal vehicle noise spectra

Frequency (Hz)	Spectrum density (dB SPL/Hz) { Tolerance (dB) Moving	Stationary	{	Stationary		
63	72.3	58.3	11.7	84.0	70.0	
80	69.3	55.0	12.7	82.0	66.7	
100	66.5	49.8	13.5	80.0	63.3	
125	63.3	45.1	14.7	78.0	60.0	
160	60.3	42.0	15.7	76.0	56.7	
200	57.5	36.8	16.5	74.0	53.3	
250	54.4	34.7	17.6	72.0	52.3	
315	51.3	32.6	18.7	70.0	51.3	
400	48.3	30.6	19.7	68.0	50.3	
500	45.4	28.7	20.6	66.0	49.3	
630	42.3	26.6	21.7	64.0	48.3	
800	39.3	24.6	22.7	62.0	47.3	± |
1000	36.5	22.8	23.5	60.0	46.3	
1250	33.3	20.6	24.7	58.0	45.3	
1600	30.3	18.6	25.7	56.0	44.3	
2000	27.5	16.8	26.5	54.0	43.3	
2500	24.4	14.7	27.6	52.0	42.3	
3150	21.3	12.6	28.7	50.0	41.3	
4000	18.3	10.6	29.7	48.0	40.3	
5000	15.4	8.7	30.6	46.0	39.3	
6300	12.3	6.6	31.7	44.0	38.3	
8000	9.3	4.6	32.7	42.0	37.3	

Tableau 2 [T2.13], p.11

–v'1P'

H.T. [T3.13]
TABLE 3
Computed sound pressure levels of spectra

Spectra Sound pressure level, unweighted (dB SPL) }	{
Moving 30 km/h	80
Moving 80 km/h	85
Moving 110 km/h	90
Stationary	75

Tableau 3 [T3.13], p.12

References

[1] CCITT Recommendation *Description and adjustment of the reference system for the determination of AEN (SRAEN)* , Yellow Book, Vol. V, Rec. P.44, ITU, Geneva, 1981.

[2] CCITT Recommendation *Measurement of the AEN value of a commercial telephone system (sending and receiving) by comparison with the SRAEN* , Yellow Book, Vol. V, Rec. P.45, ITU, Geneva, 1981.

[3] *Methods used for assessing telephony transmission performance* , Supplement No. 2, Yellow Book, Vol. V, ITU, Geneva, 1981.

[4] HOTH (D. | .): Room noise spectra at subscribers' telephone locations, *J.A.S.A* ., Vol. 12, pp. 499-504, April 1941.

[5] IEC Recommendation Publication 179, *Precision sound level meters* , 1965.

[6] CCITT Question 24/XII, Contribution COM XII-No. 120, (Noise inside light motor vehicles), Study Period 1981-1984.

[7] CCITT Question 24/XII, Contribution COM XII-No. 134, (Internal vehicle noise spectra), Study Period 1981-1984.

[8] CCITT Question 24/XII, Contribution COM XII-No. 1, (Link with mobile stations), Study Period 1981-1984.

[9] CCITT Contribution COM XII-No. 208, (Comparison of the results of vehicle noise submitted by France and BT), Study Period 1981-1984.

Supplement No. 14

SUBJECTIVE PERFORMANCE ASSESSMENT OF DIGITAL PROCESSES USING THE MODULATED NOISE REFERENCE UNIT (MNRU)

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

(quoted in Recommendation P.81)

1 Introduction

The primary purpose of this Supplement is to define a specific subjective 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 telephone network. 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.

Secondary purposes of the Supplement are to suggest ways in which the subjective test results can be treated to arrive at the assigned impairment level of a particular digital process and how this assigned impairment level can be used in transmission performance analysis.

2 Impairment reference scale for digital processes

Two reference scales that have been used for performance assessment of digital processes are a) continuous random noise (additive noise) and b) random noise with amplitude proportional to the instantaneous signal amplitude (multiplicative noise). Random noise with amplitude proportional to the instantaneous signal amplitude in terms of the Q ratio, according to the MNRU as specified in Recommendation P.81, should be used.

The reasons for this proposal are:

1) The signal processed through the MNRU is perceptually very similar in character to the signal processed through various digital processes, thus resulting, in principle, in easier assessment by test subjects, and

2) Considerable experience and information have been accumulated with the MNRU.

Note — It has not been documented that Q | represents a more suitable reference scale than continuous random noise.

3 Survey of methods

A number of methods are suitable for characterizing the performance of digital processes in terms of *Q* values. The methods deal with in this Supplement comprise listening-only tests. They are summarized in Table 1.

Other possible methods that may be mentioned are:

1) multiple paired comparisons between all systems under test and all reference condition /X_1/X_2 . | | X_i/X_j . | | X_j/R_i,

2) articulation test of MNRU conditions and digital systems in the same experiment.

These methods are not described here.

H.T. [T1.14]

TABLE 1

Indirect comparison with MNRU	Direct comparison with MNRU	
ACT method	DCR method	Equality Threshold

{ X 1 X 2 × × × X n R i }	(SSR)	{			
X 0/X 1 X 0/X 1 × × × X 0/X n X 0/R i }	(PC)	{			
X 1/R i X 2/R i × × × X n/R i }	(PC)				

Table 1 [T1.14], p.

4 Background for the test methods

4.1 *Background for the Absolute Category Rating (ACR) test method of Annex A*

The method is based on a procedure utilized in an experiment conducted by a working group of the IEEE (Institute of Electrical and Electronics Engineers) in which representatives from seven countries participated (Canada, France, Italy, Japan, Norway, the United Kingdom and the United States) [2]. The aim of this experiment was to determine whether comparable results could be obtained when the same test is performed in several countries. Speech samples in the native languages of the participating countries were processed at a central location through 38 communications circuits. The recordings of the processed speech were returned to each country for evaluation on a five-point category rating scale by native listeners.

The communications circuits included 22 references (continuous random noise, MNRU, μ-255 PCM) and 16 adaptive differential PCM (ADPCM) systems. (The type of ADPCM system used was a first order fixed predictor [3].) An important part of the data analysis was estimation of the quality of the 16 ADPCM conditions at one location given measurements of ADPCM quality elsewhere.

Results (mean opinion scores [MOS]) obtained at the different locations differed [2]. Nonetheless, analysis of the results indicated that a reasonably accurate estimate of ADPCM quality in country B is the quality measured in country A adjusted by an additive constant.

Changes in the methodology were discussed at an IEEE working group meeting in May 1982 in Paris. The methodology incorporating these changes was recommended by Study Group XII in June 1982 as a basis for evaluating candidate 32 kbit/s algorithms for CCITT standardization as Recommendation G.721 [1]. Subjective tests using the methodology were conducted under the auspices of Study Group XVIII in late 1982 with the results that a codec algorithm was selected and improvements (not related to telephone speech transmission issues) were identified. A second series of subjective tests in late 1983 confirmed that telephone speech transmission performance of the improved algorithm was suitable. (Differences between test results from the different participating organizations were also found in the 1982 and 1983 CCITT tests.)

The preceding discussion should not be taken to indicate that the subjective testing methodology is completely satisfactory: the reasons for differences found between countries [2] and [4] are thus far not explained. Nevertheless the testing methodology has the important feature of having been used by several countries.

4.2 *Background for the Degradation Category Rating (DCR) test method of Annex B*

A modification of the ACR test method, the DCR test method, is described in Annex B. Based on results from one Administration, the DCR test method provides a greater discrimination between conditions than does the ACR test method [5].

Results from an experiment conducted by another Administration do not support this conclusion [6].

5 Analysis of test results

The purpose in conducting test of digital processes is to determine their suitability for use in telephone networks. A procedure which has been used is to assign Q values, determined using the reference system of Recommendation P.81, to processes of interest. Various methods of data analysis are possible, but it appears desirable to define a single method to be used in order to assure expressing results in common terms. The provisional method is based on the use of MOS (mean opinion score) values obtained using the procedures of Annex A.

Hypothetical results obtained from a subjective test conducted according to the methodology of Annex A to this Supplement are shown in Figures 1 to 4. (Straight lines are used simply to connect data points.) Generally such results will display a saturation effect at and near the very good conditions (high MOS) and the very bad conditions (low MOS). (For high MOS, the saturation is caused by the 5-point scale and possibly by the idle circuit noise of the subjective test system without added impairments, e.g. idle circuit noise, and codec quantization distortion. For low MOS, the saturation is caused by the 5-point scale.) Experience [2] has shown that due to this saturation effect, acceptable accuracy for the determination of Q is obtained for the range of about 5 dB to 25 dB.

