International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1130
(06/2015)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Communications involving vehicles

Subsystem requirements for automotive speech services

Recommendation ITU-T P.1130

ITU-T

International Telecommunication Union

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	**Series**	**P.1100**
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than voice services	Series	P.1500

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

Recommendation ITU-T P.1130

Subsystem requirements for automotive speech services

Summary

Recommendation ITU-T P.1130 defines test methodologies for and standard behaviour of subsystems used in automotive speakerphone terminals. The purpose of this Recommendation is to provide guidance on the design and optimization of such subsystems, as well as the diagnostic capabilities needed to give a consistent and high quality of service of the overall speakerphone terminal to the users of such devices. This specification is intended to give guidance to all parties involved in the design and integration of speakerphone terminals. This Recommendation covers both narrowband and wideband systems.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1130	2015-06-29	12	11.1002/1000/12518

Keywords

Hands-free, motor vehicles, performance, quality of service, speakerphone, subsystem.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Page

Page

Recommendation ITU-T P.1130

Subsystem requirements for automotive speech services

1 Scope

The aim of this Recommendation is to define test methods and requirements for subsystems of automotive narrowband and wideband speakerphone terminals and other speech services using the speakerphone's acoustic interface. This Recommendation covers the following:

– definition of subsystems based on different system architectures;

– performance requirements for subsystems;

– diagnostic information;

– guidance on component and subsystem optimization;

– coordination of subsystems;

– both narrowband and wideband systems;

– acoustic interface requirements for other speech services such as speech recognition and application prompt playback.

The methods, the analysis and the performance parameters described in this Recommendation are based on test signals and test procedures defined in [ITU-T P.501], [ITU-T P.502], [ITU-T P.340], [ITU-T P.1100], [ITU-T P.1110], [b-ETSI ES 202 739] and [b-ETSI ES 202 740].

Several performance classes are defined for each performance parameter instead of a single pass/fail criterion. This enables easier interpretation of the performance levels achieved and provides more flexibility in specifying the required performance level of a subsystem. Lower numbered classes have more stringent requirements and represent higher performance levels.

Note that all limits in this version of the Recommendation are considered provisional due to the lack of verification data. Furthermore, the interpretation of the performance classes deserves further study.

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.100.1] Recommendation ITU-T G.100.1 (in force), *The use of the decibel and of relative levels in speechband telecommunications*.

[ITU-T G.114] Recommendation ITU-T G.114 (in force), *One-way transmission time*.

[ITU-T G.122] Recommendation ITU-T G.122 (in force), *Influence of national systems on stability and talker echo in international connections*.

[ITU-T G.131] Recommendation ITU-T G.131 (in force), *Talker echo and its control*.

[ITU-T O.41] Recommendation ITU-T O.41 (in force), *Psophometer for use on telephone-type circuits*.

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

[ITU-T P.57] Recommendation ITU-T P.57 (in force), *Artificial ears*.

[ITU-T P.58] Recommendation ITU-T P.58 (in force), *Head and torso simulators for telephonometry*.

[ITU-T P.79] Recommendation ITU-T P.79 (in force), *Calculation of loudness rating for telephone sets.*

[ITU-T P.340] Recommendation ITU-T P.340 (in force), *Transmission characteristics and speech quality parameters of hands-free telephones*.

[ITU-T P.341] Recommendation ITU-T P.341 (in force), *Characteristics of Wideband Terminals*.

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

[ITU-T P.502] Recommendation ITU-T P.502 (in force), *Objective Analysis Methods for Speech Communication Systems, Using Complex Test Signals*.

[ITU-T P.581] Recommendation ITU-T P.581 (in force), *Use of head and torso simulators (HATS) for hands-free terminal testing*.

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

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

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

[ITU-T P.863] Recommendation ITU-T P.863 (in force), *Methods for objective and subjective assessment of speech quality.*

[ITU-T P.1100] Recommendation ITU-T P.1100 (in force), *Narrowband hands-free communication in motor vehicles.*

[ITU-T P.1110] Recommendation ITU-T P.1110 (in force), *Wideband hands-free communication in motor vehicles*.

[IEC 60268-4] IEC 60268-4 (2004), *Sound system equipment – Part 4: Microphones*.

[IEC 61260] IEC 61260:2014, *Electroacoustics – Octave-band and fractional-octave-band filters – Part 1: Specifications*.

[ISO 1999] ISO 1999:2013, *Acoustics – Estimation of noise-induced hearing loss*.

[ISO 3745] ISO 3745:2012, *Acoustics – Determination of sound power levels of noise sources using sound pressure – Precision methods for anechoic and hemi-anechoic rooms.*

3 Definitions

This Recommendation defines the following terms:

3.1 artificial ear: Device incorporating an acoustic coupler and a calibrated microphone for the measurement of sound pressure and which has an overall acoustic impedance similar to that of the median adult human ear over a given frequency band.

3.2 **codec**: Combination of an analogue-to-digital encoder and a digital-to-analogue decoder operating in opposite directions of transmission in the same equipment.

3.3 **composite source signal (CSS)**: Signal composed in time by various signal elements.

3.4 **diffuse field equalization**: Equalization of the HATS sound pick-up, equalization of the difference, in dB, between the spectrum level of the acoustic pressure at the eardrum reference point (DRP) and the spectrum level of the acoustic pressure at the HATS reference point (HRP) in a diffuse sound field with the HATS absent using the reverse nominal curve given in Table 3 of [ITU-T P.58].

3.5 **eardrum reference point (DRP)**: Point located at the end of the ear canal which corresponds to the eardrum position.

3.6 **echo tail**: Full echo path as seen by the echo canceller under consideration; this includes any delay introduced by electrical, acoustical or digital subsystems included in the echopath from R2 to S2.

3.7 **freefield equalization**: The transfer characteristic of the artificial head is equalized in such a way that for frontal sound incidence in anechoic conditions the frequency response of the artificial head is flat. This equalization is specific to the HATS used.

3.8 **freefield reference point**: Point located in the free sound field, at least at a distance of 1.5 m from a sound source radiating in free air (in case of a head and torso simulator [HATS] in the centre of the artificial head with no artificial head present).

3.9 **hands-free reference point (HFRP)**: A point located on the axis of the artificial mouth, at 50 cm from the outer plane of the lip ring, where the level calibration is made, under free-field conditions. It corresponds to the measurement point 11, as defined in [ITU-T P.51].

3.10 **hands-free terminal**: Telephone set that does not require the use of hands during the communications session; examples are headset, speakerphone and group-audio terminal.

3.11 **head and torso simulator (HATS) for telephonometry**: Manikin extending downward from the top of the head to the waist, designed to simulate the sound pick-up characteristics and the acoustic diffraction produced by a median human adult and to reproduce the acoustic field generated by the human mouth.

3.12 **headset**: Device which includes telephone receiver and transmitter which is typically secured to the head or the ear of the wearer.

3.13 **MOS-LQO (mean opinion score – listening-only quality objective)**: The score is calculated by means of an objective model which aims to predict the quality for a listening-only test situation. Objective measurements made using the model given in [ITU-T P.863] give results in terms of MOS-LQO.

3.14 **MOS-TQO (mean opinion score – talking quality objective)**: The score is calculated by means of an objective model which aims at predicting the quality for a talking-only test situation. Methods generating a MOS-TQO are currently under development and not yet standardized.

3.15 **mouth reference point (MRP)**: The MRP is located on the axis and 25 mm in front of the lip plane of a mouth simulator.

3.16 **nominal setting of the volume control**: When a receive volume control is provided, this is the setting which is closest to the nominal RLR of 2 dB.

3.17 **receive loudness rating (RLR)**: The loudness loss between an electric interface in the network and the listening subscriber's ear. (The loudness loss is here defined as the weighted (dB) average of driving e.m.f. to measured sound pressure.)

3.18 send loudness rating (SLR): The loudness loss between the speaking subscriber's mouth and an electric interface in the network. (The loudness loss is here defined as the weighted (dB) average of driving sound pressure to measured voltage.)

3.19 wideband speech: Voice service with enhanced quality compared to PCM G.711 and allowing the transmission of a vocal frequency range of at least 150 Hz to 7 kHz.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

A/D	Analogue/Digital
ACR	Absolute Category Rating
AGC	Automatic Gain Control
AH,R	Attenuation range in Receive direction
AH,R,dt	Attenuation range in Receive direction during double talk
AH,R,dt_{SE}	Signal Enhancement Attenuation range in Receive direction during double talk
AH,R_{SE}	Signal Enhancement Activation range, Receive
AH,S	Attenuation range in Send direction
AH,S,dt	Attenuation range in Send direction during double talk
AH,S,dt_{SE}	Signal Enhancement Attenuation range in Send direction during double talk
AH,S_{SE}	Signal Enhancement Activation range, Send
AI	Articulation Index
BGN	Background Noise
BGNTACSE	Signal Enhancement Background Noise Transmission After Call Set-up
CNIE	Comfort Noise Injection after Enhancement
$COH_{REF-LSP}$	Coherence between test points LSP and REF
CSS	Composite Source Signal
D/A	Digital/Analogue
DEC_{DT}-$_{SE}$	Detection of Echo Components during Double-Talk after Enhancement
DFT	Discrete Fourier Transform
$DOOB_{SES}$	Discrimination against Out-Of-Band Signals after Signal Enhancement, Send
DRP	Drum Reference Point
D_{SER}	Distortion in Receive after Signal Enhancement
D_{SES}	Distortion in Send after Signal Enhancement
DTX	Discontinuous Transmission
DUT	Device Under Test
EC_{BD}	Bidirectional transport Echo Cancellation
EEB	Early Energy Balance

$ELVT_{SE}$	Signal Enhancement Echo Level Versus Time
ERL	Echo Return Loss
FFT	Fast Fourier Transform
FR_{AS}	Frequency Response Audio Subsystem
FR_{BDR}	Frequency Response Bidirectional transport, Receive
FR_{BDS}	Frequency Response Bidirectional transport, Send
FR_{MS}	Frequency Response Microphone, Send
FR_{SER}	Frequency Response Signal Enhancement Subsystem, Receive
FR_{SES}	Frequency Response Signal Enhancement Subsystem, Send
FR_{UD}	Frequency Response Unidirectional Transport
HATS	Head And Torso Simulator
HATS-HFRP	Head And Torso Simulator – Hands- Free Reference Point
HF System	Hands- Free System
HFT	Hands- Free Terminal
HVAC	Heating Ventilation Air Conditioning
$ICBN_{SE}$	Signal Enhancement Initial Convergence with Background Noise
IC_{SE}	Signal Enhancement Initial Convergence without background noise
JLR	Junction Loudness Rating
JLR_{BDR}	Junction Loudness Rating Bidirectional transport, Receive
JLR_{BDS}	Junction Loudness Rating Bidirectional transport, Send
JLR_{UD}	Junction Loudness Rating Unidirectional transport
L_{SER}	Receive signal Level Signal Enhancement
L_{SES}	Send output signal Level Signal Enhancement
LIN_{MS}	Microphone Input/Output Linearity, Send
$LIN_{REF-LSP}$	audio subsystem Linearity between REF and LSP
L_{MS}	Microphone output Level, Send
LQ_{AS}	Audio Subsystem speech Quality
LQ_{BDR}	Bidirectional transport speech Quality, Receive
LQ_{BDS}	Bidirectional transport speech Quality, Send
$LQBGN_{MS}$	Microphone speech Quality with Background Noise, Send
LQ_{MS}	Microphone speech Quality, Send
LQS_{BDR}	Bidirectional transport speech Quality Stability, Receive
LQS_{BDS}	Bidirectional transport speech Quality Stability, Send
LQS_{UD}	Unidirectional transport speech Quality Stability
LQ_{UD}	Unidirectional transport speech Quality

$L_{S,min}$	minimum activation Level (Send direction)
MOS	Mean Opinion Score
MOS-LQO_{SER}	Signal Enhancement Mean Opinion Score Listening Quality Objective, Receive
MOS-LQO_{SES}	Signal Enhancement Mean Opinion Score Listening Quality Objective, Send
MOS-LQOsw	Mean Opinion Score Listening Quality Objective, Superwideband
MRP	Mouth Reference Point
N_{AS}	Audio Subsystem idle Noise
N_{BDR}	Bidirectional transport idle Noise, Receive
N_{BDS}	Bidirectional transport idle Noise, Send
NC	Noise Criterion
NC_{BDR}	Bidirectional transport Noise Cancellation, Receive
NC_{BDS}	Bidirectional transport Noise Cancellation, Send
NC_{UD}	Unidirectional transport Noise Cancellation
NLMS	Normalized Least Mean Square
N-MOS-LQOn	Noise Quality Mean Opinion Score Listening Quality Objective, narrowband
N-MOS-LQOw	Noise Quality Mean Opinion Score Listening Quality Objective, wideband
N_{MS}	Microphone idle Noise, Send
NR	Noise Reduction
N_{UD}	Unidirectional transport idle Noise
OHC	Overhead Console
OL_{ASmax}	Audio Subsystem maximum Output Level
OVL_{AS}	Audio Subsystem Overload point
OVL_{EP}	acoustic Echo Path Overload point
OVL_{MS}	Microphone Overload point
PCM	Pulse Code Modulation
POI	Point Of Interconnection
PS	Perfect Sweep
QBGNTFSE	Enhanced Quality of Background Noise Transmission with Far-End Speech
RAL_{SE}	Signal Enhancement Activation Level, Receive
RA_{SE}	Signal Enhancement Activation, Receive
RAT_{SE}	Signal Enhancement Activation build up Time, Receive
REV_{MS}	Microphone Reverberation, Send
RFIFAP	Reference Interface Access Point
RLR	Receive Loudness Rating
Rxin	transport subsystem Receive input

Rxout	transport subsystem Receive output
SAL_{SE}	Signal Enhancement Activation Level, Send
SA_{SE}	Signal Enhancement Activation, Send
SEA_{SE}	Signal Enhancement Spectral Echo Attenuation
SII	Speech Intelligibility Index
SINR	Speech to Idle Noise Ratio
$SINR_{SER}$	Speech to Idle Noise Ratio Signal Enhancement, Receive
$SINR_{SES}$	Speech to Idle Noise Ratio Signal Enhancement, Send
SLR	Send Loudness Rating
S-MOS-LQOn	Speech Quality Mean Opinion Score Listening Quality Objective, narrowband
S-MOS-LQOw	Speech Quality Mean Opinion Score Listening Quality Objective, wideband
SNR	Signal to Noise Ratio
SNRD	SNR Difference
SNR_{M}	Microphone SNR
$SOOB_{SER}$	Spurious Out-Of-Band Signals After Signal Enhancement, Receive
SPD	Speech Pause Detection
SQPBGNE	Enhanced Speech Quality in the Presence of Background Noise
SRAIE	Speech Recognition Accuracy Indicator after Enhancement
SRW	Short Range Wireless transmission
SRWAP	Short Range Wireless Reference Access Point
SSA_{DT-SE}	Signal Enhancement Sent Speech Attenuation during Double-Talk
S_{si}(diff)	diffuse field Sensitivity
S_{si}(direct)	direct sound Sensitivity
STI	Speech Transmission Index
Sxin	transport subsystem Send input
Sxout	transport subsystem Send output
T_{AC}	Acoustic subsystem delay
T_{AS}	Audio Subsystem delay
T_{BDR}	Bidirectional Transport delay, Receive
T_{BDS}	Bidirectional Transport delay, Send
TCL_{AS}	Terminal Coupling Loss Acoustic Subsystem
TCL_{BD}	Terminal Coupling Loss Bidirectional transport subsystem
TCL_{SE}	Signal Enhancement Terminal Coupling Loss (wideband)
TCL_{W}	Weighted Terminal Coupling Loss
TCL_{WAS}	Weighted Terminal Coupling Loss Acoustic Subsystem

$TCLw_{BD}$	Weighted Terminal Coupling Loss Bidirectional Transport Subsystem
$TCLw_{SE}$	Signal Enhancement Weighted Terminal Coupling Loss (narrowband)
Test Point	subsystem access Point
T_r	Receive delay hands-free Terminal
Tr,R	built-up Time (Receive direction)
Tr,S	built-up Time (Send direction)
$T_{REF-LSP}$	audio subsystem delay between REF and LSP
T_{rtd-HF}	round trip delay Hands-Free terminal
T_s	Send delay hands-free terminal
T_{SER}	Receive delay Signal Enhancement
T_{SERTD}	Round Trip Delay Signal Enhancement
T_{SES}	Send delay Signal Enhancement
T_{UD}	Unidirectional transport delay
$TVEP_{SE}$	Signal Enhancement Echo Performance with Time Variant Echo Path
$TVEP_{SE}$-SP	Signal Enhancement Echo Performance with Time Variant Echo Path and Speech

5 Conventions

dBm: Absolute power level relative to 1 milliwatt, expressed in dB.

dBm0: Absolute power level in dBm referred to a point of zero relative level (0 dBr point).

dBm0p: Weighted dBm0, according to [ITU-T O.41].

dBm0(C): C-weighted dBm0, according to [ISO 1999].

dBov: dB relative to the overload point of a digital system according to [ITU-T G.100.1].

dBPa: Sound pressure level relative to 1 Pa, expressed in dB.

dBPa(A): A-weighted sound pressure level relative to 1 Pa, expressed in dB.

dBSPL: Sound pressure level relative to 20 µPa, expressed in dB; (94dBSPL=0dBPa).

dBV(P): P-weighted voltage relative to 1 V, expressed in dB, according to [ITU-T O.41].

dBr: Relative power level of a signal in a transmission path referred to the level at a reference point on the path (0 dBr point).

N: Newton.

Vrms: Voltage – root mean square.

cPa: Compressed Pascal, sound pressure at the output of the hearing model in the "Relative Approach" after nonlinear signal processing by the human ear.

Figures in this document are informative only, the relevant limits are derived from the associated tables.

6 How to use this Recommendation

This Recommendation uses the concept of performance classes for each parameter to be tested. It is the intention to identify the quality of the different subsystems by the classification based on the different parameters. The following general definitions apply:

Performance Class 1

Performance Class 1 characterizes an exceptionally good implementation of the relevant parameter. Performance Class 1 can typically only be achieved with special constraints or effort in terms of design, cost and time.

Performance Class 2

Performance Class 2 characterizes an implementation which most likely fulfils the requirements in [ITU-T P.1100] and [ITU-T P.1110]. Performance Class 2 can be expected to be representative for good implementations.

Performance Class 3

Performance Class 3 characterizes some weakness for the parameter under test. If a parameter is fulfilling just Performance Class 3 it may be acceptable only given that other parameters benefit from that or may compensate the weakness of this parameter, and the complete system still fulfils the requirements of [ITU-T P.1100] and [ITU-T P.1110].

Performance Class 4

Performance Class 4 characterizes an unacceptable weakness of the parameter under test.

Performance classes for individual parameters are defined for the following subsystems:

Acoustic Subsystem
Microphone Subsystem
Audio Subsystem
Signal Enhancement Subsystem
Unidirectional Transport
Bidirectional Transport
Network Transport

NOTE – A detailed interpretation of each performance class for each parameter is in the planning stage and have not yet been realized for this Recommendation.

As outlined in [ITU-T P.1100] and [ITU-T P.1110] the user scenarios as described in Annex B apply to all subsystems. The classification according to the performance classes need to take into account the impact of the different user scenarios described in Annex B on the individual parameter of the subsystem.

As a basic principle additivity of the parameters is assumed. This means that a parameter with a high performance for all subsystems will most likely lead to a good overall performance of the complete system. However, it is recognized that a weakness of one parameter in a subsystem does not necessarily lead to a weakness of this parameter for the overall system.

7 Architectures

The basic concept of this Recommendation can be seen in Figures 7-1 and 7-2. Figure 7-3 illustrates this concept by showing typical devices in the block diagram. This Recommendation uses a system architecture concept which is almost independent of the technical realization of the different subsystems. All elements in the abstract architecture may be found in technical realizations. The concept is realized in such a way that almost any technical realization of a car hands-free system can be mapped to the subsystem specification. As seen in the following sections different subsystems may be realized in one system. The signal transport between subsystems may be analogue or digital and it may be packet based as well as serial. Common to all is the access of the subsystems by accessing the subsystems using signal insertion or signal acquisition at the different signal transport subsystems associated with each subsystem by reference interfaces.

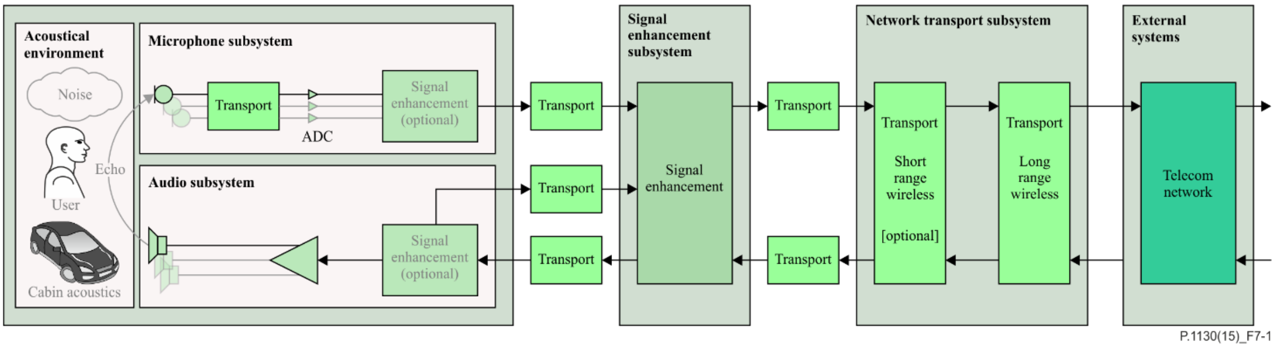


Figure 7-1 – Abstract architecture – System overview

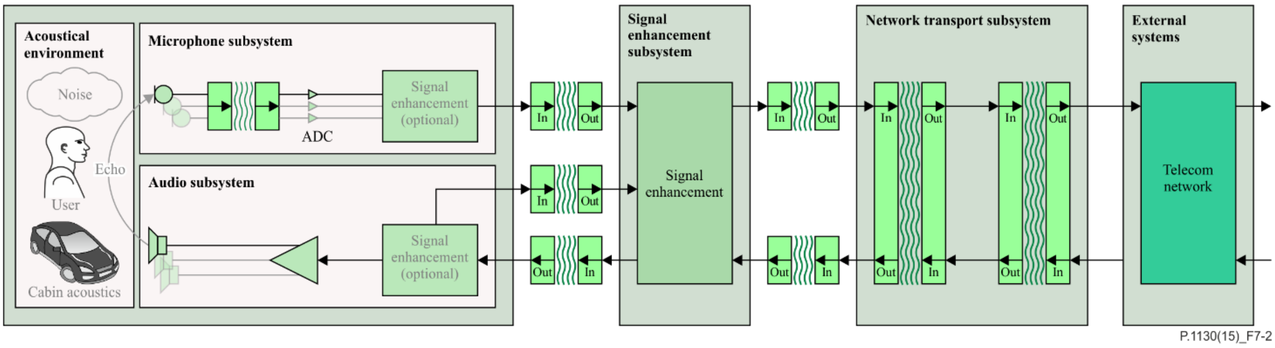


Figure 7-2 – System overview with detailed transport

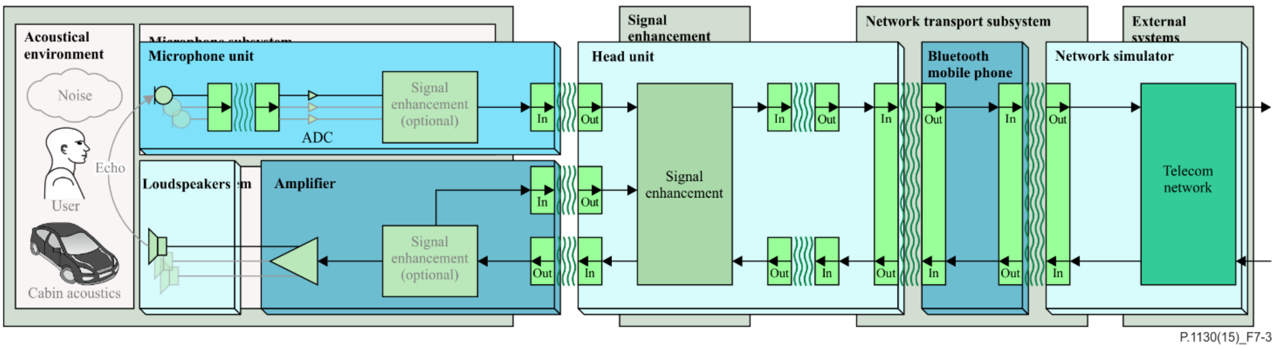


Figure 7-3 – System overview with the example of typical devices

7.1 Distributed speakerphone

The subsystems are tested as described in the different clauses of this Recommendation. The test points for the different subsystems are shown in Figures 7-4 and 7-5.

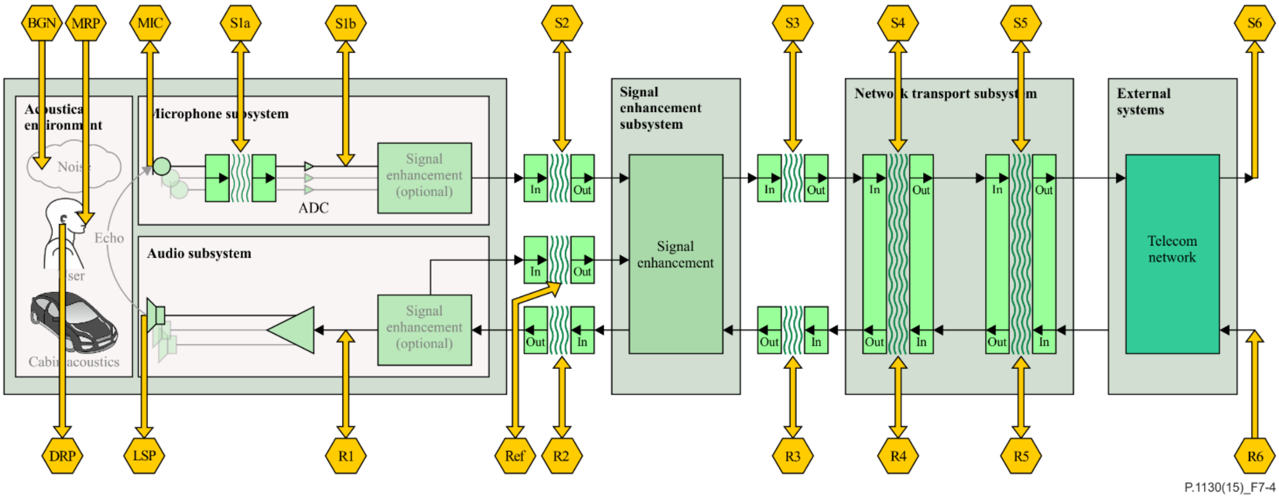


Figure 7-4 – Abstract architecture and test points

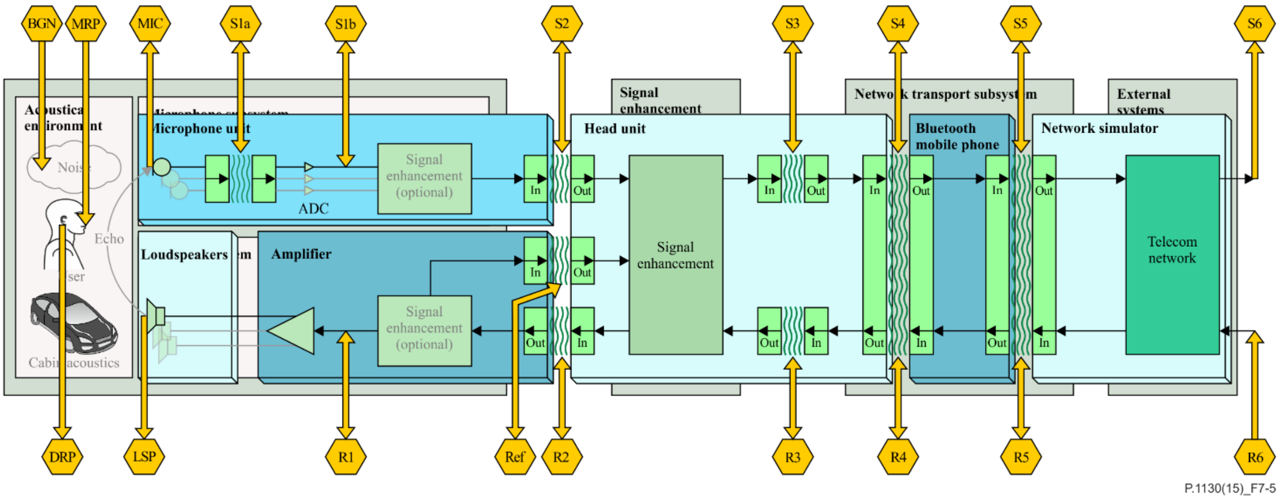


Figure 7-5 – System overview and test points; example of a typical implementation

7.2 System with integrated microphone

The system is tested as described in [ITU-T P.1100] (for narrowband systems) or in [ITU-T P.1110] (for wideband systems). If no network access is provided by the system, the system is connected to a reference interface which replaces the network access system. The requirements given in [ITU-T P.1100] and [ITU-T P.1110] apply.

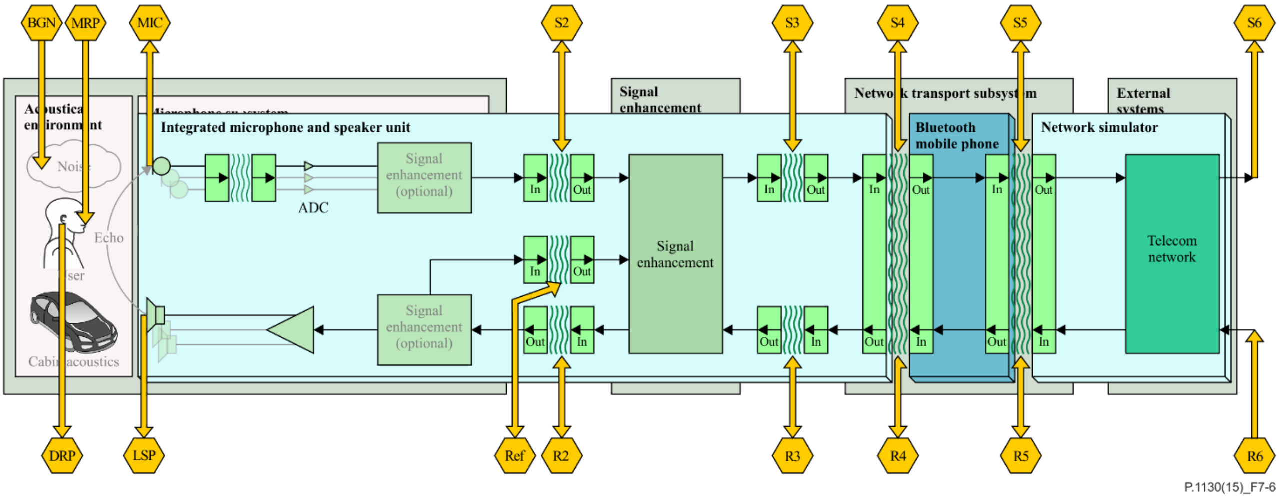


Figure 7-6 – System with integrated microphone; example of a typical implementation

The audio subsystem is tested as described in clause 8.3.3.

The network access subsystem is tested as described in clause 9.

7.3 System with integrated speaker

The system is tested as described in [ITU-T P.1100] (for narrowband systems) or in [ITU-T P.1110] (for wideband systems). If no network access is provided by the system the system is connected to a reference interface replacing the network access system. The requirements as given in [ITU-T P.1100] and [ITU-T P.1110] apply.

