

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.804

(10/2017)

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

Methods for objective and subjective assessment of
speech and video quality

Subjective diagnostic test method for conversational speech quality analysis

Recommendation ITU-T P.804

ITU-T

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

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

Recommendation ITU-T P.804

Subjective diagnostic test method for conversational speech quality analysis

Summary

Recommendation ITU-T P.804 describes a subjective methodology for assessing and diagnosing the quality of transmitted speech in a telephone conversation. In addition to a score for the overall conversation quality, the methodology yields overall quality scores for three perceivable phases in a telephone conversation: listening, speaking, and interaction, as well as scores for their corresponding seven perceptual dimensions. Four of the perceptual dimension scores represent degradation associated with the listening phase, two are associated with the speaking phase, and one is associated with the interaction phase. Each of the perceptual dimension scores are based on ratings of the amount of degradation present in one system condition. The method is designed to be used with naïve subjects. The dimension scores can be used to provide diagnostic information on the causes of system degradations. The method is meant as a complement to standard conversation tests.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.804	2017-10-29	12	11.1002/1000/13397

Keywords

Conversational speech quality evaluation, diagnostic evaluation of conversational speech quality, multi-dimensional quality assessment, subjective testing.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). 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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2017

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

Table of Contents

	Page
1 Scope.....	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Recommendation.....	1
4 Abbreviations and acronyms	1
5 Conventions	2
6 Introduction to conversational speech quality analysis	2
7 Test methodology	3
7.1 Dimension rating scales.....	3
7.2 Test design.....	4
7.3 Dimension rating scheme	5
7.4 Instruction and training.....	6
Annex A – Test instructions – translated from German – Analysis of speech quality in a conversational situation	8
Appendix I – Examples for the speaking part of the test methodology – translated from German.....	13
Appendix II – Results from an initial pilot test and second retest	14
Bibliography.....	16

Downloaded from https://www.ava.com

Recommendation ITU-T P.804

Subjective diagnostic test method for conversational speech quality analysis

1 Scope

This Recommendation describes a subjective test methodology which is able to assess and diagnose the quality of speech in a "telephone conversation" scenario. Common conversation tests, as described in [ITU-T P.800] and [ITU-T P.805], provide valid methods for the overall conversational quality, but do not give insights into reasons for possible quality losses. In addition, common conversational tests lack analytic ability, since naïve participants concentrate on the conversation flow. To circumvent these problems, this Recommendation describes a test methodology that specifically allows participants to perceive each phase of a conversation separately, in addition to a natural conversation, and yields overall conversational quality scores as well as quality scores for each phase (listening, speaking, interaction). In addition, scores for seven underlying perceptual dimensions of conversational speech quality are provided. These scores enable the analysis of conversational speech quality for diagnosis and optimization.

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 P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.805] Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*.
- [ITU-T P.806] Recommendation ITU-T P.806 (2014), *A subjective quality test methodology using multiple rating scales*.
- [ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR Absolute Category Rating

MDS	Multidimensional Scaling
MOS	Mean Opinion Score
PCA	Principal Component Analysis
RNVT	Random Number Verification Task
SCT	Short Conversation Test
SD	Semantic Differential

5 Conventions

None.

6 Introduction to conversational speech quality analysis

To provide diagnostic information about the quality of transmitted speech, ITU-T recommends using multiple rating scales in subjective experiments in [ITU-T P.806]. The approach targets assessing perpetual dimensions to give deeper insight into possible quality loss. The recommended method refers to the passive listening-only situation and gives additional information to overall quality mean opinion score (MOS) absolute category rating (ACR)-experiments, as recommended in [ITU-T P.800]. However, the listening-only situation, only partly agrees with reality in telecommunications. Thus, conversation tests to assess the quality in an interactive situation have been designed and described in [ITU-T P.805]. The recommended methods do not provide diagnostic information. For this, a set of seven perceptual dimensions for a conversational situation are proposed in [b-Köster2014]. The proposed dimensions cover the three possible phases/situations of a conversation: listening, speaking, and interaction. See [b-Guéguin2008].

To identify the relevant proposed dimensions, each phase has been analysed in detail, applying a (I) pairwise similarity experiment with a following multidimensional scaling (MDS) and a (II) a semantic differential (SD) experiment with a following principal component analysis (PCA).

Applying both methods in separate experiments in [b-Köster2014] and [b-Wältermann2010] resulted in the following set of perceptual dimensions for a conversational situation (see Table 1): the listening phase is comprised of four dimensions: *noisiness*, *discontinuity*, *coloration* and *loudness*; the speaking phase is comprised of two dimensions: *impact of one's own voice on speaking* and *degradation of one's own voice*; the interaction phase is comprised of only one dimension: *interactivity*.