International Telecommunication Union

# ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

**G.1029**

(02/2014)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Multimedia Quality of Service and performance – Generic and user-related aspects

# Voice service diagnosis framework

Recommendation ITU-T G.1029

ITU-T

International Telecommunication Union

ITU-T G-SERIES RECOMMENDATIONS

**TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS**

| | |
|---|---|
| INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS | G.100–G.199 |
| GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS | G.200–G.299 |
| INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES | G.300–G.399 |
| GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES | G.400–G.449 |
| COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY | G.450–G.499 |
| TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS | G.600–G.699 |
| DIGITAL TERMINAL EQUIPMENTS | G.700–G.799 |
| DIGITAL NETWORKS | G.800–G.899 |
| DIGITAL SECTIONS AND DIGITAL LINE SYSTEM | G.900–G.999 |
| **MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS** | **G.1000–G.1999** |
| TRANSMISSION MEDIA CHARACTERISTICS | G.6000–G.6999 |
| DATA OVER TRANSPORT – GENERIC ASPECTS | G.7000–G.7999 |
| PACKET OVER TRANSPORT ASPECTS | G.8000–G.8999 |
| ACCESS NETWORKS | G.9000–G.9999 |

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

# Recommendation ITU-T G.1029

## Voice service diagnosis framework


### Summary

Recommendation ITU-T G.1029 provides a framework and guidelines that describe how ITU-T speech quality assessment models can be used to identify common voice quality problems in live networks and how these can aid in diagnosing the cause of such problems once detected.



### History

| Edition | Recommendation | Approval | Study Group | Unique ID* |
|---|---|---|---|---|
| 1.0 | ITU-T G.1029 | 2014-02-13 | 12 | 11.1002/1000/12121 |

### Keywords

Cause analysis, ITU-T objective model, problem diagnosis, voice service.


* 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/.

## Table of Contents


| | | Page |
|---|---|---|
| 1 | Scope | 1 |
| 2 | References | 1 |
| 3 | Definitions | 2 |
| | 3.1 Terms defined elsewhere | 2 |
| | 3.2 Terms defined in this Recommendation | 2 |
| 4 | Abbreviations and acronyms | 2 |
| 5 | Conventions | 3 |
| 6 | Framework concept and architecture | 3 |
| 7 | Introduction to speech quality measurement and assessment models | 3 |
| | 7.1 Subjective and objective testing | 3 |
| | 7.2 Classes of objective models | 5 |
| | 7.3 Measurement configurations | 7 |
| 8 | Measurement data sources | 8 |
| | 8.1 Intrusive test probes | 8 |
| | 8.2 Non-intrusive test probes | 8 |
| | 8.3 Embedded probes | 8 |
| | 8.4 Additional metrics | 8 |
| | 8.5 Multi-point measurements | 9 |
| | 8.6 Signalling related measurements | 9 |
| 9 | Decision process | 9 |
| | 9.1 Thresholds | 9 |
| | 9.2 Analysing metrics derived from live signals | 9 |
| | 9.3 Mapping basic diagnostic parameters into root cause analysis | 10 |
| Appendix I – Example based on call clarity index | | 11 |
| Bibliography | | 13 |

# Recommendation ITU-T G.1029

## Voice service diagnosis framework

### 1 Scope

This Recommendation provides a framework and guidelines that describe how ITU-T speech quality assessment models can be used to identify common voice quality problems in live networks and how these can aid in diagnosing the cause of such problems once detected. The scope of this Recommendation is limited to voice service only.

The Recommendation is intended to help non voice experts tasked with managing voice service and diagnosing voice quality problems, especially when voice is regarded as "just another IP application" on the network.

The scope of this Recommendation relates to the perceived quality of the media stream; it does not include factors such as billing, availability, customer service, etc. The primary focus is point-to-point voice services; however, the methods and approaches described can also be applied to the individual legs of multi-party voice applications such as audio-conferencing.

### 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.561] Recommendation ITU-T P.561 (2002), *In-service non-intrusive measurement device – Voice service measurements*.

[ITU-T P.562] Recommendation ITU-T P.562 (2004), *Analysis and interpretation of INMD voice-service measurements*.

[ITU-T P.563] Recommendation ITU-T P.563 (2004), *Single-ended method for objective speech quality assessment in narrow-band telephony applications*.