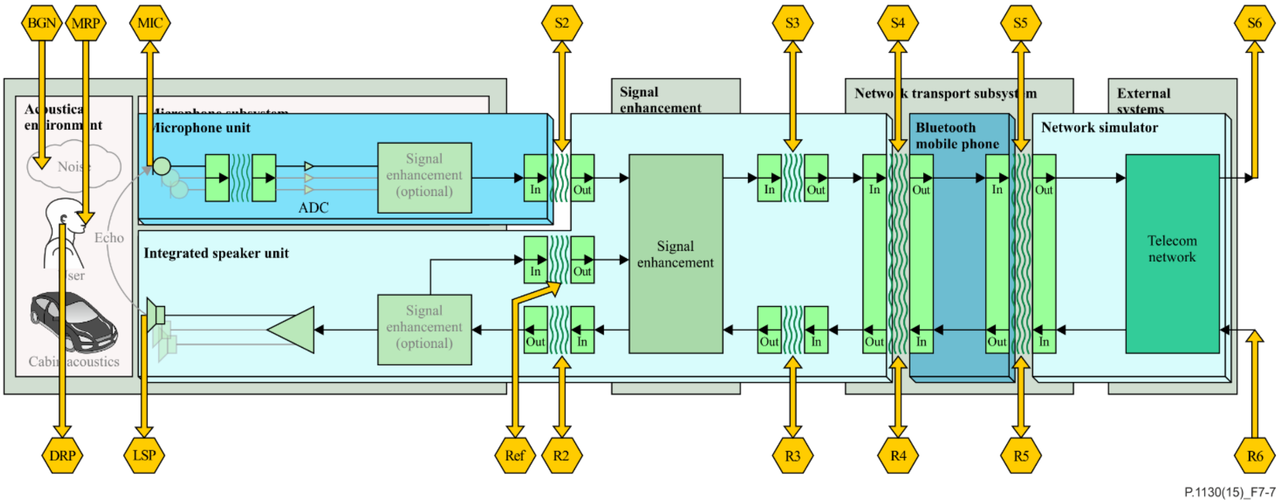


Figure 7-7 – System with integrated speaker, example of a typical implementation

The microphone subsystem is tested as described in clause 8.3.2.

The network access subsystem is tested as described in clause 9.

7.4 System with integrated microphone and speaker

The system is tested as described in [ITU-T P.1100] (for narrowband systems) or in [ITU-T P.1110] (for wideband systems). If no network access is provided by the system the system is connected to a reference interface replacing the network access system. The requirements given in [ITU-T P.1100] and [ITU-T P.1110] apply.

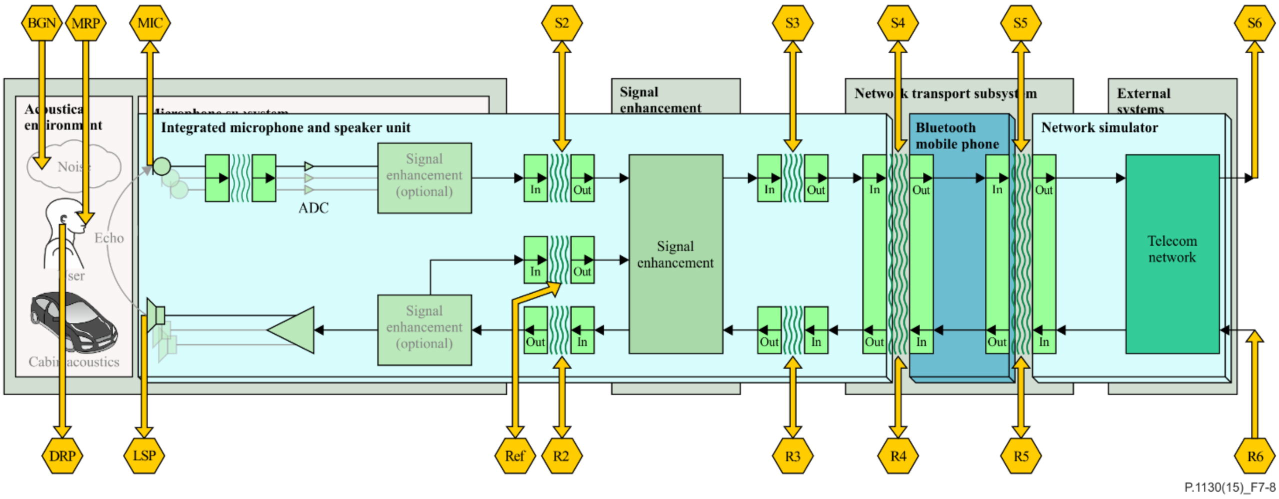


Figure 7-8 – System with integrated microphone and speaker; example of a typical implementation

Figure 7-8 shows a system with integrated microphone and speaker. The electrical access points used may be either S5/R5 (POI connected to a radio network simulator as described in [ITU-T P.1100] and [ITU-T P.1110]) or S4/R4 connected to a reference interface (e.g., SRW reference interface). In case of short range wireless transmission S4/R4 is the SRWAP.

The acoustical access points are identical to those given in [ITU-T P.1100] and [ITU-T P.1110].

8 Subsystems

8.1 System delay, subsystem delay and buffering

The system delay is measured as described in [ITU-T P.1100] for narrowband systems and [ITU-T P.1110] for wideband systems. The subsystem delay as described and measured within this specification only includes the algorithmic delay. In case a subsystem is implemented in hardware the computational delay is measured (see Annex C).

It is known that buffering which is needed to for example, take care of different frame sizes in subsystems may cause a significant increase in the system delay. Therefore, buffering and frame sizes should ideally be optimized to get a minimum system delay.

Annex C describes synchronization issues and the impact of different frame sizes and buffering in detail and gives guidance on how to optimize subsystems to achieve a mostly low system delay.

8.2 Test interface description

8.2.1 Analogue electrical interface

The analogue interface of the test system should match the impedance and the sensitivity of the system under test.

8.2.2 Digital electrical interface

The digital interface of the test system should provide a dynamic range of at least a 16 bit equivalent and should otherwise match the interface specification of the system under test.

8.2.3 Acoustic interface

8.2.3.1 Artificial mouth

The artificial mouth of the artificial head shall conform to [ITU-T P.58]. The artificial mouth is equalized at the MRP according to [ITU-T P.340].

The sound pressure level is calibrated at the HATS-HFRP so that the average level at HATS-HFRP is −25.7 dBPa. The sound pressure level at the MRP has to be corrected accordingly. The detailed description for equalization at the MRP and level correction at the HATS-HFRP can be found in [ITU-T P.581].

NOTE – The 3 dB speech level increase according to [ITU-T P.340] and applicable for all hands-free tests in the send direction is taken into account independently of the Lombard effect described below.

When testing with vehicle noise, the output level of the mouth is increased to account for the "Lombard effect". The Lombard effect refers to the change in speaking behaviour caused by acoustic noise. The level is increased by 3 dB for every 10 dB that the long-term A-weighted noise level exceeds 50 dB(A) [b-Kettler/Gierlich] (averaged over the entire background noise recording used for the measurement). This relationship is shown in the following formula:

$$I(N) = \begin{cases} 0 & \text{for} \quad N < 50 \\ 0.3(N-50) & \text{for} \quad 50 \le N \le 77 \\ 8.0 & \text{for} \quad N > 77 \end{cases}$$

where:

I = The dB increase in mouth output level due to noise level

N = The long-term A-weighted noise level measured near the driver's head position

As an example, if the vehicle noise measures 70 dB(A), then the output of the mouth would be increased by 6 dB. No gain is applied for noise levels below 50 dB(A). The maximum amount of gain that can be applied is 8 dB. Vehicle noise levels are measured using a measurement microphone positioned at the driver's head position next to the driver's outboard ear.

8.2.3.2 Artificial ear

For speakerphone hands-free terminals the ear signals of both ears of the artificial head are used. The artificial head is free-field or diffuse-field equalized (see [ITU-T P.1100] and [ITU-T P.1110]); more detailed information can be found in [ITU-T P.581].

8.2.4 Wireless interface

The interface of the test system should conform to the interface specification of the wireless system under test.

8.2.5 Test signals and signal levels

Speech-like signals are used for the measurements which can be found in [ITU-T P.501]. Detailed information about the test signal used is to be found in the corresponding clause of this Recommendation. Wherever possible the speech signals described in [ITU-T P.501], clause 7.3 are used.

NOTE – For single talk measurements, in cases where it can be shown that the subsystem signal processing does not affect the measurement result when using a shorter version of the single talk sequence of clause 7.3.2

([ITU-T P.501]), a shorter sequence consisting of two sentences may be used. In such an event the following two sentences (1 male, 1 female voice) covering the low pitch frequency of male voices and the typically higher energy in the high frequency range for female voices should be used:

"The last switch cannot be turned off" (sentence 1).

"The hogs were fed chopped corn and garbage" (sentence 6).

In general, the test signal levels are individual to the system design and vary between implementations. However if not stated otherwise, as a general rule the following test signal levels shall be used:

For digital interfaces the nominal signal level is derived by measuring the signal level at the corresponding digital input of the subsystem when applying a signal with the nominal level either at the acoustical interface of the hands-free terminal or at the POI.

Typically the nominal test signal level should be −22 dBov at the digital interface for digital systems (assuming a nominal signal level of −16 dBm0 and an overload point of 3.14 dBm0 for digital transmission systems).

For analogue interfaces a similar approach can be taken. The nominal signal level is derived by measuring the signal level at the according analogue input of the subsystem when applying a signal with nominal level either at the acoustical interface of the hands-free terminal or at the POI.

8.3 Acoustic subsystem

The acoustic subsystem and related test points are shown in Figure 8-1. This figure also shows that the acoustic subsystem can be further divided into the following subsystems:

– Microphone subsystem (send path)

– Audio subsystem (receive path).

NOTE – The acoustic environment is considered a subsystem, but no specific measurements are applied.

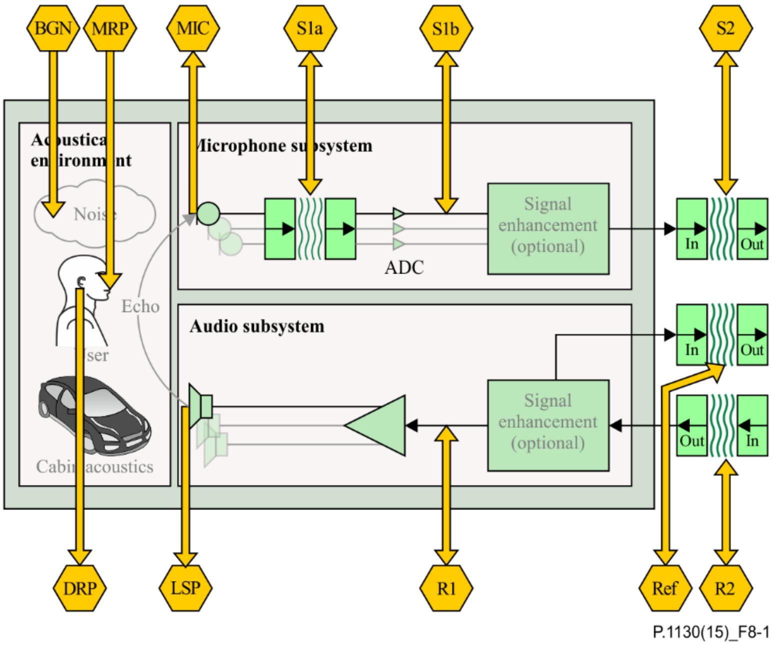


Figure 8-1 – Acoustic subsystem and test points

8.3.1 Acoustic subsystem measurements

The purpose of these measurements is to ensure proper synchronization and timing between these subsystems when realized digitally. Furthermore, the tests are intended to verify that the roundtrip echo path (i.e., R2-to-S2) will not cause failures in the signal enhancement subsystem or overall system roundtrip delay requirements.

Issues which can be observed with improper implementations are:

– echo falls out of the echo cancellation processing window causing constant echo;

– echo canceller divergence due to echo path change caused by delay drift;

– echo canceller divergence due to echo path distortions;

– dropped speech frames due to buffer overflow in non-synchronized systems;

– delay jumps due to buffer readjustments;

– difficulty carrying on a natural conversation due to excessive roundtrip delay.

8.3.1.1 Clock synchronization accuracy

8.3.1.1.1 Parameter description

The clock synchronization accuracy is measured between test points R2 and S2 as well as between R2 and Ref. If available in the system design, the clock synchronization accuracy between R2 and test points S1a and R1 are considered as well. The clock synchronization accuracy is measured based on clock drift.

The system delay t_{System} depends on the transmission method used and the delay of the reference interface. The delay t_{System} must be known and deducted from the test result.

8.3.1.1.2 Test

1) The system delay vs. time TAC (t) is measured at test points R2 and S2 as well as between R2 and Ref (in addition at S1a and R1 if present in the system).

2) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (R2) at a level according to clause 8.1.5.

3) The composite source signal is repeatedly inserted to achieve a test signal length of at least 180 s.

4) The reference signal is the original signal (test signal).

5) The test set-up is according to clause 8.3.

6) The clock drift is determined by cross-correlation analysis between the measured signal at test point S2 and Ref (and R1 and S1a if present) and the original signal. The measurement is corrected by delays which are caused by the test equipment.

7) The delay vs. time is measured by determining for each CSS burst the delay in ms; the maximum of the cross-correlation function is used for the determination.

$$T_{measure}\,(t) = T_{AC}\,(t) + T_{System}$$

8) The delay vs. time graph is analysed to estimate the clock drift in parts per million (PPM). In a first step, the raw drift *D* is calculated as a derivative of the delay vs. time series T_{AC} (t):

$$D = \frac{M}{\mathrm{Step\,Size}} * \frac{\partial T_{AC}(t)}{\partial t}$$

9) The step size is the same as the one used in the delay vs. time calculation; the normalization factor *M* depends on the scaling of the time and delay axis and usually is set to 1E6 (assuming time and delay axis are given in seconds).

NOTE – In case the raw drift *D* is noisy, a smoothing may be applied. The 20% and 80% percentiles are determined and used as a limiter. Only values inside these limits are taken into account for an average calculation of the PPM. This step may be necessary in order to compensate potential delay jumps due to buffer over/underrun at the interfaces.

8.3.1.1.3 Performance level classification based on values of this parameter

Table 8-1 – Limits for the clock drift between the interfaces

Performance Class	Clock drift in ppm
1	0 ppm
2	< 2 ppm
3	< 50 ppm
4	> 50 ppm

8.3.1.1.4 Design guidance and root-cause analysis

Zero clock drift is essential for a proper operation of all signal processing using the input signals at the different interfaces such as echo compensation. Clock drift may lead to a complete mismatch of the filters, e.g., in the echocanceller and therefore will deteriorate the echo performance of the overall system. A synchronous system design is preferable.

8.3.1.2 Acoustic subsystem delay

8.3.1.2.1 Parameter description

The delay T_{AC} is measured from test point (R2) to test point (S2). The delay T_{AC} should be minimized. The measurement is conducted at nominal volume control setting.

The system delay t_{System} depends on the transmission method used and the delay of the reference interface. The delay t_{System} must be known and deducted from the test result.

8.3.1.2.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (R2) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 8.3.

2) The delay is determined by cross-correlation analysis between the measured signal at test point (S2) and the original signal. The measurement is corrected by delays which are caused by the test equipment. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{AC} + T_{System}$$

8.3.1.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the round trip delay has to meet the requirements for the applicable performance classes for the delay T_{AS} defined in Table 8-2.

Table 8-2 – Limits for the acoustic subsystem delay for speech communication services

Performance Class	T_{AS}
1	< 12 ms
2	< 20 ms
3	< 40 ms
4	≥ 40 ms

8.3.1.2.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay low leads to higher impairment resulting even from low level echo components (see [ITU-T G.131]). Therefore, any design which provides for as low a delay in the connection as possible is preferable. Furthermore, higher delays may impair the performance of the echo canceller in the signal enhancement subsystem.

8.3.1.3 Acoustic subsystem echo path overload point

8.3.1.3.1 Parameter description

The acoustic echo path overload point OVL_{EP} of the acoustic subsystem is measured from test point (R2) to test point (S2). The measurement is conducted at a receive volume level setting of 15 dB above the nominal (i.e., RLR = 2) volume control setting.

The acoustic subsystem overload point is determined by the highest signal which is produced by the audio subsystem in the car without distorting the echo path.

8.3.1.3.2 Test

1) The test set-up is according to clause 8.3.
2) The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with the nominal average signal level applied at test point (R2). The test signal level is the average level of the complete test signal.
3) The maximum spectrum generated by the car audio system is recorded at test point (S2).
4) The coherence between the signal measured at S2 and the signal inserted at R2 is determined as an average of the coherence function in the frequency range from 100 Hz to 7 kHz for wideband and 300 Hz to 3.5 kHz for a narrowband system.

8.3.1.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the acoustic echo path overload point has to meet the requirements for the applicable performance classes for the acoustic echo path overload point defined in Table 8-3.

Table 8-3 – Limits for the acoustic echo path overload point above 0.80 coherence

Performance Class	OVL_{EP}
1	0 dB
2	−3 dB
3	−6 dB
4	−9 dB

8.3.1.3.4 Design guidance and root-cause analysis

Distortion in the acoustic echo path will prevent the acoustic echo canceller in the signal enhancement subsystem from working well and can result in echo, or temporal clipping, being heard by the person on the far end of the telephone connection. In the case of speech recognition, it can result in poor barge-in performance. These impairments can be mitigated by limiting the maximum receive volume, modifying the receive frequency response, or replacing components/subsystems causing the distortion.

8.3.1.4 Acoustic subsystem weighted terminal coupling loss ($TCLw_{AS}$) and (TCL_{AS})

8.3.1.4.1 Parameter description

This parameter measures acoustic echo loss of the acoustic subsystem. Echo loss is measured from the input to the acoustic subsystem in the receive direction (R2), to the output of the acoustic subsystem in the send direction (S2). In narrowband TCLw is measured; in wideband the unweighted TCL is measured.

8.3.1.4.2 Test

1) The test arrangement is in accordance with clause 8.3.

2) The noise level measured at the test point (idle channel noise) shall be less than −63 dBm0. The attenuation between the input to the acoustic subsystem (R2) to the output of the acoustic subsystem is measured using the compressed speech test signal as described in [ITU-T P.501], clause 7.3.3, Amendment 1. The test signal level is −10 dBm0.

3) The first 17.0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences). For the calculation, the average measured echo level at each frequency band is referred to the average level of the test signal measured in each frequency band.

4) In narrowband TCL_W is calculated according to clause B.4 of [ITU-T G.122], (trapezoidal rule). In wideband the differences between the averaged echo level and the averaged test signal level in a frequency range from 100 Hz-8 000 Hz is calculated.

8.3.1.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the echo loss through the acoustic subsystem has to meet the requirements for the applicable performance classes for the TCL_{AS} (wideband) and $TCLw_{AS}$ as defined in Table 8-4.

Table 8-4 – Limits for the acoustic subsystem TCL

Performance Class	$TCLw_{AS\ (narrowband)}$	$TCL_{AS\ (wideband)}$
1	> 46 dB	> 46 dB
2	> 20 dB	> 20 dB
3	> 6 dB	> 6 dB
4	< 6 dB	< 6 dB

8.3.1.4.4 Design guidance and root-cause analysis

Poor performance can be investigated further by repeating the measurements defined in this clause but with using different measurement points such as R1 and S1b (see Figure 8-1). Additionally, other measurement parameters such as send and receive frequency responses in the microphone and audio subclauses may prove useful in diagnosing performance issues.

8.3.1.5 Time invariance of processing

This measurement parameter is for further study. It is intended to measure time-varying effects that can negatively affect performance of the signal enhancement subsystem (e.g., adaptive beamforming, noise reduction, etc.).

8.3.2 Microphone subsystem (send path)

The microphone subsystem may consist of one or more microphones and may include a signal enhancement layer for the microphone signal as well.

NOTE – Besides the parameters used in this document the following parameters which are not currently covered might influence the quality of the microphone subsystem:

– change of speech-to-noise ratio due to soft and loud talkers
– wind buffeting noise
– specific issues with belt microphones such as structure-borne noise
– delay.

For multi-microphone solutions the following additional points might be considered:

– delay variation (over sensors and time)
– frequency response variation (over sensors and time)
– level variation (over sensors and time).

The purpose of this section is to provide guidance for positioning microphones and for optimizing the microphone performance parameters in such a way that the microphone provides optimum input signals to the speech enhancement.

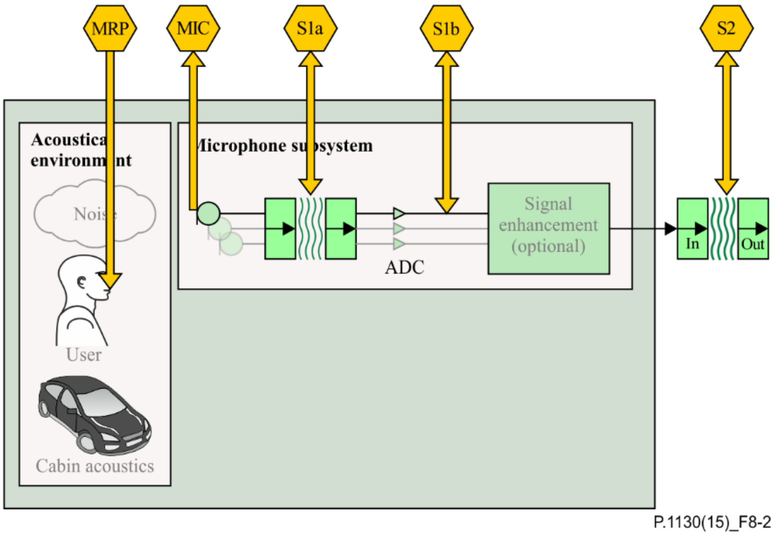


Figure 8-2 – Microphone subsystem and test points

The DUT is connected to the reference interfaces. The reference interfaces are calibrated as specified in clause 8.2.

This clause is applicable to single microphones but not to the output of microphone arrays.

8.3.2.1 Microphone measurements in the car

Positioning of hands-free microphones

The speech quality in hands-free communication is significantly affected by the positioning of the hands-free microphone. As the optimal microphone position can vary strongly depending on vehicle design as well as on specific requirements, there is no universally valid rule for the positioning of the microphone. However, there are some guidelines which should be considered. Nevertheless, in practice, this often means to find the best compromise, as not all requirements can be equally fulfilled.

– The hands-free microphone should always be placed as close as possible to the speaker, as within the near field of a sound source (in a vehicle this is up to 80-100 cm)[1] the speech level drops by $1/d^2$. In practical applications, this typically means an analogous loss in signal-to-noise ratio. For this reason, a single microphone placed nearby might give a better performance than a microphone array which is placed further away.

– There has to be a direct path between the speaker's mouth and the microphone. If this is not given, this might result in a significant decrease in signal-to-noise ratio as well as in speech quality since the speech signal becomes reverberant.

[1] The near field is characterized by the distance (measured from the sound source) where the direct sound and the reflected sound are of equal intensity. In acoustics, this distance is often referred to as critical distance.

– The direction of the highest sensitivity of the microphone should point in the direction of the speaker's mouth. If different seating positions or several speakers are to be covered by one microphone, a compromise for the microphone position has to be found, as the direction of the highest sensitivity might not cover all. However, this often means a significantly reduced performance in comparison to an optimal alignment of the microphone for a single speaker. In this case, the application of additional microphones might be considered to achieve an optimal speech quality.

– A direct airstream to the microphone, e.g., from the air conditioning, has to be avoided as the speech signal might be highly disturbed by wind buffeting.

– Saturation of the microphone by loudspeakers nearby, e.g., by a centre-speaker, has to be avoided. If necessary, the levels of the affected loudspeakers have to be reduced.

The coupling of structure-borne sound to the microphone has to be avoided.

NOTE 1 – When the microphone is measured in the car, it is recommended to use the power supply provided by the car/car hands-free system.

NOTE 2 – If the microphone is integrated digitally in the car, it is recommended to measure the microphone at a digital access point, if available. Care should be taken in order to correctly calibrate the access point.

NOTE 3 – In case the car is not available, it is recommended to integrate the microphone in a surrogate environment as close as possible to the production environment. In case wind noise may have a major influence on the microphone performance it is recommended to also simulate the wind noise.

8.3.2.1.1 Microphone overload point (in the car)

8.3.2.1.1.1 Parameter description

The microphone overload point OVL_{MS} is measured from test point (Mic) to test point (S1a) or test point (S2).

The microphone overload point is determined by the highest signal which is received by the microphone in the car without distortion. This may be the speech signal of the near-end speaker, the background noise picked up by the hands-free microphone or the signal received from the car audio system.

8.3.2.1.1.2 Test

1) The maximum spectrum is recorded by a reference microphone close to the hands-free microphone (MIC) by generating background noise, near-end speech and the signal generated by the car audio system with maximum volume.
   - The background noise conditions to be simulated are found in Annex B of [ITU-T P.1100] and [ITU-T P.1110]. The background noise producing the maximum level at the microphone is used.
   - The maximum near-end signal is generated by generating the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with a level as described in clause 8.2.3.1.
   - The maximum loudspeaker signal is generated by injecting the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] the nominal signal level at the far end and setting the audio volume control to maximum.
2) For the signal producing the maximum level (Lev_{max}) the coherence between the reference microphone positioned at MIC and S2 is determined. The coherence is determined as an average of the coherence function in the frequency range from 100 Hz to 7 kHz.
3) The measurement is repeated by reducing the signal level of the worst case signal in steps of 1 dB until the required coherence is achieved.

NOTE – The British-English single talk sequence-based tests are averaged over the complete speech sequence voice.

8.3.2.1.1.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone overload point has to meet the requirements for the applicable performance classes for the microphone overload point defined in Table 8-5.

Table 8-5 – Limits for the microphone overload point above 0.8 coherence

Performance Class	OVL re Lev_{max}
1	0 dB
2	−3 dB
3	−6 dB
4	−9 dB

8.3.2.1.1.4 Design guidance and root-cause analysis

The microphone is the first element in the transmission system of the hands-free system. Its dynamic range should be adapted to the type of car it is used in. Besides other factors this requires careful consideration of the maximum acoustical level it may be exposed to. Any amplitude distortion (clipping) introduced by the microphone may impact the speech sound quality, the transmission quality of the background noise, the performance of any noise canceller and the performance of echo cancellers.

8.3.2.1.2 Microphone frequency response (in the car)

8.3.2.1.2.1 Parameter description

The microphone frequency response in send FR_{MS} is measured from test point (Mic) to test point (S1a) or test point (S2).

8.3.2.1.2.2 Test

1) The test arrangement is according to clause 8.3.2.
2) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] is used as the test signal. Alternatively, a periodic noise signal or CS signal according to [ITU-T P.501] can be used. The correct activation of the measurement object during the measurement has to be ensured by the test lab. The artificial mouth is equalized at the MRP, the test signal level shall be −1.7 dBPa at the MRP. The test signal level is the average level of the complete test signal. Finally, the level at the HATS-HFRP is adjusted to −25.7 dBPa.
3) The measured power density spectrum at the MRP is used as the reference power density spectrum for determining the frequency response in the send direction.
4) The sensitivity frequency response is determined in third octave intervals as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz inclusive. For calculation, the average measured level is referred to the level of the reference signal in each frequency band averaged over the complete test sequence length.
5) The sensitivity is determined in dBV/Pa.

8.3.2.1.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response has to meet the requirements for the applicable performance classes for the microphone send sensitivity as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 8-6 to 8-9 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 8-6 – Tolerance mask for the send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	2	–
125	2	−6
315	2	−2
2 500	2	−2
3 150	(Note 1)	−2
5 000	(Note 1)	−2
6 300	3	−2 (Note 2)
10 000	3	−2 (Note 2)
12 500	3	–
NOTE 1 – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale. NOTE 2 – If the microphone is purely used for speech communication and not used for speech recognition or other services, an upper limit of 7 kHz is sufficient.		

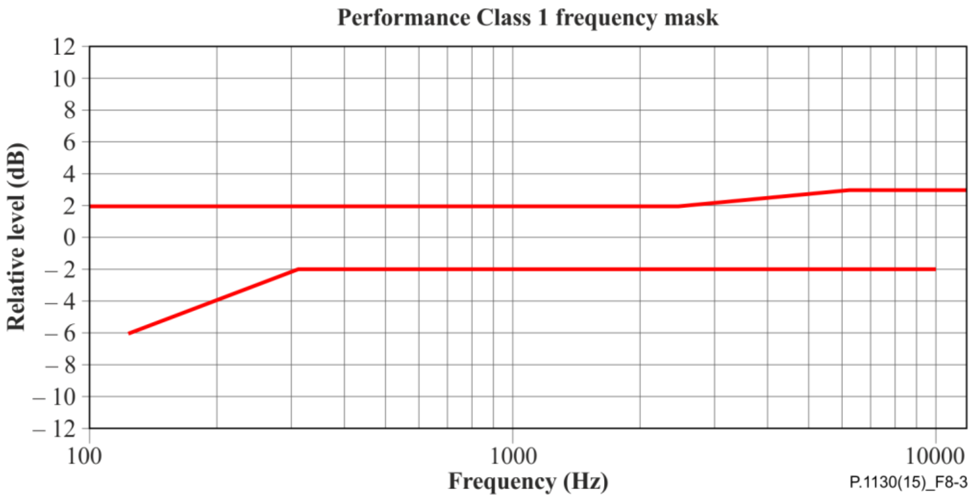


Figure 8-3 – Microphone frequency response mask (Figure is informative)

Table 8-7 – Tolerance mask for the send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	–	–
200	2	−6
500	2	−4
1 000	2	−2
2 500	4	−2
3 150	(Note)	−2
5 000	(Note)	−2
6 300	5	−6
8 000	–	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

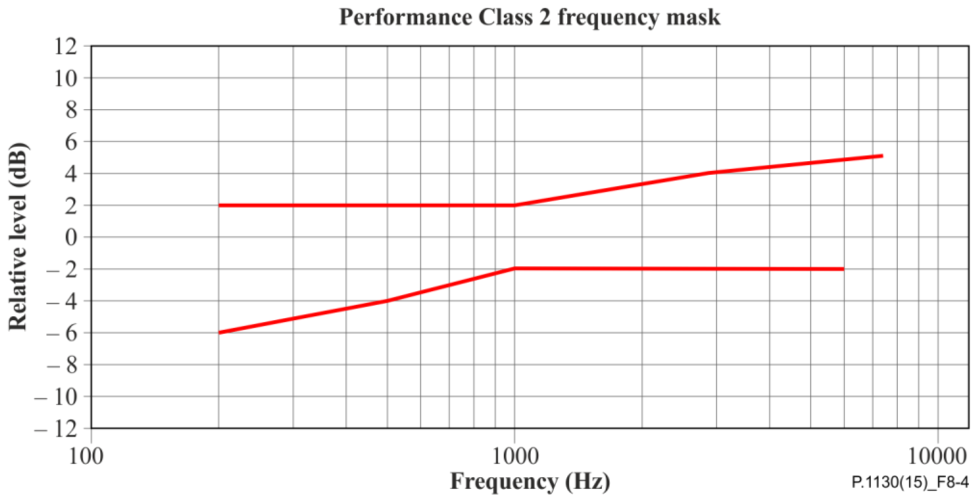


Figure 8-4 – Microphone frequency response mask (Figure is informative)

Table 8-8 – Tolerance mask for the send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	–	–
315	4	−10
500	4	−8
1 000	4	−4
2 500	8	(Note)

Table 8-8 – Tolerance mask for the send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
3 150	8	−4
5 000	8	–
6 300	8	–
8 000	–	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

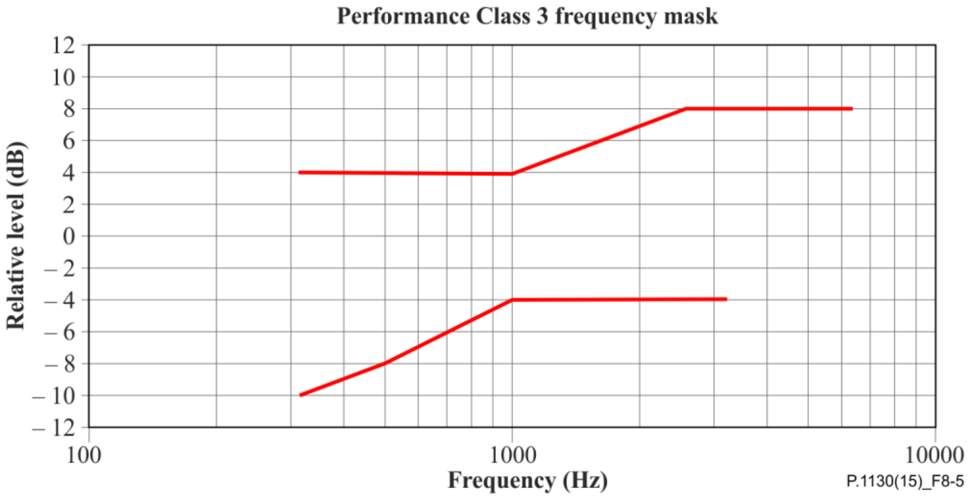


Figure 8-5 – Microphone frequency response mask (Figure is informative)

Table 8-9 – Tolerance mask for the send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	N/A	N/A
300	N/A	N/A
500	N/A	N/A
1 000	N/A	N/A
2 500	N/A	N/A
3 100	N/A	N/A
5 000	N/A	N/A
6 300	N/A	N/A
8 000	N/A	N/A
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

NOTE – Performance Class 1 mask is intended to support high quality wideband with no need to equalize the response. Performance Class 2 is intended to support high quality narrowband with no need to equalize the response. Performance Class 3 is intended to support narrowband communication and some equalization may be required.

