International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.160

(06/2008)

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

International telephone connections and circuits – Apparatus associated with long-distance telephone circuits

---

Voice enhancement devices

Recommendation ITU-T G.160

ITU-T

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
General definitions	G.100–G.109
General Recommendations on the transmission quality for an entire international telephone connection	G.110–G.119
General characteristics of national systems forming part of international connections	G.120–G.129
General characteristics of the 4-wire chain formed by the international circuits and national extension circuits	G.130–G.139
General characteristics of the 4-wire chain of international circuits; international transit	G.140–G.149
General characteristics of international telephone circuits and national extension circuits	G.150–G.159
Apparatus associated with long-distance telephone circuits	**G.160–G.169**
Transmission plan aspects of special circuits and connections using the international telephone connection network	G.170–G.179
Protection and restoration of transmission systems	G.180–G.189
Software tools for transmission systems	G.190–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
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.160

Voice enhancement devices

Summary

Recommendation ITU-T G.160 applies to the characteristics, performance and testing of voice enhancement devices (VEDs) for use in digital network-based equipment. Such devices are commonly found in mobile networks and this version of this Recommendation is only applicable for mobile networks.

A VED is defined as certain signal processing network functions (SPNFs) (such as noise reduction and acoustic echo control) in the digital transmission path that perform voice enhancement functions on voiceband signals that traverse mobile networks.

Voice enhancement functions include the control of acoustic echo generated by wireless handsets, noise reduction, and the recognition and accommodation of tandem free operation (TFO) and interworking function (IWF) signals.

The current version of this Recommendation does not include certain tests that are still under study. In particular, it does not contain any objective tests covering voice quality performance of noise reduction functions, although a placeholder for such a test is given. In the meantime, the reader should be aware that voice quality performance of noise reduction functions is not addressed by this version of this Recommendation.

Source

Recommendation ITU-T G.160 was approved on 13 June 2008 by ITU-T Study Group 16 (2005-2008) under Recommendation ITU-T A.8 procedure.

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/.

CONTENTS

Page

Recommendation ITU-T G.160

Voice enhancement devices

1 Scope

This Recommendation applies to the characteristics, performance and testing of voice enhancement devices (VEDs) for use in digital network-based equipment. Such devices are commonly found in mobile networks, and this version of the Recommendation is only applicable for mobile networks.

A VED is defined as certain signal processing network functions (SPNFs) (such as noise reduction and acoustic echo control) in the digital transmission path that perform voice enhancement functions on voiceband signals that traverse mobile networks. Possible locations for VEDs located in mobile networks are shown in Appendix I.

Voice enhancement functions covered by this Recommendation include the control of acoustic echo generated by wireless handsets, noise reduction, and the recognition and accommodation of tandem free operation (TFO) and interworking function (IWF) signals.

The purpose of this Recommendation is to define certain performance constraints for VEDs and, where appropriate, to define laboratory tests that may be performed on a VED to verify that these constraints are met.

A VED that satisfies all the constraints and passes all of the tests in this Recommendation should ensure (but will not guarantee) that the pre-existing standard of the overall network performance (e.g., the transmission of speech, voiceband data and other voiceband signals, ISDN, etc.) is not degraded when the VED is installed in the network.

This Recommendation also describes the general characteristics of VEDs and indicates what characteristics are important to provide acceptable performance in the network. Issues relating to the interaction of VEDs with other network and subscriber equipment are also discussed in this Recommendation.

This Recommendation **does not** define a standard algorithm for voice enhancement functions. Nor does it apply to either electrical echo control (governed by [ITU-T G.168]) or automatic level control (governed by [b-ITU-T G.169]).

Note that the VED may be incorporated into other network equipment, such as the transcoder or electrical echo canceller. In this case, the performance requirements of this Recommendation refer to the performance of the VED only.

Note that some terminal devices (e.g., mobile stations) incorporate voice enhancement functions (acoustic echo control, noise reduction, etc.) that are similar to those performed by the VED. This Recommendation does not apply to functions contained in these terminal devices.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.164] Recommendation ITU-T G.164 (1988), *Echo suppressors*.

[ITU-T G.168] Recommendation ITU-T G.168 (2007), *Digital network echo cancellers*.

[ITU-T P.50] Recommendation ITU-T P.50 (1999), *Artificial voices*.

[ITU-T P.501] Recommendation ITU-T P.501 (2000), *Test signals for use in telephonometry*.

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

[ITU-T P.830] Recommendation ITU-T P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs.*

[ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.*

[ITU-T Q.24] Recommendation ITU-T Q.24 (1988), *Multifrequency push-button signal reception*.

[ITU-T Q.143] Recommendation ITU-T Q.143 (1988), *Specifications of signalling system No. 5 – Line signal sender*.

[ITU-T Q.144] Recommendation ITU-T Q.144 (1993), *Specifications of signalling system No. 5 – Line signal receiver*.

[ITU-T Q.271] Recommendation ITU-T Q.271 (1988), *Specifications of signalling system No. 6 – Continuity check of the speech path: General*.

[ITU-T Q.724] Recommendation ITU-T Q.724 (1988), *Specifications of signalling system No. 7 – Telephone user part signalling procedures*.

[ITU-T V.8] Recommendation ITU-T V.8 (2000), *Procedures for starting sessions of data transmission over the public switched telephone network*.

[ITU-T V.18] Recommendation ITU-T V.18 (2000), *Operational and interworking requirements for DCEs operating in the text telephone mode*.

[ITU-T V.21] Recommendation ITU-T V.21 (1988), *300 bits per second duplex modem standardized for use in the general switched telephone network*.

[ITU-T V.25] Recommendation ITU-T V.25 (1996), *Automatic answering equipment and general procedures for automatic calling equipment on the general switched telephone network including procedures for disabling of echo control devices for both manually and automatically established calls*.

[ITU-T V.27 *ter*] Recommendation ITU-T V.27 *ter* (1988), *4800/2400 bits per second modem standardized for use in the general switched telephone network*.

[ITU-T V.32] Recommendation ITU-T V.32 (1993), *A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits*.

[IEC 61260] IEC 61260:1995, *Electroacoustics – Octave-band and fractional-octave-band filters*. <http://webstore.iec.ch/webstore/webstore.nsf/artnum/019426>

[ETSI TS 128 062] ETSI TS 128 062 V7.0.0 (2007-06), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Inband Tandem Free Operation (TFO) of speech codecs; Service description; Stage 3*. <http://webapp.etsi.org/workprogram/Report_WorkItem.asp?WKI_ID=26667>

[TIA-825-A] TIA-825-A (2003), *A Frequency Shift Keyed Modem for Use on the Public Switched Telephone Network*. <http://standardsdocuments.tiaonline.org/tia-825-a.htm>

[TIA-895-A] TIA/EIA-895-A (10/2002), *CDMA Tandem Free Operation*.
<http://tia.nufu.eu/std/TIA-895-A>

3 Terms and definitions

This Recommendation defines the following terms:

3.1 acoustic echo: Acoustic echo is the reflected signal resulting from the acoustic path between the earphone/loudspeaker and microphone of a hand-held or hands-free mobile station.

3.2 disabled VED: A VED that has been forced into a 64 kbit/s bit-sequence integrity mode of operation, by whatever means.

3.3 echo canceller (EC): A device placed in the 4-wire portion of a circuit and used for reducing cancelled-end echo present on the send path by subtracting an estimation of that echo from the cancelled-end echo.

3.4 encoded acoustic echo: Encoded acoustic echo is the acoustic echo signal that has passed through the codecs and any other signal processing in the mobile network.

3.5 laboratory tests: Tests which may be performed in a laboratory or any environment where the VED is not installed in the network and is therefore not 'in service'.

3.6 noise: For the purposes of this Recommendation, noise is defined as a slowly varying stochastic process appearing additive to the desired speech signal. Specifically, the variations in the characteristics of the noise process are such that it can be considered approximately stationary over much longer time intervals than a typical speech signal.

3.7 signal processing network device (SPND): A physical entity (e.g., chipset) that contains one or more signal processing network functions (SPNFs).

3.8 signal processing network equipment/element (SPNE): A stand-alone physical entity that supports one or more signal processing network functions (SPNFs). It may also be controlled internally and/or externally and managed through a network management system which contains managed objects, a management communications function and a management application function. SPNEs include, e.g., echo cancellers, voice enhancement devices, circuit multiplication equipment, voice gateways.

NOTE 1 – Examples of managed objects are: receive/transmit port, power supply, plug-in cards, SPNFs. The communications function provides facilities for the transport of telecommunications management network (TMN) messages to and from the management application function, as well as facilities for the transit of messages. The message communications function does not originate or terminate messages. A management application function is the origin and termination for all TMN messages.

NOTE 2 – A SPNE in this context is a combination of hardware and software that performs SPNFs.

3.9 signal processing network function (SPNF): A function within a physical entity (e.g., SPNE, SPND) that performs signal processing to provide support or services to the transport network and/or to the users. Examples include electric or acoustic echo control, noise reduction, automatic level control, digital speech interpolation, low-rate encoding, transcoding.

3.10 voice enhancement device (VED): A VED is defined as certain SPNFs (such as noise reduction and acoustic echo control) in the digital transmission path that performs voice enhancement functions on voiceband signals that traverse mobile networks.

3.11 voice enhancement: To be defined.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations:

A/D	Analogue-to-Digital converter
AEC	Acoustic Echo Control
ALC	Automatic Level Control
AMR	Adaptive Multi-Rate
BER	Bit Error Rate
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CER	Character Error Rate
CN	Core Network
CSS	Composite Source Signal
CTM	Cellular Text Modem
D/A	Digital-to-Analogue converter
DCME	Digital Circuit Multiplication Equipment
DTMF	Dual-Tone Multi-Frequency
EC	Echo Canceller
ERL	Echo Return Loss
GSM	Global System for Mobile communication
ISDN	Integrated Services Digital Network
Iub	Interface between a Radio Network Subsystem and a Node B
Iu-cs	The physical instance of Iu (Interconnection point between the Radio Network Subsystem and Core Network) towards the Circuit-Switched Service Domain of the core network
IWF	InterWorking Function
LBR	Low Bit Rate
MG	Media Gateway (CDMA-2000)
MGW	Media Gateway (3GPP)
MS	2G Mobile Station
MSC	2G Mobile Switching Centre
Nb	3G internal mobile network interface
Node-B	3G network component which serves one cell
NR	Noise Reduction
NSS	Network and Switching Subsystem
OMC	Operation and Maintenance Control
PSTN	Public Switched Telephone Network

RAN	Radio Access Network
RMS	Root Mean Square
RNC	Radio Network Controller
SPND	Signal Processing Network Device
SPNE	Signal Processing Network Equipment
SPNF	Signal Processing Network Function
TC	Transcoder (3GPP)
TCE	Transcoder Equipment
TDM/PCM	Time Division Multiplex Interface
TFO	Tandem Free Operation
TRAU	2G/3G Transcoder Unit
TrFO	Transcoder Free Operation
UE	3G User Equipment
UMSC	3G Mobile Switching Centre
UTRAN	3G Radio Access Network
VBD	VoiceBand Data
VED	Voice Enhancement Device

5 Characteristics of VEDs

5.1 General

A VED is defined as certain SPNFs (such as noise reduction and acoustic echo control) in the digital transmission path that performs voice enhancement functions on voiceband signals that traverse mobile networks. Voice enhancement functions include the control of acoustic echo (AEC) generated by wireless handsets, noise reduction (NR), and the recognition and accommodation of tandem free operation (TFO) and interworking function (IWF) signals.

If a VED is provisioned OFF, the delay in the processing path (excluding any network interface delays) should not exceed 0.5 ms.

When a VED is provisioned ON, the delay in the processing path should be kept to a minimum.

A VED may incorporate processing in both directions of transmission. However, if there is no processing present in a certain direction, then the delay in that path (excluding any network interface delays) should not exceed 0.25 ms.

NOTE 1 – VED AEC or NR functions occurring together or with other SPNFs may be algorithmically coupled. In this case, the OFF processing delay requirement applies when all such coupled SPNFs are OFF.

NOTE 2 – A VED is expected to be provisioned OFF by action of the network administration or by a network coordination mechanism at call set-up, and should be performed before the answer side goes off-hook to prevent possible step delay changes during an active call.

NOTE 3 – [b-ITU-T G.161] gives advice on the interaction of signal processing functions, such as AEC and NR and recommends their relative placement in networks.

5.2 Acoustic echo control (AEC)

Acoustic echo control is a signal processing function located in the digital transmission path, which has the objective of improving subjective speech quality by reducing encoded acoustic echo. A handset may inject acoustic echo into a digital mobile telephony system which is transmitted together with signals originating from mobile handsets to the location of the VED through the air interface, speech codecs and other signal processing and transport functions of the mobile network.

An acoustic echo control function has the following general characteristics:

- The ability to reduce acoustic echo generated in the mobile handsets.
- The ability to maintain quality of the signals originating from mobile handsets.
- The ability not to corrupt voice-band data or facsimile signals.
- The ability not to interfere with in-band network signalling tones.
- Exhibit low throughput delay.
- Provide 64 kbit/s bit-sequence integrity when disabled.

5.2.1 External disabling/enabling

The AEC function should be disabled during voice-band data and facsimile transmissions. The AEC function may be disabled by an external digital signal. If external disabling is supported, the AEC function should provide 64 kbit/s bit-sequence integrity in the externally disabled state.

5.2.2 In-band disabling

AEC functions covered by this Recommendation should be equipped with a tone disabler that conforms to clause 5 of [ITU-T G.164]. Additionally, the tone disabler should meet the requirements of [ITU-T G.164] when the disable tone is phase and/or amplitude modulated in accordance with [ITU-T V.25] and [ITU-T V.8]. When disabled, the AEC function should provide 64 kbit/s bit-sequence integrity.

5.3 Noise reduction (NR)

Noise reduction is a signal processing function located in the digital transmission path, which has the objective of reducing the noise level, thereby aiming to improve subjective speech quality as perceived by humans.

Noise reduction functions have the following general characteristics:

- The ability to modify a noise-corrupted voice signal, so as to improve the subjective quality.
- The ability to maintain the quality of the voice signal when it is not corrupted by noise.
- The ability not to corrupt voice-band data or facsimile signals.
- The ability not to interfere with in-band network signalling and information tones.
- Exhibit low through-put delay.
- Provide 64 kbit/s bit-sequence integrity when disabled.

5.3.1 External disabling/enabling

The noise reduction function should be disabled during voice-band data and facsimile transmissions. The NR function may be disabled by an external digital signal. If external disabling is supported, the noise reduction function should provide 64 kbit/s bit-sequence integrity in the externally disabled state.

5.3.2 In-band disabling

Noise reduction functions covered by this Recommendation should be equipped with a tone disabler that conforms to clause 5 of [ITU-T G.164]. Additionally, the tone disabler should meet the requirements of [ITU-T G.164] when the disable tone is phase and/or amplitude modulated in accordance with [ITU-T V.25] and [ITU-T V.8]. When disabled, the noise reduction function should provide 64 kbit/s bit-sequence integrity.

5.4 Tandem free operation (TFO)

A VED that encounters TFO signals should support tandem free operation according to [ETSI TS 128 062] and [TIA-895-A]. Once in TFO mode, a VED may optionally choose to invoke speech enhancement functions, causing the termination of TFO. The decision should be based on whether TFO or speech enhancement techniques will provide better overall speech quality. The evaluation of overall speech quality on which this decision is based is for further study. The following factors should also be taken into consideration:

- The decision to terminate TFO should take into account the apparent noise generated by the TFO signalling in the LSBs;
- VEDs should not terminate TFO for AMR wideband calls; and therefore
- VEDs need to detect TFO signalling messages in order to determine the codec type;
- VEDs should take care to avoid frequent switching between TFO and non-TFO.

NOTE – There is a need to have interoperability of different TFO standard Recommendations. The interoperability of TFO techniques described in [ETSI TS 128 062] and [TIA-895-A] is for further study.

5.5 Interworking functions (IWF)

Under study.

6 Test signals

6.1 General

For noise reduction devices, it is imperative that the speech and noise models used for evaluation represent the real-world conditions in which the noise reduction device is expected to operate. For example, in a wireless environment, the additive noise can be highly non-stationary. In [b-Gibson] it was found that simple additive white noise is turned into harmonic-related signals by mobile-radio speech coders. Additionally, harmonically rich background sounds can cause wavering or warbling in the output speech signal.

6.2 Measuring input and output signal levels

Unless otherwise noted, for the tests in this Recommendation, the input and output signals are measured using an RMS method. The power of the signal, *S(k)*, is measured in dBm0 using the following equations for A-law and μ-law encoding:

$$S(k) = \begin{cases} 3.14 + 20\log\left[\dfrac{\sqrt{\dfrac{2}{n}\displaystyle\sum_{i=k}^{k-n+1} x_i^2}}{4096}\right], & \text{for A-law encoding} \\[2ex] 3.17 + 20\log\left[\dfrac{\sqrt{\dfrac{2}{n}\displaystyle\sum_{i=k}^{k-n+1} x_i^2}}{8159}\right], & \text{for μ-law encoding} \end{cases}$$

Where x_i is the linear value of the encoded signal, *k* is the discrete time index, and *n* is the number of samples over which the measurement is made.

6.3 Gaussian white noise

Where Gaussian white noise is specified as the test signal in this Recommendation, it shall have the following characteristics, unless otherwise stated: Gaussian white noise, band-limited 300 Hz – 3400 Hz, with a crest factor of 11 dB ± 1 dB.

7 Tests

Each of the tests described in this clause is designed to evaluate different aspects of VED performance. Some tests should be performed on the overall VED, whereas others are designed to test specific functions (e.g., NR and AEC).

7.1 Overall VED tests

These tests should be performed on the overall VED, including all of the functions defined in clause 5.

7.1.1 Test 1.1 – Signalling tones test (optional)

This test applies to VEDs that are not externally disabled by the switch and which are located on the side of a switch where Signalling System No. 5, 6 and 7 tones are present.

VEDs should not distort or corrupt signalling tones in the network. This test ensures that the VED will not adversely affect signalling or continuity check tones.

Test procedure

With the VED in any initial state, the signalling tones of Table 1 are applied to the input of the VED. The level of the output from the VED is monitored.

Table 1 – Applicable signalling tones

System 5	System 6	System 7
2400 ± 15 Hz	2000 ± 30 Hz	2000 ± 30 Hz
2600 ± 15 Hz		
2400 ± 15 Hz & 2600 ± 15 Hz		

The power level M, of each signal applied, should be within the following limits:

System 5: $-16 \le M \le -2$ dBm0

NOTE 1 – This range applies to the single frequency signals f1 and f2. The level of the individual signals in the compound signal may differ from each other by not more than 5 dB. Note that when the levels of the individual signals approach –2 dBm0, the compound signal may be clipped. The nominal transmit level is specified as –9 ± 1 dBm0.

See [ITU-T Q.143] and [ITU-T Q.144].

System 6: $-18 \le M \le -6$ dBm0

NOTE 2 – The nominal transmit level is –12 ± 1 dBm0.

See [ITU-T Q.271].

System 7: $-18 \le M \le -6$ dBm0

NOTE 3 – The nominal transmit level is –12 ± 1 dBm0.

See clause 7 in [ITU-T Q.724].

The above levels are designed to ensure that the VED will operate with signals that occupy the entire range of levels given in the appropriate signalling ITU-T Recommendations.

The VED detection time should be sufficiently long to provide immunity from false operation due to voice signals. Conversely, the VED detection time should not be so long as to needlessly extend the time for the signal to appear on the network.

Requirement

The level of the signalling tones measured at the output of the VED should not vary more than ±2 dB from the level of the input after an allowed 1 second detection time.

7.1.2 Test 1.2 – DTMF tones test

The object of this test is to ensure that the VED will not interfere with the detection of DTMF tones. This test is not required if a VED will never encounter DTMF tones when it is deployed in a network. Examples of such cases are a VED integrated into other network equipment that has the ability to generate and detect DTMF tones, and a VED applied at the A-interface of a GSM wireless network.

In any test to check DTMF detection accuracy in a connection with a VED, the choice of DTMF detector may influence the percentage of detection failures. To circumvent this problem, the following test is based on a comparison of DTMF detection performance with and without the VED in-circuit. It is therefore unnecessary to specify a standard DTMF detector for the test.

The test requires a means of inputting a sequence of DTMF characters to the test circuit (e.g., sequence stored on digital audio tape, or PC file). A suitable DTMF detector (which may be PC-based for convenience) is also required, together with a means of recording the percentage of character detection failures in the replayed DTMF test sequence. A noise source is needed for the tests, and the basic test configuration is shown in Figure 1.

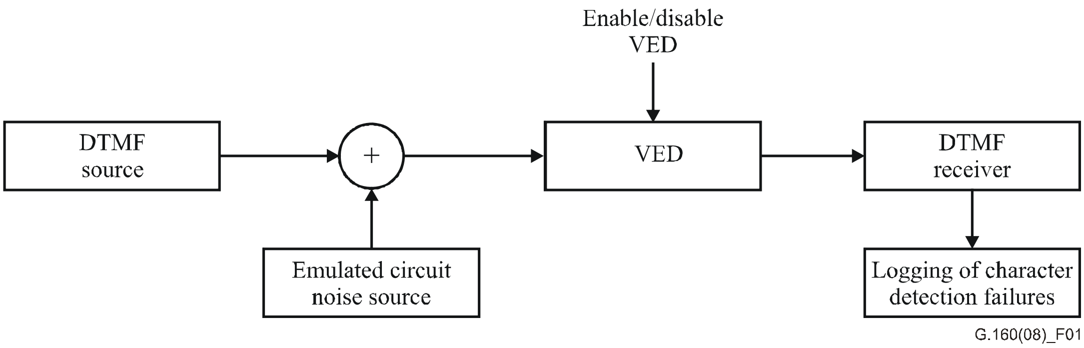


Figure 1 – Circuit configuration to check effect of VED on DTMF detection performance

The DTMF test sequence should consist of 256 different DTMF signals, arranged as 16 sets of the 16 DTMF characters (0..9, #, *, A, B, C, D). Each set has the frequency offset specified in Table 2. The replayed levels of the discrete frequencies making up the tone pairs should be as specified in Table 2. Each DTMF character should have a duration of 50 ms, with an interval between characters of 100 ms. The interval between each set should be 1 second, and the 16 sets of DTMF characters should be replayed as a continuous sequence of 256 characters. The timing relationships for this test are shown in Figure 2.

Table 2 – DTMF test sequences

Set No.	Freq. offset (%) (low freq.)	Freq. offset (%) (high freq.)	Level (dBm0) (low freq.)	Level (dBm0) (high freq.)
1	0	0	0	0
2	0	0	–18	–18
3	1.5	1.5	–10	–10
4	–1.5	–1.5	–10	–10
5	0	0	–12	–18
6	0	0	–14	–10
7	1.5	1.5	–14	–20
8	–1.5	1.5	–14	–20
9	1.5	–1.5	–6	–12
10	–1.5	–1.5	–6	–12
11	0	0	–12	–18
12	0	0	–14	–10
13	1.5	1.5	–10	–6
14	–1.5	1.5	–10	–6
15	1.5	–1.5	–18	–14
16	–1.5	–1.5	–18	–14

Annex A of [ITU-T Q.24] indicates that some Administrations require DTMF receivers to tolerate power levels per frequency up to 0 dBm.

Electrical circuit noise should be emulated by a Gaussian, white noise signal, band-limited to the frequency range 0-3.4 kHz at source. References to the level of this noise test signal relate to the mean power of the signal, measured flat (i.e., without psophometric weighting). The level of the noise should not exceed –38 dBm0.