An objective of this analysis is to determine a function $Q_2 = F(L)$ where Q_2 is the Q | value for the code and L is the line bit rate. One simple method for determining this function uses the MOS values shown in Figures 2 and 3 and can produce a graph of this function as shown in Figure 5. The method is shown in Figure 6, 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 the other L values are obtained in a similar way and the resulting set of (L_i, Q_i) gains are plotted as in Figure 5.

Analysis of test results should include statistical analysis to establish that MOS values obtained are due to the test conditions and not to other factors. Student's test may be suitable, but there is some indication that analysis of variance is more appropriate.

The principles of a method of analysis used by one organization are outlined in Annex D of this Supplement. The method uses analytic values, called fit means , calculated from subjective test results; these analytic values are similar to MOS values calculated from test results. One desirable result of the test is estimates of the Q of the processes tested. Annex D contains a method for deriving such estimates.

Values of MOS versus Q | (as per Figure 2) obtained from actual experiments are given in References [5], [7] and [18] and in Annex B.

Figures 1 et 2 Sup.14, p.14-15

Figures 3 et 4 Sup.14, p.16-17

Figure 5 Sup.14, p.18

Figure 6 Sup.14, p.19

ANNEX A
(to Supplement No. 14)

Absolute category rating (ACR)

method for subjective testing of digital processes

A.1 *Introduction*

The listening-only test method consists in principle of three parts: preparation of source tapes; processing of the source tapes to obtain stimulus tapes containing the test conditions of interest; conduct of subjective tests using the stimulus tapes. Certain steps may be combined if interchange of source/stimulus tapes between locations is not involved.

The methodology is based on the notion of simulating a connection comprising a sending system, a receiving system and an interconnection system which provides for inserting the impairment of interest (idle channel noise and quantization distortion from the MNRU and from digital processes).

Listener responses in the subjective tests are influenced by a number of sources of variation, e.g. speech material, talker voice characteristics, presentation orders, time effects, etc. Unless controlled in some way, these variables may bias the outcome of the experiment. It is therefore recommended that appropriate experimental design be applied to take this into account. Principles for experimental design may be found in textbooks on statistics.

A.2 *Preparation of source tape(s)*

The recording system consists of a tape recorder, means for injecting calibration tones and a suitably defined sending system.

A.2.1 *Tape recorder*

The tape recorder should be a high (studio) quality two-track machine. The type of equalization should be stated, but IEC is preferred. One of the tracks is used for recording the speech samples; the second channel is available for other purposes, e.g. cueing tones to allow computer start/stop control of the tape recorder. The tape recorder should be operated at 19 cm/sec.

Low print-through, low-noise tape should be used and the tape should be stored ‘‘tail-out’’ so that it is necessary to rewind the tape before it is played.

Note — The use of an A/D converter and a television cassette recorder should be considered as a means for recording and storing high quality source and test tapes.

A.2.2 *Calibration tones*

It is recommended that calibration tones be recorded on the source tape(s) to enable checking the sensitivity/frequency characteristics of the connection simulation from input to the source tape recorder to output from the stimulus presentation tape recorder. Tones should be recorded in sequence at 250, 500, 1000, 2000, 3000 and 4000 Hz of 5 seconds duration each, with a level 6 dB below the maximum r.m.s. input level of the tape recorder. These tones should be followed by a 15 second recording of a 1 kHz test tone at maximum r.m.s. input level to enable calibration of the interconnection and listening systems. This should be followed by several metres of leader tape.

A.2.3 *Sending system*

The sensitivity/frequency characteristics of sending systems of different countries are likely to differ and, thus, results of different countries may differ because of attenuation distortion. Furthermore, the performance of complex digital codec algorithms may be dependent on the shape of the sending system sensitivity/frequency characteristics. Therefore, it is desirable that at least for some of the conditions in a test the sending system characteristic be as given in Table A-1 (simulates the IRS send part without filter).

H.T. [T2.14]
TABLE A-1
IRS characteristics before adding SRAEN filter

Frequency (Hz)	S M J (dB V/Pa)	S j e (dB Pa/V)
100	—22.00	—21.00
125	—18.00	—17.00
160	—14.00	—13.00
200	—10.00	— 9.00
250	— 6.80	— 5.70
315	— 4.60	— 2.90
400	— 3.30	— 1.30
500	— 2.60	— 0.60
630	— 2.20	— 0.10
800	— 1.20	+ 0.00
1000	+ 0.00	+ 0.00
1250	+ 1.20	+ 0.20
1600	+ 2.80	+ 0.40
2000	+ 3.20	+ 0.40
2500	+ 4.00	— 0.30
3150	+ 4.30	— 0.50
4000	+ 0.00	—11.00
5000	— 6.00	—23.00
6300	—12.00	—35.00
8000	—18.00	—53.00

Tableau A-1 [T2.14], p.20

It may be desirable to include conditions for which the sending system represents a typical (average) local system according to the testing organization’s (country’s) network and/or needs. This system comprises a handset telephone set, a simulated physical cable pair, a feeding bridge and a resistive termination (e.g. 600 ohms, 900 ohms) to which the source tape recorder is connected. The telephone set can utilize a linear telephone microphone with a real voice sensitivity/frequency characteristic such that the acoustic-to-electric response of the sending system represents the organization’s average local system. It may also be desirable to include conditions obtained with a carbon telephone microphone representative of the type(s) used in the organization’s (country’s) network. (See Recommendation P.64.) The characteristics (and feeding current) should be reported. It may also be desirable to report the characteristic measured using an artificial sound source. (See Recommendations P.51 and P.64.)

A.2.4 *Recording environment*

The recording environment should be that of a quiet living room or office. The ambient room noise level should be 25-30 dBA. The noise spectrum should, if possible, have the shape of the Hoth spectrum of Supplement No. 13. Special tests may be required using other noise levels and/or spectral characteristics (e.g. typewriter noise, etc.).

The room noise characteristic should be reported in as complete a form as is possible [e.g. dBA, long-term spectrum, amplitude/time distribution, etc.].

A 30 second recording of the room noise through the local system should follow the calibration tones. This should be accomplished with a talker holding the telphone handset in a normal use manner. (Special precautions may be necessary in order to avoid breath sounds if desired.)

A.2.5 *Speech samples*

A source tape is made of 4 × C samples (4 talkers, samples consisting of training, reference and test conditions). Each sample should comprise 2 or 3 sentences separated by at least 1 second.

All samples should be different to avoid repetition of sentences during a test. When reporting test results, it may be desirable to provide a list of the sentences used (i.e. 8 × C or 12 × C sentences).

Each sample is expected to be 6-10 seconds in length. The samples should be separated by 5 seconds of silence to allow for control (e.g. turning the tape recorder on and off) and of the amount of time needed for subjects to vote.

The r.m.s. level of the speech samples (speech power while active) should be 12 dB below the r.m.s. level of the 1 kHz calibration tone in order to avoid peak clipping of the speech samples by the tape recorder and to measure in an easy way the actual r.m.s. level of the speech.

A.2.6 *Talkers*

At least 4 different talkers (2 female, 2 male) with different voice characteristics should be used. Selection of the talkers will depend on the judgement of the experimenter.

A.3 *Preparation of stimulus tape(s)*

The interconnection system will consist of the source tape recorder (resistive, 600 or 900 ohms), an input filter, a means for inserting test conditions, an output filter, and the stimulus tape recorder (resistive, 600 or 900 ohms). The characteristics of the filters should be provided.

A.3.1 *Test conditions*

The test conditions comprise the digital codec(s) of interest. The codec(s) should be defined as simply and completely as possible (e.g. A-law/µ-law, ADPCM with first order fixed predictor, etc.). This is to enable unique performance specification for codecs of the same type.

Because codecs may have different performances at different speech input levels, they should preferably be tested not only at a nominal fully-loaded condition, but also at levels below and above this level, say ± | 0 dB. These changes in input level to codecs should be ‘‘off-set’’ by corresponding adjustments of their outputs to maintain an approximately constant output level for the test. (Listening level may also affect relative performance of different digital processes. See also § A.4.4.)

The codec(s) should be tested singly (one encoding/decoding pair) and with 2, 4 and (possibly) 8 codecs connected in tandem asynchronously. (It may also be desirable to include conditions in tandem synchronously.) The codecs may be hardware or software implemented; if the latter, injected circuit noise expected for practical codecs should be included.

