ETSI EG 201 377-3 V1.1.1 (2003-06)

ETSI Guide

Speech Processing, Transmission and Quality Aspects (STQ); Specification and measurement of speech transmission quality; Part 3: Non-intrusive objective measurement methods applicable to networks and links with classes of services

Reference

DEG/STQ-00026

Keywords

interworking, quality, speech, testing, transmission, QoS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, send your comment to:
editor@etsi.org

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Guide (EG) has been produced by ETSI Technical Committee Speech Processing, Transmission and Quality Aspects (STQ).

The present document is part 3 of a multi-part deliverable covering the specification and measurement of speech transmission quality, as identified below:

Part 1: "Introduction to objective comparison measurement methods for one-way speech quality across networks";

Part 2: "Mouth-to-ear speech transmission quality including terminals";

Part 3: "Non-intrusive objective measurement methods applicable to networks and links with classes of services".

1 Scope

The present document is part 3 of a series of documents on the specification and measurement of mouth-to-ear (also end-to-end) speech transmission quality. Its main objective is to describe objective non-intrusive methods and systems for measuring speech quality in networks and to present the links between the results of such methods and the perceived quality of voice services.

The present document gives an overview of the non-intrusive methods available for measuring speech transmission quality. It applies only for narrowband (i.e. between 300 Hz and 3 400 Hz) communications, assumed between handset terminals. Its purpose is to give information and guidance primarily for operators, users, consumer organizations and regulators who wish to measure or compare the speech transmission quality provided by different networks.

The present document applies to both fixed and mobile, circuit switched or packet-based networks (including IP and ATM based networks), i.e. no specific measurement interface is needed to apply those measurement methods on networks.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication and/or edition number or version number) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

[1] ETSI ETR 250 (1996): "Transmission and Multiplexing (TM); Speech communication quality from mouth to ear for 3,1 kHz handset telephony across networks".

[2] ETSI EG 201 050 (1999): "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network".

[3] ETSI TS 101 329-5 (2002): "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 5: Quality of Service (QoS) measurement methodologies".

[4] ITU-T Recommendation E.164 (1997): "The international public telecommunication numbering plan".

[5] ITU-T Recommendation G.107 (2002): "The E-Model, a computational model for use in transmission planning".

[6] ITU-T Recommendation G.108 (1999): "Application of the E-model: A planning guide".

[7] ITU-T Recommendation G.109 (1999): "Definition of categories of speech transmission quality".

[8] ITU-T Recommendation G.122 (1993): "Influence of national systems on stability and talker echo in international connections".

[9] ITU-T Recommendation G.212 (1988): "Hypothetical reference circuits for analogue systems".

[10] ITU-T Recommendation G.223 (1988): "Assumptions for the calculation of noise on hypothetical reference circuits for telephony".

[11] ITU-T Recommendation G.763 (1998): "Digital circuit multiplication equipment using G.726 ADPCM and digital speech interpolation".

[12] ITU-T Recommendation O.41 (1994): "Psophometer for use on telephone-type circuits".

[13] ITU-T Recommendation P.56 (1993): "Objective measurement of active speech level".

[14] ITU-T Recommendation P.561 (2002): "In-service non-intrusive measurement device - Voice service measurements".

[15] ITU-T Recommendation P.562 (2000): "Analysis and interpretation of INMD voice-services measurements".

[16] ITU-T Recommendation P.862 (2001): "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs".

[17] ITU-T Recommendation Y.1540 (2002): "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[18] IEEE 743: "IEEE Standard Equipment Requirements and Measurement Techniques for Analog Transmission Parameters for Telecommunications".

[19] IETF RFC 1889: "RTP: A Transport Protocol for Real-Time Applications" Audio-Video Transport Working Group, H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson. January 1996.

[20] IETF RFC 2678: "IPPM Metrics for Measuring Connectivity" J. Mahdavi, V. Paxson. September 1999.

[21] IETF RFC 2680: "A One-way Packet Loss Metric for IPPM" G. Almes, S. Kalidindi, M. Zekauskas. September 1999.

[22] ETSI EG 201 377-1: "Speech Processing, Transmission and Quality Aspects (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

bark: frequency unit in the perceptual domain; e.g. frequencies at 3, 4 and 5 Bark are perceived as equally-spaced

cepstrum: cepstrum of a signal is defined as the inverse Fourier transform of the logarithm of the power spectrum of that signal

NOTE: Linear distortions of a signal (e.g. delay, echo) are additive in the cepstral domain.

cognitive: pertaining to higher layers of human reception; e.g. interpretation of speech

perceptual: pertaining to lower layers of human reception; e.g. processing of sound signals

psycho-acoustic: pertaining to acoustic processing particular to the human sound perception system; e.g. masking of adjacent frequency components

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AMR	Adaptive Multi Rate codec
ANSI	American National Standards Institute
CCI	Call Clarity Index
DC	Direct Current
DCME	Digital Circuit Multiplication Equipment
DTX	Discontinuous Transmission
EFR	Enhanced Full Rate
ETR	ETSI Technical Report
ETSI	European Telecommunications Standards Institute
FR	Full Rate
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile communications (originally French - Groupe Spécial Mobile)
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
INMD	In-service Non-intrusive Measurement Device
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ITU(-T)	International Telecommunications Union (Telecommunication standardization Sector)
MOS	Mean Opinion Score
NINA	Non-Intrusive Network Assessment
NiQA	Non-intrusive Quality Assessment
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PSOM	Perceptual Single-ended Objective Measure
PSTN	Public Switched Telephone Network
RFC	Request For Comment
RLR	Receiving Loudness Rating
RR	Receiver Report
RRM	Radio Resource Management
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
SLA	Service Level Agreement
SLR	Sending Loudness Rating
SR	Sender Report
TDM	Time Division Multiplexing
TS	ETSI Technical Specification
UDP	User Datagram Protocol
VAD	Voice Activity Detector
VoIP	Voice over the Internet Protocol
VTQM-E	Voice Transmission Quality from Mouth to Ear

4 Quick definition and overview of existing methods

4.1 Introduction

In order to measure speech transmission quality on a regular basis, it is necessary to avoid the complicated and expensive procedure of subjective determination, and to use objective systems instead. Today there are several systems and methods in use, some of which are in a rather experimental state, while others are commercially available products. The following non-exhaustive list shows some situations related to transmission over networks where such objective measurement methods are being applied:

- *Mobile Communications:* In mobile communication systems (e.g. GSM), speech quality measurement campaigns can unveil coverage problems, base station failures (e.g. handover problems), etc.

- *Speech compression devices:* In networks with speech compression devices (e.g. codecs, DCME), speech quality can be severely impacted due to the interaction of such devices with each other and with effects like noise, echoes, etc. Monitoring mouth-to-ear speech quality can detect such problems.

- *Voice over IP:* The characteristics of packet-based networks (including IP) are different from switched networks. For voice over IP, the most critical parameters are delay and the degree of packet loss. Monitoring speech quality can help to improve service quality. In order to do so, however, objective methods shall be able to cope with such impairments.

- *Cascade of networks and/or analogue interfaces:* In today's liberalized environment, it is increasingly likely that a call is routed through several networks. On its way, the speech signal could be compressed and expanded repeatedly (tandeming of DCME pairs), or undergo several A/D and D/A conversions. The impact of such cascading on speech quality is virtually impossible to predict, but can be assessed by mouth-to-ear speech quality measurements.

- *Private networks (e.g. corporate networks, closed user groups) interconnected with the public ISDN/PSTN*: After the pre-installation transmission planning (e.g. according to EG 201 050 [2]), it might be highly desirable to measure and monitor the resulting speech quality in order to obtain feedback on the quality and reliability of the transmission planning process. This monitoring can also be used to report data needed to verify whether the terms of contractual Service Level Agreements (SLA) are respected.

Although all assessments of overall speech quality are ultimately subjective because they depend on the user's opinion, a distinction is made between:

- subjective methods, which involve real time user assessment; and

- objective methods, which use stored information on the user's assessment and therefore involve some degree of calibration.

Objective methods for the evaluation of speech quality fall into three categories:

a) *Comparison Methods:* Methods based on the comparison of transmitted speech signal and a known reference; e.g. PESQ (ITU-T Recommendation P.862 [16]).

b) *Absolute Estimation Methods:* Methods based on the absolute estimation of the speech quality (i.e. there is no known reference signal); e.g. INMD (ITU-T Recommendation P.561 [14]).

c) Transmission Rating Models: Methods that derive a value for the expected speech quality from knowledge about the network; e.g. ETSI Model (ETR 250 [1], ITU-T Recommendation G.107 [5]).

The classification of assessment methods is depicted in figure 1.

Practical implementations of test equipment may include combinations of these methods. The focus of the present document is on absolute estimation methods (non-intrusive methods). The other categories are only covered in short overviews (for further detail, see EG 201 377-1 [22]), although they may be preferable for certain applications.