[ITU-T P.564] Recommendation ITU-T P.564 (2007), *Conformance testing for voice over IP transmission quality assessment models*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[ITU-T P.800.1] Recommendation ITU-T P.800.1 (2006), *Mean Opinion Score (MOS) terminology*.

[ITU-T P.800.2] Recommendation ITU-T P.800.2 (2013), *Mean opinion score interpretation and reporting*.

[ITU-T P.862] Recommendation ITU-T 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] Recommendation ITU-T P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO*.

[ITU-T P.862.2] Recommendation ITU-T P.862.2 (2007), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*.

[ITU-T P.863] Recommendation ITU-T P.863 (2011), *Perceptual objective listening quality assessment*.

[ITU-T Y.1540] Recommendation ITU-T Y.1540 (2011), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[IETF RFC 3550] IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications*.

# 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:

| | |
|---|---|
| CCI | Call Clarity Index |
| CQO | Conversational Quality Objective |
| ERL | Echo Return Loss |
| INMD | In-line Non-intrusive Measurement Device |
| IP | Internet Protocol |
| LQO | Listening Quality Objective |
| $LQO_N$ | Listening Quality Objective narrowband |
| $LQO_W$ | Listening Quality Objective wideband |
| $LQO_{SWB}$ | Listening Quality Objective super-wideband |
| LQS | Listening Quality Subjective |
| MOS | Mean Opinion Score |
| PCM | Pulse Code Modulation |
| PLC | Packet Loss Concealment |
| PSTN | Public Switched Telephone Network |
| RMON | Remote network Monitoring |
| RTCP | Real Time Control Protocol |
| RTCP XR | RTP Control Protocol extended Reports |
| RTP | Real time Transport Protocol |
| SIP | Session Initiation Protocol |
| UDP | User Datagram Protocol |
| VOIP | Voice Over IP |

## 5 Conventions

None.

## 6 Framework concept and architecture

This Recommendation provides a set of high level guidelines. It provides an introduction to speech quality measurement and assessment models, measurement data sources and general considerations relating to defining decision thresholds and mapping diagnostic values to the root cause of voice service problems.

Appendix I is provided as an example and is used to illustrate some of the concepts described in this Recommendation.

The ITU-T intends to provide two more Recommendations addressing, in detail, a framework for invoking diagnostic functions and a framework for mapping diagnostic parameters to root cause parameters, as follows:

– Framework for invoking diagnostic functions: Identification of decision parameters that can be used to detect the presence of a media quality problem in a network so that a diagnostic function can be invoked.

– Framework for mapping diagnostic parameters to root cause parameters: This work item is concerned with mapping basic diagnostic parameters into root cause analysis parameters.

Figure 1 shows the roadmap for the development of these two Recommendations.

