



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.832

(05/2000)

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

Methods for objective and subjective assessment of
quality

Subjective performance evaluation of hands-free terminals

ITU-T Recommendation P.832

(Formerly CCITT Recommendation)

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.

Subjective performance evaluation of hands-free terminals

Summary

This ITU-T Recommendation describes methods and procedures for conducting subjective performance evaluations of hands-free terminals.

The use of hands-free terminals in communication has numerous advantages for the telephone users, especially for all "non-traditional" types of terminals such as car phones, computer/laptop-type terminals and others. Due to the complex acoustical situation a big variety of signal processing which may be non-linear and/or time variant is expected. ITU-T Recommendation P.340 describes measurement techniques for hands-free terminals, ITU-T Recommendation P.581 describes the use of the HATS for the evaluation of terminals, ITU-T Recommendations P.501 and P.502 describe measurement signals and analysis procedures. Using these methods a minimum performance of hands-free terminals should be ensured. However, there is always the possibility that those tests do not address fully the impact of all kinds of signal processing in a hands-free terminal and their impact on speech transmission quality.

Subjective testing is a commonly used method of assessing the performance of terminals, including digital speech codecs, voice-operated signal processing, echo cancellation, noise reduction and other types of signal processing. This ITU-T Recommendation defines methods for the subjective evaluation all kinds of hands-free terminals.

Source

ITU-T Recommendation P.832 was prepared by ITU-T Study Group 12 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

Keywords

Hands-free terminals, speech transmission quality, subjective performance.

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 Conference (WTSC), 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 WTSC 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.

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 2001

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 General.....	1
1.1 Scope.....	1
1.2 References.....	1
1.3 Terms and definitions	2
1.4 Abbreviations.....	2
2 Overview on test procedures.....	3
3 General considerations.....	4
3.1 Hands-free parameters to evaluate.....	4
3.2 General considerations about test equipment and calibration.....	5
3.3 Selection of subjects	5
4 Conversational test procedure.....	6
4.1 Purpose.....	6
4.1.1 Benefits.....	7
4.1.2 Drawbacks	7
4.2 Test parameters	7
4.3 Test set-up.....	8
4.4 Description of test procedure.....	9
4.5 Reference conditions.....	10
5 Double talk test procedure	10
5.1 Purpose.....	10
5.1.1 Benefits.....	10
5.1.2 Drawbacks	10
5.2 Test parameters	10
5.3 Test set-up.....	12
5.4 Description of test procedure.....	13
5.5 Reference conditions.....	13
6 Third-party listening test procedure.....	13
6.1 Purpose.....	13
6.1.1 Benefits.....	14
6.1.2 Drawbacks	14
6.2 Test parameters and scaling.....	15
6.3 Test set-up and recording parameters	16
6.4 Description of test procedure.....	19
6.5 Reference conditions.....	21

	Page
Annex A – Corpus of the source signals.....	21
A.1 Size and parameters of the corpus	21
A.2 Design of each script.....	22

Subjective performance evaluation of hands-free terminals

1 General

1.1 Scope

This ITU-T Recommendation describes procedures to be used to assess the subjective performance of hands-free terminals. The methods defined here may be used to assess the extent to which a hands-free terminal operates effectively for speech. This ITU-T Recommendation does not define specific values for hands-free terminal parameters (e.g. convergence time of echo cancellers) to yield satisfactory subjective performance.

The procedures defined here may also be appropriate for evaluating the subjective performance of other types of terminals and signal processing devices.

A complete subjective evaluation of hands-free telephones can be performed by the combination of three types of tests: conversational test, double talk test and third-party listening test (listening only test).

In general the evaluation of hands-free phones performance must take into account conversational interactions between subjects; a conversational test is the only type of subjective test which allows such an evaluation.

If a more detailed evaluation of a hands-free terminal is needed, it is recommended to perform double talk test and/or third-party listening tests additionally.

1.2 References

The following 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; all 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.