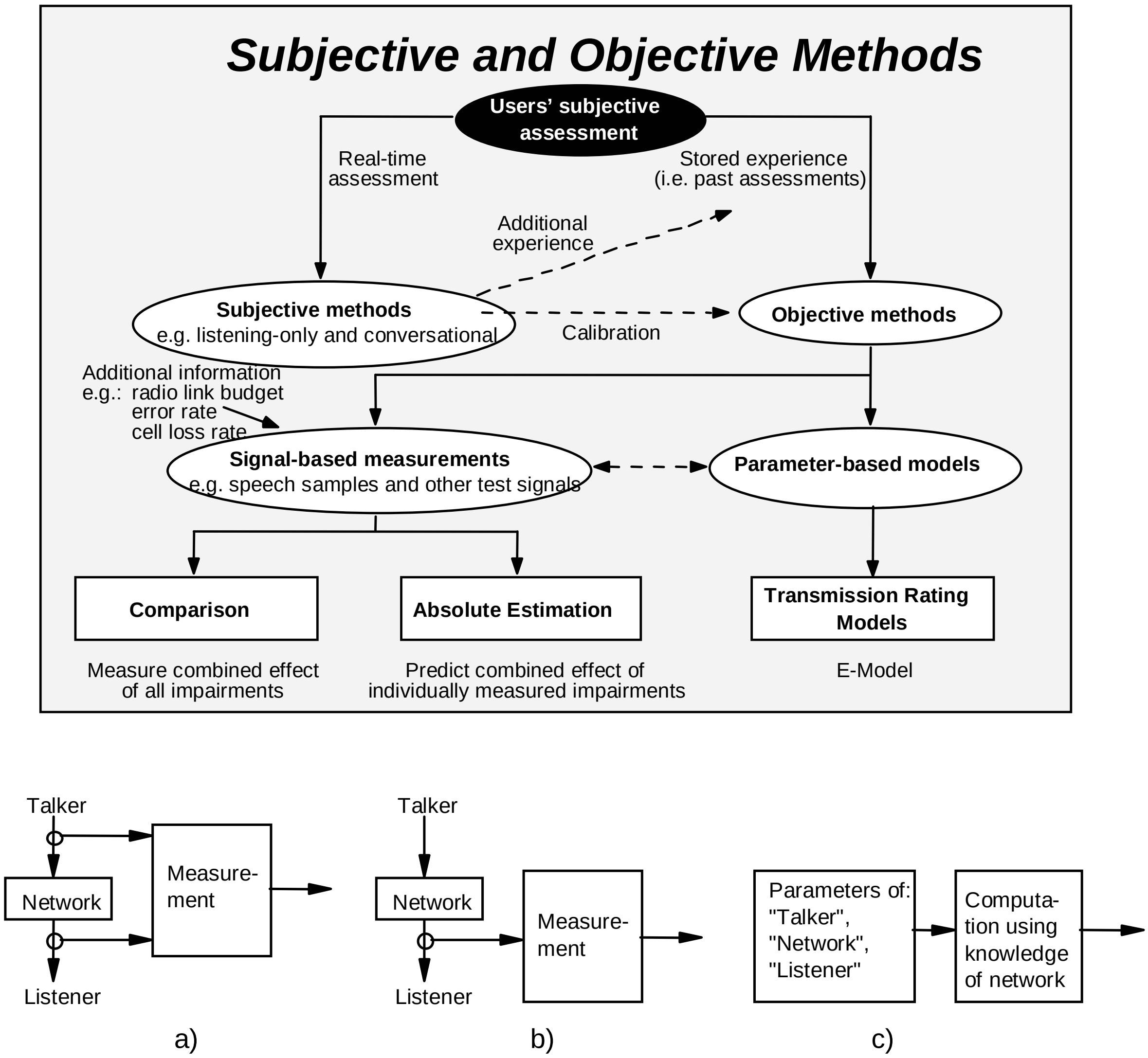


Figure 1: Classification of measurement methods showing: a) Comparison methods, b) Absolute estimation methods, c) Transmission rating models

4.2 Non-intrusive methods

Non-intrusive methods analyse signals (from real or test calls) without a known reference, in contrast to the intrusive methods described in clause 4.3. They belong to the class of absolute estimation methods (see figure 1b) The most known systems for non-intrusive measurement are the in-service, non-intrusive measurement devices (or INMDs) described in ITU-T Recommendation P.561 [14] and dedicated to the supervision of E1/T1 trunks (classes A, B and C) as well as IP links (class D).

ITU-T Recommendation P.561 [14] requires for INMDs the measurement of the following parameters:

Speech and noise characterization:

- active speech level: average signal level during active (non-silent) intervals of a speech connection;
- noise level (psophometric weighted): average signal level during speech pauses, weighted to account for psychoacoustic perception;
- speech activity factor: ratio of active speech time to total elapsed time.

Echo characterization:

- speech echo path delay (single or multiple reflection measurement): time delay of echo in received signal;
- and either echo loss (single or multiple reflection measurement): frequency-weighted attenuation of echo path;
- or echo path loss (single or multiple reflection measurement): unweighted attenuation of echo path;
- or speech echo path loss (single or multiple reflection measurement): unweighted attenuation of speech echo path.

For the INMDs of class D, extra measurement functions related to the analysis of the Real-time Transport Protocol (RTP) are required:

- 1-point IP packet delay variation;
- IP packet loss ratio.

But non-intrusive methods can be used to measure or evaluate several other parameters (related to the voice signal itself or to the analysis of the transport signalling) whose list is given in clause 6.

New, recently developed, algorithms give even the possibility for non-intrusive systems to evaluate the perceived speech quality, using psycho-acoustical models without a reference signal. Within ITU-T SG12/Q.9 and 16, the suitability of those approaches is currently investigated, and a standardization process is about to begin with an invitation for model submission for developing a future ITU-T Recommendation on single-ended (i.e. without reference signal) quality assessment algorithms. Some of those algorithms are described in annex A to the present document.

4.3 Intrusive methods

Intrusive speech transmission quality measurement methods consist in establishing test calls between two measurement units or probes, to send tests signals from one probe to the other and to analyse the received signals knowing the original input signal. Several parameters related to speech transmission quality can be evaluated by this means, like transmission delay (one way or round trip), echo, attenuation, frequency shifting, etc. And of course speech quality itself, generally expressed on a 5 level mean opinion scale (from 1: bad, to 5: excellent).

The measurement methods (models) for speech quality use two signals as their input, namely an original signal (reference pattern) and the corresponding output signal after its transition through the network or system under test.

The signal processing within objective methods based on the comparison of speech samples can be structured into three major steps as follows:

- pre-processing;
- psycho-acoustic modelling;
- speech quality estimation model.

For details on the building blocks of those methods, see EG 201 377-1 [22].

The current international standard for objective one-way speech quality evaluation is the PESQ (for Perceptual Evaluation of Speech Quality) model, standardized in 2001 by the ITU-T under Recommendation P.862 [16].

New studies on end-to-end speech quality assessment methods are going on and will soon lead to new generations of algorithms, able to cope with specific aspects not covered by PESQ:

- objective mouth to ear speech quality assessment including terminals;
- objective speech quality assessment for wideband speech;
- objective speech quality assessment for talker dependencies (including the effect of echo and of transmission delay).

5 Comparison of non-intrusive methods with intrusive methods

The number of parameters that both families of measurement methods can assess is very large, but they do not address the same part of a link:

- the intrusive techniques are better for the assessment of parameters directly linked to the perception of speech quality by the users;
- whereas the non-intrusive techniques are better adapted for the assessment of the network quality of service.

The intrusive methods make use of their test signals that are transmitted through the channel under test. Thanks to having access to both the input and output signal, the intrusive methods can typically achieve very high correlation with subjective tests. The drawback, however, is the usage of network resources for the assessment, which leads to a compromise between a limitation of the cost of those occupied resources (to avoid also possible congestions) and the need for a minimal number of measurements.

This kind of measurement is simple and well suited for assessment with access on both terminations of a link, but it is difficult to implement when competitors' networks are interconnected.

Non-intrusive techniques monitor live traffic, without modifying or delaying the transmitted signals, to determine quality received by the customer. This is a positive approach to large scale monitoring as it does not use network capacity. A key benefit of the non-intrusive methods is their ability to monitor large numbers of customers at low operational cost.

This simplifies the measurement process because measurements can be performed in-service during ordinary calls, thus eliminating the need for specific test calls and for equipment at the talker side. In turn, this allows a large number of measurements to be performed, and thus to have a much better statistical view of the quality than with intrusive methods.

Accurate quality assessment by the non-intrusive approach presents a far more complex picture than the intrusive methods since one makes use only of the output signal: no special or dedicated test signal is required i.e. the input signal can be everything (time answering machine, news, human conversation, etc).