8.3.2.1.2.4 Design guidance and root-cause analysis

NOTE – Ideally, the response characteristics of the microphone should be flat in the frequency range of wideband transmission (100 Hz-7 kHz). However, especially in the presence of background noise, a bandwidth limitation may be desirable. No explicit recommendation can be given here since such limitation would depend on the level and spectral content of the background noise and ideally should be adaptive. If however, a bandwidth limitation is introduced, it should be made at both the high and low frequencies.

8.3.2.1.3 Microphone idle channel noise (in the car)

8.3.2.1.3.1 Parameter description

The microphone idle channel noise in send N_{MS} is measured from test point (Mic) to test point (S1a) or test point (S2).

8.3.2.1.3.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with an average signal level of −1.7 dBPa is applied at the MRP. The test signal level is the average level of the complete test signal. Finally, the level at the HATS-HFRP is adjusted to −25.7 dBPa. The output level at the hands-free microphone is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) For the noise measurement, no test signal is used. However, all sources which potentially contribute to noise at the output of the microphone such as GSM noise, electrical noise introduced by the car must be considered. In order to ensure a reliable activation of active microphone arrangements, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of 4 composite source signals according to [ITU-T P.501]. The spectrum of the test at the MRP is calibrated under free-field conditions. The level of the activation signal is −25.7 dBPa measured at the HATS-HFRP.

3) The test arrangement is according to clause 8.3.2.

   The idle channel noise is measured at the microphone subsystem output in the frequency range between 100 Hz and 8 kHz. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers shall be taken into account; the time window must be shifted accordingly. The length of the time window is 1 second which is the averaging time for the idle channel noise. The test laboratory has to ensure the correct activation of the microphone/microphone arrangement during the measurement. If the microphone arrangement is deactivated during measurement, the measurement window has to be cut to the duration when the microphone remains activated.

   The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hann window is used.

   If it is known that the microphone stays activated without any activation signal, no activation signal is required. In this case, a simple noise measurement is conducted.

4) The idle channel noise is determined by A-weighting and referring to the reference speech signal level as determined with speech sequence.

5) Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis which has a noise bandwidth of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3-rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from $2^{(-1/6)}f$ to $2^{(+1/6)}f$). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum up to 3.4 kHz in narrowband and up to 6.8 kHz in wideband.

8.3.2.1.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 8-10.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 8-10 – Limits for the idle channel noise re. reference speech signal level

Performance Class	SINR
1	> 40 dB
2	> 30 dB
3	> 25 dB
4	≤ 25 dB

8.3.2.1.3.4 Design guidance and root-cause analysis

Be sure that the receive path is deactivated. Check shielding and check microphone preamplifier should not demodulate RF into NF.

8.3.2.1.4 Microphone SNR (in the car)

8.3.2.1.4.1 Parameter description

The SNR measurement is based on individual broadband estimations of the speech signal power and the noise signal.

NOTE 1 – It is recognized that fan noise, which varies from car to car and depends upon the relative positioning of the microphone and fan, may contribute significantly to the noise perceived by the far-end user. In order to determine the impact of the level and spectral content of this noise under different operating conditions, a noise test as described below may be used.

NOTE 2 – The method described requires a stationary type of background noise and therefore it should be applied for constant driving conditions or conditions where the background noise can be considered to be stationary.

The microphone signal to noise ration SNR_M is measured at test point (S1a) or test point (S2). Test point (S2) should only be used if test point (S1a) is not accessible.

8.3.2.1.4.2 Test

1) The test arrangement is according to clause 8.3.2.

2) The speech test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. To account for the Lombard effect, the test signal level is adjusted to the level specified in clause 8.2; the test signal level is measured as "active speech level" according to [ITU-T P.56].

3) For the measurement speech and background noise are recorded simultaneously while simulating the background noise.

4) The actual measurement is conducted with the hands-free microphone.

5) The SNR is calculated according to Appendix II.

NOTE – The method is applicable in the SNR range from −10 dB to +30 dB. Care should be taken that the resulting SNR in the measurement is within this SNR-range.

The user scenarios as defined in Annex B have to be applied.

8.3.2.1.4.3 Performance level classification based on values of this parameter

Performance requirements are for further study.

NOTE – It is not yet clear whether this method can be used to adequately predict the desired performance.

Background noises for driving speed ≤ 80 km/h

To claim compliance with a certain performance class the SNRD has to meet the requirements for the applicable performance classes for the SNRD defined in Table 8-11.

Table 8-11 – Limits for the SNRD

Performance Class	SNRD
1	FFS
2	FFS
3	FFS
4	FFS

Background noises for driving speed ≤ 120 km/h

To claim compliance with a certain performance class the SNRD has to meet the requirements for the applicable performance classes for the SNRD defined in Table 8-12.

Table 8-12 – Limits for the SNRD

Performance Class	SNRD
1	FFS
2	FFS
3	FFS
4	FFS

8.3.2.1.4.4 Design guidance and root-cause analysis

Insufficient SNRD can be caused by an insufficient directivity of the microphone used, by the wrong orientation of the microphone, by the wrong microphone placement (e.g., too far away from the speaker) or by too high background noise at the microphone location. This measurement parameter could be used to help optimize the hands-free microphone design (position and orientation etc.) within the vehicle.

8.3.2.1.5 Microphone send reverberation (SRV)

8.3.2.1.5.1 Parameter description

The microphone reverberation in send REV_{MS} is measured from test point (MRP) to test point (S1a) or test point (S2).

A measure for determining the subjectively perceived reverberation is defined in [ITU-T P.340]. The early energy balance (EEB) is a measure which seems to correlate to the subjectively perceived reverberation when transmitting reverberant speech over hands-free terminals. The EEB is defined as:

$$EEB = 10 \log \left\{ \frac{\int_0^{35\text{ ms}} h^2(t)dt}{\int_0^{5\text{ ms}} h^2\,(t)dt} \right\} \text{ dB}$$

8.3.2.1.5.2 Test

1) The test arrangement is according to clause 8.3.2.
2) The test signal should be composed of an activation signal (e.g., voiced sound of the CSS) and a maximal-length sequence of several periods as described in [ITU-T P.340]. The period length must be longer than the length at the impulse response of the system under test.
3) The impulse response is calculated from the average of several periods, beginning with the second period.
4) The delay introduced by the system under test and the test system has to be compensated for.
5) From this the early energy balance (EEB) is calculated.

8.3.2.1.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the EEB has to meet the requirements for the applicable performance classes for the EEB defined in Table 8-13.

Table 8-13 – Limits for the EEB

Performance Class	EEB
1	FFS
2	FFS
3	FFS
4	FFS

8.3.2.1.5.4 Design guidance and root-cause analysis

Too high reverberation may lead to hollow sounding speech. The reverberation may be caused by the positioning of the microphone(s) too close to reflecting surfaces (typically windows) or by insufficient damping of reflecting surfaces in a car. Additional damping of surfaces or changing the microphone position may help to correct the problem.

8.3.2.1.6 Microphone send input/output linearity

8.3.2.1.6.1 Parameter description

The microphone send input/output linearity LIN_{MS} is measured from test point (Mic) to test point (S1a) or test point (S2).

The microphone linearity is determined by the measurement of the A-weighted output level between input and output in the level range from −40 dB to +6 dB re. to the nominal input signal level and referring the measured level to the input signal level.

8.3.2.1.6.2 Test

1) The test arrangement is according to clause 8.3.2.
2) The spectrum at MIC is recorded by a reference microphone close to the hands-free microphone (MIC).
3) The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The correct activation of the measurement object during the measurement has to be ensured by the test lab. The artificial mouth is equalized at the MRP; the test signal level shall be −30 dB, −20 dB, −10 dB, 0 dB and +8 dB re to −1.7 dBPa at the MRP. The test signal level is the average level of the complete test signal.
4) The A-weighted output level is determined at test point S2 and referred to the input signal level.
5) The measurement is repeated for all input signal levels.

NOTE – For microphones that include adaptive processing, it may be required to conduct this measurement with the car background noise present. In such an event it may be required to measure the speech level according to Appendix II. Care should be taken not to use this method when SNR < −10 dB.

8.3.2.1.6.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class, for all input signal levels the measured level difference between input and output has to meet the requirements for the applicable performance classes defined in Table 8-14.

Table 8-14 – Limits for the LIN_{MS}

Performance Class	LIN_{MS}
1	< 0.5 dB
2	FFS
3	FFS
4	FFS

8.3.2.1.6.4 Design guidance and root-cause analysis

8.3.2.1.7 Microphone send speech quality

8.3.2.1.7.1 Parameter description

The microphone listening speech quality in send LQ_{MS} is measured from test point (MRP) to test point (S1a) or test point (S2).

The test is intended to determine any impairment of the listening speech quality introduced by the microphone.

8.3.2.1.7.2 Test

The tests method to be used is [ITU-T P.863].

NOTE 1 – [ITU-T P.863] does not provide a wideband mode. Wideband systems are evaluated on a superwideband scale. Therefore, the MOS requirements are given in MOS-LQOsw. See [ITU-T P.863] for more information.

NOTE 2 – [ITU-T P.863] provides a narrowband and a superwideband mode. In this document, for narrowband the narrowband mode of ITU-T P.863 is used leading to a maximum MOSn of approximately 4.2 evaluated in a narrowband context. In superwideband mode this would lead to a maximum score of an MOSsw of around 3.6. In wideband, the superwideband mode is used leading to a maximum score of approximately an MOSsw of 4.5.

1) The test signals used are the English test sequences as specified in [ITU-T P.501], Annex B (2 male speakers, 2 female speakers, two sentences each). The test signal level is the nominal signal level measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.

   The original speech signal is used as the reference signal for the determination of the speech quality.

2) The test arrangement is according to clause 8.3.2. For wideband systems MOS-LQOsw is determined. For narrowband systems MOS-LQOn is determined.

   The calculation is made using the signal recorded at test point (S2).

8.3.2.1.7.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality in send has to meet the requirements for the applicable performance classes given in the table below.

Table 8-15 – Limits for the microphone listening speech quality in send

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

NOTE – It is recognized that high MOS scores measured at the subsystem level may not necessarily lead to the best overall system performance. For some codecs or other signal processing in the overall system a degraded signal at the subsystem level may lead to better overall performance.

8.3.2.1.7.4 Design guidance and root-cause analysis

No impairment on listening speech quality should be introduced by the microphone. Any degradation of listening speech quality will deteriorate the overall listening speech quality and should be avoided. Therefore, the listening speech quality under quiet conditions is checked. Potential degradations of the microphone listening speech quality in send may be caused by noise (e.g., circuit noise), non-optimum frequency response characteristics, additional reverberation or impairments introduced by signal processing such as non linearities.

8.3.2.1.8 Microphone send speech quality with background noise

8.3.2.1.8.1 Parameter description

The microphone speech quality in send with background noise $LQBGN_{MS}$ is measured at test point (S1a) or test point (S2) with the reference point (Mic) for the unprocessed speech + noise signal.

The speech quality in background noise is tested based on [b-ETSI ES 202 396-3]. The test method described leads to three MOS-LQO quality numbers:

N-MOS-LQO: Transmission quality of the background noise

S-MOS-LQO: Transmission quality of the speech

G-MOS-LQO: Overall transmission quality

For this test, the speech level is adjusted at the MRP to take into account the Lombard effect. The level adjustment is calculated according to clause 8.1.

8.3.2.1.8.2 Test

1) The test arrangement is given in clause 8.3.2.
2) According to the specification of the manufacturer/test lab, the background noise is played back. The background noise should be applied for at least 5 s in cases where speech enhancement is used to adapt noise reduction algorithms in advance of the test.
3) The near-end speech signal consists of eight sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be found in [ITU-T P.501]. The preferred language for wideband is French since the objective method was validated with the French language. The preferred language in narrowband is English. The test signal level is the signal level according to clause 8.1 measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%. Three signals are required for the tests:
   – The clean speech signal is used as the undisturbed reference (see [b-ETSI ES 202 396-3]).
   – The speech plus undisturbed background noise signal is recorded at the microphone position (Mic) using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
   – The send signal is recorded at test point (S2).

N-MOS-LQO, S-MOS-LQO and G-MOS-LQO are calculated as described in [b-ETSI ES 202 396-3].

8.3.2.1.8.3 Performance level classification based on values of this parameter

Background noises for driving speed ≤ 80 km/h

To claim compliance with a certain performance class the listening speech quality with background noise in send has to meet the requirements for the applicable performance classes given in the tables below.

Table 8-16 – Limits for the listening speech quality with background noise in send: N-MOS

Performance Class	N-MOS-LQOn	N-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Table 8-17 – Limits for the listening speech quality with background noise in send: S-MOS

Performance Class	S-MOS-LQOn	S-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Table 8-18 – Limits for the listening speech quality with background noise in send: G-MOS

Performance Class	G-MOS-LQOn	G-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Background noises for driving speed ≤ 130 km/h

To claim compliance with a certain performance class the listening speech quality with background noise in send has to meet the requirements for the applicable performance classes given in the table below.

Table 8-19 – Limits for the listening speech quality with background noise in send: N-MOS

Performance Class	N-MOS-LQOn	N-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

Table 8-20 – Limits for the listening speech quality with background noise in send: S-MOS

Performance Class	S-MOS-LQOn	S-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

Table 8-21 – Limits for the listening speech quality with background noise in send: G-MOS

Performance Class	G-MOS-LQOn	G-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

NOTE – It is recognized that high MOS scores measured at the subsystem level may not necessarily lead to the best overall system performance. For some codecs or other signal processing in the overall system a degraded signal at the subsystem level may lead to better overall performance.

8.3.2.1.8.4 Design guidance and root-cause analysis

The microphone send speech quality with background noise mainly depends on the SNR at the hands-free microphone. Other factors which may influence the speech quality in background noise are microphone placement (which may help to improve the SNR) and the microphone frequency response characteristics. Further impact can be expected by wind buffeting caused, e.g., by poorly designed outlets for air conditioning or heating systems.

A methodology for more detailed investigation of background noise modulations due to noise cancelling algorithms can be found in Appendix I.

8.3.3 Audio subsystem (receive path)

The audio subsystem and its related access points are shown in Figure 8-6. The audio subsystem may consist of one or more loudspeakers with associated amplifiers and an associated signal enhancement system. In comparison to other subsystems the audio subsystem may provide an output signal (reference channel) for the echo canceller which is part of the signal enhancement subsystem (see clause 8.3).

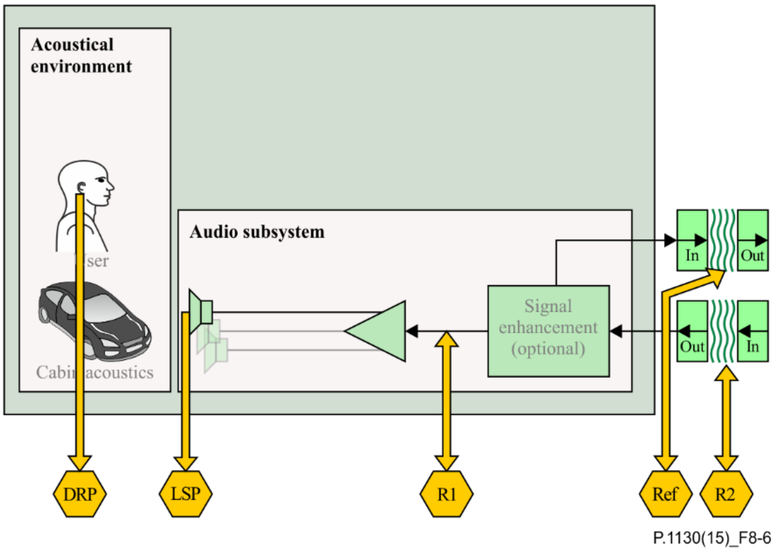


Figure 8-6 – Audio subsystem and test points

NOTE – All measurements and requirements are made for stand-still or low speed conditions. Any speed or noise-dependent equalization, compression or other functionality is not covered by this document because this functionality is considered to be very car specific and no general recommendation can be given.

8.3.3.1 Audio subsystem delay

8.3.3.1.1 Parameter description

The delay T_{AS} is measured from test point (R2) to test point (DRP). The delay T_{AS} should be minimized. The measurement is conducted at nominal volume control setting.

The system delay t_{System} depends on the transmission method used and the delay of the reference interface. The delay t_{System} must be known and deducted from the test result.

8.3.3.1.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (R2) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 8.3.3.

2) The delay is determined by cross-correlation analysis between the measured signal at the DRP and the original signal. The measurement is corrected by delays which are caused by the test equipment. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{AS} + T_{System}$$

8.3.3.1.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the delay T_{AS} defined in Table 8-22.

Table 8-22 – Limits for the audio subsystem delay for speech communication services

Performance Class	T_{AS}
1	< 7 ms
2	< 12 ms
3	< 30 ms
4	≥ 30 ms

8.3.3.1.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: higher delay low leads to higher impairment resulting even from low level echo components (see [ITU-T G. 114]). Therefore, any design providing as low a delay in the connection as possible is preferable. Furthermore, higher delays may impair the performance of the echo canceller in the signal enhancement subsystem.

8.3.3.2 Audio subsystem sensitivity frequency response

8.3.3.2.1 Parameter description

The frequency response FR_{AS} is measured from test point (R2) to test point (DRP). The measurement is conducted at nominal volume control setting.

The sensitivity response on the audio subsystem should be mostly flat in the entire frequency range.

8.3.3.2.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The test signal level is the nominal signal level, the level is averaged over the complete test signal.

NOTE – If the signal enhancement subsystem includes an equalization function for the receive path the equalization curve should be determined and applied to the test signal inserted. Such a pre-equalized signal is used for the measurements.

The measured power density spectrum at test point (R2) interface is used as the reference power density spectrum for determining the audio subsystem sensitivity.

2) The test arrangement is according to clause 8.3.3. The audio subsystem frequency response is determined in third octave intervals, as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz inclusive. In each third octave band, the level of the measured signal is referred to the level of the reference signal averaged over the complete test sequence length.

3) The sensitivity is determined in dBPa/V.

8.3.3.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response has to meet the requirements for the applicable performance classes for the unidirectional transport FR_{AS} defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 8-23 to 8-30 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 8-23 – Tolerance mask for the wideband audio subsystem sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	3	−6
200	3	−6
315	3	−3
6 300	3	−3
7 600	3	−6
8 000	3	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

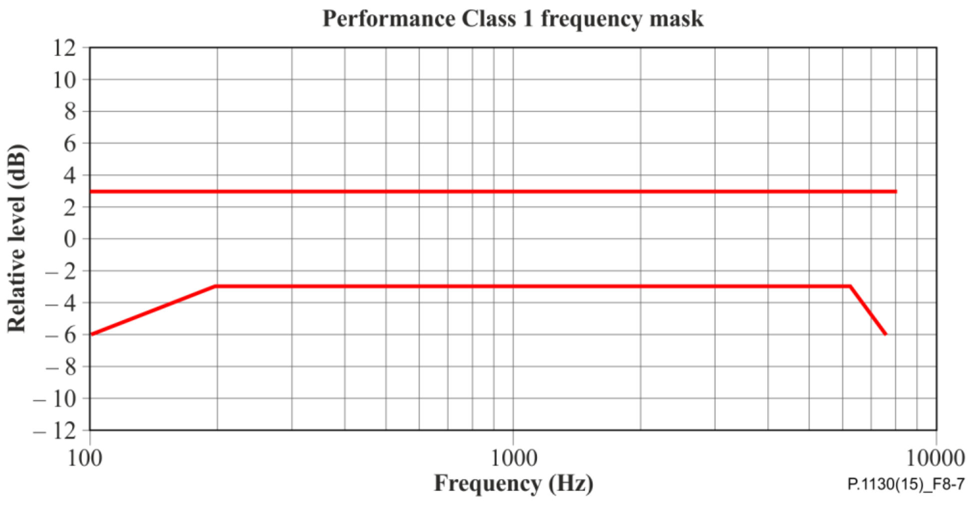


Figure 8-7 – Audio subsystem frequency response mask (Figure is informative)

Table 8-24 – Tolerance mask for the wideband audio subsystem sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	8	–
250	8	−9
315	8	−6
400	6	−6
5 000	6	−6
7 000	6	−9
8 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

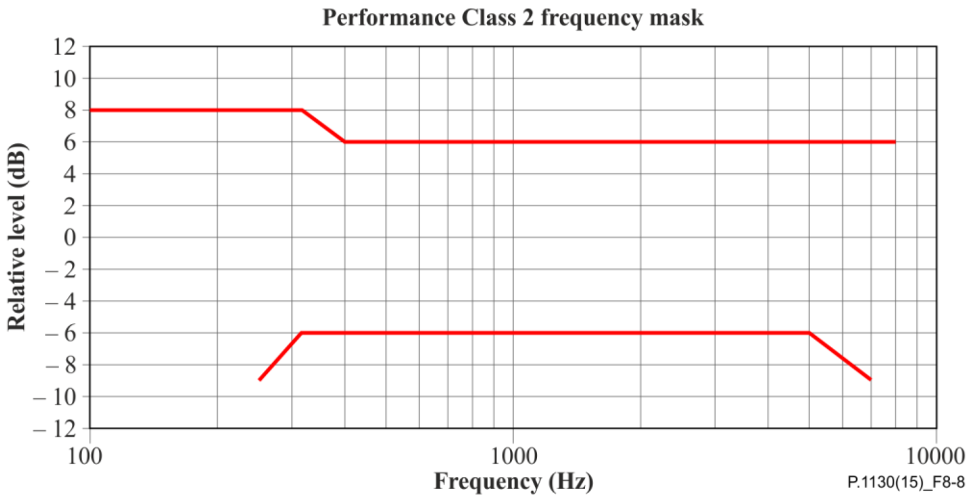


Figure 8-8 – Audio subsystem frequency response mask (Figure is informative)

Table 8-25 – Tolerance mask for the wideband audio subsystem sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	8	–
250	8	–
315	8	−9
400	6	−6
5 000	6	−9
7 000	6	−12
8 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

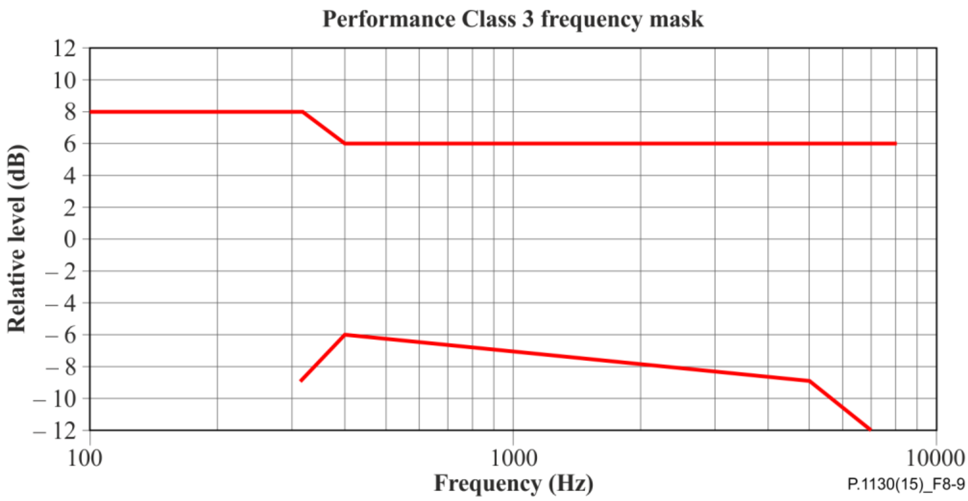


Figure 8-9 – Audio subsystem frequency response mask (Figure is informative)

Table 8-26 – Tolerance mask for the wideband audio subsystem sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	N/A	N/A
250	N/A	N/A
315	N/A	N/A
400	N/A	N/A
5 000	N/A	N/A
7 000	N/A	N/A
8 000	N/A	N/A
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

Table 8-27 – Tolerance mask for the audio subsystem sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	3	−6
250	3	(Note)
315	3	−3
400	3	−3
3 100	3	−3
3 400	3	−6
4 000	3	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

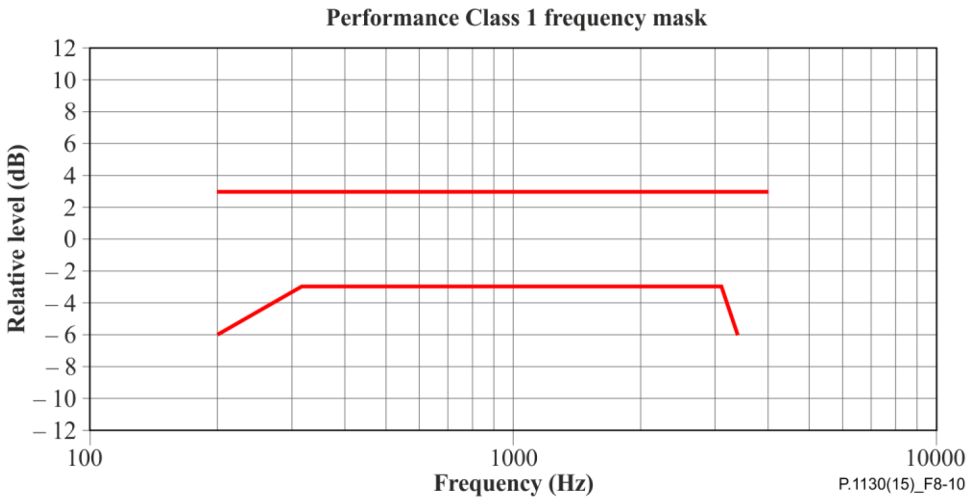


Figure 8-10 – Audio subsystem frequency response mask (Figure is informative)

Table 8-28 – Tolerance mask for the narrowband audio subsystem sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	6	–
250	6	–
315	6	−9
400	6	−6
3 100	6	−6
4 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

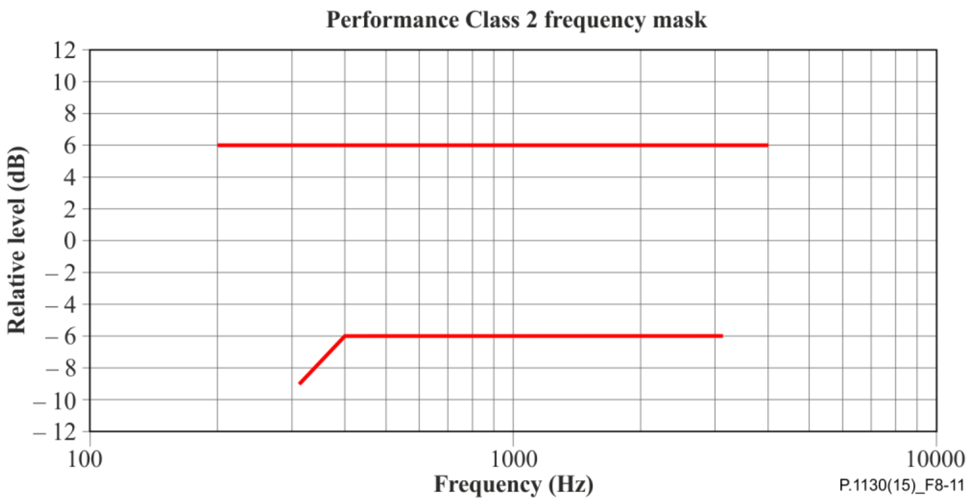


Figure 8-11 – Audio subsystem frequency response mask (Figure is informative)

Table 8-29 – Tolerance mask for the narrowband audio subsystem sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	6	–
250	6	–
315	6	−12
400	6	−8
3 100	6	−8
4 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

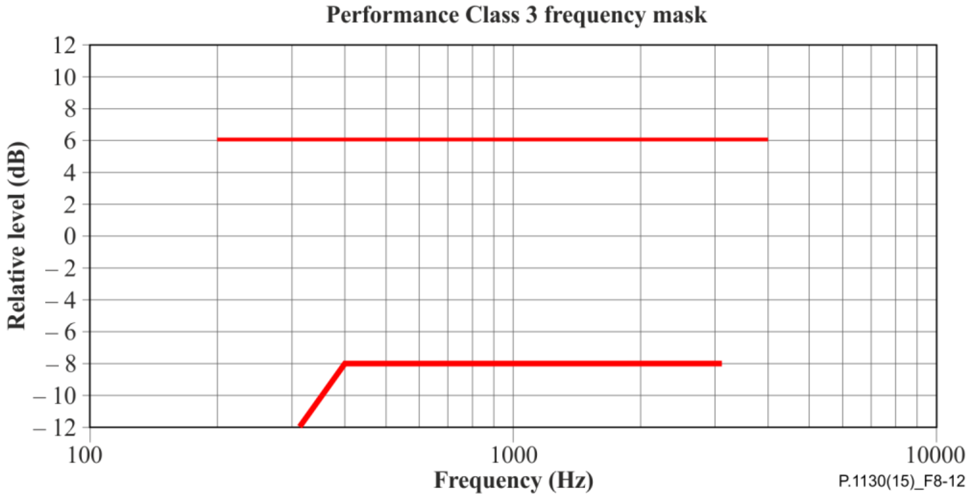


Figure 8-12 – Audio subsystem frequency response mask (Figure is informative)

Table 8-30 – Tolerance mask for the narrowband audio subsystem sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
200	N/A	N/A
315	N/A	N/A
1 000	N/A	N/A
1 300	N/A	N/A
1 600	N/A	N/A
2 000	N/A	N/A
3 100	N/A	N/A
4 000	N/A	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

8.3.3.2.4 Design guidance and root-cause analysis

Deviations from flat frequency response characteristics may cause degradation of the listening speech quality in the receive direction or may result in insufficient listening speech quality. A flat frequency response mostly contributes to good customer experience in the car.

