



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.862

(02/2001)

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

Methods for objective and subjective assessment of
quality

**Perceptual evaluation of speech quality (PESQ):
An objective method for end-to-end speech
quality assessment of narrow-band telephone
networks and speech codecs**

ITU-T Recommendation P.862

(Formerly CCITT Recommendation)

14-00000

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.

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

Summary

This Recommendation describes an objective method for predicting the subjective quality of 3.1 kHz (narrow-band) handset telephony and narrow-band speech codecs. This Recommendation presents a high-level description of the method, advice on how to use it, and part of the results from a Study Group 12 benchmark carried out in the period 1999-2000. An ANSI-C reference implementation, described in Annex A, is provided in separate files and form an integral part of this Recommendation. A conformance testing procedure is also specified in Annex A to allow a user to validate that an alternative implementation of the model is correct. This ANSI-C reference implementation shall take precedence in case of conflicts between the high-level description as given in this Recommendation and the ANSI-C reference implementation.

This Recommendation includes an electronic attachment containing an ANSI-C reference implementation of PESQ and conformance testing data.

Source

ITU-T Recommendation P.862 was prepared by ITU-T Study Group 12 (2001-2004) and approved under the WTSA Resolution 1 procedure on 23 February 2001.

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.

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 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 ITU.

CONTENTS

Page

1	Introduction	1
2	Normative references.....	1
3	Abbreviations	1
4	Scope	2
5	Conventions.....	4
6	Overview of PESQ	4
7	Comparison between objective and subjective scores.....	6
7.1	Correlation coefficient.....	6
7.2	Residual errors.....	6
8	Preparation of processed speech material.....	7
8.1	Source material.....	7
8.1.1	Choice of source material.....	7
8.1.2	ITU-T Temporal structure and duration of source material.....	7
8.1.3	Filtering and level calibration	8
8.2	Addition of background noise	8
8.3	Processing through system under test.....	8
9	Selection of experimental parameters	8
10	Description of PESQ algorithm.....	9
10.1	Level and time alignment pre-processing (Figure 3)	13
10.1.1	Computation of the overall system gain.....	13
10.1.2	IRS filtering.....	13
10.1.3	Time alignment	13
10.2	Perceptual model (Figures 4a and 4b).....	15
10.2.1	Precomputation of constant settings.....	15
10.2.2	IRS-receive filtering.....	15
10.2.3	Computation of the active speech time interval	15
10.2.4	Short-term Fast Fourier Transform	16
10.2.5	Calculation of the pitch power densities	16
10.2.6	Partial compensation of the original pitch power density for transfer function equalization	16
10.2.7	Partial compensation of the distorted pitch power density for time-varying gain variations between distorted and original signal	16
10.2.8	Calculation of the loudness densities	16
10.2.9	Calculation of the disturbance density	17
10.2.10	Cell-wise multiplication with an asymmetry factor	17

	Page
10.2.11 Aggregation of the disturbance densities over frequency and emphasis on soft parts of the original	17
10.2.12 Zeroing of the frame disturbance for frames during which the delay decreased significantly	18
10.2.13 Realignment of bad intervals.....	18
10.2.14 Aggregation of the disturbance within split second intervals	18
10.2.15 Aggregation of the disturbance over the duration of the speech signal (around 10 s), including a recency factor.....	18
10.2.16 Computation of the PESQ score.....	18
Annex A – Reference implementation of PESQ and conformance testing	19
Electronic attachment: ANSI-C reference implementation of PESQ and conformance testing data.	

ITU-T Recommendation P.862

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs¹

1 Introduction

The objective method described in this Recommendation is known as "Perceptual Evaluation of Speech Quality" (PESQ). It is the result of several years of development and is applicable not only to speech codecs but also to end-to-end measurements.

Real systems may include filtering and variable delay, as well as distortions due to channel errors and low bit-rate codecs. The PSQM method as described in ITU-T P.861 (February 1998), was only recommended for use in assessing speech codecs, and was not able to take proper account of filtering, variable delay, and short localized distortions. PESQ addresses these effects with transfer function equalization, time alignment, and a new algorithm for averaging distortions over time. The validation of PESQ included a number of experiments that specifically tested its performance across combinations of factors such as filtering, variable delay, coding distortions and channel errors.

