5i’

SECTION 7

SUBJECTIVE OPINION TESTS

Recommendation P.80

METHODS FOR SUBJECTIVE DETERMINATION

OF
TRANSMISSION QUALITY

1 Introduction

This Recommendation contains advice to Administrations on conducting subjective tests in their own laboratories. The tests carried out in the CCITT Laboratory by using reference systems are described in Section 3 of this Volume.

In the course of developing items of telephone equipment, it is necessary to conduct various kinds of specialized tests to diagnose faults and shortcomings; such tests dedicated to the study of specific aspect of transmission quality are not discussed here. The present purpose is to indicate methods that have been found suitable for determining how satisfactory given telephone connections may be expected to be if offered as such for use by the public.

The methods indicated here are intended to be generally applicable whatever the form of any degrading factors present. Examples of degrading factors include transmission loss (often frequency dependent), circuit and room noise, sidetone, talker echo, nonlinear distortion of various kinds, propagation time, deleterious affects of voice-operated devices and changes in characteristics of telephone sets, including loudspeaking sets. Combinations of two or more of such factors have to be catered for.

2 Recommended methods

To be applicable for such a wide range of types of degrading factor given in § 1, the assessment method must reproduce as far as possible all the relevant features present when customers converse over telephone connections. Suitable methods are referred to as ‘‘ Conversation Tests ´out by British Telecom are given in Supplement No. 2 at the end of this volume.

If the rather large amount of effort needed is available and the importance of the study warrants, transmission quality can be determined by service observations and recommended ways of performing these, including the questions to be asked when interviewing customers, are given in Recommendation P.82.

A disadvantage of the service observation method for many purposes is that little control is possible over the detailed characteristics of the telephone connections being tested. A method that largely overcomes this disadvantage but retains many of the advantages of service observations is that used by the AT&T Co. and termed SIBYL (refer to Supplement No. 5, Volume V, *Red Book*). According to this method, members of the staff of Bell Laboratories volunteer to allow a small proportion of their ordinary internal calls to be passed through special arrangements which modify the normal quality of transmission according to a test programme. If a particular call has been so treated the volunteer is asked to vote by dialling one of a set of digits to indicate his opinion. In this way all results are recorded by the controlling computer and complete privacy is retained.

This Recommendation was numbered P.74 in the *Red Book* .

3 Supplementary methods

Under certain conditions, it is permissible to dispense with the full conversation method and to use one-way listening-only tests Suitable conditions apply for using a listening test when the degrading factor(s) under study affect the subjects only in their listening role. Attenuation/frequency

distortion and nonlinear distortion caused by quantizing have been studied successfully by listening tests but it would be unwise to study the effects of sidetone, for example, by this method. Listening-only tests may also be misleading when assessing the effects of a factor, like circuit noise, when the magnitude of the degradation caused is substantial. In any case, sufficient comparison with the results from full conversation tests should be made before the results from listening-only tests are accepted as reliable.

Recommendation P.81 (2) the use of a wideband MNRU as the reference system in terms of which subjective performance of wideband digital processes should be expressed

---

This Recommendation was numbered P.70 in the *Red Book* .FE

MODULATED NOISE REFERENCE UNIT (MNRU)

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

The CCITT,

considering

(a) that the use of digital processes (64 kbit/s PCM A-law or μ-law, A/D/A encoder pairs, A/μ-law or μ/A-law converters, digital pads based on 8-bit PCM words, 32 kbit/s ADPCM, etc.) in the international telephone network has grown rapidly over the past several years, and this growth is expected to continue;

(b) that new digital processes are being standardized, e.g. 64 kbit/s 7 kHz wideband ADPCM;

(c) that there is a need for standard tools to measure the quantization distortion performance of digital processes [for example, 32 kbit/s ADPCM (Recommendation G.721) and 64 kbit/s 7 kHz wideband codec (Recommendation G.722)], so that the tools can be used for estimating the subjective transmission performance of international connections containing digital processes;

(d) that an objective speech quality assessment method has not yet been established;

(e) that, at the present time, subjective tests incorporating reference system conditions represent the only suitable method for measuring the speech transmission performance of digital processes;

(f) that expressing results in terms of a common reference system may facilitate comparison of subjective test results obtained at different laboratories,

recommends

(1) the use of a narrow-band Modulated Noise Reference Unit (MNRU) as the reference system in terms of which subjective performance of telephone bandwidth digital processes should be expressed;

This specification is subject to future enhancement and therefore should be regarded as provisional.

Note 1 — The MNRU can be realized using laboratory equipment or by computer simulation. Further information on the MNRU is given in the references listed at the end of this Recommendation.

Note 2 — The listening-only method presently proposed when using the MNRU in subjective tests is described in Supplement No. 14 at the end of this volume. See Recommendation P.80, § 3, for precautions concerning the use of listening-only tests.

Note 3 — Objective measurement methods which suitably reflect subjective quantization distortion performance of various types of digital processes do not exist at present. (For example, the objective technique of Recommendation G.712, based on sine-wave and band-limited noise measurements, are designed for PCM and do not measure appropriately the distortion induced by other systems such as ADPCM.) The artificial voice described in Recommendation P.50 may be relevant. Even if an objective method is developed, subjective tests will be required to establish correlation of subjective results/objective results for particular digital process types.

Note 4 — The wideband MNRU without noise shaping as described in this Recommendation is recommended noise path after the multiplier (see Supplement No. 15), to shape the correlated noise spectrum. Some Administrations suggest the use of such a filter while others do not.

1 Introduction

The MNRU was originally devised to produce distortion subjectively similar to that produced by logarithmically companded PCM systems [1]. This approach was based on the views:

1) that network planning would require extensive subjective tests to enable evaluation of PCM system performance over a range of compandor characteristics, at various signal levels and in combination with various other transmission impairments (e.g. loss, idle circuit noise, etc.) at various levels, and

2) that it would be as reliable and easier to define a reference distortion system, itself providing distortion perceptually similar to that of PCM systems, in terms of which the performance of PCM systems could be expressed. This requires extensive subjective evaluation of the reference system when inserted in one or more simulated telephone connections, but leads to the possibility of simplified subjective evaluation of new digital processing techniques.

Various organizations (Administrations, scientific/industrial organizations), as well as the CCITT itself, have made extensive use of the MNRU concept for evaluating the subjective performance of digital processes (in arriving at Recommendations G.721 and G.722, for example). A modified version for use in evaluating codecs of wider bandwidth (70-7000 Hz) is now common practice. However, the actual devices used, while based on common principles, may have differed in detail, and hence the subjective results obtained may also have differed. (Differences in subjective testing methodology are also relevant.) The purpose of this Recommendation is to define the narrow-band and wideband versions of the MNRU as completely and in as much detail as possible in order to minimize the effects of the device, and of its objective calibration procedures, on subjective-test results.

2 General description

Simplified arrangements of the MNRU are shown in Figure 1a/P.81 for the narrow-band version and Figure 1b/P.81 for the wideband version. Speech signals entering from the left are split between 2 paths, a signal path and a noise path. The signal path provides an undistorted (except for bandpass filtering) speech signal at the output. In the noise path, the speech signal instantaneously controls a multiplier with an applied gaussian noise ‘‘carrier’’ which has a uniform spectrum between 0 Hz and a frequency at least twice the cutoff frequency of the lowpass portion of the bandpass filter. The output of the multiplier consisting of the noise modulated by the speech signal, is then added to the speech signal to produce the distorted signal.

The attenuators and switches in the signal and noise paths allow independent adjustment of the speech and noise signal levels at the output. Typically, the system is so calibrated that the setting of the attenuator (in dB) in the noise path represents the ratio of instantaneous speech power to noise power, when both are measured at the output of the band-pass filter (terminal OT).

For this Recommendation, the decibel representation of the ratio is called Q_N for the narrow-band version and Q_W for the wideband version.

Figure 1a/P.81, p.

Figure 1b/P.81, p.

3 Performance specifications

3.1 *General*

The specifications in this section apply both to hardware implementations and software simulations.

For practical implementations, the actual signal levels and noise levels may be increased or decreased to meet special needs. In such cases, the level requirements detailed below will have to be modified accordingly.

3.2 *Signal path*

The requirements under this heading refer to the MNRU with infinite attenuation in the noise path of Figures 1a/P.81 and 1b/P.81; separate resistive terminations at the terminals T5 and T6 (unlinked) will achieve this.

The frequency response of the signal path (i.e. between terminals IT and OT of Figures 1a/P.81 and 1b/P.81) should be within the limits of Figure 2a/P.81 for the circuit of Figure 1a/P.81 and within the limits of Figure 2b/P.81 for the circuit of Figure 1b/P.81.

The loss between terminals IT and OT for a 0 dBm, 1 kHz input sine wave should be 0 dB. Over the input level range +10 dBm to —50 dBm, the loss should be 0 dB ± 0.1 dB.

Any harmonic component should be at least 50 dB below the fundamental at the system output (terminal OT in Figures 1a/P.81 and 1b/P.81) for any fundamental frequency between 125 Hz and 3000 Hz in a narrow-band system and 100 Hz and 6000 Hz in a wideband system.

The idle noise generated in the signal path must be less than —60 dBm, measured at terminal OT, in order to conform with § 3.4.

It is recommended that the level of speech signals applied to the terminals IT should be less than —10 dBm (mean power while active, i.e. mean active level according to Recommendation P.56) in order to avoid amplifier peak-clippings of the signal, and be greater than —30 dBm to ensure sufficient speech signal-to-noise ratio.

3.3 *Noise path*

The requirements under this heading refer to the MNRU with infinite attenuation inserted into the signal path of Figures 1a/P.81 and 1b/P.81; separate resistive terminations at the terminals T1 and T2 (unlinked) will achieve this.

3.3.1 *Linearity as function of input level*

With a Q_N setting of 0 dB in the circuit of Figure 1a/P.81, or a Q_W setting of 0 dB in the circuit of Figure 1b/P.81, as the case may be, the noise level at the system output (terminal OT) should be numerically equal to the sine wave level at the input terminal (terminal IT). A correspondence within ± 0.5 dB should be obtained for input levels from +5 dBm to —45 dBm, and for input frequencies from 125 Hz to 3000 Hz in a narrow-band system and 100 Hz to 6000 Hz in a wideband system.