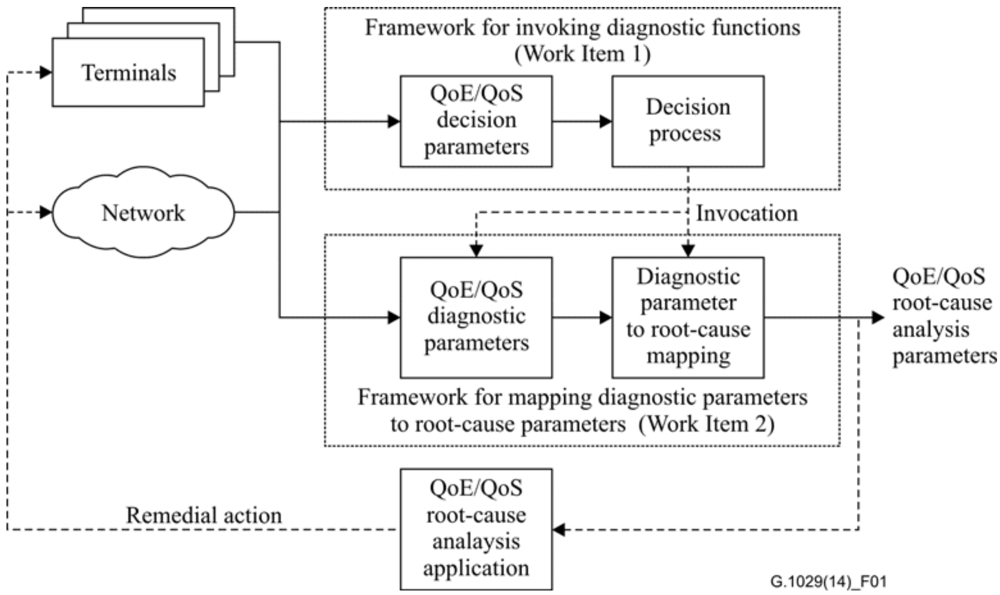


**Figure 1 – Roadmap for further development of framework Recommendations**

## 7 Introduction to speech quality measurement and assessment models

### 7.1 Subjective and objective testing

The perceived quality of a voice call is a subjective quantity. This means that the baseline for voice and audio quality is the subjective opinion of human listeners. One person's impression of "good" may be quite different to another person's impression, but neither is incorrect. Communications systems are therefore designed and tested against an "average" person's perception of voice quality. This is often summarized by the term mean opinion score (MOS). There are essentially two complementary approaches to measuring the quality of a voice signal as it will be perceived by the end user. These are:

– Subjective testing

– Objective testing

These two approaches are described in the following clauses.

### 7.1.1 Subjective testing

Subjective tests aim to find the average user's perception of the voice quality delivered by a communications system. This is done by asking a panel of users a directed question and providing them with a limited response choice. For example, to determine the listening quality of a voice signal, users are asked to rate "the quality of the speech" on a five-point discrete scale from bad to excellent as described in [ITU-T P.800].

The mean opinion score (MOS) for a particular test condition is calculated by averaging the votes of all subjects for that particular condition. A subjective test will typically contain many different conditions. Therefore, such tests take a long time to perform and the results are influenced by a wide range of factors.

### 7.1.2 Objective testing

Objective testing techniques measure physical properties of a system. Objective perceptual models map these physical properties to a predicted subjective score. In comparison with subjective testing, objective measurements are fast, inexpensive and repeatable. Significant work has led to objective prediction techniques that can be used in situations where it is impractical to perform formal subjective testing.

### 7.1.3 MOS notation

Voice subjective experiments are generally designed to either measure conversational quality or listening (one-way) quality. Conversational experiments investigate how effects such as delay, echo and level affect the ability of two people to carry out a conversation. Listening-only experiments are concerned with the perceived quality delivered by one side of a link and take into account factors such as distortion due to voice compression and packet loss. This distinction is also true for objective speech quality models and they can be designed to predict either listening quality or conversational quality.

The ITU-T has defined terms to denote the difference between conversational and listening quality MOS values and the difference between MOS values from subjective experiments and MOS predictions made by objective models. The ITU-T notation defined in [ITU-T P.800.1] is shown in Table 1. The suffixes N (narrowband), W (wideband) and SWB (super-wideband)[1] are also sometimes added to the notation to denote whether the subjective experiment, or predicted experiment, was conducted in a narrowband (300-3 400 Hz), wideband (50-7 000 Hz) or super-wideband (50-14 000 Hz) context respectively. For example, listening quality assessed in a narrowband context subjective experiment would be denoted as MOS-$LQS_N$.

The reader is also directed to [ITU-T P.800.2], which introduces some of the more common types of mean opinion score (MOS) and describes the minimum information that should accompany MOS values to enable them to be correctly interpreted.

[1] The SWB suffix is not explicitly mentioned in [ITU-T P.800.1], but the use of such extensions is encouraged where appropriate.

**Table 1 – ITU-T MOS notation**

| Description | Notation |
|---|---|
| Subjectively assessed listening quality | MOS-LQS |
| Subjectively assessed conversational listening quality | MOS-CQS |
| Objectively assessed listening quality | MOS-LQO |
| Objectively assessed conversational listening quality | MOS-CQO |

## 7.2 Classes of objective models

In general, objective voice quality models fall into one of four classes: full-reference, reduced-reference, no-reference or parametric. The following clauses introduce the main classes of models and introduce some of the relevant ITU-T Recommendations. Table 2 provides an overview of these Recommendations.