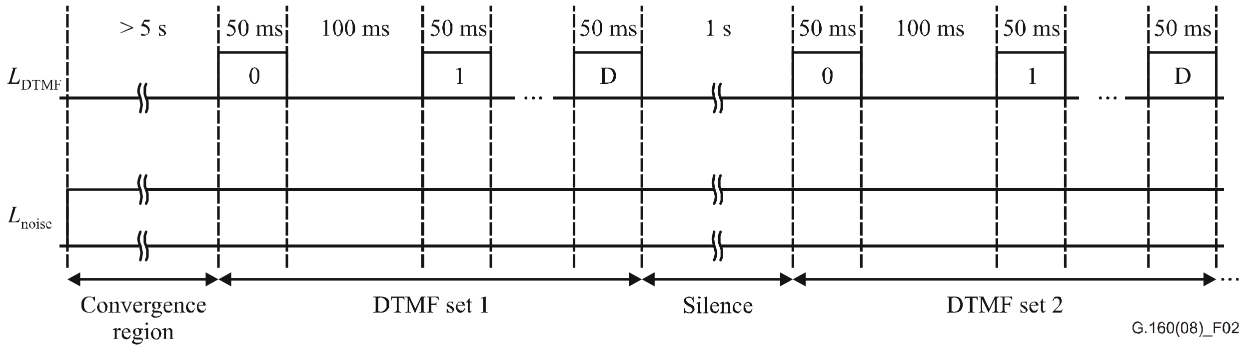


Figure 2 – Timing relationships for the DTMF tones test (two sets out of 16)

Test procedure

1) With the VED disabled, replay the test signal comprising the sum of DTMF test sequence and emulated circuit noise over the test circuit. The test signal should start with the noise-only segment of at least 5 seconds. This will allow the VED to converge during step 3. Record the percentage of detection failures and detection errors.

2) Repeat step 1 a sufficient number of times to satisfactorily establish the standard of DTMF detection performance. Detection with VED disabled should be reliable. If this is not the case, lower the noise level and repeat steps 1 and 2.

3) Repeat steps 1 and 2, but with the VED enabled. If DTMF detection performance is noticeably worse with the VED enabled, it must be assumed that the VED degrades the transmission of DTMF signals.

7.1.3 Test 1.3 – Text telephone, data modem and facsimile test (under study)

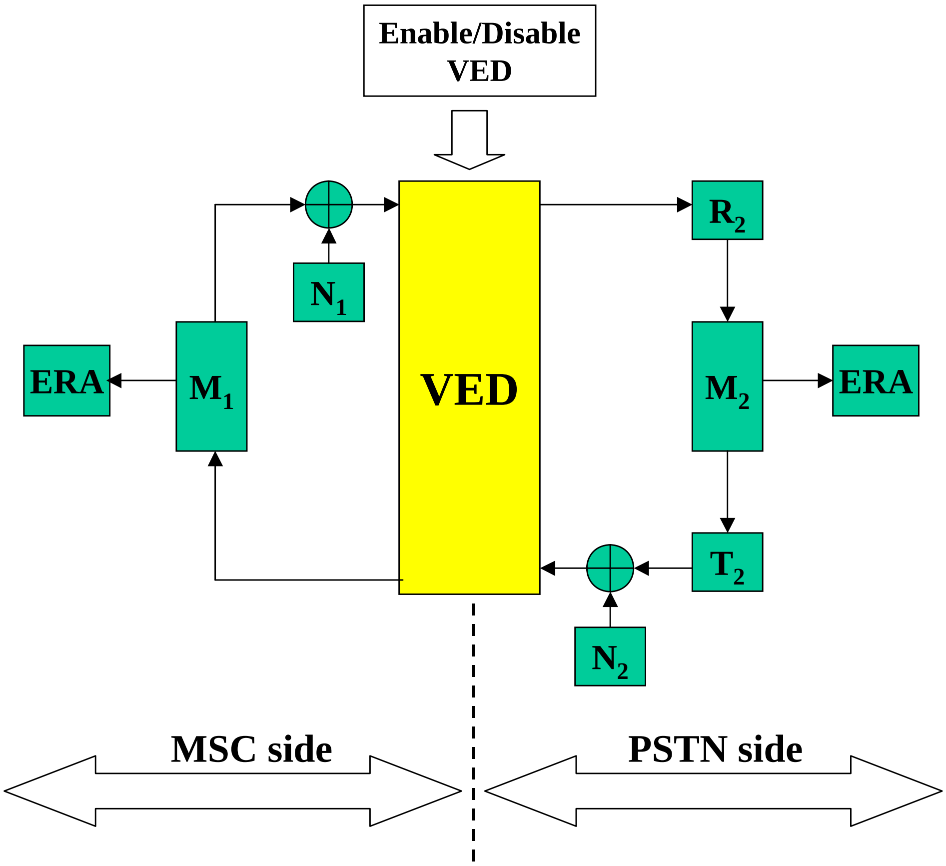
This test will ensure that a VED will not degrade voice-band data signals (text telephone, data modem and facsimile) used in mobile networks. Table 3 is a practical representation of modems most commonly used in GSM networks that should be used for testing. In a GSM network, all modems except cellular text modems (CTMs) are located on the PSTN side of the mobile switching centre (MSC). CTM modems are located in the voice path of the GSM network, e.g., at the A-interface as shown in Figure I.18. The intent of this test is to allow a VED vendor to demonstrate acceptable performance with a range of modem types used in GSM networks, and for network operators to gain a level of confidence that existing modem traffic will not be impaired. The performance criteria for this test depend on the type of modem being tested. For [ITU-T V.32] and [ITU-T V.21] data modems, bit error rate (BER) is used to measure performance (measured synchronously where possible). For textphone modems, however, character error rate (CER) is more appropriate. The performance criteria combined with minimum volume of data to be transmitted for each type of modem are given in Table 3.

This test is not required if a VED will never encounter text telephone, data modem and facsimile signals when it is deployed in a network. An example of such a case is a VED integrated into another network equipment that has the ability to bypass the VED for the treatment of text telephone, data modem and facsimile signals (e.g., using [b-ITU-T V.150.1] procedures).

Table 3 – Minimum set of modems to be used in Test 1.3 (relevant to GSM and 3G mobile networks only)

Application	Modem type	Data rate	Answer tone	Performance criteria	Minimum volume of data
GSM Circuit Switched Data	V.32	9600 bit/s 2400 bit/s	ANS	BER	10000 bytes
GSM Circuit Switched Data	V.21	300 bit/s		BER	10000 bytes
GSM Text Telephony Transcoding CTM/V.18 and V.18/CTM	V.18 TIA 825-A Baudot V.21 V.23 EDT	300 bit/s 45.45, 50 baud 300 bit/s 75/1200 bit/s 110 bit/s	ANS None ANS (optional) ANS (optional) ANS (optional)	CER	1000 characters
GSM Voice-band text data	ETSI Cellular Text Modem (CTM)	400 bit/s	None	CER	1000 characters
GSM Facsimile	V.27 *ter*	4800 bit/s	ANS		2 pages

ANS 2100 Hz Answer Tone as defined in [ITU-T V.25]
BER Bit Error Rate
CER Character Error Rate
NOTE – The CTM is described in [b-ETSI TS 126 226].



ERA	Error Rate Analyser
VED	Voice Enhancement Device
M_1, M_2	Modems listed in Table 3
N_1, N_2	Emulated Circuit Noise
R_2	Receive Attenuator
T_2	Transmit Attenuator

NOTE 1 – Modem M1 represents a 4-wire modem located in the MSC that is controlled using AT commands from the MS or UE.

NOTE 2 – Modem M2 represents a 2-wire text telephone, data modem or facsimile machine connected to the PSTN.

NOTE 3 – Noise N1 represents background noise that may be present when using a CTM adapter for a text telephone. N1 is not required for testing data modem and facsimile.

Figure 3 – Test set-up for Test 1.3 – Text telephone, data modem and facsimile test

The values of the settings for all modems should be as follows:

R_2 = 6 dB to simulate access loss.

T_2 = 3 dB to 9 dB (3 dB is the nominal level, 9 dB simulates a 6 dB level offset).

M_1, M_2 = modem data transmission levels should be between –8 dBm and –20 dBm.

N_1, N_2 = set to produce signal-to-noise ratios of not less than 25 dB, and no noise.

Test procedure

The sequence for this test is to measure the performance of each modem type with the VED disabled and compare this performance with the VED enabled. The philosophy is that when enabled, the VED should not degrade performance compared to that when disabled. The VED is placed in the test configuration of Figure 3. The test procedure is to first disable the VED and allow the modems to train. The modems are then operated for the time it takes to send the minimum volume of test data as defined for each modem type given in Table 3 (e.g., bit error rate, character error rate, etc.). This performance is noted and the test is then repeated with the VED enabled.

Requirement

With the VED enabled, for the conditions specified above, the percentage of data errors should not increase compared with when the VED is disabled, when data is exchanged between the two terminals for the given minimum volume of data transmitted.

7.2 Noise reduction tests

These tests should be performed on the noise reduction function only, with all other functions disabled. Parameters Q_m (required) and L_{C_min} (optional) are specific for a particular noise reduction function. The parameter Q_m is the specified level of noise reduction in dB. The optional parameter L_{C_min} specifies a lower bound on input noise power levels, below which the function is not required to achieve the specified Q_m.

7.2.1 Test 2.1 – Noise reduction signal level test

This test is intended to ensure that a NR function included in a VED produces the stated amount of noise reduction and does not modify speech signal levels beyond acceptable limits.

Test procedure

The test set-up is shown in Figure 4 and the sequence of events is shown in Figure 5. The test procedure is to apply a noise signal during the entire test, starting at time t_0. During the time period t_0 to t_1, the NR function is allowed to converge and reach the specified amount of noise reduction (Q_m). Note that the level of the noise signal should be high enough to allow a reduction by Q_m dB to be observable in the PCM signal. In Figure 4, the levels in points A, B, C and D are measured in the unit of dBm0.

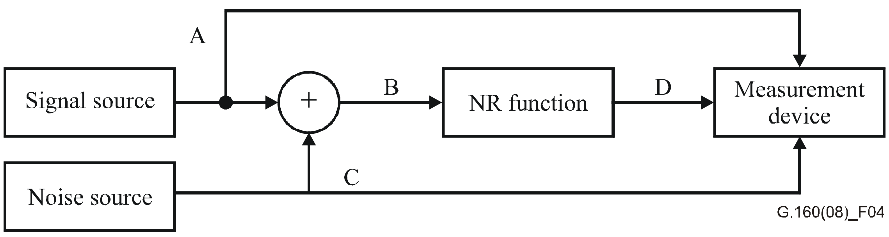


Figure 4 – Test set-up

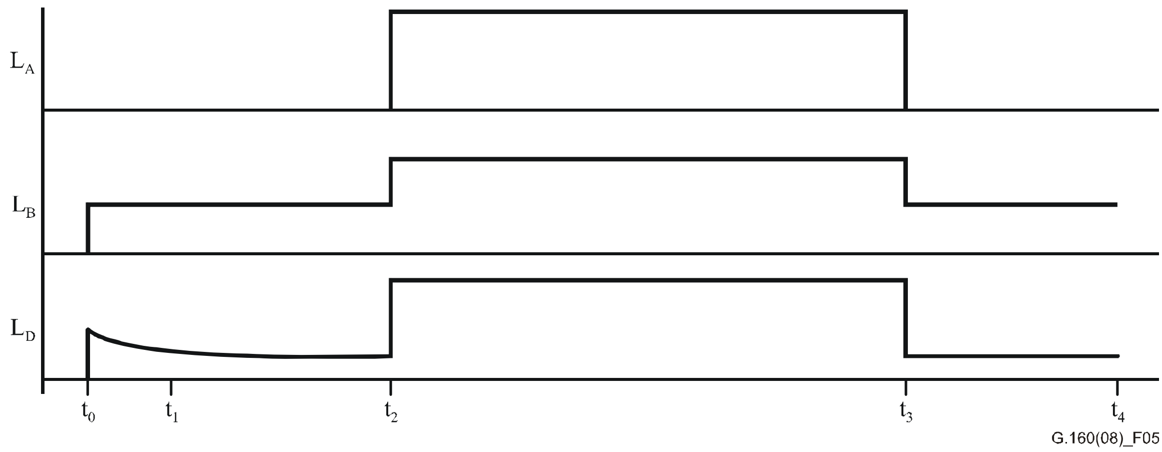


Figure 5 – Signal description

During the period t_1 to t_2 and t_3 to t_4, the levels at the output of the NR function (point D) and at the output of the noise generator (point C) are measured and noise attenuation factors are computed as $Q_n = L_C - L_D$. At time t_2, a measurement signal is added and kept for a period of at least 22 seconds from t_2 to t_3. In order to make the contribution from the noise signal to the signal level measured in point D negligible, the level of the measurement signal should be at least 20 dB above the level of the reduced noise. During the period t_2 to t_3, the output level of the NR function (point D) and at the measurement signal level (point A) are measured. The level difference is computed as $Q_s = L_A - L_D$.

Note that noise reduction functions may introduce a delay into the signal path. This delay needs to be compensated for before comparing input and output levels of the function.

White Gaussian noise should be used as the noise signal and the male and female artificial voice passages from [ITU-T P.50] as the measurement signals in this test. Concatenated, the artificial voice passages are approximately 22 seconds in length. The RMS method specified in clause 6.2 should be used for level measurements. Other measurement methods are for further study.

The time instants t_0 to t_4 are defined as follows: $t_0 = 0$, $t_1 = 3$ s, $t_2 = 10$ s, $t_3 = 32$ s, $t_4 = 39$ s.

Requirements

For all RMS signal levels –6 dBm0 ≥ L_A ≥ –30 dBm0, noise levels max[L_{C_min} , –60 dBm0 + Q_m] < $L_C < L_A$ – SNR and SNR values given by:

$$\text{SNR} \geq 12 \text{ dB} \qquad \text{if } Q_m \geq 8 \text{ dB}$$

$$\text{SNR} \geq 20 \text{ dB} - Q_m \qquad \text{if } Q_m < 8 \text{ dB.}$$

1) During the period t_1 to t_2 and t_3 to t_4, the noise should be attenuated by the amount specified for the particular setting of the NR function under the test ± 3 dB, i.e., $Q_m - 3 < Q_n < Q_m + 3$. To verify this, the RMS signal level measurement is computed twice according to the definition of clause 6.2. First across the signal period from t_1 to t_2 ($k = t_2$, $n = t_2 - t_1$) and then from t_3 to t_4 ($k = t_4$, $n = t_4 - t_3$).

2) During the period from t_2 to t_3, the signal levels measured at the output of the NR device (point D) and at the output of the measurement signal generator (point A) should satisfy $-3 < Q_s < 2$. To verify this, the RMS signal level is computed according to the definition of clause 6.2 across the signal period, from t_2 to t_3 ($k = t_3$, $n = t_3 - t_2$).

NOTE 1 – Three level measurements using the RMS method as defined in clause 6.2 are made in this test. Short-term signal level comparisons are likely not to meet Requirement 2 above in all cases and are not required to do so.

NOTE 2 – Noise attenuation for noise sources, other than white noise, may vary from these test results.

7.2.2 Test 2.2 – Noise reduction convergence/reconvergence test

This test is intended to ensure that a NR function included in a VED produces the stated amount of noise reduction in response to a step change in noise level after a maximum allowed convergence time.

Test procedure

The test set-up is the same as shown in Figure 4 except that the signal source is not applicable. The test procedure is to apply a white Gaussian noise signal abruptly changed in level as shown in Figure 6.

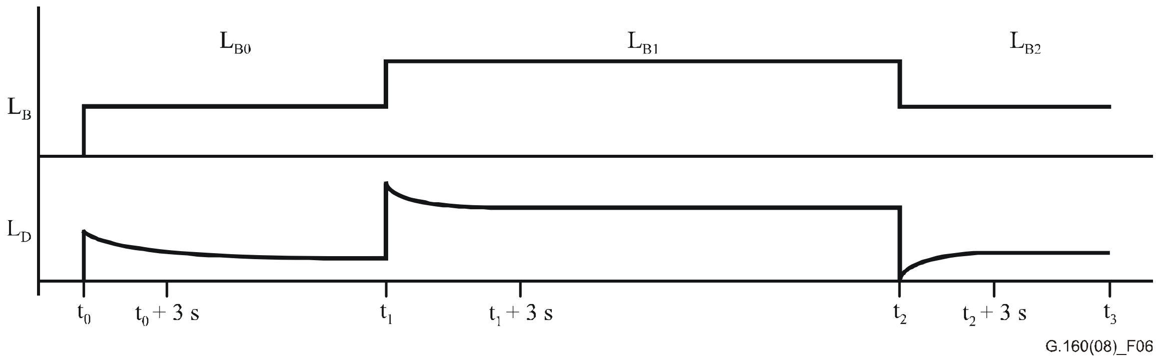


Figure 6 – Signal levels

The L_B RMS noise powers for the indicated time periods are as follows:

For $< t_0$, L_B = quiet

For t_0 to t_1, $\max[L_{C\_min}, -60 \text{ dBm0} + Q_m] < L_{B0} < -30$ dBm0 where $t_1 = t_0 + 6$ seconds

For t_1 to t_2, $L_{B1} = L_{B0} + 12$ (range) dBm0 where $t_2 = t_0 + 12$ seconds

For t_2 to t_3, $L_{B2} = L_{B0}$ (range) dBm0 where $t_3 = t_0 + 18$ seconds

Optionally, other level combinations may be tested.

Requirements

3.0 seconds after the input noise level, L_B, changes as indicated in Figure 6, the NR output noise level, L_D, should be attenuated by the amount specified for the particular setting of the NR function under the test ±3 dB, i.e., $Q_m - 3 < Q_n < Q_m + 3$.

7.2.3 Test 2.3 – Noise reduction voice quality test (under study)

This test is meant to assess perceived listening quality of a noise reduction algorithm. For this purpose, it is based on the use of objective methods (to be determined), which will predict the subjective quality as specified in [ITU-T P.835].

This test is split into 3 parts.

7.2.3.1 Test 2.3a

This test evaluates the ability of the NR function to maintain the quality of the voice signal at high SNR.

Test method: TBD

High SNR values: SNR ≥ 18 dB

7.2.3.2 Test 2.3b

This test evaluates the ability of the NR function to improve the quality of the voice signal at moderate SNR.

Test method: TBD

Moderate SNR values: 18 dB > SNR ≥ 12 dB

7.2.3.3 Test 2.3c

This test evaluates the ability of the NR function to maintain or improve the quality of the voice signal at low SNR.

Test method: TBD

Low SNR values: SNR < 12 dB

NOTE – Future development of this test should take into account the optional provision for the parameter L_{C_min} as suitable test and requirement details are developed.

7.3 Acoustic echo control tests

These tests should be performed on the acoustic echo control function only, with all other functions disabled.

7.3.1 Possible AEC test set-ups

Possible test set-ups are shown in Figures 7 and 8. Echo path models are shown within dotted boxes. Because returned acoustic echo is always encoded in real network implementations, encoded echo should be simulated in the echo path model by using appropriate low bit rate (LBR) codecs both in the receive and send directions of the echo path. The choice of an appropriate LBR codec or codecs should be based on the supported codecs in a target mobile network.

In addition to LBR codecs, the echo path model should consist of a single ideal reflection with an appropriate echo return loss (ERL) and a radio access delay simulation functionality. The total echo path delay will be equal to the radio access delay plus twice LBR codec delay.

NOTE – Although the single ideal reflection is not representative of the acoustic echo that will be present in a practical handset, where the signal will be filtered, it is considered sufficient for the purpose of this test. The real acoustic echo path impulse response varies considerably from case to case (hands-free, handset, different mobiles, etc.). Also it has typically a non-linear component that cannot be predicted. Having a band-pass filter in the echo path would bring the echo path frequency response closer to that of real acoustic echo paths. However, to include an optimal design would necessitate the collection of numerous real-world data so that a representative band-pass filter could be designed. One possibility would be to use the modified IRS receive filter before the ERL block and the modified IRS send filter after the summing block. This would result in increased complexity since the modified IRS filter of [b-ITU-T G.191] requires up- and down-sampling. Furthermore, the attenuation due to any IRS filtering would then need to be compensated for in the ERL block. The use of a more representative echo path in this test is for further study.

The acoustic echo control defined in this Recommendation is located in the network, while the dispersive part of the echo path is formed by the acoustic path in a handset. Hence, various equipments like base station controller (BSC), base transceiver station (BTS) and mobile station (MS), in the GSM system, appear in the echo path. This equipment introduces a considerable delay (in the order of 180 ms round-trip) contributing to a minimum possible echo delay seen by acoustic echo control. Acoustic echo control may take this pure delay into account and hence be able to control echo in a delay range between Δ_1 and Δ_2, where $\Delta_2 > \Delta_1 > 0$. The values of Δ_2 and Δ_1 should be selected in accordance with the delay characteristics of the mobile system where the acoustic echo control is deployed as well as the placement of the VED in the system. Detailed delay calculations for a GSM system can be found in [b-ETSI TR 126 975].

If a VED can provide a [b-ITU-T G.711] connection towards a radio access network, the test set-up depicted in Figure 7 may be used. If a VED can only provide a LBR encoded connection towards a radio access network, the test set-up depicted in Figure 8 may be used instead. Figure 9 shows pre-processing stages required for the signal and noise sources of Figures 7 and 8. The modified IRS send filter is required to model the frequency response of the handset at each end of the connection. White noise at –65 dBm0 is added to the speech signal to represent the typical noise floor due to electrical noise in the system.

No external control is required between a VED and a test system. A connection from S_{in} (or a connection representing S_{in} signal) to a measurement device is required in the comfort noise test. Otherwise, the test set-up follows Figure 8.