In counterpart, since the measured signal is from real traffic, the correlation with the perception of the real users involved in the real traffic monitored is better than with an intrusive method, where the correlation is better with the mean opinion of all users.

Another advantage of non-intrusive techniques is that they give to network operators the opportunity to monitor not only the quality of service in their network, but also of their interconnection with other networks, without being obliged to interfere with these networks (no probe installed, no traffic generated).

Developments of new non-intrusive soft agents (including single-ended quality assessment algorithms) that can be implemented in terminals makes the non-intrusive measurement techniques more representative of the user experience.

As far as the measurement of echo is concerned, both techniques have drawbacks:

- intrusive methods have difficulties to measure with accuracy short echo delays with analogue interfaces;
- non-intrusive methods do not measure the end to end delay and attenuation.

A summary of the relative advantages and drawbacks of both techniques exposed above is given in table 1.

Table 1

	Intrusive techniques	Non-intrusive techniques
Best fitted for the assessment of	Parameters directly linked to the perception of speech quality by the users	Network quality of service
Correlation with the subjective perception	High	Medium
Use of test signals	Mandatory	No, but non-intrusive methods can be applied to test calls
Affects network loading	Yes	No
Interfaces	End-user interface (probes), electrical or acoustical	Network node (probes) or inside terminals (agents)
Supervision of the interconnection with a competitor's network	Needs an access to a subscriber line of this competitor	Supervision at the point of interconnection
End-users supervised with 1 probe	One probe per line tested	One INMD card per T1/E1 transmission unit
Measure of echo	Yes (difficult for short delays on 2-wire analogue interfaces)	Yes, because they are always at a 4-wire point
Measure of end to end one-way delay possible if	The two measuring-points are often not at the same position. In this case a time-synchronisation at the measuring-points is needed. If they can be reached with a well known line with better quality and the same delay in both directions, e.g. ISDN, it is possible to make a "loop-delay-reference-measurement" of this back-line. Afterward one can subtract the one-way-delay of this back-line. This method is used and protected by Deutsche Telekom. Otherwise, the probes need to be synchronised with an external clock, e.g. with GPS.	Only possible if there is detectable echo from the end of the connection from which the delay can be estimated. However echo from the end of the connection may be confused with echo from other sources in other parts of the connection

6 Parameters measured in a non-intrusive way and linked to quality of service in networks

In this clause, one will find a list of the most important information that can be measured in a non-intrusive way, with precisions on how they are generally assessed (measurement location, measurement unit and range, link with other parameters and with subjective perception).

6.1 Parameters related to transport protocol or signalling

6.1.1 Originating and terminating address digits (ITU-T Recommendation E.164 and/or IP)

For TDM INMDs (i.e. from classes A, B or C of ITU-T Recommendation P.561 [14]), the originating and terminating address (ITU-T Recommendation E.164 format [4]) are decoded from signalling information associated with the channel being observed or via data interface to the switch processor.

For Ethernet access INMDs (i.e. from class D of ITU-T Recommendation P.561 [14]) or IP probes, not only the ITU-T Recommendation E.164 [4] digits are available, but also the IP addresses and port numbers of the ends of the IP part of the path. In the case of an IP transit between PSTN networks, the IP address will be the same for all calls (i.e. the one of the gateway). This information can be decoded from the analysis of the signalling protocol packets in the headers of the RTP packets.

It is generally recommended that the measurement devices give all the digits of the phone numbers. But for security and privacy reasons, they can also only give the first ones (concerning the country, or the area).

This information is useful for identifying performance with a particular end user, access line, local switching office, or network route.

6.1.2 Facility or circuit identification

For facility access INMDs, the facility or circuit identification is usually the system group or digroup and channel codes. The INMDs can be primed with this information at the time of bridged access. For switch access INMDs, the circuit termination information is obtained via data interface to the switch processor. This information can also be completed by identification of the PCM time slot of the communication.

This information is useful for identifying performance with a particular network element or circuit.

6.1.3 Time and duration of connection and of call

The time of connection is usually the time at which the measured connection or stream is placed in the active state. The duration of the connection is usually the interval between the connect and disconnect times. This information is useful as traffic data.

The duration of the call is usually the interval between the time when both ends are off hook and the time when one end puts down the receiver.

Times and durations are generally expressed in seconds, but some devices can go down to an accuracy of 1 ms.

Additionally, correlation on connections with short duration and poor performance can indicate severe network failures.

6.1.4 Customer identification

For dedicated circuits, the end user or inter-exchange customer identification is usually collected by INMDs at the time of bridged access. This information may be useful for reporting customer performance as well as sharing performance data with the customer.

6.1.5 Connection state

These measurements may be useful for determining circuit connection performance and points of abnormally high connection failures. Connection state could include, for example, "answered", "busy", and "ring no answer".

INMDs generally give also useful information on call setup performance, like the post dialling delay.

6.1.6 Data analysis and reports

Data analysis and reports are useful for network characterization as well as network management and maintenance reports.

6.1.7 Jitter

This is a specific characteristics of IP networks, that can not be obtained with traditional TDM INMDs. The jitter corresponds to the variation of the transmission time of RTP/UDP/IP packets and is generally understood as the mean deviation from the mean transmission time (but the measurement devices can also give information on the maximum deviation or statistics on the 95 % measurements closest to the mean delay, etc.). It can be measured:

- either from a derivation of the measurement of the intervals between the arrival times of the packets (one-point jitter, called 1-point IP packet delay variation in ITU-T Recommendation P.561 [14]);
- or by measuring the transmission delay of each packet (two-point jitter, requires synchronization between the sending and receiving sides).

If the measurement is performed inside an IP terminal (a gateway, or an IP-Phone, etc.), it will occur before the synchronization process run by the de-jitter buffer of the terminal.

Jitter is expressed in milliseconds.

The one-point jitter method will give accurate results for network performance only if the packets are sent without jitter at the interface into the network. This may be true for dedicated terminals but will not be true for applications running on PCs.

6.1.8 Packet loss and out of order packet rates

IP packet loss ratio, as defined in ITU-T Recommendation Y.1540 [17], is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest. It corresponds to the proportion (expressed in %) of speech frames within RTP packets that did not reach the point where the measurement is performed.

It can also be called one-way packet loss (see IETF RFC 2680 [21]).

To measure the packet loss due to congestions or TTL (time to live) expirations of packets, the packets must be analysed at the input of the device where the measurement is performed. If the goal is to assess the overall packet loss, caused by both the network and the device (packets can be discarded by a jitter buffer, for instance), the packets must be analysed just before being decoded.

In the specific case of a measurement at an IP terminal level, another possibility to estimate the overall packet loss rate is to measure the proportion of packets arriving out of order and their individual extra-delay (or jitter). Together with the characteristics (depth and dynamic) of the jitter buffer of the terminal, this information is potentially enough to know how many packets will be discarded by the buffer.

Packet loss rate can be measured as a mean value over a certain duration or on a whole conversation. It can be also interesting to have supplementary statistics on the distribution of the packet loss.

The Internet exhibits certain specific types of behaviour (e.g. bursty packet loss) that can affect the performance seen by the users as well as the operators. The loss pattern or loss distribution is a key parameter that determines the performance observed by the users for certain real-time applications such as packet voice and video. For the same loss rate, two different loss distributions could potentially produce widely different perceptions of performance. Using the base loss metric defined in IETF RFC 2680 [21], the IETF is working on the definition for two derived metrics "loss distance" and "loss period", and the associated statistics that together capture loss patterns experienced.

Example of such statistics are: number of time periods (e.g. 1 s long) with a packet loss rate higher than x times the overall mean value, number and location in time of bursts of more than x consecutively lost packets, etc.

As a complement to the IP packet loss ratio, statistics on the order of arrival of IP packets can be useful to estimate the quality of a link.

6.1.9 RTCP round trip delay

The round-trip delay from a given source to a given destination corresponds to the following situation: the source sent the first bit of a packet to destination at a given time T, the destination received that packet, then immediately sent a packet of the same type back to the source, and the source received the last bit of that packet at a time T+dT. DT corresponds to the round trip delay.

If the IP terminals involved in the communications monitored by a non-intrusive measurement device use the RTCP protocol to share information on the quality of the link, this information can be used for the measurement of the round trip delay, which is the duration between the time when a terminal sends a RTCP ticket (SR) and the time when it receives an answer from the far end terminal (RR). Packet loss and jitter can be partially assessed (see 6.1.7 and 6.1.8) this way.

There are two potential problems:

- the transmission delays in IP networks are not symmetric. So it is not correct to consider that the one-way delay in any direction is equal to half of the round trip delay;
- the delay measured is the delay associated with the control protocol and this is not necessarily the same as the delay of the voice path.

Beware: the transmission delays in IP networks are not symmetric. So it is not exact to consider that the one-way delay in any direction is equal to half of the round trip delay.