**Table 2 – Summary of ITU-T objective voice quality model standards**

| Model | Scope / brief definition | Input | Output |
|---|---|---|---|
| **[ITU-T P.863]** | Full-reference model (LQO) | Reference and degraded 16-bit PCM signals, 8 and 48 kHz | $\text{MOS-LQO}_N$, $\text{MOS-LQO}_{SWB}$ |
| **[ITU-T P.862]**<br>**[ITU-T P.862.1]**<br>**[ITU-T P.862.2]** | Full-reference model (LQO) | Reference and degraded 16-bit PCM signals, 8 and 16 kHz | $\text{MOS-LQO}_N$, $\text{MOS-LQO}_W$ |
| **[ITU-T P.561]** | No-reference model (CQO) | PCM signals (for INMD classes A, B and C)<br>IP packets (for INMD class D) | Speech, noise and echo characterization |
| **Annex A of [ITU-T P.562]** | No-reference model (CQO) | ITU-T P.561 output | $\text{MOS-CQO}_N$ |
| **[ITU-T P.563]** | No-reference (LQO) | 8 kHz 16-bit PCM signal | $\text{MOS-LQO}_N$ |
| **[ITU-T P.564]** | Parametric (LQO) | RTP, UDP and IP headers | $\text{MOS-LQO}_N$, $\text{MOS-LQO}_W$ |

NOTE – This table lists only what is currently available and it is envisioned that this table will be updated.

### 7.2.1 Full-reference models

Full-reference objective models measure the impact on perceived voice quality of one or more network elements by comparing two versions of a test signal. As shown in Figure 2, the first signal is a copy of the original signal that was injected into the system under test; the second is the received signal, which has typically been degraded. Full-reference models can be used to measure the quality of a single network element, such as a codec, or an entire link, such as a mobile radio link. [ITU-T P.863] is an example of a full-reference objective voice model that predicts a listening quality MOS or MOS-LQO [ITU-T P.863].

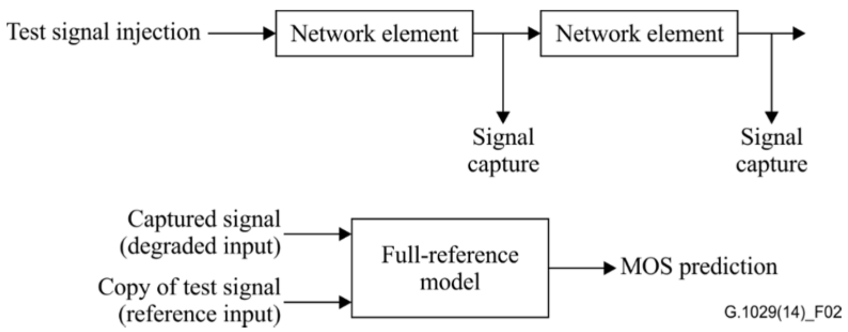


**Figure 2 – Diagram of an intrusive measurement configuration and a full-reference model**

### 7.2.2 Reduced-reference models

Reduced-reference models are similar to full-reference models, but use a reduced set of information about the original test signal rather than an exact copy. Reduced-reference models are typically used in applications where the reference signal is live traffic rather than a predetermined test signal, as shown in Figure 3. This means that information about the reference signal must be transmitted to the point of assessment if a comparison is to be made between the two.

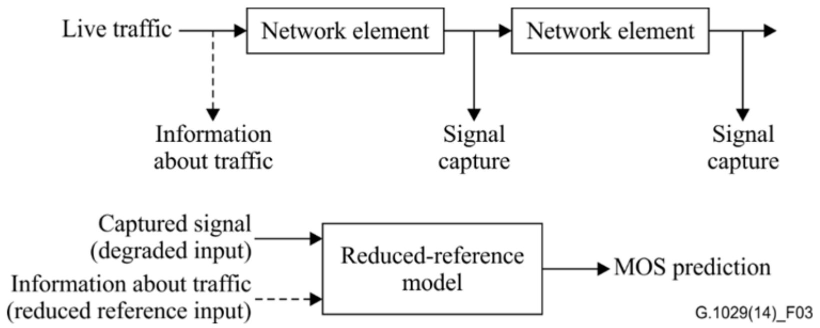


**Figure 3 – Diagram of a non-intrusive measurement configuration and a reduced-reference model**

### 7.2.3 No-reference models

No-reference, or single-ended, models base their assessment on a single input signal and are typically used to monitor the quality of live traffic as shown in Figure 4. [ITU-T P.563] is an example of a no-reference objective speech model that predicts a listening quality MOS, or MOS-LQO [ITU-T P.563]; [ITU-T P.562] includes an example of a no-reference model that predicts conversational quality MOS, or MOS-CQO [ITU-T P.562].