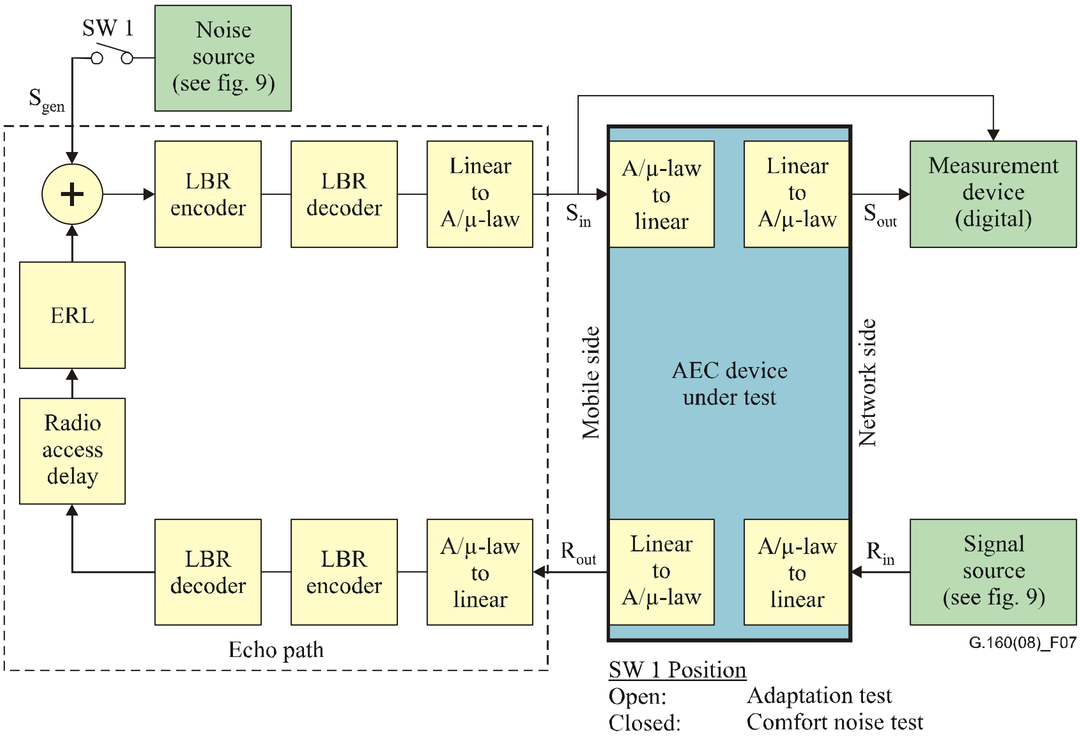


Figure 7 – Test set-up in which AEC provides G.711 connection to mobile direction

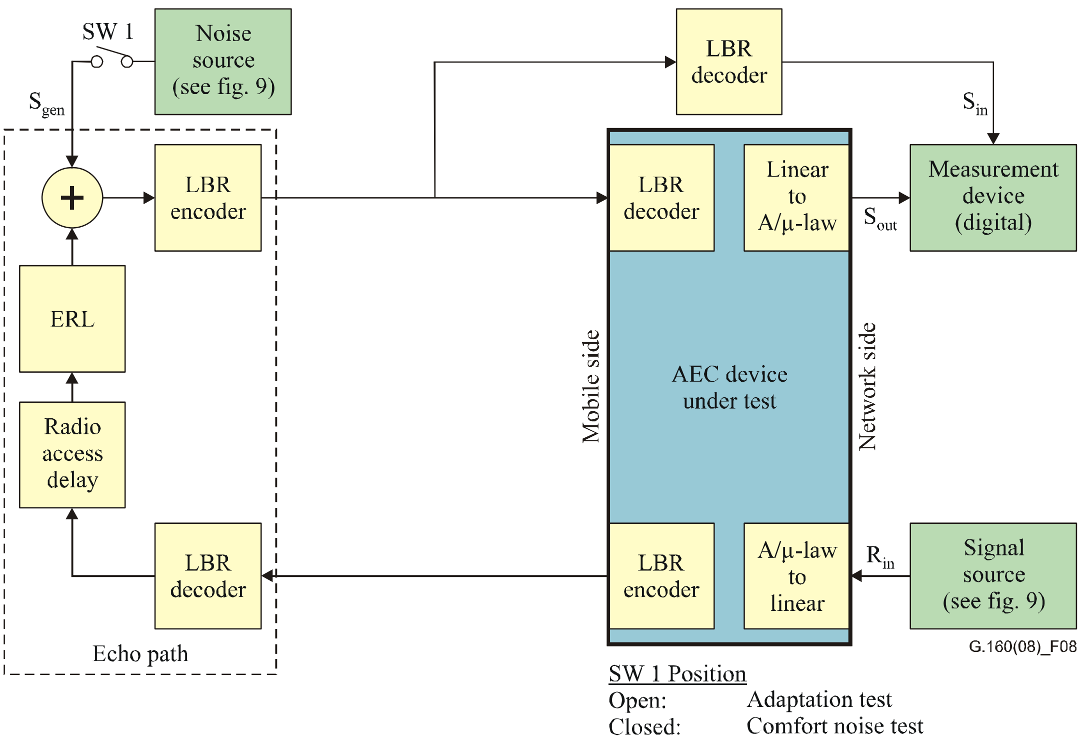


Figure 8 – Test set-up in which AEC provides LBR encoded connection to mobile direction

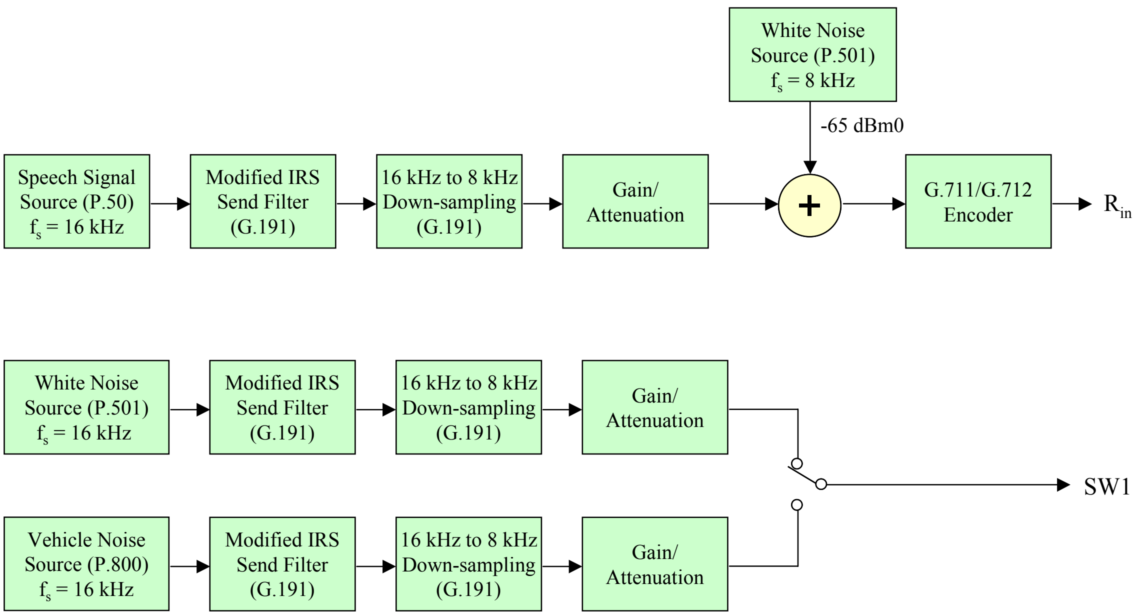


Figure 9 – Details of signal and noise source pre-processing

7.3.2 Test 3.1 – Adaptation test

This test is meant to ensure that the AEC adapts rapidly to all combinations of input signal levels and echo paths, and that the returned echo level is sufficiently low.

The adaptation characteristics are measured by making use of the artificial speech signals defined in [ITU-T P.50]. A single period of the test signal consists of the concatenation of the female and male artificial speech signals, and has a duration of about 22 seconds. The test signal should be filtered by an appropriate filter to model the frequency response of the handset. The modified IRS send filter, according to Annex D of [ITU-T P.830] is recommended for this purpose. This is available as part of the [b-ITU-T G.191] software tool library. After the filtering, the test signal should be scaled to a desired level L_{Rin}. Finally white noise, defined in clause 5.1.2 of [ITU-T P.501], with the level of approximately –65 dBm0 should be added to it to create a low noise floor and also replace periods of digital zeros.

NOTE 1 – When using the [b-ITU-T G.191] software tools, to avoid an excessive saturation of the [ITU-T P.50] signal at high input levels, it is important to carry out IRS filtering and down-sampling before the gain. IRS filtering attenuates the signal, and having a gain before the filtering would result in even more clipping because the original [ITU-T P.50] signal would be scaled to over –4 dBm0. A noise floor of –65 dBm0 represents electrical noise which has typically a flat spectrum. White noise should be inserted after the gain/attenuation of [ITU-T P.50].

The measurement of the adaptation characteristics starts at t_0 simultaneously with the start of the test signal as shown in Figure 10. The test signal should be continuously repeated through the test. The level at R_{in} is L_{Rin} that is measured over a single period of the test signal using the RMS method described in clause 6.2

NOTE 2 – L_{Rin} should be measured after the [b-ITU-T G.711]/[b-ITU-T G.712] encoder shown in Figure 9. For high levels of L_{Rin}, the [ITU-T P.50] signal will be clipped. This will make the level adjustment process non-linear and may require several attempts to achieve the desired level.

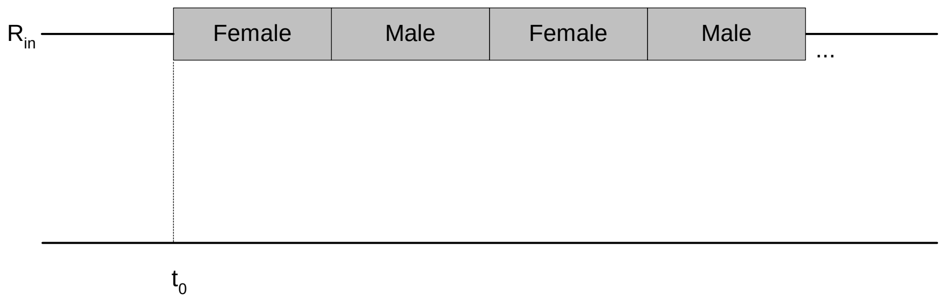


Figure 10 – Measurement time relationships

Requirement

For all values $L_{Rin} \geq -30$ dBm0 and ≤ -4 dBm0 and for all values of ERL ≥ 30 dB and echo path delay, $\Delta_1 \leq t_d \leq \Delta_2$ ms, the level L_{Sout} at S_{out} should be less than or equal to that shown in Figure 11. The level L_{Sout} is measured as a function of time using a meter conforming to the characteristics described in clause 6.4.1.2.1 of [ITU-T G.168].

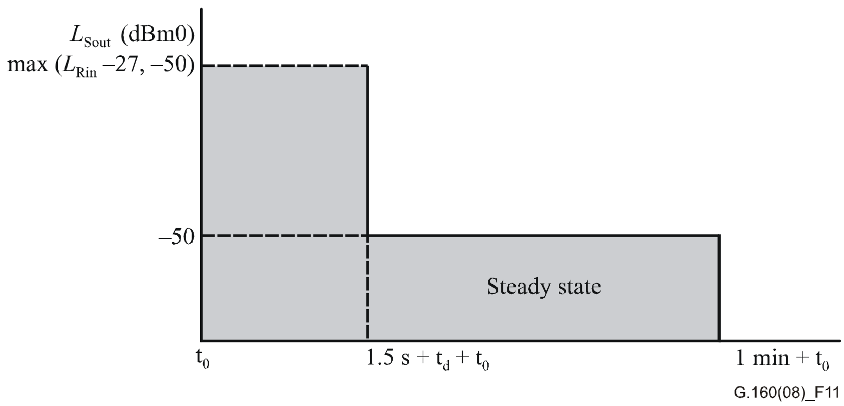


Figure 11 – Adaptation characteristics

7.3.3 Test 3.2 – Comfort noise test (level test mandatory, spectral matching optional)

This test is meant to ensure that the AEC is able to provide a comfort noise signal at S_{out} that matches noise received at S_{in}. It also tests whether the canceller is able to adjust the level and magnitude spectrum of this comfort noise signal to compensate for changes in the level and in the magnitude spectrum of input noise. Magnitude spectra should be measured for three one-octave sub-bands, having the centre frequencies of 500 Hz, 1000 Hz, and 2000 Hz ([IEC 61260] clause 4.4.3, class 0). The steps of this test should be applied in sequence.

NOTE 1 – Examples of one-octave sub-band filters suitable for this test are given in Appendix IV.

A white noise source, defined in clause 5.1.2 of [ITU-T P.501], is applied at $R_{in.}$ Another white noise source, defined in clause 5.1.2 of [ITU-T P.501], and an internal moving vehicle noise source, defined in clause A.1.1.2.2.2 of [ITU-T P.800], with levels (measured at S_{gen}) between –50 dBm0 and –30 dBm0 are applied at S_{gen} in sequence. The noise signals should be filtered by an appropriate filter (as per the adaptation test) to model the frequency response of the handset. The modified IRS filter according to Annex D of [ITU-T P.830] is recommended for this purpose. This is available as part of the [b-ITU-T G.191] Software Tool Library. After filtering, the noise signals should be scaled to the desired level at $L_{Sgen,}$ see Figure 9.

The timing relationship of the various signals for this test is shown in Figure 12. The test consists of the application of the following steps in sequence. The steps are associated with different regions in Figure 12.

Region-1

- Apply an internal vehicle noise source at S_{gen}. Set L_{Sgen} to a value between –50 dBm0 and –30 dBm0.
- Set L_{Rin} to silence (< –60 dBm0).
- Measure, L_{Sin}, denoted as $L_{Sin,\,1}$, and the average magnitude spectrum of the noise signal at S_{in} during the silence period.

Region-2

- Apply a White noise source at S_{gen} for 6 s during the silence period. Set L_{Sgen} to a value between –50 dBm0 and –30 dBm0. There should be an absolute difference of 10 dB compared to the vehicle noise level. This is to verify that the AEC will adapt to changes of noise level. It also reduces the number of test cycles, since two levels are tested during a single test run.
- Measure L_{Sin}, denoted as $L_{Sin,\,2}$, and the average magnitude spectrum of the White noise signal at S_{in} during the silence period.

Region-3

- Set L_{Rin} to –3 dBm0.
- Measure the average magnitude spectrum at S_{out} and L_{Sout}, denoted as $L_{Sout,\,2}$, after 3 s.

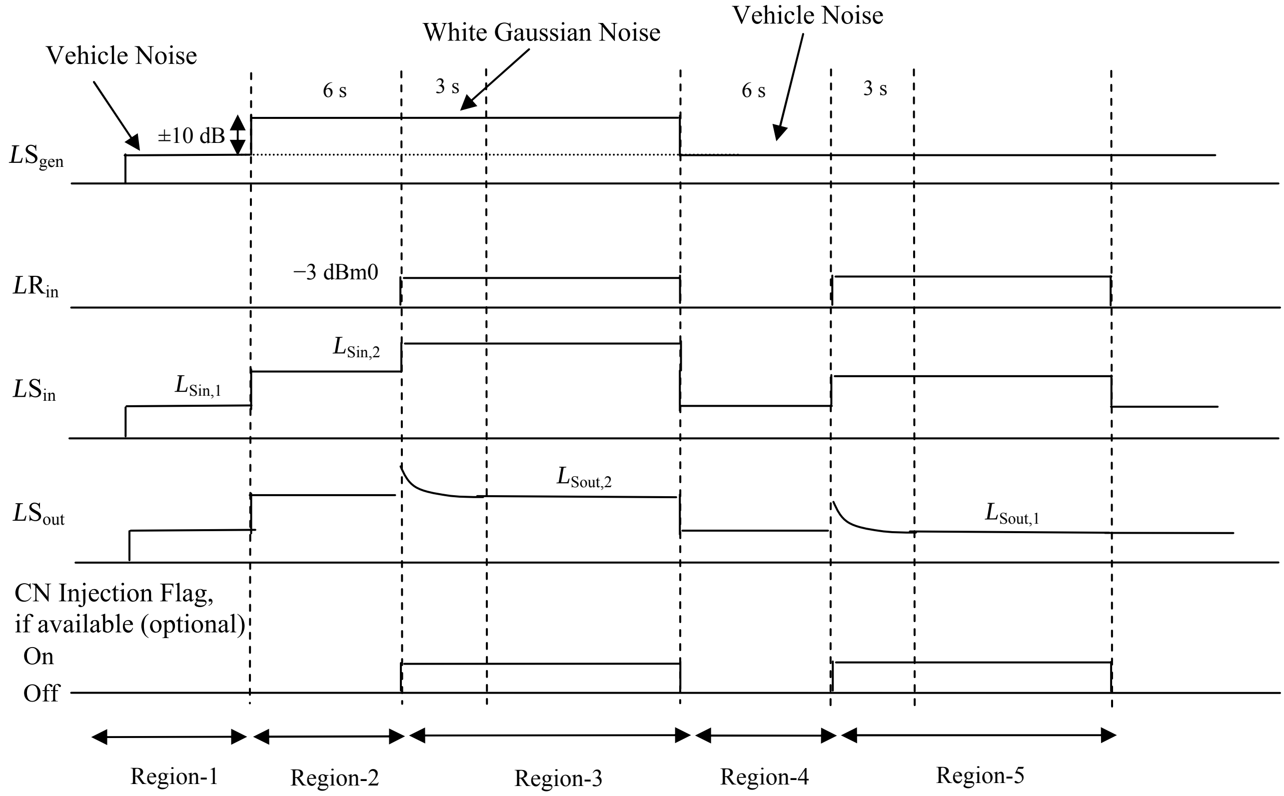
Region-4

- Set L_{Rin} to silence (< –60 dBm0) again.
- Apply the vehicle noise source at S_{gen} for 6 s during the silence period. Set L_{Sgen} to the value used in Region-1.

Region-5

- Set L_{Rin} to –3 dBm0.
- Measure the average magnitude spectrum at S_{out} and L_{Sout}, denoted as $L_{Sout,\,1}$, after 3 s.

NOTE 2 – If a flag is available that forces the injection of comfort noise at S_{out}, then this flag may optionally be activated in performing steps 5 and 7 above. Activating the flag will ensure that comfort noise is injected continuously without interruption so that the quality of the comfort noise can be more easily measured.



NOTE – The S_{in} signal is the sum of the S_{gen} signal and the acoustic echo resulting from R_{in}. Hence, the level of S_{in} is higher than the level of S_{gen} when R_{in} is not silence.

Figure 12 – Signal level timing relationships for the AEC comfort noise test

Requirement

For all values of L_{Sgen}, and for ERL = 20 dB and echo path delay, $\Delta_1 \le t_d \le \Delta_2$ ms, L_{Sout} should be within 4.0 dB of L_{Sin}, when measured over a 1.4 s window. That is, $L_{Sin,2} - 4.0$ dB $\le L_{Sout,2} \le L_{Sin,2} + 4.0$ dB and $L_{Sin,1} - 4.0$ dB $\le L_{Sout,1} \le L_{Sin,1} + 4.0$ dB.

Optional: L_{Sout} in each of the three sub-bands should be within 6 dB of L_{Sin} in the same sub-bands when measured over a 1.4 s window. Moreover, the overall level value and the level values of the sub-bands should hold as long as the noise source at S_{gen} stays the same, and sequential 1.4 s-measurements should be made over a period of 7 seconds to check this.

NOTE 3 – Any discontinuous transmission (DTX) feature of the low bit rate codecs should be turned off during this test. This is because the use of white noise as the test signal may cause the DTX feature to operate causing non-linearities in the echo path.

NOTE 4 – An ERL of 20 dB, an R_{in} signal of white noise and an R_{in} signal level of –3 dBm0 have been chosen to exercise the comfort noise feature of the AEC device. These conditions may not represent the scenario that an AEC device will encounter in practice.

Appendix I

Information on the function and location of VEDs in mobile networks

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

I.1 VED functionality

Figure I.1 shows the individual functions that may be included in a VED. This diagram is for illustration purposes, and different VEDs may not contain all of these functions. Furthermore, the location of each function in the VED with respect to the others may be different. For example, the NR and AEC functions may be transposed. Some of the functions are covered by other Recommendations as indicated.

NOTE – For a VED to be TFO compliant, all elements in Figure I.1 need to be transparent to TFO.

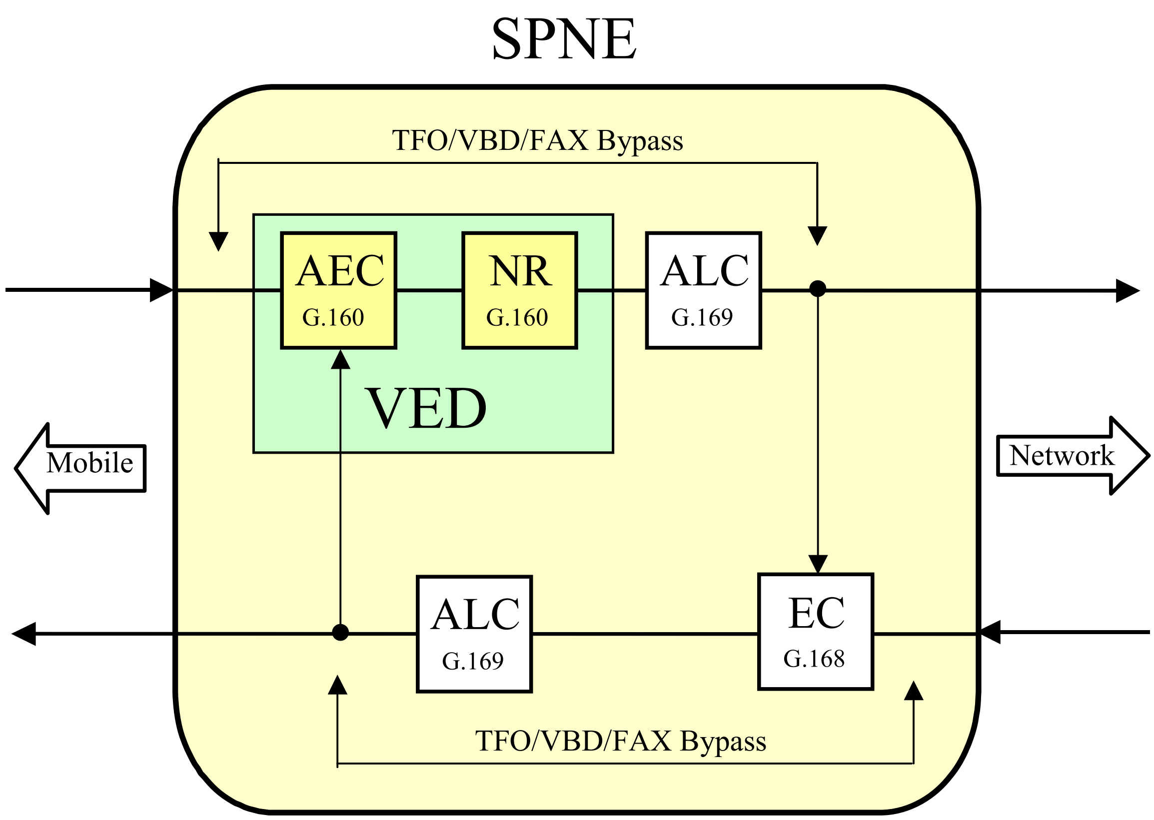


Figure I.1 – Example block diagram showing the presence of a VED in a SPNE

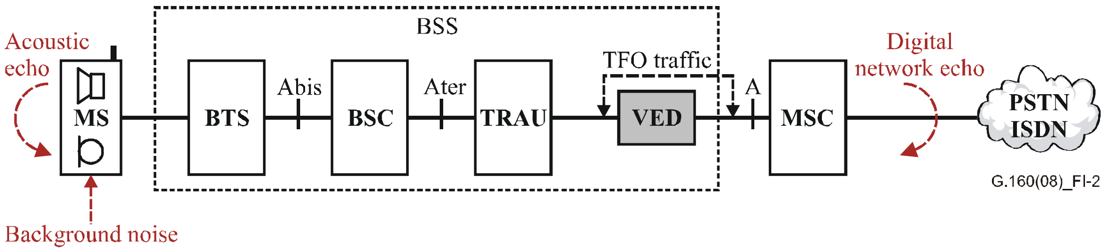
I.2 VED location

The following diagrams show examples of the possible locations of a VED in 2G and 3G mobile network architectures.

If external disabling is supported, the VED function should provide 64 kbit/s bit-sequence integrity in the externally disabled state. In this case, depending on the locations of a VED in 2G and 3G mobile network architectures, either a functional entity of the radio subsystem or a functional entity of the core network will control the signal processing functions contained in a VED on a per call basis.

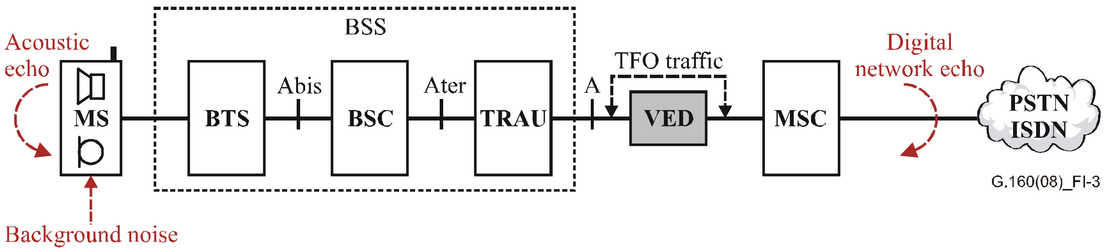
[b-ITU-T Q.55] and [b-ITU-T Q.56] describe such control interfaces that may be used for a TDM bearer environment.

[b-ITU-T Q.115.0] with the H.248 SPNE control package describes the protocol for the control of SPNEs associated with a media gateway.