8.3.3.3 Audio subsystem speech quality

8.3.3.3.1 Parameter description

The audio subsystem listening speech quality LQ_{AS} is measured from test point (R2) to test point (DRP). The measurement is conducted at nominal and at maximum volume control setting.

The test is intended to determine any impairment of the listening speech quality introduced by the unidirectional transport.

8.3.3.3.2 Test

The tests method to be used is [ITU-T P.863].

NOTE – [ITU-T P.863] does not provide a wideband mode. Wideband systems are evaluated on a superwideband scale. Therefore, the MOS requirements are given in MOS-LQOsw. See [ITU-T P.863] for more information.

1) The test signals used are the English test sequences specified in [ITU-T P.501] (2 male speakers, 2 female speakers, two sentences each). The test signal level is the nominal signal level, measured at test point (R2), the test signal level is measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.

   The original speech signal is used as the reference signal for the determination of the speech quality.

NOTE – If the signal enhancement subsystem includes an equalization function for the receive path the equalization curve should be determined and applied to the test signal inserted. Such a pre-equalized signal is used for the measurements.

2) The test arrangement is according to clause 8.3.3. For wideband systems MOS-LQOsw is determined. For narrowband systems MOS-LQOn is determined.

   The calculation is made using the signal recorded at test point (DRP).

8.3.3.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality has to meet the requirements for the applicable performance classes given in the tables below.

Table 8-31 – Limits for the audio subsystem listening speech quality at nominal volume control setting

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

Table 8-32 – Limits for the audio subsystem listening speech quality at maximum volume control setting

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

8.3.3.3.4 Design guidance and root-cause analysis

No additional impairment on listening speech quality should be introduced by the audio subsystem. Any degradation of listening speech quality will deteriorate the overall listening speech quality for the user in the car and should be avoided. Therefore, the listening speech quality under quiet conditions is checked.

8.3.3.4 Audio subsystem idle channel noise

8.3.3.4.1 Parameter description

The audio subsystem transport idle channel noise in N_{AS} is measured from test point (R2) to test point (DRP). The measurement is conducted at nominal and at **maximum** volume control setting.

8.3.3.4.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with the nominal average signal level [ITU-T P.50] is applied at test point (R2). The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) For the noise measurement, no test signal is used. However, all sources which potentially contribute to noise at the output of the unidirectional transport as GSM noise, electrical noise introduced by the car must be considered. In order to ensure a reliable activation, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of 4 composite source signals according to [ITU-T P.501]. The level of the activation signal is –16 dBm0.

3) The test arrangement is according to clause 8.3.3.

   The idle channel noise is measured at the output in the frequency range between 20 Hz and 20 kHz. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers shall be taken into account; the time window must be shifted accordingly. The length of the time window is 1 second which is the averaging time for the idle channel noise. The test laboratory has to ensure the correct activation of the audio subsystem during the measurement. If the audio subsystem is deactivated during measurement, the measurement window has to be cut to the duration when the Audio subsystem remains activated.

   The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hann window is used.

   If it is known that the audio subsystem stays activated without any activation signal, no activation signal is required. In this case, a simple noise measurement is conducted.

   The idle channel noise is determined by A-weighting and referring to the reference speech signal level as determined with the British-English test sentence.

Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3-rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^ (–1/6)*f* to 2^ (+1/6)*f*). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum up to 20 kHz.

8.3.3.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR as defined in Table 8-33 and Table 8-34.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 8-33 – Limits for the idle channel noise re. reference speech signal level at nominal volume control setting

Performance Class	SINR
1	> 70 dB
2	> 60 dB
3	> 40 dB
4	≤ 40 dB

Table 8-34 – Limits for the idle channel noise re. reference speech signal level at maximum volume control setting

Performance Class	SINR
1	> 70 dB
2	> 60 dB
3	> 40 dB
4	≤ 40 dB

8.3.3.4.4 Design guidance and root-cause analysis

Check shielding, connectors and cabling with respect to potentially induced interfering signals. Check the quality of the amplifiers used to insert signals and the preamplifiers used to receive signals. Check power amplifiers. Verify that preamplifiers do not demodulate RF into NF; if up-sampling is used check artefacts due to up-sampling.

8.3.3.5 Audio subsystem output level at maximum level setting

8.3.3.5.1 Parameter description

The audio subsystem maximum output level OL_{ASmax} is measured from test point (R2) to test point (DRP). The measurement is conducted at maximum volume control setting.

The audio subsystem maximum output level is determined with nominal input signal level at R2 and maximum volume control setting.

8.3.3.5.2 Test

1) The test set-up is according to clause 8.3.3.
2) The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with the nominal average signal level applied at test point (R2). The test signal level is the average level of the complete test signal.
3) The maximum spectrum generated by the car audio system with maximum volume is recorded at the DRP.
4) The A-weighted sound pressure level is determined at the DRP in the frequency range from 100 Hz to 7 kHz.

8.3.3.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the audio subsystem maximum output level has to meet the requirements for the applicable performance classes for the audio subsystem maximum output level as defined in Table 8-35.

Table 8-35 – Limits for the audio subsystem maximum output level

Performance Class	OL_{ASmax}
1	> 84 dB(A)
2	>78 dB(A)
3	>72 dB(A)
4	<72 dB(A)

8.3.3.5.4 Design guidance and root-cause analysis

Insufficient output level will lead to poor speech intelligibility for the driver in the car. This may be especially problematic with high background noise, e.g., at high driving speeds. In general luxury cars or cars with good, low noise acoustical design are less affected by this problem compared to cars with less acoustical treatment. The audio system design should ensure good speech intelligibility/comprehension with low listening effort at typical driving conditions.

8.3.3.6 Audio subsystem overload point

8.3.3.6.1 Parameter description

The audio subsystem overload point OVL_{AS} is measured from test point (R2) to test point (LSP). The measurement is conducted at level setting of 15 dB above nominal volume and maximum control setting.

The audio subsystem overload point is determined by the highest signal which is produced by the audio subsystem in the car without distorting.

8.3.3.6.2 Test

1) The test set-up is according to clause 8.3.3.
2) The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with the nominal average signal level applied at test point (R2). The test signal level is the average level of the complete test signal.
3) The maximum spectrum generated by the car audio system with maximum volume is recorded with a reference microphone at the LSP.

4) The coherence between the signal measured at the LSP and the signal inserted at R2 is determined as an average of the coherence function in the frequency range from 100 Hz to 7 kHz for wideband and 300 Hz to 3.5 kHz for a narrowband system. The coherence is determined by a spectral analysis which has a bandwidth of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent).

NOTE – In case of multiple loudspeakers it is recommended to position the reference microphone close to a midrange speaker (for narrowband systems) or find a position as close as possible between the midrange and the high-frequency speaker (for wideband systems).

8.3.3.6.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the audio subsystem overload point has to meet the requirements for the applicable performance classes for the audio subsystem overload point defined in Table 8-36.

Table 8-36 – Limits for the audio subsystem overload point above 0.95 coherence

Performance Class	OVL re Lev_{max}
1	0 dB
2	−3 dB
3	−6 dB
4	−9 dB

NOTE – This table is used for both volume control settings independently.

8.3.3.6.4 Design guidance and root-cause analysis

Distortion in the audio subsystem will mainly cause two problems: audible distortion to the driver in the car and nonlinearities in the echopath. While the user may tolerate some distortion, the impact on the echo canceller may be more severe. Any non-linearity in the echo path may impair the convergence performance of the echo canceller and may lead to intolerable echo or switching caused by the NLP in the echo canceller. In such cases either the maximum output volume should be limited or a less distorting audio subsystem is required.

8.3.3.7 Audio subsystem delay between REF and LSP

8.3.3.7.1 Parameter description

The delay $T_{REF\text{-}DRP}$ is measured from test point (REF) to test point (LSP). The delay $T_{REF\text{-}LSP}$ should be minimized. The measurement is conducted at a nominal volume control setting.

8.3.3.7.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (R2) at a level according to clause 8.1.5.

   The reference signal is the signal measured at REF.

   The test set-up is according to clause 8.3.3.

2) The delay is determined by cross-correlation analysis between the measured signal at the LSP and the signal measured at REF. The measurement is corrected by delays which are caused

by the test equipment. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{REF\text{-}LSP} + T_{System}$$

8.3.3.7.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the delay $T_{REF\text{-}LSP}$ defined in Table 8-37.

Table 8-37 – Limits for the audio subsystem delay between REF and LSP

Performance Class	$T_{REF\text{-}LSP}$
1	< ±3 ms
2	< ±6 ms
3	< ±10 ms
4	≥ ±10 ms

8.3.3.7.4 Design guidance and root-cause analysis

Low delay is for a proper functionality of the echo canceller in the signal enhancement layer. Any delay between LSP and REF may impair the performance of the echo canceller and therefore should be avoided or minimized.

8.3.3.8 Audio subsystem linearity between reference output and LSP

8.3.3.8.1 Parameter description

The audio subsystem linearity $LIN_{REF\text{-}LSP}$ is measured from test point (REF) to test point (LSP).

The microphone linearity is determined by the measurement of the A-weighted output level between at LSP in the level range from −40 dB to +6 dB re. to the nominal input signal level and referring the measured level to the signal level measured at REF.

The measurement is conducted at nominal and at maximum volume control setting.

8.3.3.8.2 Test

1) The test arrangement is according to clause 8.3.3.

2) The test signal is inserted at (R2). The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The correct activation of the measurement object during the measurement has to be ensured by the test lab. The test signal level shall be −40 dB, −30 dB, −20 dB, −10 dB, 0 dB and +6 dB re. to the nominal signal level. The test signal level is the average level of the complete test signal.

3) The volume control is set to nominal.

4) The signals are recorded simultaneously at test points (LSP and REF).

5) The A-weighted output level is determined at test point S2 and referred to the input signal level.

6) The measurement is repeated for all input signal levels.

7) The measurement is repeated for maximum volume control setting.

8.3.3.8.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the measured level difference between REF and LSP has to meet for all input signal levels the requirements for the applicable performance classes defined in Table 8-38.

Table 8-38 – Limits for the $LIN_{REF\text{-}LSP}$

Performance Class	$LIN_{REF\text{-}LSP}$
1	< 0.2 dB
2	FFS
3	FFS
4	FFS

8.3.3.8.4 Design guidance and root-cause analysis

8.3.3.9 Coherence between reference output and LSP

8.3.3.9.1 Parameter description

The REF – LSP coherence is measured from test point (REF) to test point (LSP).

The $COH_{REF\text{-}LSP}$ is determined between (REF) and (LSP) output in the level range from −40 dB to +6 dB re. to the nominal input signal level.

The measurement is conducted at nominal and at maximum volume control setting.

8.3.3.9.2 Test

1) The test arrangement is according to clause 8.3.3.
2) The test signal is inserted at (R2). The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The correct activation of the measurement object during the measurement has to be ensured by the test lab. The test signal level shall be −40 dB, −30 dB, −20 dB, −10 dB, 0 dB and +6 dB re. to the nominal signal level. The test signal level is the average level of the complete test signal.
3) The volume control is set to nominal.
4) The signals are recorded simultaneously at test points (LSP and REF). The coherence between the LSP and REF is determined. The coherence is determined as an average of the coherence function in the frequency range from 100 Hz to 7 kHz.
5) The measurement is repeated for all input signal levels.
6) The measurement is repeated for maximum volume control setting.

8.3.3.9.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the $COH_{REF\text{-}LSP}$ has to meet the requirements for the applicable performance classes defined in Table 8-39.

Table 8-39 – Limits for the $COH_{REF-LSP}$

Performance Class	$COH_{REF-LSP}$
1	0.95
2	FFS
3	FFS
4	FFS

8.3.3.9.4 Design guidance and root-cause analysis

8.4 Signal enhancement subsystem

The signal enhancement subsystem attempts to improve signals captured and played at the acoustic interface. Network-side problems can also be reduced or eliminated. Figure 7-4 shows how this subsystem fits into the overall telecommunications system. This subsystem could be physically located in either the terminal device or a network device. However, the terminal device is usually a better location for technical reasons. Common signal processing functions provided by this subsystem include:

– Send direction
  - acoustic echo cancellation
  - noise reduction
  - equalization
  - automatic level control
  - fixed gain
  - limiter

– Receive direction
  - bandwidth extension
  - equalization
  - automatic level control
  - automatic level enhancement
  - fixed gain
  - limiter

The next subclause describes the test set-up for measuring the performance of the signal enhancement subsystem. Subsequent subclauses provide information related to each of the measurement parameters. This information includes a parameter description, test method, requirements, design guidance and root-cause analysis.

8.4.1 Test set-up

The test set-up for measuring performance of the signal enhancement subsystem consists of:

1) speech, noise test signals.

2) A set-up that enables playback and recording of these test signals by either:

   1 file-based approach – passing pre-recorded test signals to and from the signal enhancement layer;

2 interactive convolution-based approach – simulating the interaction of the signal enhancement subsystem and other subsystems (e.g., acoustic interface) using pre-recorded impulse responses including also time-variant effects of other subsystems;

3 vehicle-based approach – using a real vehicle with appropriate access points.

NOTE 1 – When using the file-based approach care must be taken to ensure that the delay between R3 and S2 is representative of the real system under test.

Speech and noise test signals at the input of the signal enhancement subsystem are processed to capture the effects of other subsystems involved in the end-to-end connection (e.g., vehicle microphones, speech coders, etc.).

In general the subsystem testing can be made either file-based, interactive convolution-based or, vehicle-based.

The signal enhancement subsystem shall be tested with the set of signals acquired or simulated for the individual car under test. In case of signal enhancement subsystem comparison, e.g., for the purpose of system qualification for different cars, a predefined set of test signals representative of the cars under consideration can be defined and used.

NOTE 2 – The current version of this Recommendation does not yet include reference test signals for testing the signal enhancement subsystem. However, it is still planned to include such reference test signals in a future version of this Recommendation.

Test signals are passed to test points S2 and R3 which represent the inputs to the signal enhancement layer. The processed output from the signal enhancement layer is retrieved from test points S3 and Ref and analysed.

Additionally, for the interactive convolution-based approach the output at REF is also convolved with an impulse response simulating the acoustic echo path in the vehicle. The simulated acoustic echo is then mixed with the send test signal before being presented to test point S2. This allows any changes of the test signal in the receive direction (e.g., gain change due to the signal enhancement layer) to be dynamically reflected in the input test signal at S2 which is what happens in the real world.

If not stated otherwise, all signal processing functions including equalization are active.

Where test signals are inserted depends on the approach used for testing:

– File-based mode – all test signals are applied to the interfaces as shown in Figure 8-13.
– Interactive convolution-based mode – All test signals are generated (and inserted) based on convolution and applied to the interfaces as shown in Figure 8-13.

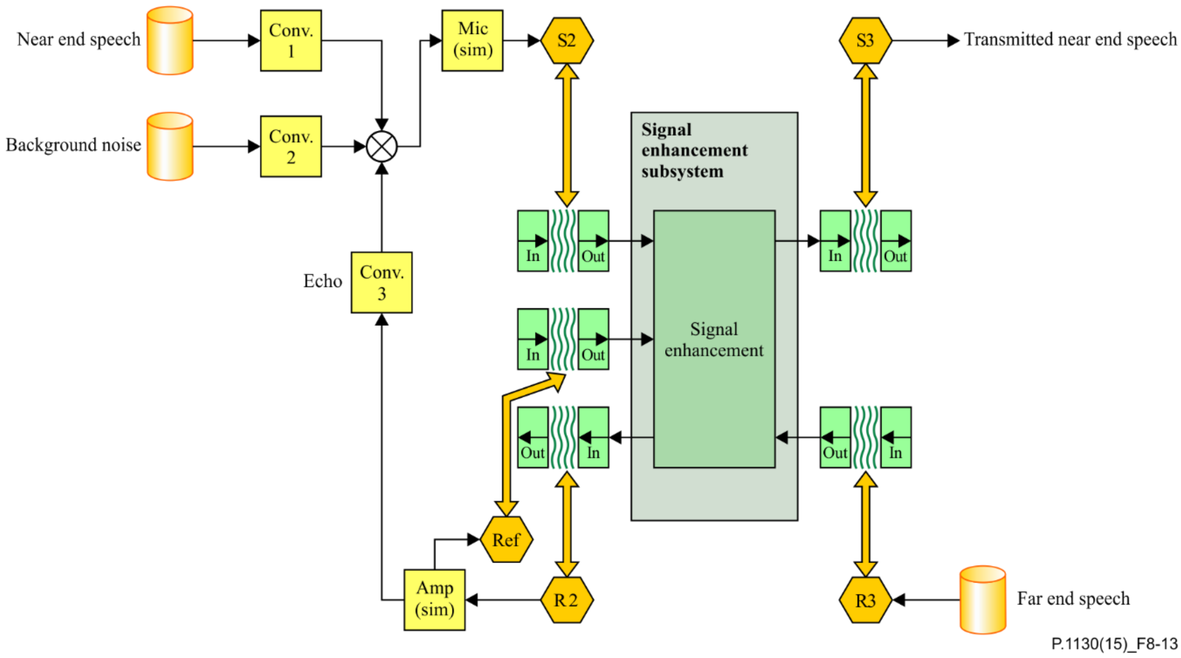


Figure 8-13 – Signal enhancement subsystem with simulation paths

– Vehicle-based approach – All test signals are inserted at interfaces MRP, BGN and R3 (see Figures 8-13, 7-4)

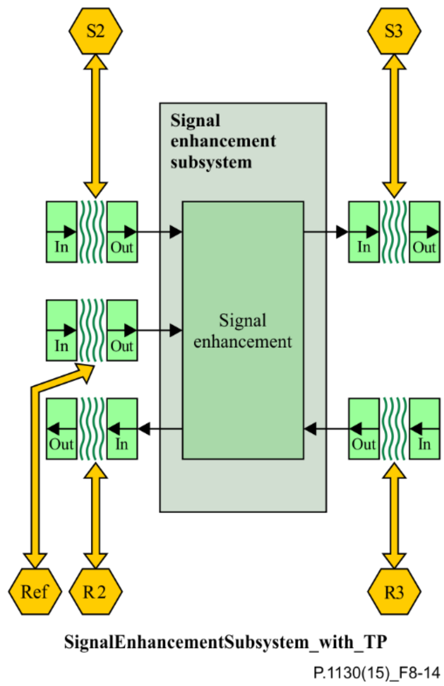


Figure 8-14 – Signal enhancement subsystem and test points

NOTE 1 – Some signal Enhancement subsystems might not use test point Ref.

NOTE 2 – Ideally, the access to the signal enhancement subsystem is made using the S3/R3 interface. However, if these interfaces are not available the S4/R4 or S5/R5 interface can be used applying the same requirements.

8.4.2 Round-trip delay of signal enhancement

The round-trip delay of signal enhancement T_{SERTD} consists of the delay in send direction the T_{SES} plus the delay in receive direction T_{SER}.

8.4.2.1 Performance level classification based on values of this parameter (T_{SERTD})

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the signal enhancement delay T_{SERTD} defined in Table 8-40.

Table 8-40 – Limits for the round-trip delay of signal enhancement for speech communication services

Performance Class	T_{SERTD}
1	≤ 25 ms
2	< 60 ms
3	< 100 ms
4	≥ 100 ms

8.4.2.2 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay leads to higher impairment resulting even from low level echo components (see [ITU-T G. 131]). Therefore, any design providing as low a delay in the connection as possible is preferable.

8.4.3 Signal enhancement delay in send direction (T_{SES})

8.4.3.1 Parameter description

The delay in send direction T_{SES} is measured from test point (S2) to test point (S3). The delay T_{SES} should be minimized.

The system delay t_{System} depends on the delay of the reference interface used. The delay t_{System} must be known and deducted from the test result.

8.4.3.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (S2) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 8.4.1.

2) The delay is determined by cross-correlation analysis between the measured signal at the electrical access point (S3) and the original signal. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{SES} + T_{System}$$

8.4.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the signal enhancement delay T_{SES} defined in Table 8-41.

Table 8-41 – Limits for the signal enhancement delay in send direction

Performance Class	T_{SES}
1	≤ 20 ms
2	< 50 ms
3	< 70 ms
4	≥ 70 ms

8.4.3.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay leads to higher impairment resulting even from low level echo components (see [ITU-T G.131]). Therefore, any design providing as low a delay in the connection as possible is preferable.

8.4.4 Signal enhancement delay in receive direction (T_{SER})

8.4.4.1 Parameter description

The signal enhancement delay in receive direction T_{SER} is measured from test point (R3) to test point (R2).

The system delay t_{System} depends on the delay of the reference interface used. The delay t_{System} must be known and deducted from the test result.

8.4.4.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is applied at test point (R3) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

2) The test arrangement is according to clause 8.4.1.
3) The delay is determined by cross-correlation analysis between the measured signal at test point (R2) and the original signal. The measurement is corrected by delays which are caused by the test equipment.
4) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{SER} + T_{System}$$

8.4.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the receive direction has to meet the requirements for the applicable performance classes for the signal enhancement delay in receive T_{SER} defined in Table 8-42.

Table 8-42 – Limits for the signal enhancement delay in receive direction

Performance Class	T_{SER}
1	< 5 ms
2	< 10 ms
3	< 30 ms
4	≥ 30 ms

8.4.4.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay leads to higher impairment resulting even from low level echo components (see [ITU-T G.131). Therefore, any design providing as low a delay in the connection as possible is preferable.

8.4.5 Send signal level (L_{SES})

8.4.5.1 Parameter description

The send signal level L_{SES} is measured at test point (S3).

The send signal level is system specific and describes the required output level needed to properly access the network transport subsystem.

8.4.5.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501], recorded at the output of the microphone subsystem S2. The test signal level is the nominal signal level, the level is averaged over the complete test signal.
2) For the calculation, the averaged level at test point (S2) is used. In wideband the send level is determined from 100 Hz-8 kHz; in narrowband from 100 Hz to 4 kHz measured as active speech level according to [ITU-T P.56].
3) The level is measured and referenced to the nominal signal level required by the network transport subsystem.

8.4.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the L_{SES} in send has to meet the requirements for the applicable performance classes for the signal enhancement L_{SES} defined in Table 8-43.

Table 8-43 – Limits for the signal enhancement L_{SES} in send

Performance Class	L_{SES}
1	0 ± 0.5 dB
2	0 ± 3 dB
3	0 ± 6 dB
4	> 0 ± 6 dB

NOTE – Measured values may need to be compensated for amplification or attenuation of gain stages occurring after the signal enhancement subsystem.

8.4.5.4 Design guidance and root-cause analysis

Any deviation of the JLR_{SES} from 0 dB may result either in insufficient or too high speech levels and thus may deteriorate the performance of the interconnected devices and should be avoided.

8.4.6 Signal enhancement automatic gain control in send

It is recognized that a single level measurement in send does not completely cover all the requirements that a signal enhancement subsystem should fulfil. Amongst others automatic level control might be used. Other measurement procedures are needed to characterize the different type of AGC mechanisms used to compensate for varying microphone signal levels due to soft or loud talker, different passenger positions etc. This is for further study.

8.4.7 Receive signal level L_{SER}

8.4.7.1 Parameter description

The receive signal level L_{SER} is measured at test point (R2).

The receive signal level is system specific and describes the required output level needed to properly access the audio subsystem.

8.4.7.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501], recorded at the output of the network transport subsystem R3. The test signal is the nominal signal level.
2) The test arrangement is according to clause 8.3.1. For the calculation, the averaged level at test point (R2) is used. In wideband the receive level is determined from 100 Hz-8 kHz; in narrowband from 100 Hz to 4 kHz measured as active speech level according to [ITU-T P.56].
3) The level is measured and referenced to the nominal signal level required by the audio subsystem, measured as active speech level according to [ITU-T P.56].

8.4.7.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the receive signal level at nominal setting of the volume control has to meet the requirements for the applicable performance classes for the signal enhancement L_{SER} defined in Table 8-44.

Table 8-44 – Limits for the signal enhancement receive signal level referenced to the nominal signal level required by the audio subsystem

Performance Class	L_{SER}
1	0 ± 0.5 dB
2	0 ± 3 dB
3	0 ± 6 dB
4	> ± 6 dB

8.4.7.4 Design guidance and root-cause analysis

Any deviation of the L_{SER} from 0 dB may result either in insufficient or too high speech levels and thus may deteriorate the performance of the interconnected devices and should be avoided. As a result, overload or audible noise may occur in the audio subsystem.

8.4.8 Signal enhancement automatic gain control in receive

It is recognized that a single level measurement in receive does not completely cover all the requirements that a signal enhancement subsystem should fulfil. Amongst others automatic level control might be used. These may be controlled by the incoming signal level or by the interior noise level in the car. Other measurement procedures are needed to characterize the different type of AGC mechanisms. This is for further study.

8.4.9 Signal enhancement send sensitivity frequency response (FR_{SES})

8.4.9.1 Parameter description

The send frequency response FR_{SES} is measured from test point (S2) to test point (S3). This parameter measures the sensitivity frequency response of the signal enhancement subsystem in send.

8.4.9.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501], recorded at the output of the microphone subsystem S2. The test signal level is the nominal signal level. The level is averaged over the complete test signal.

   The measured power density spectrum at the MRP interface is used as the reference power density spectrum for determining the signal enhancement subsystem send sensitivity frequency response.

2) The test arrangement is according to clause 8.4.1. In wideband the send sensitivity frequency response is determined in third octave intervals, as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz, inclusive. In narrowband it is determined for frequencies from 200 Hz to 4 kHz. In each third octave band, the level of the measured signal is referred to the level of the reference signal averaged over the complete test sequence length.

3) The sensitivity is determined in dBV/Pa.

8.4.9.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response in send has to meet the requirements for the applicable performance classes for the signal enhancement subsystem FR_{SES} as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 8-45 to 8-52 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 8-45 – Tolerance mask for the wideband signal enhancement subsystem send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	3	−10
125	3	−5
200	3	−3
1 000	3	−3
6 300	5	−3
7 600	5	−6
8 000	5	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

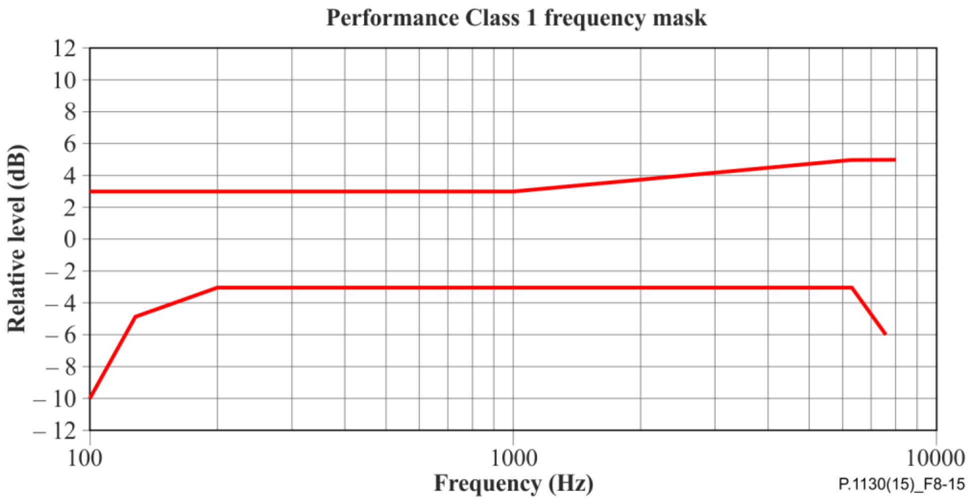


Figure 8-15 – Signal enhancement frequency response mask (Figure is informative)

Table 8-46 – Tolerance mask for the wideband signal enhancement subsystem send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	4	–
125	4	−10
200	4	−4
1 000	4	−4
5 000	(Note)	−4
6 300	9	−7
8 000	9	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

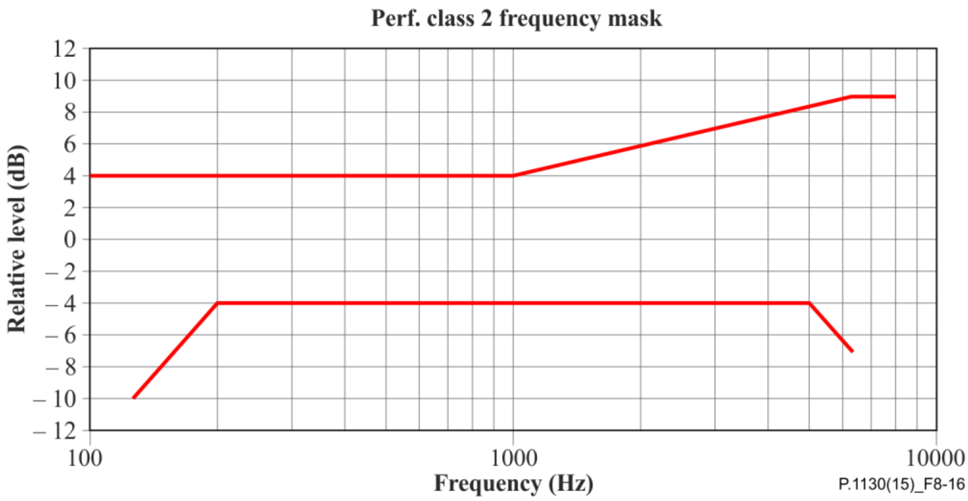


Figure 8-16 – Signal enhancement frequency response mask (Figure is informative)

Table 8-47 – Tolerance mask for the wideband signal enhancement subsystem send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	6	–
125	6	−15
200	6	−6
1 000	6	−6
5 000	(Note)	−6
6 300	9	−8
8 000	–	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

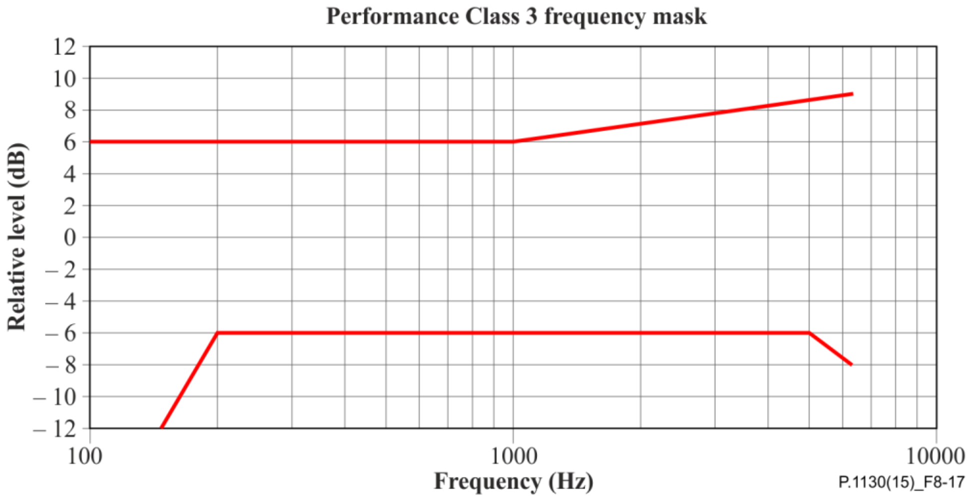


Figure 8-17 – Signal enhancement frequency response mask (Figure is informative)

Table 8-48 – Tolerance mask for the wideband signal enhancement subsystem send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	N/A	N/A
125	N/A	N/A
200	N/A	N/A
1 000	N/A	N/A
5 000	N/A	N/A
6 300	N/A	N/A
8 000	N/A	N/A
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