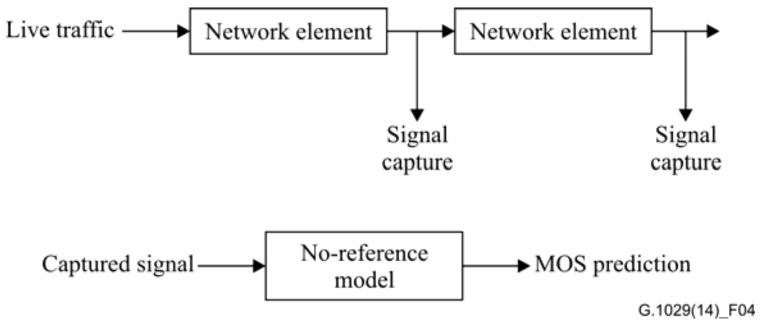


**Figure 4 – Diagram of a non-intrusive measurement configuration and a no-reference model**

### 7.2.4 Parametric models

Parametric models are designed to predict the impact of transmission impairments on the voice quality perceived by the end user of a system. Parametric models are no-reference models, but analyze parameters related to the underlying transmission system rather than the actual voice signal itself. For example, in the case of a voice over IP (VoIP) system, the main parameters will be derived from the packet loss and jitter characteristics of the link. Parametric voice quality models base their predictions on the assumption that the payload contains a typical, well-conditioned voice signal.

Parametric models are useful in VoIP applications because simple measurements such as packet loss and jitter often correlate poorly with the quality perceived by the end-user. For example, 1% bursty packet loss may produce a number of mutes that will substantially degrade the voice quality perceived by the end-user; whereas 1% uniform packet loss may be largely concealed if the end-point includes an effective packet loss concealment (PLC) algorithm.

An example of a parametric model is provided in [ITU-T P.564], which is concerned with a class of MOS-LQO prediction models that only analyse the IP/UDP/RTP header portion of the VoIP packets [ITU-T P.564]. This approach has several benefits including low-operational complexity and resilience to payload encryption. Note that [ITU-T P.564] does not specify a unique parametric model; it specifies minimum performance criteria that must be met by a model to achieve conformance with the Recommendation.

## 7.3 Measurement configurations

Objective measurement techniques can be categorized as either intrusive (active) or non-intrusive (passive). The distinction between classes of model and measurement configuration is often unclear because full-reference models are generally used in an intrusive configuration whilst no-reference and parametric models are used in a non-intrusive configuration. However, there can be exceptions to this rule, and so it is worth retaining the distinction.

### 7.3.1 Intrusive (active) measurement

In an intrusive measurement configuration a test signal is injected into the system under test and then captured and assessed at a later point as shown in Figure 2.

Intrusive measurement is generally performed using a full-reference model because the test signal is known *a priori* and a copy can be stored at the assessment point. However, there is no reason in principal why a reduced-reference or no-reference model could not be used to measure the quality of the signal at the assessment point.

Intrusive measurement is sometimes referred to as "active" measurement and can be used during the development, commissioning and routine monitoring of a communications service. Intrusive measurement configurations are also highly suitable for laboratory based testing because it is easy to isolate network components and inject and capture test signals.

### 7.3.2 Non-intrusive (passive) measurement

Non-intrusive measurement configurations are generally used to monitor live traffic, as shown in Figure 4.

Non-intrusive measurement is typically performed using a no-reference or parametric model, although a reduced-reference model can also be used as described above. It is important that no-reference models can operate reliably over a very wide range of input signals because the input signal is not restricted to a carefully controlled test signal as it is with intrusive measurement. Parametric models avoid this issue because they do not use the voice signal in the quality calculation, but use an assumed voice signal instead.

In VoIP testing, non-intrusive probes have visibility of all traffic at the monitoring point, including non-voice services. This can be useful in situations where the traffic profile is causing quality problems. For example, if VoIP traffic is transmitted with the wrong class-of-service marking, it may suffer from excessive jitter or packet loss due to the presence of other, non-voice traffic with the same marking.