3.3.2 *Noise spectrum*

For a narrow-band system, when Q_N is set to 0 dB, input sine waves applied to terminal IT in Figure 1a/P.81 with levels from +5 to —45 dBm and frequencies from 125 Hz to 3000 Hz should result in a flat noise system spectrum density at the output of the multiplication device (terminal T3 of Figure 1a/P.81) within ± 1 dB over the frequency range 75 Hz to 5000 Hz. The spectrum density should be measured with a bandwidth resolution of maximum 50 Hz.

For a wideband system, when Q_W is set to 0 dB, input sine waves applied to terminal IT in Figure 1b/P.81 with levels from +5 to —45 dBm and frequencies from 100 Hz to 6000 Hz should result in a flat noise system spectrum density at the output of the multiplication device (terminal T3 of Figure 1b/P.81) within ± 1 dB over the frequency range 75 Hz to 10 000 Hz. The spectrum density should be measured with a bandwith resolution of maximum 50 Hz.

3.3.3 *Amplitude distribution*

The amplitude distribution of the noise at the system output should be approximately gaussian.

Note — A noise source consisting of a gaussian nose generator followed by a peak clipper with a flat spectrum from near zero to 20 kHz will produce a satisfactory output noise at terminal OT.

Figure 2a/P.81 p.

Figure 2b/P.81, p.

3.3.4 *Noise attenuators*

The loss of the noise attenuator(s) i.e. between terminals T4 and T5 in Figures 1a/P.81 and 1b/P.81, should be within ± 0.1 dB of the nominal setting. The attenuator(s) should at least allow Q_N and Q_W settings in the range —5 dB to 45 dB, i.e. a 50 dB range.

3.4 *Combined path*

The requirements under this heading refer to the MNRU with both speech and noise paths simultaneously in operation.

With Q_N or Q_W (as the case may be) set to zero, and the input terminated by an equivalent resistance, the idle noise generated in the combined path should be less than —60 dBm when measured at the system output (terminal OT).

References

[1] LAW (H. | .), SEYMOUR (R. | .): A reference distortion system using modulated noise, *The Institute of Electrical Engineers* , pp. 484-485, November 1962.

Bibliography

CCITT — Contribution COM XII-No. 63, *Some considerations on specifications for modulated noise reference unit* , NTT, Japan, Study Period 1981-1984.

CCITT — Contribution COM XII-No. R4, pp. 71-79, Study Period 1981-1984.

CCITT — Contribution COM XII-No. 119, *Description and method of use of the modulated noise reference unit (MNRU/MALT)* , France, Study Period 1981-1984.

(Melbourne, 1988)

1 Introduction

1.1 *Purpose*

The purpose of this Recommendation is to describe a subjective listening test method which can be used to compare the performance of Digital Circuit Multiplication Equipment (DCME) and packetized voice systems

Many of the degradations found in DCME or packetized voice systems have not been tested before and their effects on other systems in the network are unknown. Therefore the only definitive method is the conversation test where the effects of non-linearity, delay, echo, etc. and their interactions can be verified.

For DCME systems, degradations can include not only the effects of variable bit-rate coding, DSI gain (channel allocation), clipping, freezeout and noise contrast, but also those due to non-linearities in the speech detection system, such that the system may function differently for different speech input levels or activity factors. For packetized voice systems the subjective effect, for example, of ‘‘lost packets’’ is unknown.

Listening tests play an important preliminary role in the assessment, and can supply useful information serving to narrow the range of conditions needing a complete conversation test. Moreover, listening tests of the effects of the impairments

This Recommendation was numbered P.77 in the *Red Book* .FE

METHOD FOR
EVALUATION OF SERVICE FROM

THE STANDPOINT OF SPEECH TRANSMISSION QUALITY

(Geneva, 1976; amended at Malaga-Torremolinos, 1984)

1 General

The CCITT recommends that Administrations make use of telephone users’ surveys in the manner of Recommendation E.125 [1] as a means of measuring speech transmission quality on international calls.

Such surveys being call-related (in this instance to the last international call made) can be conducted either by the full use of the

Recommendation E.125 questionnaires (where other valuable information is obtained on users’ difficulties, e.g. knowing how to make the call, difficulties in dialling or understanding tones, etc.) or by making use of those questions solely related to transmission quality which appear in Annex A.

Note — The evaluation of the transmission performance may be altered by difficulties in setting-up call. Hence the response to incomplete questionnaires should be considered with some reservation.

2 Conduct of surveys

In order to make valid comparisons between data collected in different countries, Recommendation E.125 should be strictly adhered to. Specifically the preamble to the Recommendation, the notes of intended use of the questionnaires and the precise order and wording of the questions should be rigidly followed. In some cases, however, an exception will be made and Question 10.0 will be replaced by the wording indicated in Annex B (detailed information is given in [3]).

Note — This alternative version has the advantage of simplifying the classification of responses to open end probes by experts, as well as increasing the sensitivity to some types of impairments such as delay. These advantages should be weighed against the additional interview time which may be required.

3 Treatment of results

To provide quantitative information suitable for comparisons, the subjective assessments (e.g. those obtained from Question 9.0 of Annex A) of excellent, good, fair or poor (see Note) should be accorded scores of 4, 3, 2 and 1, respectively and a mean opinion score (MOS) calculated for all associated responses. Similarly for all those experiencing difficulty (under Question 10.0 of Annex A or, alternatively, Question 10.0 of Annex B) a percentage of the total responses should be calculated. These two criteria of MOS and percentage difficulty are now internationally recognized and have been measured under many different laboratory simulated connections and practical situations.

The results can be classified in a number of ways, e.g. in terms of the call-destination countries or by nature/composition of the connection i.e. cable/satellite circuits, presence or otherwise of echo suppressors etc. Typical methods of presentation of the results are shown in [2], in this case for several countries. It should be noted that in all presentations it is essential to show the number of responses.

Note — Among the reasons which lead to the limitation of users’ opinions of transmission quality to four classes, i.e. excellent, good, fair and poor, is the following. The experience gained in human factor investigations has shown that when a question which requires a selection from several produced by DCME, in association with an evaluation of the effects of delay added by the DCME, using the echo tolerance method described in Recommendation G.131, can give a good indication of the overall performance of such systems and allow reasonable comparisons to be made. In addition, the

---

different classifications is posed in aural form, e.g. by face-to-face interview or by telephone as with Recommendation E.125, the respondent is frequently unable to carry a clear mental separation of more than four

categories. As a consequence, he is unable to draw on his short-term memory and judgement ability in a sufficiently precise manner to avoid confusion and gives an unreliable response. This restriction does not apply to other situations where a written presentation of the choices is used, in which case frequently five or more classes may be appropriate and shown to yield reliable responses.

ANNEX A

(to Recommendation P.82)

Extract from the questionnaire annexed to Recommendation E.125

Reproduced below are the questions relating to transmission quality which appear in the questionnaire annexed to Recommendation E.125.

The CCITT recommends that this Annex should be used when customers' general impressions of transmission performance are required.

9.0

Which of these four words comes closest to describing the quality of the connection during conversation?

9.1 — *excellent*

9.2 — *good*

9.3 — *fair*

9.4 — *poor* .bp

10.0 *Did you or the person you were talking to have difficulty in talking or hearing over that connection?*

(If answer is ''yes'') probe for nature of difficulty, but without suggesting possible types of difficulty, and copy down answers verbatim: e.g. ''*Could you describe the difficulty a little more?*''

At end of interview, categorize the answers in terms of the items below:

10.1 — low volume

10.2 — noise or hum

10.3 — distortion

10.4 — variations in level, cutting on and off

10.5 — crosstalk

10.6 — echo

10.7 — complete cut off

10.8 — other (specify)

Note — Responses to Questions 10.1 to 10.8 are only obtained from customers who have expressed difficulty in Question 10.0.

ANNEX B

(to Recommendation P.82)

Alternative version for Question 10.0 of questionnaire

delay evaluation should determine whether or not the use of DCME in a network setting will require additional echo control. This listening test method will not provide results useful for generating network application rules based on factors such as the quantizing distortion unit (qdu). Future improvements of the test will allow such results to be obtained.

____

annexed to Recommendation E.125

Studies at AT&T have shown that the verbatim responses describing impairments (requested after Question 10.0 of Annex A) are often too imprecisely worded to permit accurate classification by interviewers who are not experienced in transmission studies. A typical solution to this problem has been to convene a panel of experts to classify the responses, a method which may become impractical as the size and number of user reaction tests increases. This annex presents an alternative approach developed in 1976 and used widely since then by AT&T to measure customer's perceptions of transmission quality on domestic and international telephone connections. The approach involves a more complicated technique of probing for impairments which simplifies the ultimate task of classifying the responses. The alternative of Question 10.0 is reproduced below.

The CCITT recommends that this annex should be used for diagnostic purposes only.

10.0 *Did you have any difficulty talking or hearing over that connection?*

Do not probe: If the person volunteers an explanation, write it down.

On question 10.1-10.8, attempt to read entire text before respondent replies.

10.1 *Now I'd like to ask some specific questions about the connection.*

If the person has already described difficulty, add:

(*In view of what you've already said, some of these may seem repetitious, but please bear with me*). *First, during your conversation on that call, did you hear your own voice echoing back, or did your own voice sound hollow to you?*

10.1.1 — echo hollow (own voice)

10.1.2 — neither

10.1.3 — don't remember/not sure

10.1.4 — other (specify)

10.2 *Did you hear another telephone conversation on the telephone network at the same time as your own?*

10.2.1 — other conversation

10.2.2 — no

10.2.3 — don't remember/not sure

10.2.4 — other (specify)

10.3 *Now I'd like you to think about the voice of the person you were talking to. Was the volume of the voice low as if the person were faint and far away; did the voice fade in and out; or was the voice interrupted or chopped up at times?*

10.3.1 — low volume

10.3.2 — fading

10.3.3 — chopping

10.3.4 — none

Evaluation of DCME in tandem with other DCME has not been considered at this stage nor have the effects of systems using encoding at different rates. This Recommendation will subsequently be updated when information on these specific points becomes available.

