

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.564

(11/2007)

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

Objective measuring apparatus

Conformance testing for voice over IP transmission quality assessment models

ITU-T Recommendation P.564

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
Transmission performance and QoS aspects of IP end-points	Series	P.1000

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

Document released under license

ITU-T Recommendation P.564

Conformance testing for voice over IP transmission quality assessment models

Summary

ITU-T Recommendation P.564 specifies the minimum criteria for objective speech quality assessment models that predict the impact of observed IP network impairments on the one-way listening quality experienced by the end-user in IP/UDP/RTP-based telephony applications (3.1-kHz narrow-band in the main body, 7 kHz wideband in Annex B). Models compliant with this Recommendation predict mean opinion scores (MOS) on the ACR listening quality scale. It is expected that the primary applications for such models are monitoring of transmission quality for operations and maintenance purposes, and measurements in support of service level agreements (SLAs) between service providers and their customers. P.564-conformant models may be deployed both in endpoint locations and at mid-network monitoring points.

This Recommendation includes an electronic attachment containing the test vectors.

...

Source

ITU-T Recommendation P.564 was approved on 13 November 2007 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

Keywords

ACR, conformance, model, MOS, P.862, VoIP.

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 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 2008

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
2	References.....	2
3	Definitions	3
4	Abbreviations and acronyms	3
5	Conventions	4
6	Application scenarios.....	4
6.1	General comments	4
6.2	Network problem diagnosis.....	4
6.3	IP Centrex	4
6.4	Managed IP service SLA.....	5
6.5	Call quality monitoring for reporting	5
6.6	Active test calls.....	5
7	Modes of operation	5
7.1	Overview	5
7.2	Dynamic operation	6
7.3	Static operation.....	6
7.4	Embedded operation.....	6
7.5	Reference endpoints	7
8	Inputs	7
9	Outputs.....	8
9.1	Speech quality (mandatory).....	8
9.2	Interpretation of quality predictions	8
9.3	Diagnostics (optional)	9
9.4	Compliance statement (mandatory).....	10
10	Conformance test methodology	11
10.1	Overview	11
10.2	Test vector generation	14
10.3	Processing the test vectors.....	17
10.4	Performance analysis.....	17
11	Conformance accuracy requirements	19
	Annex A – File formats and naming conventions.....	21
A.1	General	21
A.2	Test vector files	21
A.3	Test vector quality score files.....	21
	Annex B – Extension of P.564 to wideband voice over IP quality assessment models	22
	Appendix I – Example statement of compliance	23
	Electronic attachment: Test vectors for narrow-band and wideband applications	

ITU-T Recommendation P.564

Conformance testing for voice over IP transmission quality assessment models

1 Scope

This Recommendation specifies minimum criteria for objective speech quality assessment models that predict the impact of observed IP network impairments on the one-way listening quality experienced by the end-user in IP/UDP/RTP-based 3.1-kHz narrow-band telephony applications. An extension to 7 kHz wideband telephony is also provided in Annex B.

It is expected that the primary applications for such models are monitoring of transmission quality for operations and maintenance purposes, and measurements in support of service level agreements (SLAs) between service providers and their customers.

Models compliant with this Recommendation predict mean opinion scores (MOS) on the ACR listening quality scale. Their performance is based and estimated on the MOS scale (as defined in [ITU-T P.800]¹). The primary quality prediction made by such a model is not based on the payload of the RTP stream being analysed, but assumes a typical, or generic, voice payload. Some additional diagnostic outputs may be based on the payload, if available. A model compliant with this Recommendation should always take the voice codec into account. If any input parameter is unavailable and an assumed value is used, this fact should be reported.

A model compliant with this Recommendation cannot provide a comprehensive end-to-end evaluation of transmission quality because its scores can only reflect the impairments on the IP network being measured that may only be part of the end-to-end connection.

The effects of speech level, acoustic background noise, delay, sidetone, echo and other impairments related to the payload are not reflected in the scores computed by such a model. Therefore, it is possible to have high scores with a model compliant with this Recommendation, yet have a poor quality of the connection overall.

The accuracy criteria herein were derived with the intent to avoid the frequent occurrence of "false positive" or "false negative" errors. This is an especially important consideration when a model compliant with this Recommendation is used, for example, in assessing compliance with SLAs.

The criteria for the model described in this Recommendation are applicable to devices that may reside anywhere within the packet transport network, including edge devices. As such, each will be able to use only information present at the location deployed. However, this does include information that can be extracted from RTCP-SR, RR and XR.

The accuracy criteria are based on a comparison of a model's performance with the P.862 perceptual evaluation of speech quality (PESQ) algorithm using the output mapping defined in [ITU-T P.862.1] for 3.1-kHz narrow-band telephony, hereafter referred to as P.862.1. Hence, compliance to this Recommendation shall only be claimed for factors, technologies or applications that are within the scope of P.862, or for which the operation of P.862 has been verified against subjective test data.

The speech test material described in clause 10.2.1 is provided in an electronic attachment and forms an integral and normative part of this Recommendation.

¹ These scores can be translated into *Ie-eff* values using the formulae given in Annex B of [ITU-T G.107] (equation B-4) and in Appendix I of [ITU-T G.107] (equations I-1, I-2 and I-3).