Non-intrusive measurement is sometimes referred to as "passive" measurement.

# 8 Measurement data sources

## 8.1 Intrusive test probes

Intrusive test systems typically inject and capture at the edges of the network being tested, although it is possible to make multiple measurements from the same test signal as shown in Figure 2.

Drive test systems are typically used to measure the voice quality of mobile systems as a function of geographical location. They comprise one or more central servers and a number of special mobile test devices. During each test call, a known test signal is sent in each direction and the received signal is processed by a full-reference model such as [ITU-T P.862] or [ITU-T P.863] to produce MOS-LQO values for each direction. Multiple MOS-LQO measurements may be made during each test call to improve measurement accuracy.

VoIP test units are typically located at different points around the network and are configured to make scheduled test calls to each other. Such testing allows the ability of an IP network to carry VoIP traffic to be tested prior to the deployment of the actual VoIP system or at times when the system would otherwise be inactive. Such test arrangements can also be used to monitor live systems, although measurements are limited to the test traffic and not the customer traffic.

## 8.2 Non-intrusive test probes

Non-intrusive, or passive, probes are generally used to monitor live network traffic.

In a VoIP system, copies of packets may be taken at monitoring locations by means of a network tap or a mirror port configured on a switch or a router. VoIP streams can then be analysed using no-reference and/or parametric models to produce MOS-LQO and/or MOS-CQO values in addition to other voice diagnostic parameters, such as voice level, noise level, echo etc.

Non-intrusive VoIP probes are typically located at key demarcation points in the network, such as public switched telephone network (PSTN) gateways, session initiation protocol (SIP) peering points, customer edge locations and conference bridges. Some vendors make signal quality measurements within the VoIP end-point. The methods by which such measurements are made available vary from vendor to vendor, but the aim is generally to send the measurements to a central point where the data can be analysed. The RTP control protocol extended reports (RTCP XR) mechanism defined in [b-IETF RFC 3611] describes a mechanism whereby measurements made at VoIP end-points are transported along the path taken by the voice packets and can thus be collected by monitoring probes at mid-network locations [b-IETF RFC 3611].

## 8.3 Embedded probes

In addition to statistics collected at the monitoring locations, many network elements are capable of making transmission measurements and reporting them to a central data collector. A number of vendors support the collection of IP flow based metrics, and some have been standardized by the IETF in the form of the remote network monitoring (RMON) family of RFCs.

## 8.4 Additional metrics

Quality problems caused by transmission impairments represent an important class of diagnostic information, and it is recommended that signal related metrics, such as MOS, voice level, noise level, echo, etc. are accompanied by associated transmission statistics.

For example, in a VoIP system statistics relating to packet loss, jitter, out-of-sequence packets, round-trip delay and class-of-service markings all provide useful information when diagnosing problems that are the result of IP transmission problems. Such metrics are outside the scope of this Recommendation, but the reader is directed to the measurements defined in [ITU-T Y.1540] and the real-time control protocol (RTCP) measurements defined in [IETF RFC 3550].

## 8.5 Multi-point measurements

The location of transmission problems that affect voice quality can be isolated by collecting and comparing measurements from multiple monitoring locations in the network. However, when performing such analysis, it is important to ensure that values such as MOS-LQO scores have been calculated using the same method or their comparison may be misleading.

## 8.6 Signalling related measurements

Signalling monitoring forms an important part of monitoring a VoIP deployment, but is outside the scope of this Recommendation. The ITU-T E-Series of Recommendations specify metrics that relate to signalling performance.

# 9 Decision process

Designing the optimum strategy for launching diagnostic processes will depend on the specific application, however the following points should be considered.

## 9.1 Thresholds

Detailed diagnostic or troubleshooting activities will generally be launched when one or more thresholds have been violated. Such thresholds may have absolute values or relative values, for example to detect deviations from normal operation.

Many underlying problems will often only affect a subset of all calls. Hence applying thresholds to global averages across all calls may mean that decision thresholds are not violated because the problem traffic is too diluted.

Invoking diagnostic analysis for every call that violates a threshold runs the risk of overwhelming the diagnostic function and generating "false-positive" alarms. This is especially true in the case of objective MOS values and other signal metrics derived from no-reference analysis of live calls (see clause 9.2). Thresholding the output of a parametric model might be appropriate in situations where there is little or no tolerance for transmission related quality problems.