10.3.5 — don't remember/not sure

10.3.6 — other (specify)

10.4 *How did the voice of the person your were talking to sound to you: did it echo or sound hollow and tinny; or did it sound fuzzy or unnatural?*

10.4.1 — echo, hollow

10.4.2 — fuzzy, unnatural

10.4.3 — none

10.4.4 — don't remember/not sure

10.4.5 — other (specify)

10.5 *Now let me describe three kinds of noise. Tell me if you noticed any of these noises during your conversaiton: a rushing or hissing sound; a frying and/or sizzling, crackling sound; or a humming or buzzing sound?*

10.5.1 — rushing, hissing

10.5.2 — frying and/or sizzling, cackling

10.5.3 — humming, buzzing

10.5.4 — none

10.5.5 — don't remember/not sure

10.5.6 — other (specify)

10.6 *Now let me describe three more kind of noise. Tell me if you noticed any of these during your conversation: a clicking sound; a series of musical tones or beeps; or a continuous high-pitched tone?*

10.6.1 — clicking

10.6.2 — tones or beeps

10.6.3 — high-pitched tone

10.6.4 — none

10.6.5 — don't remember/not sure

10.6.6 — other (specify)

10.7 *Did the other person seem slow to respond, as if there were delay or time lag in the conversation?*

10.7.1 — yes

10.7.2 — no

10.7.3 — don't know

10.7.4 — other (specify)

This Recommendation confines itself solely to listening tests; a separate Recommendation P.85, on conversation tests, will be formulated when sufficient information on evaluation techniques is available. Alternatively, this Recommendation may be revised to include conversation test methods.

____

10.8 *Would you please try to remember the background noise in the area around your telephone (e.g. noise from air-conditioning plant unit, road traffic, office equipment or other people talking) when you made the call. Which of the following categories best describes it?*

10.8.1 — very noisy

10.8.2 — noisy

10.8.3 — quiet

10.8.4 — very quiet

10.8.5 — other (specify)

10.9 *Which of the categories listed below best describes the extent to which you heard your own voice through your telephone when you were talking?*

10.9.1 — could not hear it

10.9.2 — could hear it now that you have drawn my attention to it

10.9.3 — did notice it — not loud

10.9.4 — did notice it — loud

10.9.5 — other (specify)

10.10 *Was there anything else about the connection you'd like to mention?*

Yes — What? (Write in)

Coding instructions:

— is there a written comment?

— does the comment apply to this call?

— does it mention an impairment?

— has it been mentioned already?

— other (specify)

Note — The responses to the specific questions are only obtained from customers who have expressed difficulty in Question 10.0. This may prevent the diagnosis of certain impairments (the bias produced is more serious than that mentionned at the end of Annex A).

References

[1] CCITT Recommendation *Inquiries among users of the international telephone service* , Red Book, Vol. II, Rec. E.125, ITU, Geneva, 1985.

[2] CCITT — Question 2/XII, Annex 2, Contribution COM XII-No. 1, Study Period 1977-1980, Geneva, 1977.

1.2 *Definitions*

1.2.1 **digital circuit multiplication equipment (DCME)**

A general class of equipment which permits concentration of a number of 64 kbit/s PCM encoded input speech circuits onto a reduced number of transmission channels.

This equipment allows an increase in the circuit capacity of the system. The capacity of speech and voiceband data can both be increased by the use of DCME.

---

[3] CCITT — Question 2/XII, Annex, Contribution COM XII-No. 171, Study Period 1977-1980, Geneva, August 1979.

Recommendation P.84

SUBJECTIVE LISTENING TEST METHOD FOR EVALUATING

DIGITAL CIRCUIT MULTIPLICATION AND PACKETIZED VOICE | SYSTEMS

The specifications in this Recommendation are subject to future enhancement and therefore should be regarded as provisional.

1.2.2 **digital circuit multiplication system** (DCMS)

A telecommunication system comprised of two or more DCME terminals connected by a digital transmission system providing a pool of bearer channels. The DCMS supports:

i) 64 kbit/s clear channels for ISDN services (can be used in the bearer pool),

ii) voiceband data (dial-up) up to and including 9600 bit/s V.29. Group III facsimile is also included under this heading,

iii) voice services in the frequency range 300-3400 Hz, carried at 56 or 64 kbit/s,

iv) 64 kbit/s clear (not ISDN dial-up),

v) sub-64 kbit/s digital data.

1.2.3 *Circuit versus packet mode*

Internally the DCME may employ a circuit or a packet mode for the transmission of speech or data. In the circuit mode, bearer channels are derived by providing suitable time slots on the transmission facility interconnecting the DCME terminal equipment. In the packet mode virtual bearer channels are created and the speech or data samples are put into one or more packets of fixed or variable length. The packets are addressed to the destination circuit and transmitted in a virtual channel on the transmission facility one at a time. Thus, in the circuit mode the transmission facility can be thought of as carrying a number of bearer channels multiplexed together, while in the packet mode the facility is thought of as a single high speed channel logically divided into virtual channels which transmits packets one at a time.

1.2.4 **single clique working (point-to-point operation)**

The system of two DCMEs interconnected by one set of bearer channels. This working of a DCME is the most efficient mode of operation for a DCMS. It utilizes the maximum bearer pool capacity and the minimum inter-DCME control information. It is an exclusive mode of operation. Another term for point-to-point is circuit-based DCMS. Figure 1/P.84 shows an example of point-to-point or circuit-based DCMS.

Figure 1/P.84, p.

1.2.5 **multi-clique working (point-to-multipoint operation)**

A single DCME working to more than one DCME each on a point-to-point destination basis; designations are split and are therefore not interactive. Multi-clique working reduces the traffic handling capacity compared with

point-to-point
operation, due to a reduction in bearer capacity. Single clique working is the equivalent of point-to-point operation.

1.2.6 **multi-destination operation**

Many DCMEs working over a common bearer capacity pool, enabling interactive working. This is the equivalent of a TDMA satellite system. Traffic handling capacity is drastically reduced since the bearer becomes very small, due to inter-DCME control messages and inter-terminal operation reducing the bearer capacity. Another term for multi-destination DCMS is network-based DCMS. Figure 2/P.84 shows an example of this.

Figure 2/P.84, p.

1.2.7 **low rate encoding (LRE)**

Speech coding methods with bit rates less than 64 kbit/s, e.g. the 32 kbit/s ADPCM transcoder, (Recommendation G.721). This is one technique commonly used in DCME to increase the circuit capacity.

1.2.8 **digital speech interpolation (DSI)**

This is a technique whereby advantage can be taken of the inactive periods during a conversation, creating extra channel capacity. Speech activity is typically 30-40%, on average, which can produce a DSI gain of up to 3 | | , but generally in the range of 2 | | to 2,5 | | .

1.2.9 **LRE gain, DSI gain, DCME gain**

LRE gain is the factor by which the 64 kbit/s rate of the incoming circuits is reduced when LRE is used for coding within the DCME. For example, when a transcoder conforming to Recommendation G.721 is used, the LRE gain will equal 2. The LRE gain is 1 when no transcoding is used.

DSI gain is the ratio of the number of active speech input circuits to the number of bearer channels used to transport this speech, where the same encoding rate is used for circuits and bearer channels. The DSI gain is constrained by the number of input circuits and the speech activity factor and other input speech characteristics. The DSI gain is 1 when DSI is not used.

The DCME gain is the product of the LRE and DSI gain factors.

1.2.10 **DCME overload**

The instant when the number of instantaneously active input circuits exceeds the maximum number of ‘‘normal’’ bearer channels available for DSI.

1.2.11 **freezeout**

The condition when an input circuit becomes active with speech and cannot be immediately assigned to a bearer channel, due to lack of availability of such channels.

1.2.12 **freezeout fraction**

The percentage of speech lost, obtained by averaging over all input circuits for a given time interval, e.g. one minute.

1.2.13 **transmission overload**

The condition when the freezeout fraction goes beyond the value set in accordance with the speech quality requirements.

1.2.14 **clipping**

An impairment occurring in DSI systems employing speech detectors whereby the detector, due to the time it takes to recognize that speech is present, can cut off (‘‘clip’’) the start of the speech utterance. Competitive clipping is the impairment caused by the overload control strategy which allows

freezeout to occur when bearer channels are temporarily unavailable. Another name for the competitive clipping overload control strategy is sample dropping

1.2.15 **variable bit rate (VBR)**

An overload control strategy often used to cope with traffic peaks and hence freezeout problems. Temporary, additional bearer channels (overload channels) are created. Several VBR techniques are available:

i) Graceful overload is one technique to reduce the bit rate. For example, a 4-bit sample 32 kbit/s ADPCM channel can be reduced on demand to a minimum of a 3-bit sample 24 kbit/s, and the VBR will average across the DCMS somewhere between 3 and 4 bits. The dynamic load control (DLC) will operate when the predicted traffic loading rises above a preset VBR.

ii) Permanent 3-bit allocation set on block of channels. These channels operate solely in a 3-bit mode.

The different reduction techniques available have different subjective performances.

1.2.16 **queuing**

An overload control strategy employing buffer memory in the DCME transmitter to store speech samples while waiting for a bearer channel to become available.

1.2.17 **dynamic load control (DLC)**

An overload control strategy in which the DCMS signals to the associated switch that the traffic load the switch is generating, or is predicted to generate, cannot be transmitted satisfactorily by the DCMS and that the switch should reduce its demand on the DCMS by a holding signal sent to the circuits when they become idle.

1.2.18 **load carrying capacity**

The load carrying capacity is defined as the maximum offered speech load plus ‘‘overhead’’ load (see § 1.2.19) that the DCME channels can carry without forced loss of any speech samples. DCME overload is defined to occur when the instantaneously offered load exceeds the carrying capacity of the DCME bearer channels.

1.2.19 **applied and offerd load**

The applied load consists of the speech bursts entering the DCME on the active circuits. Thus, applied load is a function of the number of active circuits and the speech activity on the circuits.