6.1.10 Type and configuration of codec

The signalling protocols on IP communications carry information for the decoding process at the output of the IP part of the transit, amongst which:

- the type of voice coder (generally negotiated before the connection, but likely to change during, and so coded in all RTP packet headers): standard name and bit rate;
- whether a silence detection (or a VAD) has been implemented during the coding and requires a comfort noise generation at the decoding.

These parameters (and their potential changes in time) should be given for each communication in both transmission directions (since, in packet networks, it is possible to use different coders in the two directions of a communication).

6.1.11 Quantity of transmitted data

In the case of communications on IP using a silence detection scheme at the input of the IP part of the path, the data rate will vary in time since information will be transmitted only during active speech periods. The measurement device can count the number of voice frames (or bytes) received on both directions and report it. The ratio between this value and the theoretical amount of data sent during the call at the nominal coding rate will give a good estimate of the speech activity factor without needing to analyse the voice signal itself (see 6.2.3).

6.1.12 IP one-way connectivity

Connectivity is the basic stuff from which the Internet is made. Therefore, metrics determining whether pairs of hosts (IP addresses) can reach each other must form the base of a measurement suite. Several such metrics are defined in IETF RFC 2678 [20], some of which serve mainly as building blocks for the others.

6.1.13 Packet arrival descriptors

To facilitate other descriptions of IP performance, some low level information can be provided by INMDs. This information is required for some of the measures described above and therefore does not introduce any further processing requirements. The information provided includes:

- Packet arrival time to 1ms resolution;
- RTP sequence number as defined in IETF RFC 1889 [19];
- RTP Time stamp as defined in IETF RFC 1889 [19].

6.2 Parameters related to the signal analysis

6.2.1 Signal classification

The service classification (data, voice, etc.) can be determined. Also, more specific information such as data baud rate, analogue versus digital, signal format, etc. might be detectable by signal processing and pattern recognition techniques. This information is useful as traffic data. Furthermore, this classification will be necessary for measurement of connection state and data parameters.

Signals that can be present on a circuit are speech, data, fax, idle circuit, music, reorder, ring back, no circuit tones, etc.

The requirements for speech classification in INMDs are similar to those of DCMEs. ITU-T Recommendation G.763 [11] has recommendations for data/speech classification for DCMEs. INMD equipment compliant with ITU-T Recommendation P.561 [14] should adhere to these recommendations.

6.2.2 Active speech level

The active speech level is the root mean square (r.m.s.) of the speech spurt amplitudes, and is expressed in dBm. The active speech level shall be the speech level averaged over speech spurt intervals including any hangover time and does not include measurements during conversational pauses. By including hangover time, any inter syllabic pauses are included in the r.m.s. value. Typically, these measurements are estimated by sampling or integrating the active speech signal. This definition is consistent with active speech level as defined in ITU- T Recommendation P.56 [13].

The speech level measurement interval must be long enough to predict an accurate active speech level within one dB of a measurement made over the entire call duration. The measurement interval should include a minimum of 20 seconds of active speech.

6.2.3 Speech activity factor

Speech activity factor is the ratio of active time to total time elapsed during a measurement, usually expressed as a percentage. The active time is the aggregate of all intervals of time when speech is deemed to be present.

This parameter is very useful in helping to interpret other measurements.

6.2.4 Noise level (psophometric weighting)

The psophometric weighting is specified by ITU-T Recommendation G.223 [10].

The noise level is the average root mean square (r.m.s.) of the weighted noise amplitude expressed in dBmp (ITU-T Recommendation G.212 [9]). The noise level is the psophometric weighted noise level averaged over speech pause intervals excluding any hangover time. By excluding hangover time, r.m.s. noise measurements are not corrupted by any residual speech. Level averaging for noise should comply with ITU-T Recommendation O.41 [12].

The noise level measurement interval shall be long enough to predict an active noise level within one dB of a measurement made over the entire call duration. It is currently assumed that typically, this measurement interval is one minute or longer, although this needs to be verified.

6.2.5 Noise classification

The noise signal (psophometric weighted) can be analysed to estimate its stationarity. As a result it can be classified as:

- Stationary.
- Hybrid.
- Non-stationary.

A well known example of non-stationary noise is the so called "impulsive" noise.

Indeed, for this kind of noise it is rather difficult to determine a set of parameters which model the non-stationary noise and are well suited to automatic measurements.

It can be added that a correct classification depends on the "quality" of the noise. If a segment contains "mixed" noise, the algorithm could detect either one correct kind of noise or classify depending on the frequencies of the noises and on their levels.

The thresholds suggested (provisionally) in ITU-T Recommendation P.561 [14] to classify "active" noise could be -67 dBmp and; for impulsive noise, -38 dBmp, respectively.

6.2.6 Speech echo path delay

The delay of the echo corresponds to the time (in ms) taken by the original signal at a given measurement point to come back as an echo (i.e. a replica of the original signal, delayed, weakened and potentially distorted) at the same measurement point.

The measurement devices should be able to detect single and/or multiple echo paths in both directions.

6.2.6.1 Single reflections

The speech echo path delay shall be the value from the zero reference at the point of measurement to the peak absolute amplitude of the time domain impulse response of the echo path.

6.2.6.2 Multiple reflections

The speech echo path delay, of any one reflection, shall be the value from the zero reference at the point of measurement to the peak absolute amplitude of the time domain impulse response of the echo path.

A measurement device should be able to differentiate between reflections that are separated by 10 ms or more.

6.2.6.3 Methods of calculating speech echo path delay

There are currently two known measurement techniques for speech echo path delay - correlation and adaptive filter analysis. It is likely that each technique will be appropriate to a different range of values.

6.2.7 Attenuation of the echo signal

The attenuation of the echo signal corresponds to the difference between the levels of the original and of the echo signals. It is expressed in decibels.

ITU-T Recommendation P.561 [14] defines three different relevant parameters and their measurement methods:

- **Echo Loss** (a-b), as defined in ITU-T Recommendation G.122 [8], can be calculated from the frequency response of the echo path using a frequency weighting. This Echo Loss figure has been found to give better agreement with subjective opinion for individual connections than an unweighted echo path loss. However, for large samples of actual connections it has been found that the two methods give very similar means and standard deviations.
- **Echo path loss** is derived from the impulse response of the echo path, without frequency weighting.
- The **speech echo path loss** is the r.m.s. ratio of incident to reflected speech signals. It can only be calculated if the speech echo path delay is known. The speech echo path delay is used to determine when the r.m.s. of the reflected speech signal is calculated.

6.2.7.1 Single reflections

The Echo Loss shall be the integral of the weighted frequency response of the echo path.

The echo path loss shall be the integral of the total impulse response (in the frequency domain) of the echo path. Using Parseval's theorem this is equivalent to mean sum of squares of the impulse response (in the time domain).

6.2.7.2 Multiple reflections

The Echo Loss of any one reflection shall be the integral of the weighted frequency response for that reflection.

The echo path loss of any one reflection shall be the integral of the unweighted frequency response for that reflection.

6.2.8 3 kHz flat noise level

The 3 kHz flat noise level is measured in a way similar to the psophometric noise level in 6.2.4, except that, the psophometric weighting is changed to flat weighting. The noise level is weighted according to the IEEE 743 [18] standard for measuring analogue voice frequency circuits. When digital signal processing techniques are used, any DC component shall be filtered out. The 3 kHz noise measurements are useful for correlating the psophometric noise level to determine the type and spectral content of the noise present on the circuit.

6.2.9 Saturation clipping

The percentage of active speech which has been saturated (amplitude "clipped", causing non-linear distortion) can be detected by looking for occurrence of the maximum positive and negative values of the PCM codes.

6.2.10 Double talk

Double talk is a condition whereby, for whatever reason, one party in a telephone connection starts talking before the other has finished. This is a naturally occurring phenomenon - where one participant wishes to interrupt the flow of speech from the other, but occurs more frequently as the propagation time of the connection increases. Monitoring the occurrence of double talk provides a useful indicator linked to customer perception of performance. A high level of double talk can be caused by a high transmission delay, for instance.

Double talk is generally reported as a percentage of the measurement interval.

6.2.11 Syllable clipping

Syllable clipping occurs when part of a speech burst is missing.

It can be due to an aggressive or failing voice activity detection during speech coding in a DCME or a VoIP gateway, or when the number of active signals temporarily exceeds the number of available channels in busy periods (freeze out).

6.2.12 One-way transmission

Temporary loss of one direction of the transmission has a severe effect on customer perception of quality. By an analysis of activity and noise levels, it may be possible to detect even short durations of one-way transmission. The ability to detect permanent one-way transmission will depend on whether the faulty circuit is monitored when the connection is made. This will depend upon the rate of scanning, since neither party will hold the circuit for very long. One-way transmission can occur in some genuinely connected calls (e.g. when a user is listening to a recording) and the probabilities of this situation will be taken into account in deciding whether to classify apparent one-way transmission as a fault or not.

6.2.13 Crosstalk

Crosstalk on a connection is said to occur when a talker participating in another conversation over a different circuit is heard on the connection. This occurs when there is excessive mutual coupling between circuits. "Near end" crosstalk is where the disturbing talker is located at the same end of the group of circuits as the disturbed listener. "Far end" crosstalk occurs when the disturbing talker and disturbed listener are at opposite ends of the group of circuits. It may be possible to detect crosstalk by an analysis of speech levels and speech statistics but it must be realized that there are distinct problems in differentiating far-end crosstalk from background talkers or from multiple talkers in three-way links.

6.2.14 Stability loss

The risk of the echo loss reaching low values at any frequency in the range 0 kHz to 4 kHz should be as small as practicable. The stability loss is the lowest value of loss in the frequency band to be considered.

6.2.15 Voice Quality and distortion

Voice quality is understood here as the level of acceptance expressed by subjective listeners exposed to a given level of signal distortion (due to several causes like low bit rate coding, comfort noise generation, saturation clipping, etc.). It is generally expressed as a prediction of the received quality on a mean opinion score (MOS) scale from 1 to 5.

The evaluation of voice quality itself by objective methods, once restricted to end-to-end techniques requiring the comparison of a degraded signal to a known reference (see EG 201 377-1 [22]), is now possible using a single-ended technique, i.e. without a reference signal.

Standardization of such a measure is under study in ITU-T Recommendation SG 12/Q.9.

Examples of single-ended psycho-acoustical models are given in annex A to the present document.

6.2.16 Other degradations

Other specific degradations of the voice signal, e.g. metallic Voice, whistling, crackling, can be characterized and their impact on subjective quality assessed in a non-intrusive way.

By analysing the characteristics of a given signal in the time and in the frequency domains, it is possible, using pattern recognition techniques and neural networks, to detect in this signal the presence of some specific events, like metallic voice, whistling, crackling, with a high accuracy. It is also possible to assess the level of subjective annoyance due to these phenomena if a correlation can be found between a subjective score and a combination of values for temporal or spectral characteristics.

7 Guidance on how to use measurement results to derive quality of service ratings

7.1 Use of individual thresholds

INMDs continuously monitor the network and have the potential to generate vast amounts of data.
ITU-T Recommendation P.562 [15] (from which most of the text of this clause has been extracted) explains how to interpret INMDs measurement results (restricted to classes A, B and C for the time being) to predict quality of service. It addresses in particular the use of thresholds on measured parameters: how individual voice-service measurements should be interpreted and how measurements from many calls can be collated.

Each of the measured parameters allow some aspect (or aspects) of network performance to be determined or predicted for that particular call. The parameters measured by an INMD characterize the network connection from each talker to the INMD equipment in that direction only. The network connection in the opposite direction, INMD to listener, is not measured. The only exception to this is the measurement of the echo-path which provides some information about the network connection from the INMD to source of reflected echo (usually the 4-wire to 2-wire hybrid) and back to the INMD. This means that the majority of impairments in the receive path, INMD to listener, cannot be detected by an INMD.

Some aspects of network performance that can be derived from single voice-service measurement parameters are shown in table 2. Any assumptions made in deriving network performance are also listed.

As example, if we take the first line of table 2, the measurement of active speech level can be used to derive a value for Sending Loudness Rating, assuming that the speaker's vocal level lies in a normal range (i.e. no saturation).

Table 2

Parameter Measured with INMD	Aspect of network performance	Assumptions to derive network performance from the measured parameter
Active speech level	Network SLR	Speaker's vocal level
Psophometric noise level	Circuit noise level introduced by the network	Room noise level
Echo loss Echo path loss Speech echo path loss	Operation or presence of echo cancellers Hybrid performance	
Echo path delay	Transmission delay of the connection	Local delay
Speech activity factor	Accuracy of other parameters (e.g. 90 % activity in both directions could mean noise is being classified as speech) Type of call (e.g. recorded message)	Normal conversational habits
Front end clipping	Performance of voice activity detectors (e.g. in DCMEs)	
Saturation clipping	Amplitude clipping and distortion	
Double-talk	Rough indicator of delay of the connection	Normal conversational habits
One-way transmission	Short network outages	Two-way conversation

It must be kept in mind that, the content of table 2, like all data coming from ITU-T Recommendation P.562 [15], concerns only PSTN factors.

When interpreting INMD voice-service measurements call results should ideally be viewed as a set. Investigating single parameters in isolation may give rise to misleading conclusions regarding the quality of the connection. In addition to this, data from a single call is prone to variations in customers' voices and customers' equipment which should be taken into account when considering the data. The examples below show how possible measurements can be misinterpreted.

EXAMPLE 1: Measurements from a call show that the echo path loss is low and the echo-path delay is low. The low echo path loss indicates that there is significant echo present and the call is assumed to be of poor quality.

However, since the echo path delay is also low the user hears the echo only as sidetone, and perceives the call as good quality. This call would be correctly classified when using a customer opinion model such as the E-model [5].

EXAMPLE 2: Measurements from a call show no discernible noise present and average active speech levels. This initially appears to be a good quality call. Inspection of the speech activity factor reveals speech activity of over 90 % in both directions.

The cause of this abnormal speech pattern may be due to high levels of noise being interpreted as speech or high levels of echo being interpreted as speech. In fact this call may be severely degraded and could require investigation.

Collating measurement data helps to reduce variations due to customers' voices and equipment. When collating measurement parameters, consideration should be given to the purpose for which the information is being gathered.

Once data on a single parameter has been gathered for a particular grouping (country, carrier, link), it should be processed using the following procedure:

Stage 1 Exclude all invalid measurements.

Stage 2 A sample size should be used that gives statistically valid results.

Stage 3 Calculate the sample mean and sample standard deviation.

Stage 4 Calculate the percentage of valid measurements that exceed maximum and minimum pre-set threshold values.

Stage 5 Report analysis data.

Table 3 lists recommended threshold values for the most common parameters measured with INMDs.

Table 3

Parameter	Recommended threshold values	
	minimum	**Maximum**
Active speech level	-35 dBm0	-6 dBm0
Psophometric noise level	None	-50 dBmp
Echo path loss	15 dB (see note 1) or 35 dB (see note 2)	None
Echo path delay (round-trip) (see note 3)	None	40 ms (see note 1) or 800 ms (see note 2)
NOTE 1: For connections without echo cancellers. NOTE 2: For connections with echo cancellers. NOTE 3: This is the sum of the near and far end delay values.		

7.2 Use of models

Models attempt to map objective measures of network performance to subjective opinions. A customer opinion model for INMDs should therefore be able to relate the network performance (as represented by the objective measurements such as speech level, echo loss, etc.) to customer perceived performance (represented by an opinion score).

Benefits of using a model to interpret INMD measurements include:

1) The identification of combination effects that are incorrectly classified when using individual measures.

2) Reduction in data volume (a single figure now represents the measured quality compared to many individual measurements).

3) The model encapsulates expert knowledge about the effects of impairments on customer perception.

Several examples of such models are given in this clause.

7.2.1 E-Model

7.2.1.1 The historical model

The E-model has been developed by ETSI (ETR 250 [1]) at the beginning of the 90s. It was chosen some years afterwards by the ITU-T to become an international recommendation known as ITU-T Recommendation G.107 [5].

It is based on the assumption (rather well verified in practice) that each family of quality degradation can be associated with separate impairment factors, and that all these impairment factors are additive on a psychological scale. The sum of the different impairment factors is an overall transmission rating factor called R and is comprised between 0 and 100. For a more detailed description of the E-model, please refer to ITU-T Recommendation G.107 [5].

7.2.1.2 Links with network planning and classes of service

ITU-T Recommendations G.108 [6] and G.109 [7] provide guidance for using subjective ratings from the E-model to do network planning. The guidance provided by these two recommendations could be used to assess the acceptability of the performance of a network or route or to plan changes to a network.

7.2.1.3 Adaption to non-intrusive measurements

The E-model has not been specifically developed to fit non-intrusive measurement results, but annex B of ITU-T Recommendation P.562 [15] provides equations for mapping INMDs parameters to some of the parameters used by the E-model. To get ratings for the performance of the end-to-end loss, noise and echo performance of connections though, the RLR and receive loss of connections must be estimated from network averages because they are not included in the INMD's measurements.

Analyses of the ratings produced using this mapping have shown that they do accurately measure the performance of connections. When ratings derived from multiple INMD measurements from a network are averaged, they can be used to accurately evaluate the performance of the network. This will provide network planners with a useful tool for determining how changes to the loss, noise, echo loss, or delay in their network will affect performance.

Because the E-model can also additively include the impact of other impairments not measured by the INMD on performance, the mapping provided in ITU-T Recommendation P.562 [15] can be used to assess how adding new technologies to a network will affect performance. This technique can be used to determine how adding echo control devices, low bit rate codecs, digital circuit multiplication systems (DCMEs), or other technologies to connections will change performance.

7.2.1.4 Adaption to VoIP

An attempt to adapt the E-model to VoIP is described in annex E of TS 101 329-5 [3], based on an IP Metrics based Voice Quality Monitoring Algorithm called VQMon and developed by Telchemy®. This method, also candidate for the standardization of a future standard for the evaluation of voice quality of VoIP services based on parametric information ongoing under Q.16/12 at the ITU-T, allows for the calculation of specific impairment factors associated with the following degradations occurring in IP networks:

- packet loss and its distribution;
- packet delay variation;
- codec;
- one-way delay.

The individual impairment factors are then combined together, taking account of recency effects, to get an overall factor that may be used as input to the E-Model in order to calculate an R factor for the call. If other parameters required for the E-Model are unavailable then they should be set to their default values.

VQMon has been designed to be applied in both end-points (e.g. IP Phones, Gateways) and mid-stream (e.g. SLA monitoring points).

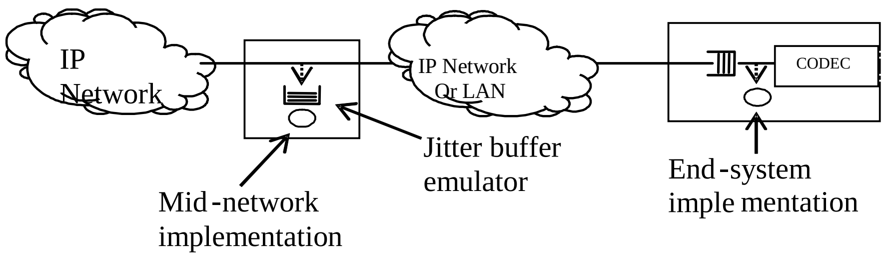


Figure 2: VQMon implementations

VQmon has three elements: a packet loss distribution model, a computation of call quality metrics and a jitter buffer emulator, (required only in mid-network implementations).

7.2.2 CCI (Call Clarity Index)

CCI is described in ITU-T Recommendation P.562 [15] for interpreting INMD measurements. It provides a prediction of call quality, from the live customer traffic, on a call by call basis. CCI takes the INMD measurements and translates these to subjective opinion.

This translation to subjective opinion has been achieved using extensive modelling of cognitive processes. The modelling has been refined by calibrating against subjective tests.

The overall operation of the CCI is to use the non-intrusive measurement parameters in conjunction with assumptions about the network and the users at either end to predict the signals arriving at each user's ear. These predicted signals, along with knowledge of the human auditory system, are then transformed into conversational speech quality opinion predictions of the call as perceived by each user.

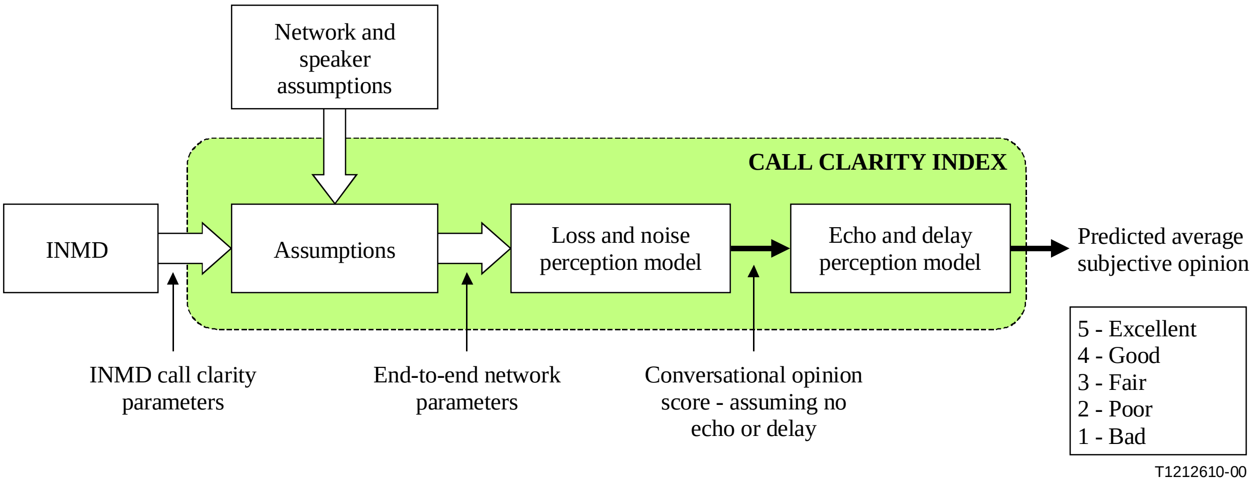


Figure 3: The functional blocks that form the CCI

7.2.3 Other parametric models

PsyVoIP, a development from Psytechnics™ (ITU-T Recommendation COM 12-D49), can non-intrusively measure, from within the IP network, the speech quality delivered by a VoIP network to the end users. PsyVoIP has been designed to provide a fast, accurate monitoring capability anywhere along the VoIP connection from the within a gateway, across the IP network, to any VoIP terminating device.

The structure of psyVoIP is shown in figure 4.

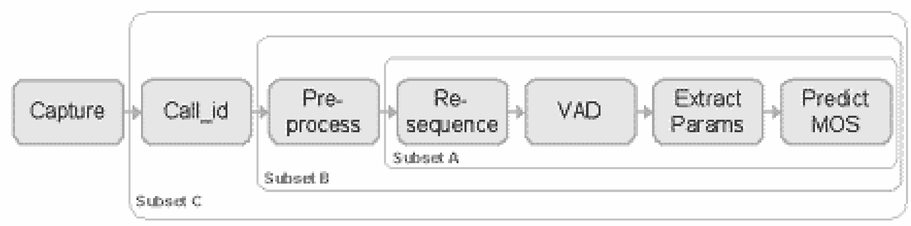


Figure 4: psyVoIP structure

The purpose of each block is as follows:

Capture	captures all packets or captures only relevant packets, such as those belonging to a particular call
Call_id	Identifies if packet belongs to a call and if so, which call
Pre process	extracts information required by rest of model so packet can be discarded
Re sequence	Accounts for out of sequence packets
VAD	enables psyVoIP to distinguish between speech and silence intervals during a conversation. Packet loss and jitter have less effect on quality during silence
Extract Params	extracts the statistical descriptors required to predict MOS
Predict MOS	when enough packets have arrived to make a quality prediction valid, this performs its MOS calculation

ETSI is not aware of the existence of any other model of this kind. PsyVoIp is, together with VQMon from Telchemy® (see clause 7.2.1) one of the two candidates for the standardization of a future standard for the evaluation of voice quality of VoIP services based on parametric information ongoing under Q.16/12 at the ITU-T.

7.2.4 Impact of the duration and location of speech degradations

Considering subjective voice quality, when listening to a call, one must take into account several aspects: the annoyance caused by the type and the duration of impairments, as well as their frequencies (of each individual kind and of all kinds of impairments) and their temporal locations in the call. The impact of the impairment location is known as recency effect; it means that the global judgment of the listeners at the end of a call depends on the duration of the call as well as on the temporal locations of the impairments (the later, the worse the global judgment will be). Although all these considerations seem trivial, it is not obvious how to measure voice quality by taking them into account, with no reference signal.

7.3 Miscellaneous specific usages of non-intrusive measurements

7.3.1 Detection and characterization of specific defaults

This approach has been studied by France Telecom R&D, and a good summary of the subjective methods followed for this study can be found in ITU-T Recommendation COM 12-23.

Figure 5 shows the different steps necessary to derive measurement methods for the detection and characterization of specific defaults from subjective knowledge.

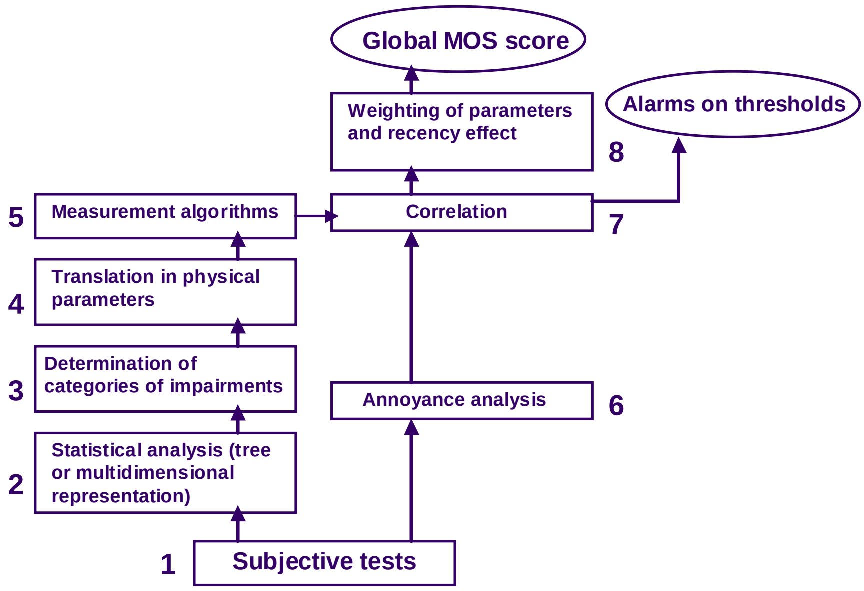


Figure 5: Building blocks of the development of non-intrusive measurement methods from subjective test results

Steps 1, 2 and 3 correspond to the elaboration of a library of typical impairments perceived by users of IP telephony. In step 4 each type of impairment is characterized by physical indicators to ensure a reliable detection in the speech signal. Step 5 corresponds to the detection of the library's impairments in the speech signal by the means of different signal processing algorithms. The parameters recommended in ITU-T Recommendation P.561 [14] can also be evaluated at this step of the process. The rating of the annoyance due to the voice degradations on a subjective scale is then necessary (step 6). Each impairment in the library is now characterized by physical and annoyance indicators. A correlation step (block 7) is then necessary to use objective measurement results to model the users' perception of voice quality. The last step determines a MOS score for each transmission direction by weighting all the impairments encountered in a call depending on their occurrence and location.

It is very useful to have a combination of both overall subjective and individual objective indicators as outputs from a quality estimation process: the MOS. score is directly linked to the perception of the quality by the users, and so gives a direct overall indication of the quality of service delivered. If computational models like the E-model (see 7.2.1) or the Call Clarity Index (see 7.2.2) are used rather than psycho-acoustical ones, parameters like the ITU-T Recommendation P.561 [14] ones will be accurately taken into account, but not necessarily the others like the ones taken in the library elaborated in steps 1 to 3; the alarms on thresholds are needed to undertake actions to improve the quality, based on a detailed knowledge of the cause of the impairments. Having only a single quality score, without this information, has no interest for practical network supervision purposes.

The results of this study showed the existence of 3 main impairment categories:

- "Clipping, reducing".
- "Metallic, robot voice, beep".
- "Whistling, breath, crackling".

Additionally, these results show that the impairments located outside speech activity periods are not perceived when there is no background noise and when some packet loss concealment mechanisms are used.

The analysis of the signal characteristics of the stimuli used during such subjective tests has then been used as the basis for the development of new non-intrusive measurement techniques of voice quality following steps 4 to 8 of figure 5.

7.3.2 Detection of the type of codec

Different studies have been run by the University of Kiel (Germany) concerning the possibility of distinguishing between codecs or families of codecs using INMDs.

A first study, presented at Eurospeech 2001, addressed a classification of coders in two classes, depending on their bit rate:

- G.711, G.726 and G.728 on one side (bit rates higher than 16 kbits/s);
- Other coders (GSM, G.729, G.723.1, IS-54) on the other.

The classification method used is gaussian, applied on signal features like the means and standard deviations of the prediction gain and the spectral flatness. The accuracy of this method reaches a recognition rate of 97 % for signals with at least 16 s of active speech.

A second study, presented at DAGA, took advantage of the specific characteristics of GSM-FR coder to develop an algorithm able to detect the presence of this coder. Indeed, the GSM-FR coder introduces a spectral attenuation around 2 700 Hz. The error rate of this algorithm is below 5 %.

7.3.3 Detection of comfort noise

Also presented at DAGA (see Bibliography), an algorithm has been developed by the University of Kiel to detect non-intrusively the presence of comfort noise inserted by DTX systems.

The first step of this algorithm consists in separating speech and pause segments from the signal by means of a VAD. Then, both parts are brought into the spectral domain using a discrete Fourier transform. Spectral points during speech which belong to background noise are localized and compared with the noise frequencies during pauses. Based on the correlation level obtained, a decision is made as to whether comfort noise is present (correlation < 0,6) or not.

Though tested on a relatively small database, such an algorithm seems promising, since it reaches error rates below 4 %.

7.3.4 Dynamic choice of AMR coding rate and optimization of bandwidth

The approach described in this clause has been developed by a Japanese company called Genista (see Bibliography), who specialize in Radio Resource Management (RRM) functions like link quality analysis and system-level resource allocation.

The idea of this product, called Sonogen, is the following: first the quality of the channel is monitored continuously and non-intrusively based on perceptual metrics like a MOS score, then the bit rate is adjusted based on real time statistical process control.

By using this perceptual rate adaptation solution for the Adaptive Multi Rate (AMR) codec, GSM operators have the ability to deliver a stable voice service to subscribers, competitive with wireline tool-quality services with less network equipment than the current Enhanced Full-Rate (EFR) codec.

In addition, such a solution enables AMR to offer the opportunity for rural coverage improvements and deeper in-building coverage due to the greater robustness of the adaptive full-rate channel. By including AMR into 3G network expansion plans and allowing for gradual handset penetration, operators can deliver capacity requirements with significantly less infrastructure, reducing capital investment and operating costs. AMR also frees up traffic channels for dedication to packet data (GPRS), improving the throughput and reducing the delay of mobile Internet services.

Annex A (informative): Examples of Specific Systems

Within ITU-T SG12/Q.9, the suitability of various non-intrusive approaches is under investigation, and a standardization process is ongoing for developing a future ITU-T Recommendation on single-ended (i.e. without reference signal) quality assessment algorithms before the end of 2003.

This annex lists the candidates for this standardization process, with a brief description of each one of them. These algorithms have been presented by several laboratories, but it is possible that the final standard algorithm will be a combination of several candidates.

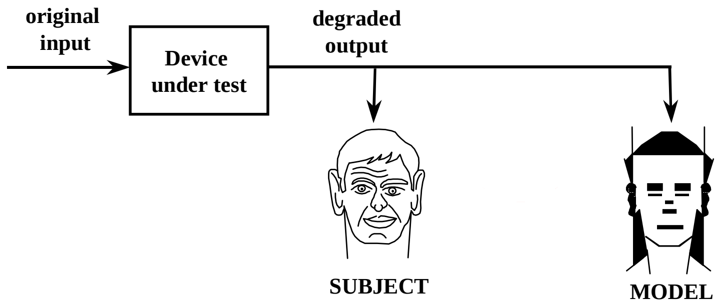
A.1 NiQA (Non-intrusive Quality Assessment)

This method, that was proposed by Psytechnics™ at the ITU-T Recommendation SG 12 in October 2001 (ITU-T Recommendation COM 12-D48), follows an earlier joint proposal by KPN and BT at the ITU-T Recommendation SG 12 in December 2000 (ITU-T Recommendation COM 12-11), extending CCI to take account of a full range of distortions.

It uses the following ingredients:

- A speech production model.
- A speech perceptual and cognitive model as used in PESQ [16].
- A set of distortion detectors that are optimized to detect specific kinds of common distortion as found in telecommunication systems.

The method uses the speech production model to find signal parts which can not be produced by the human vocal tract. Furthermore it finds signal parts where special kinds of distortions occur. The perceptual and cognitive models are used to estimate the impact on the perceived quality of the signal parts that are found by the speech and distortion models.



NOTE: A computer model of the subject, consisting of a perceptual and a cognitive model, is used to interpret distortions.

Figure A.1: Overview of the basic philosophy used in NiQA

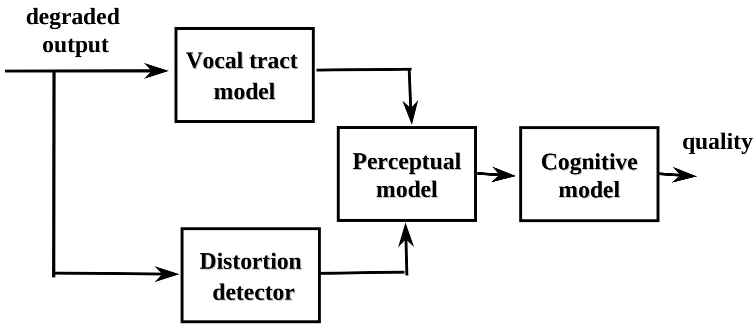


Figure A.2: Block Diagram of NiQA algorithm

A.2 NINA (Non-intrusive Network Assessment)

NINA is developed by SwissQual Inc. and has been presented in a white contribution to ITU-T Recommendation SG.12 in June 2001 (ITU-T Recommendation COM 12-27). It uses only the degraded signal on the receiving side to assess the quality of the same. It is based on the same fundamental human perception principles as used in intrusive models.

The NINA algorithm works as follows:

- The speech signal is separated from the input audio signal.
- The reference is estimated by separating the degradations from the input signal.
- Time clipping, like lost frames in VoIP are detected and interpolated.
- The reconstructed reference is used as an input for intrusive quality assessment.
- The impact of speech codec on the speech quality is evaluated.
- The long-term parameters of an input speech signal like: background noise, speech activity, interferences, echo residuals are calculated.
- The intermediate results are fed into a quality evaluation algorithm, which evaluates the grade of speech quality degradation.

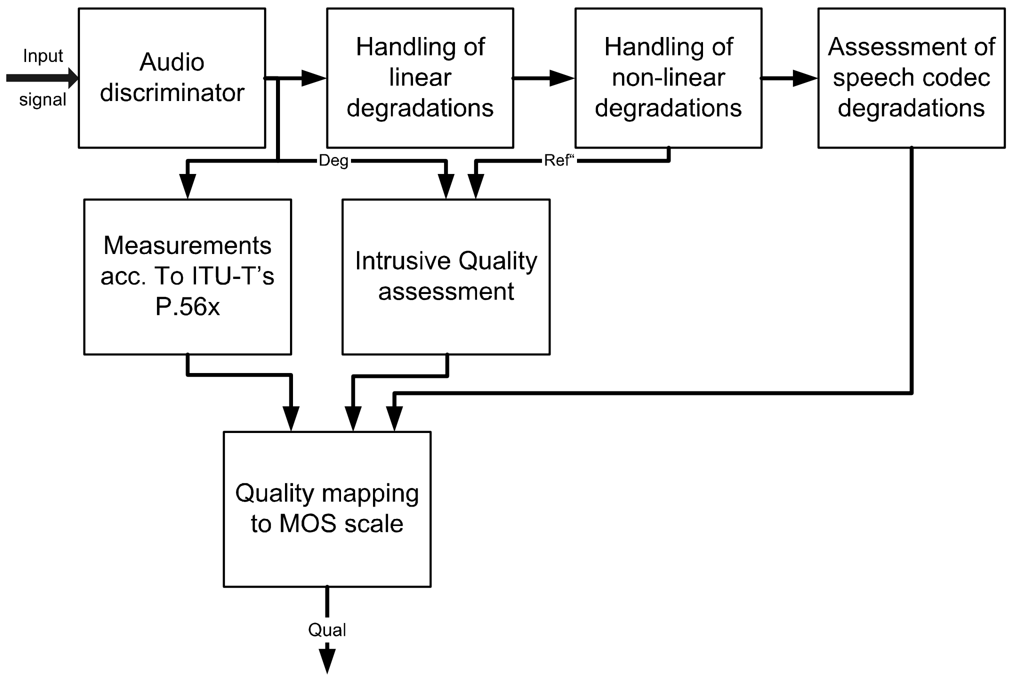


Figure A.3: Block Diagram of NINA algorithm

A.3 PSOM (Perceptual Single-ended Objective Measure)

PSOM is developed by France Telecom R&D and has been presented in a delayed contribution to ITU-T Recommendation SG.12 in May 2003 (ITU-T Recommendation COM 12-D79).

The PSOM algorithm works as follows:

- An auditory transform is first applied to the input audio signal. The output of the auditory transform is a time-frequency representation that approximates fundamental psychoacoustic properties.
- A speech segregation stage separates the speech stream from the additive noise and detects the occurrence of non-speech components like music or signalling tones.
- Speech and noise streams are separately processed by a characterization stage. In the case of speech signal, the likelihood measures of various perceptual parameters values are estimated from the statistical model of speech. Furthermore, the characterization stage computes a set of disturbance measures from the noise signal, like the energy level and the spectral characteristics of background noise.
- The likelihood measures of speech and the disturbance measures of noise are merged into a quality prediction stage, which maps them to an objective speech quality grade.

A schematic diagram of PSOM algorithm is shown in figure A.4. The "invalid input" flag is set when the objective quality grade is non-meaningful because of the presence of non-speech signals (other than background noise). In such case, the quality grade is set to "-99".

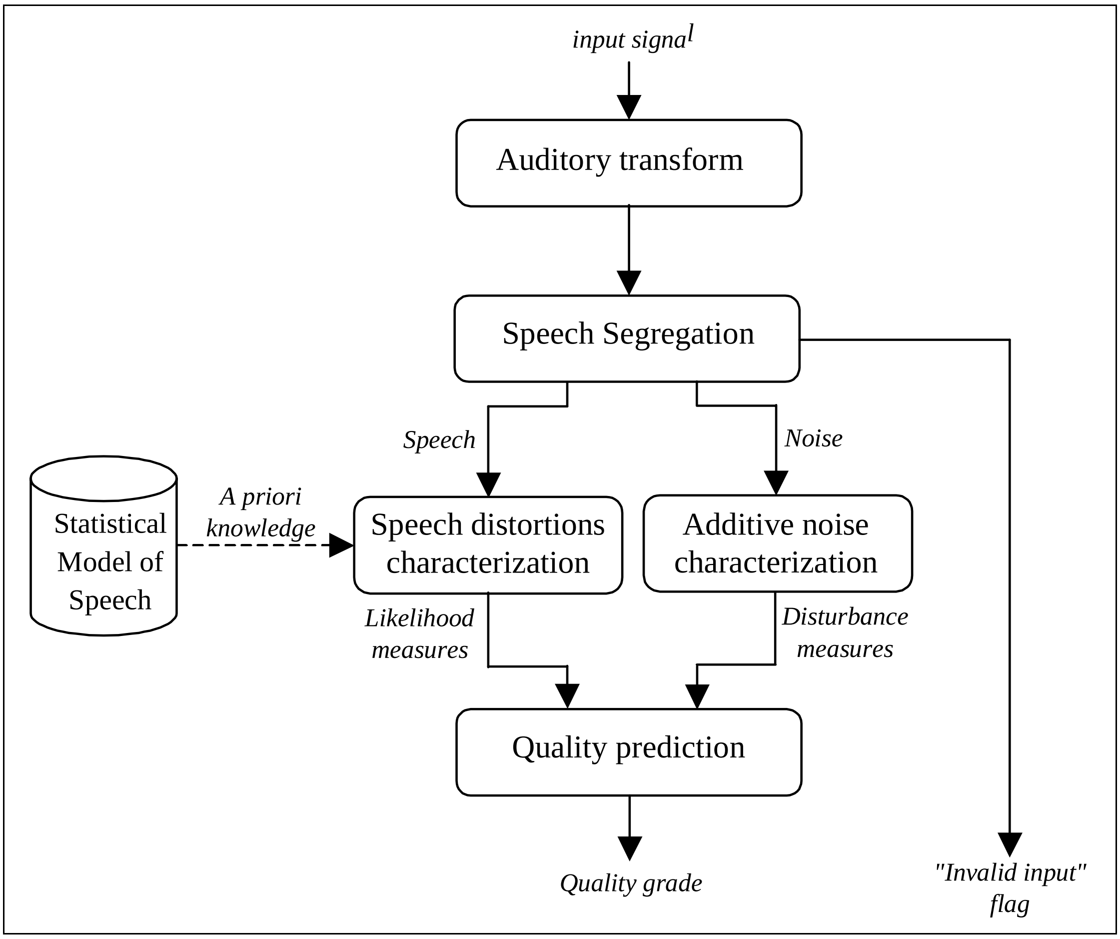


Figure A.4: Block Diagram of PSOM algorithm

Annex B (informative): Bibliography

ITU-T Recommendation COM 12-11 (2001-2004) KPN, The Netherlands, and BT, United Kingdom: "Call for proposals for a single ended speech quality measurement method for non intrusive measurements on live voice traffic".

ITU-T Recommendation COM 12-23 (2001-2004) France Telecom: "Non-Intrusive assessment of perceived voice quality impairments".

ITU-T Recommendation COM 12-27 (2001-2004) SWISSQUAL Inc.: "Non-intrusive speech quality measurement".

ITU-T Recommendation COM 12-D48 (2001-2004) Psytechnics (UK) (Q9, 16/12): "NiQA - Non-intrusive speech Quality Assessment".

ITU-T Recommendation COM 12-D49 (2001-2004) Psytechnics (UK) (Q9, 16?12): "Non-intrusive monitoring of speech quality in voice over IP networks".

ITU-T Recommendation COM 12-D79 (2001-2004) France Telecom (France): "PSOM - Perceptual Single-ended Objective Measure".

Eurospeech 2001, Thorsten Ludwig, Ulrich Heute (ICST, University of Kiel, Germany): "Detection of digital Transmission Systems for Voice Quality Measurements".

DAGA, Thorsten Ludwig (ICST, University of Kiel, Germany): "Speech Quality Evaluation in Telephone Networks".

Genista™ documentation on web site.

NOTE: "http://www.genista.com".

History

Document history			
V1.1.1	April 2003	Membership Approval Procedure	MV 20030606: 2003-04-08 to 2003-06-06
V1.1.1	June 2003	Publication	