One means of achieving a balance is to define a threshold in terms of the proportion of calls that have violated an underlying threshold. For example, an alarm might be triggered if more than 10% of traffic in a given analysis period had a MOS-LQO value lower than 3.5. Such an approach can highlight problems that affect only a subset of calls, whilst still providing a degree of aggregation and protection from false-positive events.

## 9.2 Analysing metrics derived from live signals

Non-intrusive measurements provide an effective method of identifying systemic or underlying problems from live traffic measurements, i.e., problems that are consistently present, but it is important to ensure that a sufficient number of calls are analyzed before drawing any conclusions and invoking further diagnostic processes. Many factors relating to the quality of live customer calls naturally vary from call to call, for example the talker's speech characteristics, their acoustic environment, etc. This means that objective MOS values produced by a no-reference model will also naturally vary from call to call. The same is also true of other signal metrics calculated from live traffic, such as voice level, noise level and echo return loss.

This inherent variability in signal related metrics means that inferences about system performance drawn from them for a single call are likely to be unreliable, and a suitable sample of calls should be used instead. However, it should be emphasized that since the main purpose of this type of analysis is to identify systemic issues, it is not necessary to analyse all calls – merely to ensure that a sufficient number of calls have been analysed. If a problem is systemic, analysis of a sub-sample of calls will be sufficient to identify the problem.

The minimum sample size used in a decision process will depend on the decision mechanism and the natural variability of the metric being considered. Analysis of the standard deviation of a sample should provide an indication of how many samples are required to gain reliable results. The use of sampling can be applied both in the monitoring process that is used to invoke diagnostic functions, and in the diagnostic process itself.

For example, if a contact centre were to change their headsets to models that were incorrectly matched to the agents' terminals, the conversational voice quality would be reduced due to sub-optimal voice levels. Such a systemic problem could be rapidly detected by analyzing a few hundred calls made to the contact centre, despite the fact that many thousands of calls could have been made in the same time interval. Further diagnostics, for example analysis of factors that contribute to conversational quality such as delay, echo, noise level and voice level, could also be based on a sub-sample of calls.

### 9.3 Mapping basic diagnostic parameters into root cause analysis

The following factors should be considered when mapping between basic diagnostic parameters and the possible root cause of a problem:

- What problems have been seen before and the likelihood of them occurring in the system being monitored
- Which diagnostic values are sensitive to a given underlying problem, and which are not
- The direction of problem voice flows and whether the equipment at the source of the flow is likely to generate a particular problem
- What proportion of flows are likely to be affected
- Equipment or configuration that is common to flows that have violated a particular threshold

Table I.2 in Appendix I illustrates how [ITU-T P.561] and [ITU-T P.562] diagnostic parameters can be associated with common voice quality problems. For example, incorrectly set pads at interconnect points or handsets incorrectly matched to terminal devices are both common causes of very high or very low voice levels. For a given set of flows with an abnormal voice level, the likelihood of these two candidate root causes can be further separated by determining whether the source of the traffic is a gateway or a particular group of phones.

The ability to group sets of measurements by common attributes can greatly assist in identifying the underlying root cause of a problem. This could be the type of equipment, source of configuration or physical location, depending on the particular root cause. For example, the ability to isolate all traffic flowing to and from a PSTN gateway or session border controller will greatly assist in detecting the root cause of common problems associated with these devices.

Finally, it is important to recognize that the exact relationship between diagnostic parameters and possible root causes will often depend on the exact design and configuration of the voice service being monitored.

# Appendix I

## Example based on call clarity index

(This appendix does not form an integral part of this Recommendation.)

### Introduction to call clarity index

The call clarity index (CCI) specified in Annex A of [ITU-T P.562] produces a MOS-CQO estimate of the conversational quality experienced at each end of a voice connection. The CCI output values are based on measurements of the voice level and the noise level in each direction, the echo return loss (ERL) at both ends of the connection and the roundtrip delay between the two ends. The inputs to CCI are taken from an in-line non-intrusive measurement device (INMD) conforming to [ITU-T P.561].

### Decision metric

The percentage of calls with a CCI of $Y_C^A$ and $Y_C^B$ exceeding a MOS-CQO threshold.

NOTE 1 – $Y_C^A$ denotes conversational speech quality as perceived from end A.

NOTE 2 – $Y_C^B$ denotes conversational speech quality as perceived from end B.

### Thresholds and alerting strategies

In common with any alerting strategy involving live traffic, alerting based on a threshold violation by a single call is not recommended when monitoring traffic using CCI [ITU-T P.562]. Instead, it is generally more appropriate to raise an alert when a certain proportion of calls have violated a MOS-CQO threshold.

Some factors that reduce conversational quality are outside the control of the service provider, for example if a user makes a call from a noisy location; however, such events are unlikely to affect all calls, whereas systematic problems will, by definition, impact a large proportion of calls handled by the problematic equipment or link.

### Diagnostic parameters

The [ITU-T P.561] inputs to the CCI algorithm are all useful diagnostic parameters and are listed in Table I.1.

**Table I.1 – Diagnostic parameters associated with ITU-T P.562 call clarity index**

| Name | Description | Units |
|---|---|---|
| $SL^A$ | INMD measured active speech level in direction A → B | dBm0 |
| $SL^B$ | INMD measured active speech level in direction B → A | dBm0 |
| $NL^A$ | INMD measured psophometric noise level in direction A → B | dBm0p |
| $NL^B$ | INMD measured psophometric noise level in direction B → A | dBm0p |
| $EL^A$ | INMD measured echo path loss for echo path A → B → A | dB |
| $EL^B$ | INMD measured echo path loss for echo path B → A → B | dB |
| $ED^A$ | INMD measured echo path delay for echo path A → B → A | ms |
| $ED^B$ | INMD measured echo path delay for echo path B → A → B | ms |

### Interpretation and possible root causes

A key factor in isolating the root cause of voice quality problems is the ability to analyse calls based on their source and destination. In particular, information about the route a call has taken before arriving at monitoring point that has raised an alert can be essential in diagnosing the root cause of the alert.

If a particular route displays a large proportion of calls with a low CCI score, then further analysis is recommended by looking at the [ITU-T P.561] values associated with each CCI score. Note that low CCI scores will often be caused by a combination of factors, e.g., the presence of both echo and delay, or the combination of low voice levels and high noise levels.

Some common causes of low a CCI score and the associated diagnostic parameters are listed in Table I.2.

**Table I.2 – Diagnosing low call clarity index values**

| Diagnostic | Example values | Possible causes |
|---|---|---|
| Presence of echo with a delay exceeding a few tens of milliseconds. | EL < 40 dB and<br>$ED^A + ED^B > 40$ ms | • Absence of echo cancellation or suppression equipment.<br>• Echo control equipment erroneously disabled.<br>• Echo tail length in echo canceller too short to echo path.<br>• Acoustic echo from devices. |
| Large round-trip delay. | $ED^A + ED^B > 500$ ms | • Excessive jitter on VoIP networks.<br>• Unexpected call routing. |
| Very high or very low voice level. | SL < −32 dBm0 or<br>SL > −8 dBm0 | • Incorrectly set pads at interconnect points or gateways.<br>• Handsets incorrectly matched to terminal devices. |
| Low signal-to-noise ratio due to low voice levels or high noise levels. | SL < −32 dBm0 or<br>NL > −50 dBm0p | • Large proportion of calls made from a noisy environment.<br>• Also see low voice level. |

# Bibliography


[b-IETF RFC 3577] IETF RFC 3577 (2003), *Introduction to the Remote Monitoring (RMON) Family of MIB Modules*.

[b-IETF RFC 3611] IETF RFC 3611 (2003), *RTP Control Protocol Extended Reports (RTCP XR)*.

[b-IETF RFC 3919] IETF RFC 3919 (2004), *Remote Network Monitoring (RMON) Protocol Identifiers for IPv6 and Multi Protocol Label Switching (MPLS)*.

[b-IETF RFC 5101] IETF RFC 5101 (2008), *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information*.

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | General tariff principles |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| **Series G** | **Transmission systems and media, digital systems and networks** |
| Series H | Audiovisual and multimedia systems |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Terminals and subjective and objective assessment methods |
| Series Q | Switching and signalling |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects and next-generation networks |
| Series Z | Languages and general software aspects for telecommunication systems |

Printed in Switzerland
Geneva, 2014