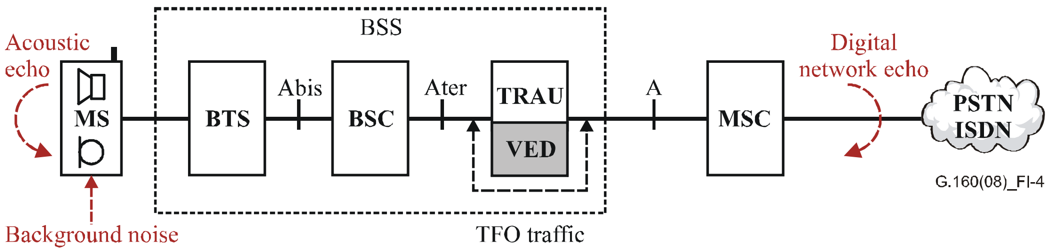
NOTE – Dedicated configuration located on the A interface. If TFO protocol is not terminated, the VED will provide a transparent channel for each connection in TFO mode. The referring OMC is the BSS-OMC.

Figure I.2 – Block diagram 2G mobile network configuration 1



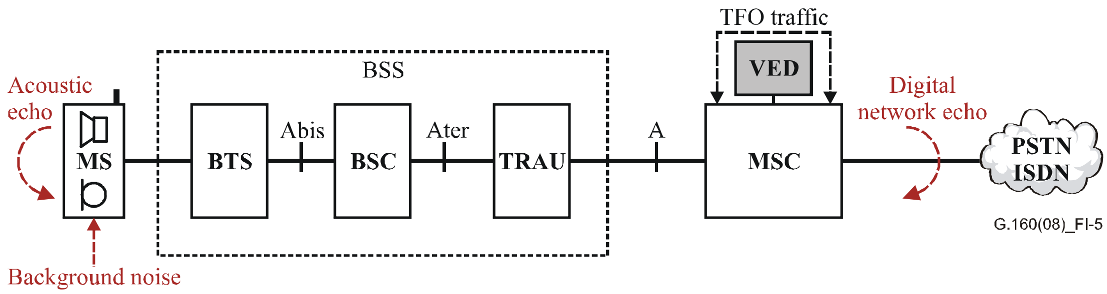
NOTE – Dedicated configuration located on the A interface. If TFO protocol is not terminated, the VED will provide a transparent channel for each connection in TFO mode. Same location as Figure I.2, but, in this case, the referring OMC is the MSC-OMC.

Figure I.3 – Block diagram 2G mobile network configuration 2



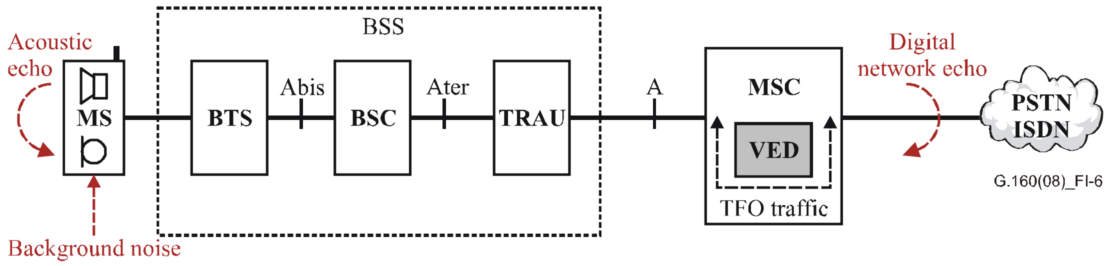
NOTE – Dedicated configuration located on the TRAU equipment. No interaction with TFO protocol. The referring OMC is the BSS-OMC.

Figure I.4 – Block diagram 2G mobile network configuration 3



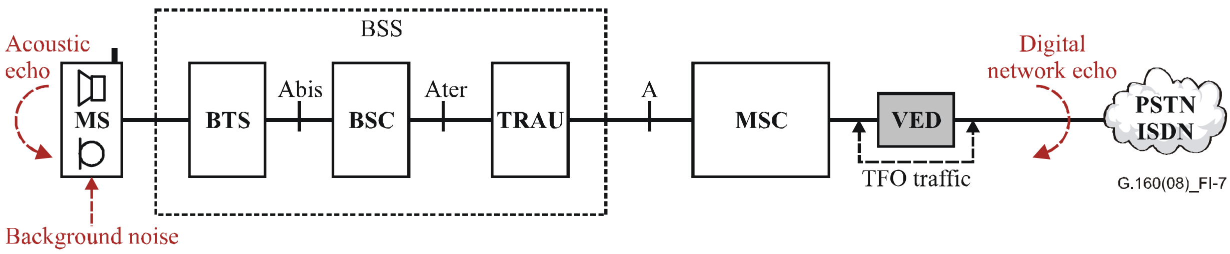
NOTE – Dedicated or pool configuration with external equipment connected to MSC. If TFO protocol is not terminated, the VED will provide a transparent channel for each connection in TFO mode. The referring OMC is the MSC-OMC.

Figure I.5 – Block diagram 2G mobile network configuration 4



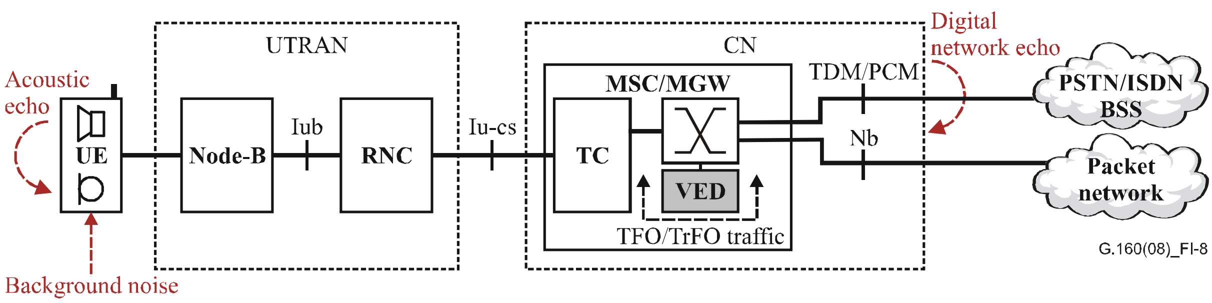
NOTE – Dedicated or pool configuration located inside MSC. If TFO protocol is not terminated, the VED will provide a transparent channel for each connection in TFO mode. The referring OMC is the MSC-OMC.

Figure I.6 – Block diagram 2G mobile network configuration 5



NOTE – Dedicated located on the PSTN interface. This configuration does not allow voice enhancement for mobile-to-mobile calls. No interaction is possible with TFO protocol. The referring OMC is the MSC-OMC.

Figure I.7 – Block diagram 2G mobile network configuration 6

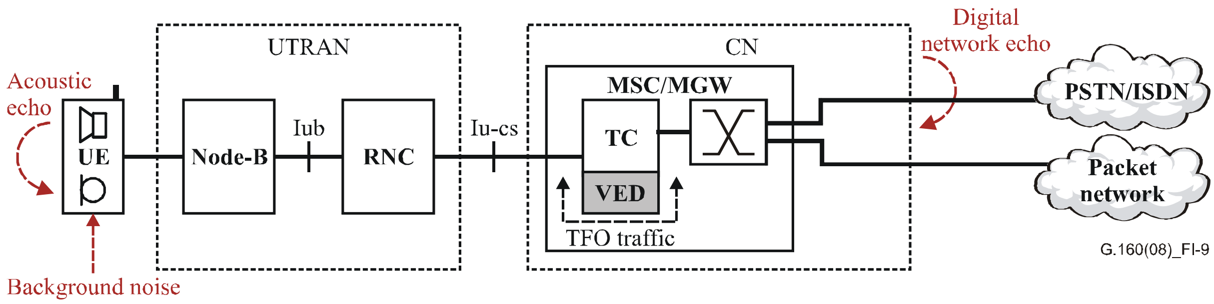


NOTE 1 – VED in dedicated or pool configuration located inside MGW. The referring OMC is the MGW-OMC.

NOTE 2 – Nb interface, TFO and TrFO are only supported in 3GPP Release 4 and onwards networks. The term MGW is used with Release 4 and onwards networks.

NOTE 3 – Transcoder (TC) is not present during TrFO operation.

Figure I.8 – Block diagram of VED location in UMTS 3G mobile network configuration 1

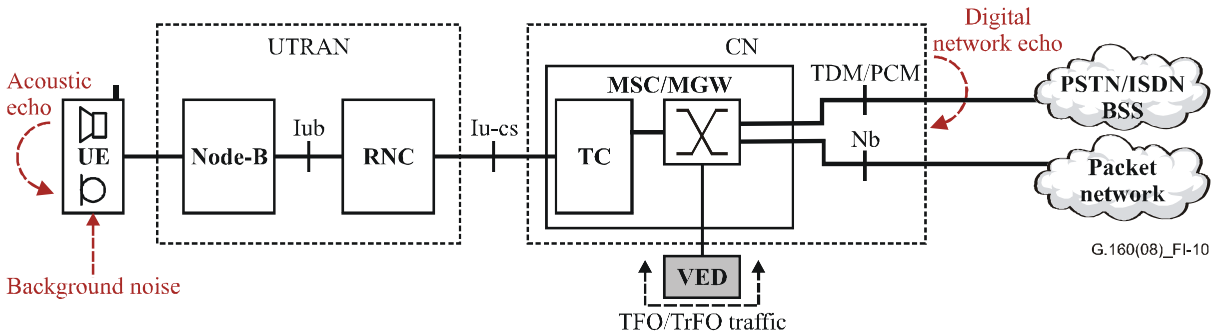


NOTE 1 – Dedicated configuration located on the TC equipment. The referring OMC is the MGW-OMC. In this scenario, the VED may be located anywhere within the TC, and its exact location will determine whether or not it needs to implement TFO detection.

NOTE 2 – Nb interface, TFO and TrFO are only supported in 3GPP Release 4 networks and onwards. The term MGW is used with Release 4 and onwards networks.

NOTE 3 – Transcoder (TC) is not present during TrFO operation.

Figure I.9 – Block diagram of VED location in UMTS 3G mobile network configuration 2

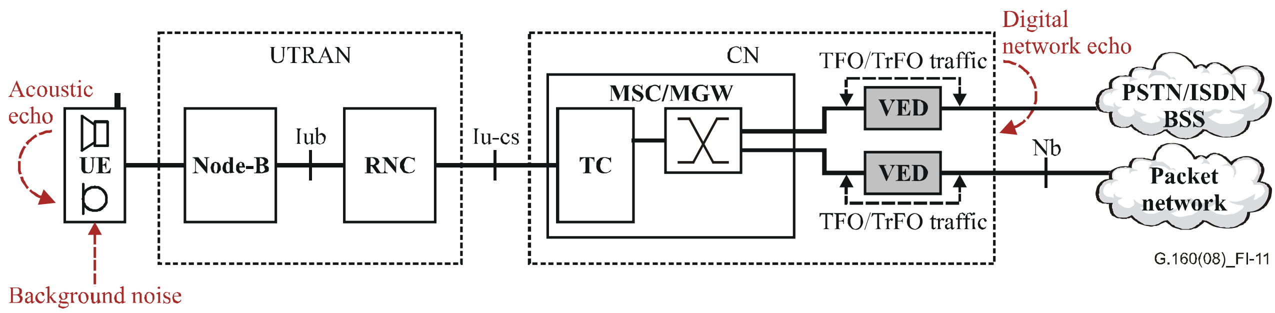


NOTE 1 – VED in dedicated or pool configuration with external equipment connected to MGW. The referring OMC is the MGW-OMC.

NOTE 2 – Nb interface, TFO and TrFO are only supported in 3GPP Release 4 and onwards networks. The term MGW is used with Release 4 and onwards networks.

NOTE 3 – Transcoder (TC) is not present during TrFO operation.

Figure I.10 – Block diagram of VED location in UMTS 3G mobile network configuration 3



NOTE 1 – Dedicated configuration located on the PSTN interface. On packet-oriented interface, there is the possibility to have compressed speech. In such a case, the VED will provide a transparent channel. The referring OMC is the MSC/MGW-OMC.

NOTE 2 – Nb interface, TFO and TrFO are only supported in 3GPP Release 4 and onwards networks. The term MGW is used with Release 4 and onwards networks.

NOTE 3 – Transcoder (TC) is not present during TrFO operation.

Figure I.11 – Block diagram of VED location in UMTS 3G mobile network configuration 4

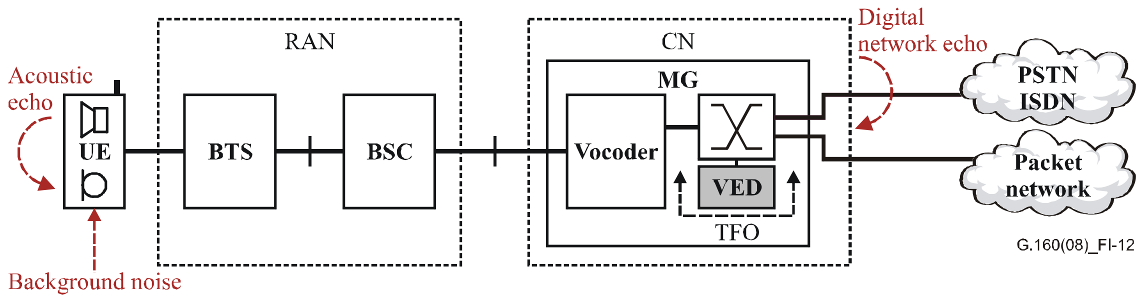


Figure I.12 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 1

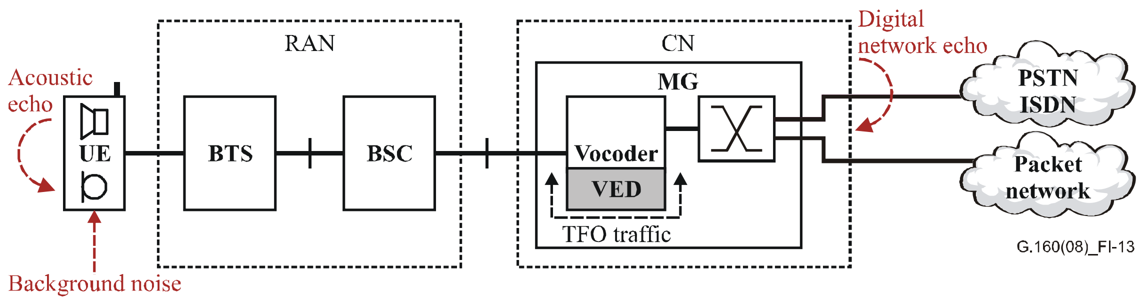


Figure I.13 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 2

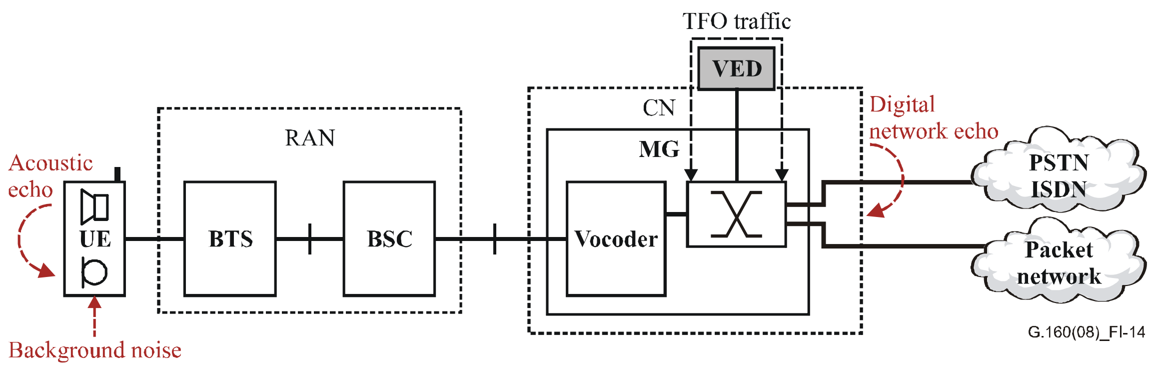


Figure I.14 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 3

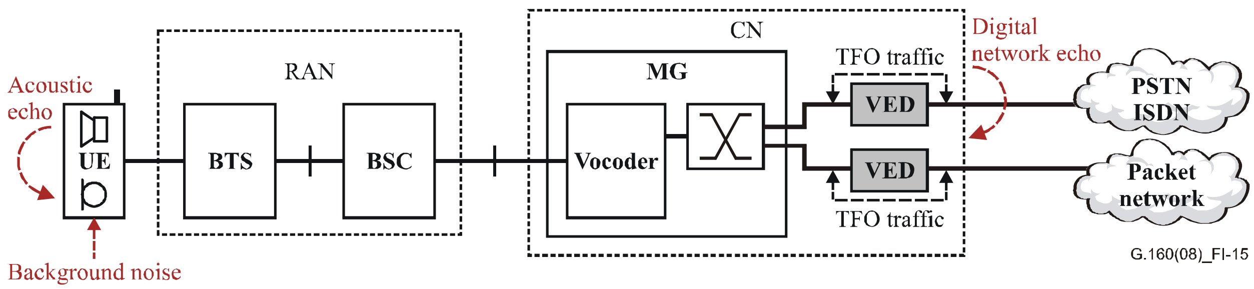
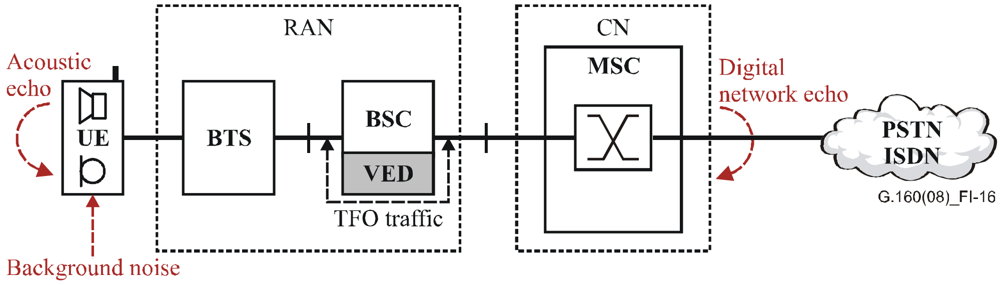
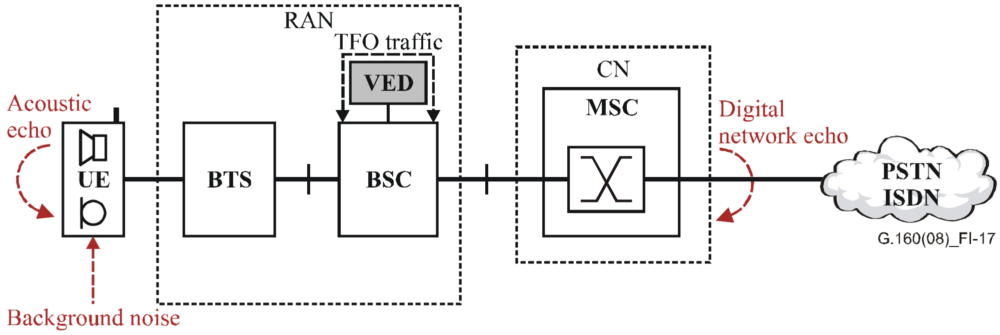


Figure I.15 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 4



NOTE – The CDMA selection function and vocoder are both in BSC in this case.

Figure I.16 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 5



NOTE – The CDMA selection function and vocoder are both in BSC in this case.

Figure I.17 – Block diagram of VED location in CDMA2000 3G mobile network, configuration 6

I.3 CTM operation

Figure I.18 show examples of the possible locations of CTM adapters in 2G and 3G mobile network architectures.

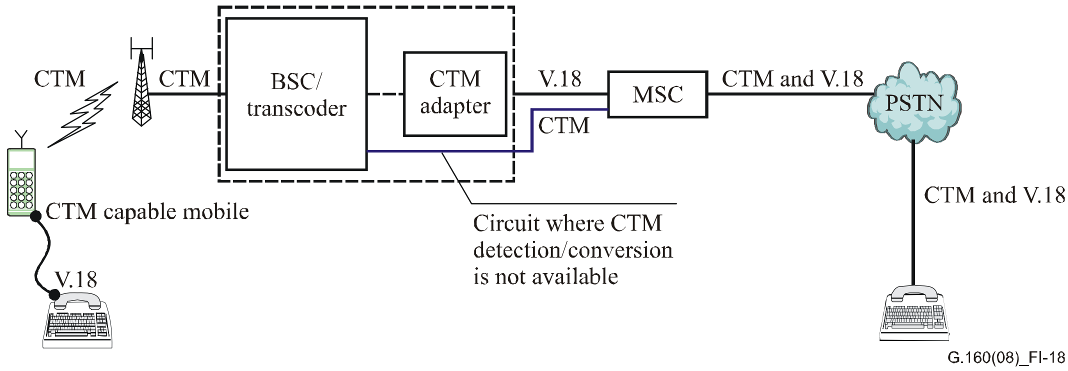


Figure I.18 – CTM adapter present on specific A-interface circuits – 2G mobile network

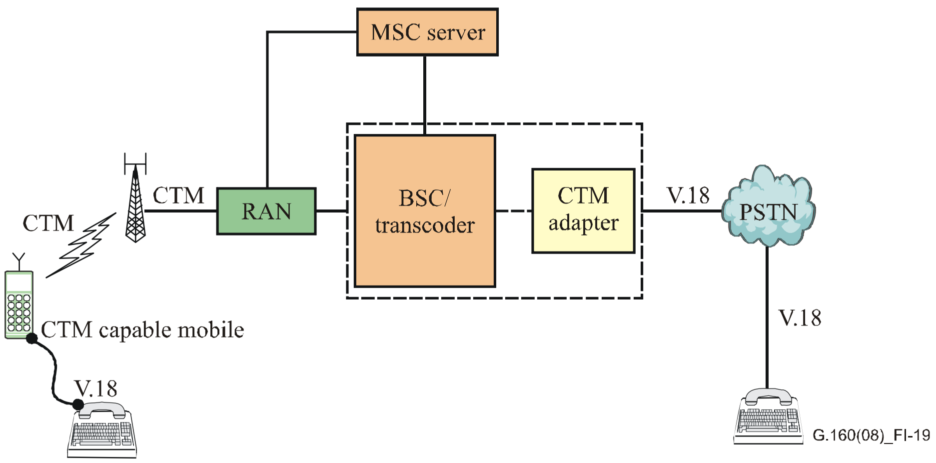


Figure I.19 – CTM adapter present in media gateway – 3G mobile network

Appendix II

Objective measures for the characterization of the basic functioning of noise reduction algorithms

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

II.1 Introduction

This appendix presents an objective methodology for characterizing the basic effect of noise reduction (NR) methods. Three objective measures are specified for characterizing NR solutions. Recommendation target value ranges in specified conditions are given for the measures to serve as basic guidelines for proper functioning of NR methods.

The methodology monitors the basic functioning of an NR solution in terms of signal-to-noise ratio improvement (*SNRI*) and total noise level reduction (*TNLR*). While *SNRI* is measured during speech activity, focusing on the effect of NR on speech signal, *TNLR* estimates the overall level of noise reduction, experienced both during speech and speech pauses. In addition, a delta measurement (*DSN*) is computed to reveal speech attenuation or undesired speech amplification caused by an NR solution.

The proposed methodology is a further development of one included in [b-ETSI TS 126 077] and [b-TIA/EIA-IS-853]. A detailed description of the earlier methodology is incorporated in the named specifications.

NOTE – "Noise suppression" in the named standard specifications refers to methods that reduce the background noise impact in the desired speech signal, with no limitation as to their capability of reducing the noise level either only outside of or also during speech activity. The terms "noise suppression" and "noise reduction" are used interchangeably in this appendix.

II.2 Notations

The following notations are used in this appendix:

- The operator NR(·) corresponds to applying the NR algorithm on the input speech.
- The clean speech original signals are referred to as s_i, i = 1 to I, where I is the total number of speech files used.
- The noise original signals are referred to as n_j, j = 1 to J, where J is the total number of noise files used.
- The noisy speech test signals are referred to as $d_{ij} = s_i + \beta_{ij}(\text{SNR})\, n_j$, i = 1 to I, j = 1 to J, where d_{ij} is built by adding s_i and n_j with a pre-specified SNR as presented below.
- The processed signal is referred to as $y_{ij} = \text{NR}(d_{ij})$.
- The reference signal in the calculations shall be the noisy speech test signal d_{ij} itself.
- The notation $Log(\cdot)$ indicates the decimal logarithm.
- $\beta_{ij}(\text{SNR})$ is the scaling factor to be applied to the background noise signal n_i in order to have a ratio **SNR** (in dB) between the clean speech signal s_i and n_j. The use of the scaling factor is explained below in clause II.3, in conjunction with the generation of test samples.
- Frames are classified by their average speech power (high, medium, low); these classes are named $\{k_{sph}\}$, $\{k_{spm}\}$, $\{k_{spl}\}$.
- Frames containing only noise form a separate class, named $\{k_{nse}\}$.

- The determination of which frames contain active speech is to be carried out with reference to [b-ITU-T P.56] and is related to the classification of the frames into the presented categories, or *speech power classes* which is explained below.

II.3 Test signals

The test material should manifest at least the following extent:

- Clean speech utterance sequences: 6 utterances from 4 speakers – 2 male and 2 female – totalling 24 utterances
- Noise sequences:
  - Car interior noise, 120 km/h, fairly constant power level
  - Street noise, slowly varying power level

Special care should be taken to ensure that the original samples fulfil the following requirements:

- The clean speech signals are of a relatively constant average (within sample, where 'sample' refers to a file containing one or more utterances) power level
- The noise signals are of a short-time stationary nature with no rapid changes in the power level and no speech-like components

The test signals should cover the following background noise and SNR conditions:

- Car noise at 6 dB, 12 dB and 18 dB SNR
- Street noise at 6 dB, 12 dB and 18 dB SNR

The samples should be digitally filtered before NR processing by an appropriate filter to become representative of a real cellular system frequency response. This filtering should be carried out before the scaling of the samples to be discussed below.

NOTE – Appropriate digital filters to make the test material representative of a real cellular mobile station frequency response, which are available as part of the G.191 software tool library are the following:

- The modified IRS filter according to [b-ITU-T P.48].
- The MSIN (mobile station in) filter. The transfer function of the MSIN filter has been defined based on measurements by British Telecom. Even though the MSIN filter attempts mainly to describe the spectral modifications of background noise, in the ETSI STC SMG11#8 Plenary (Jan 1999; Helsinki, Finland), it has been discussed that this filter may also be applied to the speech signal.

The test samples corresponding to the named SNR conditions shall be generated according to the following procedure:

- The clean speech material shall be scaled to the active speech level –26 dBov with the [b-ITU-T P.56] speech voltmeter, one file at a time, each file including a sequence of one (to four) utterance(s) from one speaker.
- A silence period of 2 s is inserted at the beginning of each of the resulting files to make up augmented clean speech files.
- Within each noise type and SNR condition, a noise sequence is selected for every augmented clean speech utterance file, each with the same length as the corresponding speech file, and each noise sequence is stored in a separate file.
- Each of the noise sequences is scaled to a dBov level leading to the SNR condition corresponding to the β_{ij}(SNR) value in each of the test cases by applying the RMS level based scaling according to [b-ITU-T P.56].
- Augmented clean speech and noise files are then added sample by sample (overload values being clipped) to create the needed set of noisy speech files with predefined SNR, sample-rate, bits/sample, file-length.

II.4 Objective measures for characterization of NR algorithm effect

II.4.1 Categorization of speech frames into speech power classes

The objective metric for measuring SNR improvement due to an NR algorithm, or *SNRI*, that is presented below, is calculated separately in three groups of speech signal frames that represent power-gated constituents of active speech signal. These groups of frames are called *speech power classes* or just *power classes* in the following. Hence, the *SNRI* measure is calculated separately in frames of high, medium and low power. These categories are used to characterize the effect of the NR processing on speech, allowing the distinction of the effect on strong, medium and weak speech. In addition to calculating the SNR improvement separately on the three categories, they are used to form an aggregate measure.

In addition to the division into three frame classes containing speech activity, a fourth frame class is needed to be able to calculate the power levels in the noise periods lying amongst speech activity. These power levels are used both in the calculation of the *SNRI* metrics and in the calculation of the overall change in noise power level (*TNLR*) due to NR processing.

A frame length of 10 ms, i.e., 80 samples (for speech sampled at 8 kHz), is used in the analysis since it has been found the most efficient to describe changes in the signal caused by NR processing.

To determine which frames belong to high, medium and low power classes of active speech and which present pauses in the speech activity (noise only), the active speech level (in dB) (sp_lvl) of the noise free speech $s_i(n)$ is first determined according to [b-ITU-T P.56]. Thereafter, the frames are classified into the four classes based upon a comparison of the power in each frame to predefined threshold values relative to the active speech level. As a result, four number sequences are obtained to represent the three speech classes and the speech pause class, respectively: $\{k_{sph}\}$, $\{k_{spm}\}$, $\{k_{spl}\}$, $\{k_{nse}\}$.

The threshold values to be related to the active speech level are presented in Table II.1. To be able to apply the threshold values given here, the speech should be first normalized at –26 dBov using the [b-ITU-T P.56] speech voltmeter. Figure II.1 presents an example of the classification of a clean speech sample into power classes.

Table II.1 – Power threshold values for speech power classes

Threshold	Explanation	Value
th_h	Lower bound for high speech power class	–1 dB
th_m	Lower bound for medium speech power class	–10 dB
th_l	Lower bound for low speech power class	–16 dB
th_nh	Higher bound for speech pause class	–25 dB
th_nl	Lower bound for speech pause class	–40 dB

The following notes are made concerning the formulation of the speech power classes:

- The lower bound for the power of the noise-only class of frames is motivated by a desire to restrict the analysis to noise frames that are among or close to speech activity, hence excluding long pauses from the analysis. This makes the analysis concentrate increasingly on the effects encountered during speech activity, applying to the computation of both *SNRI* and *DSN*. On the other hand, the lower bound is not used in the computation of *TNLR*, as this should indicate the overall reduction in noise level experienced both during short speech pauses amongst speech activity and long speech pauses when no speech is immediately continuing.

- In low SNR conditions, the noise power level may occur to be higher than the lower bound of some of the speech power classes. However, even in this case, the information of the effect on the low power portions of speech may be informative.

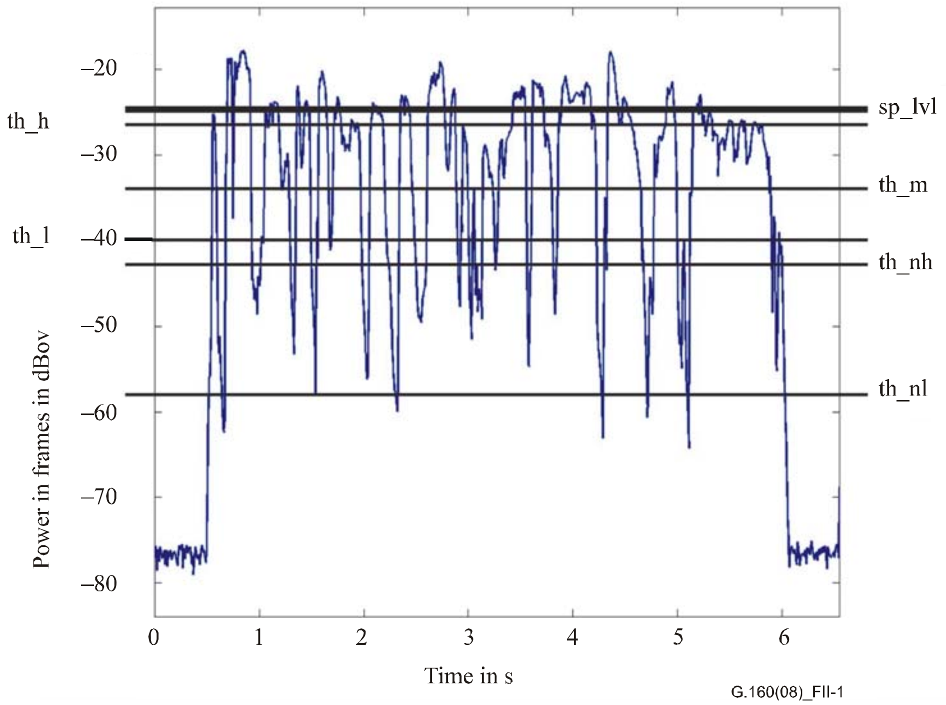


Figure II.1 – Power in frames of clean speech as a function of time – Classification into power classes

II.4.2 Assessment of SNR improvement – *SNRI*

The SNR improvement measure, *SNRI*, measures the SNR improvement achieved by the NR algorithm. As said above, SNR improvement is calculated in three *speech power classes* to obtain an evaluation of the effect separately for strong, medium and weak speech. An aggregate measure is obtained in addition via weighted averaging.

The calculation is here presented for the high power speech class. A similar procedure is used to obtain the respective figures for the medium and low power speech classes.

For each background noise condition *j* and for each speech sample *i*, construct a noisy speech input signal d_{ij} as follows:

$$d_{ij}(n) = \beta_{ij}\ n_j(n) + s_i(n),$$

where β_{ij} depends on the SNR condition according to the procedure described above. Now, the noise suppressed output speech signal can be denoted as:

$$y_{ij} = \mathrm{NR}\ (d_{ij})$$

In this way, the SNR of the output and input speech signals can be expressed as:

$$SNR\mathrm{out\_h}_{ij} = 10 \cdot \mathrm{Log}\left\{ \max\left[\xi, \frac{10^{\frac{1}{K_{sph}} \sum_{l=1}^{K_{sph}} \mathrm{Log}\left(\xi + \sum_{n} y_{ij}^2(l,n)\right)}}{10^{\frac{1}{K_{nse}} \sum_{m=1}^{K_{nse}} \mathrm{Log}\left(\xi + \sum_{p} y_{ij}^2(m,p)\right)}} - 1 \right] \right\} \qquad \text{(II.1)}$$

$$SNR\mathrm{in\_h_{ij}} = 10 \cdot \mathrm{Log}\left\{ \max\left[\xi, \frac{10^{\frac{1}{K_{sph}} \sum_{l=1}^{K_{sph}} \mathrm{Log}\left(\xi + \sum_{n} d_{ij}^{2}(l,n) \right)}}{10^{\frac{1}{K_{nse}} \sum_{m=1}^{K_{nse}} \mathrm{Log}\left(\xi + \sum_{p} d_{ij}^{2}(m,p) \right)}} - 1 \right] \right\} \quad \text{(II.2)}$$

$$SNRI\_h_{ij} = SNRout\_h_{ij} - SNRin\_h_{ij}, \quad \text{(II.3)}$$

where K_{sph} is the total number of speech frames and K_{nse} is the total number of noise frames lying amongst speech activity and $\xi > 0$ is a constant that should be set at 10^{-5}.

The summation with respect to indices *n* and *p* is carried out in frames of 80 samples. The index, *n*, relates to speech frames, while the index, *p*, relates to noise frames where the frame power is between the higher (th_nh) and lower (th_nl) bounds for speech pause class (see Table II.1).

$SNRI\_m_{ij}$ and $SNRI\_l_{ij}$ are computed correspondingly for medium and low power frames.

Finally, the aggregate measure for signal-to-noise ratio improvement is computed as below.

$$SNRI_{\mathrm{ij}} = \frac{1}{K_{sph} + K_{spm} + K_{spl}} \left(K_{sph} \cdot \mathrm{SNRI\_h_{ij}} + K_{spm} \mathrm{SNRI\_m_{ij}} + K_{spl} \mathrm{SNRI\_l_{ij}} \right) \quad \text{(II.4)}$$

$$SNRI_{\mathrm{j}} = \frac{1}{\mathrm{I}} \sum_{i=1}^{I} SNRI_{\mathrm{ij}} \quad \text{(II.5)}$$

$$SNRI = \frac{1}{\mathrm{J}} \sum_{j=1}^{J} SNRI_{\mathrm{j}} \quad \text{(II.6)}$$

In addition, measures for the SNR improvement in the high, medium and low power speech classes (SNRI_h, SNRI_m, SNRI_l, respectively) shall be recorded based on the following formulae:

$$SNRI\_h = \frac{1}{J} \sum_{j=1}^{J} SNRI\_h_j = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{I} \sum_{i=1}^{I} SNRI\_h_{ij} \quad \text{(II.7)}$$

$$SNRI\_m = \frac{1}{J} \sum_{j=1}^{J} SNRI\_m_j = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{I} \sum_{i=1}^{I} SNRI\_m_{ij} \quad \text{(II.8)}$$

$$SNRI\_l = \frac{1}{J} \sum_{j=1}^{J} SNRI\_l_j = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{I} \sum_{i=1}^{I} SNRI\_l_{ij} \quad \text{(II.9)}$$

It is, in addition, informative to record separately the noise type specific SNR improvement measures, namely, *SNRI_hj, SNRI_lj, SNRI_mj* and *SNRIj* for each background noise condition *j*.

II.4.3 Assessment of total noise level reduction – *TNLR*

The total noise level reduction measure, or *TNLR,* relates to the capability of the NR method to attenuate the background noise level measured during both speech activity and speech pauses. As the number of frames in the speech pause class during speech activity is typically relatively small compared to the number of frames during long speech pauses, *TNLR* mainly measures the capability of an NR to reduce noise during long speech pauses.

The *TNLR* measure is calculated as follows:

For each background noise condition *j* and for each speech sample *i*, construct a noisy input signal d_{ij} as follows:

$$d_{ij}(n) = \beta_{ij}\ n_j(n) + s_i(n),$$

where β_{ij} depends on the SNR condition according to the procedure described above. The noise suppressed output speech signal can be denoted as:

$$y_{ij} = \text{NR}\ (d_{ij})$$

In this way, the total noise level reduction measure can be expressed as:

$$TNLR_{ij} = \frac{1}{K_{pse}} 10 \cdot \sum_{m=1}^{K_{pse}} \left[\text{Log}\left(\xi + \sum_q y_{ij}^2(m,q) \right) - \text{Log}\left(\xi + \sum_q d_{ij}^2(m,q) \right) \right] \tag{II.10}$$

Where K_{pse} is the total number of noise frames during both short speech pauses amongst speech activity and long speech pauses between utterances; and

$\xi > 0$ is a constant that should be set at 10^{-5}

$$TNLR_j = \frac{1}{\text{I}} \sum_{i=1}^{I} TNLR_{ij} \tag{II.11}$$

$$TNLR = \frac{1}{\text{J}} \sum_{j=1}^{J} TNLR_j \tag{II.12}$$

The summation with respect to index *q* is carried out in noise frames of 80 samples. The index, *q*, relates to noise frames with frame power less than the higher bound for speech pause class, th_nh (see Table II.1).

Furthermore, it is informative to record separately the noise type specific *TNLR* measures, or $TNLR_j$, for each background noise condition *j*.

II.4.4 Comparison of *SNRI* and *NPLR* – *DSN*

One of the targets in noise suppression is to maintain the power level of the speech signal so as not to attenuate the level of the speech signal together with the noise signal in the NR processing. Furthermore, an improvement in the SNR of a noisy speech signal may be obtained in a simple amplification of the speech signal, the amplification being typically only applied to the high-energy portions of the signal. As both the attenuation and amplification of speech are not desirable in NR, they should be detected.

Both the attenuation and amplification of a speech signal due to an NR can be measured by studying the balance between the SNR improvement and the noise level reduction obtained during speech activity. For this purpose, noise power level reduction, or *NPLR*, is measured during short speech pauses lying amongst speech activity to study the capability of an NR to reduce noise from speech. This measure is computed otherwise as *TNLR*, i.e., according to equations II.10, II.11 and II.12, except that the computation is only carried out between the higher (th_nh) and lower (th_nl) bounds for speech pause class specified in Table II.1. In this way, *NPLR* provides a counterpart for *SNRI*, these metrics forming together the basis for the evaluation of the balance. As *NPLR* is used as an internal variable, it is not reported.

SNRI-to-NPLR difference (*DSN*), comprising a comparison of the *SNRI* and *NPLR* measures, is therefore proposed as a measure to acquire an indication of possible speech attenuation or speech amplification produced by the tested NR method and is formulated as:

$$DSN = SNRI + NPLR \tag{II.13}$$

Note that *NPLR* is typically negative, thus the *DSN* quantity should get values close to zero. If the *NPLR* parameter assumes clearly higher absolute values than *SNRI*, making *DSN* clearly negative, the NR solution turns out to produce speech level attenuation. On the other hand, if *DSN* becomes clearly positive, revealing that *SNRI* indicates SNR improvement without a decrease in noise level, the speech signal has been amplified.

It is informative to record separately the noise type specific *DSN* measures, or DSN_j, for each background noise condition j.

II.4.5 Block diagram presentation of proposed measures

A block diagram of the calculation for the proposed metrics is provided in Figure II.2. It should be noted that an appropriate prefiltering method (see clause II.3) should be used in each studied case to make the speech data representative of real network conditions.

If the measurement of the NR device ("noise reduction") is made in a network system implementation, the interfaces for playing back the NR input signal ("*Playback IF*") and recording ("*Record IF*") the NR output signal can be utilized. The rest of the processing and calculation stages can be conducted off-line.

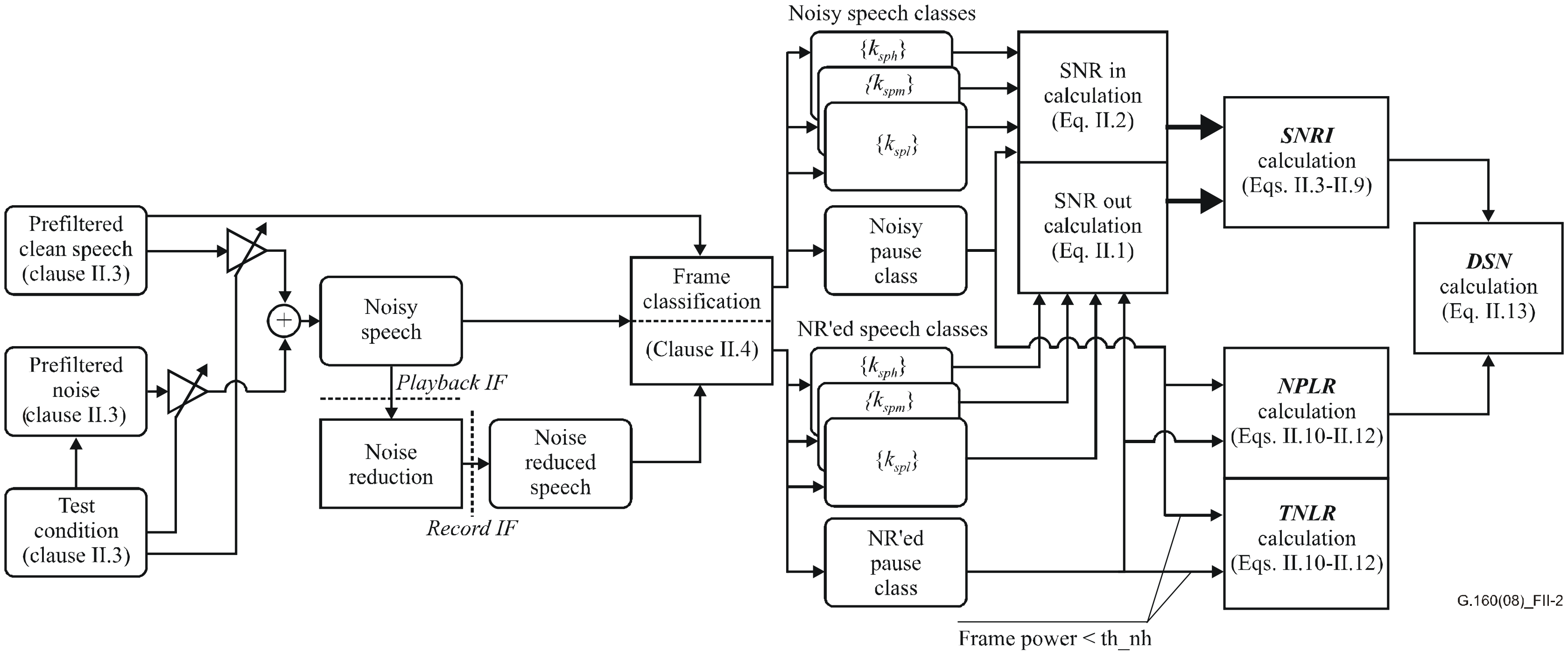


Figure II.2 – Block diagram of calculation of *SNRI*, *TNLR*, *NPLR* and *DSN* measures for noise reduction algorithms

II.5 Recommendation values for *SNRI, TNLR and DSN*

The objective measures of signal-to-noise ratio improvement (*SNRI*), total noise level reduction (*TNLR*), and *DSN* defined above are to be used to characterize the performance of the NR solution. Objectives are defined for these measures in Table II.2. These measures can be used to provide an indication of the benefit produced by an NR solution but should be interpreted as informative.

It is noted that NR solutions may function in different ways in varying usage conditions in terms of the objective metrics presented above. Therefore, for *SNRI, TNLR* and *DSN,* the performance objectives are defined for average values of these measures over all test conditions.

Table II.2 – NR algorithm performance targets in *SNRI, TNLR* and *DSN* metrics

Objective quality measure/test condition	Performance objective
SNRI *Assessment*: To be evaluated according to clause II.4 using a predefined set of material, comprising speech mixed with stationary car noise and street noise in the SNR conditions defined in clause II.3.	SNRI ≥ 4 dB as an average over all test conditions
TNLR *Assessment*: To be evaluated according to clause II.4 using a predefined set of material, comprising speech mixed with stationary car and street noise in the SNR conditions defined in clause II.3.	TNLR ≤ –5 dB as an average over all test conditions
DSN *Assessment*: To be evaluated according to clause II.4 using a predefined set of material, comprising speech mixed with stationary car noise and street noise in the SNR conditions defined in clause II.3.	–4 dB ≤ DSN ≤ 3 dB as an average over all test conditions

Meeting the presented performance objectives can be regarded as indicating that the NR method under study is capable of producing significant benefit over not using noise reduction. On the other hand, different tuning of an NR algorithm design may be beneficial in different usage scenarios, e.g., with regard to the number of speech transcoding and/or voice enhancement processing stages in the speech path. Hence, no strict requirements on the NR performance in the specified metrics are set.

Appendix III

Properties of level measurement devices

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

III.1 Introduction

Several tests, specified in this Recommendation, require measurement of signal levels. This appendix describes some properties of the devices used for level measurement.

The popular methods used for level measurement are:

1 RMS method

$$y_r(k) = \frac{1}{n} \sum_{i=k-n+1}^{k} x_i^2$$

where x_i is the i-th sample of signal x, k is the discrete time index, n is the number of samples over which the measurement is made (window length).

2 Exponential averaging as described in clause 6.4.1.2.1 of [ITU-T G.168]

$$y_e(k) = \left(1 - \exp\left(-\frac{1}{f_s\tau}\right)\right) x^2(k) + \exp\left(-\frac{1}{f_s\tau}\right) y_e(k-1)$$

where f_s is the sampling frequency and τ is the time constant of the exponential filter in seconds.

In the following, it is assumed that the sampling frequency f_s = 8 kHz.

In both methods, the signal x can be first filtered to extract the speech band. One such a prefilter is specified in clause 6.4.1.2.1 of [ITU-T G.168] and referenced to as frequency-weighting network (FWN). The frequency-weighting network is used to provide a greater attenuation of the frequencies outside the band of interest, i.e., the speech band. Frequency response of FWN vanishes at zero frequency, and it therefore mitigates DC effects due to the bias induced by A-law encoding.

Both of the measurement devices can be seen as linear (low pass) filters applied to squared signal samples, possibly preceded by the frequency-weighting network, see Figure III.1.

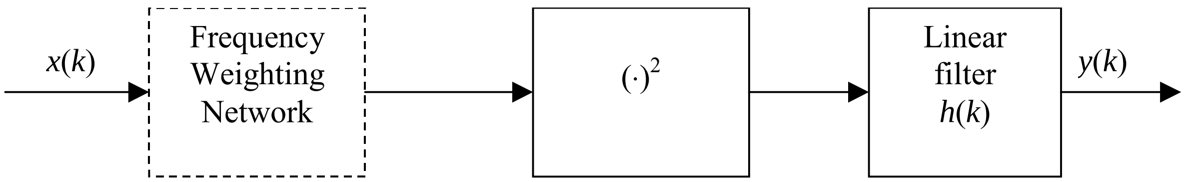


Figure III.1 – Level measurement device

The impulse response, $h(k)$, is in the case of RMS method:

$$h(k) = \begin{cases} \frac{1}{n}, & 0 \le k < n \\ 0, & \text{otherwise,} \end{cases}$$

and in the case of exponential averaging:

$$h(k)=\begin{cases}\left(1-\exp\left(-\frac{1}{f_s\tau}\right)\right)\exp\left(-\frac{k}{f_s\tau}\right), & k\geq 0\\ 0, & \text{otherwise.}\end{cases}$$

The impulse responses are shown for $n=\tau f_s=10$ in Figure III.2.

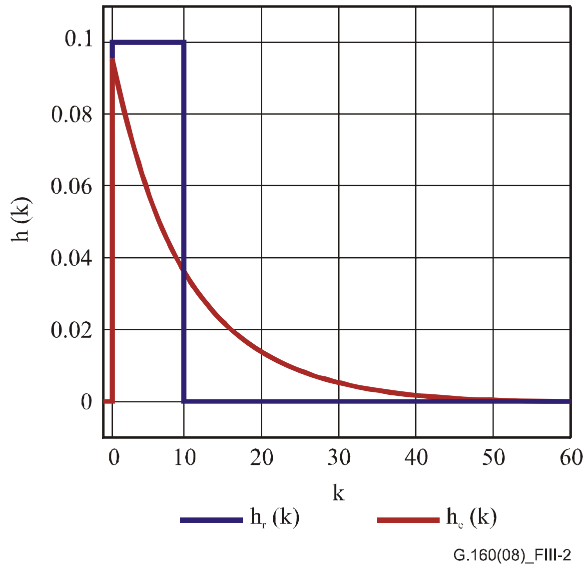


Figure III.2 – Impulse responses of the measurement filters

Both of the measurement devices, thus, sum a number of squared input signal samples to produce the level estimate. The difference is in weighs given to the squared samples in summation. The exponential averaging gives highest weight to the most recent signal samples, while the RMS method weights the signal samples inside the measurement window equally.

The signal levels are usually expressed in dBm0. The equations to compute level in dBm0 from *y* for A-law and µ-law encoding are:

$$S(k)=3.14+20\log\left[\frac{\sqrt{2y(k)}}{4096}\right] \quad \text{A-law encoding}$$

$$S(k)=3.17+20\log\left[\frac{\sqrt{2y(k)}}{8159}\right] \quad \mu\text{-law encoding}$$

III.2 Measurement of signal level

In several cases of interest, the noise process is Gaussian. Let us therefore first investigate what would be the output of the level measurement device if its input is zero mean white Gaussian noise. It is known that the sum of *n* squared normalized independent Gaussian, $N(a,\sigma)$, variables $z=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x-a)^2$ is a random variable that obeys χ^2-distribution with *n* degrees of freedom. The corresponding probability density function is given by:

$$f(z) = \frac{e^{-\frac{z}{2}} z^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}$$

where $\Gamma(\bullet)$ denotes the gamma function, $\Gamma(r) = \int_0^\infty e^{-t} t^{r-1} dt$.

Mean value of the χ^2-distribution is n and variance $2n$.

Hence, probability density function of the mean square value, y_r, above, computed over n samples of zero mean white Gaussian process with variance σ^2 is given by:

$$f(y) = \frac{e^{-\frac{yn}{2\sigma^2}} y^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{2\sigma^2}\right)^{\frac{n}{2}}$$

and has mean σ^2 and variance $\frac{2\sigma^4}{n}$.

The distribution of y_r, plotted with respect to deviation from its mean value in dB scale is shown in Figure III.3.

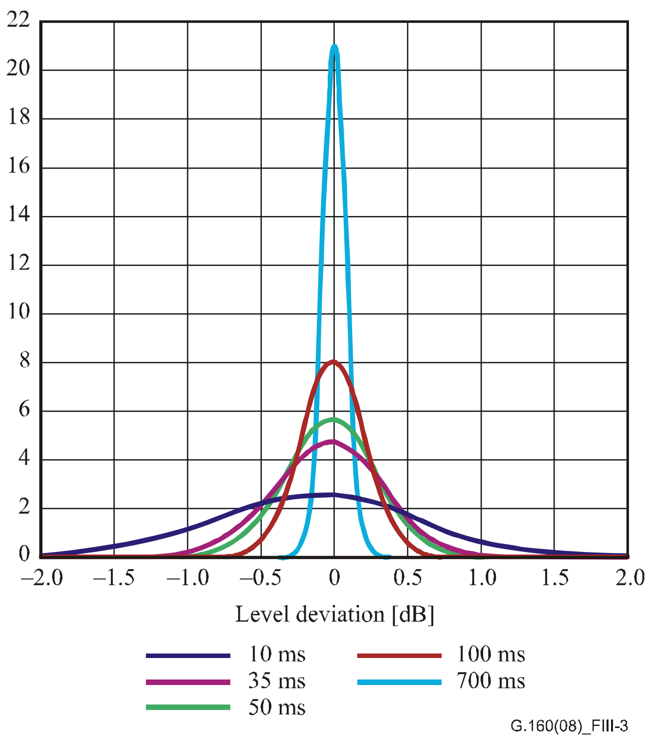


Figure III.3 – chi-squared distribution

The width of the interval inside what the level estimate can be expected with probabilities 0.98, 0.99 and 0.995 is plotted in Figure III.4 and tabulated for some window lengths in Table III.1.

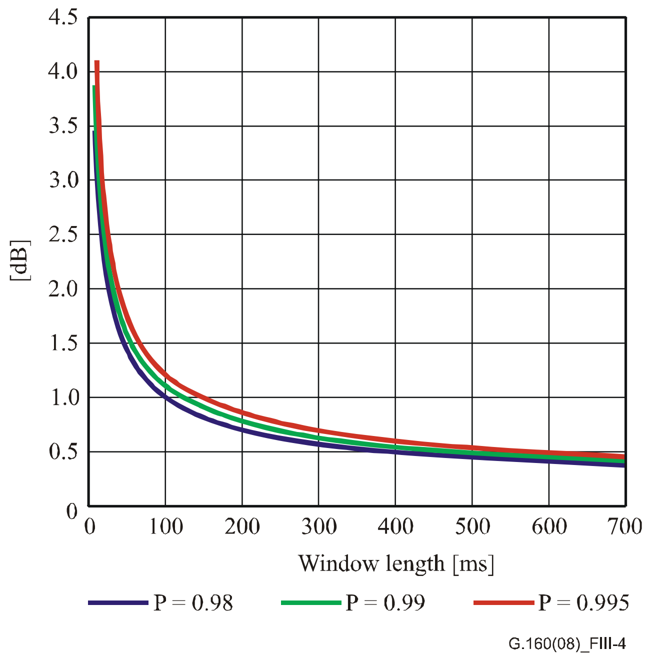


Figure III.4 – Level estimate uncertainty as function of window length

Table III.1 – Level estimate uncertainty for some window lengths

Window length	**Probability**		
	0.98	**0.99**	**0.995**
35 ms	1.71 dB	1.90 dB	2.07 dB
70 ms	1.21 dB	1.34 dB	1.56 dB
350 ms	0.54 dB	0.60 dB	0.65 dB
700 ms	0.38 dB	0.42 dB	0.46 dB

The smaller is the window length, n, the wider is the distribution of y_r and the more uncertainty there is in the provided estimates. And the other way around, the larger the window length, the narrower is the interval inside what the level estimate falls with certain probability. It is important to note that the level estimate of a random signal is another random signal and the bounds given above are probabilistic rather than deterministic.

To further illustrate the above results, let us plot a level measurement for 40-second long realization of band-limited white Gaussian noise with true level of –40 dBm0. The graph is computed using the measurement device with exponential averaging and 35 ms time constant in Figure III.5. One can see that the estimate fluctuates randomly around the true value.

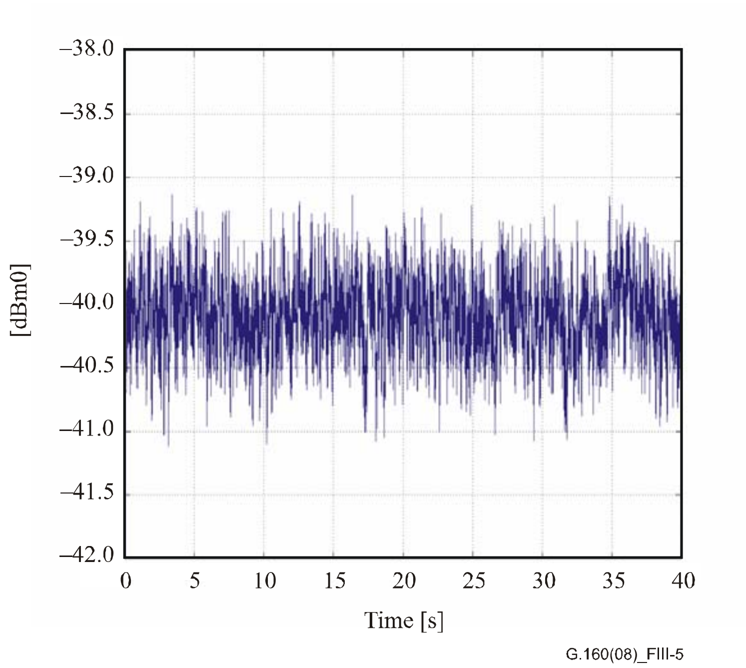


Figure III.5 – Level measurement of a white Gaussian noise realization using exponential averaging and 35 ms time constant

Let us call the difference between maximum and minimum values of the estimated level variation at the output of the level measurement device. In the following, the variation is studied by simulations.

In Figure III.6, the variation is plotted as a function of window length or time constant for measurement devices with and without the frequency weighting network. The numerical values are tabulated for some time constants (window lengths) in Table III.2. It should be kept in mind that the graphs are obtained from a single noise realization and do not represent absolute bounds. The variation is computed as the difference between the maximum and minimum reading between 10 and 40 seconds. The first 10 seconds are left out to exclude any impact of transients.

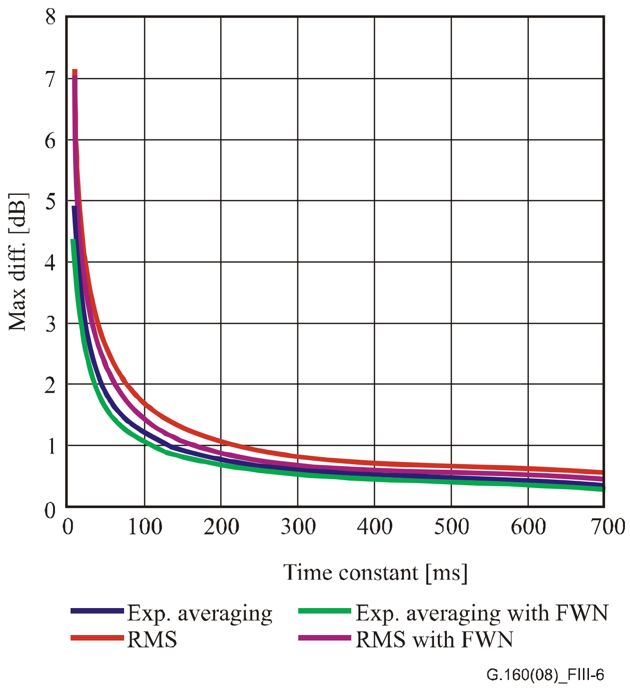


Figure III.6 – Variation in level measurements of white Gaussian noise as a function of time constant

It can be seen that the graphs in Figure III.6 above follow the shape of those in Figure III.4. The RMS method tends to give a slightly larger variation, as compared to the exponential averaging method. Application of the frequency weighting network tends to reduce the variation in both methods.

Table III.2 – Variation in level measurement results for some time constants

Time constant (window length)	Exp. averaging	Exp. averaging with FWN	RMS	RMS with FWN
35 ms	2.22 dB	2.01 dB	3.18 dB	2.94 dB
70 ms	1.52 dB	1.32 dB	2.12 dB	1.88 dB
350 ms	0.56 dB	0.48 dB	0.76 dB	0.61 dB
700 ms	0.36 dB	0.31 dB	0.58 dB	0.46 dB

The above illustrates achievable accuracy when measuring level of a stationary zero mean Gaussian noise process. In practice, the signal to measure may have mean different from zero and the signal may not be stationary or Gaussian. In those cases, the above analysis is not directly applicable but can be used as an illustration. In particular, when measuring the level of a non-stationary process, the measurement device needs to follow the changes in the statistical properties of the process.

III.3 Transition time for measurement device with exponential averaging

The other interesting parameter characterizing the measurement device is transition time, i.e., the time it takes to adjust the reading after the input level has changed.

As shown above, the level measurement device with exponential averaging can be seen as a linear filter applied to squared signal. The impulse response of the filter is:

$$h(k)=\left(1-\exp\left(-\frac{1}{f_s\tau}\right)\right)\exp\left(-\frac{k}{f_s\tau}\right)u(k),$$

and the response to unit step, $u(k)$, is:

$$g(k)=\left(1-\exp\left(-\frac{k+1}{f_s\tau}\right)\right)u(k)$$

Consider a piecewise constant signal, changing its level from a to b at time t = 0, as illustrated in Figure III.7.

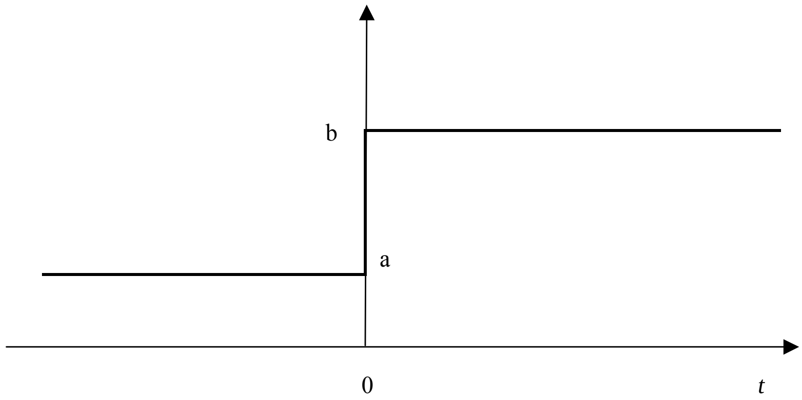


Figure III.7 – Piecewise constant signal

The signal in Figure III.7 can be formally written as:

$$s(k)=a+(b-a)u(k)$$

The reaction of the exponential filter to this signal will be:

$$y(k)=a+(b-a)\left(1-\exp\left(-\frac{1+k}{\tau f_s}\right)\right)u(k)$$

Let us define β-vicinity of b as the interval $[\beta b, b]$. In logarithmic scale, this translates to the interval $[(10\log_{10} b+10\log_{10}\beta), 10\log_{10} b]$ dB. The signal y will approach the new level b monotonically, and hence the time that it takes for y to reach into β-vicinity of b can be computed from the following equation:

$$a+(b-a)\left(1-\exp\left(-\frac{1+k}{\tau f_s}\right)\right)=\beta b$$

Solving above for k gives:

$$k=-\tau f_s \ln\left(\frac{b(1-\beta)}{b-a}\right)-1$$

For example, let us compute the time required to reach 0.5 dB vicinity of the final level in case of a 20 dB level change upward using measurement device with 35 ms time constant. Select a = 1 and b = 100. Then $\beta=10^{-0.5/10}=0.8913$, and from the above formula we find that k = 617, which corresponds to 76 ms with 8 kHz sampling rate.

Let us now make the same calculation for level change of 20 dB downward. Select $a = 100$ and $b = 1$. As the output of the level measurement device now approaches the true level in downward direction, we need to select $\beta = 10^{0.5/10} = 1.22$, and from the above formula we find that $k = 1875$ sample times or 233 ms with 8 kHz sampling rate.

We can also compute the free fall rate for the measurement device with exponential averaging. Consider the signal $1 - u(k)$, i.e., a signal changing from $a = 1$ to $b = 0$ at time zero. The response of the exponential filter to this signal would be:

$$1 - g(k) = \exp\left(\frac{k+1}{f_s \tau}\right)$$

In decibel scale we obtain:

$$10\log_{10}(1 - g(k)) = 10\log_{10}\left(\exp\left(\frac{k+1}{f_s \tau}\right)\right) = -\frac{10}{f_s \tau \ln 10}(k+1)$$

Hence, at the falling edge the exponential filter output decreases with the rate $\frac{10}{f_s \tau \ln 10}$ dB per sample or $\frac{10}{\tau \ln 10}$ dB per second if the input is removed. For example, if τ = 35 ms, the rate is 124 dB/s, and for τ = 700 ms the rate is 6.2 dB/s.

Using the above formulae, one can find out that during the time equal to time constant, the output of the exponential measurement device reaches into 2 dB vicinity of the target level when adjusting up. When adjusting downwards, the output decreases by 4.34 dB during the time equal to time constant.

To illustrate the above calculations, Figure III.8 shows the reaction of the level measurement device with exponential averaging for 35 ms and 700 ms time constants to instant changes in signal level. The signal to measure is a piece-wise constant with level –50 dBm0 from 0 to 5 seconds, then the level is raised to be –30 dBm0 between 5 and 10 seconds and the level is lowered again to be –50 dBm0 between 10 and 20 seconds.

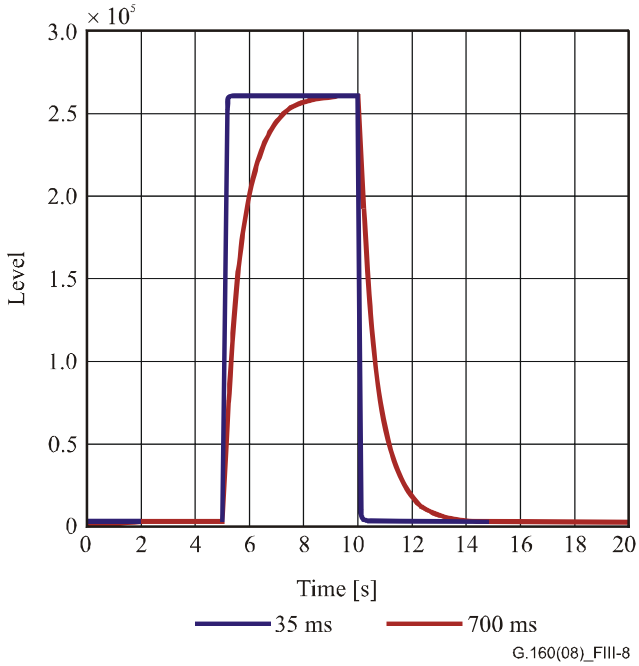


Figure III.8 – Reaction of exponential level measurement device to step changes (linear sale)

Figure III.9 shows the graphs from Figure III.8 in logarithmic (dB) scale.

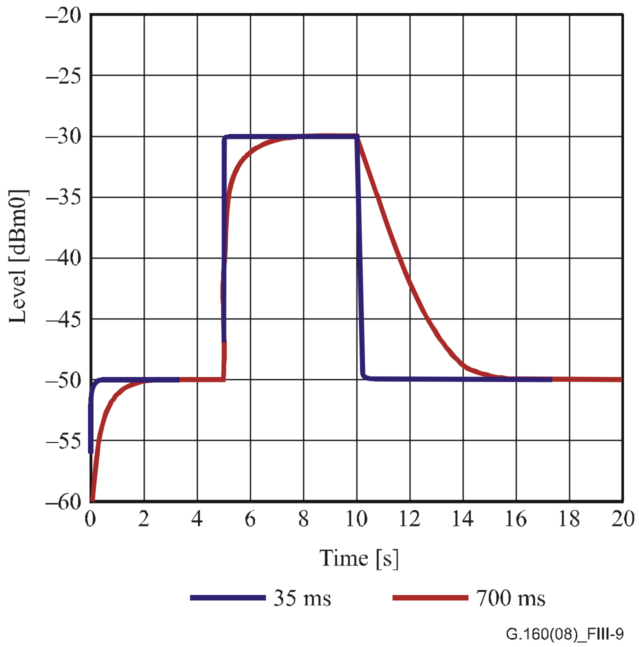


Figure III.9 – Reaction of exponential level measurement device to step changes (logarithmic scale)

It can be seen that even though the reaction on linear scale is exponential with equal shape on both upward and downward fronts, the symmetry is lost when plotting the same on logarithmic scale. In particular, it can be seen that the time it takes for the output from the measurement device to reach into X dB vicinity of the target value is different on up-front and down-front, being longer for adjustment at the falling edge.

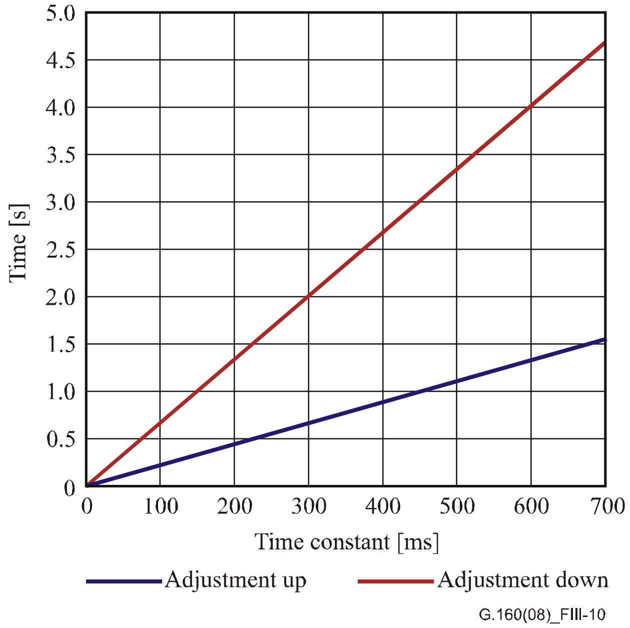


Figure III.10 – Transition time into 0.5 dB vicinity from the target as a function of time constant

Define the transition time as the time required for the reading to reach within 0.5 dB vicinity of the true value. Figure III.10 shows the transition times as a function of the time constant. Note that downward transition time in addition to the time constant also depends on the extent of the level change. The upper line in Figure III.10 shows downward transition time for 20 dB level change. For a larger level change, the transition time will be correspondingly longer.

III.4 Transition time for RMS measurement device

The RMS level measurement device has finite memory, equal to the window length, and consequently it reacts to step changes faster. The measured level has adjusted to the new input value after n/f_s s as illustrated in Figures III.11 and III.12.

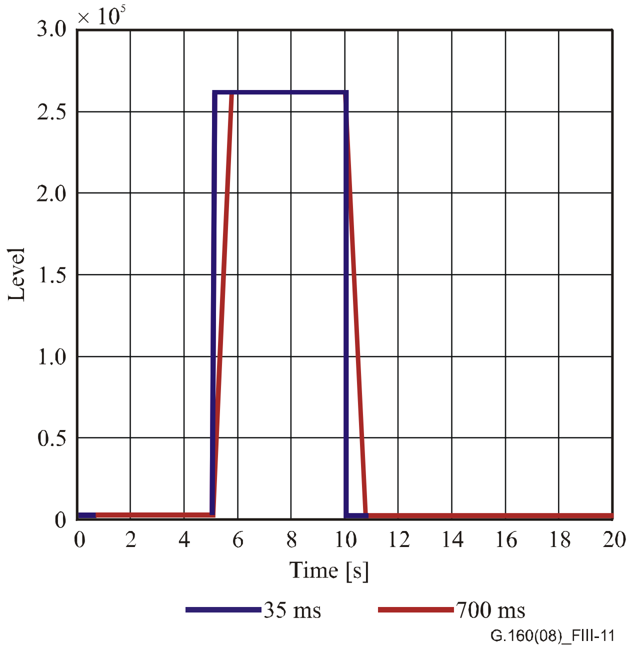


Figure III.11 – Reaction of RMS level measurement device to step changes (linear scale)

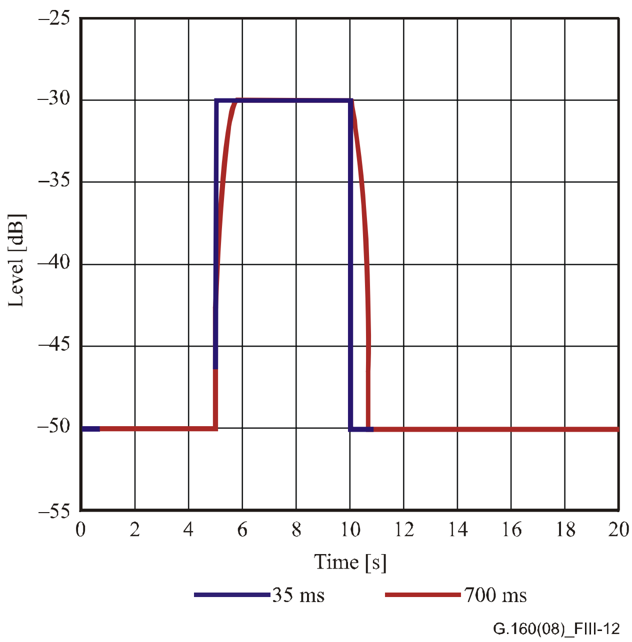


Figure III.12 – Reaction of RMS level measurement device to step changes (logarithmic scale)

III.5 Comparing the levels of two signals (exponential averaging)

Another complication arises when comparing levels of two signals. Usually good results can be obtained when constant levels are compared. During transients, however, the results may not be trustworthy. To illustrate this, let us compute the level of a signal rising from –30 to –10 dBm0 at time 0.5 seconds (black line in Figure III.13). We also compute the level of the copy of the same signal delayed by 1 ms (grey line in Figure III.13). In both cases, the measurement device with 35 ms time constant is used.

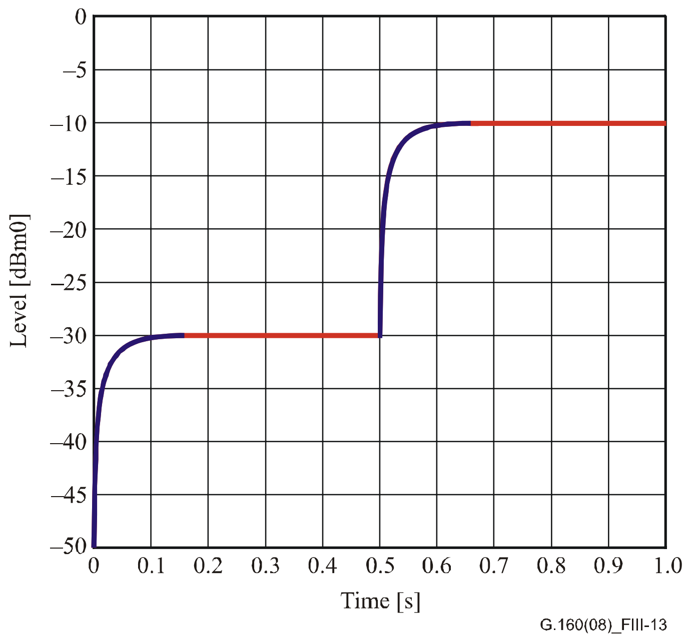


Figure III.13 – Measured levels of two signals delayed by 1 ms

The levels, if plotted together are almost indistinguishable. Subtracting one curve from the other will however result in difference by several dB during the transients (t = 0 and t = 0.5 s) as shown in Figure III.14.

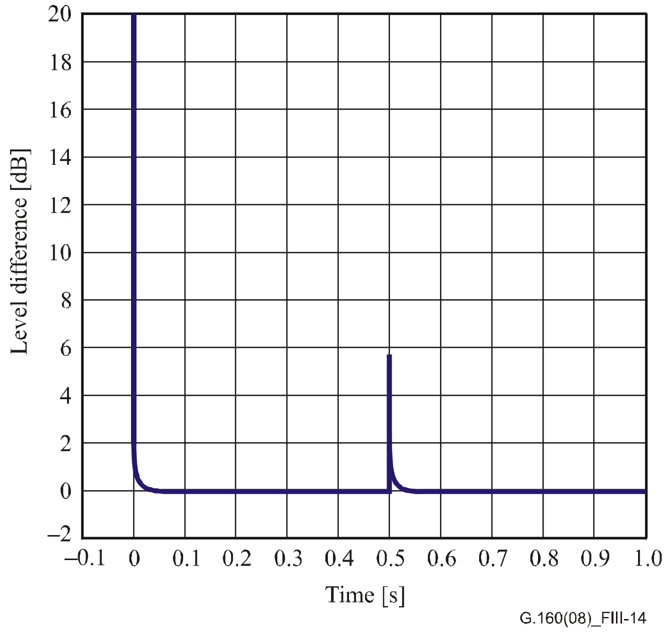


Figure III.14 – Level differences reflected in Figure III.13

Similar peaks will occur if the device is used to compare the input and output of a signal processing device, i.e., small and subjectively not noticeable differences during transients may lead to large differences when comparing the measured levels of the two signals.

III.6 Comparing the levels of two signals (RMS method)

The same effect is present when using the RMS measurement device as illustrated in Figures III.15 and III.16. The measurement device with 35 ms window is used to compute the graphs on the figures.

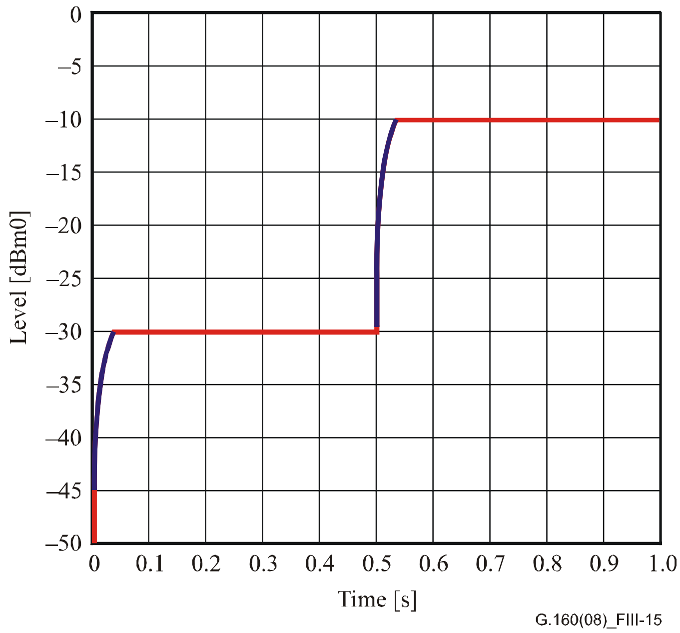


Figure III.15 – Measured levels of two signals delayed by 1 ms with RMS method

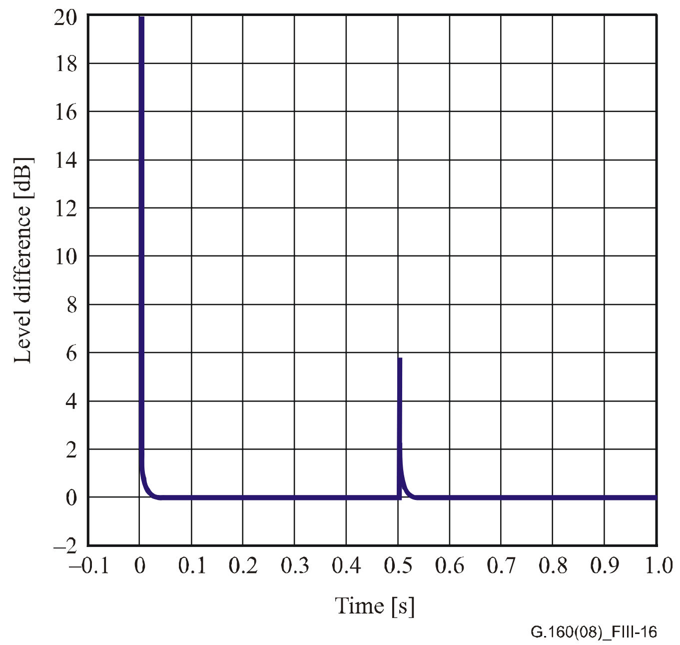


Figure III.16 – Level differences reflected in Figure III.15

The result is similar to that with the exponential level measurement device. Again, during transients the difference between the two level measurements shows considerable peaks.

III.7 Artificial speech signal in noise

Let us now consider concatenated male and female artificial voice passages from [ITU-T P.50] in noise. Figure III.17 shows the level measurements of the noise, artificial voice and an output from an "ideal noise reduction device" using the exponential averaging measurement device with 35 ms time constant. By "ideal noise reduction device", we mean a device that attenuates the noise by 4 dB after 2 seconds and that lets the artificial speech signal to pass unaltered.

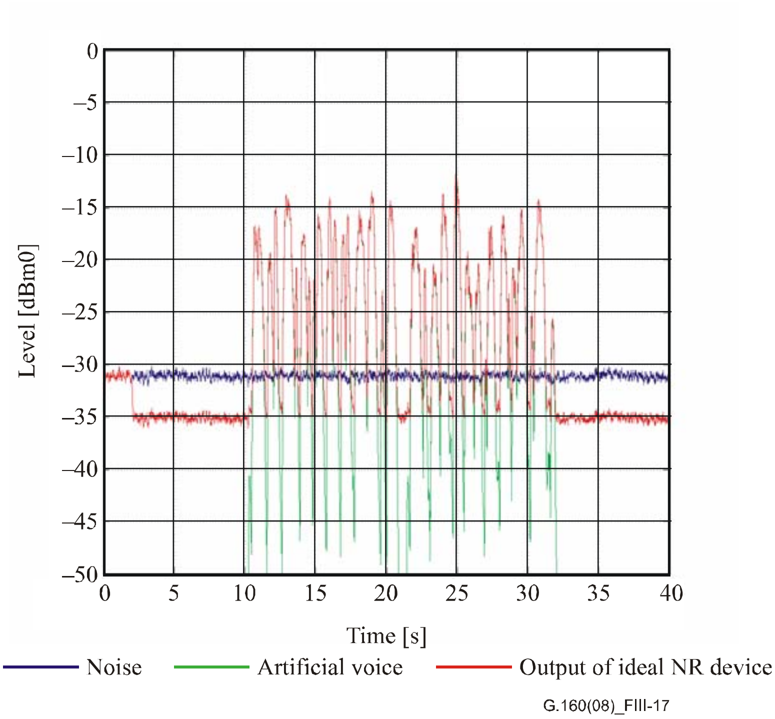


Figure III.17 – Level measurement using exponential averaging and 35 ms time constant

One can see that the level measurements of the clean signal and the noise-attenuated signal differ considerably. The difference is particularly large during the intervals when the artificial signal is weaker than the reduced noise, but even the neighbouring stronger parts of the artificial speech signal are affected because of the exponential transients due to the measurement device.

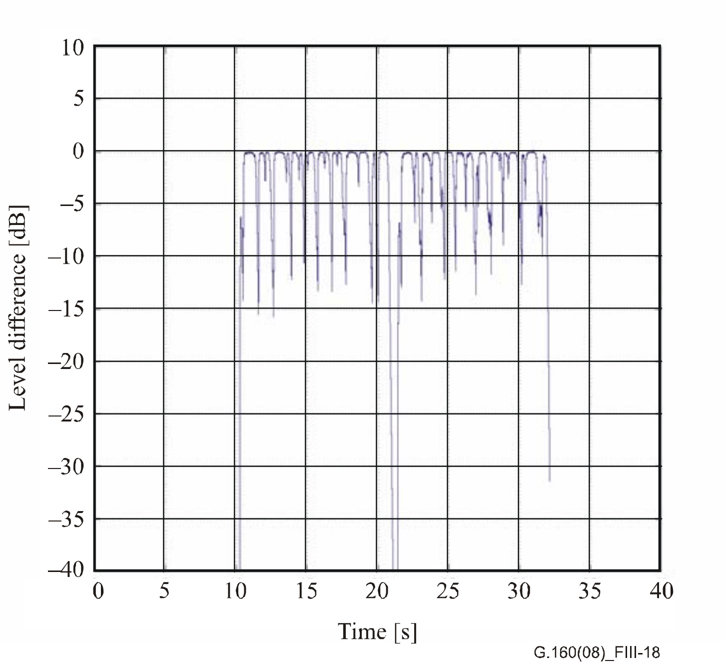


Figure III.18 – Difference between the NR output and clean signal from Figure III.17

Figure III.18 shows the difference between the levels of the clean signals (L_A) and the output of the noise reduction device (L_D). The difference shows large peaks, and this effect needs to be taken into account when formulating any test requirements. In fact, the peaky nature of the result turns the 35 ms measurement device to be not feasible in this type of measurement.

The corresponding plots for level measurement device with 700 ms time constant are shown in Figures III.19 and III.20.

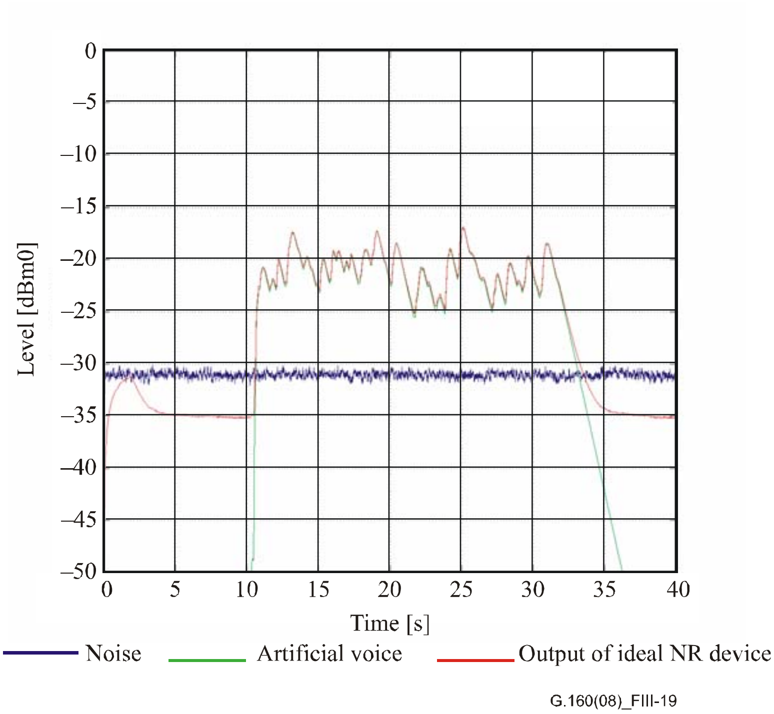


Figure III.19 – Level measurement using exponential averaging and 700 ms time constant

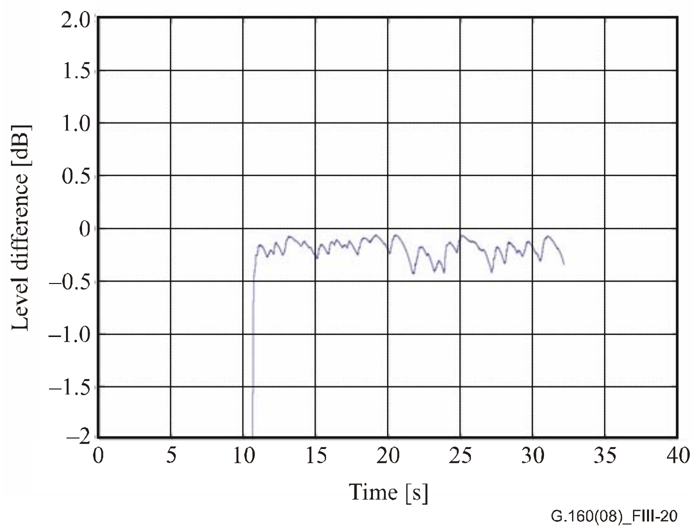


Figure III.20 – Difference between the NR output and clean signal from Figure III.19

One can see that now the inaccuracy caused by the measurement device is less than 0.5 dB making the measurement feasible.

From Figure III.10 we can, however, find that the transition time with 700 ms time constant is 1.5 s in upward and 4.7 s in downward direction (for 20 dB adjustment). This means that for the measurements one needs long guard times to allow the measurement result to stabilize.

Similar results can be obtained for the RMS method. Figure III.21 shows the level measurements of the noise, artificial voice and an output from an "ideal noise reduction device". The window length is 35 ms.

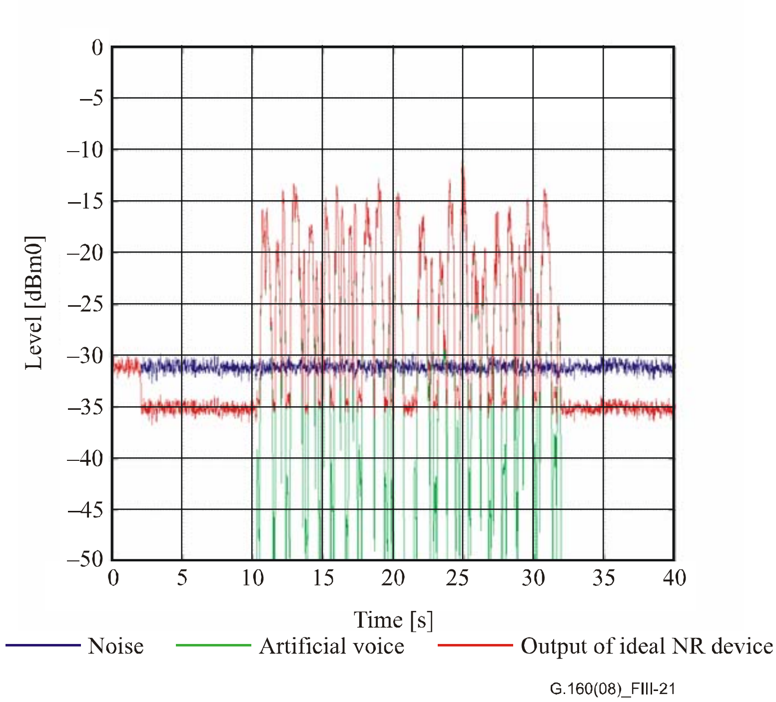


Figure III.21 – Level measurement using RMS method with 35 ms window

Figure III.22 shows the difference between the levels of the clean signals (L_A) and the output of the noise reduction device (L_D). The difference shows even larger peaks than those in Figure III.18 and, as previously, this effect needs to be taken into account when formulating any test requirements. Again, the peaky nature of the result turns the 35 ms measurement device to be not feasible in this type of measurement.

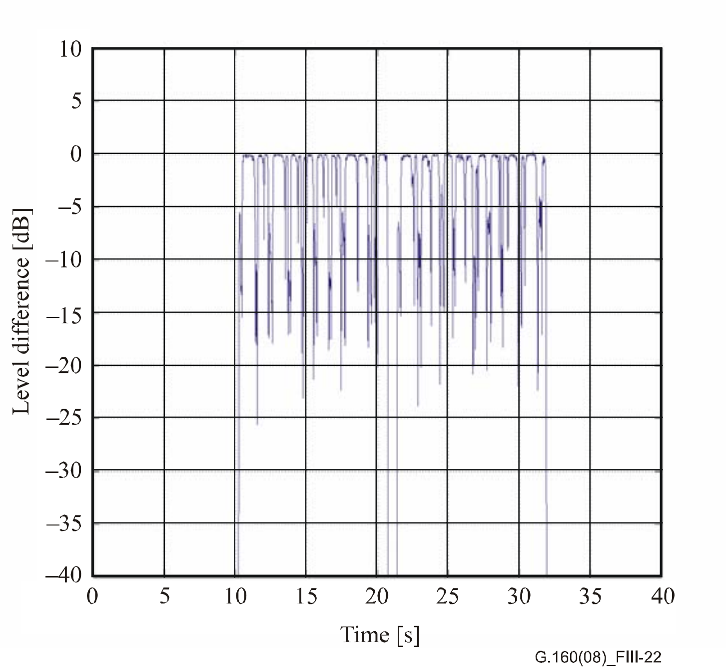


Figure III.22 – Difference between the NR output and clean signal from Figure III.21

The corresponding plots for level measurement device with 700 ms window are shown in Figures III.23 and III.24.

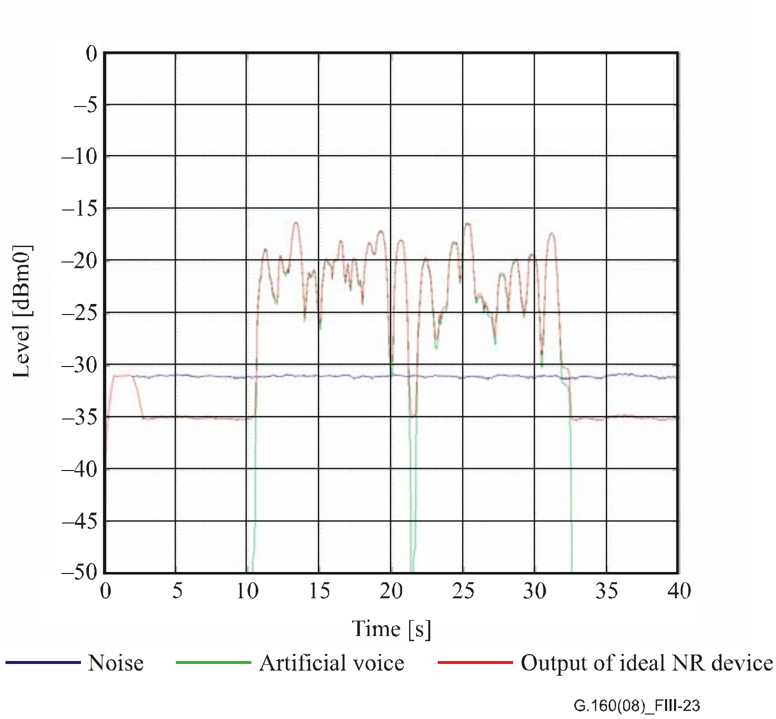


Figure III.23 – Level measurement using RMS method with 700 ms window

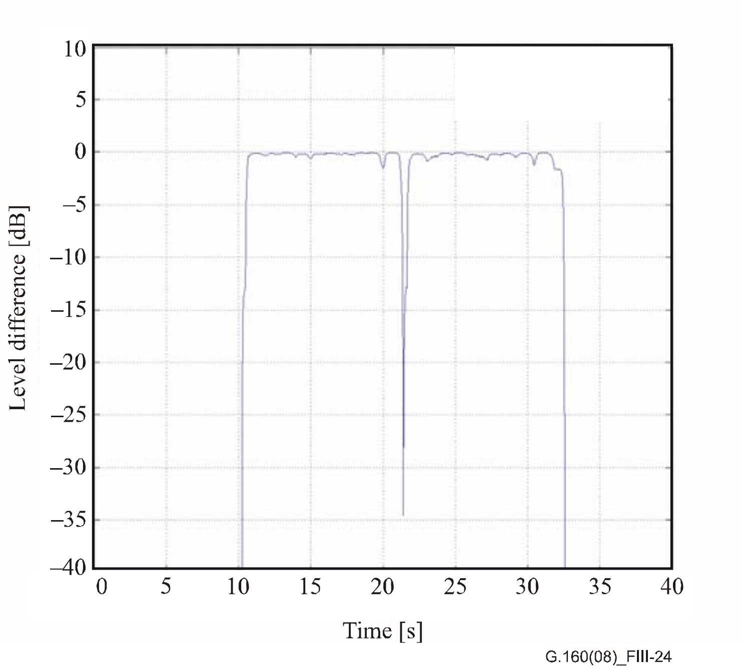


Figure III.24 – Difference between the NR output and clean signal from Figure III.23

It can be seen that the level difference computed with RMS method shows large peaks even if 700 ms window is used.

III.8 Conclusions

Level estimate of a random signal is another random signal, and a proper characterization of it is probabilistic rather than deterministic. In particular, no deterministic bounds to the deviation of the measurements from the true value can be given.

Choice of the window length in case of RMS method or time constant in case of exponential averaging for the level measurement device is a trade-off between measurement accuracy and capability to follow changes in signal level. In general, long time constant gives accurate level measurements but has pure tracking capabilities and vice versa. For any particular test, the choice of the level measurement device is thus a trade-off between accuracy and length of the transition time. Whatever choice is made, the requirements of the test should be aligned with the properties of the measurement device.

Appendix IV

One-octave filters for the AEC comfort noise test

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

IV.1 Introduction

The measure of spectral matching in the AEC comfort noise test utilizes three one-octave filters designed in accordance with [IEC 61260]. The centre frequencies (f_m) of the three filters are at 500 Hz, 1000 Hz and 2000 Hz. The lower band-edge frequencies (f_L) and the upper band-edge frequencies (f_U) of them are shown in Table IV.I. The attenuation characteristics of the filters are required to satisfy the class 0 characteristics of [IEC 61260]. Table IV.II gives the class 0 attenuation characteristics, and Figure IV.1 illustrates the characteristics in pictorial form.

This appendix provides two possible realizations of the three one-octave filters. Both realizations satisfy the required characteristics. One realization is infinite impulse response (IIR), and the other is finite impulse response (FIR).

Table IV.1 – Lower and upper band-edge frequencies of the sub-band filters

f_m (Hz)	f_L (Hz) = $f_m / \sqrt{2}$	f_U (Hz) = $f_m \times \sqrt{2}$
500	353.5534	707.1068
1000	707.1068	1414.2136
2000	1414.2136	2828.4271

Table IV.2 – Attenuation characteristics of the class 0 sub-band filters from [IEC 61260]

Normalized freq. f / f_m	Min; max attenuation limits (dB)	Normalized freq. f / f_m	Min; max attenuation limits (dB)
2^{0}	–0.15 ; +0.15	$2^{\pm 1/2}$	+2.3 ; +4.5
$2^{\pm 1/8}$	–0.15 ; +0.2	$2^{\pm 1}$	+18.0 ; +∞
$2^{\pm 1/4}$	–0.15 ; +0.4	$2^{\pm 2}$	+42.5 ; +∞
$2^{\pm 3/8}$	–0.15 ; +1.1	$2^{\pm 3}$	+62 ; +∞
$< 2^{+1/2}$	–0.15 ; +4.5	$\geq 2^{+4}$	+75 ; +∞
$> 2^{-1/2}$	–0.15 ; +4.5	$\leq 2^{-4}$	+75 ; +∞

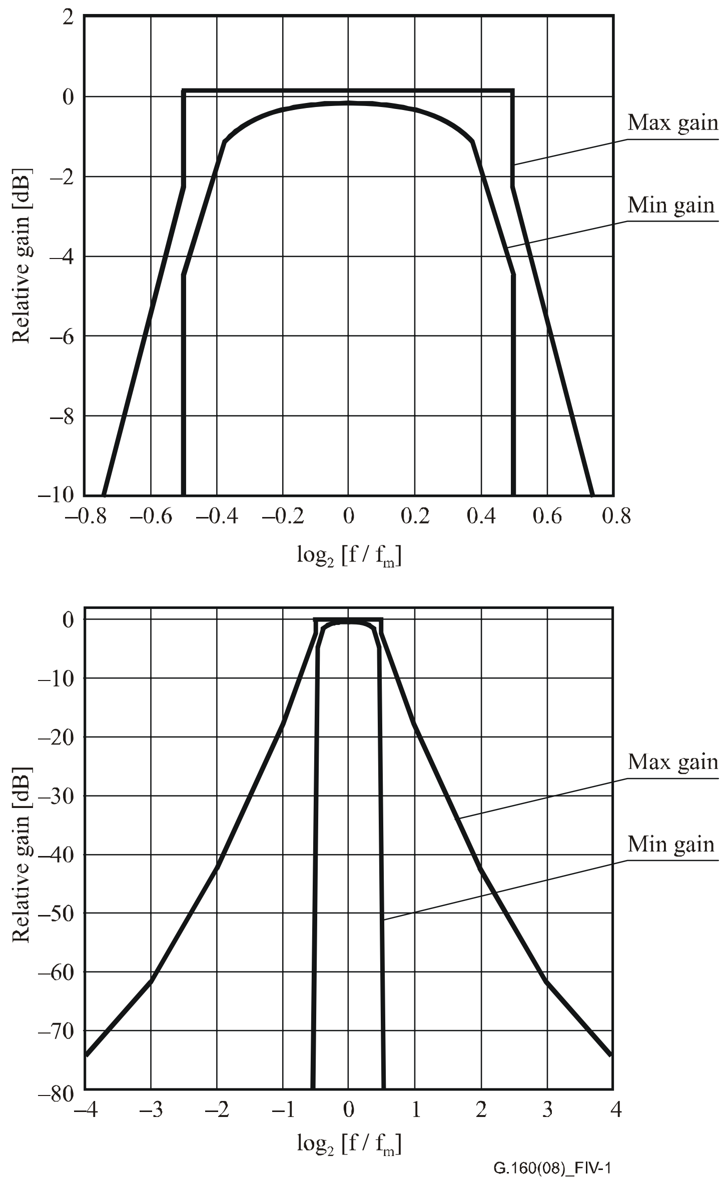


Figure IV.1 – Specification of the class 0 attenuation characteristics for the octave-band filters

IV.2 IIR realization

The IIR realizations of the one-octave filters are Butterworth filters. The orders of the Butterworth filters are 10. They are specified in the z-transfer function form as shown below. Conversion to difference equations is needed before applying them in the time domain.

$$H_{500}(z) = \frac{0.1341\left(1 - z^{-2}\right)}{\left(1 - 1.6172z^{-1} + 0.8939z^{-2}\right)} \frac{0.1341\left(1 - z^{-2}\right)}{\left(1 - 1.8673z^{-1} + 0.9424z^{-2}\right)} \frac{0.1263\left(1 - z^{-2}\right)}{\left(1 - 1.7447z^{-1} + 0.8346z^{-2}\right)}$$
$$\frac{0.1263\left(1 - z^{-2}\right)}{\left(1 - 1.5548z^{-1} + 0.7598z^{-2}\right)} \frac{0.1237\left(1 - z^{-2}\right)}{\left(1 - 1.6197z^{-1} + 0.7527z^{-2}\right)}$$

$$H_{1000}(z) = \frac{0.2542\left(1 - z^{-2}\right)}{\left(1 - 1.5926z^{-1} + 0.8841z^{-2}\right)} \frac{0.2542\left(1 - z^{-2}\right)}{\left(1 - 0.8287z^{-1} + 0.8142z^{-2}\right)} \frac{0.2295\left(1 - z^{-2}\right)}{\left(1 - 1.3562z^{-1} + 0.6862z^{-2}\right)}$$
$$\frac{0.2295\left(1 - z^{-2}\right)}{\left(1 - 0.8762z^{-1} + 0.5859z^{-2}\right)} \frac{0.2218\left(1 - z^{-2}\right)}{\left(1 - 1.0887z^{-1} + 0.5564z^{-2}\right)}$$

$$H_{2000}(z)=\frac{0.4654\left(1-z^{-2}\right)}{\left(1+1.0427z^{-1}+0.7652z^{-2}\right)}\frac{0.4654\left(1-z^{-2}\right)}{\left(1-0.7465z^{-1}+0.7408z^{-2}\right)}\frac{0.4005\left(1-z^{-2}\right)}{\left(1+0.6690z^{-1}+0.4183z^{-2}\right)}$$
$$\frac{0.4005\left(1-z^{-2}\right)}{\left(1-0.3926z^{-1}+0.3832z^{-2}\right)}\frac{0.3820\left(1-z^{-2}\right)}{\left(1+0.1356z^{-1}+0.2359z^{-2}\right)}$$

The attenuation characteristics of the filters are given in Tables IV.3-IV.5, respectively.

Table IV.3 – Attenuation characteristics of the filter centred at 500 Hz

Frequency	Attenuation
f0 = 500 Hz	0 dB
$2^{\pm\frac{1}{8}}$ f0 = 545.3 Hz, 458.5 Hz	0 dB
$2^{\pm\frac{1}{4}}$ f0 = 594.6 Hz, 420.4 Hz	0 dB
$2^{\pm\frac{3}{8}}$ f0 = 648.4 Hz, 385.6 Hz	0.1 dB, 0.4 dB
$2^{\pm\frac{1}{2}}$ f0 = 707.1 Hz, 353.6 Hz	2.1 dB, 4.2 dB
$2^{\pm 1}$ f0 = 1000 Hz, 250 Hz	33.1 dB, 34.2 dB
$2^{\pm 2}$ f0 = 2000 Hz, 125 Hz	81.3 dB, 73.1 dB
$2^{\pm 3}$ f0 = 4000 Hz, 62.5 Hz	288.4 dB, 108.2 dB
2^{-4} f0 = 31.3 Hz	145.4 dB

Table IV.4 – Attenuation characteristics of the filter centred at 1000 Hz

Frequency	Attenuation
f0 = 1000 Hz	0 dB
$2^{\pm\frac{1}{8}}$ f0 = 1090.5 Hz, 917 Hz	0 dB
$2^{\pm\frac{1}{4}}$ f0 = 1189.2 Hz, 840.9 Hz	0 dB
$2^{\pm\frac{3}{8}}$ f0 = 1296.8 Hz, 771.1 Hz	0.1 dB, 0.4 dB
$2^{\pm\frac{1}{2}}$ f0 = 1414.2 Hz, 707.1 Hz	2.1 dB, 4.1 dB
$2^{\pm 1}$ f0 = 2000 Hz, 500 Hz	38.6 dB, 31.9 dB
$2^{\pm 2}$ f0 = 4000 Hz, 250 Hz	297.7 dB, 72.4 dB
2^{-3} f0 = 125 Hz	104.3 dB
2^{-4} f0 = 62.5 Hz	137.8 dB

Table IV.5 – Attenuation characteristics of the filter centred at 2000 Hz

Frequency	Attenuation
f0 = 2000 Hz	0 dB
$2^{\pm\frac{1}{8}}$ f0 = 2181 Hz, 1834 Hz	0 dB
$2^{\pm\frac{1}{4}}$ f0 = 2378.4 Hz, 1681.8 Hz	0 dB
$2^{\pm\frac{3}{8}}$ f0 = 2593.7 Hz, 1542.2 Hz	0.5 dB
$2^{\pm\frac{1}{2}}$ f0 = 2828.4 Hz, 1414.2 Hz	3 dB
$2^{\pm 1}$ f0 = 4000 Hz, 1000 Hz	236 dB, 27.7 dB
2^{-2} f0 = 500 Hz	64.4 dB
2^{-3} f0 = 250 Hz	97.4 dB
2^{-4} f0 = 125 Hz	127.9 dB

IV.3 FIR realization

The coefficients of the FIR realization of the three one-octave filters are tabulated in Tables IV.6-IV.8. The lengths of the three filters, starting from the lowest sub-band filter, are 160, 80 and 40. The magnitude responses of the filters are shown in Figure IV.2.

Table IV.6 – Filter coefficients for the one-octave sub-band filter centred at 500 Hz, dividing coefficients by 2^{15} needed to obtain unity gain in passband

$h_{0,159}$	0	$h_{20,139}$	2	$h_{40,119}$	12	$h_{60,99}$	–15
$h_{1,158}$	0	$h_{21,138}$	4	$h_{41,118}$	46	$h_{61,98}$	41
$h_{2,157}$	0	$h_{22,137}$	3	$h_{42,117}$	82	$h_{62,97}$	215
$h_{3,156}$	0	$h_{23,136}$	–3	$h_{43,116}$	102	$h_{63,96}$	474
$h_{4,155}$	0	$h_{24,135}$	–14	$h_{44,115}$	88	$h_{64,95}$	744
$h_{5,154}$	–1	$h_{25,134}$	–28	$h_{45,114}$	31	$h_{65,94}$	920
$h_{6,153}$	–1	$h_{26,133}$	–41	$h_{46,113}$	–65	$h_{66,93}$	903
$h_{7,152}$	0	$h_{27,132}$	–49	$h_{47,112}$	–184	$h_{67,92}$	625
$h_{8,151}$	1	$h_{28,131}$	–46	$h_{48,111}$	–295	$h_{68,91}$	82
$h_{9,150}$	2	$h_{29,130}$	–32	$h_{49,110}$	–367	$h_{69,90}$	–656
$h_{10,149}$	4	$h_{30,129}$	–6	$h_{50,109}$	–372	$h_{70,89}$	–1449
$h_{11,148}$	5	$h_{31,128}$	26	$h_{51,108}$	–300	$h_{71,88}$	–2115
$h_{12,147}$	6	$h_{32,127}$	56	$h_{52,107}$	–163	$h_{72,87}$	–2475
$h_{13,146}$	5	$h_{33,126}$	77	$h_{53,106}$	9	$h_{73,86}$	–2396
$h_{14,145}$	3	$h_{34,125}$	82	$h_{54,105}$	170	$h_{74,85}$	–1836
$h_{15,144}$	0	$h_{35,124}$	69	$h_{55,104}$	278	$h_{75,84}$	–857

Table IV.6 – Filter coefficients for the one-octave sub-band filter centred at 500 Hz, dividing coefficients by 2^{15} needed to obtain unity gain in passband

$h_{16,143}$	–3	$h_{36,123}$	45	$h_{56,103}$	306	$h_{76,83}$	373
$h_{17,142}$	–4	$h_{37,122}$	17	$h_{57,102}$	251	$h_{77,82}$	1620
$h_{18,141}$	–4	$h_{38,121}$	–3	$h_{58,101}$	143	$h_{78,81}$	2632
$h_{19,140}$	–1	$h_{39,120}$	–6	$h_{59,100}$	34	$h_{79,80}$	3198

Table IV.7 – Filter coefficients for the one-octave sub-band filter centred at 1000 Hz, dividing coefficients by 2^{15} needed to obtain unity gain in passband

h_{0}	0	h_{20}	25	h_{40}	6397	h_{60}	–12
h_{1}	0	h_{21}	164	h_{41}	3241	h_{61}	34
h_{2}	–1	h_{22}	176	h_{42}	–1714	h_{62}	139
h_{3}	–1	h_{23}	–131	h_{43}	–4792	h_{63}	153
h_{4}	1	h_{24}	–591	h_{44}	–4230	h_{64}	52
h_{5}	8	h_{25}	–744	h_{45}	–1312	h_{65}	–63
h_{6}	12	h_{26}	–326	h_{46}	1250	h_{66}	–97
h_{7}	6	h_{27}	340	h_{47}	1840	h_{67}	–56
h_{8}	–5	h_{28}	612	h_{48}	948	h_{68}	–6
h_{9}	–8	h_{29}	286	h_{49}	81	h_{69}	8
h_{10}	4	h_{30}	–30	h_{50}	69	h_{70}	–3
h_{11}	6	h_{31}	429	h_{51}	502	h_{71}	–8
h_{12}	–28	h_{32}	1487	h_{52}	557	h_{72}	0
h_{13}	–82	h_{33}	1806	h_{53}	17	h_{73}	10
h_{14}	–92	h_{34}	164	h_{54}	–601	h_{74}	11
h_{15}	–12	h_{35}	–2898	h_{55}	–734	h_{75}	4
h_{16}	112	h_{36}	–4950	h_{56}	–368	h_{76}	–1
h_{17}	163	h_{37}	–3671	h_{57}	63	h_{77}	–1
h_{18}	90	h_{38}	746	h_{58}	204	h_{78}	0
h_{19}	–6	h_{39}	5265	h_{59}	92	h_{79}	0

Table IV.8 – Filter coefficients for the one-octave sub-band filter centred at 2000 Hz, dividing coefficients by 2^{15} needed to obtain unity gain in passband

h_0	0	h_{10}	49	h_{20}	12793	h_{30}	–23
h_1	–2	h_{11}	353	h_{21}	–3428	h_{31}	278
h_2	2	h_{12}	–1181	h_{22}	–8461	h_{32}	103
h_3	24	h_{13}	–652	h_{23}	2501	h_{33}	–194
h_4	–11	h_{14}	1223	h_{24}	1897	h_{34}	–12
h_5	7	h_{15}	–60	h_{25}	138	h_{35}	–6
h_6	–55	h_{16}	2974	h_{26}	1114	h_{36}	0
h_7	–185	h_{17}	327	h_{27}	–1201	h_{37}	21
h_8	224	h_{18}	–9899	h_{28}	–736	h_{38}	–1
h_9	179	h_{19}	1493	h_{29}	407	h_{39}	–1

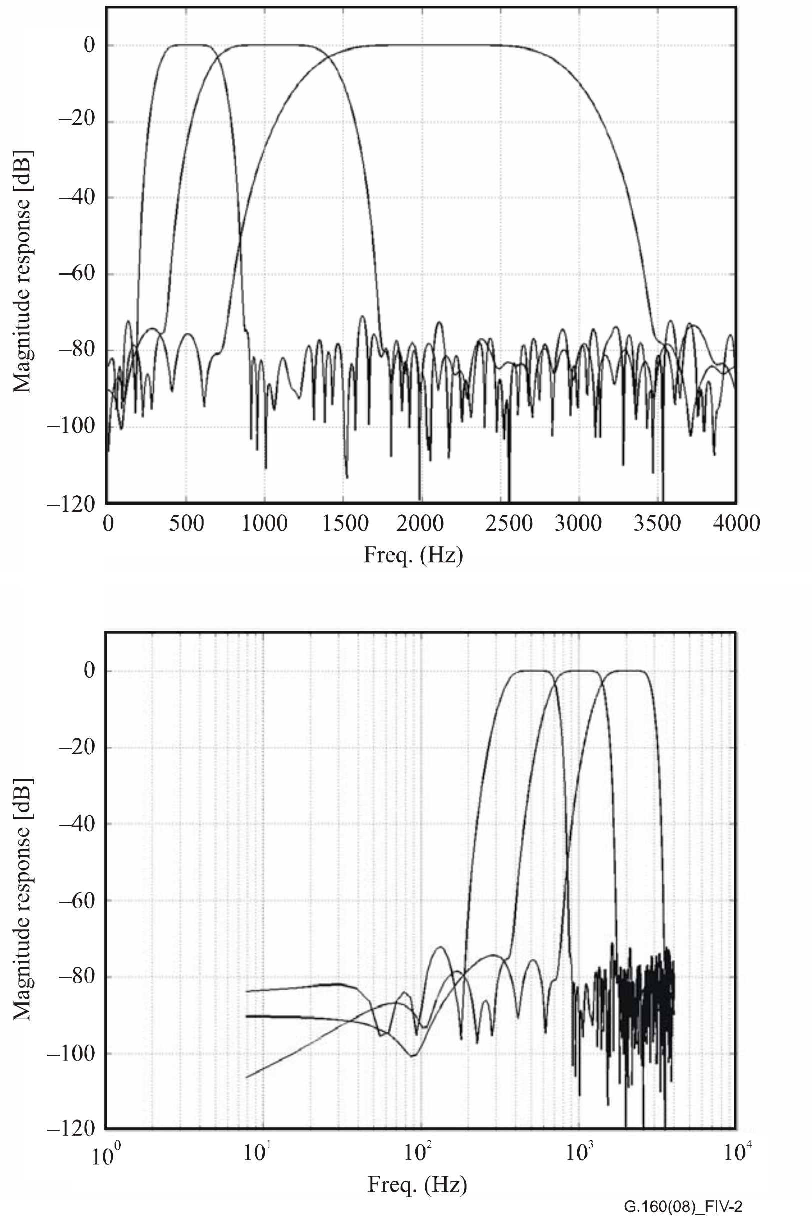


Figure IV.2 – Magnitude responses of the three one-octave sub-band filters, the filter lengths of the three sub-band filters, starting from the first octave filter, are 160, 80 and 40

Bibliography

[b-ITU-T G.161] Recommendation ITU-T G.161 (2004), *Interaction aspects of signal processing network equipment*.

[b-ITU-T G.169] Recommendation ITU-T G.169 (1999), *Automatic level control devices*.

[b-ITU-T G.191] Recommendation ITU-T G.191 (2005), *Software tools for speech and audio coding standardization*.

[b-ITU-T G.711] Recommendation ITU-T G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

[b-ITU-T G.712] Recommendation ITU-T G.712 (2001), *Transmission performance characteristics of pulse code modulation channels*.

[b-ITU-T P.48] Recommendation ITU-T P.48 (1988), *Specification for an intermediate reference system*.

[b-ITU-T P.56] Recommendation ITU-T P.56 (1993), *Objective measurement of active speech level*.

[b-ITU-T Q.55] Recommendation ITU-T Q.55 (1999), *Signalling between signal processing network equipments (SPNE) and international switching centres (ISC)*.

[b-ITU-T Q.56] Recommendation ITU-T Q.56 (2001), *Signalling between signal processing network equipment (SPNE) and international switching centres (ISC) over an IP network*.

[b-ITU-T Q.115.0] Recommendation ITU-T Q.115.0 (2002), *Protocols for the control of signal processing network elements and functions*.

[b-ITU-T V.23] Recommendation ITU-T V.23 (1988), *600/1200-baud modem standardized for use in the general switched telephone network*.

[b-ITU-T V.150.1] Recommendation ITU-T V.150.1 (2003), *Modem-over-IP networks: Procedures for the end-to-end connection of V-series DCEs*.

[b-TIA/EIA-IS-853] TIA TR45 specification TIA/EIA-IS-853, *Noise Suppression Minimum Performance for AMR*. <http://global.ihs.com/search_res.cfm?RID=NS&input_doc_number=TIA%2FEIA%2FIS%2D853 >

[b-ETSI TS 126 077] ETSI TS 126 077 V5.0.0 (2002-06), *Digital cellular telecommunications system (Phase 2+): Universal Mobile telecommunications System (UMTS); Minimum Performance Requirements for Noise Suppresser Application to the AMR Speech Encoder*. <http://webapp.etsi.org/workprogram/Report_WorkItem.asp?WKI_ID=16935>

[b-ETSI TR 126 975] ETSI TR 126 975 V6.0.0 (2004-12), *Digital cellular telecommunications system (Phase 2+): Universal Mobile telecommunications System (UMTS); Performance characterization of the adaptive Multi-Rate (AMR) speech codec*. <http://webapp.etsi.org/workprogram/Report_WorkItem.asp?WKI_ID=21816>

[b-ETSI TS 126 226] ETSI TS 126 226 V7.0.0 (2007-06), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Cellular text telephone modem; General description*. <http://webapp.etsi.org/workprogram/Report_WorkItem.asp?WKI_ID=26458>

[b-Gibson] Gibson, Jerry D., and Budagavi Madhukar (1998), *Speech Coding in Mobile Radio Communications*, Proceedings of the IEEE, Vol. 86, No. 7; July. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=00681370 >

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	Telephone transmission quality, telephone installations, local line networks
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, 2009