For the single codec(s) conditions, the line bit rate should be the design value and, if possible, line bit rates both above (to ensure subjective saturation) and below (to ensure degraded performance). These conditions may be useful in assigning a performance level(s) to the codec(s). (For example, a nominal 32 kbit/s ADPCM algorithm might also be tested at 16, 24, 40 and 48 kbit/s.)

The tandem conditions should utilize the codec(s) at the design line bit rate(s).

Codec conditions with line errors should be included. Bit error rates covering the range $10^{\mathrm{D}}\mathrm{lF261}^{3}$ to $10^{\mathrm{D}}\mathrm{lF261}^{6}$ should be used.

A.3.2 *Reference conditions*

Reference conditions which should be included are *Q* | values within the range 5 dB to 25 dB with a minimum of 4 steps. (It may also be desirable to include *Q* values of 0 dB and 30 dB.)

It is desirable to include injected circuit noise values to provide SNRs within the range 5 dB to 45 dB with a minimum of 4 steps. (SNR is the dB ratio of speech power in milliwatts while active to injected circuit noise in milliwatts; the circuit noise conditions should be band-limited by filters having the same characteristics as the filter of the MNRU.) Note that the 45 dB ratio could be dependent on the inherent system noise, e.g. noise from the source tape preparation process, noise from the source tape recorder, etc.

Source conditions should also be included. (These are obtained by removing the injected idle circuit noise.)

The purpose of including the injected circuit noise conditions is to enable the relating of test results to results available on the effects of loss and circuit noise (Question 4/XII) and to allow use of the test results in subjective opinion model studies (Question 7/XII).

Other reference conditions can be included at the discretion of the testing organization. For example, particular organizations may have available information from previous tests of A-law/μ-law companded PCM, and it may be desirable to include some PCM conditions to allow comparison with previous results.

A.3.3 *Calibration*

The insertion loss of the interconnection circuit should be 0 dB at 1 kHz between the resistive source/termination. This should apply for the better conditions e.g. $Q = 25$ dB, SNR = 45 dB and the test codec(s) operated at design line bit rate(s).

The r.m.s. level of the 1 kHz calibration tone at the input to the inter-connection circuit should be 3 dB below the codec(s) overload level (which should be quoted). This will ensure that r.m.s. level of the speech samples will be 15 dB below the codec(s) r.m.s. sinewave overload level.

With the above calibration, the injected circuit noise levels in dBm across the output resistive termination should be adjusted to an appropriate level relative to the output 1 kHz calibration tone level in dBm. Note that in particular the circuit noise impairment should be present during the speech sample idle periods but not before and after the speech sample.

The stimulus tape recorder calibration should be the same as that for the source tape recorder.

A.3.4 *Stimulus tape(s)*

Stimulus tapes should begin with the 1 kHz calibration tone recorded (without introduced impairments), 12 practice conditions and then the test and reference conditions.

The practice conditions should be selected to introduce the test subjects to the test format and range of speech quality. These conditions should consist of each of the four talkers with 3 practice conditions.

The basic test and reference conditions will be 4 (i.e. number of talkers) times the number of nominal conditions. These conditions should appear in random order. There should be at least 2 stimulus tapes with different random orders. (These could be used in different tests with different subject groups.)

It may also be desirable to include replication of at least some of the test/reference conditions. However, this may not be possible for a practical subjective test size.

The timing of conditions in the stimulus tapes is the same as that for the source tapes, e.g. approximately 6-10 seconds (2 or 3 sentences) with each condition separated by 5 seconds of silence.

The calibration tones on the source tape need not appear on the stimulus tape (except for 1 kHz calibration tone as noted above). However, the calibration tone levels should be measured at the interconnection system output resistive termination so that the system sensitivity/frequency characteristics can be measured and reported for all condition types.

A.4 *Testing procedure*

A.4.1 *Listeners*

The preferred number of listeners is 32, assigned equally to each tape. At least 12 test subjects should be used. It is desirable that the subjects be selected to represent the typical customer population (e.g. half of the group females and half males, ages approximating the population distribution of ages, normal hearing, etc.).

A.4.2 *Listening system*

For reasons given in the first paragraph of § A.2.3, the receiving system characteristic should be as given in Table A-1 (simulates the IRS receive part without filter).

It would be desirable if the listening system simulated the organization's typical (e.g. average) local system representing the central office source impedance, feeding bridge, physical cable pair and the handset telephone set. The electric-to-acoustic sensitivity/frequency characteristic of the listening system should be determined (see Recommendation P.64). Sidetone in the listening system should be suppressed.

A.4.3 *Listening environment*

The listening handset(s) should be located in a room with an ambient room noise level 40 dBA, preferably 25-30 dBA (simulating a quiet office or living room). The noise spectrum should, if possible, have the shape of the Hoth spectrum of Supplement No. 13. The actual ambient room noise level and spectrum, if different from the above, should be reported.

A.4.4 *Speech level*

The 1 kHz calibration tone on the stimulus tape when played through the listening system should be adjusted such that reproduction occurs at a level of —3 dBPa as measured with the artificial ear recommended by the CCITT. (See Recommendation P.51.) This will result in a speech level of about —15 dBPa which is close to the preferred level. It may also be desirable to include conditions with a 10 dB lower level and 10 dB higher level since the listening level may affect the relative performance of different digital processes.

A.4.5 *Test instructions*

Test subjects will be provided with a written set of instructions which will also be read to them (either by the test administrator or by means of a tape recording). The instructions should be given before the practice conditions. Subjects should not be instructed that the practice conditions represent the full range of quality to be encountered in the test. After the practice conditions, there should be sufficient time allowed for answering possible questions by the subjects.

The subjects should be instructed to rate the conditions according to the five point quality scale as follows:

Score Quality rating 5 Excellent 4 Good 3 Fair 2 Poor 1 Bad

In countries for which English is not the native language, the appropriate terms in the native language should be used.

Before the listening test is conducted, it is necessary to carry out practice sessions to ensure full adaptation of listeners to the test conditions and obtain a stable evaluation.

There is some indication that a speech level of —5 dBPa (1 kHz tone level of +7 dBPa) would be more suitable than —15 dBPa for discrimination between coder conditions.

A.4.6 *Data collection*

Subjects' responses can be recorded by computer, on paper or by such other means as are appropriate. If paper and pencil are used, the response to each condition should be recorded on a separate card so that the subject is not looking at a previous opinion while making a new judgement.

A.5 *Results reports*

Reporting all of the raw data may be desirable but results in excessive documentation. Therefore, it may be appropriate to combine data across talkers and report the number of ratings in each of the 5 categories for each condition type, e.g. Q = 15 dB, SNR = 25 dB, etc. (Conclusions resulting from an analysis of the study of possible talker effects should be reported.) In addition, mean opinion scores (MOSs), standard deviation, 95 percent confidence intervals and other statistics computed by the organizations in analyzing the data should be reported.

Other items which should be reported are as follows:

a) microphone type;

b) sensitivity/frequency characteristic of the sending system (Recommendation P.64);

c) description of recording room and ambient noise levels;

d) measurement and adjustment procedure for speech levels;

e) sensitivity/frequency characteristics of the interconnection system for all test/reference condition types;

f) sensitivity/frequency characteristic of the listening system (Recommendation P.64);

g) description of the listening room and ambient noise level;

h) method of recording test subject opinions;

i) description of subject group including age, sex, population, prior experience and, if possible, audiometric threshold;

j) handset dimensions.

Bibliography for Annex A

KIRK (R. | .): Experimental design procedures for the behavioral sciences, *Brooks/Cole Publishing Company* , Belmont California, 1968.

CCITT Recommendation P.64.

CCITT Recommendation P.74.

ANNEX B
(to Supplement No. 14)

Subjective perfomance assessment of digital encoders

using the degradation category rating procedure (DCR)

(Contribution of the French Administration)

B.1 *Introduction*

A listening-only test method has been drafted by CCITT SG XII to assess the subjective quality of digital encoders (see Annex A). This procedure, Absolute Category Rating test (ACR), leads to a low sensitivity in distinguishing among good telephone quality coders (within the range of quality of 6-8 bit PCM coders). If higher sensitivity is needed we propose to use a modified version of that procedure, which can be defined as a Degradation Category Rating test (DCR). For image testing CCIR [6] recommends two alternative methods, absolute category ratings and degradation category ratings. The DCR procedure, which in particular uses an annoyance scale and a high quality reference before each judgement, seems to be suitable for evaluating good quality images. Therefore this method has been adapted to evaluate speech quality.