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 G.107] ITU-T Recommendation G.107 (2005), *The E-model, a computational model for use in transmission planning*.
- [ITU-T G.191] ITU-T Recommendation G.191 (2005), *Software tools for speech and audio coding standardization*.
- [ITU-T G.711] ITU-T Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.
- [ITU-T G.729] ITU-T Recommendation G.729 (2007), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)*.
- [ITU-T G.799.1] ITU-T Recommendation G.799.1/Y.1451.1 (2004), *Functionality and interface specifications for GSTN transport network equipment for interconnecting GSTN and IP networks*.
- [ITU-T G.1020] ITU-T Recommendation G.1020 (2006), *Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks*.
- [ITU-T G.1050] ITU-T Recommendation G.1050 (2007), *Network model for evaluating multimedia transmission performance over Internet Protocol*.
- [ITU-T H.248.30] ITU-T Recommendation H.248.30 (2007), *Gateway control protocol: RTCP extended performance metrics packages*.
- [ITU-T H.460.9] ITU-T Recommendation H.460.9 (2002), *Support for online QoS-monitoring reporting within H.323 systems*.
- [ITU-T P.56] ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level*.
- [ITU-T P.501] ITU-T Recommendation P.501 (2007), *Test signals for use in telephony*.
- [ITU-T P.800] ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.800.1] ITU-T Recommendation P.800.1 (2006), *Mean Opinion Score (MOS) terminology*.
- [ITU-T P.862] ITU-T Recommendation P.862 (2001), *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 P.862.1] ITU-T Recommendation P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO*.
- [ITU-T P.862.2] ITU-T Recommendation P.862.2 (2007), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*.

[ITU-T Y.1541]	ITU-T Recommendation Y.1541 (2006), <i>Network performance objectives for IP-based services</i> .
[IETF RFC 768]	IETF RFC 768 (1980), <i>User Datagram Protocol</i> .
[IETF RFC 791]	IETF RFC 791 (1981), <i>Internet Protocol</i> .
[IETF RFC 3550]	IETF RFC 3550 (2003), <i>RTP: A Transport Protocol for Real-Time Applications</i> .
[IETF RFC 3555]	IETF RFC 3555 (2003), <i>MIME Type Registration of RTP Payload Formats</i> .
[IETF RFC 3611]	IETF RFC 3611 (2003), <i>RTP Control Protocol Extended Reports (RTCP XR)</i> .

3 Definitions

This Recommendation defines the following term:

3.1 speech sample: A series of one or more speech utterances.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR	Absolute Category Rating
IP	Internet Protocol
LAN	Local Area Network
MOS	Mean Opinion Score
MOS-LQO _N	Mean Opinion Score, Listening Quality Objective measurement in Narrow-band audio context
OOS	Out Of Sequence
PBX	Private Branch eXchange
PCAP	Packet CAPture
PDV	Packet Delay Variation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
RFC	Request For Comments
ROI	Region Of Interest
RR	Receiver Report
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
SLA	Service Level Agreement
SR	Sender Report
UDP	User Datagram Protocol
VAf	Voice Activity Factor
VoIP	Voice over Internet Protocol
XR	eXtended Report

5 Conventions

The conventions used in mathematical equations in this Recommendation are defined at the appropriate points in the text.

6 Application scenarios

6.1 General comments

When using a P.564 device as a midpoint monitor, it is assumed that it will have visibility of all packets in the data stream of interest. Note, however, that due to the nature of IP networks, each packet is routed individually with no guarantee of any path consistency. In situations where not all packets are visible, scores may not represent the actual user experience.

6.2 Network problem diagnosis

By continuously monitoring network performance, midpoint probes, based on a model compliant with this Recommendation, can continuously gather and report information on service quality. This information can then be used to classify and isolate network problems.

6.2.1 Temporal correlation

It may be possible to correlate service quality events with other network problems that caused the change in quality. In this application, it is useful to know the approximate time interval during which service quality was degraded (e.g., start time and duration).

6.2.2 Impairment correlation

It may also be possible to identify events occurring on multiple call streams that are similar in nature, to help identify common causes of problems within a network. In this application, diagnostic information can help to characterize the problem.

6.2.3 Location correlation

The location of problem links in a network may be identified by monitoring the quality of the same stream at multiple points (see Figure 6-1).

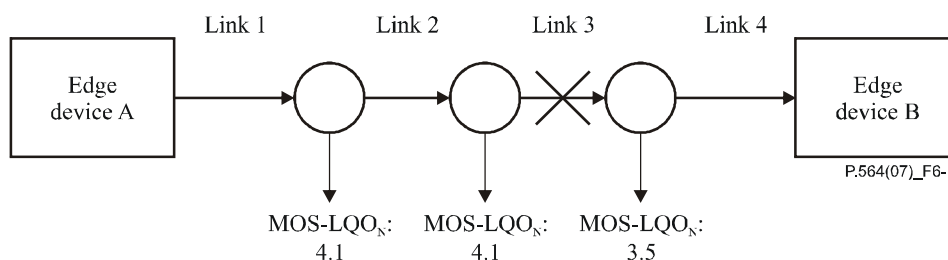


Figure 6-1 – Using multiple monitoring points to identify problem links

6.3 IP Centrex

IP Centrex (or hosted PBX) service quality management can be facilitated by monitoring at both the service provider, customer network demarcation point, and the endpoint (e.g., IP phone). This helps the service provider to understand whether problems are located within their network or the customer LAN.