The offered load consists of the applied load plus any additional load (overhead) generated by the DCME messages and control information. The offered load is the load presented to the DCME bearer channels. If the offered load is less than the load-carrying capacity of the channels, then all the offered load is carried by the DCME. However, if the offered load exceeds the capacity of the bearer channels, then, depending upon the overload strategy of the DCME, some of the offered load will be lost through competitive clipping (sample dropping). The DCME may employ variable bit rate coding so that, should the freezeout fraction exceed some preset limit, the DCME can momentarily increase the load-carrying capacity of the bearer channels (creation of overload channels) in order to accommodate the extra load. Dynamic load control may also be used to limit the applied load.

The instantaneous load is a function of the statistics of the input speech and the DCME overhead traffic , and is difficult to characterize mathematically. However, the long-term time average applied load can be calculated as follows:

$$L_a = N \frac{\alpha}{\alpha + \beta} ,$$

where L_a is the average applied load, α is the average speech burst length, β is the average silence length, and N is the number of circuits in use. The term $\alpha/(\alpha + \beta)$ is equal to the average speech activity. The applied load is measured at the input to the DCME on the circuits. Thus, the average load on the DCME can be externally controlled by varying the number of circuits in use, N , or the speech activity factor , $\alpha/(\alpha + \beta)$.

Similarly, average offered load is a useful concept, and it can be calculated from this formula:

$$L_o = N \frac{\alpha (k + 1)}{\alpha + \beta} + G ,$$

where L_o is the average load offered to the bearer channels, the term k is a constant which accounts for the ‘‘stretching’’ effect that the speech detector has on the activity factor, and the term G is a load factor that accounts for the system overhead traffic (e.g. control messages). Thus, the average offered load, L_o, will almost always be larger than the average applied load, L_o.

1.3 *Test philosophy*

In order for a test to satisfactorily evaluate DCME performance the test methodology should meet certain conditions. These are as follows:

i) the method should use principles, procedures, and instrumentation that are acceptable to CCITT;

ii) the method should be adaptable to different languages and should yield results that are comparable to previous test results;

iii) the method should permit DCME performance to be compared subjectively (or objectively) to reference conditions. Examples of suitable reference conditions are hypothetical reference connections (HRCs), white noise and speech correlated noise. The HRCs should model the facilities the DCME is designed to replace, when these facilities are known. The results of the comparisons should permit making ‘‘equivalence statements’’ about the DCME, e.g. a DCME system is subjectively equivalent to x asynchronously tandemed 64 kbit/s PCM systems. Ideally, the method should yield results from which a network application rule can be derived;

iv) the DCME should be tested with a realistic traffic load simulator and circuit-under-test signal conditions applied. Most of the transitory impairments arise when the DCME is operating in the range of applied load which forces the use of DSI. Therefore, to subjectively measure the effects of these impairments it is necessary to vary the applied load on the DCME up to and including the maximum design load. The clipping produced by the speech detector is affected by the type of signal being transmitted on the circuit under test. Therefore, only a realistic speech signal which also contains appropriate additive noise should be used on the circuit under test;

v) in most instances DCME is designed to be used in the network as a replacement for an existing facility. If the DCME introduces more delay than the facility replaced, then this additional delay will reduce the echo tolerance (grade of service) unless it is compensated for by the use of extra echo control measures magnitude of the reduction in the echo tolerance that will occur without extra echo control can be determined and hence a decision taken as to the need for additional echo control measures.

vi) The methodology should, ideally, yield results which can be used to produce new opinion models or modify existing models.

1.4 *Description of DCME*

Annex A contains a detailed description of the characteristics of DCME that can be evaluated with this methodology. This section contains a brief summary of these characteristics.

The test methodology applies to two types of DCME: one type which uses DSI only to obtain a DSI gain and a second type which uses a combination of LRE and DSI to obtain both a LRE gain and a DSI gain. The test methodology accounts for the operation of the speech detector, recognizing that speech clipping is an impairment that may occur even though the DCME is not overloaded.

The test methodology is applicable to DCME employing any one or a combination of three methods of overload control: 1) sample dropping or competitive clipping, 2) variable bit rate, and 3) queuing. The test plan also allows for testing of DCME having DLC capability.

The test methodology recognizes that many of the impairments produced by DCME occur only when a load is applied, and therefore provision is made to apply a controlled load to the DCME under test. The load is varied between zero and 100% of circuit capacity. Use of the packet mode in the DCME converts it into a packetized voice system, and this test methodology is applicable to these systems. At the present time only point-to-point (and possibly point-to-multipoint) DCME are covered by this methodology.

2 Source recordings

2.1 *Apparatus and environment*

The talker should be seated in a quiet room having a volume of between 40 and 120 cubic meters and a reverberation time of less than 500 ms (preferably in the range 200 to 300 ms). The room noise level must be below 30 dBA with no dominant peaks in the spectrum.

Speech should be recorded from an Intermediate Reference System (IRS), as specified in Recommendation P.48, or an equivalent circuit. The IRS is chosen as it is well documented and can be implemented by all laboratories. The IRS should be calibrated according to Recommendation P.64.

The recording equipment should be of high quality and of the type agreed to by the test. The equipment selected should be capable of providing at least a 40 dB signal-to-noise ratio. A suitable system might consist, for example, of a high-quality digital audio tape recording system.

All the source speech material should be recorded so that the active speech level, as measured according to Recommendation P.56, is approximately 23 dB below the peak overload level of the recording system. This will assure that the speech peaks will not overload the recording system.

2.2 *Speech material*

The speech material should consist of a sequence of simple, meaningful, short sentences, chosen at random because easy to understand (from current non-technical literature or newspapers, for example). Very short and very long

sequences
should be avoided, the aim being that each sequence when spoken should have a duration of at least 30 s and the duration of any two sequences should differ by no more than 5 s. Administrations can use one of two approaches:

i) to have as many different sequences as there are conditions (an example of suitable material from which sequences may be constructed is contained in Annex B), or

ii) to have a more limited number, e.g. 10 sequences per talker, where combinations of two sequences can be used (this is shown in detail in Annex C).

Because of the opinion scales to be used the first approach is recommended. Enough sequences should be available to cater for all the test conditions, plus a sufficient number for use in a practice session.

2.3 *Procedure*

At least three sentences should be used for each sequence. A silent period containing only circuit noise of approximately one second should procede the first sequence and the sequence should end with a similar silent period containing only the circuit noise. One of the inter-sentence pauses containing circuit noise should last one to two seconds. Otherwise, the talker should speak so that pauses occur naturally.

To facilitate the processing of the recorded speech through the DCME, i.e. to allow for the starting and stopping of the recorders between sequences and to allow time for adjusting the DCME for the next test condition, sequences should be separated by a 5 seconds gap on the tape. Therefore, the recorded source sequences will have the pattern on the tape shown in Figure 3/P.84.

Figure 3/P.84, p.

Sequences should be played back to listeners beginning with the one second silent period. After the sequence has ended, a 5 s period of complete silence should be provided to permit the listener to vote.

Talkers should pronounce the sequence of sentences fluently but not dramatically and have no speech deficiencies such as ‘‘stutter’’.

At least two male-female pairs of talkers shall be used, and more pairs are desirable if the test-time permits.

The method of presentation of the source sequences will be by randomization of talkers by blocks; as shown in the following example:

Block 1 Block 2 Block 3 Block n Talker 1 2 3 4 3 4 1 2 1 3 2 4 2 3 1 4

where talkers 1 and 2 are male and talkers 3 and 4 are female.

2.4 *Calibration signals and speech levels*

When the recordings have been made, the active speech level of each speech sequence (excluding the preceding and following silent periods) should be measured, preferably according to Recommendation P.56. If necessary, the speech should then be re-recorded onto the *right* channel of a second system with the necessary gain adjustments, so that all the sequences will be brought to the same speech level, namely 23 dB below the peak overload level of the recording system.

Thirty seconds of 1000 Hz tone should be inserted at the re-recording stage, at an r.m.s. level 17 dB above the active speech level, i.e. 6 dB below the peak overload level of the recording system: the peak level of this tone will be 3 dB higher still. This tone can then be used later to adjust the r.m.s. input speech level to be 20 dB below the overload point of the DCME (a peak/r.m.s. of tone of 3 dB with the speech level 17 dB below the r.m.s. tone level will give the 20 dB figure).

The *left* channel of the source recording should contain a 1000 Hz tone at a level 23 dB below the peak overload level and of 0.5 s duration, recorded about 0.5 s before the start and after the end of each sequence. These two signals may be used as checking and control signals in the processing of the source sequences through the DCME under test.

3 Simulating system load

3.1 *Requirements for a generic voice load simulator*

Digital Circuit Multiplication Equipment (DCME), by definition, is used to gain an advantage in the number of circuits multiplexed onto a digital transmission facility. With this advantage, however, comes potential degradation of transmission quality when carried loads exceed that for which the DCME was engineered. Thus, a rigorous performance evaluation of DCME includes studying the behaviour of the DCME under conditions of no load, engineered load, and overload. Because the transmission performance of DCME under load depends critically upon the load characteristics, it is necessary to use *known and controlled* simulated loads in order to properly assess DCME performance. This section describes the generic requirements for a voice load simulator for the purpose of facilitating DCME performance evaluations under conditions that are meaningful. Use of voice load simulators with the generic requirements described here will also enable the comparison of results from different studies of various DCME.

Note 1 — The load simulator specified here is to be used for the performance evaluation of DCME using Digital Speech Interpolation (DSI). This excludes Type A DCME, for which load is not an issue by virtue of the fixed time-slot assignment of the channels.

Note 2 — The load simulator specified here is an ‘‘external’’ simulator that produces simulated speech signals so as to exercise many circuits being multiplexed onto a digital transmission facility. Prototype DCME frequently use ‘‘internal’’ load simulation of ‘‘trunk needs service’’ requests that simulate the output of multiple speech detector circuits and thus compete for transmission capacity, even though no simulated signals are actually transmitted; only the ‘‘live’’ channel under test is actually transmitting. This type of simulator can be very useful in the lab, but is not treated here because certain assumptions would have to be made regarding the performance characteristics of the associated speech detector simulation.