This Supplement first describes the adaptation of the DCR procedure to speech. Then the sensitivity of the method is compared with that of the ACR procedure on the same circuits. Only the differences between ACR and DCR procedure are presented here. One can refer to Annex A for common points which are not covered in this Annex.

B.2 *Degradation category rating procedure (DCR)*

B.2.1 *Speech samples*

Each configuration is evaluated by means of judgements upon four talkers reading two different samples. Each sample should comprise two sentences separated by at least one second. These two samples (S1, S2), hence four different sentences, should be selected from a wider corpus composed of phonetically balanced sentences so that the mean score obtained in evaluating MNRU circuits for these four sentences is about the same as that obtained for the wider corpus. Therefore the corpus consists of eight samples defined as follows:

talker T1 reading samples S1, S2

talker T2 reading samples S1, S2

talker T3 reading samples S1, S2

talker T4 reading samples S1, S2.

This results in a repetition of the two samples during the test. But we feel that this is not so critical for the procedure where a degradation is evaluated with regard to a reference, especially for good telephone quality where the intelligibility of speech is nearly perfect. The use of different samples for each configuration as is done in ACR experiments could be one of the reasons for this procedure's lack of sensitivity.

B.2.2 *Reference conditions*

Reference conditions should include multiplicative noise with Q | values within the range of 10 to 30 dB with a minimum of four steps. (It may also be desirable to include Q values of 5 dB and 35 dB).

A high quality reference should be chosen to be inserted before each judgement. Usually source conditions are used, i.e. samples with no more degradation than those introduced by sending systems and limitations of frequency bandwidth. Four ''null pairs'' (A-A) are included to check the quality of anchoring of the listeners' judgements.

B.2.3 *Stimulus presentation*

The stimuli are presented to listeners by pairs (A-B) or repeated pairs (A-B-A-B) where A is the high quality reference sample and B the same sample processed by a codec. The purpose of the reference sample is to anchor each judgement of the listeners. Using a reference and subjective judgements with respect to that reference is quite a common procedure in psychoacoustics. It tends to result in a good sensitivity for the overall evaluation by listeners. Samples A and B should be separated by 0.75 s and in a repeated pair procedure (A-B-A-B) the separation between the two pairs should be 2 s.

It seems that the classical order effect observed in a one-sample listening test (ACR for example) is not observed with the DCR procedure. Thus, only one random order of presentation can be used. Therefore the basic test and reference conditions will be eight times (four talkers $\times$ two samples) the number of nominal conditions.

The timing for the response of listeners is the same as for the ACR test, i.e. 5 s between each presentation (pair or repeated pairs).

B.2.4 *Test instructions*

The subjects should be instructed to rate the conditions according to the five point degradation category scale as follows:

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.

B.3 *Comparison between the sensitivity of an ACR and a DCR procedure for the same coder configurations*

Tables B-1 and B-2 summarize the results obtained with ACR test and DCR test respectively for the evaluation of three 32 kbit/s ADPCM algorithms.

Figures B-1, B-2 and B-3 show the mean opinion score (MOS) and degradation mean opinion store (DMOS) obtained by the same conditions with the two procedures (ACR and DCR respectively).

From these figures one can note:

— a good agreement between the results obtained with the two procedures;

— a larger spread of the DMOS obtained for MNRU circuits with Q values ranging from 10 dB to 35 dB, and a good anchoring of the judgements of listeners (‘‘null pairs’’ have obtained a score of 4.98);

— a higher sensitivity of the DCR procedure in the range of good telephone quality (20 < Q < 35 dB).

These sensitivities can be quantified by means of a statistical multiple comparison test. When an *a posteriori* comparison of codecs is needed a Tuckey [7] honestly significant difference (HSD) test can be applied effectively. The HSD test is designed to make all pairwise comparisons among the means and to determine the significance of the differences in the mean values. Under identical conditions ($\alpha = 0.01$, $k = 2$, $N = 225$, fixed mode) the HSD limit value ($q_{\alpha,k,N}$) is 3.70 and since the residual errors for ACR and DCR procedures are about the same (0.42), two means can be declared as significantly different if:

$\Delta = |$
X_i |
| (em
X_j |
| > 1.21

This difference, expressed in Q value, corresponds to:

H.T. [T3.14]

Range in Q (dB)	ACR test Δ	Range in Q (dB)	DCR test Δ
15 — 20	1.48	15 — 20	1.07
20 — 25	1.87	20 — 25	1.14
25 — 30	3.00	25 — 30	1.36

Tableau [T3.14], p.21

This means that the resolution of the DCR test may be twice that of the ACR test in terms of Q value in the range of good telephone quality.

B.4 *Conclusion*

A good agreement between the results obtained with the two procedures (ACR and DCR) has been found. The presence of a reference before each judgement for the DCR procedure ensures a good anchoring of the listener’s rating and consequently a larger spread of the degradation mean opinion score (DMOS) obtained by the coders. The evaluation of the coders based on the same speech samples leads to a better precision for the DCR procedure at a price, of course, of a decrease of the importance of the effort made to comprehend the samples in the overall quality judgement. Therefore the degradation category rating procedure seems well adapted to evaluate good telephone quality coders.

H.T. [T4.14]

TABLE B-1

Mean opinion scores (MOS) and 95% confidence intervals (INT) for ACR test

Test conditions	X		Y		Z	
	MOS	INT	MOS	INT	MOS	INT
PCM	3.81	0.45	3.89	0.13	4.16	0.13
PCM 2A	3.99	0.13	4.10	0.13	3.90	0.14
PCM 4A	3.35	0.12	4.02	0.14	3.70	0.14
PCM 8A	3.39	0.14	3.48	0.14	3.46	0.12
PCM 10^{D}lF261^{4}	3.31	0.15	3.55	0.14	3.15	0.16
PCM 10^{D}lF261^{3}	1.90	0.15	1.78	0.13	2.10	0.17
PCM + 10 dB	3.94	0.15	4.02	0.12	4.14	0.11
PCM — 15 dB	3.49	0.16	3.60	0.14	3.41	0.16
ADPCM	3.60	0.15	3.41	0.13	3.65	0.12
ADPCM 2A	3.72	0.13	3.30	0.12	3.38	0.13
ADPCM 4A	3.14	0.13	2.85	0.13	2.63	0.13
ADPCM 8A	2.51	0.14	2.09	0.14	2.23	0.15
ADPCM 2T	3.77	0.12	3.33	0.13	3.42	0.13
ADPCM 4T	3.86	0.14	3.01	0.14	3.80	0.13
ADPCM 10^{D}lF261^{4}	3.54	0.11	3.28	0.12	2.81	0.15
ADPCM 10^{D}lF261^{3}	2.88	0.16	2.55	0.15	1.93	0.13
ADPCM + 10 dB	3.80	0.14	3.55	0.14	3.61	0.13
ADPCM — 15 dB	3.20	0.15	3.02	0.15	2.92	0.14
ADPCM, C 2A	2.44	0.16	2.62	0.16	2.23	0.14
ADPCM, C 4A	2.13	0.15	2.14	0.13	1.90	0.13
ADPCM, C 8A	1.98	0.14	1.84	0.13	1.59	{
0.12						
S/N 40						
3.52						
0.15						
S/N 35						
3.18						
0.17						
S/N 25						
2.04						
0.15						
S/N 15						
1.23						
0.09						
Q 10						
1.41						
0.10						
Q 15						
2.34						
0.11						
Q 20						
3.04						
0.10						
Q 25						
3.61						
0.09						
Q 30						
3.96						
0.09						
}						

Note 1 — Votes combined across four speakers and two sentences.

Note 2 — Number of votes = 128 except for *Q* where *N* = 256.

Tableau B-1 [T4.14], p.22

BLANC

H.T. [T5.14]

TABLE B-2

Degradation mean opinion scores (DMOS) and 95% confidence intervals (INT) for DCR test

Test conditions	X		Y		Z	
	DMOS	INT	DMOS	INT	DMOS	INT
PCM	4.35	0.10			4.41	0.11
PCM 8A	3.48	0.16			3.33	0.15
PCM 10^{D}lF261^{3}	2.21	0.11			2.25	0.14
ADPCM	4.33	0.11	4.22	0.11	4.05	0.12
ADPCM 8A	2.63	0.14	2.35	0.14	2.38	0.17
ADPCM 10^{D}lF261^{3}	3.14	0.16	2.83	0.14	1.85	0.14
ADPCM 4T	4.29	0.10	3.69	0.14	4.09	{
0.13						
Q 15						
1.99						
0.15						
Q 20						
2.97						
0.17						
Q 25						
3.89						
0.18						
Q 30						
4.66						
0.10						
Q 35						
4.81						
0.09						
Origin						
4.98						
0.03						
}						

Note 1 — Votes combined across four speakers and two sentences.