It is recommended that PESQ be used for speech quality assessment of 3.1 kHz (narrow-band) handset telephony and narrow-band speech codecs.

2 Normative references

The following ITU-T Recommendations and other reference 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.

- ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- ITU-T P.810 (1996), *Modulated noise reference unit (MNRU)*.
- ITU-T P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs*.
- ITU-T P-series Supplement 23 (1998), *ITU-T coded-speech database*.

3 Abbreviations

This Recommendation uses, the following abbreviations:

ACR	Absolute Category Rating
CELP	Code-Excited Linear Prediction
DMOS	Degradation Mean Opinion Score
HATS	Head And Torso Simulator

¹ This Recommendation includes an electronic attachment containing an ANSI-C reference implementation of PESQ and conformance testing data.



IRS	Intermediate Reference System
LQ	Listening Quality
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure

4 Scope

Based on the benchmark results presented within Study Group 12, an overview of the test factors, coding technologies and applications to which this Recommendation applies is given in Tables 1 to 3. Table 1 presents the relationships of test factors, coding technologies and applications for which this Recommendation has been found to show acceptable accuracy. Table 2 presents a list of conditions for which the Recommendation is known to provide inaccurate predictions or is otherwise not intended to be used. Finally, Table 3 lists factors, technologies and applications for which PESQ has not currently been validated. Although correlations between objective and subjective scores in the benchmark were around 0.935 for both known and unknown data, the PESQ algorithm cannot be used to replace subjective testing.

It should also be noted that the PESQ algorithm does not provide a comprehensive evaluation of transmission quality. It only measures the effects of one-way speech distortion and noise on speech quality. The effects of loudness loss, delay, sidetone, echo, and other impairments related to two-way interaction (e.g. centre clipper) are not reflected in the PESQ scores. Therefore, it is possible to have high PESQ scores, yet poor quality of the connection overall.

Table 1/P.862 – Factors for which PESQ had demonstrated acceptable accuracy

Test factors
Speech input levels to a codec
Transmission channel errors
Packet loss and packet loss concealment with CELP codecs
Bit rates if a codec has more than one bit-rate mode
Transcodings
Environmental noise at the sending side (See Note.)
Effect of varying delay in listening only tests
Short-term time warping of audio signal
Long-term time warping of audio signal
Coding technologies
Waveform codecs, e.g. G.711; G.726; G.727
CELP and hybrid codecs ≥ 4 kbit/s, e.g. G.728, G.729, G.723.1
Other codecs: GSM-FR, GSM-HR, GSM-EFR, GSM-AMR, CDMA-EVRC, TDMA-ACELP, TDMA-VSELP, TETRA

Applications
Codec evaluation
Codec selection
Live network testing using digital or analogue connection to the network
Testing of emulated and prototype networks
NOTE – When environmental noise is present the quality can be measured by passing PESQ the clean original without noise, and the degraded signal with noise.

Table 2/P.862 – PESQ is known to provide inaccurate predictions when used in conjunction with these variables, or is otherwise not intended to be used with these variables

Test factors
Listening levels (See Note.)
Loudness loss
Effect of delay in conversational tests
Talker echo
Sidetone
Coding technologies
Replacement of continuous sections of speech making up more than 25% of active speech by silence (extreme temporal clipping)
Applications
In-service non-intrusive measurement devices
Two-way communications performance
NOTE – PESQ assumes a standard listening level of 79 dB SPL and compensates for non-optimum signal levels in the input files. The subjective effect of deviation from optimum listening level is therefore not taken into account.

Table 3/P.862 (For further study) – Factors, technologies and applications for which PESQ has not currently been validated

Test factors
Packet loss and packet loss concealment with PCM type codecs (See Note 1.)
Temporal clipping of speech (See Note 1.)
Amplitude clipping of speech (See Note 2.)
Talker dependencies
Multiple simultaneous talkers
Bit-rate mismatching between an encoder and a decoder if a codec has more than one bit-rate mode
Network information signals as input to a codec
Artificial speech signals as input to a codec
Music as input to a codec