INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

P.84

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(03/93)

TELEPHONE TRANSMISSION QUALITY

SUBJECTIVE OPINION TESTS

SUBJECTIVE LISTENING TEST METHOD FOR EVALUATING DIGITAL CIRCUIT MULTIPLICATION AND PACKETIZED VOICE SYSTEMS

ITU-T Recommendation P.84

(Previously "CCITT Recommendation")

FOREWORD

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

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

ITU-T Recommendation P.84 was revised by the ITU-T Study Group XII (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

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 Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

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

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

© ITU 1994

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

CONTENTS

Page

1 Introduction 1
1.1 Purpose 1
1.2 Test philosophy 1
2 Source recordings 2
2.1 Apparatus and environment 2
2.2 Speech material 2
2.3 Procedure 3
2.4 Calibration signals and speech levels 4
3 Simulating system load 4
3.1 Requirements for a generic voice load simulator 4
3.2 Determining load capacity of tested systems 5
3.3 Controlling load applied to tested systems 5
4 Processing of the speech 6
5 Test design 6
5.1 Test No. 1 – Effect of applied load 8
5.2 Test No. 2 – Effect of digital errors in the DCME control channel 9
6 Listening test procedure 9
7 Analysis of results 10
Annex A – Description of Digital Circuit Multiplication Equipment 10
A.1 Definitions 10
A.2 Digital Speech Interpolation (DSI) 13
A.3 Speech Detection 13
A.4 DCME load 13
A.5 Overload strategies 14
A.6 Silence reconstruction methods 16
A.7 Circuit versus packet mode 16
A.8 Packet reconstruction 17
Annex B – Speech material used to construct speech sequences 17
Annex C – Instructions on the use of a limited number of sentences 19

SUBJECTIVE LISTENING TEST METHOD FOR EVALUATING DIGITAL CIRCUIT MULTIPLICATION AND PACKETIZED VOICE SYSTEMS[1)]

(Melbourne, 1988, amended at Helsinki, 1993)

1 Introduction

1.1 Purpose

The purpose of this Recommendation is to describe a subjective listening test method which can be used to evaluate the speech quality of Digital Circuit Multiplication Equipment (DCME) or packetized voice systems. This Recommendation is intended for use with DCME systems, such as those described in Recommendation G.763, which use digital speech interpolation (DSI) techniques. For subjective test evaluations of DCME systems not using digital speech interpolation techniques, described in Recommendation G.726 [40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)], use Recommendation P.83.

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 service to narrow the range of conditions needing a complete conversation test. This listening test method will not provide results for generating network application rules based on factors analogous to the quantizing distortion unit (qdu). Future improvements of the test will allow such results to be obtained.

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 of these specific points becomes available.

This Recommendation confines itself solely to listening tests; a separate recommendation, 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.

1.2 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 those from other tests performed using this Recommendation;

---

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

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

2 Source recordings

2.1 Apparatus and environment

See B.1.1/P.80 through B.1.3/P.80.

2.2 Speech material

The format of the speech material must be suitable for the opinion scale being used in the test. This will normally be the listening-quality scale, but optionally the listening-effort scale will be used (see clause 5).

When the listening-quality scale is used, the following requirements apply:

i) The speech material should consist of short passages (called segments), chosen at random (from current non-technical literature or newspapers for example), easy to understand, and more or less self-contained in meaning.

ii) Each segment when spoken naturally should have a duration not less than 9 seconds and not more than 11 seconds.

iii) Each segment should consist of at least three "sentences" in a broad sense, that is, parts that can naturally be separated by pauses in speaking, but connected in meaning to what precedes and follows within the segment.

iv) Within each segment there should be at least one pause which naturally, in view of the meaning and structure of the text, would last for 1 to 2 seconds. The other pauses must be of natural length, since unnaturally long or short pauses may well be interpreted as reducing the quality of the speech. The simplest way of ensuring this is to construct the script of each text with either a special mark or the beginning of a new paragraph at the point where the 1 to 2 second pause is desired. The talkers recording the segments can then be instructed to read consecutively, making sure that they pause for a second or two at the marked point, and pausing naturally at other points.