Table 8-49 – Tolerance mask for the narrowband signal enhancement subsystem send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	3	−10
315	3	−3
1 000	3	−3
1 300	5	−3
1 600	6	−3
2 000	6	−3
3 100	6	−3
3 400	(Note)	−6
4 000	3	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

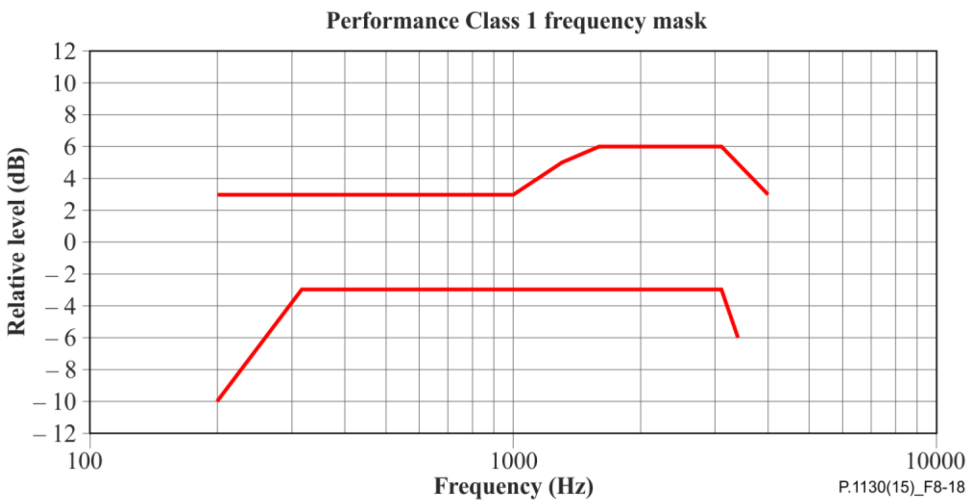


Figure 8-18 – Signal enhancement frequency response mask (Figure is informative)

Table 8-50 – Tolerance mask for the narrowband signal enhancement subsystem send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	4	–
315	4	−10
1 000	4	−4
1 300	6	−4
1 600	7	−4
2 000	8	−4
3 100	8	−4
4 000	4	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

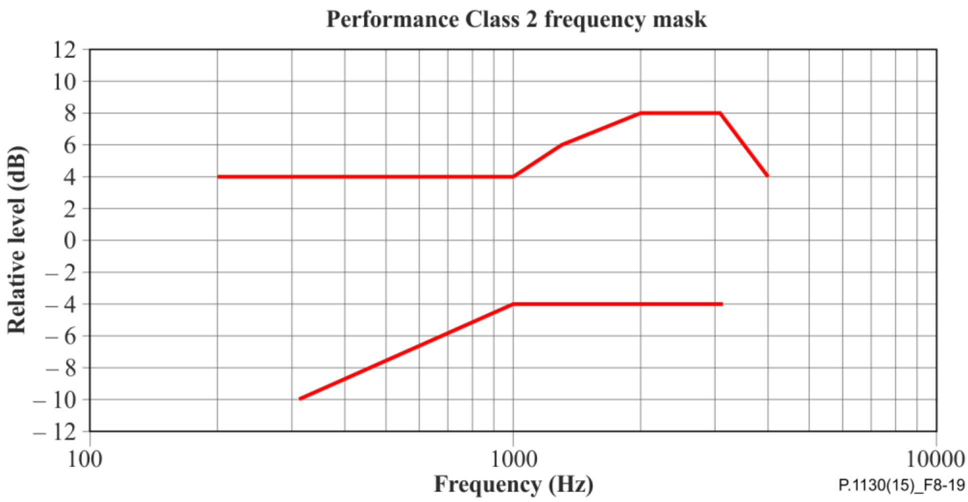


Figure 8-19 – Signal enhancement frequency response mask (Figure is informative)

Table 8-51 – Tolerance mask for the narrowband signal enhancement subsystem send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	6	–
315	6	−10
1 000	6	−6
1 300	6	−6
1 600	(Note)	−6
2 000	9	−6
3 100	9	−10
4 000	–	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

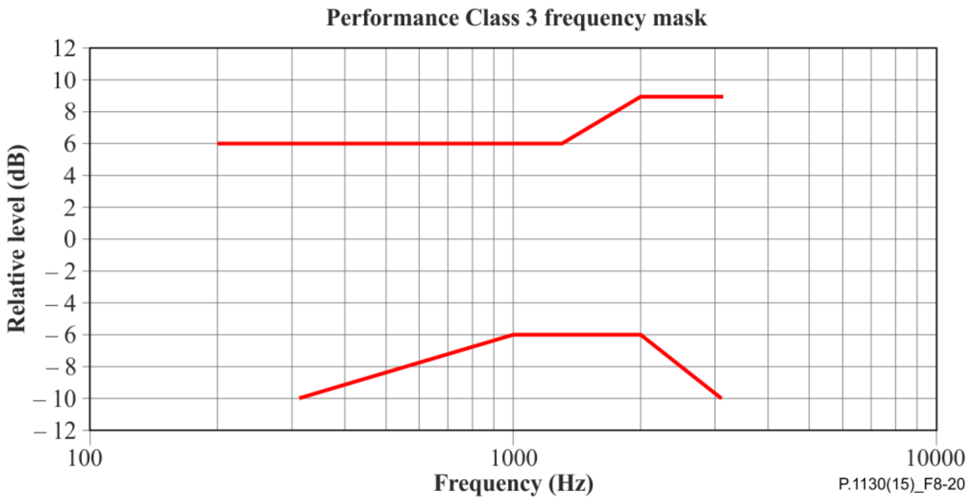


Figure 8-20 – Signal enhancement frequency response mask (Figure is informative)

Table 8-52 – Tolerance mask for the narrowband signal enhancement subsystem send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
200	N/A	N/A
315	N/A	N/A
1 000	N/A	N/A
1 300	N/A	N/A
1 600	N/A	N/A
2 000	N/A	N/A
3 100	N/A	N/A
4 000	N/A	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

8.4.9.4 Design guidance and root-cause analysis

Deviations from the desired frequency response characteristic may cause degradation of the listening speech quality in the send direction or may result in insufficient listening speech quality in the presence of background noise which affects the listing speech quality perceived by the far-end subscriber.

8.4.10 Signal enhancement receive sensitivity frequency response (FR_{SER})

8.4.10.1 Parameter description

The receive frequency response FR_{SER} is measured from test point (R3) to test point (R2). This parameter measures the sensitivity frequency response of the signal enhancement subsystem in receive.

8.4.10.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501], recorded at the output of the network transport

subsystem R3. The test signal level is the nominal signal level. The level is averaged over the complete test signal.

The measured power density spectrum at test point (R3) interface is used as the reference power density spectrum for determining the signal enhancement subsystem receive sensitivity frequency response.

2) The test arrangement is according to clause 8.4.1. In wideband the send sensitivity frequency response is determined in third octave intervals, as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz, inclusive. In narrowband it is determined for frequencies from 200 Hz to 4 kHz. In each third octave band, the level of the measured signal is referred to the level of the reference signal averaged over the complete test sequence length.

3) The sensitivity is determined in dBV/V.

4) The measured result is corrected by the measured sensitivity frequency response characteristics of the audio subsystem in order to take into account any equalization potentially performed in the audio subsystem as well as the loudspeaker response characteristics. Alternatively, a simulated R2 to DRP simulation could be used. In such a case the simulated output signal would be referred to the measured power density spectrum at test point (R3) interface.

8.4.10.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response in receive has to meet the requirements for the applicable performance classes for the signal enhancement subsystem FR_{SER} as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 8-53 to 8-60 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 8-53 – Tolerance mask for the wideband receive enhancement subsystem receive sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	3	−6
200	3	−3
6 300	3	−3
7 600	3	−6
8 000	3	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

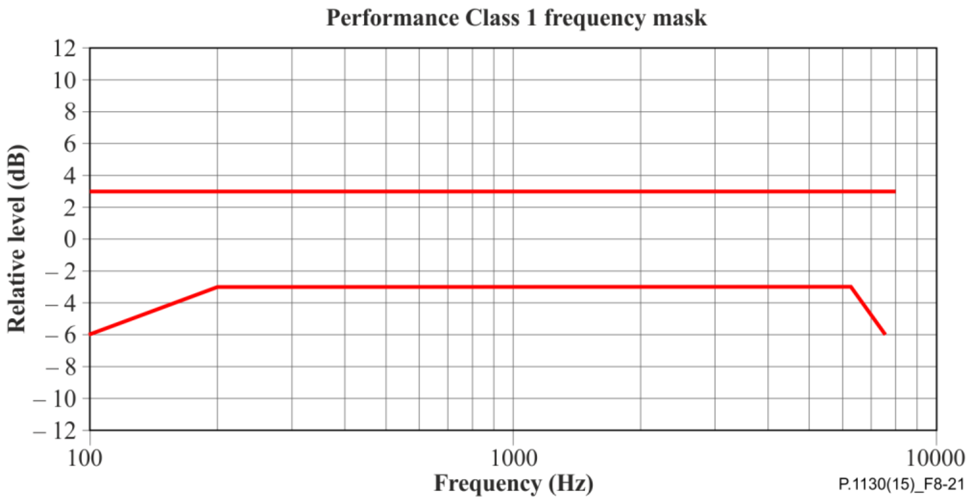


Figure 8-21 – Signal enhancement frequency response mask (Figure is informative)

Table 8-54 – Tolerance mask for the wideband signal enhancement subsystem receive sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
125	8	–
200	8	−12
250	8	−9
315	7	−6
400	6	−6
5 000	6	−6
6 300	6	−9
8 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

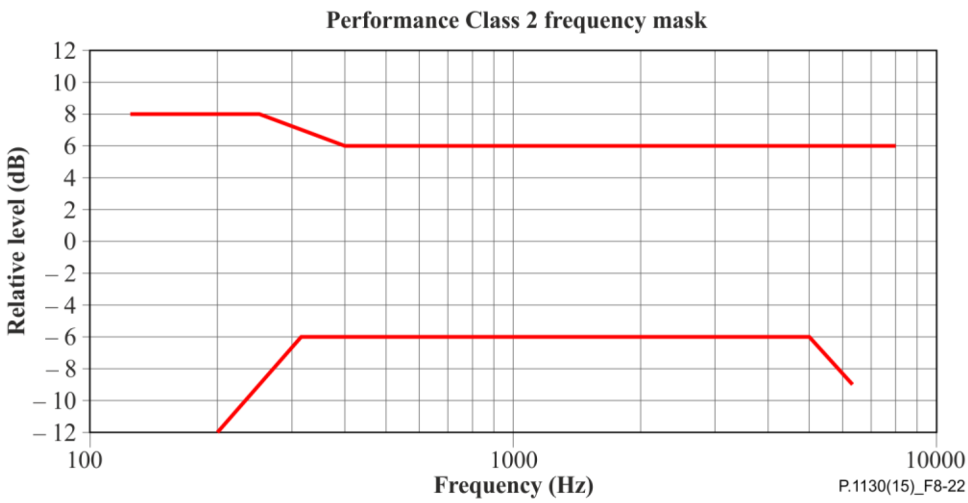


Figure 8-22 – Signal enhancement frequency response mask (Figure is informative)

Table 8-55 – Tolerance mask for the wideband signal enhancement subsystem receive sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	8	–
200	8	–
250	8	–
315	8	−9 dB
400	6	−6 dB
5 000	6	−9 dB
7 000	6	−12 dB
8 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

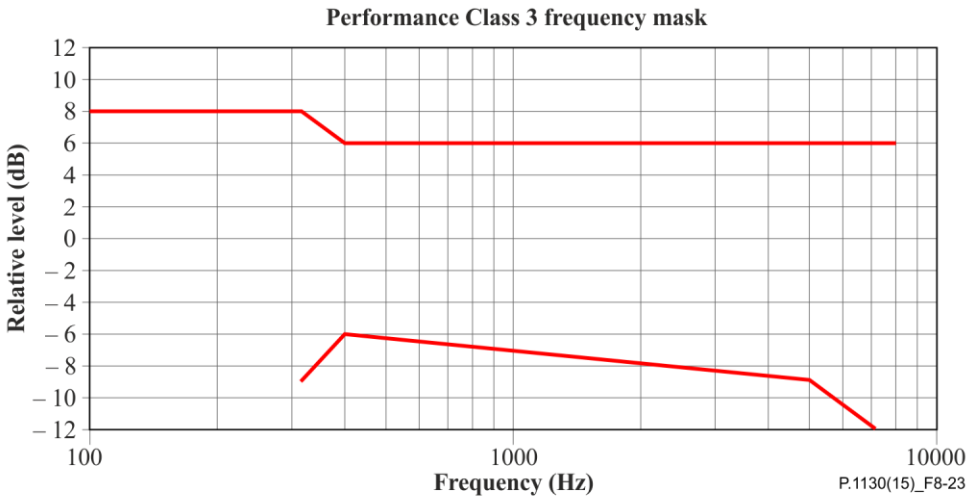


Figure 8-23 – Signal enhancement frequency response mask (Figure is informative)

Table 8-56 – Tolerance mask for the wideband signal enhancement subsystem receive sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
100	N/A	N/A
125	N/A	N/A
200	N/A	N/A
1 000	N/A	N/A
5 000	N/A	N/A
6 300	N/A	N/A
8 000	N/A	N/A
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

Table 8-57 – Tolerance mask for the narrowband signal enhancement subsystem receive sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	3	−6
250	3	(Note)
315	3	−3
400	3	−3
3 100	3	−3
3 400	3	−6
4 000	3	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

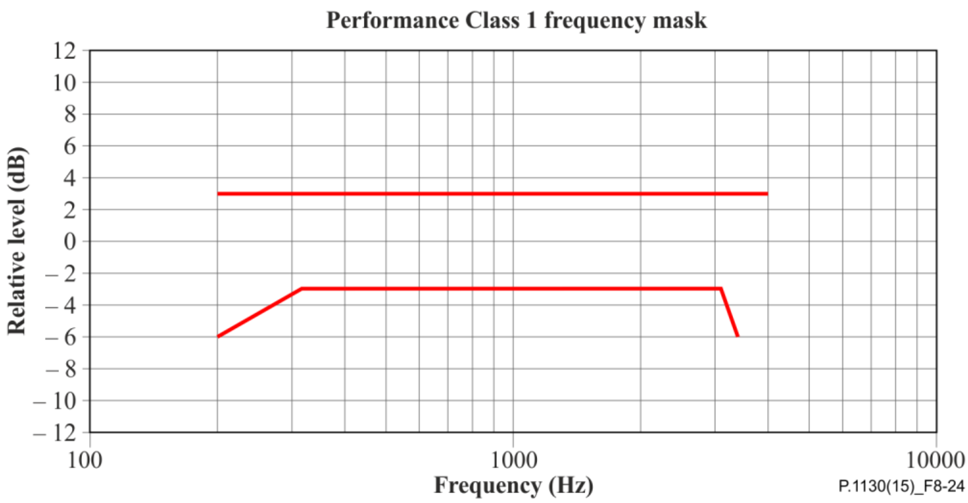


Figure 8-24 – Signal enhancement frequency response mask (Figure is informative)

Table 8-58 – Tolerance mask for the narrowband signal enhancement subsystem receive sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	6	–
250	6	–
315	6	−9
400	6	−6
3 100	6	−6
4 000	6	–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

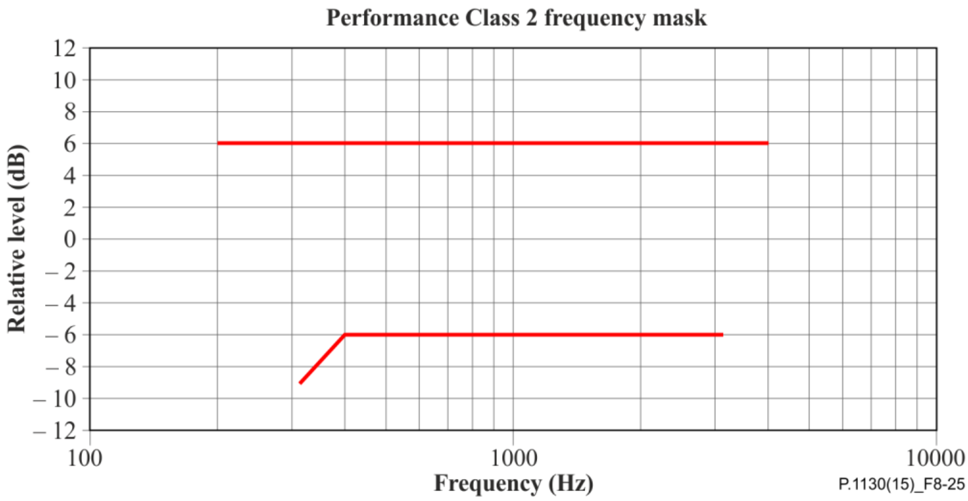


Figure 8-25 – Signal enhancement frequency response mask (Figure is informative)

Table 8-59 – Tolerance mask for the narrowband signal enhancement subsystem receive sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	6	–
250	6	–
315	6	−12
400	6	−8
3 100	6	−8
4 000		–
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

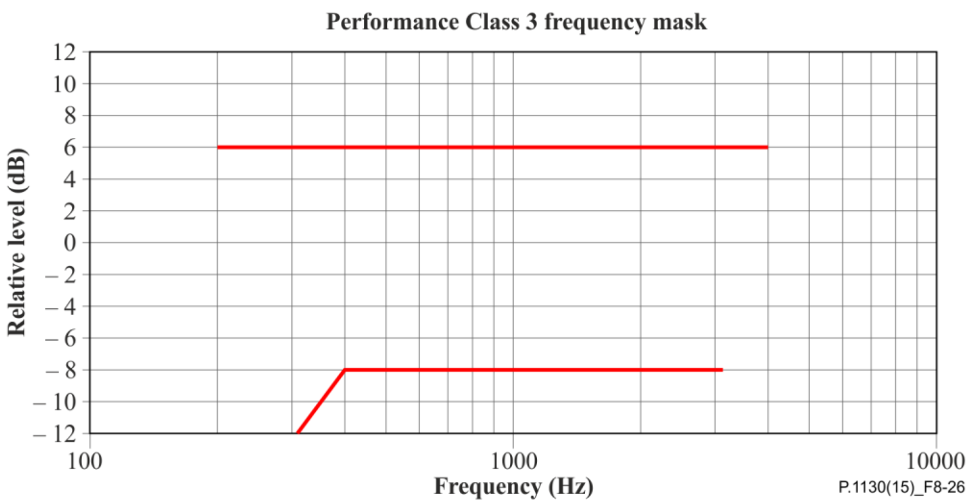


Figure 8-26 – Signal enhancement frequency response mask (Figure is informative)

Table 8-60 – Tolerance mask for the narrowband signal enhancement subsystem receive sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit (dB)	Lower limit (dB)
200	N/A	N/A
315	N/A	N/A
1 000	N/A	N/A
1 300	N/A	N/A
1 600	N/A	N/A
2 000	N/A	N/A
3 100	N/A	N/A
4 000	N/A	
NOTE – The limits for intermediate frequencies lie on a straight line drawn between the given values on a linear (dB) – logarithmic (Hz) scale. All sensitivity values are expressed in dB on an arbitrary scale.		

8.4.10.4 Design guidance and root-cause analysis

Deviations from the desired frequency response characteristic may cause degradation of the listening speech quality for the subscriber in the car.

Careful design should take into account:

– not too strong equalization in order to avoid overload at certain frequencies;

– no counteraction equalization between audio subsystem and signal enhancement.

8.4.11 Signal Enhancement Send Speech Quality (MOS-LQO$_{SES}$)

8.4.11.1 Parameter description

This parameter predicts speech quality under quiet conditions in the send direction after the signal enhancement subsystem. The signal enhancement speech quality in send direction MOS-LQO$_{SES}$ is measured from test point (S2) to test point (S3). In order to determine the send speech quality including the speech coder applied in the system under consideration the signal acquired at test point (S3) is processed using the appropriate speech coder.

NOTE – This measurement method does not apply to systems including frequency shifting technologies.

8.4.11.2 Test

1) The test signal used is composed of eight sentences of two male and two female British English speakers according to Annex B.3.3 of [ITU-T P.501] and concatenated as described in [ITU-T P.863]. The test signal is recorded at the output of the microphone subsystem S2. The test signal level at the MRP is the nominal signal level. The test signal level is measured as "active speech level" according to [ITU-T P.56].

   The original speech signal is used as the reference signal for the determination of the speech quality.

   In wideband the superwideband mode of the model is used. In narrowband the superwideband mode is applied as well however, applying a narrowband reference signal as described in [ITU-T P.863].

2) The test arrangement is according to clause 8.4.1.

   The calculation is made using the signal recorded at S3.

3) In a second step the signal acquired at S3 is processed by the speech coder used in the hands-free implementation under consideration.

4) The one-way speech quality is determined as MOS-LQON (narrowband) and MOS-LQOSW (wideband).

8.4.11.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality in send has to meet the requirements for the applicable performance classes given in the table below.

Table 8-61 – Limits for the signal enhancement listening speech quality in send, measured at test point S3

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

Table 8-62 – Limits for the signal enhancement listening speech quality in send, measured after speech coding using the signal acquired at test point S3

Performance Class	MOS-LQOn	MOS-LQOsw
1	See Note	See Note
2	FFS	FFS
3	FFS	FFS
4	FFS	FFS

NOTE – In order to achieve Performance Class 1 the MOS-LQO value should be equivalent to one measured for the codec and bitrate under consideration when using a clean speech signal without processing by the signal enhancement.

8.4.11.4 Design guidance and root-cause analysis

For further study.

8.4.12 Signal enhancement receive speech quality ($MOS\text{-}LQO_{SER}$)

8.4.12.1 Parameter description

This parameter predicts speech quality in the receive direction after the signal enhancement subsystem. The signal enhancement speech quality in receive direction $MOS\text{-}LQO_{SER}$ is measured from test point (R3) to test point (R2)

NOTE 1 – This measurement method does not apply to systems including frequency shifting technologies such as artificial bandwidth extension.

NOTE 2 – The measurement may be repeated including the filtering with the impulse response measured from R2 to DRP in order to investigate the impact of the acoustical transmission on the measured MOS-LQO.

8.4.12.2 Test

1) The test signal used is composed of eight sentences of two male and two female British English speakers according to Annex B.3.3 of [ITU-T P.501] and concatenated as described

in [ITU-T P.863], and recorded at the output of the network transport subsystem R3. The test signal level is the nominal signal level. The test signal level is measured as "active speech level" according to [ITU-T P.56].

The original speech signal is used as the reference signal for the determination of the speech quality.

In wideband the superwideband mode of the model is used. In narrowband the superwideband mode is applied as well however, applying a narrowband reference signal as described in [ITU-T P.863].

2) The test arrangement is according to clause 8.4.1.

   The calculation is made using the signal recorded at R2. The measured signal is corrected by the measured sensitivity frequency response characteristics of the audio subsystem in order to take into account any equalization potentially performed in the audio subsystem as well as the loudspeaker response characteristics.

3) The one-way speech quality is determined as MOS-LQON (narrowband) and MOS-LQOSW (wideband).

8.4.12.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality in receive has to meet the requirements for the applicable performance classes given in the table below.

Table 8-63 – Limits for the signal enhancement listening speech quality in receive

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

8.4.12.4 Design guidance and root-cause analysis

For further study.

8.4.13 Signal enhancement send intelligibility (SIE)

8.4.13.1 Parameter description

This parameter is for further study. It is intended to predict speech intelligibility in the send direction after the signal enhancement subsystem. A measure of intelligibility is desirable because some automotive speech services (e.g., eCall) should be optimized for speech intelligibility/comprehension.

8.4.13.2 Test method

For further study.

NOTE – The test method should consider both approaches that use signal-based measures (e.g., STI, AI, SII, etc.) and ASR-based approaches.

8.4.13.3 Performance level classification based on values of this parameter

For further study.

8.4.13.4 Design guidance and root-cause analysis

For further study.

8.4.14 Signal enhancement receive intelligibility (RIE)

8.4.14.1 Parameter description

This parameter predicts speech intelligibility in the receive direction after the signal enhancement subsystem. A measure of intelligibility is desirable because some automotive speech services (e.g., eCall) should be optimized for speech intelligibility/comprehension.

8.4.14.2 Test

For further study.

NOTE – The test method should consider both approaches that use signal-based measures (e.g., STI, AI, SII, etc.) and ASR-based approaches.

8.4.14.3 Performance level classification based on values of this parameter

For further study.

8.4.14.4 Design guidance and root-cause analysis

For further study.

8.4.15 Signal enhancement send speech to idle channel noise ratio ($SINR_{SES}$)

8.4.15.1 Parameter description

This parameter measures the idle channel noise, when no significant acoustic noise is present in the vehicle, in the send direction after the signal enhancement subsystem. The signal enhancement idle channel noise in send N_{SES} is measured at test point (S3).

8.4.15.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] – recorded at the output of the microphone subsystem S2 is applied at test point (S2). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

   The measured power density spectrum at test point (S3) is used as the reference power density spectrum for determining the signal enhancement subsystem SINR. The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) The idle channel noise is determined from the pauses in the speech signal.

3) The test arrangement is described in clause 8.4.1. If implemented any equalizer function is activated.

4) The idle channel noise is measured at test point S3. In narrowband the frequency range between 100 Hz and 4 kHz is used. In wideband the frequency range from 100 Hz to 8 kHz is used. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers or reverberation influence shall be taken into account; the time window must be shifted accordingly. The length for the time window is the time between 2 consecutive speech samples which is 0.5 seconds, and which is also the averaging time for the idle channel noise. The test laboratory has to ensure that the terminal is activated during the measurement. If the subsystem is deactivated during the measurement, the measurement window has to be cut to the duration while the subsystem remains activated.

   The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hanning window is used.

5) In narrowband the idle channel noise is determined by A-weighting. Spectral peaks are measured in the frequency domain. The average noise spectrum used for determining the spectral peak should be calculated as the arithmetic mean of the noise spectrum values when stated in dBm0(P), dBm0(A) respectively in wideband.

6) The weighted noise signal level is referenced to the speech signal level measured at S2.

8.4.15.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 8-64.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 8-64 – Limits for the idle channel noise re. reference speech signal level with equalizer function activated

Performance Class	SINR
1	> 40 dB
2	> 30 dB
3	> 20 dB
4	≤ 20 dB

8.4.15.4 Design guidance and root-cause analysis

For further study.

8.4.16 Signal enhancement receive speech to idle channel noise ratio ($SINR_{SER}$)

8.4.16.1 Parameter description

This parameter measures the idle channel noise, when there is no significant noise coming from the network or far-end terminal, in the receive direction after the signal enhancement subsystem. The signal enhancement idle channel noise in receive N_{SER} is measured at test point (R2).

8.4.16.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] – recorded at the output of the network transport subsystem R3 is applied at test point (R3). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

   The measured power density spectrum at test point (R2) is used as the reference power density spectrum for determining the signal enhancement subsystem SINR. The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) The idle channel noise is determined from the pauses in the speech signal.

3) The test arrangement is described in clause 8.4.1. If implemented any equalizer function is activated.

4) The idle channel noise is measured at test point (R2). In narrowband the frequency range between 100 Hz and 4 kHz is used. In wideband the frequency range from 100 Hz to 8 kHz is used. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers or reverberance influence shall be taken into account; the time window must be shifted accordingly. The length for the time window

is the time between 2 consecutive speech samples which is 0.5 seconds, and which is also the averaging time for the idle channel noise. The test laboratory has to ensure that the terminal is activated during the measurement. If the subsystem is deactivated during the measurement, the measurement window has to be cut to the duration while the subsystem remains activated.

The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hanning window is used.

5) The idle channel noise is A-weighted. Spectral peaks are measured in the frequency domain. The average noise spectrum used for determining the spectral peak should be calculated as the arithmetic mean of the noise spectrum values when stated in dBm0(A).

6) The weighted noise signal level is referenced to the speech signal level measured at S2.

8.4.16.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 8-65.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 8-65 – Limits for the idle channel noise re. reference speech signal level with equalizer function activated

Performance Class	SINR
1	> 40 dB
2	> 30 dB
3	> 20 dB
4	≤ 20 dB

8.4.16.4 Design guidance and root-cause analysis

For further study.

8.4.17 Discrimination against out-of-band signals after send enhancement ($DOOB_{SES}$)

8.4.17.1 Parameter description

This parameter measures the degree to which speech and noise signal frequencies, which are outside the transmission channel passband at the input to the signal enhancement subsystem, cause noise in the send direction at the output of the signal enhancement subsystem. The discrimination against out-of-band signals in send $DOOB_{SES}$ is measured at test point (S3).

8.4.17.2 Test

1) The test arrangement is in accordance with clause 8.4.1.

2) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] – recorded at the output of the microphone subsystem S2 is applied at test point (S2). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

   The measured power density spectrum at test point (S3) is used as the reference power density spectrum for determining the signal enhancement subsystem DOOB.

3) In order to ensure a reliable activation of the subsystem, an activation signal is generated before the actual measurement starts. The activation signal consisting of the female speaker of the short conditioning sequence described in clause 7.3.7 of [ITU-T P.501] recorded at the output of the microphone subsystem S2 is applied at test point (S2). The activation level shall be the nominal signal level. The level of the activation signal is averaged over the complete activation sequence signal.

4) Directly after the activation signal, the actual test signal is inserted. The test signal recorded at the output of the microphone subsystem S2 is applied at test point (S2). The test signal is inserted exactly after the activation signal. The duration of the test signal amounts to 200 ms.

5) The test signal is a white Gaussian noise, band-limited from 4.6 kHz to 8 kHz in narrowband and 8 kHz to 10 kHz in wideband with nominal signal level. The level of the test signal is averaged over the complete test signal sequence.

6) For the analysis, a rectangular window is used which is adapted to the test signal duration (200 ms). Any "ringing" of filters or receivers or reverberance influence shall be taken into account; the time window must be shifted accordingly. In narrowband the signal level is determined in the frequency range from 300 Hz to 3.4 kHz at test point (S2). The level of the reference signal, band-limited from 300 Hz to 3.4 kHz is determined at test point (S2) as well. In wideband for both signals 100 Hz to 8 kHz is used. The in-band reference signal level is determined using a speech level voltmeter according to [ITU-T P.56].

7) The ratio (in dB) between the reference level and the measured signal level is determined.

NOTE 1 – With low sensitivity in the send direction, the measured noise level may already exceed the required minimum out-of-band level.

NOTE 2 – This measurement method does not apply to systems including frequency shifting technologies.

8.4.17.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the discrimination against out-of-band signals after signal enhancement has to meet the requirements for the applicable performance classes for the $DOOB_{SES}$ defined in Table 8-66.

Table 8-66 – Limits for the discrimination against out-of-band signals after send enhancement

Performance Class	$DOOB_{SES}$
1	> 40 dB
2	> 35 dB
3	> 30 dB
4	≤ 30 dB

8.4.17.4 Design guidance and root-cause analysis

For further study.

8.4.18 Spurious out-of-band signal after receive enhancement ($SOOB_{SER}$)

8.4.18.1 Parameter description

This parameter measures the degree to which speech signal frequencies within the transmission channel passband cause spurious noise outside the passband at the output of the signal enhancement subsystem in the receive direction. The signal enhancement spurious out-of-band signal in receive $SOOB_{SER}$ is measured at test point (R2).

8.4.18.2 Test

1) The test arrangement is in accordance with clause 8.4.1.

2) The test signal is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] recorded at the output of the network transport subsystem R3 applied at test point (R3). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

   The measured in-band power density spectrum at test point (S3) is used as the reference power density spectrum for determining the signal enhancement subsystem DOOB.

3) In narrowband the level of the out-of-band signal is determined between 4.6 and 8 kHz. In wideband the level of the out-of-band signal is determined between 8.6 and 16 kHz. The in-band reference level is determined by measuring the in-band signal level between 300 Hz and 3.4 kHz in narrowband and 100 Hz to 8 kHz in wideband. The in-band reference signal level is determined using a speech level voltmeter according to [ITU-T P.56].

4) The ratio (in dB) between the reference level and the measured signal level is determined.

NOTE 1 – This measurement method does not apply to systems including frequency shifting techniques such as artificial bandwidth extension.

NOTE 2 – In case of fluctuating noise levels, multiple reference noise measurements should be made and the worst case result should be used as a reference.

NOTE 3 – With low sensitivity in the receive direction, the measured noise level may already exceed the required minimum out-of-band level.

8.4.18.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the spurious out-of-band signals after signal enhancement has to meet the requirements for the applicable performance classes for the $SOOB_{SER}$ defined in Table 8-67.

Table 8-67 – Limits for the discrimination against out-of-band signals after send enhancement

Performance Class	$SOOB_{SER}$
1	> 40 dB
2	> 35 dB
3	> 30 dB
4	≤ 30 dB

8.4.18.4 Design guidance and root-cause analysis

For further study.

8.4.19 Distortion in send after enhancement (D_{SES})

8.4.19.1 Parameter description

This parameter measures speech signal distortion in the send direction after the signal enhancement subsystem. The signal enhancement distortion in send D_{SES} is measured at test point (S3).

8.4.19.2 Test

1) The test arrangement is according to clause 8.4.1. The test signal is a sinusoidal signal with a frequency of 300 Hz, 500 Hz, 1 kHz and 2 kHz (for wideband only) is applied at test point

(S2). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

2) In order to ensure a reliable activation of the subsystem, an activation signal is generated before the actual measurement starts. The activation signal consisting of the female speaker of the short conditioning sequence described in clause 7.3.7 of [ITU-T P.501] is applied at test point (S2). The activation level shall be the nominal signal level. The level of the activation signal is averaged over the complete activation sequence signal.

3) The test signal is inserted immediately after the activation sequence. The test signal duration is 200 ms.

4) For the analysis, a Hann window is used which is adapted to the duration of the test signal (200 ms).

5) The harmonic distortion produced by the signal enhancement subsystem is measured at test point (S3).

8.4.19.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the discrimination distortion after signal enhancement has to meet the requirements for the applicable performance classes for the D_{SES} defined in Table 8-68.

Table 8-68 – Limits for the signal enhancement subsystem distortion

Performance Class	D_{SES}
1	< 0.1%
2	< 0.5%
3	< 1%
4	> 1%

8.4.19.4 Design guidance and root-cause analysis

For further study.

8.4.20 Distortion in receive after enhancement (D_{SER})

8.4.20.1 Parameter description

This parameter measures speech signal distortion in the receive direction after the signal enhancement subsystem. The signal enhancement distortion in receive D_{SER} is measured at test point (R2).

8.4.20.2 Test

1) The test arrangement is according to clause 8.4.1. The test signal is a sinusoidal signal with a frequency of 300 Hz, 500 Hz, 1 kHz and 2 kHz (for wideband only) is applied at test point (R3). The test signal level is the nominal signal level. The level is averaged over the complete test signal.

2) In order to ensure a reliable activation of the subsystem, an activation signal is generated before the actual measurement starts. The activation signal consisting of the female speaker of the short conditioning sequence described in clause 7.3.7 of [ITU-T P.501] is applied at test point (R3). The activation level shall be the nominal signal level. The level of the activation signal is averaged over the complete activation sequence signal.

3) The test signal is inserted immediately after the activation sequence. The test signal duration is 200 ms.

4) For the analysis, a Hann window is used which is adapted to the duration of the test signal (200 ms).

5) The harmonic distortion produced by the signal enhancement subsystem is measured at test point (R2).

8.4.20.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the discrimination distortion after signal enhancement has to meet the requirements for the applicable performance classes for the D_{SER} defined in Table 8-69.

Table 8-69 – Limits for the signal enhancement subsystem distortion

Performance Class	D_{SER}
1	< 0.1%
2	< 0.5%
3	< 1%
4	> 1%

8.4.20.4 Design guidance and root-cause analysis

For further study.

8.4.21 Signal enhancement weighted terminal coupling loss ($TCLw_{SE}$) and (TCL_{SE})

8.4.21.1 Parameter description

This parameter measures acoustic echo loss during receive single-talk. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3). In narrowband TCLw is measured in wideband the unweighted TCL is measured.

8.4.21.2 Test

1) The test arrangement is in accordance with clause 8.4.1.

2) The noise level measured at the electrical access point (idle channel noise) shall be less than −63 dBm0. The attenuation between the input to the signal enhancement subsystem (R3) to the output of the signal enhancement subsystem is measured using the compressed speech test signal as described in [ITU-T P.501], clause 7.3.3, Amendment 1. The test signal level is −10 dBm0.

3) The first 17.0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences). For the calculation, the average measured echo level at each frequency band is referred to the average level of the test signal measured in each frequency band.

4) In narrowband TCL_W is calculated according to clause B.4 of [ITU-T G.122], (trapezoidal rule). In wideband the differences between the averaged echo level and the averaged test signal level in a frequency range from 100 Hz-8000 Hz is calculated.

8.4.21.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the echo loss after signal enhancement has to meet the requirements for the applicable performance classes for the TCL_{SE} (wideband) and $TCLw_{SE}$ defined in Table 8-70.

Table 8-70 – Limits for the signal enhancement subsystem TCL

Performance Class	$TCLw_{SE\ (narrowband)}$	$TCL_{SE\ (wideband)}$
1	> 50 dB	> 50 dB
2	> 46 dB	> 46 dB
3	> 40 dB	> 40 dB
4	< 40 dB	< 40 dB

8.4.21.4 Design guidance and root-cause analysis

For further study.

8.4.22 Signal enhancement echo level versus time ($ELVT_{SE}$)

8.4.22.1 Parameter description

This parameter measures acoustic echo loss as a function of time during receive single-talk. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3).

8.4.22.2 Test

1) The test arrangement is in accordance with clause 8.4.1.
2) The test signal consists of a periodically repeated composite source signal according to [ITU-T P.501] with an average level of −5 dBm0 as well as an average level of −25 dBm0. The echo signal is analysed during a period of at least 2.8 s, which represents 8 periods of the CS signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the level analysis of the reference signal.
3) The measurement result is displayed as attenuation versus time. The exact synchronization between input and output signal has to be guaranteed.
4) In addition, the test is repeated with the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The average test signal level is −16 dBm0. The echo signal is analysed during the complete test signal. The integration time for the level analysis shall be 35 ms, the analysis is referred to the analysis of the same signal but using an integration time of 1 s.

NOTE – When testing using CSS, the analysis is conducted only during the active signal part, the pauses between the composite source signals are not analysed. The analysis time is reduced by the integration time of the level analysis (35 ms).

8.4.22.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the $ELVT_{SE}$ defined in Table 8-71.

Table 8-71 – Limits for the signal enhancement subsystem echo level vs. time

Performance Class	$ELVTw_{SE}$
1	≤ 3 dB
2	≤ 6 dB
3	≤ 9 dB
4	> 9 dB

8.4.22.4 Design guidance and root-cause analysis

For further study.

8.4.23 Signal enhancement spectral echo attenuation (SEA_{SE})

8.4.23.1 Parameter description

This set of parameters measures the frequency response of the acoustic echo path during receive single-talk. The echo path frequency response is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3).

8.4.23.2 Test

1) The test arrangement is in accordance with clause 8.4.1. The noise level measured at the electrical access point (idle channel noise) shall be less than −63 dBm0.
2) The test signal is the compressed speech signal as described in [ITU-T P.501], clause 7.3.3, Amendment 1. The test signal level is −10 dBm0.
3) The power density spectrum of the measured echo signal is referred to the power density spectrum of the original test signal. The analysis is conducted using FFT with 8 k points (48 kHz sampling rate or equivalent, Hann window).
4) The spectral echo attenuation is analysed in the frequency domain in dB.

8.4.23.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the LVT_{SE} defined in Table 8-72 for narrowband systems, and in Table 8-73 for wideband systems.

Table 8-72 – Limits for the signal enhancement subsystem spectral echo attenuation in narrowband

Frequency (Hz)	Performance Class 1	Performance Class 2	Performance Class 3	Performance Class 4
100	≤ –20 dB	FFS	FFS	FFS
200	≤ –30 dB	FFS	FFS	FFS
300	≤ –38 dB	FFS	FFS	FFS
800	≤ –34 dB	FFS	FFS	FFS
1 500	≤ –33 dB	FFS	FFS	FFS
2 600	≤ –24 dB	FFS	FFS	FFS
4 000	≤ –24 dB	FFS	FFS	FFS

Table 8-72 – Limits for the signal enhancement subsystem spectral echo attenuation in narrowband

Frequency (Hz)	Performance Class 1	Performance Class 2	Performance Class 3	Performance Class 4
NOTE 1 – All sensitivity values are expressed in dB on an arbitrary scale. NOTE 2 – The limit at intermediate frequencies lies on a straight line drawn between the given values on a log (frequency) – linear (dB) scale.				

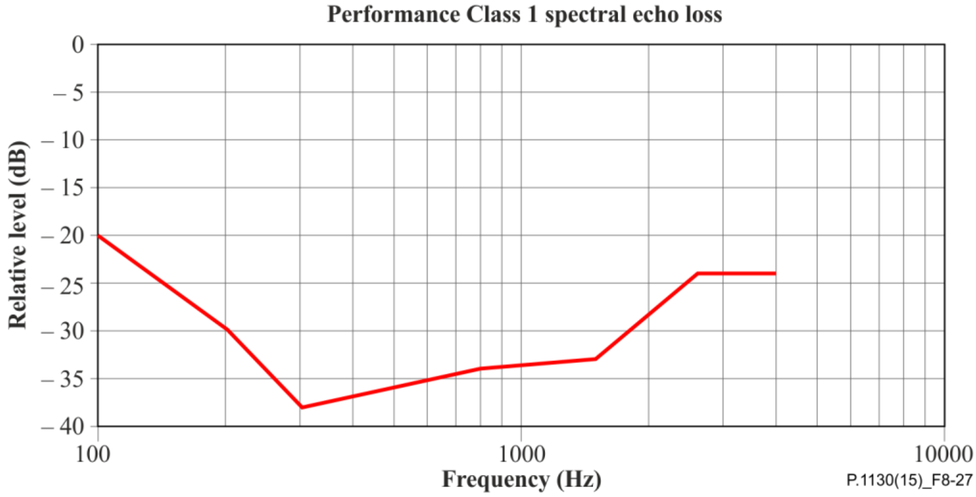


Figure 8-27 – Spectral echo loss mask Performance Class 1 (Figure is informative)

Table 8-73 – Limits for the signal enhancement subsystem spectral echo attenuation in wideband

Frequency (Hz)	Performance Class 1	Performance Class 2	Performance Class 3	Performance Class 4
100	≤ –41 dB	FFS	FFS	FFS
1 300	≤ –41 dB	FFS	FFS	FFS
3 450	≤ –46 dB	FFS	FFS	FFS
5 200	≤ –46 dB	FFS	FFS	FFS
7 500	≤ –37 dB	FFS	FFS	FFS
8 000	≤ –37 dB	FFS	FFS	FFS
NOTE 1 – All sensitivity values are expressed in dB on an arbitrary scale. NOTE 2 – The limit at intermediate frequencies lies on a straight line drawn between the given values on a log (frequency) – linear (dB) scale.				

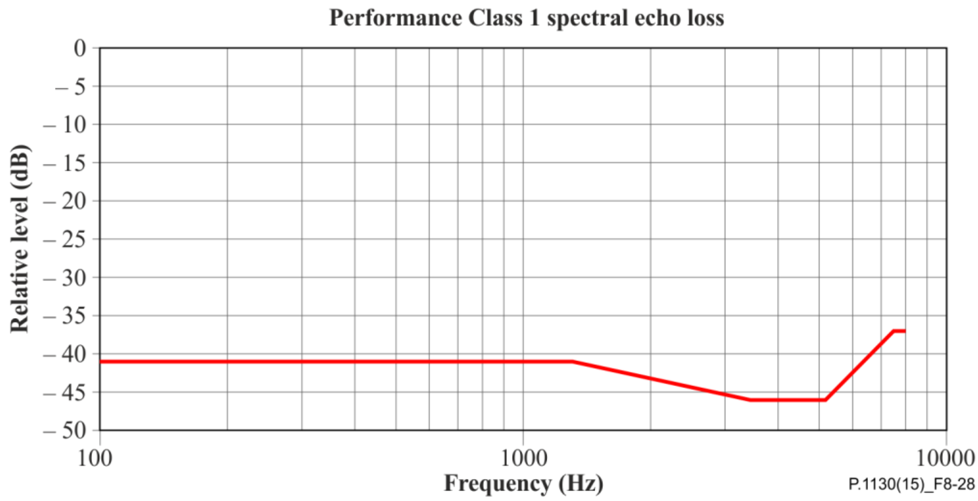


Figure 8-28 – Spectral echo loss mask Performance Class 1 (Figure is informative)

8.4.23.4 Design guidance and root-cause analysis

Spectral echo components exceeding the limits as given in this measurement may cause audible echo impairments in certain frequency ranges. A potential cause may be not optimum adjustment of the echo canceller in conjunction with the non-linear processor used in the echo canceller to suppress residual echo. In case of sub-band echo cancellation the individual adjustment in each frequency band should be checked. Care should be taken not to reduce double-talk performance, e.g., if the non-linear processing is improved.

8.4.24 Signal enhancement initial convergence without background noise (IC_{SE})

8.4.24.1 Parameter description

This parameter measures acoustic echo loss as a function of time immediately after call set-up when there is no background noise. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3).

8.4.24.2 Test

1) The test arrangement is in accordance with clause 8.4.1. The noise level measured at the electrical access point (idle channel noise) shall be less than −63 dBm0.

2) The test signal is applied immediately after setting up the call and setting the volume control to its maximum.

3) The test signal is a composite source signal according to [ITU-T P.501] repeated periodically. The average signal level is −16 dBm0. The echo signal is analysed over a period of at least 5 s. The analysis integration time is 35 ms, the analysis is referred to the level analysis of the reference signal.

4) In addition, the test is repeated with the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The average test signal level is –16 dBm0. The echo signal is analysed over a period of at least 5 s. The analysis integration time is 35 ms; the analysis is referred to the level analysis of the reference signal.

5) The measurement is displayed as echo attenuation vs time, measured signal and reference signal have to be synchronized in time.

NOTE 1 – The analysis of the CSS is performed only on the active signal parts, the pauses between the bursts of the composite source signal are not analysed. The analysis time is reduced by the time constant of the level analysis due to the integration time of 35 ms.

NOTE 2 – The required performance for speech signal should be achieved for different starting points of the speech signal.

8.4.24.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the IC_{SE} defined in Table 8-74.

Table 8-74 – Limits for the signal enhancement subsystem initial convergence

Time (ms)	Performance Class 1	Performance Class 2	Performance Class 3	Performance Class 4
200	≥ 6 dB	FFS	FFS	FFS
1 000	≥ 20 dB	FFS	FFS	FFS
1 200	≥ 40 dB	FFS	FFS	FFS

8.4.24.4 Design guidance and root-cause analysis

For further study.

8.4.25 Signal enhancement initial convergence with background noise ($ICBN_{SE}$)

8.4.25.1 Parameter description

This parameter measures acoustic echo loss as a function of time immediately after call set-up when background noise is present. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3).

8.4.25.2 Test method

1) The test arrangement is in accordance with clause 8.4.1.

2) All of the background noise conditions defined by the user scenarios in Annex B shall be tested. For the test at the output of the microphone subsystem (S2) is recorded and inserted as the background noise at test point (S2).The background noise is played back at least 5 s before the start of the actual measurement. This allows time for some adaptive algorithms in the signal enhancement subsystem, which are constantly monitoring the microphone signal, to stabilize (e.g., AGC, NR). The test is conducted under simulated constant driving conditions.

3) The test signal is applied immediately after setting up the call and setting the volume control to its maximum.

4) The test signal is a composite source signal according to [ITU-T P.501] repeated periodically. The average signal level is −16 dBm0. The echo signal is analysed over a period of at least 5 s. The analysis integration time is 35 ms; the analysis is referred to the level analysis of the reference signal.

5) In addition, the test is repeated with the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The average test signal level is −16 dBm0. The echo signal is analysed over a period of at least 5 s. The analysis integration time is 35 ms, the analysis is referred to the level analysis of the reference signal.

6) The measurement is displayed as echo vs. time. The echo level above the background noise is determined.

NOTE 1 – The analysis of the CSS is performed only on the active signal parts, the pauses between the bursts of the composite source signal are not analysed. The analysis time is reduced by the time constant of the level analysis due to the integration time of 35 ms.

NOTE 2 – The required performance for the speech signals should be achieved for different starting points of the speech signal.

8.4.25.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the $ICBN_{SE}$ defined in Table 8-75.

Table 8-75 – Limits for the signal enhancement subsystem echo level above background noise

Time (ms)	Performance Class 1	Performance Class 2	Performance Class 3	Performance Class 4
100	≤ 10 dB	FFS	FFS	FFS
1 500	≤ 0 dB	FFS	FFS	FFS

8.4.25.4 Design guidance and root-cause analysis

For further study.

8.4.26 Signal enhancement echo performance with time variant echo path ($TVEP_{SE}$)

8.4.26.1 Parameter description

This parameter measures acoustic echo loss when the echo path changes during the measurement. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3). For the test a time variant echo path needs to be realized. Appendix III describes a method how to record and apply time variant impulse responses.

8.4.26.2 Test

1) Before conducting the test, the echo canceller should be fully converged.

2) The test arrangement is in accordance with clause 8.4.1.

3) The time variant echo path simulation is time-synchronized with the playback of the test signals. The simulation starts rotating synchronously with the start of the test signals.

4) The test signal consists of a periodically repeated composite source signal according to [ITU-T P.501] with an average level of −5 dBm0 as well as an average level of −25 dBm0. The echo signal is analysed during a period of at least 2.8 s, which represents 8 periods of the CS signal. The integration time for the level analysis shall be 35 ms; the analysis is referred to the level analysis of the reference signal.

5) In addition, the test is repeated with the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The average test signal level is −16 dBm0. The echo signal is analysed over a period of at least 5 s. The analysis integration time is 35 ms; the analysis is referred to the level analysis of the reference signal.

6) The measurement result is displayed as attenuation versus time. The exact synchronization between input and output signal has to be guaranteed.

NOTE – When using the CSS, the analysis is conducted only during the active signal part, the pauses between the composite source signals are not analysed. The analysis time is reduced by the integration time of the level analysis (35 ms).

8.4.26.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the $TVEP_{SE}$ defined in Table 8-76.

Table 8-76 – Limits for the signal enhancement subsystem echo maximum deviation of echo level vs. maximum echo level

Performance Class	$TVEP_{SE}$
1	≤ 3 dB
2	≤ 6 dB
3	≤ 9 dB
4	> 9 dB

8.4.26.4 Design guidance and root-cause analysis

For further study.

8.4.27 Signal enhancement echo performance with time variant echo path and speech ($TVEP_{SE}$-SP)

8.4.27.1 Parameter description

This parameter measures acoustic echo loss when the echo path changes during the measurement and the British-English single talk sequence ([ITU-T P.501]) is used as the test signal. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction (R3), to the output of the signal enhancement subsystem in the send direction (S3). For the test a time variant echo path needs to be realized. Appendix III describes a method how to record and apply time variant impulse responses.

8.4.27.2 Test

1) Before conducting the test the echo canceller should be fully converged.
2) The echo path remains constant.
3) The test arrangement is in accordance with clause 8.4.1.
4) The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The average test signal level is −16 dBm0. The echo signal is analysed during the complete test signal. The integration time for the level analysis shall be 35 ms. The measurement result is calculated as level vs. time. The echo level is determined under steady state conditions and stored as reference.
4) Now the echo path variation is started.
5) The test is repeated with the time variant echopath active and the British-English speech [ITU-T P.501]. The average test signal level is −16 dBm0. The echo signal is analysed during the complete test signal. The integration time for the level analysis shall be 35 ms. The measurement result is calculated as level vs. time.
6) The difference of the echo level between the reference and the measured echo loss with the time variant echo path active is determined.

7) The measurement result is displayed as attenuation vs. time. The exact synchronization between the two measured signals has to be guaranteed.

8.4.27.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the TVEP$_{SE}$-SP defined in Table 8-77.

Table 8-77 – Limits for the signal enhancement subsystem echo maximum deviation of echo level vs. maximum echo level

Performance Class	TVEP$_{SE}$-SP
1	≤ 3 dB
2	≤ 6 dB
3	≤ 9 dB
4	> 9 dB

8.4.27.4 Design guidance and root-cause analysis

For further study.

8.4.28 Signal enhancement send activation (SA$_{SE}$)

8.4.28.1 Parameter description

This parameter measures the minimum speech signal level required to activate the send transmission path (S2) up to the output of the signal enhancement subsystem in the send direction (S3).

The activation in the send direction is mainly determined by the built-up time T_r,, S-$_{,min}$ and the minimum activation level (SAL$_{SE}$-$_{,min}$). The minimum activation level is the level required to remove the inserted attenuation in the send direction during idle mode. The built-up time is determined for the test signal burst which is applied with the minimum activation level.

8.4.28.2 Test

The structure of the test signal is shown in Figure 8-29. The test signal consists of CSS components according to [ITU-T P.501] with increasing level for each CSS burst.

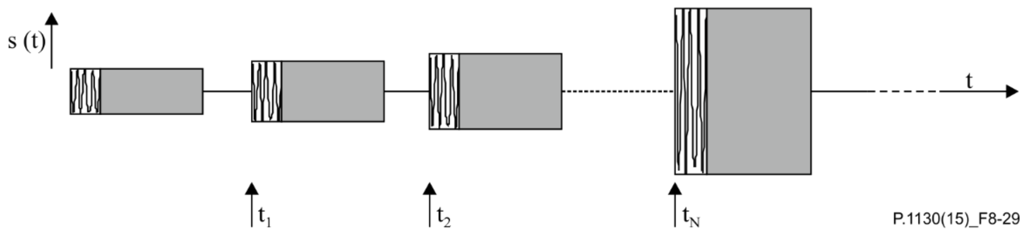


Figure 8-29 – Test signal to determine the minimum activation level and the built-up time

The settings of the test signal are as follows.

Table 8-78 – Settings of the CSS in send

	CSS duration/ pause duration	Level of the first CS signal (active signal part at the MRP)	Level difference between two periods of the test signal
CSS to determine switching characteristic in send direction	248.62 ms/451.38 ms	−20 dB rel. to nominal signal level (Note 1)	1 dB
NOTE 1 – The level of the active signal part corresponds to an average level of the nominal signal level when −24.7 dBPa is produced at the MRP for the CSS according to [ITU-T P.501], assuming a pause of 101.38 ms. NOTE 2 – The signal level at the MRP is corrected at the HATS-HFRP.			

It is assumed that the pause length of 451.38 ms is longer than the hang-over time so that the test object is back to idle mode after each CSS burst.

1) The test arrangement is described in clause 8.4.1.
2) The signal is applied at test point (S2).
3) The level of the transmitted signal is measured at test point (S3). The measured signal level is referred to the test signal level and displayed versus time. The levels are calculated from the time domain using an integration time of 5 ms.
4) The minimum activation level is determined from the CSS burst which indicates the first activation of the test object. The time between the beginning of the CSS burst and the complete activation of the test object is measured.

NOTE – If the measurement using the CS signal does not allow to clearly identify the minimum activation level, the measurement may be repeated by using the one syllable word described in [ITU-T P.501], clause 7.3.4 instead of the CS signal.

8.4.28.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the SAL_{SE}-$_{min}$ and $T_{r,,}$ S_{-min} defined in Table 8-79.

Table 8-79 – Limits for the signal enhancement minimum activation level and build-up time

Performance Class	SAL_{SE-min}	$T_{r,,}$ $S_{–min}$
1	≤ 20 dB re. nominal signal level	≤ 5ms
2	≤ 17 dB re. nominal signal level	≤ 50 ms
3	≤ 10 dB re. nominal signal level	≤ 100 ms
4	> 10 dB re. nominal signal level	> 100 ms

8.4.28.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables of a user's speech for a wide range of speech levels. A poor performance in this measurement may lead to initial syllable or even complete word clipping. In such an event speech detection algorithms or double-talk detection functions in the echo canceller should be optimized.

8.4.29 Signal enhancement receive activation (RA_{SE})

8.4.29.1 Parameter description

This parameter measures the minimum speech signal level required to activate the receive transmission path (R3) up to the output of the signal enhancement subsystem in the receive direction (R2).

The activation in the receive direction is mainly determined by the built-up time $T_{r,}$, R_{-min} and the minimum activation level ($RAL_{SE\text{-}min}$). The minimum activation level is the level required to remove completely any attenuation inserted during the idle mode. The built-up time is determined from the level variation of the transmitted test signal which is applied with a minimum activation level.

8.4.29.2 Test

The signal construction is shown in Figure 8-29. The test signal settings are as follows.

Table 8-80 – Settings of the CSS in receive

	CSS duration/pause duration	Level of the first CS signal (active signal part)	Level difference between two periods of the test signals
CSS to determine switching characteristics in receive direction	248.62 ms/451.38 ms	−22.7 dBm rel. to nominal signal level (Note)	1 dB
NOTE – The level of the active signal part corresponds to an average level of −24 dBm re. to the nominal signal level for the CSS according to [ITU-T P.501], assuming a pause of 101.38 ms.			

1) The test arrangement is in accordance with clause 8.4.1.
2) The signal is applied at test point (R3).
3) The transmitted signal is recorded at R2.
4) The measured result is corrected by the measured sensitivity frequency response characteristics of the audio subsystem in order to take into account any equalization potentially performed in the audio subsystem as well as the loudspeaker response characteristics. Alternatively, a simulated R2 to DRP simulation could be used. In such a case the simulated output signal would be referred to the measured power density spectrum at test point (R3) interface.
5) The measured signal level is referred to the test signal level and displayed versus time. The integration time of the level analysis used should be 5 ms.
6) The minimum activation level is determined from the CSS burst indicating the first activation of the test object. The duration between the beginning of this CSS burst and the complete activation of the terminal is measured.

NOTE – If the measurement using the CS signal does not allow to clearly identify the minimum activation level, the measurement may be repeated by using a one syllable word described in [ITU-T P.501], clause 7.3.4 instead of the CS signal.

8.4.29.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the RAL_{SE}-min and $T_{r,}$, R_{-min} defined in Table 8-81.

Table 8-81 – Limits for the signal enhancement minimum activation level and build-up time

Performance Class	$RAL_{SE\text{-}min}$	$T_{r,,} R_{\text{-}min}$
1	≤ 20 dB re. nominal signal level	≤ 5ms
2	≤ 17 dB re. nominal signal level	≤ 50 ms
3	≤ 10 dB re. nominal signal level	≤ 100 ms
4	> 10 dB re. nominal signal level	> 100 ms

8.4.29.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables of a user's speech for a wide range of speech levels. A poor performance in this measurement may lead to initial syllable or even complete word clipping. In such an event speech detection algorithms or double-talk detection functions in the echo canceller should be optimized.

8.4.30 Signal enhancement send attenuation range (AH,S_{SE})

8.4.30.1 Parameter description

This parameter measures the range of attenuation in the send direction during single-talk due to the signal enhancement subsystem.

The attenuation range in the send direction is determined by applying the test signal in the send direction at S2 after the signal enhancement subsystem was activated in the receive direction at R3. During the measurement, the attenuation range in the send direction (AH,S) and the built-up time in the send direction ($T_{r,,}$ S) is determined.

8.4.30.2 Test

The structure of the test signals is shown in Figure 8-30. It consists of periodically repeated composite source signal bursts used for activating the receive direction and the voiced sound used to measure the send direction.

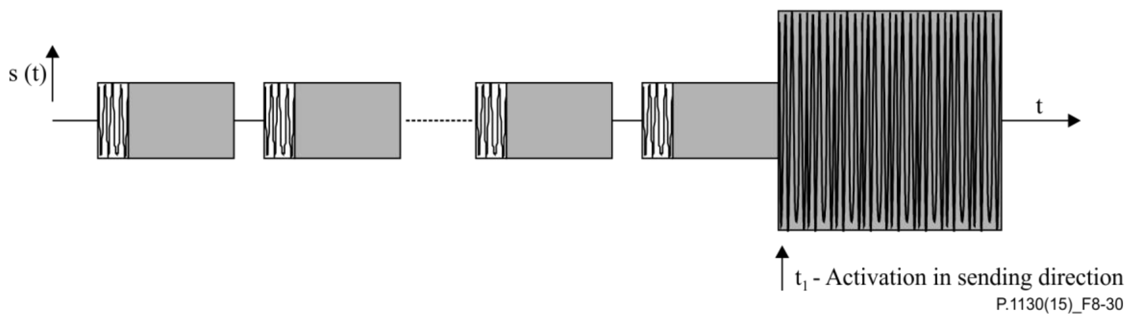


Figure 8-30 – Structure of the test signal for measuring the attenuation range

The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement should be constant during the measurement.

1) The test arrangement is in accordance with clause 8.4.1.

2) The test signal used is according to Figure 8-30; the receive direction is activated first. The measurement parameters are as follows.

Table 8-82 – Signal levels for double-talk tests in send and receive

	Receive direction	Send direction
Average signal level (including 101.38-ms pauses)	Nominal signal level	–
Active signal part	Nominal signal level −1.3 dB	Nominal signal level −1.3 dB

The level in the receive direction is determined at R3.

3) The signals are applied at test points (R3) and (S2).

4) The transmitted signal in send is recorded at S3.

5) The level is determined as level versus time calculated from the time domain. The integration time of the levels analysis is 5 ms. The attenuation range is determined by calculating the difference between the measured level between the beginning of the test signal at S3 (t_1 in Figure 8-30) until complete activation in the send direction at S3.

8.4.30.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the AH, S-min and $T_{r,,}$ S$_{\mathbf{-min}}$ defined in Table 8-83.

Table 8-83 – Limits for the signal enhancement minimum activation level and build-up time

Performance Class	AH,S$_{\mathbf{-min}}$	$\mathbf{T_{r,,}}$ S$_{\mathbf{-min}}$
1	≤ 10 dB re. nominal signal level	≤ 15ms
2	≤ 20 dB re. nominal signal level	≤ 50 ms
3	≤ 20 dB re. nominal signal level	≤ 100 ms
4	> 20 dB re. nominal signal level	> 100 ms

8.4.30.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables in a conversational situation where the conversational partner responds immediately to his other partner. A poor performance in this measurement may lead to initial syllable or even complete word clipping and such answers might get lost. In such an event mainly the double-talk detection function and the NLP implementation in the echo canceller is the cause of the problem and should be optimized.

8.4.31 Signal enhancement receive attenuation range (AH,R$_{SE}$)

8.4.31.1 Parameter description

This parameter measures the range of attenuation in the receive direction during single-talk due to the signal enhancement subsystem.

The attenuation range in the receive direction is determined by applying the test signal in the receive direction at R3 after the signal enhancement subsystem was activated in the send direction at S2. During the measurement, the attenuation range in the receive direction (AR,S) and the built-up time in the receive direction ($T_{r,,}$ R) is determined.

8.4.31.2 Test

The structure of the test signals is shown in Figure 8-31. It consists of periodically repeated composite source signal bursts used for activating the receive direction and the voiced sound used to measure the send direction.

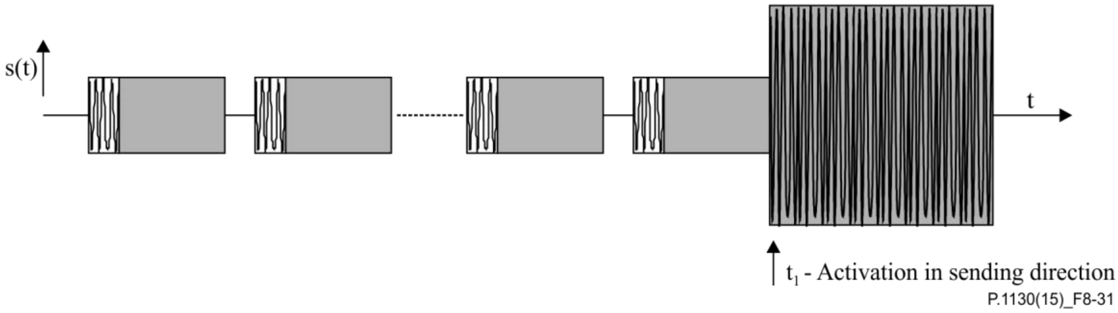


Figure 8-31 – Structure of the test signal for measuring the attenuation range

The test signals are synchronized in time at S2. The delay of the test arrangement should be constant during the measurement.

The structure of the test signal is shown in Figure 8-31. CSS bursts are used for activating the send direction and the voiced sound is used to measure the receive direction. The delay of the test arrangement should be constant during the measurement.

1) The test arrangement is in accordance with clause 8.4.1.
2) The test signal shown in Figure 8-31 is used; the send direction is activated first.

   The measurement parameters are as follows.

Table 8-84 – Signal levels for double-talk tests in send and receive

	Receive direction	Send direction
Average level (including 101.38-ms pauses)	–	Nominal signal level
Active signal part	Nominal signal level −1.3 dB	Nominal signal level −1.3 dB

   The level in the receive direction is determined at the electrical reference point.

3) The signals are applied at test points (R3) and (S2).
4) The transmitted signal in receive is recorded at R2.
5) The level is determined as level versus time calculated from the time domain. The integration time of the level analysis is 5 ms. The attenuation is determined by calculating the difference between the beginning of the measured test signal in the receive direction at R2 (t_1 in Figure 8-31) and the complete activation in the receive direction.

8.4.31.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the AH, R-min and $T_{r,,\ R}$S**-min** defined in Table 8-85.

Table 8-85 – Limits for the signal enhancement minimum activation level and build-up time

Performance Class	$A_{H,R-min}$	$T_{r,,R-min}$
1	≤ 10 dB re. nominal signal level	≤ 15ms
2	≤ 20 dB re. nominal signal level	≤ 50 ms
3	≤ 20 dB re. nominal signal level	≤ 100 ms
4	> 20 dB re. nominal signal level	> 100 ms

8.4.31.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables in a conversational situation where the conversational partner responds immediately to his other partner. A poor performance in this measurement may lead to initial syllable or even complete word clipping and such answers might get lost. In such an event any speech activated attenuation in receive or eventually the double-talk detection functions in the echo canceller may be the cause of the problem and should be optimized.

8.4.32 Signal enhancement attenuation range in send during double-talk ($A_{H,S,dtSE}$)

8.4.32.1 Parameter description

This parameter measures the range of attenuation in the send direction during double-talk due to the signal enhancement subsystem.

The attenuation range in send is measured from S2 to S3.

The requirements apply to both nominal and maximum settings of the receive volume control.

The requirements apply to nominal signal levels in the send and receive directions as well as for the level combinations +6 dB (re. nominal level) in the send/−6 dB (re. nominal level) in receive; and +6 dB (re. nominal level) in receive/−6 dB (re. nominal level) in send. Furthermore, the test is conducted with nominal levels but with maximum setting of the volume control.

8.4.32.2 Test

The test signal to determine the attenuation range during double talk is shown in Figure 8-32. The test signal to determine the attenuation range during double talk is the double-talk speech sequence as defined in clause 7.3.5 of [ITU-T P.501]. The competing speaker is always inserted as the double-talk sequence sdt(t) in receive and is used for analysis. The test signals are synchronized in time at the S2 interface. The delay of the test arrangement should be constant during the measurement.

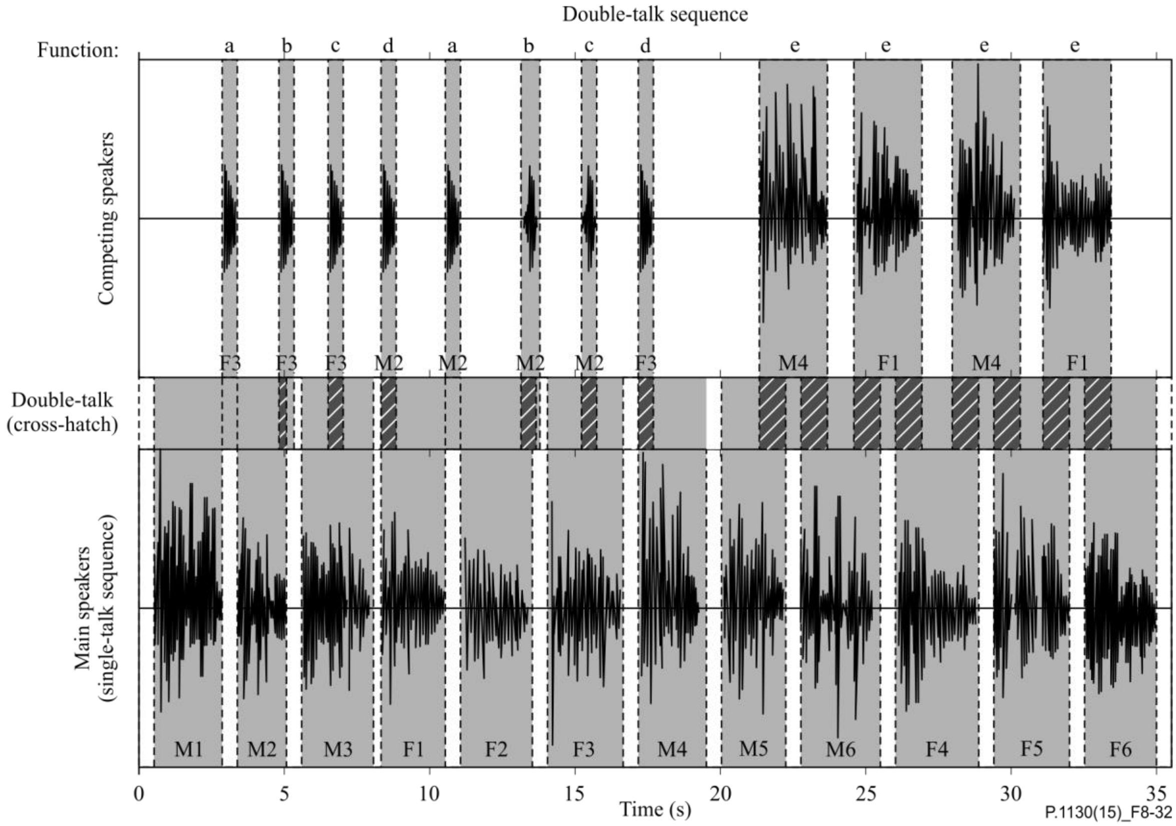


Figure 8-32 – Double-talk test sequence with overlapping CS-signals in send and receive direction

The settings for the test signals are as follows:

Table 8-86 – Timing of the double-talk sequences

	Receive direction	**Send direction**
Average signal level	Nominal signal level	Nominal signal level

The tests are repeated with maximum volume control setting in the receive direction.

1) The test arrangement is in accordance with clause 8.4.1. Before the actual test a training sequence for the echo canceller consisting of the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with nominal signal level is applied to R3.

2) When determining the attenuation range in the send direction the signal measured at S2 is referred to the test signal inserted.

3) The attenuation range during double talk is determined as described in Appendix III of [ITU-T P.502]. The double-talk performance is analysed for each word and sentence produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

4) The test is repeated for all level combinations as defined in the parameter description.

8.4.32.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the AH, S,dt_{SE} defined in Table 8-87.

Table 8-87 – Categorization of double-talk capability according to [ITU-T P.340]

Category (according to [ITU-T P.340])	1	2a	2b	2c	3
	Full duplex capability	Partial duplex capability			No duplex capability
AH,S,dt_{SE} [dB]	≤ 3	≤ 6	≤ 9	≤ 12	> 12

The requirements apply to both nominal and maximum settings of the receive volume control.

8.4.32.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables and complete words in conversational double situations. A poor performance in this measurement may lead to initial syllable or even complete word clipping and such information sent by the conversational partner might get lost during double talk. In such an event mainly the double-talk detection functions and the NLP implementation in the echo canceller is the cause of the problem and should be optimized.

8.4.33 Signal enhancement attenuation range in receive during double-talk (AH,Rdt_{SE})

8.4.33.1 Parameter description

This parameter measures the range of attenuation in the receive direction during double-talk due to the signal enhancement subsystem.

The attenuation range in receive is measured from R3 to R2.

The requirements apply to both nominal and maximum settings of the receive volume control.

The requirements apply to nominal signal levels in the send and receive directions as well as for the level combinations +6 dB (re. nominal level) in the send/−6 dB (re. nominal level) in receive; and +6 dB (re. nominal level) in receive/−6 dB (re. nominal level) in send. Furthermore, the test is conducted with nominal levels but with maximum setting of the volume control.

8.4.33.2 Test

The test signal to determine the attenuation range during double talk is shown in Figure 8-33. The test signal to determine the attenuation range during double talk is the double-talk speech sequence as defined in clause 7.3.5 of [ITU-T P.501]. The competing speaker is always inserted as the double-talk sequence sdt(t) in receive and is used for analysis. The test signals are synchronized in time at the S2 interface. The delay of the test arrangement should be constant during the measurement.

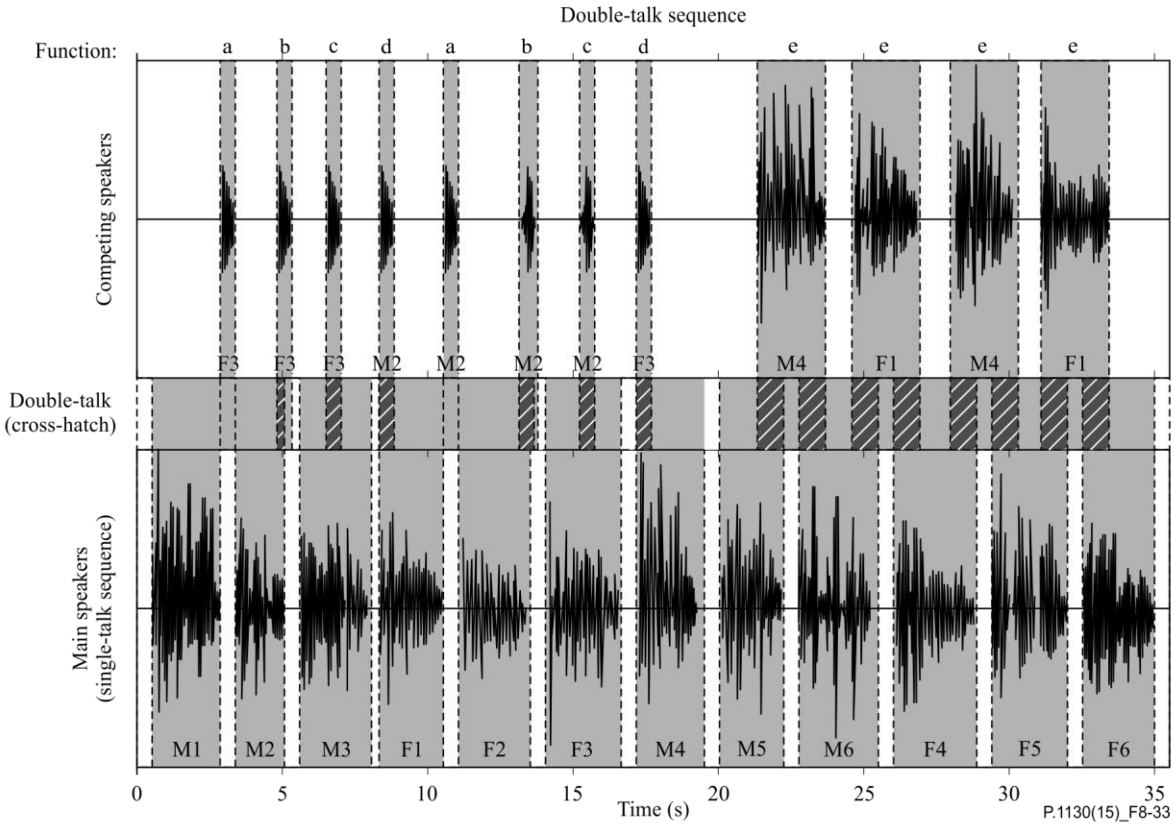


Figure 8-33 – Double-talk test sequence with overlapping CS-signals in receive and send direction

The settings for the test signals are as follows:

Table 8-88 – Timing of the double-talk sequences

	Receive direction	Send direction
Average signal level	Nominal signal level	Nominal signal level

The tests are repeated with maximum volume control setting in the receive direction.

1) The test arrangement is in accordance with clause 8.4.1. Before the actual test a training sequence for the echo canceller consisting of the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with nominal signal level is applied to R3.

2) When determining the attenuation range in the receive direction the signal measured at the R2 interface of the signal enhancement subsystem is referred to the test signal inserted.

3) The attenuation range during double talk is determined as described in Appendix III of [ITU-T P.502]. The double-talk performance is analysed for each word and sentence

produced by the competing speaker. The requirement has to be met for each word and sentence produced by the competing speaker.

4) The test is repeated for all level combinations as defined in the requirements.

8.4.33.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the AH, R,dt_{SE} defined in Table 8-89.

Table 8-89 – Categorization of double-talk capability according to [ITU-T P.340]

Category (according to [ITU-T P.340])	**1**	**2a**	**2b**	**2c**	**3**
	Full duplex capability	Partial duplex capability			No duplex capability
AH,R,dt_{SE} [dB]	≤ 3	≤ 5	≤ 8	≤ 10	> 10

The requirements apply to both nominal and maximum settings of the receive volume control.

8.4.33.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables and complete words in conversational double situations. A poor performance in this measurement may lead to initial syllable or even complete word clipping and thus information sent by the conversational partner might get lost during double talk. In such an event any AGC function, speech controlled activation or the double-talk detection functions in the echo canceller may be the cause of the problem and should be optimized.

8.4.34 Detection of echo components during double-talk after enhancement ($DEC_{DT\text{-}SE}$)

8.4.34.1 Parameter description

This parameter measures acoustic echo loss during double-talk. Echo loss is measured from the input to the signal enhancement subsystem in the receive direction, to the output of the signal enhancement subsystem in the send direction.

The echo components are measured from S2 to S3.

The echo attenuation during double talk is based on the parameter talker echo loudness rating ($TELR_{dt}$). It is assumed that the terminal at the opposite end of the connection provides nominal loudness rating (SLR + RLR = 10 dB). "Echo loss" is the echo suppression provided by the hands-free terminal measured at the electrical reference point. Under these conditions, the performance classification is applicable (more information can be found in Annex A of [ITU-T P.340]).

8.4.34.2 Test

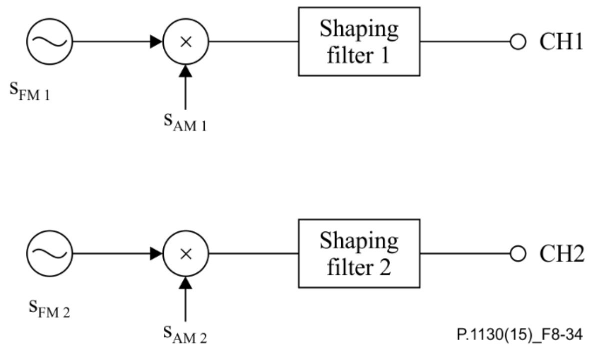
1) The test arrangement is in accordance with clause 8.4.1.

2) The double-talk signal consists of a sequence of orthogonal signals which are realized by voice-like modulated sine waves spectrally shaped similar to speech. The measurement signals used are shown in Figure 8-34. A detailed description can be found in [ITU-T P.501].

The signals are fed simultaneously in the send and receive directions. The level in the send direction is the nominal signal level at S2, the level in the receive direction is the nominal signal level at R3.

3) The test signal is measured at the S3. The measured signal consists of the double-talk signal which was fed in at S2 and the echo signal. The echo signal is filtered by a comb filter using mid-frequencies and bandwidth according to the signal components of the signal in the receive direction (see [ITU-T P.501]). The filter will suppress frequency components of the double-talk signal.

4) In each frequency band which is used in the receive direction the echo attenuation can be measured separately. The requirement for the corresponding category is fulfilled if in any frequency band the echo signal is either below the signal noise or below the required limit. If echo components are detectable, the classification is based on Table 8-90. In narrowband the echo attenuation is to be achieved for each individual frequency band from 200 Hz to 3 450 Hz according to the different categories. In wideband the echo attenuation is to be achieved from 200 Hz to 6 950 Hz according to the different categories.



$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t n * F_{01,2})\ ;\ n = 1, 2, ...$$

$$s_{AM1,2}(t) = \sum A_{AM1,2} * \cos(2\pi t F_{AM1,2})\ ;$$

In wideband the settings for the signals are as follows, in narrowband the frequencies marked in grey are used:

Receiving direction			**Sending direction**		
f_m(Hz)	**$f_{mod(fm)}$(Hz)**	**F_{am}(Hz)**	**f_m(Hz)**	**$f_{mod(fm)}$(Hz)**	**F_{am}(Hz)**
125	±2.5	3	150	±2.5	3
250	±5	3	270	±5	3
500	±10	3	540	±10	3
750	±15	3	810	±15	3
1 000	±20	3	1 080	±20	3
1 250	±25	3	1 350	±25	3
1 500	±30	3	1 620	±30	3
1 750	±35	3	1 890	±35	3
2 000	±40	3	2 160	±35	3
2 250	±40	3	2 400	±35	3

Receiving direction			Sending direction		
f_m(Hz)	$f_{mod(fm)}$(Hz)	F_{am}(Hz)	f_m(Hz)	$f_{mod(fm)}$(Hz)	F_{am}(Hz)
2 500	±40	3	2 650	±35	3
2 750	±40	3	2 900	±35	3
3 000	±40	3	3 150	±35	3
3 250	±40	3	3 400	±35	3
3 500	±40	3	3 650	±35	3
3 750	±40	3	3 900	±35	3
4 000	±40	3	4 150	±35	3
4 250	±40	3	4 400	±35	3
4 500	±40	3	4 650	±35	3
4 750	±40	3	4 900	±35	3
5 000	±40	3	5 150	±35	3
5 250	±40	3	5 400	±35	3
5 500	±40	3	5 650	±35	3
5 750	±40	3	5 900	±35	3
6 000	±40	3	6 150	±35	3
6 250	±40	3	6 400	±35	3
6 500	±40	3	6 650	±35	3
6 750	±40	3	6 900	±35	3
7 000	±40	3			
NOTE – Parameters of the shaping filter: F ≥ 250 Hz: low pass filter, 5 dB/oct; f < 250 Hz,: high pass filter.					

Figure 8-34 – Parameters of the two test signals for double-talk measurement based on AM-FM modulated sine waves

NOTE – Some implementations may fail this requirement due to perceptually-based spectral filters which allow low levels of the double-talk signal to leak into the analysis window used for measuring echo. If it can be demonstrated that failures are not caused by echo then the device under test (DUT) is considered compliant with this requirement.

8.4.34.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the $DEC_{DT\text{-}SE}$ defined in Table 8-90.

Table 8-90 – Categorization of double-talk capability according to [ITU-T P.340]

Category (according to [ITU-T P.340])	**1**	**2a**	**2b**	**2c**	**3**
	Full duplex capability	*Partial duplex capability*			*No duplex capability*
Echo loss [dB]	≥27	≥23	≥17	≥11	<11

8.4.34.4 Design guidance and root-cause analysis

This measurement ensures a sufficient suppression of echo components in conversational double situations. A poor performance in this measurement may lead to annoying echo components which may disturb the conversational partner during double talk. In such an event typically the double-talk detection functions in the echo canceller and NLP implementations are the cause of the problem and should be optimized.

8.4.35 Signal enhancement sent speech attenuation during double-talk ($SSA_{DT\text{-}SE}$)

8.4.35.1 Parameter description

This parameter measures the range of attenuation in the send direction immediately after the onset double-talk due to the signal enhancement subsystem. It is intended to complement the "Enhanced Attenuation Range in Send during Double-Talk" measurement parameter by verifying the attenuation range at the onset of double-talk is consistent with that measured during double-talk.

The sent speech attenuation range during double talk is measured from S2 to S3.

The requirements apply to both nominal and maximum settings of the receive volume control.

The requirements apply to nominal signal levels in send and receive directions as well as for the level combinations +6 dB (re. nominal level) in send/–6 dB (re. nominal level) in receive and +6 dB (re. nominal level) in receive/–6 dB (re. nominal level) in send. Furthermore, the test is conducted with nominal levels but with maximum setting of the volume control.

8.4.35.2 Test

1) The test arrangement is in accordance with clause 8.4.1.

2) The double-talk signal consists of a sequence of orthogonal signals which are realized by voice-like modulated sine waves spectrally shaped similar to speech. The measurement signals used are shown in Figure 8-34. A detailed description can be found in [ITU-T P.501].

   The signals are fed simultaneously in the send and receive directions. The level in the send direction at S2 is the nominal signal level, the level in the receive direction at S3 is the nominal signal level.

3) The test signal is measured at S3 (send direction). The measured signal consists of the double-talk signal which was fed in by the artificial mouth and the echo signal. The double-talk signal (send signal) is filtered out by a comb filter using mid-frequencies and bandwidth according to the signal components of the signal in the send direction (see [ITU-T P.501]). The filter will suppress frequency components of the echo signal.

4) In each frequency band which is used in the send direction, the sent speech attenuation, $A_{H,S,dt}$, can be measured separately. The requirement for each category is fulfilled if in each frequency band the attenuation of the signal in the send direction is below the required limit. If attenuation is detectable, the classification is based on Table 8-91. In narrowband the sent speech attenuation $A_{H,S,dt}$ is to be achieved for each individual frequency band from 200 Hz

to 3 550 Hz according to the different categories. In wideband the echo attenuation is to be achieved for each individual frequency band from 200 Hz to 6 900 Hz according to the different categories.

5) The test is repeated for all level combinations as defined in the requirements.

8.4.35.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class of this parameter after signal enhancement has to meet the requirements for the applicable performance classes for the AH, S,dt_{SE} defined in Table 8-91.

Table 8-91 – Categorization of double-talk capability according to [ITU-T P.340]

Category (according to [ITU-T P.340])	**1**	**2a**	**2b**	**2c**	**3**
	Full duplex capability	*Partial duplex capability*			*No duplex capability*
AH,S,dt_{SE} [dB]	≤3	≤6	≤9	≤12	>12

The requirements apply to both nominal and maximum settings of the receive volume control.

8.4.35.4 Design guidance and root-cause analysis

This measurement ensures a proper transmission of initial syllables and complete words in double-talk situations. A poor performance in this measurement may lead to initial syllable or even complete word clipping and thus information sent by the conversational partner might get lost during double talk. In such an event the double-talk detection function and the NLP implementation in the echo canceller may be the cause of the problem and should be optimized.

8.4.36 Signal enhancement speech-to-echo ratio (SpERE)

8.4.36.1 Parameter description

This measurement parameter is for further study. It provides a measurement of the send speech level relative to the echo level at the output of the signal enhancement subsystem. One possible test method proposed is to use temporally-weighted terminal coupling loss with varying echo path to calculate the echo portion.

8.4.37 SNR improvement after enhancement (SNRIE)

8.4.37.1 Parameter description

This measurement parameter is for further study. It provides a measurement of the send speech level relative to the noise level at the output of the signal enhancement subsystem.

8.4.38 Signal enhancement background noise transmission after call set-up (BGNTACSE)

8.4.38.1 Parameter description

This parameter predicts the transmitted noise quality after call set-up at the output of the signal enhancement subsystem. The measurement is intended to avoid audible noise bursts which might occur at the beginning of a call, e.g., in the adaptation phase of noise cancellers.

The signal enhancement background noise transmission is measured from S2 to S3.

8.4.38.2 Test

1) The test arrangement is given in clause 8.4.1.
2) According to the specification of the manufacturer/test laboratory, the background noise is played back. The test should be carried out during a constant driving situation. The signal is simultaneously in the send directions. The level in the send direction at S2 is the nominal signal level.
3) The signal enhancement subsystem is reset. The test is started after resetting the system.
4) The test signal is measured at S3 (send direction). The transmitted signal in the send direction is recorded at least 1 s before the start of the test and for at least 7 s after the test has started. The analysis range is chosen at 8 s including an initial pause of 1 s before the test was started.
5) The recorded signal is analysed using the relative approach (see [ITU-T P.1100]).

8.4.38.3 Performance level classification based on values of this parameter

Table 8-92 – Signal enhancement background noise transmission after call set-up

Performance Class	BGNTACSE first signal peak	$BGNTACSE_n$
1	≤ 3 cp/cPa	≤ 6 cp/cPa
2	≤ 6 cp/cPa	≤ 15 cp/cPa
3	≤ 15 cp/cPa	≤ 30 cp/cPa
4	> 15 cp/cPa	> 30 cp/cPa

8.3.38.4 Design guidance and root-cause analysis

This test is intended to avoid audible artefacts of signal processing adjustment/adaptation dealing background noise at the beginning of a call. Audible impairments such as noise bursts, tonal components or other are detected by this measurement. Readjustment of the noise or echo cancellation parameters may help to improve the performance.

8.4.39 Enhanced speech quality in the presence of background noise (SQPBGNE)

8.4.39.1 Parameter description

This parameter predicts speech quality under noisy conditions in the send direction after the signal enhancement subsystem.

The enhanced speech quality in the presence of background noise is measured from S2 to S3.

8.4.39.2 Test

1) The test arrangement is given in clause 8.4.1.
2) According to the specification of the manufacturer/test laboratory, the background noise is inserted at test point [S2]. The background noise should be applied for at least 5 s in order to adapt noise reduction algorithms in advance of the test. The background noise signal level is according to the background noise signal level measured at S2 when driving with 80 km/h and 130 km/h.
3) The near-end speech signal consists of 8 sentences of speech (2 male and 2 female talkers, 2 sentences each). Appropriate speech samples can be found in [ITU-T P.501]. The preferred language for wideband is French since the objective method was validated with the French language. The preferred language in narrowband is English. The test signal level is the nominal signal when simulating 89 km/h driving conditions. For simulating 130 km/h driving

conditions, the speech level is adjusted at test point (S2) taking into account the Lombard effect. The test signal level is the signal level according to clause 8.1 measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.Three signals are required for the tests:

– The clean speech signal is used as the undisturbed reference (see [b-ETSI ES 202 396-3]).
– The speech plus undisturbed background noise signal is recorded at the terminal's microphone position using an omnidirectional measurement microphone with a linear frequency response between 50 Hz and 12 kHz.
– The send signal is recorded at the electrical reference point.

S-MOS-LQO, N-MOS-LQO and G-MOS-LQO are calculated as described in [b-ETSI ES 202 396-3].

8.4.39.3 Performance level classification based on values of this parameter

Background noises for driving speed ≤ 80 km/h

To claim compliance with a certain performance class the listening speech quality with background noise in send has to meet the requirements for the applicable performance classes given in the table below.

Table 8-93 – Limits for the listening speech quality with background noise in send: N-MOS

Performance Class	N-MOS-LQOn	N-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Table 8-94 – Limits for the listening speech quality with background noise in send: S-MOS

Performance Class	S-MOS-LQOn	S-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Table 8-95 – Limits for the listening speech quality with background noise in send: G-MOS

Performance Class	G-MOS-LQOn	G-MOS-LQOw
1	> 4.0	> 4.0
2	> 3.0	> 3.0
3	> 2.5	> 2.5
4	≤ 2.5	≤ 2.5

Background noises for driving speed ≤ 130 km/h

To claim compliance with a certain performance class the listening speech quality with background noise in send has to meet the requirements for the applicable performance classes given in the table below.

Table 8-96 – Limits for the listening speech quality with background noise in send: N-MOS

Performance Class	N-MOS-LQOn	N-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

Table 8-97 – Limits for the listening speech quality with background noise in send: S-MOS

Performance Class	S-MOS-LQOn	S-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

Table 8-98 – Limits for the listening speech quality with background noise in send: G-MOS

Performance Class	G-MOS-LQOn	G-MOS-LQOw
1	> 3.5	> 3.5
2	> 2.5	> 2.5
3	> 2.0	> 2.0
4	≤ 2.0	≤ 2.0

NOTE – It is recognized that high MOS scores measured at the subsystem level may not necessarily lead to the best overall system performance. For some codecs or other signal processing in the overall system a degraded signal at the subsystem level may lead to better overall performance.

8.4.39.4 Design guidance and root-cause analysis

In order to further investigate the nature of impairments introduced on the background noise the background noise distortion measure as described in Appendix I might be used. This method can be used in addition to measurement of the enhanced speech quality in the presence of background noise for investigating the root causes of background noise impairments in more detail and optimizing noise reduction algorithms. The objective noise distortion measure found in Appendix I measures objectively the background noise distortion based only on the input background noise and the filtered background noise after the noise reduction algorithm. Such a noise distortion measure can be applied to help optimize the parameterization of the noise reduction algorithm under test in the sense of background noise quality.

In general, the setting of the parameters should be optimized to allow a good performance at different driving speeds and with different background noises. Other optimization criteria are for further study.

8.4.40 Enhanced quality of background noise transmission with far-end speech (QBGNTFSE)

8.4.40.1 Parameter description

This parameter measures the variation in noise level at the output of the signal enhancement subsystem in the send direction due to the presence/absence of speech in the receive direction.

The enhanced quality of background noise transmission with far-end speech is measured from S2 to S3. The receive signal is inserted at R3.

8.4.40.2 Test

1) The test arrangement is given in clause 8.4.1.
2) All of the background noise conditions defined by the user scenarios in Annex B shall be tested.
3) First the measurement is conducted without inserting the signal at the far end (R3). At least 10 s of noise are recorded. The background signal level versus time is calculated using a time constant of 35 ms. This is the reference signal.
4) In a second step the same measurement is conducted but with the insertion of the CS-Signal at the far end (R3). The exactly identical background noise signal is applied. The background noise signal must start at the same point in time which was used for the measurement without the far-end signal. The background noise should be applied for at least 5 seconds in order to allow adaptation of the noise reduction algorithms. After at least 5 seconds a composite source signal according to [ITU-T P.501] is applied in the receive direction with a duration of ≥ 2 CSS periods. The test signal level is the nominal signal level.
5) The send signal is recorded at test point S3. The test signal level versus time is calculated using a time constant of 35 ms.
6) The level variation in the send direction is determined during the time interval when the CS-Signal is applied and after it stops. The level difference is determined from the difference of the recorded signal levels vs. time between reference signal and the signal measured with far-end signal.

8.4.40.3 Performance level classification based on values of this parameter

The test is carried out by applying the composite source signal in the receive direction. During and after the end of composite source signal bursts (representing the end of far-end speech simulation), the signal level in the send direction should not vary by more than 10 dB (during transition to transmission of background noise without far-end speech).

To claim compliance with a certain performance class of this parameter the signal enhancement has to meet the requirements for the applicable performance classes given in the table below for all background noises as specified in Annex B.

Table 8-99 – Limits for the enhanced quality of background noise transmission with far-end speech

Performance Class	QBGNTFSE
1	≤ 3 dB
2	≤ 10 dB
3	≤ 15 dB
4	> 15 dB

8.4.40.4 Design guidance and root-cause analysis

The background noise from the near end should be transmitted continuously without interruption or temporal attenuations in all conversational situations. Interruptions may be caused by the NLP of the echo canceller or other signal processing. Optimized parameter settings might help to improve the situation.

8.4.41 Comfort noise injection after enhancement (CNIE)

8.4.41.1 Parameter description

This set of parameters measures the difference in noise spectrum due to comfort noise injection at the output of the signal enhancement subsystem. The measurement is intended to avoid audible noise bursts which might occur at the beginning of a call, e.g., in the adaptation phase of noise cancellers.

The comfort noise injection after enhancement is measured at S3.

8.4.41.2 Test

The test arrangement is in accordance with clause 8.4.1.

Background noise is played back. All noise conditions defined by the user scenarios in Annex B shall be tested.

The test signal is applied in the receive direction at R3 consisting of an initial pause of 10 s and a periodical repetition of the composite source signal in the receive direction (duration ≥ 10 s) with nominal level to enable comfort noise injection.

The transmitted signal is recorded in the send direction at test point S3.

The power density spectrum measured in the send direction during the initial pause of the test signal (8 k FFT/48 kHz sampling rate or equivalent, averaged over ≥ 5 s) is referred to the power density spectrum determined during the period with the periodical repetition of the composite source signals in the receive direction (8 k FFT/48 kHz sampling rate or equivalent, averaged over ≥ 5 s). Spectral differences between both power density spectra are analysed and compared.

The level of the transmitted signal in the send direction is determined during the initial pause of the test signal in the receive direction and referred to the level of the transmitted signal in the send direction determined during the application of the test signal in the receive direction. Both levels are calculated using A-weighting.

8.4.41.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the comfort noise injection after enhancement has to meet the requirements for the applicable performance classes given in the tables below.

Table 8-100 – Spectral Limits for the comfort noise injection after enhancement

Frequency [Hz]	Performance Class 1	Performance Class 1	Performance Class 2	Performance Class 2	Performance Class 3	Performance Class 3	Performance Class 4	Performance Class 4
	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit
200	3 dB	–3 dB	6 dB	–6 dB	12 dB	–12 dB	N/A	N/A
800	3 dB	–3 dB	6 dB	–6 dB	12 dB	–12 dB	N/A	N/A
800	2 dB	–2 dB	5 dB	–5 dB	10 dB	–10 dB	N/A	N/A
2 000	2 dB	–2 dB	5 dB	–5 dB	10 dB	–10 dB	N/A	N/A
2 000	2 dB	–2 dB	3 dB	–3 dB	6 dB	–6 dB	N/A	N/A
4 000	2 dB	–2 dB	3 dB	–3 dB	6 dB	–6 dB	N/A	N/A
6 000*	2 dB	–2 dB	3 dB	–3 dB	6 dB	–6 dB	N/A	N/A

NOTE 1 – All sensitivity values are expressed in dB on an arbitrary scale.

NOTE 2 – The limit at intermediate frequencies lies on a straight line drawn between the given values on a log (frequency) – linear (dB) scale.

NOTE 3 – *wideband only

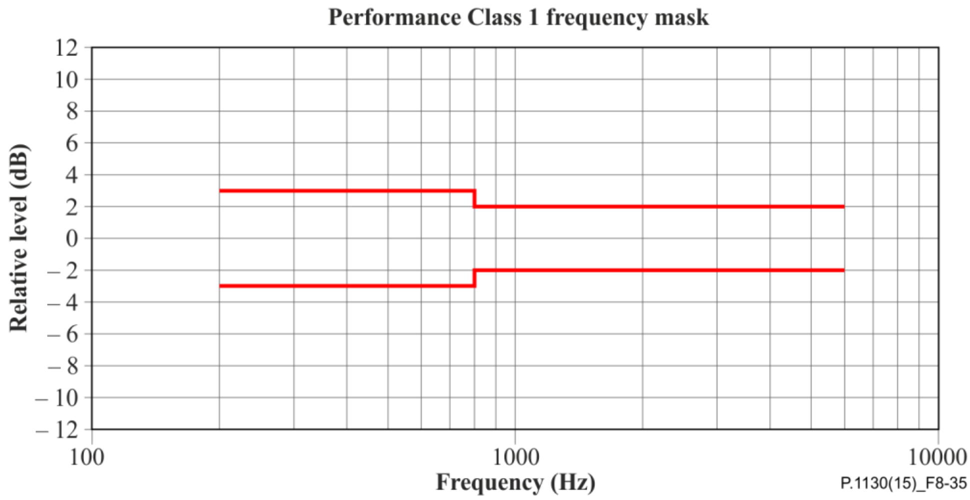


Figure 8-35 – Spectral limits for the comfort noise injection after enhancement for Performance Class 1 (Figure is informative)

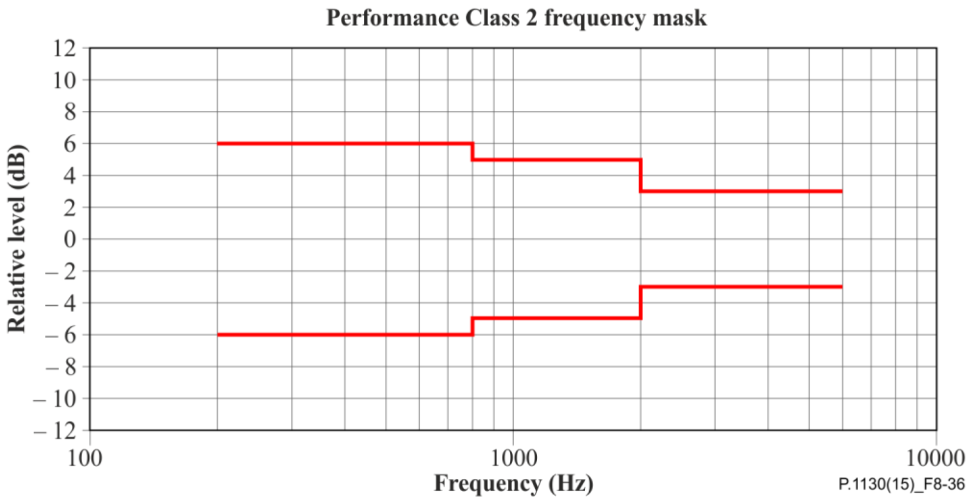


Figure 8-36 – Spectral limits for the comfort noise injection after enhancement for Performance Class 2 (Figure is informative)

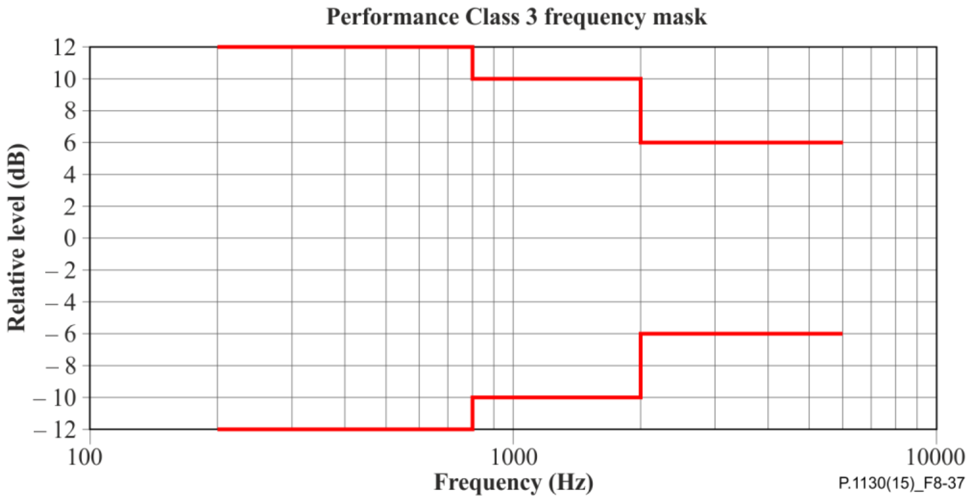


Figure 8-37 – Spectral limits for the comfort noise injection after enhancement for Performance Class 3 (Figure is informative)

Table 8-101 – Limits for the comfort noise injection after enhancement

Performance Class	CNIE upper limit	CNIE lower limit
1	1 dB	–3 dB
2	2 dB	–5 dB
3	5 dB	–10 dB
4	> 5 dB	< 10 dB

8.4.41.4 Design guidance and root-cause analysis

It is desirable that any comfort noise if used- matches the original background noise before and after periods of comfort noise injection as closely as possible. This test focuses only on level and spectral content of the background noise which are considered the most important parameters of any comfort noise insertion. The parameter settings of the comfort noise insertion which is typically attributed to the echo canceller should be adjusted accordingly for all types of background noises as listed in Annex B.

8.4.42 Speech recognition accuracy indicator after enhancement (SRAIE)

8.4.42.1 Parameter description

This parameter is for further study. It predicts speech recognition (SR) accuracy at the output of the signal enhancement subsystem. Some approaches which have been proposed include signal-based (e.g., weighted SNR, AI, etc.) and reference speech recognizer-based (e.g., % correct, % confusions, etc.) methods.

9 Bidirectional signal transport including network transport

The purpose of this section is to ensure linearity (as far as possible) and time invariance of a bidirectional signal transport over a wired or a short range wireless transmission system. Any bidirectional signal transport system including the network transport system in conjunction with the devices it is integrated in should not introduce any additional signal processing. Any other function than pure speech data transmission should be avoided.

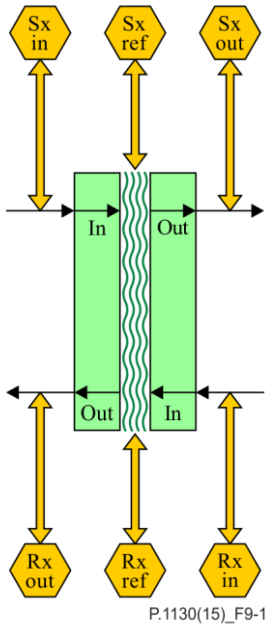


Figure 9-1 – Bidirectional transport subsystem

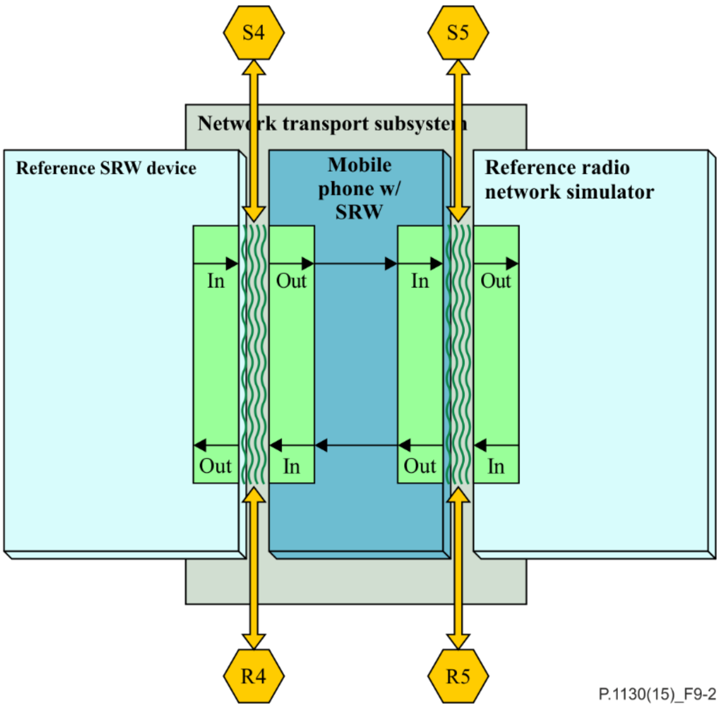


Figure 9-2 – Network transport subsystem with reference interfaces attached

Figure 9-1 shows the general bidirectional transport subsystem. The general concept of testing is as follows: Any transport system consists of a transmitter, a transport medium and a receiver. The transport system under consideration can consist of one of the following types:

Transmitter, Medium & Receiver (Sxin->Sxout, Rxin->Rxout)

(see clause 8.2)

Transmitter & Medium (Sxin->Sxref, Rxin->Rxref)

In this case the receiver at the send side is part of the transport system under test and the transmitter is part of the reference interface of the test equipment used to access Sxref.

Medium & Receiver (Sxref->Sxout, Rxref->Rxout)

In this case the transmitter at the receive side is part of the transport system under test and the receiver is part of the reference interface of the test equipment used to access Rxref.

Medium only (Sxref, Rxref)

In case of measuring the medium only reference transmitters and reference receivers are used.

Figure 9-2 shows the network transport subsystem including the short range wireless transmission SRW. Typically the measurements for the network transport are made from S4 (Sxin) to S5 (Sxout) and from R5 (Rxin) to R4 Rxout). Reference interfaces are used to connect the device to a simulated radio network (radio network simulator) and the SRW (SRW reference interface). In this case S4 and R4 define the SRWAP, S5 and R5 define the network access point POI (input and output of the reference speech coder of the radio network simulator).

The DUT is connected to the reference interfaces. The reference interfaces are calibrated as specified in clause 8.1.

9.1 Bidirectional transport delay in send direction

9.1.1 Parameter description

The delay in send direction T_{BDS} is measured from test point (Sxin) to test point (Sxout) or to the POI (input of the reference speech coder of the system simulators) test point (Sxout). The delay T_{BDS} should be minimized.

The system delay t_{System} depends on the transmission method used, the delay of the reference interface and the network simulator. The delay t_{System} must be known and deducted from the test result.

9.1.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point (Sxin) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 9.1.

2) The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{BDS} + T_{System}$$

9.1.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the bidirectional transport delay T_{BDS} defined in Table 9-1.

Table 9-1 – Limits for the bidirectional transport delay in send direction for speech communication services

Performance Class	T_{BDS}
1	< 2 ms
2	< 10 ms
3	< 30 ms
4	≥ 30 ms

NOTE – It is well understood that current technologies used for short range wireless transmission do not fulfil class 1 and 2 requirements. Delays observed with such systems are in the range of 8-10 ms in narrowband and 20-30 ms in wideband.

9.1.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay leads to higher impairment resulting even from low level echo components (see [ITU-T G.13]). Therefore, any design providing as low a delay in the connection as possible is preferable.

9.2 Clock drift in send direction

9.2.1 Parameter description

Due to unsynchronized clocks between subsystems clock drift might occur. The clock drift in the send direction is measured from test point (Sxin) to test point (Sxout). The clock drift should be minimized.

9.2.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. A sequence of CS signals (active signal length = 250 ms) is repeated for 120 s in order to analyse clock drift and any other time-variant delay. The pause length between two CS bursts is 100 ms and 1.2 s after every fourth burst in order to simulate a speech pause, which may lead to buffer adjustments. The test signal is inserted at test point (Sxin) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 9.1.

2) A cross correlation analysis versus time is carried out over the whole 120 s sequence between the received and the original test signal. The duration of the measurement (120 s) is indicated on the x-axis, the result of the cross correlation analysis (delay) is plotted on the y-axis.

3) The resulting clock drift within an analysis time range of at least 60 s is calculated as follows:

$$clock\ drift\,[ppm] = \frac{delay\ drift\,[s]}{analysis\ duration\,[s]} \cdot 1 \cdot 10^{6}$$

9.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the clock drift in the send direction has to meet the requirements for the applicable performance classes for the bidirectional clock defined in Table 9-2.

Table 9-2 – Limits for the bidirectional transport clock drift in send direction for speech communication services

Performance Class	Clock drift in ppm
1	0
2	< 10
3	< 50
4	≥ 50

9.2.4 Design guidance and root-cause analysis

No or at least low clock drift is essential for a seamless conversational performance especially for highly interactive conversations. Unsynchronized clocks and resulting delay drift lead to additional buffering increasing the overall delay and to potential loss of speech frames impacting the listening speech quality. Therefore, all synchronized designs are preferable.

9.3 Bidirectional transport delay in receive direction

9.3.1 Parameter description

The delay in receive direction T_{BDR} is measured from test point (Rxin) to test point (Rxout).

The system delay t_{System} depends on the transmission system, the delay of the reference interface and on the network simulator used. The delay t_{System} must be known and deducted from the test result.

9.3.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is applied at test point (Rxin) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

2) The test arrangement is according to clause 9.1.

3) The delay is determined by cross-correlation analysis between the measured signal at the SRWR and the original signal. The measurement is corrected by delays which are caused by the test equipment.

4) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{BDR} + T_{System}$$

9.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the receive direction has to meet the requirements for the applicable performance classes for the bidirectional transport delay T_{BDR} defined in Table 9-3.

Table 9-3 – Limits for the bidirectional transport delay in receive direction for speech communication services

Performance Class	T_{BDR}
1	< 2 ms
2	< 10 ms
3	< 30 ms
4	≥ 30 ms

NOTE – It is well understood that current technologies used for short range wireless transmission do not fulfil class 1 and 2 requirements. Delays observed with such systems are in the range of 8-10 ms in narrowband and 20-30 ms in wideband.

9.3.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay leads to higher impairment resulting even from low level echo components (see [ITU-T G.131]). Therefore, any design providing as low a delay in the connection as possible is preferable.

9.4 Clock drift in receive direction

9.4.1 Parameter description

Due to unsynchronized clocks between subsystems clock drift might occur. The clock drift in the receive direction is measured from test point (Rxin) to test point (Rxout). The clock drift should be minimized.

9.4.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. A sequence of CS signals (active signal length = 250 ms) is repeated for 120 s in order to analyse clock drift and any other time-variant delay. The pause length between two CS bursts is 100 ms and 1.2 s after every fourth burst in order to simulate a speech pause, which may lead to buffer adjustments. (Rxin) at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 9.1.

2) A cross correlation analysis versus time is carried out over the whole 120 s sequence between the received and the original test signal. The duration of the measurement (120 s) is indicated on the x-axis, the result of the cross correlation analysis (delay) is plotted on the y-axis.

3) The resulting clock drift within an analysis time range of at least 60 s is calculated as follows:

$$clock\ drift\ [ppm] = \frac{delay\ drift\ [s]}{analysis\ duration\ [s]} \cdot 1 \cdot 10^6$$

9.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the clock drift in the receive direction has to meet the requirements for the applicable performance classes for the bidirectional clock defined in Table 9-4.

Table 9-4 – Limits for the bidirectional transport clock drift in receive direction for speech communication services

Performance Class	Clock drift in ppm
1	0
2	< 10
3	< 50
4	≥ 50

9.4.4 Design guidance and root-cause analysis

No, or at least low, clock drift is essential for a seamless conversational performance especially for highly interactive conversations. Unsynchronized clocks and resulting delay drift lead to additional buffering increasing the overall delay and to potential loss of speech frames impacting the listening speech quality. Therefore, all synchronized designs are preferable.

9.5 Bidirectional transport send junction loudness ratings

9.5.1 Parameter description

The send junction loudness rating JLR_{BDS} is measured from test point (Sxin) to test point (Sxout).

The JLR describes any amplification or attenuation on the network transport layer. The send junction loudness rating JLR_{BDS} should be 0 dB.

9.5.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signal level is the nominal signal level; the level is averaged over the complete test signal.

The measured power density spectrum at test point interface (Sxin)is used as the reference power-density spectrum for determining the bidirectional transport send sensitivity.

2) The test arrangement is according to clause 9.1. For wideband the bidirectional send sensitivity is calculated from each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1-20. For narrowband bands 4-17 are used for calculation.

For the calculation, the average measured level at the electrical reference point for each frequency band is referred to the average test signal level measured in each frequency band at the reference interface.

3) The sensitivity is expressed in dBV/V, the junction loudness rating JLR_{BDS} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors W_J for JLR according to Table A.2 of [ITU-T P.79]. For narrowband bands 4-17 are used for the calculation.

9.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR in send has to meet the requirements for the applicable performance classes for the bidirectional transport JLR_{BDS} defined in Table 9-5.

Table 9-5 – Limits for the bidirectional transport JLR in send

Performance Class	JLR_{BDS}
1	0 ± 0.5 dB
2	0 ± 3 dB
3	0 ± 6 dB
4	> ± 6 dB

9.5.4 Design guidance and root-cause analysis

Any deviation of the JLR_{BDS} from 0 dB may result either in insufficient or too high speech levels and thus may deteriorate the performance of the interconnected devices. and should be avoided.

However, it is recognized that different SRW implementations deviate significantly (±6 dB) from this level. In order to maintain a good system performance, appropriate corrections have to be integrated in the hands-free implementation to ensure a good system performance. Based on these tests, appropriate level corrections can be inserted.

9.6 Bidirectional transport receive loudness ratings

9.6.1 Parameter description

The receive junction loudness rating JLR_{BDR} is measured from test point (Rxin) to test point (Rxout).

The JLR describes any amplification or attenuation on the network transport layer. The send junction loudness rating JLR_{BDR} should be 0 dB.

9.6.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signal is the nominal signal level.

2) The test arrangement is according to clause 9.1. For the calculation, the averaged level at test point (Rxin) interface is used. In wideband the test point receive sensitivity is determined by the bands 1-20 according to Table A.2 of [ITU-T P.79]. For narrowband systems bands 4-17 are used.

For the calculation, the average signal level of each frequency band is referred to the signal level of the reference signal measured in each frequency band.

3) The sensitivity is expressed in dBV/V, the junction loudness rating in receive JLR_{BDRv} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors W_J for JLR according to Table A.2 of [ITU-T P.79]. For narrowband systems bands 4-17 are used for the calculation.

9.6.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR in receive has to meet the requirements for the applicable performance classes for the bidirectional transport JLR_{BDR} defined in Table 9-6.

Table 9-6 – Limits for the bidirectional transport JLR in receive

Performance Class	JLR_{BDR}
1	0 ± 0.5 dB
2	0 ± 3 dB
3	0 ± 6 dB
4	> ± 6 dB

9.6.4 Design guidance and root-cause analysis

Any deviation of the JLR_{BDR} from 0 dB may result either in insufficient or too high speech levels and thus may deteriorate the performance of the interconnected devices. and should be avoided.

However, it is recognized that different SRW implementations deviate significantly (±6 dB) from this level. In order to maintain a good system performance, appropriate corrections have to be integrated in the hands-free implementation to ensure a good system performance. Based on these tests, appropriate level corrections can be inserted.

9.7 Bidirectional transport linearity in send direction

9.7.1 Parameter description

The linearity of the send junction loudness rating JLR_{BDS} is measured from test point (Sxin) to test point (Sxout).

The test is aimed to detect any amplitude non-linearities including AGC or companding.

9.7.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signals are in the range of –40 dBV to 5 dBV in steps of 5 dB relative to the nominal signal level, measured at test point (Sxin). The test signal level is the average level of the complete test signal.

   The measured power density spectrum at test point (Sxin) is used as the reference power-density spectrum for determining the send sensitivity.

2) The test arrangement is according to clause 9.1. In wideband test point send sensitivity is calculated from each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1-20. For narrowband systems bands 4-17 are used.

   For the calculation, the average measured level at the electrical reference point for each frequency band is referred to the average test signal level measured in each frequency band at test point (Sxin) reference interface.

3) In wideband the sensitivity is expressed in dBV/V, the bidirectional junction loudness rating JLR_{BDS} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors in the send direction according to Table A.2 of [ITU-T P.79]. For narrowband systems bands 4-17 are used for the calculation.

9.7.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR in send has to meet the requirements for the applicable performance classes defined in Table 9-7. The bidirectional transport JLR_{BDS} deviation is measured relative to the measured JLR_{BDSnom} measured with nominal signal level.

Table 9-7 – Limits for the bidirectional transport linearity in the send direction

Performance Class	Input signal level range	JLR_{BDS} re JLR_{BDSnom}
1	–40 dB/+5 dB	0 ± 0.5 dB
2	–40 dB/+5 dB	0 ± 3 dB
3	–40 dB/+5 dB	0 ± 6 dB
4	–40 dB/+5 dB	> ± 6 dB

9.7.4 Design guidance and root-cause analysis

Any AGC or companding on the network transport should be avoided. Additional non-linearities of this type may counteract algorithms implemented in the signal processing subsystem and deteriorate its performance.

9.8 Bidirectional transport linearity in receive direction

9.8.1 Parameter description

The linearity of receive junction loudness rating JLR_{BDR} is measured from test point (Rxin) to test point (Rxout).

The test is aimed to detect any amplitude non-linearities including AGC or companding.

9.8.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signals are in the range of –40 dBV to 5 dBV in steps of 5 dB relative to the nominal signal level measured at test point (Rxin). The test signal level is the average level of the complete test signal.

   The measured power density spectrum at test point (Rxin) is used as the reference power-density spectrum for determining the receive sensitivity.

2) In wideband the test arrangement is according to clause 9.1. The test point receive sensitivity is calculated from each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1-20. For narrowband bands 4-17 are used.

   For the calculation, the average measured level at the test point for each frequency band is referred to the average test signal level measured in each frequency band at the electrical reference interface.

3) The sensitivity is expressed in dBV/V, the bidirectional transport junction loudness rating JLR_{BDS} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors in the send direction according to Table A.2 of [ITU-T P.79]. For narrowband bands 4-17 are used for the calculation.

9.8.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR in receive has to meet the requirements for the applicable performance classes defined in Table 9-8. The bidirectional transport JLR_{BDR} deviation is measured relative to the measured JLR_{BDRnom} measured with nominal signal level:

Table 9-8 – Limits for the bidirectional transport linearity in the receive direction

Performance Class	Input signal level range	JLR_{BDR} re JLR_{BDRnom}
1	–40 dB/+5 dB	0 ± 0.5 dB
2	–40 dB/+5 dB	0 ± 3 dB
3	–40 dB/+5 dB	0 ± 6 dB
4	–40 dB/+5 dB	> ± 6 dB

9.8.4 Design guidance and root-cause analysis

Any AGC or companding on the network transport should be avoided. Additional non-linearities of this type may counteract algorithms implemented in the signal processing subsystem and deteriorate it's performance.

9.9 Bidirectional transport send sensitivity frequency response

9.9.1 Parameter description

The send frequency response FR_{BDS} is measured from test point (Sxin) to test point (Sxout). The send sensitivity response on the network transport should be mostly flat in the entire frequency range in order not to interfere with any (wanted) response characteristic in the signal processing subsystem or the acoustical frontend.

9.9.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The test signal level is the nominal signal level, the level is averaged over the complete test signal.

   The measured power density spectrum at test point (Sxin) interface is used as the reference power density spectrum for determining the bidirectional transport send sensitivity.

2) The test arrangement is according to clause 9.1. In wideband the bidirectional send sensitivity is determined in third octave intervals, as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz, inclusive. In narrowband it is determined for frequencies from 200 Hz to 4 kHz. In each third octave band, the level of the measured signal is referred to the level of the reference signal averaged over the complete test sequence length.

3) The sensitivity is determined in dBV/V.

9.9.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response in send has to meet the requirements for the applicable performance classes for the bidirectional transport FR_{BDS} as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 9-9 to 9-16 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 9-9 – Tolerance mask for the bidirectional transport wideband send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
6 200	0.5	–0.5
7 000	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-10 – Tolerance mask for the bidirectional transport wideband send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
6 200	2	–2
7 000	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-11 – Tolerance mask for the bidirectional transport wideband send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
6 200	4	–4
7 000	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-12 – Tolerance mask for the bidirectional transport wideband send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	NA	NA
6 200	NA	NA
7 000	NA	NA
NOTE – There is no frequency tolerance mask requirement for class 4.		

Table 9-13 – Tolerance mask for the bidirectional transport narrowband send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
3 100	0.5	–0.5
3 500	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-14 – Tolerance mask for the bidirectional transport narrowband send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
3 100	2	–2
3 500	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-15 – Tolerance mask for the bidirectional transport narrowband send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
3 100	4	–4
3 500	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-16 – Tolerance mask for the bidirectional transport narrowband send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	NA	NA
3 100	NA	NA
3 500	NA	NA
NOTE – There is no frequency tolerance mask requirement for class 4.		

9.9.4 Design guidance and root-cause analysis

Deviations from a flat frequency response characteristic may cause degradation of the listening speech quality in the send direction or may result in insufficient listening speech quality in the presence of background noise which affects the listening speech quality perceived by the far-end subscriber.

9.10 Bidirectional transport receive sensitivity frequency response

9.10.1 Parameter description

The receive frequency response FR_{BDR} is measured from test point (Rxin) to test point (Rxout).

9.10.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The test signal is the nominal signal level, measured at test point (Rxin) and averaged over the complete test signal sequence.
2) The test arrangement is according to clause 9.1. For wideband the bidirectional transport receive sensitivity is determined in third octave intervals as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz, inclusive. In narrowband it is determined for frequencies from 200 Hz to 4 kHz. In each third octave band, the level of the measured signal is referred to the level of the reference signal, averaged over the complete test sequence length.

3) The sensitivity is determined in dBV/V.

9.10.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response in send has to meet the requirements for the applicable performance classes for the bidirectional transport FR_{BDR} as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables 9-17 to 9-24 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 9-17 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
6 200	0.5	–0.5
7 000	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-18 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
6 200	2	–2
7 000	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-19 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
6 200	4	–4
7 000	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-20 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	NA	NA
6 200	NA	NA
7 000	NA	NA
NOTE – There is no frequency tolerance mask requirement for class 4.		

Table 9-21 – Tolerance mask for the bidirectional transport narrowband receive sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
3 100	0.5	–0.5
3 500	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-22 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
3 100	2	–2
3 500	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-23 – Tolerance mask for the bidirectional transport wideband receive sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
3 100	4	–4
3 500	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-24 – Tolerance mask for the bidirectional transport narrowband receive sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	NA	NA
3 100	NA	NA
3 500	NA	NA
NOTE – There is no frequency tolerance mask requirement for class 4.		

9.10.4 Design guidance and root-cause analysis

Deviations from a flat frequency response characteristics may cause degradation of the listening speech quality in the receive direction or may result in insufficient listening speech quality in the presence of background noise which affects the listening speech quality perceived by the subscriber in the car.

9.11 Bidirectional transport idle channel noise in send

9.11.1 Parameter description

The bidirectional transport idle channel noise in send N_{BDS} is measured from test point (Sxin) to test point (Sxout).

9.11.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with nominal signal level is applied at test point (Sxin). The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) For the noise measurement, no test signal is used. However, all sources which potentially contribute to noise at the output of the bidirectional transport as GSM noise, electrical noise introduced by the car must be considered. In order to ensure a reliable activation, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of 4 composite source signals according to [ITU-T P.501]. The level of the activation signal is –16 dBm0.

3) The test arrangement is according to clause 9.1.

   The idle channel noise is measured at the output in the frequency range between 100 Hz and 8 kHz. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers shall be taken into account; the time window must be shifted accordingly. The length of the time window is 1 second which is the averaging time for the idle channel noise. The test laboratory has to ensure the correct activation of the bidirectional transport during the measurement. If the bidirectional transport is deactivated during measurement, the measurement window has to be cut to the duration when the transport system remains activated.

   The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hann window is used.

   If it is known that the bidirectional transport stays activated without any activation signal; no activation signal is required. In this case, a simple noise measurement is conducted.

4) The idle channel noise is determined by A-weighting and referring to the reference speech signal level as determined with speech sequence.

5) