



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.840

(11/2003)

SERIES P: TELEPHONE TRANSMISSION QUALITY,
TELEPHONE INSTALLATIONS, LOCAL LINE
NETWORKS

Methods for objective and subjective assessment of
quality

**Subjective listening test method for evaluating
circuit multiplication equipment**

ITU-T Recommendation P.840

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30 P.300
Transmission standards	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	Series	P.80 P.800
Audiovisual quality in multimedia services	Series	P.900

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation P.840

Subjective listening test method for evaluating circuit multiplication equipment

Summary

This Recommendation describes a subjective listening test method which can be used to evaluate the speech quality of Circuit Multiplication Equipment (CME). It is intended for use with CME systems, such as those described in ITU-T Recs G.763, G.767, G.768 (DCME), G.765 (PCME) and G.769/Y.1242 (IP-CME) which use digital speech interpolation (DSI) techniques.

In this version, the scope is expanded to more recent speech coders implemented in CME. Updating the Recommendation includes tandemming situations and comfort noise test configuration.

The new Appendix I gives guidance for conversation tests.

Source

ITU-T Recommendation P.840 was approved by ITU-T Study Group 12 (2001-2004) under the ITU-T Recommendation A.8 procedure on 13 November 2003.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. 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 Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

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

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2004

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

Page

1	Scope	1
1.1	Purpose	1
1.2	Test philosophy	1
2	References.....	2
3	Definitions	3
4	Abbreviations.....	3
5	Source recordings	4
5.1	Apparatus and environment.....	4
5.2	Speech material	4
5.3	Procedure.....	5
5.4	Calibration signals and speech levels	6
6	Simulating system load.....	6
6.1	Requirements for a generic voice load simulator	6
6.2	Determining load capacity of tested systems	7
6.3	Controlling load applied to tested systems	8
7	Processing of the speech.....	8
8	Test design.....	9
8.1	Test No. 1 – Effect of applied load.....	9
8.2	Test No. 2 – Effect of digital errors in the DCME control channel	11
9	Listening test procedure.....	12
10	Analysis of results	12
Annex A	– Description of circuit multiplication equipment.....	12
A.1	Definitions	12
A.2	Digital Speech Interpolation (DSI).....	15
A.3	Speech detection.....	16
A.4	CME load.....	16
A.5	Overload strategies	17
A.6	Silence reconstruction methods	19
A.7	Circuit versus packet mode	19
A.8	Packet reconstruction.....	20
Annex B	– Speech material used to construct speech sequences	20
Annex C	– Instructions on the use of a limited number of sentences.....	22
Appendix I	– Guidance for conversation tests	22

ITU-T Recommendation P.840

Subjective listening test method for evaluating circuit multiplication equipment¹

1 Scope

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 Circuit Multiplication Equipment (CME). It is intended for use with CME systems, such as those described in ITU-T Recs G.763, G.764, G.765, G.767, G.768 or as described in G.769/Y.1242, which use digital speech interpolation (DSI) techniques. For subjective test evaluations of CME systems not using digital speech interpolation techniques, channel transcoder systems described in ITU-T Rec. G.761 (*General characteristics of a 60-channel transcoder equipment*) or ITU-T Rec. G.762 (*General characteristics of a 48-channel transcoder equipment*), use ITU-T Rec. P.830.

In addition to the listening tests described here, conversational tests are also recommended in order to thoroughly assess the combined effects of delay, echo, non-linearities, etc. Appendix I gives some guidance for conversation tests.

For CME 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.

It may, therefore, be useful to perform additional tests such as those provided by ITU-T Rec. P.831. 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. Future improvements of the test will allow such results to be obtained.

Evaluation of CME in tandem with other CME is considered in this revised version, but some effects of systems using encoding at different rates need to be completed. This Recommendation will subsequently be updated when information on these specific points becomes available.