When the listening-effort scale is used, the following requirements apply:

v) The speech material should consist of single meaningful sentences, easy to understand, chosen at random (from current non-technical literature or newspapers for example) and assembled into groups (called segments). The number of sentences in every segment should be the same (three is the recommended number).

vi) There should be no obvious connection of meaning between one sentence and another in the same segment. This precaution is to reduce the context information within the segment to a minimum, so as not to inflate the opinion scores artificially.

vii) The pauses between the sentences must be at least one second in length. This ensures that the listeners perceive the sentences as isolated from each other in meaning, and puts the DCME system to the test in respect of the unbridged gaps in speech. Moreover, each segment should have a duration not less than 9 seconds and not more than 11 seconds including the pauses.

This structure can be attained by either:

– giving timing cues to talkers at the recording stage [see Section 2.5.8.1 d) of the *Handbook on Telephonometry*]; or

– editing the recordings afterwards.

Either of two approaches may be used for the listening-quality scale:

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

ii) to have a more limited number, e.g. 10 segments per talker, where combinations of two segments can be used (this is shown in detail in Annex C). In this case, additional precautions would have to be taken in the analysis of variance of results from the tests.

The first approach is essential when the listening-effort scale is used, because listening-effort scores are affected when the listener has heard the same sentences before. Enough segments must be available to cater for all the test conditions, plus a sufficient number for use in a practice session.

2.3 Procedure

A silent period containing only circuit noise of approximately one second should proceed each segment and the segment should end with a similar silent period containing only the circuit noise.

To facilitate the processing of the recorded speech through the DCME, i.e. to allow for the starting and stopping of the recorders between segments and to allow time for adjusting the DCME for the next test condition, segments should be separated by a 5 second gap on the tape. Therefore, the recorded source segments will have the pattern on the tape shown in Figure 1 (this is an example for the listening quality scale; for the listening effort scale all pauses must be at least 1 second). Note that if the speech is stored digitally (e.g. on a disk based system) the segments will not require these 5 second gaps.

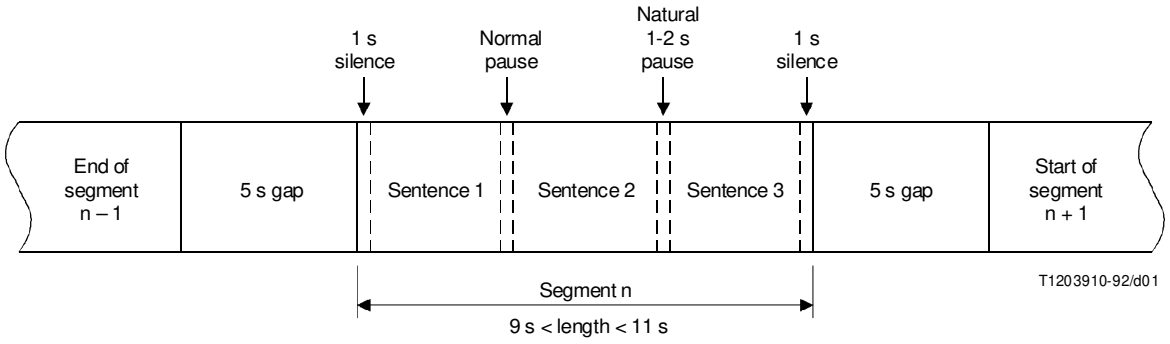


FIGURE 1/P.84

Example of a speech segment in the format required for the listening quality scale, containing three sentences

The recording procedure detailed in Section 2.5.8.1 (Listening Tests) of the *Handbook on Telephonometry* should be followed. Only the part dealing with recording through an IRS is required for this Recommendation.

NOTE – When this recording is made through a physical IRS, the sidetone path of the IRS should be set to 12 dB STMR. This helps to stabilize the speech level of the talker.

Segments should be played back to listeners complete with the silent period. After the segment has ended, a sufficient length of complete silence should be provided to permit the listener to vote.

Talkers should pronounce the segment of sentences fluently but not dramatically and have no speech deficiencies such as "stutter". (See also B.1.6/P.80.)

2.4 Calibration signals and speech levels

See B.1.7/P.80 and B.1.8/P.80.

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 rigourous 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 know and control simulated loads in order to properly assess DCME performance. This subclause 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.

Either the load simulator or the DCME system itself must be programmed so as to record, for each individual segment played through it, the proportion of time during that segment for which the "live" channel was ON, or equivalent information. This quantity x_i will be called the concomitant variable for segment i [see Section 2.5.2 f) of the *Handbook on Telephonometry*].

NOTES

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 (see A.1), for which load is not an issue by virtue of the fixed time-slot assignment of the channels.

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 internal 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 subclause):

- 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 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.

3.1.2.2 Talk-spurt characteristics

The characteristics of talk-spurts are specified in Recommendation P.59.

3.1.2.3 Silence (gap) characteristics

The characteristics of silences are specified in Recommendation P.59.

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 dBm0 is provisionally recommended.

3.1.2.5 Properties of the simulated speech

The artificial voice signal of Recommendation P.50 shall be used, as a basis for simulating the characteristics of human speech. This signal can then be switched on/off according to the talk-spurt and silence duration statistics described in Recommendation P.59.

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}, where L_{max} equals the product of the maximum number of circuits, N_{max}, possible and the average speech activity. 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, 0.75, 1.0 and 1.2.

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 activity circuits which come closest to achieving this desired value.

For example, if L_{max} = 48 and if a partial load of L_i = 0.75, L_{max} is desired and the speech activity factor of 28% is assumed, then the number of active circuits, N_{active} is calculated thus:

$$N_{active} = c_i \frac{L_{max}}{\text{(speech activity factor)}} = 0.75 \frac{48}{0.28} = 129 \text{ active circuits}$$

In this example, 129 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:

$$P_{data} = \frac{\text{Number of input circuits active with data}}{\text{Total number of active circuits}} \times 100\%$$

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) while operating the DCME at the agreed load, and record the output from the circuit under test in a predetermined arrangement (explained in clause 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 2. 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.

The idle circuit noise (ICN) must be added to the speech prior to processing through the DCME. This noise is electrically summed with the source speech immediately before the A/D converter at the DCME input, as shown in Figure 2. It is essential that the r.m.s. input level of the speech to the DCME is adjusted using the 1 kHz tone on the source recording and not by measuring the level of the speech plus noise at the input of the A/D.

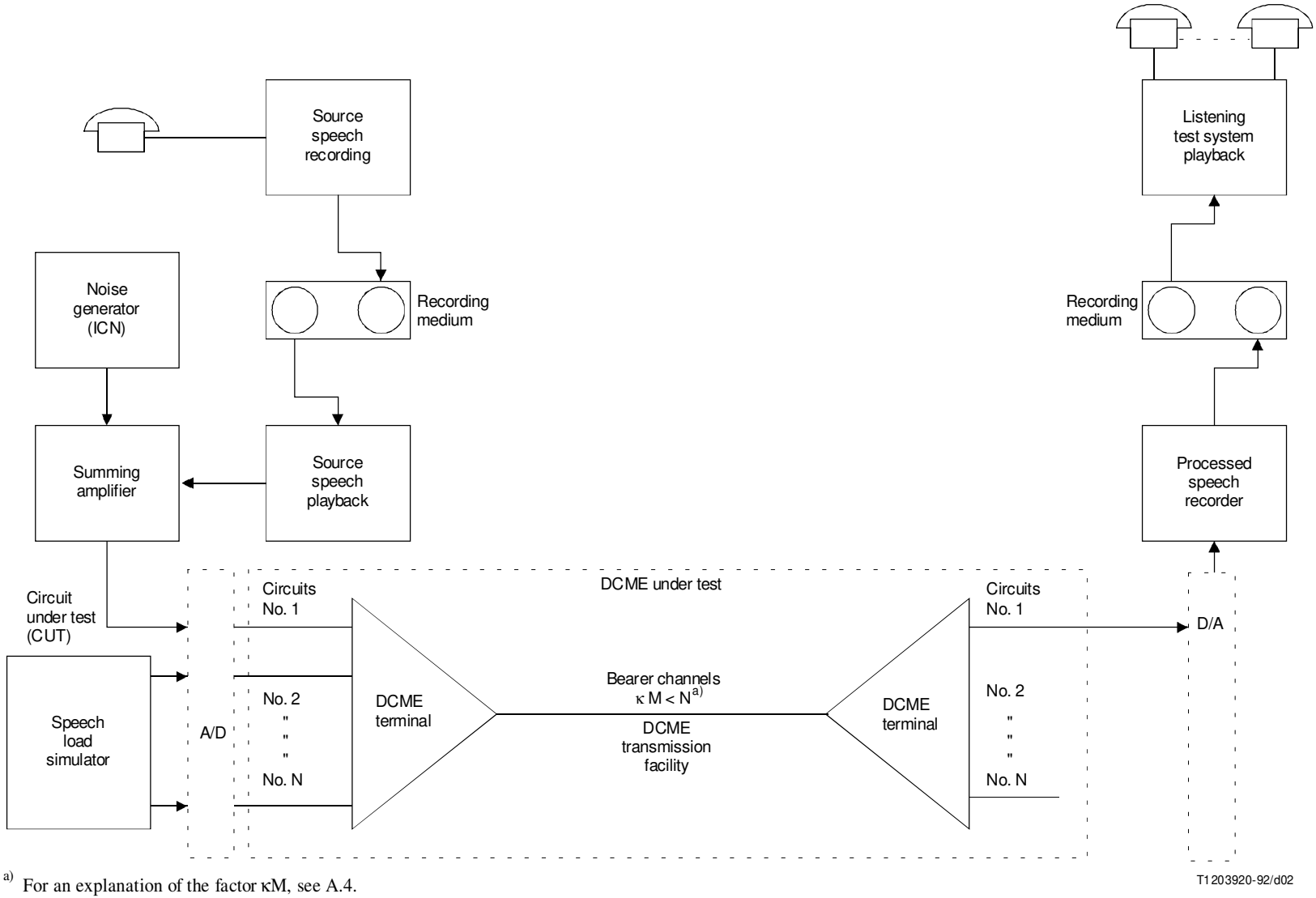
All the processed outputs will be on the left channel of the recording medium. The corresponding original signal will be simultaneously recorded on the right channel. The 1 kHz tone will be available both in its original form (right channel) and as processed by passing through the DCME under test (left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

The 1 kHz tone on the source recording (see 2.4) will be used to adjust the r.m.s. input speech level to be 20, 30 or 38 dB below the peak overload point of the DCME coder.

5 Test design

Two 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.

In both these tests, white noise equivalent to –68 dBmp at the input to the IRS receive end should be added to reduce noise contrast effects at the onset of speech utterances. The IRS must be used with sidetone present (12 dB STMR), as this sounds more natural to the subjects.



a) For an explanation of the factor κM, see A.4.

FIGURE 2/P.84

Testing DCME

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. If performing both listening effort and listening quality tests then either the subjects or speech material must be different. If they are the same, the listening effort test must be performed first as it is essential in listening effort tests to use material which has not been heard before.

The parameters for testing are as follows:

a) *DCME parameters*

1) DCMEs under test: N (at least 1)
2) DCME loads: four values (0, 0.75, 1.0, 1.2 (see 3.2)
3) speech activity factor: one value (28%)
4) active circuit speech characteristic: one value (see 3.1)
5) circuit under test (CUT) idle circuit noise (ICN): two values (–70 and –45 dBm0p)
6) input speech level to CUT: three values (20, 30 and 38 dB below DCME coder peak overload)
7) output listening levels[2)]: 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 segments: one value
2) MNRU: four values (5-35 dB in 10 dB steps)
3) SNR(p): 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:

$4 \times 2 \times 3 \times 3 \times 4 \times N = 288 \times N$ DCME conditions

plus

$12 \times 3 \times 4 = 144$ reference conditions.

This totals (assuming N = 1 DCME):

432 test conditions + 36 practice = 468 conditions.

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

The sequence in Table 1 will be repeated for each of 4 talkers and 3 listening levels to create 12 test sequences: A to L inclusive. A practice sequence P will also be created. The test sequence A to L inclusive plus P can then be ordered for playback in the listening test according to the procedure described in clause 6.

Assuming each condition takes 15 s to present and obtain a vote, total test time is about 2 hours.

---

2) This is a set up using the calibration tone on the source tape, not by measuring the level of the processed speech.

TABLE 1/P.84

Basic sequence (assumes 1 DCME for testing)

Condition	Load	ICN (dBm0p)	Input[a] (dB)	Q (dB)	SNR (dB)	HRC
1	0.00	–70	20			
2	0.75	–70	20			
3	1.00	–70	20			
4	1.20	–70	20			
5	0.00	–45	20			
6	0.75	–45	20			
7	1.00	–45	20			
8	1.20	–45	20			
9	0.00	–70	30			
10	0.75	–70	30			
11	1.00	–70	30			
12	1.20	–70	30			
13	0.00	–45	30			
14	0.75	–45	30			
15	1.00	–45	30			
16	1.20	–45	30			
17	0.00	–70	38			
18	0.75	–70	38			
19	1.00	–70	38			
20	1.20	–70	38			
21	0.00	–45	38			
22	0.75	–45	38			
23	1.00	–45	38			
24	1.20	–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 peak overload level.

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 × 96 plus 48 reference conditions. With N = 1, the test time is 36 minutes.

NOTE – If this test is to use the listening effort scale, either different subjects or different speech material must be used for the reason given in 5.1.

6 Listening test procedure

See B.2/P.80, B.3/P.80 and B.4/P.80.

7 Analysis of results

As well as the opinion score, Y_{lq} or Y_{le}, collected from each subject for each segment replayed, the associated value x (see 3.1) for that segment will be also known and should also be taken into account in the analysis. The usual analysis of variance becomes in fact an analysis of co-variance, with x as the concomitant variable. This increases the precision of the results and conclusions, effectively eliminating the uncertainty due to the random instantaneous fluctuation of system load about its intended mean, by adjusting all mean scores to what they would have been had that mean been maintained exactly.

The same technique should be used in the test with digital errors, by regarding the actual error rate for each segment as the concomitant variable.

Annex A

Description of Digital Circuit Multiplication Equipment

(This annex forms an integral part of this Recommendation)

A detailed description of the characteristics of DCME that can be evaluated with this methodology is given in this annex. Some definitions are given below. For a more comprehensive list of definitions related to DCME, see 2/G.763.

A.1 Definitions

For the purpose of this Recommendation the following definitions apply:

A.1.1 digital circuit multiplication equipment (DCME): Digital Circuit Multiplication Equipment (DCME) is defined in Recommendation G.763 as "A general class of equipment which permits concentration of a number of 64 kbit/s PCM encoded input trunk channels on a reduced number of transmission channels.". 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 customer's 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 discussion, we conclude that there are three basic types of DCME. These are:

- *Type A* – Uses only low rate encoding (LRE, < 64 kbit/s see A.1.4) 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.83: other methods (e.g. 48 kbit/s vocoding) may require new subjective test methods.

- *Type B* – Uses only digital speech interpolation (DSI, see A.1.5) 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 drive 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 conforms to Recommendation G.726 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.

This Recommendation is applicable to Types B & C DCME only.

A.1.2 digital circuit multiplication system (DCMS): A telecommunications system comprised of two or more DCME terminals where each terminal contains a transmit unit and a receive unit 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) including Group III facsimile;

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

iv) 64 kbit/s clear (pre-assigned not ISDN dial-up);

v) sub-64 kbit/s digital data.

A.1.3 circuit versus packet mode: Internally the DCME may employ a circuit or a packet mode for the transmission of speech of 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 a single high speed channel logically divided into virtual channels which transmits packets one at a time.

A.1.4 low rate encoding (LRE): A voice-band signal encoding method, e.g. ADPCM, which results in a bit rate less than 64 kbit/s, e.g. either 40 kbit/s, 32 kbit/s, 24 kbit/s or optionally 16 kbit/s. Conversion between speech signals encoded in PCM at 64 kbit/s and those encoded in ADPCM must be carried out by means of transcoding processes given in Recommendation G.726.

A.1.5 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 3:1, but generally in the range of 2:1 to 2.5:1.

A.1.6 bearer channel (BC): A bearer channel is a unidirectional, digital, transmission path from the transmit unit of one DCME to the receive unit of a second associated DCME used to carry concentrated traffic between two DCMEs.

NOTES

1 A number of bearer channels in each direction of transmission form the both-way link required between two DCMEs. This link may be, for example, a 2048 kbit/s system.

2 A bearer channel may have any of the following instantaneous bit rates: either 64, 32, 24 or optionally 16 kbit/s.

A.1.7 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 the 32 kbit/s portion of Recommendation G.726 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.

A.1.8 DCME overload: The condition when the number of instantaneously active input circuits exceeds the maximum number of bearer channels available for DSI.

A.1.9 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.

A.1.10 freezeout fraction: The percentage of speech lost, obtained by averaging over all input circuits for a given time interval, e.g. one minute. (See 2.23/G.763 and 15.2.3.2/G.763 for a precise definition of freezeout fraction.)

A.1.11 transmission overload: The condition when the average bits per sample goes beyond the value set in accordance with the speech quality requirements.

A.1.12 clipping

i) Speech detector clipping is due to the time it takes to recognize that speech is present, can cut off ("clip") the start of the speech utterance.

ii) Competitive clipping is caused by the overload control strategy which allows freezeout to occur when bearer channels are temporarily unavailable.

iii) Control channel clipping resulting from simultaneous activity detection in more than one input circuit within the same DCME frame or servicing higher order priority messages such as 64 kbit/s unrestricted circuit establishment.

A.1.13 variable bit rate (VBR): The capability of the encoding algorithm to dynamically switch between either 32 and 24 kbit/s or also optionally between 24 and 16 kbit/s for speech traffic under control of the DCME.

This provides an overload control strategy to cope with traffic peaks and hence freezeout problems.

Temporary, additional bearer channels (overload channels) with reduced signal-to-noise ratio are created. Graceful degradation is provided through the use of rapid rotation where the lower encoding rate is cyclically moved through all bearer channels carrying speech and is only applied to any given bearer channel for a short interval of time (typically ≤10 ms for DCME or 16 ms for packet systems). The frequency of occurrence of the lower bit rate is a function of the ensemble DCME input circuit activity.

A.1.14 queueing: An overload control strategy employing buffer memory in the DCME transmitter to store speech samples while waiting for a bearer channel to become available.

A.1.15 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.

A.1.16 load carrying capacity: The load carrying capacity is defined as the maximum offered speech load plus "overhead" load (see A.1.17) that the DCME channels can carry without entering the DCME overload mode (see A.1.8).