3.1.1 *Parameters*

A generic Voice Load Simulator (VLS) for DCME performance evaluation has the following attributes (the parametric specification of which are detailed later in this section):

- talk-spurt characteristics,
- silence (gap) characteristics,
- background noise-fill for silent periods,
- spectral properties of the simulated speech,
- amplitude characteristics,
- physical interface, including idle-circuit specifications.

The above are a minimum set of parameters that may have to be expanded as required; for example, time variation of the number of simulated calls might have to be studied, at which time a pertinent specification would have to be added. Also, only simulated speech signals are discussed. It may be desirable to add simulated tones, signalling frequencies, and voiceband data of various types at a later date.

3.1.2 *Requirements*

3.1.2.1 *General*

These requirements apply to a generic VLS testing a DCME. Accordingly, the DCME must receive digital signals from the VLS that simulate multiple and independent sources of speech similar to that which is observed in telephone networks. To meet the ‘‘multiple and independent’’ condition, it will be assumed that the VLS output is to several T1 or CEPT interfaces.

Where possible, existing Recommendations have been used in deriving these requirements. The most notable exception are the requirements associated with speech activity and the underlying statistical distributions of talk-spurts and silent periods (gaps). For these, the current technical literature was surveyed; the results of [1] being both recent and based on conversational speech, are used here.

3.1.2.2 *Talk-spurt characteristics*

The probability density function (p.d.f.) of talk-spurt durations is modeled by two weighted geometric p.d.f.'s:

$$f_t(k) = C_1(1-U_1)U_1^{k-1} + C_2(1-U_2)U_2^{k-1}, \quad k = 1, 2, 3, \ldots$$

where

$C_1 = 0.60278$ $U_1 = 0.92446$

$C_2 = 0.39817$ $U_2 = 0.98916.$

Every increment of the variable *k* is equal to 5 ms in time. The cumulative distribution function of talk-spurt durations is shown in Figure 4/P.84. The average talk-spurt duration is $\alpha = 227$ ms.

Figure 4/P.84, p.

3.1.2.3 *Silence (gap) characteristics*

The p.d.f. of silence durations is also modeled by two weighted geometric p.d.f.'s:

$$f_s(k) = D_1(1-W_1)W_1^{k-1} + D_2(1-W_2)W_2^{k-1}, \quad k = 1, 2, 3, \ldots$$

where

$D_1 = 0.76693$ $W_1 = 0.89700$

$D_2 = 0.23307$ $W_2 = 0.99791.$

The cumulative distribution function of silence (gap) durations is shown in Figure 4/P.84.

The average silence duration of β = 596 ms, combined with the 227 ms talk-spurt duration average, yields a long-term speech activity factor of 27.6 percent.

3.1.2.4 *Background noise-fill for silent periods*

Noise should be inserted into the silent periods (gaps) so that the performance of DSI in the presence of noise can be studied. It is desirable to have the noise level adjustable; a default value of —58 dbm0p is provisionally recommended.

3.1.2.5 *Properties of the simulated speech*

The artificial voice signal of Recommendation P.51 shall be used as a basis for simulating the characteristics of human speech. Supplement No. 7 to the Series P Recommendations describes a possible generation process of the artificial voice according to Recommendation P.51. This signal can then be switched on/off according to the talk-spurt and silence duration statistics described in §§ 3.1.2.3 and 3.1.2.4.

3.1.2.6 *Physical interface*

The load simulator should have T1 and/or CEPT outputs which have physical, electrical, coding, frame structure, alignment, and signalling characteristics as per Recommendations G.703, G.704, G.711 and G.732 (2048 kbit/s) or G.733 (1544 kbit/s).

3.2 *Determining load capacity of tested systems*

The average applied load equals the product of the number of circuits in use, N , and the average speech activity. The load capacity of the tested system equals the maximum load that the system is designed to handle, L_{max}. The load capacity can be determined by:

i) obtaining the manufacturer’s specifications,

ii) calculation.

After the load capacity is determined, the partial loads at which the system will be tested can be determined. The partial loads are:

$$L_i = c_i L_{max}$$

where

c_i = 0.0, 0.50, 0.75 and 1.0.

3.3 *Controlling load applied to tested systems*

The load applied to the DCME can be changed by varying N and the activity factor. For these tests the speech activity factor will be assumed constant at 28%. Therefore, to obtain a partial load, L_i, it is necessary to calculate the number of active circuits which come closest to achieving this desired value.

For example, if L_{max} = 48 and if a partial load of L_i = 0.50 L_{max} is desired and the speech activity factor of 28% is assumed, then the number of active circuits, N_{active} , is calculated thus:

$$N_{active} = c_i \frac{L_{max}}{f(\textit{speech activity factor})} = 0.5$$

$$\frac{8}{.28} = 86 \text{ active circuits.}$$

In the test, 86 circuits would carry speech load and the remainder would be idled.

Note — The following items are for future study:

a) Should DCME loads include voiceband data as well as speech? The effect of voiceband data traffic on speech quality is an important issue in the evaluation of DCME performance. Data percentage is defined as follows:

$$P_{\text{data}} = \frac{\textit{umber of input circuits active with data}}{\textit{otal number of active circuits}} \times 100\%$$

b) Some Administrations report that speech activity on their real circuits averages about 36% when using a highly sensitive speech detector having a short hangover time of about 30 ms. Is it desirable to modify the speech load requirements given in § 3.1, and, if so, what values are recommended?

c) Fractional values of speech load are given in § 3.2. Some DCME may operate so as to display significant changes in performance at different fractional load points. Should the fractional load points be changed to accommodate this type of operation, and, if so, what changes are recommended?

4 Processing of the speech

The DCME testing laboratory will take the source recordings, replay them through the circuit under test of the agreed DCME (using the calibration tone to set the agreed input level), operating the DCME at the agreed load, and record the output from the circuit under test in a predetermined arrangement (explained in § 5). The recorded outputs will then be used to perform the listening test. The DCME being tested must be connected to the load simulator and to the recording and playback equipment as shown in Figure 5/P.84. It may be necessary to make provision for special A/D and D/A interfaces to permit the selected load simulator and recording equipment to be connected to the DCME.

All the processed outputs will be on the *left* channel of the recording medium. The corresponding original signal will be simultaneously recorded on the *right* channel. The 1 kHz tone will be available both in its original form (*right* channel) and as processed by passing through the DCME under test (*left* channel).

The 1 kHz tone on the source recording (see § 2) will be used to adjust the r.m.s. input speech level to be 20, 30 or 38 dB below the overload point of the DCME coder.

Figure 5/P.84, p.

5 Test design

Three separate tests are proposed to evaluate different aspects of DCME performance. The first verifies the effect of various loads on the performance. The second verifies the effect of errors in the DCME digital control channel. The

third
test calculates the effect that the DCME delay has on the echo tolerance. This last test will be done using Recommendation G.131 and does not involve subjective testing.

5.1 *Test No. 1: Effect of applied load*

This test may be conducted twice, once to obtain a quality rating and (optionally) a second time to obtain a listening effort rating. The parameters for testing are as follows:

a) *DCME test parameters:*

1. DCMEs under test: *N*

2. DCME loads: four values (0, 0.5, 0.75, 1.0) (see § 3.2)

3. speech activity factor: one value (28%)

4. active circuit speech characteristics: one value (see § 3.1)

(—77 and —45 dBm0p)

6. input speech level to CUT: three values (20, 30 and 38 dB below DCME coder overload)

7. output listening levels: at least three values (preferred and preferred ±10 dB)

8. talkers: four talkers, i.e. 2 male and 2 female.

b) *Reference parameters*

1. original source sequences: one value

2. MNRU: four values (5-35 dB in 10 dB steps)

3. SNR: three values (20, 30 and 40 dB)

4. reference connections (HRCs): approximately four different cases to be decided by test team

5. listening levels: three levels (see above)

6. talkers: four talkers, i.e. 2 male and 2 female.

For the stated set of parameters the number of test condition is:

$$4 \times 2 \times 3 \times 3 \times 4 \times N = 288 \times N \text{ DCME conditions}$$

plus

$$12 \times 3 \times 4 = 144 \text{ reference conditions.}$$

This totals (assuming N = 1 DCME):

$$432 \text{ test conditions} + 36 \text{ practice} = 468 \text{ conditions.}$$

The set of test conditions should be divided into about 13 segments (12 test + 1 practice) of 36 conditions with the conditions within each segment put into a random order. Table 1/P.84 lists the conditions in a basis non-randomized segment.

5. circuit under test (CUT) idle circuit noise (ICN): two values

---

Time permitting, use of a third noise level of —58 dBm0p is suggested. This will permit a better characterization of the effect different noise levels have on the DCME.

The basic balanced segment in Table 1/P.84 will be repeated for each of 4 talkers and 3 listening levels to create 12 test segments: A thru L. A practice segment P will also be created. The test segments A thru L plus P can then be ordered for playback in the listening test according to the procedure described in § 6.

Assuming each condition takes 35 s to present and obtain a vote, total test time is about 4.5 hours.

5.2 *Test No. 2: Effect of digital errors in the DCME control channel*

The preceding test was done assuming that the digital transmission facility is operated error-free. Under real conditions errors will occur and errors in the DCME control channel may cause momentary disruption of the voice circuits. To determine the effect of digital errors on performance, Test No. 1 should be repeated while random errors at a rate of 10^{-3} are injected into the control channel. For this test only one listening level (preferred) is necessary, so the total number of test conditions is $N \times 96$ plus 144 reference conditions. With $N = 1$, the test time is 2.3 hours.

H.T. [T1.84]

TABLE 1/P.84

Basic segment (assumes 1 DCME for testing)

Condition	Load	BCN (dBm0p)	Input | ua) (dB)	Q (dB)	SNR (dB)	HRC
1	0.00	—77	20			
2	0.50	—77	20			
3	0.75	—77	20			
4	1.00	—77	20			
5	0.00	—45	20			
6	0.50	—45	20			
7	0.75	—45	20			
8	1.00	—45	20			
9	0.00	—77	30			
10	0.50	—77	30			
11	0.75	—77	30			
12	1.00	—77	30			
13	0.00	—45	30			
14	0.50	—45	30			
15	0.75	—45	30			
16	1.00	—45	30			
17	0.00	—77	38			
18	0.50	—77	38			
19	0.75	—77	38			
20	1.00	—77	38			
21	0.00	—45	38			
22	0.50	—45	38			
23	0.75	—45	38			
24	1.00	—45	38			
25			20			Original
26			20	5		
27			20	15		
28			20	25		
29			20	35		
30			20		20	
31			20		30	
32			20		40	
33			20			HRC1
34			20			HRC2
35			20			HRC3
36			20			HRC4

ICN idle circuit noise

a) dB below DCME coder overload level.

tableau 1/P.84 [T1.84], p.

5.3 *Test No. 3: Effect of delay*

In this test, using Recommendation G.131, the intent is to calculate the transmission delay through the DCME, then determine if the delay will require the use of additional echo control measures. The answer to this question requires that we define the connections that the DCME will be used to provide, then determine the echo tolerance of these connections assuming that conventional transmission facilities are used in place of the DCME, and then finally determine the reduction in the echo tolerance that will occur by inserting the DCME into the connections. If the reduction in tolerance falls below acceptable limits then additional echo control measures will be required if the DCME is used.

6 Listening test procedure

6.1 *Apparatus, calibration and environment*

The listening room should meet the same conditions as the recording room with the exception that the environmental noise should be set to 45 dBA (Hoth spectrum — Supplement No. 13, at the end of this fascicle.

The IRS receiving end (Recommendation P.48) or equivalent circuit will be used. The IRS should be calibrated according to Recommendation P.64.

The gain of the system should be set in such a way that the 1 kHz tone played back from the recordings produces a sound pressure of 7 dBPa when measured on the IEC 318 artificial ear (Recommendation P.51). Thus the speech level at that point will also be —10 dBPa (84 dB SPL) for undistorted speech which is close to the ‘‘preferred listening level’’.

6.2 *Instructions to subjects*

The instructions are given in Annex D. When the subjects have read these instructions, they should listen to the practice conditions and give their response to each sample. No suggestions should be made to them that the practice conditions exhaust the range of qualities that they can expect to hear. Questions about procedure or about the meaning of the instructions should be answered, but any technical questions must be met with the response, ‘‘We cannot tell you anything about that until the test is finished’’.

6.3 *Opinion scale*

The methods agreed to are both of the single stimulus type based on the mandatory ‘‘quality’’ scale and the optional ‘‘listening effort’’ scale.

6.3.1 *Opinions based on the ‘‘quality’’ scale*

The following five categories should be used for the quality test:

— Excellent

— Good

— Fair

— Poor

— Bad

or equivalent depending on language. (Supplement No. 2, at the end of this fascicle.

6.3.2 *Opinions based on the effort required to understand the meaning of sentences (listening effort scale)*

The following five categories should be used for the optional listening effort test:

— complete relaxation possible, no effort required;

— attention necessary, no appreciable effort required;

— moderate effort required;

— considerable effort required;

— no meaning understood with any feasible effort.

or equivalent according to language. (Supplement No. 2, at the end of this fascicle.)

Note 1 — It is expected that quality and listening effort scales are correlated. Therefore it is not generally required to use both scales. However, if, in a particular case, it is desirable to obtain ratings on both scales, the test should first be performed by using the listening effort scale and then duplicated using the quality scale. This order of presentation is particularly important if the same listeners and the same speech sources are used in both tests.

Note 2 — The rating scales associated with the categories defined in §§ 6.3.1 and 6.3.2 are assumed to be linear interval scales. It is recommended to bring this assumption to the attention of the subjects in the test instructions, either in words or by presenting numbers of numerical scales in

the written instructions. Examples of how this can be done is given in Annex D. Alternatively, the scale can have more than 5 grades (e.g. 7 or 11 grades) with the same five verbal definitions at equal distances. An additional possibility is to define the end points of the scale separately (e.g. Ideal and Unusable). These defined end points then serve as anchoring points but are not supposed to be used for the rating. Examples of such alternative subjective scales are found in Annex E.

6.4 *Sequence of operations*

The 12 test plus 1 practice segments (A-L plus P) should be played back according to the augmented latin-squares:

Quality test Optional listening effort test P CABD . | | P ABDC . | | P DBAC . | | P DCAB . | | P ADCB . | | P BDCA . | | P BCDA . | | P CABD . | |

In these squares, each row is used for each group of listeners, who may listen either simultaneously or separately. The segments are played back in the given order within each row. A pause will naturally occur between one segment and the next, while the right place on the recording medium is being found and possibly the calibration is checked; this pause will also be welcomed by the listeners.

6.5 *Listeners*

The listeners used in the tests should be drawn at random from the population of telephone service customers. About 40 but not less than 30 listeners should be solicited.

6.6 *Data collection*

Subject's responses may be collected by any convenient method: pencil and paper, press-buttons controlling lamps recorded by the operator, or automatic data-logging equipment, for example. But whatever method is used, care must be taken that subjects should not be able to observe other subjects' responses, nor should they be able to see the record of their own previous responses. Apart from the inevitable memory and practice effects, each response should be independent of every other.

7 Statistical analysis and reporting of results

After the test is finished and all subject responses are collected, the experimenter will assign numerical scores to the responses as follows:

Response Score Excellent 5 Good 4 Fair 3 Poor 2

Bad 1 Complete relaxation possible, no effort required 5 Attention necessary, no appreciable effort required 4 Moderate effort required 3 Considerable effort required 2 No meaning understood with any feasible effort 1

The numerical mean score (over subjects) should be calculated for each condition, and these means listed (this is required so that effects due to male and female speech can be seen).

As a further aid to rapid review of results, graphs should be prepared according to the formats shown in Figure 6/P.84.

Note especially that the averaging of male and female results is here proposed purely to reduce the output to manageable proportions, and does not imply that this step would be warranted for the detailed study and interpretation of the results (unless the significance tests justify it).

Calculation of separate standard deviations for each condition is not recommended. Confidence limits should be evaluated and significance tests performed by conventional analysis-of-variance techniques.

Figure 6/P.84, p.

ANNEX A
(to Recommendation P.84)

Description of
digital circuit multiplication equipment

A.1 *Definition of DCME*

Digital circuit multiplication equipment (DCME) is defined in § 1.2.1. A working definition may be: any digital transmission method that derives more voicegrade circuits than is possible using equipment conforming to Recommendation G.711. For our purposes the term circuit may at times refer to a circuit between two switching points (trunk) or between the customers premises and a switching point (loop). At other times it may refer to an end-to-end digital connection. The circuit may also be physical or virtual. The term voicegrade means that the bandwidth of the circuit is nominally 3.1 kHz. We will attempt to avoid confusion by using suitable qualifiers, when necessary, to describe the kind of circuit we mean.

Based on the above definitions we conclude that there are three basic types of DCME. These are:

Type A — Uses only LRE (low rate encoding, < | 4 kbit/s) to obtain a circuit multiplier larger than 1. Some LRE methods (e.g., 32 kbit/s ADPCM) are amenable to the subjective testing methods described in Recommendation P.70; other methods (e.g. 48 kbit/s vocoding) may require new subjective test methods.

Type B — Uses only digital speech interpolation (DSI) to obtain a circuit multiplier larger than 1. DSI is defined in § A.2. By definition the digital coding used in Type B DCME to derive a circuit, operates at 64 kbit/s and conforms to Recommendation G.711. Thus, the coding provides a circuit multiplier of unity. During periods of DCME overload any of several overload strategies may be used to resolve the contention for channels. The three basic overload strategies are defined in § A.5. For example, during momentary periods of overload the channel coding rate may be reduced to increase the channel capacity. However, this recoding action is attributed to the DSI and the circuit multiplier larger than 1 thus obtained is credited to the DSI.

Type C — Combination of Types A and B. This hybrid type employs LRE to obtain a circuit multiplier larger than 1, and then DSI to obtain an additional circuit multiplier larger than 1. For example, if the LRE comforms to Recommendation G.721 32 kbit/s ADPCM, then the coder has a circuit multiplier of $\kappa = 2$. The DSI may increase this multiplier by a further factor of 2 or 3, depending upon the DCME. The total multiplier, 4 to 6, is equal to the product of the LRE and DSI multipliers.

A.2 *Digital speech interpolation (DSI)*

Digital speech interpolation, is defined in § 1.2.8. A working definition of DSI may be: any method for assigning a voicegrade bearer channel on demand for the transmission of speech at the onset of the speech burst (talk-spurt). The bearer channel comes from a pool maintained by the DCME and the speech comes from an active circuit connected to the DCME. When the speech burst stops the channel is either:

i) relinquished and put back into the pool, or

ii) kept assigned to the circuit as long as the pool is not empty and the channel is not needed to service another circuit.

In the above context the term ‘‘bearer channel’’ refers to the transmission paths between the DCME terminals, which are used to carry the traffic on the circuits connected to the DCME. By definition, a bearer channel has the same bandwidth as a circuit, i.e. voicegrade. Bearer channels may be derived using time, space or even frequency or wavelength division multiplexing of the transmission medium used by the DCME. The transmission media may be copper wire, coaxial cable, radio path or fibre.

A.3 *Speech detection*

To perform DSI, the DCME must contain a speech detector The speech detector monitors the circuits and determines when speech is present and when it is not. When speech is declared present the DCME attempts to assign an available bearer channel to the circuit. If no channel is available the DCME then invokes its overload strategy. When the speech burst ends the speech detector may provide some ‘‘hangover’’ to avoid tail-end clipping of the burst. Hangover extends the effective length of the burst.

‘‘Fill-in ´employed to bridge or eliminate the silence gaps less than a certain length between speech bursts. Fill-in does not extend the length of individual bursts the way hangover does, but requires a processing delay equal to the maximum filled-in gap. Both hangover and fill-in increase the activity factor of the speech on the bearer channels.

To avoid front-end clipping of the speech burst, the speech detector sometimes employs delay of a few milliseconds to give it time to decide whether speech is present.

Clipping or mutilation of the speech burst (both front-end and possibly tail-end) may occur because the speech detector makes false or late decisions. The operation of the speech detector and thus the clipping performance of the DCME is a function of many factors characterizing the signal on the circuits, such as the signal level, signal-to-noise ratio, and echo path loss.

A.4 *Definition of load*

The frequency of DCME overloading is a function of the load on the system. The system load consists of the speech bursts generated on the incoming circuits plus DCME generated overhead traffic. Since the speech burst activity on the circuits varies from moment to moment, the load also has short-term variations.

In defining load we must distinguish between the applied load and the offered load. The applied load is the speech bursts entering the DCME on the circuits in use. Thus, applied load is a function of the number of circuits in use and the speech activity on the circuits. The offered load consists of the applied load plus any additional load generated by the DCME. The offered load is the load presented to the DCME channels. It should be evident that the offered load is usually larger than the applied load, because:

i) the speech detector increases the activity factor, since it adds fill-in or hangover to speech bursts;

ii) ‘‘overhead’’ information may have to be transmitted on the channels along with the speech samples.

While the load varies continuously, subject to the statistics of the speech and the circuit activity, if we assume that the number of circuits in use, N , is a constant over some period of time in which we are observing the operation of the DCME, then the average applied and offered loads becomes useful concepts. Formulas for the average loads are defined in § 1.2.19. While these formulas are somewhat simplistic and do not capture the information concerning the variance of the load about the average, they do allow useful insight into the operation of the DCME.

The load carrying capacity of the DCME channels is also an important consideration. The load carrying capacity is defined as the maximum offered speech plus ‘‘overhead’’ load that the DCME channels can carry. If the offered load is less than the load carrying capacity of the channels, then all the offered load is carried by the DCME. However, if the offered load exceeds the capacity of the channels, then depending upon the overload strategy of the DCME, (see § A.5) some of the offered load will be lost through sample dropping , or variable bit rate coding will be used to momentarily increase the load carrying of the channels so that they can accommodate the extra load. Thus, overloading is defined to occur when the offered load exceeds the carrying capacity of the DCME channels.

In a sample dropping system the load capacity is fixed and is simply κM , where M is the number of 64 kbit/s equivalent channels provided and κ is the LRE factor which accounts for the difference in bit rates between the circuits (always 64 kbit/s) and the channels. If 32 kbit/s LRE is used on the channels, for example, then $\kappa = 2$. If LRE is not used then $\kappa = 1$. If variable bit rate (VBR) coding is used then the load capacity of the DCME is not fixed, and overloading may be avoided by temporarily creating extra bearer channels. If the coding rate drops from 32 to 16 kbit/s, for example, then during the period VBR is active $\kappa = 4$.

In these examples the number of channels available to carry speech is assumed to be constant. However, in DCME that carries voiceband data and other tones on the circuits, DSI cannot be used on these signals. The result is that these continuous signals capture channels for full-time use, reducing the pool of channels available for carrying speech.

By using the average load equations and the concept of load capacity, we can illustrate in Figure A-1/P.84 the load curves for a sample dropping type C DCME. The slope of the offered load curves depends upon the speech activity factor. $\alpha/(\alpha + \beta)$, and the speech detector ‘‘stretch’’ factor, k . Load

curves for three different activity factors are shown. If the number of circuits in use, N , is less than $N_{m\backslash di\backslash dn} = \kappa M — G = 43$ then the DSI will never activate, even if the momentary speech activity factor goes to unity on all active circuits. Since the DCME-carried load cannot exceed $\kappa M = 48$, as the average offered load, L_o, gets closer and closer to the maximum capacity, the frequency of overloading (sample dropping) will increase as the moment-to-moment fluctuations in the speech activities push the offered load above the limit.

Figure A-1/P.84, p.

Figure A-2/P.84 illustrates the load curves for a variable bit rate type C system which recodes at 16 kbit/s during overload. In this example, when the offered load exceeds $\kappa M = 48$ the coding rate is dropped from 32 to 16 kbit/s on the bearer channels. The capacity is thus increased to $\kappa M = 96$. The extra capacity absorbs the momentary overload and prevents sample dropping (freezeout) from occurring. If the offered load exceeds 96 then sample dropping will have to occur, because further VBR (e.g. down to 8 kbit/s) is not provided for in this example.

Figure A-2/P.84, p.

Thus, in summary, as long as N $N_{\mathrm{m\backslash di\backslash dn}}$ the DCME will not need to use the DSI function, because all circuits will have access to a bearer channel. Overload will not occur until the offered load exceeds the load carrying capacity. In overload, the DCME will start dropping samples or will queue the samples, in which case κ will not change, or the DCME will decrease the coding rate, in which case κ will increase, thus momentarily increasing the capacity of the DCME.

A.5 *Overload strategies*

When a number of active circuits connected to the DCME exceeds the number of available channels, the DCME will experience momentary overloads; an increase in speech bursts will sometimes require more channels than are available. When this happens the DCME must invoke its ‘‘overload strategy’’. The strategy is designed to deal with the issue of how best to share the channel pool. A number of basic strategies are possible:

Type 1 — Competitive clipping or speech sample dropping . In this strategy, defined in § 1.2.14, samples are dropped from the front end of the speech burst that unsuccessfully bids for a channel. Sample dropping continues until a channel is available or the burst ends normally. Perceptually, the effects of front-end sample dropping and front-end clipping, the latter caused by the speech detector, should be the same, even though they have different causes. Theoretically, however, they are not entirely the same, because front-end clipping is more likely to affect low-level parts of the signal, whereas freezeout affects all levels with equal probability.

Type 2 — Variable bit rate coding . This strategy, defined in § 1.2.15, employs embedded speech coding algorithms or other means to effectively multiply the number of bearer channels momentarily available to the circuits to carry the offered load. Since a lowering of the bit rate will have the effect of increasing the quantization noise produced by the coders, the perceptual effect of variable rate coding will be momentary increases in quantizing noise, i.e. reductions in Q (for a discussion of Q , see Recommendation P.81, § 2).

Type 3 — Queueing . This strategy, defined in § 1.2.16, employs buffers (memories) for the speech burst samples to occupy while waiting for a channel. The perceptual effect of pure queueing, without buffer overflow, is a time shift of the speech bursts. No samples are lost, and there is no increase in noise. The impairment introduced can be called ‘‘ silence duration modulation queued will begin somewhat later in time relative to its predecessor burst than it would have without queueing. Also the succeeding burst may be perceived as beginning somewhat sooner. Since the buffers must, of necessity, be finite this strategy cannot be employed alone, but it must be coupled with either sample dropping or variable rate coding. Thus, a queueing system can have speech mutilation or recoding noise as well as time shifting.

Type 4 — Dynamic load control . An overload control strategy, defined in § 1.2.17, in which the DCME signals to the associated switch that the traffic load which the switch is generating, or is predicted to generate, cannot be transmitted satisfactorily by the DCME, and the switch should reduce its demand on the DCME by a holding signal sent to the circuits when they become idle.

A.6 *Silence reconstruction methods*

Since the DCME does not transmit silences between speech bursts at the receiving end, the silences must be artificially recreated. Several different methods for doing this are possible. The simplest is to insert a white noise at a fixed level in the receiver during silences. Careful selection of the level is necessary to avoid noise contrast, that is, an apparent and annoying contrast between the noise in the silences and the background noise during speech bursts. Other methods are possible which attempt to adapt the noise level automatically to the circuit conditions; these methods require careful filtering and estimation of source noise power.

A.7 *Circuit versus packet mode*

Internally the DCME may employ a circuit or a packet mode for the transmission of speech bursts. In the circuit mode, bearer channels are derived by providing suitable time slots on the transmission facility interconnecting the DCME terminal equipment. In the packet mode, the speech burst samples are put into one or more packets of fixed or variable length. The packets are addressed to the destination circuit and transmitted over the transmission facility one at a time. Thus, in the circuit mode the transmission facility can be thought of as carrying a number of channels multiplexed together, while in the packet mode the facility is thought of as a single high speed channel which transmits packets one at a time.

In the packet mode, performance of the system depends on how the packets are serviced. Two servicing methods are:

a) All packets from all circuits enter a first-in first-out (FIFO) queue and are serviced by the high speed channel one at a time. Each packet is treated independently. Each packet experiences a variable delay in arriving at the receiving end that is a function of the fill of the FIFO queue. If packets arrive too late, after a given reconstruction delay, they will be lost or discarded by the receiver. This is called packet dropping and it is a function of the system load. Packet dropping can cause speech mutilation at any point in the burst. It gives rise to ‘‘mid-burst’’ sample dropping. Packets can also be dropped in the FIFO queue if it experiences overflow. The fill of the queue is monitored and the overload strategy is invoked when necessary to prevent excessive packet dropping.

b) Once a circuit has seized the high speed channel for transmission of a packet all the packets on the circuit for that burst are transmitted before the high speed line is free to transmit another circuit’s packets. Thus the circuit is ‘‘cut-through’’ during the burst. Cut-through operation avoids mid-burst speech sample loss. However, since only one

circuit at a time can use the high speed channel, other circuits with packets to transmit must await their turn. The packets must be queued while they await the channel. Load-dependent queueing delays must be equalized at the receiving end. This is usually done by employing some form of time stamp on the packet. The possibility always exists that packet queues will overflow before the packets can be transmitted. When this happens the overload strategy is invoked to prevent excessive packet dropping.

Packet mode introduces more delay than a non-packet mode DCME. The extra delay has three components. The first is the packetization time rate. The second is the reconstruction delay loss. The third is packet queueing delay

In summary, use of packet mode rather than circuit mode may introduce these additional performance-affecting aspects:

i) mid-burst sample dropping,

ii) additional delay equal to the sum of the packetization and reconstruction delays,

iii) packet queueing delay.

A.8 *Packet reconstruction*

In a packet mode, system loss of a packet presents the receiver with a dilemma, namely, what to use in place of the speech samples carried in the lost packet. Several methods are employed and they have different performance consequences. One method is to insert noise samples in place of the lost speech samples. Another method repeats samples in a previous packet to replace the lost samples. Other methods are also employed.

A.9 *Circuit versus network systems*

With the above definitions in mind there appears to be yet another way to classify DCME. We can talk about DCME using non-switched channels and DCME using switched channels. The first type, non-switched channels, is called a circuit-based DCME. The second type, using switched channels, is called a connection-based DCME.

A circuit-based system would be used to provide circuits, either trunks or loops. All switching is done outside the DCME. The connection-based system incorporates circuit- or packet-switching and thus is more properly thought of as a network solution rather than a circuit solution.

The testing of a connection-based DCME is likely to be more complicated than is the testing of a circuit-based DCME. One reason is that the size of a connection-based system may make it difficult to test in a laboratory. Another reason is that loading such a system with a controlled load is difficult.

ANNEX B
(to Recommendation P.84)

Speech material used to construct speech sequences

(The following narratives are examples used by Bell |
Communications Research)

ORWELL

George Orwell began his classic novel 1984 with, ‘‘It was a bright cold day in April,’’ but he gave no further hint as to what the weather might be during the fateful year. From the succession of untoward weather events that marked 1983, many have come to believe that the world’s weather has undergone an unprecedented change for the worse and that we might be headed for a series of natural disasters this year to match the demise of free democratic thought and speech described in Orwell’s book.

Since we do not have the ability to predict what individual weather events might occur during 1984, let us turn the calendar back a hundred years and see what happened throughout the country in 1884. The year opened with the arrival of arctic air from northern Canada which drove the thermometer down to —40° | at Rockford, Illinois, and to —25° | at Indianapolis, Indiana, both records that still stand. Sub-zero temperatures penetrated into the South, and a hard freeze hit citrus groves in Florida.

In early February, heavy rains falling on a deep snow cover caused the Ohio River to flood. Crests were of record height from Cincinnati to the river’s mouth at Cairo, Illinois.

Late February brought an outbreak of tornados in the South and the Ohio Valley, where some sixty individual funnels descended. More than 420 were killed, and more than 1000 injured. Nothing approached this visitation in severity or extent until the tornado outbreak in April in Durango, Colorado, for seventy-six days ending April 16.

In May, out-of-season rainstorms in the deserts of the Southwest caused widespread floods. Rail traffic from Salt Lake City to the south was interrupted for three weeks, and the Rio Grande River flooding at El Paso, Texas, caused $1 million in damage.

Heavy frosts occurred in late May, when the thermometer dropped to 22° | in Massachusetts, and snow fell in Vermont on Memorial Day.

California got more heavy rain in June; Los Angeles had 1.39 inches and San Francisco 2.57 inches, both all-time June records. And as a result of rain in Wisconsin the flooding Chippewa River did more than $1.5 million in damages and left 2,000 homeless at Eau Claire.

The great Oregon snow blockade followed 34 inches of snowfall at Portland in the middle of December. Rail communication was cut off from the east and south for many days, and mail from California had to come by ocean steamer.

If you think the weather that made so many headlines in 1983 was unprecedented, hark back to 1884. We do not know whether El Nino was active then or whether some other atmospheric or oceanic force was the culprit. All we can do now is wait and see what 1984 brings.

FOG

One of winter’s most spectacular sights is a smokelike fog that rises from openings in the arctic ice fields and occasionally appears above the open waters of unfrozen lakes and harbors in our temperate zone. Various names for the phenomenon are ‘‘frost smoke’’, ‘‘sea smoke’’, ‘‘steam fog’’, ‘‘warm water fog’’, and ‘‘water smoke’’. The fog is caused by the passage of a stream of arctic or polar air with a temperature near zero Fahrenheit over unfrozen water. Within the lower forty-eight states, it occurs principally over unfrozen areas of the Great Lakes and over harbor waters of the north Atlantic coast.

‘‘Sea smoke’’ occurs because the vapor pressure at the surface of the water is greater than that in the air above. Water vapor evaporates into the air faster than the air can accommodate it. The excess moisture condenses and forms a layer of fog, like steam or smoke rising off the water. Usually a clear space exists between the water’s surface and the bottom of the fog, and its upper limit is generally 10 to 25 feet. If an atmospheric inversion develops near the water’s surface, the fog may be confined there and becomes thick, resulting in a hazard to navigation.

If the air temperature is severely cold, —20° | or below, the rising moisture may form ice crystals in the layer of air just above the water. This is called ‘‘frost smoke’’, and it makes a beautiful sight, especially when sunlight glitters on the thin ice needles.

‘‘Steam fog’’ can occur over lakes and streams in the autumn following a clear, still night during which the air has cooled. The differences in vapor pressures cause the warm water to steam into the cold air, and whole valleys and basins can be covered with a thin layer of fog while the hillside remains clear.

ANNEX C
(to Recommendation P.84)

Instructions on the use of a limited number of sentences

(Contribution by the Swedish Telecommunication Administration)

If N sentences per talker are used there will be $N(N-1)$ possible sentence combinations per talker. The first 16 results are tabulated below:

N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$N(N-1)$	2	6	12	20	30	42	56	72	90	110	132	156	182	210	240	272

Either of two reasons for wanting to limit the number of sentences can be put forth:

— the wish to save time by not having to author lists of more than 2×85 sentence combinations per talker. Separate recording of all the combinations is of course still needed unless sophisticated editing equipment for digital types is at hand, or

— the need to organize the test in a way that fulfills the requirements for an analysis of variance.

Depending on which of the motives above is invoked, different methods can be adopted. These are:

1) All possible $N(N-1)$ sentence combinations per talker are recorded.

a) The same N sentences are used for all 4 talkers. The same sentence pair should then not be used for the same test conditions from talker to talker, in order to avoid possible systematic interaction between test conditions and phonetic content, or

b) Four different sets of N sentences (N 1, N 2, N 3 and N 4) are authored. Then no precautions corresponding to a) are needed. However, interaction will still be possible and uncontrolled.

2) To allow for an analysis of variance, subjects must judge the same speech material for all test conditions and all talkers. The number of sentences will then be limited to $M \times 2$ where M is the number of pairs that will be used in the test. If $M = 1$ the test may appear too tedious for the subjects and the phonetic coverage may be insufficient. If an analysis of variance is to be justified, and the test is still to be practically possible, an expansion of the number of presentations is therefore recommended. $M = 2$ or 3 should be enough. This will lengthen the test time for each subject, but experience shows that tests of 2.5 hours per subject are quite possible. Adjustments for such an expansion must then be made when deciding the presentation order.

ANNEX D
(to Recommendation P.84)

Instructions to subjects

D.1 *Quality scale — DCME test*

In this test we are evaluating systems that might be used for telecommunications service between separate places.

You are going to hear a number of samples of speech reproduced in the earpiece of the handset. Each sample will consist of a 30 to 35 seconds long sequence of three or more sentences.

Please listen to the complete sequence, then indicate your opinion of the overall sound quality. If you hear any noises or other interference in the pauses before, between or following the sentences you should include the effect of this interference in your judgement of the overall quality.

For indicating your opinion you are requested to use the following 5-point rating scale:

Score Quality opinion 5 Excellent 4 Good 3 Fair 2 Poor 1 Bad or Unsatisfactory

After listening to a sample sequence, either (1) please write down on your response sheet a score, or (2) please press the appropriate button which on this rating scale represents your opinion of the sound quality of the sample just heard.

After you have given your opinion there will be a short pause before the next sample begins.

For practice, you will first hear ‘‘*n* ’’ samples and give an opinion on each; then there will be a break to make sure that everything is clear.

From then on you will have a break after every ‘‘*k* ’’ samples. There will be a total of ‘‘*t* ’’ samples in the test. The test will last a total of about ‘‘*time* ’’ hours.

D.2 *Listening effort scale — DCME test*

In this test we are evaluating systems that might be used for telecommunications service between separate places.

You are going to hear a number of samples of speech reproduced in the earpiece of the handset. Each sample will consist of a 30 to 35 seconds long sequence of three or more sentences.

Please listen to the complete sequence, then indicate your opinion of the effort required to understand the meaning of the sentences.

For indicating your opinion you are requested to use the following 5-point rating scale:

Score Listening effort opinion 5 Complete relaxation possible, no effort required 4 Attention necessary, no appreciable effort required 3 Moderate effort required 2 Considerable effort required 1 No meaning understood with any feasible effort

After listening to a sample sequence, either (1) please write down on your response sheet a score, or (2) please press the appropriate button which on this rating scale represents your opinion of the effort required to understand the meaning of the sample just heard.

After you have given your opinion there will be a short pause before the next sample begins.

For practice, you will first hear ‘‘*n* ’’ samples and give an opinion on each; then there will be a break to make sure that everything is clear.

From then on you will have a break after every ‘‘*k* ’’ samples. There will be a total of ‘‘*t* ’’ samples in the test. The test will last a total of about ‘‘*time* ’’ hours.

ANNEX E
(to Recommendation P.84)

Examples of other subjective scales

E.1 *Eleven-grade quality scale*

10 9 8 7 6 Excellent Good The number 10 denotes a reproduction that is perfectly faithful to the ideal. No improvement is possible.

5 4 3 2 1 0 Fair Poor Bad The number 0 denotes a reproduction that has no similarity to the ideal. A worse reproduction cannot be imagined. (See IEC Report 268-13, Annex A.)

E.2 *Seven point quality scale*

Score Quality description 6 Ideal circuit 5 Excellent circuit. Possible to relax completely during call, very agreeable 4 Good circuit. Necessary to pay attention, but not necessary to make a special effort. Agreeable circuit

3 Fair circuit. A moderate, but not too great, effort is necessary. Not a very agreeable circuit 2 Poor circuit. Listening is possible, but somewhat difficult. Listening disagreeable 1 Bad circuit. Can be used only with great difficulty. Listening very disagreeable 0 Very bad circuit. Practically unusable (See CCIR Report 751, Volume VIII.3, 1986.)

E.3 *Five-grade impairment scale*

5 Imperceptible.

4 Perceptible, but not annoying.

3 Slightly annoying.

2 Annoying.

1 Very annoying.

(See Supplement No. 14, Annex B.)

Reference

[1] LEE and UN: A study of ON-OFF characteristics of conversational speech, *IEEE Trans. Comm.* , Vol. COM-34, No. 6, June 1986.

Blanc

MONTAGE: PAGE 234 = BLANCHE