Note 2 — Number of votes = 128.

Tableau B-2 [T5.14], p.23

Figure B-1, p.24

Figure B-2, p.25

Figure B-3, p.26

ANNEX C
(to Supplement No. 14)

Threshold method for direct comparison of digital encoders

with a modulated noise reference unit (MNRU)

C.1 *Introduction*

By direct comparison of a digital system with an MNRU it is possible to assess the *Q* value which equals the performance of the system under test. The method described here leads to a threshold of equality defined as the 50% preference level between the MNRU and the digital system.

The threshold method is expected to give stable and precise results even for high quality digital processes. For wideband digital encoders the use of a wideband MNRU as described in Annex A of Recommendation P.81 is recommended.

C.2 *Testing procedure*

A listening-only test procedure is used. A signal pair consisting of a reference signal and a test signal is presented to listeners, who are then asked to indicate which of the signals in the pair they judge to have the highest quality (preference rating). Subjective equivalent SNR (*Q*) is defined as the reference SNR corresponding to the intersection point of the regression curve of the preference scores at the 50% preference level. An example of *Q* obtained with hypothetical preference scores is shown in Figure C-1.

Figure C-1, p.

C.3 *Presentation of signals*

Reference signal A and test signal B are arranged in an equal number of A-B pairs and B-A pairs, and presented in random order. Several distortion levels spaced, for example, at 2 dB intervals, are introduced to the reference signal so that the range of preference scores extends from 20% to 80%, where the 50% preference lies in the middle of the distortion range. A timing diagram of the presentation is shown in Figure C-2.

Figure C-2, p.

The subject is required to make a judgement and respond by saying ‘‘A is better’’ or ‘‘B is better’’ (forced choice). The response ‘‘A equals B’’, or ‘‘No difference’’ is forbidden. The duration of the presentation should be limited to about six minutes in order not to tire the listeners. More listening samples may be presented after a suitable rest period. At least two, preferably four or five replications (repetitions of identical presentations) are recommended.

Note — If the MNRU is available in hardware and the SNR can be easily changed between presentations, a simplified procedure can be used. In this case the balancing to equally perceived quality is done by the subject. The adjustment is made during the pause between the pairs. The reference is always presented first. Presentation continues until the subject reports that the equality threshold has been reached.

C.4 *Speech sources*

It is necessary to use short sentences spoken by at least two males and two females, preferably four or six of each; different sentences are required for each speaker. The duration should be 2.5-5 seconds for speech and less than 10-15 seconds for music signals. Clicks at the beginning and end of the samples must be avoided. A linear microphone of sufficient bandwidth should be used to record the source signals in a sound-absorbent room having an ambient noise of less than 20 dBA and a reverberation time of less than 0.3 seconds in the band 125-8000 Hz. If digital recording equipment is used, the quantizing noise level should be less than the noise level in 14-bit linear PCM.

C.5 *Listening environment*

A high-fidelity sound reproduction system should be used for the listening test. When listening is carried out with loudspeakers, the reproduction equipment should be of studio-quality and the listening room should conform to CCIR Report 797. If headphones are used, diotic (binaural) listening is preferable. The bandwidth should be at least as wide as that of the digital system under test.

C.6 *Listeners*

Although it is preferred that listeners should be selected according to the description in the ACR method (see § A.4.1), this is not a strict condition in the pair comparison test. If the purpose of the listening test is to obtain the opinions of untrained listeners, untrained subjects are necessary. However, if this is not the purpose of the test, then trained listeners can be used and the reliability of the listening test can be extended by increasing the number of replications for each listener. The minimum number of listeners is six, but preferably twelve or more. Several subjects may listen simultaneously but it must be ensured that their responses are obtained independently.

C.7 *Reliability*

Since variations in preference score in subjective tests are assumed to conform to a *t* -distribution, the score variation width *r* which yields 95% reliability at score *u* (0*u*) over the number of trials (i.e. the number of repetitions for each presentation pair multiplied by the number of subjects number of source signals) is presented in equation (C-1).

$r = \pm$ | fIt (n — 1, 0.05) $\times$

$$\sqrt{f\,Iu\,(1\tilde{}(\mathrm{em}\,\tilde{}\,\mathrm{f\,Iu}\,)/(n\ \tilde{}(\mathrm{em}\,\tilde{}1)} \tag{C-1}$$

If n | equals 96 and u | equals 0.5 (preference score is 50%), r | equals ± | 0%.

ANNEX D

(to Supplement No. 14)

Principles of a method used by one organization in analyzing

digital codec performance (Bell Communications Research, Inc.)

Hypothetical mean opinion score (MOS) results obtained from a subjective test conducted according to the methodology of Annex A are shown in Figures 1 to 4 of this Supplement. (Straight lines are used simply to connect data points.) Generally such results will display a saturation effect at and near the very good conditions (high MOS) and the very bad conditions (low MOS). (For high MOS, the saturation is caused by the 5-point scale and possibly by the idle circuit noise of the subjective test system without added impairments, e.g. idle circuit noise, Q , codec quantization distortion. For low MOS, the saturation is caused by the 5-point scale.)

An analytic method of data analysis used by Bell Communications Research, Inc. provides a value called ‘‘ fit mean ´condition [8]. The fit means are then used in analysis of the data. (Fit means and MOSs are nearly equal over the mid-range of MOS values; however, fit means are not numerically constrained for extremely good and extremely bad conditions as are MOSs.) Plots of test results in terms of fit means will be similar to those of Figures 1 to 4, but will numerically exhibit greater spread.

The objective of the analysis is to determine a function:

$$Q_S = f(L) \tag{D-1}$$

where

Q_S = Q | alue for the codec quantization distortion ,

L = Line bit rate (e.g. in kbit/s).

(A simple linear relation may be possible in some cases while other cases may require a more complex function.)

The codec Q | an then be estimated from Equation (D-1).

Determination of Equation (D-1) needs to take into account in an appropriate manner the saturation effects discussed earlier. For example, the codec design line bit rate may correspond to the middle data point of Figure 3 for which there appears to be a modest saturation effect.

Similarly, the equivalence function (SNR *vs* Q) internal to the test may need to be considered. [This function is determined from appropriate functions fitted to the curves (in terms of fit means) similar to the curves of Figures 1 and 2.]

An important consideration in the analysis method is obtaining predicted performance values (fit means, Q) approximating as closely as possible the actual performance values (fit means similar in form to the curves of Figures 3 and 4 or Q values obtained by converting the fit mean values to Q values using an appropriate function fitted to fit mean data similar in form to the data of Figure 2). (For present purposes, it is assumed that for asynchronously tandemed codecs the combining law is 15 $\log_{10} n$, where n is the number of tandemed identical codecs. It may also be desirable to include the determination of the combining law in the analysis.)

The Q | alues obtained for a digital process according to the procedure described above can be used in various ways to assess the effect of quantization distortion on telephone connection performance. The subjective opinion model of Supplement No. 3 is in terms of corrected reference equivalent and idle circuit noise level. This model requires that the Q for an overall connection be converted to an equivalent SNR which can then be converted to an idle circuit noise level based on knowledge of the speech levels for connections of interest.

Equivalence functions used for this purpose have been found to vary (see [9], [10], [11], [12] and [13]). It is not clear if there exists a unique equivalence function which can be agreed on, and what that equivalence function should be. (Perhaps a basic equivalence function should be based on conversational test results.)

The Q | alues obtained for digital processes can also be used as a basis for specifying codec network application rules in terms of the number of asychronously tandemed 8-bit, μ-255 codecs. For the model of Supplement No. 3, the relation between the number of such codecs and Q (based on a 15 $\log_{1\backslash d0} n$ law) is as follows:

Number	1	2	3	4	5	6	8	10	12	14
Q | dB)	37	32.5	30	28	26.5	25.5	23.5	22	21	20

According to the model of Supplement No. 3, a 7-bit, μ-255 PCM codec would correspond to about 2.5 asynchronously tandemed 8-bit, μ-255 codecs. (Note that the value of $Q = 37$ dB is 3-4 dB greater than the minimum S/D values of Recommendation G.712 [15]; it is assumed that average 8-bit systems perform at the higher value.)

References

[1] CCITT Recommendation *32 kbit/s adaptive differential pulse code modulation (ADPCM)* , Vol. III, Rec. G.721, ITU.

[2] GOODMAN (D. | .), NASH (R. | .): 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.

[3] DAUMER (W. | .), CAVANAUGH (J. | .): A subjective comparison of selected digital codecs for speech, *Bell System Technical Journal* , Vol. 57, No. 9, November 1978.

[4] RICHARDS (D. | .), BARNES (G. | .): 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.

[5] COMBESCURE (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.

[6] CCIR Doc. 11/17 *Subjective assessment of the quality of television pictures* | EBU), Study Period 1978-1982.

[7] TUCKEY: The problem of multiple comparisons, *Ditton* , Princeton University, Ed. 1953.

[8] CAVANAUGH (J. | .), HATCH (R. | .), SULLIVAN (J. | .) Models for the subjective effects of loss, noise and echo on telephone connections, *Bell System Technical Journal* , Vol. 55, November 1976.

[9] CCITT Contribution COM XII-No. 24 (Study on determination of subjectively equivalent noise (NTT)), Study Period 1981-1984.

[10] CCITT Contribution COM XII-No. 61 (Subjectively equivalent noise for linear and carbon microphone originated speech signals — Bell Northern Research, Canada), Study Period 1981-1984.

[11] CCITT Contribution COM XII-No. 124 (Application of information index to quantizing noise in PCM — France), Study Period 1981-1984.

[12] CCITT Contribution COM XII-No. 130 (Subjective equivalence functions for consideration in a quantization distortion opinion model — Bell Northern Research, Canada), Study Period 1981-1984.

[13] CCITT Contribution COM XII-No. 162 (Transmission performance of digital systems — COMSAT), Study Period 1981-1984.

[14] CCITT Recommendation *Transmission impairments* , Yellow Book, Vol. III.1, Rec. G.113, ITU, Geneva, 1981.

[15] CCITT Recommendation *Performance characteristics of PCM channels audio frequencies* , Yellow Book, Vol. III.3, Rec. G.712, ITU, Geneva, 1981.

[16] CCITT Contribution COM XII-No. 222 (Evaluation of speech transmission quality by ACR and DCR methods — USSR Telecommunications Administration), Study Period 1985-1988.

This should not be interpreted as the number of qdus [14] to be used for international planning; the relation between the number of codecs and Q applies for the model of Supplement No. 3 which has been used in planning studies in the United States.

[17] CCITT Contribution COM XII-No. 223 (Results of determination of typical dependence of average subscriber's evaluation from *Q* values — USSR Telecommunications Administration) Study Period 1985-1988.

[18] CCITT Contribution COM XII-R5, page 49 and pages 76-78, June 1985.

Supplement No. 15

WIDEBAND (7 kHz) MODULATED NOISE REFERENCE

UNIT (MNRU) WITH NOISE SHAPING

(Melbourne, 1988)

(Quoted in Recommendation P.81)

(Contribution of NTT Japan)

1 Introduction

The configuration of a wideband MNRU takes into account the following three points with respect to its use as a common reference signal:

a) The procedure for generating the reference signal should be simple and clear;

b) The speech quality characteristics of the reference signal should be similar to those of the test signals;

c) It should be possible to control grades of degradation arbitrarily.

Compared to the narrow-band MNRU, the wideband MNRU has an enlarged bandwidth (70-7000 Hz) and a fixed noise spectrum shaping filter which lessens the high-frequency range noise, thus making the noise spectrum in the reference signal resemble that in wideband encoders.

2 Arrangement of the wideband MNRU

The basic arrangement of the wideband MNRU is shown in Figure 1. Wideband Gaussian noise instantaneously multiplied by source speech is fed to a spectrum shaping filter. The source speech and the shaped-spectrum noise are band-limited and attenuated/amplified to obtain the desired SNR, and then both are added to produce the distorted signal.

Figure 1 Sup.15, p.

3 Spectrum shaping filter

The noise spectrum is shaped with a 1st order auto-regressive filter, whose diagram is shown in Figure 2. A computer simulation using a sampling frequency of 16 kHz and a bandwidth of 7 kHz yields a filter coefficient of 0,8 which approximates the long-term speech spectrum envelope of wideband source signals.

Figure 2 Sup.15, p.

4 Band-pass filter

The bandwidth of the wideband MNRU should correspond to that of wideband speech encoders. The provisional frequency response requirements for the 7 kHz band-pass filter are shown in Figure 3 based upon the present output filter in Recommendation P.81.

Figure 3 Sup.15, p.

5 Signal-to-noise ratio

In a computer simulation, SNR can be calculated using mean power with a time constant of 8-20 milliseconds after band-pass filtering of both the source speech and noise. When the SNR is set using a sinusoidal source signal for the MNRU equipment, measurement of noise power should actually be carried out using an r.m.s. volt meter, rather than by simply calibrating the loss or gain of the noise channel based on the sinusoidal signal.

6 Other specifications

Sampling frequency: | It is recommended that the sampling frequency be more than twice the upper pass-band frequency; with respect to the test of Recommendation G.722 encoders, a sampling frequency of 16 kHz is recommended.

Noise source: | For MNRU equipment, sufficiently wideband Gaussian noise should be used. If the noise is computer-generated, a full flat spectrum over one half the sampling frequency band is necessary. Amplitudes greater than three r.m.s. value should be clipped.

Supplement No. 16

GUIDELINES FOR PLACEMENT OF MICROPHONES AND LOUDSPEAKERS IN TELEPHONE CONFERENCE ROOMS [1]

AND FOR GROUP AUDIO TERMINALS (GATs)

(Malaga-Torremolinos, 1984; amended Melbourne, 1988)

(Quoted in Recommendations G.172 and P.30)

1 General

The following guidelines provide basic rules for assessing the acoustics of telephone conference rooms and for installing group audio terminals consistent with maximum speech intelligibility and easy talker recognition.

2 Conference room acoustics — General requirements

The design and installation of a telephone conferencing system or group audio terminals which meet reasonable cost and performance specifications involve numerous judgements and trade-offs. These guidelines will enable the planner and installation engineer to assess the acoustics of a room, to make the necessary choices and decisions, to install the appropriate equipment properly and thereby provide satisfactory service.

The audio portion of a group audio terminal consists of terminal equipment with microphones and loudspeakers installed in conference rooms and interconnected by an audio transmission facility. This transmission facility may be either public switched telephone connections or private line facilities.

In both public and private systems, transmission is frequently interconnected via a multipoint conference bridge so that each room can communicate simultaneously with any of the other locations. When this is done, it is most important that the bridge be located at the electrical loss center of the network in order to minimize level contrast between the speech originating in the different rooms.

Formerly supplement No. 25 to Fascicle III.1 *(Red Book)* .

Unlike telephony between handsets, the acoustic properties of the conference room and the placement of microphones in the room critically determine the level, the speech signal-to-ambient-noise ratio and the reverberant quality (rain-barrel effect) of the transmitted speech. Particularly in multipoint conferences, these three factors are easily judged and critically commented upon by users.

In general, the larger, noisier and more reverberant a room is, the less suitable it will be for group communication. The presence of noise and/or reverberation in the transmitted speech results in a system whose performance is unsatisfactory. In extreme cases, experience has demonstrated that excess noise in one room, e.g. from overflying aircraft, can temporarily block transmission between all rooms of a multipoint system. Excess reverberation results in such hollowness to the received speech that talkers become difficult to recognize and understand, causing users to fatigue easily and refuse to use the system.

In principle, any room is suitable for group communication if these guidelines are followed. However, the guidelines will dictate that in a noisy or reverberant room, talkers must speak so close to microphones that they might as well use handsets. The user requesting the installation must then choose one or more of the following options:

1) select another conference room;

2) acoustically treat the room; or

3) accept the close microphone/talker distances dictated by the guidelines.

Several very important criteria must be fulfilled simultaneously to assure satisfactory audio performance of a telephone conference system. The balance of this section describes the determination of these criteria. Briefly, these criteria are:

1) A room suitable for a normal face-to-face conference must be selected.

2) A noise dependent microphone/talker distance must be determined.

3) A reverberation dependent microphone/talker distance must be determined.

4) The microphones and loudspeakers must be positioned in accordance with both these distances.

3 Ambient noise level considerations

The ambient noise level requirements for conference rooms of increasing size and number of conferences are given in Table 1. As the room size and number of conferences increase, the participants will sit further apart. Consequently, for comfortable talking and listening, the ambient noise level in the room must decrease as the group size increases.

H.T. [T1.16]

TABLE 1

Ambient noise level limits for conference rooms

Room description Maximum sound level meter reading }	{ Acoustic environment	
Conference room for 50 people Very quiet, suitable for large conferences at tables 6-9 m in length }	35	{
Conference room for 20 people Quiet, satisfactory for conferences at tables 4.5 m in length }	40	{
Conference room for 10 people Satisfactory for conferences at tables 1.5-2.5 m in length }	45	{

{
Conference room for 6 people
}
Satisfactory for conferences at tables 1.0-1.5 m in length
}
	50	{

Table 1 [T1.16], p.

Noise measurements as stipulated in Table 1 should be performed at the conference table with the room in normal operation but unoccupied. These noise measurements should be performed at least 0.6 m away from any surface.

Noise measurements in dBA can be made with a sound level meter employing A-weighting, a reference pressure level of 20 µ Pascal and otherwise conforming to Recommendation P.54. A-weighting is used in these guidelines since it approximates the annoyance level of noise to the human ear.

The maximum microphone/talker distance is limited by ambient noise. Figure 1 shows the maximum distance between a talker and a microphone which ensures a marginally acceptable signal-to-noise ratio of 20 dB in the transmitted speech. No attempt should be made to ignore or increase this distance beyond that determined in Figure 1. As an example, with an ambient noise level of 50 dBA, Figure 1 shows that the *maximum* distance ($D_{\mathrm{m\backslash da\backslash dx}}$) from talker to microphone for *marginal* acceptability is 0.5 m. Figure 1 applies to omnidirectional microphones. When directional microphones, e.g., cardioid or bidirectional are used, the $D_{\mathrm{m\backslash da\backslash dx}}$ value determined in Figure 1 can be increased by 50 percent.

Figure 1 Sup.16, p.

If more than one microphone is used to cover more than two or three talkers, and all microphones are active at the same time, then the amount of room noise picked up by the microphones and transmitted on the circuit will increase. How much it will increase is not completely predictable but a useful approximation is that the apparent noise level will rise 3 dB each time the number of microphones is doubled. This apparent rise in the effective noise level can be taken into account by adding it to the measured noise level before using Figure 1 to determine $D_{\mathrm{m\backslash da\backslash dx}}$.

4 Reverberation considerations

Most rooms for telephone conferencing have acoustical characteristics which cannot be altered, thus the quality of sound transmitted from the room can only be controlled by microphone placement. When the microphone is close to the talker, the greatest percentage of sound picked up comes directly from the talker, reverberation in the room would exert relatively little influence. As the distance between the microphone and the talker increases, the direct sound level reaching the microphone decreases 6 dB for each doubling of the distance, whereas the average level of the reverberant sound remains more nearly constant.

The critical distance (D_c) of a room is a useful concept to describe a room. It is the distance from a sound source (talker, loudspeaker) at which the direct sound energy from the source equals the reverberant energy reflected off all room surfaces (walls, ceiling, furnishings, floor). Critical distances in conference rooms are typically in the range of 0.2 to 1.5 meters.

The critical distance can be expressed as:

$$D_c = 0.056 \sqrt{\frac{V}{T_R}} \text{ meters}$$

(see ISO 35u)

where

V is the volume of the room in cubic meters,

T_R is the reverberation time of room in seconds.

As the ratio of direct-to-reverberant sound energy decreases with increasing microphone/talker separation, reproduced speech becomes less intelligible, of poorer quality, difficult to recognize and fatiguing to listen to. It acquires a hollowness which sounds as if the person were speaking from the bottom of a rain-barrel. For good performance, microphones should be placed at no more than half the critical distance (0.5 D_c) from talkers. This usually requires installing multiple microphones on the conference table or lavaliere microphones on conferees, and definitely rules out placing microphones in the ceiling. Many installations for group communication have failed because microphones were installed in ceilings without regard to the above acoustic requirements.

When directional (cardioid or bidirectional) microphones are used, the distance between microphones and talkers may be increased by 50 percent, to three-quarters of the critical distance (0.75 D_c). For best results, talkers must sit in front of cardioid (heart shape) microphones; they may sit on either side of a vertically mounted bidirectional microphone with a cosine (figure-eight shape) sensitivity pattern. Table 2 gives typical microphone/talker separation distances for small (60-300 m^2 of wall, ceiling and floor surface area) and large (300-1000 m^2) rectangular conference rooms, together with the estimated critical distance (D_c). Areas in square meters are used in these guidelines, since they are much more relevant to conference room acoustics than are the often quoted room volumes.

H.T. [T2.16]

TABLE 2

Typical microphone/talker separation (meters)

Conference room	Omnidirectional microphone	Directional microphone	Critical distance
{ Small room (60-300 m^2) moderate room treatment | ua) }	0.3	0.5	0.6
{ Large room (300-1000 m^2) some room treatment | ua) }	0.6	0.9	1.2
{ considerable room treatment | ua) }	0.9	1.4	1.8

a) In this context, a room with moderate treatment might have an accoustic ceiling and a carpet on the floor; one with some treatment might have either an accoustic ceiling or a carpet; while a room with considerable treatment might have heavy, lined drapes covering half the wall area in addition to a high-quality suspended acoustic ceiling and a thick carpet with underfelt.

Table 2 [T2.16], p.

Microphones with an attached, adjustable strap which can be hung around the neck of the user.

5 Microphone type and placement

As stated earlier, when omnidirectional microphones are used the microphone/talker distance must be less than the maximum distance ($D_{m\backslash da\backslash dx}$) determined from Figure 1 to ensure adequate signal-to-noise ratio. When directional microphones are used, the microphone/talker distance can be increased but must be less than 1.5 $D_{m\backslash da\backslash dx}$.

Also stated earlier, when using omnidirectional microphones, the microphone/talker separation must be less than half the critical distance to ensure highly-intelligible, easily-recognizable, nonreverberant speech. When directional microphones are used, the microphone/talker distance can be increased but must be less than 0.75 D_c.

Microphones must be placed to satisfy *both* | he above rules; in other words the microphone/talker distance must not exceed the smaller distance.

So that all talkers can satisfy the above microphone/talker criteria, more than one microphone is usually required. Typically one microphone for every 3 talkers is necessary. For each doubling of the number of microphones, the effective noise level in the room will increase by 3 dB. Thus, in the example of § 3 if four microphones were used, the reading of 50 dBA would be raised to an effective value of 56 dBA. The noise determined, $D_{m\backslash da\backslash dx}$ from Figure 1 would thus be reduced to 25 cm. Clearly, lavaliere microphones would provide a practical solution to keeping talkers within 25 centimeters of a microphone.

6 Loudspeaker placement

The requirements for placing loudspeakers in a conference room are much less critical than those for microphones. It is generally considered good practice to limit the distance from any listener in the room to the nearest loudspeaker to not more than twice the critical distance.

Loudspeakers should be distributed in the ceiling, on the walls, or on the conference table to ensure a minimum sound pressure level of 65 dBA at listener positions. If there is significant noise, the sound pressure level should be at least 20 dB above the ambient noise level. More ‘‘presence’’ and less ‘‘voice-on-high’’ effect is achieved when the loudspeakers are placed on or in the edge of the conference table.

Ceiling mounted loudspeakers are usually simpler to install and less conspicuous. Generally, loudspeakers installed in a visible grid, suspended, acoustic panel ceiling should be placed approximately 0.6 meters outside the edge of the conference table. Best results are obtained when the loudspeakers are *not* installed symetrically but somewhat randomly. This prevents exciting pronounced room modes of vibration.

Reference

[1] *Teleconference center construction guidelines* , Bell System Technical Reference, PUB 42903, May 1980, American Telephone and Telegraph Co.

Supplement No. 17

DIRECT LOUDNESS BALANCE AGAINST | THE INTERMEDIATE REFERENCE SYSTEM (IRS)

FOR THE SUBJECTIVE DETERMINATION OF LOUDNESS RATINGS

(*Melbourne, 1988*)

(Quoted in Recommendations P.78)

(Contribution from China)

1 Introduction

In the subjective determination of loudness ratings according to Recommendation P.78, the wideband fundamental reference system NOSFER should be always used in addition to the Intermediate Reference System (IRS). The main reason for using the indirect method for the subjective determination of loudness ratings is the difficulty to hold two handsets, one of the IRS and the other of the unknown system, in one hand during balance. Since 1982, the CCITT Laboratory and some other laboratories have tried to use the direct loudness balance method for the subjective determination of loudness ratings using a cut-out handset. Results show that not only can the test be simplified, but also the discrepancies of the test results can be reduced considerably. Typically the standard deviation of the test results is only half of that using the Recommendation P.78 technique. Furthermore, the introduction of NOSFER in the subjective determination of loudness ratings is no longer necessary.

This Supplement describes the essential arrangement used in the direct loudness balance method

2 Method

2.1 *Handset*

The IRS sending handset with its microphone is mounted in a loudness rating guard-ring position (LRGP) support. However, the handle along with the microphone holder of the IRS receiving handset may be cut away, if necessary, to facilitate holding both an unknown handset and the IRS cut-out receiver piece in one hand during the subjective balance for the RLR or OLR.

2.2 *Speech volume*

Experiments show that the average reading of a VU meter connected to the output of the IRS sending system is about —1.7 dB while an operator is speaking into the microphone of the IRS sending handset at the LRGP using the ‘‘standard volume’’ (see Recommendation P.72, Red Book). This value will be different if a different volume meter is used. Experiment results show that it is not necessary to establish the individual relationship between the ‘‘standard volume’’ and the reading of a meter connected to the output of the IRS sending system for each of the operators.

Because the bandwidth of the IRS sending system is limited, the fluctuation of the needle of the meter is larger than in the case of a wideband system while the talker is active. However, it is not difficult for the operator to control his volume within 1 or 2 dB using his own rule of reading.

2.3 *Listening level*

The loss inserted into the overall IRS connection is fixed at 18 dB, because this value is close to the ‘‘X2’’ value (refer to Recommendation P.78) determined by the recent subjective test team of the CCITT Laboratory as well as those of other laboratories.

2.4 *Test arrangements*

The test arrangements for the determination of SLR, RLR, OLR and JLR are shown in Figure 1 to Figure 4.

2.5 *Balance method*

The ‘‘margin’’ method is used. The details are similar to the subjective determination of R25 equivalent, see Recommendation P.72 (Red Book).

In the determination of RLR and OLR, the operator tends automatically to apply more force to the cut handset fitted to his ear because he holds the cut handset by his fingers directly, while at the same time holding the handle part of the ‘‘unknown’’ handset. This is why the test results of RLR and OLR found in some laboratories are about 1 to 2 dB larger (quieter) than those using the Recommendation P.78 method. This effect can be eliminated if the operator is told that his ear must feel the same force whether the earcap of the handset of an ‘‘unknown’’ system or the earcap of the cut handset of the IRS is applied to his ear.

BLANC

Figure 1 Sup.17, p. 35

Figure 2 Sup.17, p. 36

Figure 3 Sup.17, p. 37

Figure 4 Sup.17, p. 38

Supplement No. 18

COMPARISON OF THE READINGS GIVEN ON SPEECH BY METERS CONFORMING TO RECOMMENDATIONS P.52 AND P.56

1 Introduction

This Supplement gives information on the internationally coordinated ‘‘round-robin’’ experiment which compared the readings of standard recordings given by VU [1], ARAEN [2] and peak programme [3] meters conforming to Recommendation P.52 and the speech voltmeter conforming to Recommendation P.56.

2 Composition of the tapes

The recorded material was made in three languages: English, Polish and Singhalese. Male and female talkers were also required in order to give a wide range of frequencies and timbre of voice.

Each talker recorded a list of 5 short sentences with approximately a 5-second gap between each sentence and the next. This was followed by a short passage of prose of aboute 1 minute’s duration. Each pair of talkers then held a conversation in their mother tongue lasting several minutes. To stimulate the conversation, each talker was given a standard set of picture cards normally used in conversation experiments (Supplement No. 2).

The speech levels recorded were not altered because the tapes were intended to test the capability of the various meters and methods used in measuring. In practice this gave a range of about 15 dB.

3 Frequency responses

Two frequency responses were chosen for this experiment: the IRS sending end (Recommendation P.48) and a wideband flat response (±0.5 dB over the frequency range 100 to 6300 Hz).

4 Information about the tapes

The tapes were prepared on a 2-track (2-channel) tape recorder with IEC equalisation at a speed of 7.5 ips.

The format of each track was as follows:

At the beginning of each tape there was a 1 kHz tone that lasted for approximately 12 s (considered as the reference level). This was followed by a period of silence that lasted for approximately 5 s, which was then followed by the first speech condition to be measured. Each speech condition was followed by 5 s of silence and then 4 s of tone to indicate that the next speech condition began after a few more seconds of silence.

The contents of the tapes is shown in full in Table 1.

The output of the replay tape recorder was set in such a way that the level of the 1 kHz tone at the beginning of each track fell in the following range: 0 dBm to —10 dBm across 600 ohms (i.e. —2.2 dBV to —12.2 dBV).

5 Results

When used for measuring continuously spoken speech, the VU meter is specified to be read by taking the average of the peak deflections approximately every 10 seconds after excluding the two or three highest readings. For the ARAEN meter the readings are specified to be interpreted according to the CCITT rule of observing the reading which is exceeded on the average once every three seconds.

There is naturally an error due to the human element in interpreting any single reading.

The results from the ‘‘round-robin’’ experiment are shown in Table 2. No obvious differences were observed between the variables in the experiment, i.e. language, bandwidth, speech material and talkers, and therefore the results presented in Table 1 have been averaged over all variables, including observers. All readings from meters conforming to Recommendation P.52 are compared to the meter conforming to Recommendation P.56.

As can be seen from the results there is a wide variation. For the VU meter the range is some 6 dB. It must be borne in mind that this range is larger if the individual readings of the observers are taken into consideration, especially as some laboratories used more than one observer. In fact, the total range is slightly greater than 8 dB. These findings are consistent with those stated in [4].

In general, there appears to be consistency within a laboratory but inconsistency between laboratories. For example, for the USA there is an average difference between laboratories of nearly 6 dB.

The findings are similar for both the ARAEN meter and peak programme meter (PPM).

The results obtained from meters conforming to Recommendation P.56 showed, in general, a variation of less than 1 dB between all observers.

These results can be compared with older data from British Telecom and show that the ‘‘world’’ average for the VU meter agrees favourably, but for the ARAEN meter there is a difference of some 3 dB. However, the new results from British Telecom are consistent with this older data.

It is obvious that care is needed when comparing results between countries using meters conforming to Recommendation P.52 and the results from this experiment give guidelines to the differences to be expected.

BLANC

H.T. [T1.18]

TABLE 1

Contents of the tapes

{
Tape 1 — Track 1 (wideband)
}

Condition	Talker	Language	Speech material
1 kHz tone			
1	male	english	short sentences
2	male	english	narrative
3	male	english	conversation
4	female	polish	short sentences
5	female	polish	narrative
6	female	polish	conversation
7	female	singhalese	short sentences
8	female	singhalese	narrative
9	female	singhalese	conversation

Tableau 1 [T1.18], p.39

H.T. [T2.18]
TABLE 2
Comparison of readings made on meter conforming to P.52 and P.56
(Readings in dB relative to reading of meter conforming to Rec. P.56)

Country	VU	ARAEN	PPM
USA AT&T Bell Labs.	 +4.5 —1.2		
Sweden Telecom Admin LME	 +0.7 —1.1	 +0.3	
Australia Telecom Australia	 +0.3		
Norway Telecom Admin	 +0.4		
PR of China 1sr Research Inst.	 +2.9		
UK STL British Telecom	 —1.3	 +1.4 +5.0	 +10.3
CCITT	0.0	+1.8	
France CNET —2.1 | ua) A-weighted +1.7 corrected }	 {		
Japan NTT		 +3.0	
Average	+0.7 | ub)	+2.3	+10.3

a) French results were made with ‘‘A-weighting’’ and a correction was made to eliminate this effect.

b) The ‘‘world’’ average used the corrected French result.

Tableau 2 [T2.18], p.40

References

[1] CCITT *Volume Meter Standardised in the United States of America* , *Termec VU Meter* , Supplement No. 11, White Book, Volume V, 1969.

[2] CCITT *ARAEN Volume Meter or Speech Voltmeter* , Supplement No. 10, White Book, Volume V, 1969.

[3] CCITT *Modulation Meter Used by the British Broadcasting Corporation* , Supplement No. 12, White Book, Volume V, 1969.

[4] CCITT *Comparison of the readings given on conversational speech by different types of volume meter* , Supplement No. 14, White Book, Volume V, 1969.

[5] RICHARDS (D. | .): Telecommunication by speech, page 59, *Butterworths,* London, 1973.