A.1.17 applied and offered 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. The DCME may employ variable bit rate coding so that 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. 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 may be lost through competitive clipping (sample dropping).

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 circuit input to the DCME. Thus, the average load on the DCME can be externally controlled by varying the number 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 is for the "stretching" effect that the DCME 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 always be larger than the average applied load, L_a.

A.2 Digital Speech Interpolation (DSI)

Digital speech interpolation is defined in A.1.5. 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 path between the DCME terminals which are used to carry the traffic on the circuits connected to the DCME.

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" is another speech detector function sometimes 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 usually 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 DCME load

The frequency of DCME overloading is a function of the load on the system. The system load consists of 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, both defined in A.1.17.

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 become useful concepts. Formulas for the average loads are defined in § A.1.17. 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 in A.1.16. If the offered load is less then 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) variable bit rate coding will be used to momentarily increase the load carrying of the channels so that they can accommodate the extra load, or some of the offered load will be lost through sample dropping. Thus, overloading is defined to occur when the load exceeds the normal 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 bearer channels. If 32 kbit/s LRE is used on the bearer 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 sample dropping 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 bearer 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 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 $N_{min} = \kappa M\text{-}G = 43$ 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 sample dropping will increase as the moment-to-moment fluctuations in the speech activities push the offered load above the limit.

Figure A.2 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.

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

A.5 Overload strategies

When a number of active circuits connected to the DCME exceeds the number of available normal 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 A.1.12, 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 due to the speech detector is of short duration and is more likely to affect low-level parts of the signal, but front-end clipping due to freezeout is of longer duration. Freezeout, for cases other than front end sample dropping, affects all levels with equal probability.

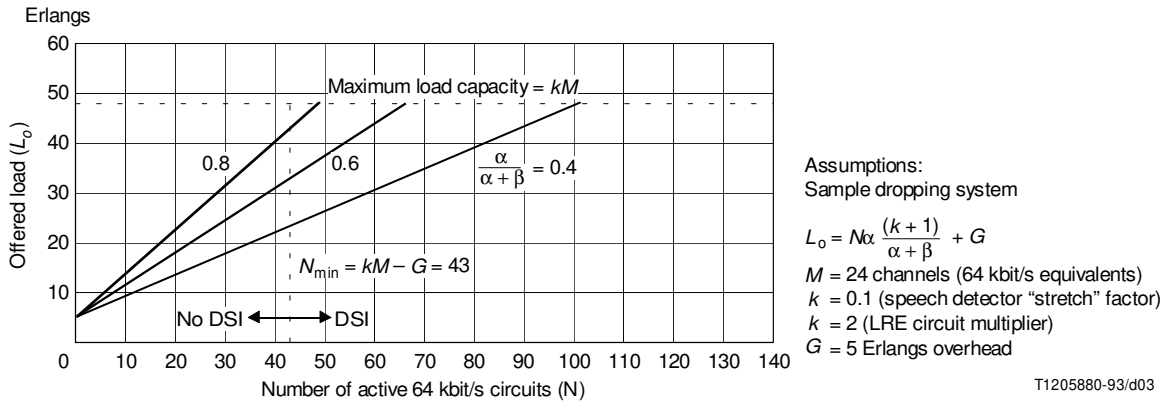


FIGURE A.1/P.84

Example of load curves for type-C, sample dropping, DCME

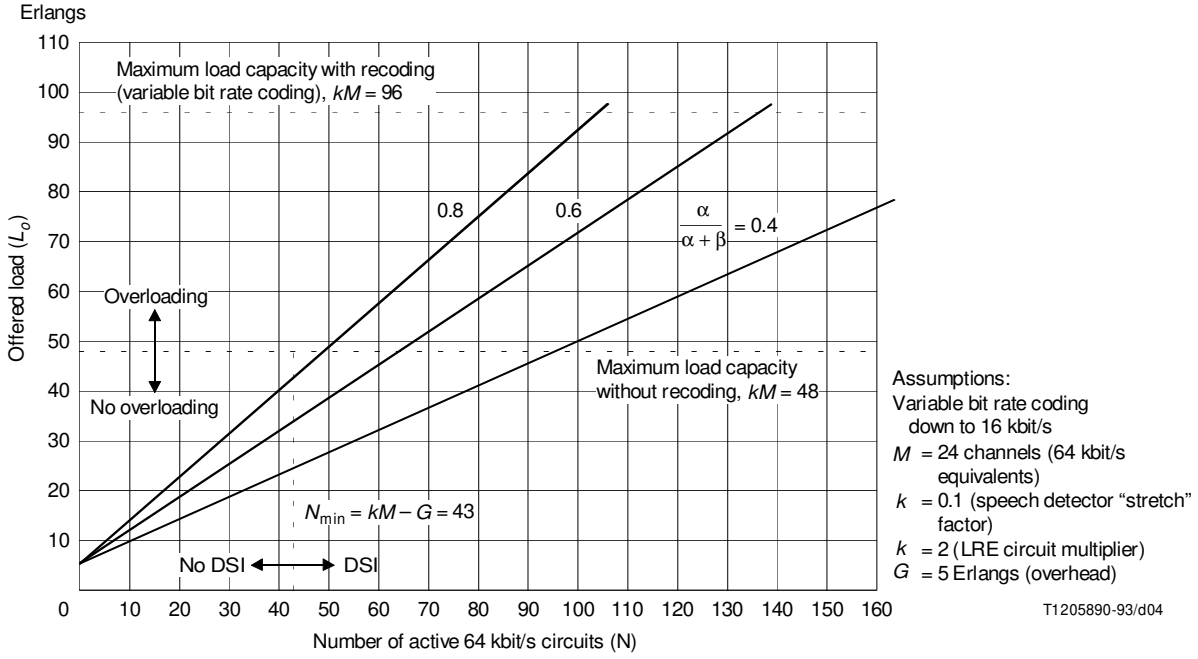


FIGURE A.1/P.84

Example of load curves for type-C, sample dropping, DCME

Type 2 – Variable bit rate coding

This strategy, defined in A.1.13, employs multiple bit rate speech coding algorithms (G.726 or G.727) 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 2/P.81). The subjective effect is minimized by rapidly moving the use of the lower bit rate coding among all active speech bearer channels.

Type 3 – Queueing

This strategy, defined in A.1.14, 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". From the listener's point of view a given speech burst when queued will begin somewhat later in time relative to its predecessor burst than it would have without queueing. Also the successful 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 A.1.15, 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.

A.6 Silence reconstruction methods

Since a DCME with loading sufficient to require interpolation among bearer channels 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 lost. 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. Packetization time is a function of packet length and circuit coding rate. The second is the reconstruction delay. Reconstruction delay is chosen to minimize the probability of packet loss. The third is packet queueing delay. All three components may be load-dependent and thus variable.

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 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.

Annex B

Speech material used to construct speech sequences

(This annex forms an integral part of this Recommendation)

(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 heading 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 °F at Rockford, Illinois, and to –25 °F 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 heights 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 Durango, Colorado, for seventy-six days ending 16 April.

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 °F 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 2000 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 Niño 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 harbours 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 harbour waters of the north Atlantic coast.

"Sea smoke" occurs because the vapour pressure at the surface of the water is greater than that in the air above. Water vapour 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°F 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 vapour 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

Instructions on the use of a limited number of sentences

(This annex forms an integral part of this Recommendation)

(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 fulfils 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 sentence is used for all 4 talkers. The same sentence pair should then not be used for the same test conditions from talkers to talker, in order to avoid possible systematic interaction between test conditions and phonetic content; or

   b) Four different sets of N sentence (N1, N2, N3 and N4) 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 × 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 tats of 2.5 hours per subject are quite possible. Adjustments for such an expansion must then be made when deciding the presentation order.