1.2 Test philosophy

In order for a test to satisfactorily evaluate CME 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 ITU-T;
- 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;
- iii) the method should permit CME 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 CME is designed to replace, when these facilities are known. The results of the comparisons should permit making "equivalence statements" about the CME,

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

e.g., a CME 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 CME should be tested with a realistic load simulator and circuit-under-test signal conditions applied. Most of the transitory impairments arise when the CME 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 CME 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 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- ITU-T Recommendation G.703 (2001), *Physical/electrical characteristics of hierarchical digital interfaces*.
- ITU-T Recommendation G.704 (1998), *Synchronous frame structures used at 1544, 6312, 2048, 8448 and 44 736 kbit/s hierarchical levels*.
- ITU-T Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.
- ITU-T Recommendation G.726 (1990), *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*.
- ITU-T Recommendation G.728 (1992), *Coding of speech at 16 kbit/s using low-delay code excited linear prediction*.
- ITU-T Recommendation G.729 (1996), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)*.
- ITU-T Recommendation G.732 (1988), *Characteristics of primary PCM multiplex operating at 2048 kbit/s*.
- ITU-T Recommendation G.733 (1988), *Characteristics of primary PCM multiplex operating at 1544 kbit/s*.
- ITU-T Recommendation G.761 (1988), *General characteristics of a 60-channel transcoder equipment*.
- ITU-T Recommendation G.762 (1988), *General characteristics of a 48-channel transcoder equipment*.
- ITU-T Recommendation G.763 (1998), *Digital circuit multiplication equipment using G.726 ADPCM and digital speech interpolation*.
- ITU-T Recommendation G.764 (1990), *Voice packetization – Packetized voice protocols*.
- ITU-T Recommendation G.765 (1992), *Packet circuit multiplication equipment*.

- ITU-T Recommendation G.767 (1998), *Digital circuit multiplication equipment using 16 kbit/s LD-CELP, digital speech interpolation and facsimile demodulation/remodulation.*
- ITU-T Recommendation G.768 (2001), *Digital circuit multiplication equipment using 8 kbit/s CS-ACELP.*
- ITU-T Recommendation G.769/Y.1242 (2002), *Circuit multiplication equipment optimized for IP-based networks.*
- ITU-T Recommendation P.50 (1999), *Artificial voices.*
- ITU-T Recommendation P.59 (1993), *Artificial conversational speech.*
- ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality.*
- ITU-T Recommendation P.810 (1996), *Modulated noise reference unit (MNRU).*
- ITU-T Recommendation P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs.*
- ITU-T Recommendation P.831 (1998), *Subjective performance evaluation of network echo cancellers.*

3 Definitions

This Recommendation defines the following term:

3.1 Q: The ratio, in dB, of speech power to modulated noise power in the Modulated Noise Reference Unit, as described in ITU-T Rec. P.810.

4 Abbreviations

This Recommendation uses the following abbreviations:

A/D	Analogue-to-Digital
ADPCM	Adaptive Differential Pulse Code Modulation
BC	Bearer Channel
CME	Circuit Multiplication Equipment
CMS	Circuit Multiplication System
CUT	Circuit Under Test
D/A	Digital-to-Analogue
DCME	Digital Circuit Multiplication Equipment
DCMS	Digital Circuit Multiplication System
DLC	Dynamic Load Control
DSI	Digital Speech Interpolation
FIFO	First-In First-Out
HRC	Hypothetical Reference Connection
ICN	Ideal Circuit Noise
IP-CME	Circuit Multiplication Equipment optimized for IP-based network
IRS	Intermediate Reference System

ISDN	Integrated Services Digital Network
LRE	Low Rate Encoding
MNRU	Modulated Noise Reference Unit (ITU-T Rec. P.810)
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
SNR	Signal-to-Noise Ratio
STMR	SideTone Masking Rating
VBR	Variable Bit Rate
VLS	Voice Load Simulator

5 Source recordings

5.1 Apparatus and environment

See B.1.1/P.800 through B.1.3/P.800.

5.2 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 8).

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 of 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.