- [1] ITU-T Recommendation P.10 (1998), *Vocabulary of terms on telephone transmission quality and telephone sets*.
- [2] ITU-T Recommendation P.340 (2000), *Transmission characteristics and speech quality parameters of hands-free terminals*.
- [3] ITU-T Recommendation P.501 (2000), *Test signals for use in telephony*.
- [4] ITU-T Recommendation P.502 (2000), *Objective test methods for speech communication systems using complex test signals*.
- [5] ITU-T Recommendation P.51 (1996), *Artificial mouth*.
- [6] ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level*.
- [7] ITU-T Recommendation P.57 (1996), *Artificial ears*.
- [8] ITU-T Recommendation P.58 (1996), *Head and torso simulator for telephony*.
- [9] ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

- [10] ITU-T Recommendation P.810 (1996), *Modulated noise reference unit (MNRU)*.
- [11] ITU-T Recommendation P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs*.
- [12] ITU-T Recommendation P.581 (2000), *Use of head and torso simulator (HATS) for hands-free terminal testing*.
- [13] ITU *Handbook on Telephonometry*, 2nd edition, Geneva 1992.

1.3 Terms and definitions

This ITU-T Recommendation defines the following terms:

1.3.1 double talk: When near-end and far-end speech occur simultaneously at a given point, typically the terminal under test.

1.3.2 near end: The end of a network connection to which the HFT, whose characteristics are evaluated, is attached.

1.3.3 far end: The end of the network which is opposite to the near end.

1.3.4 syllable clipping or temporal clipping: Loss of speech energy caused by voice/speech activated devices. For echo cancellers, the primary source of temporal clipping is the NLP. In this instance, clipping does not refer to amplitude limiting.

1.3.5 third-party listening test: A listening-only subjective test (see ITU-T Recommendation P.800) in which the listener hears as an "ear witness" the acoustical recordings of the connection under evaluation. In conventional listening-only tests, the listener is positioned at one end of the connection under study.

1.3.6 conversation test: A subjective test in which two participants have a conversation, as described in Annex A/P.800 and in the *Handbook on Telephonometry*.

1.3.7 double talk test: A subjective test in which the participants are forced to talk simultaneously while simultaneously listening for impairments (e.g. echo).

1.3.8 untrained subject: See 3.3.1.

1.3.9 experienced subject: See 3.3.2.

1.3.10 experts: See 3.3.3.

1.3.11 ear signal: Signal recorded in the ear canal of a listener's ear.

1.4 Abbreviations

This ITU-T Recommendation uses the following abbreviations:

ACR	Absolute Category Rating
DCR	Degradation Category Rating
DMOS	Degradation Mean Opinion Score
GAT	Group Audio Terminal
HATS	Head And Torso Simulator (Recommendation P.58)
HFT	Hands-free Terminal
LRGP	Loudness Rating Guard Position
MNRU	Modulated Noise Reference Unit

MOS	Mean Opinion Score
MRP	Mouth Reference Point
NLP	Non-Linear Processor

2 Overview on test procedures

The test procedures suitable for the assessment of speech quality performance of hands-free terminals can be classified into three categories:

- 1) Conversational tests (see clause 4).
- 2) Double talk tests (see clause 5).
- 3) Third-party listening tests (see clause 6).

NOTE – Although headsets allow a "hands-free" operation and may be tested in general using the same procedures, they are not covered by this Recommendation.

In order to give guidance for the selection of the appropriate test procedure, information about advantages and/or disadvantages of a specific test procedure is listed in Table 1.

Table 1/P.832 – Advantages and disadvantages of different test procedures

	Advantages	Disadvantages
Conversational tests	• Very close to a real conversation • Preparation time is relatively short (compared to third-party listening tests)	• Subjects tend to have different behaviour in a conversation (due to culture, personality, etc.) which creates more response variability in assessing speech quality aspects • Since subjects have to concentrate on both running the conversation and taking care of the quality performance, they may be less sensitive to performance or quality • Devices under test and simulation tools must be available at the testing lab and must run in real time
Double talk tests	• Preparation time is relatively short (compared to third-party listening tests) • Evaluation of double talk capability in more detail than in conversational tests • Due to standardized dialogue structures, individual behaviour depending on culture and/or personality affecting double talk is reduced	• Subjects have to concentrate of both reading their text and taking care of the quality performance • Devices under test and simulation tools must be available at the testing lab and must run in real time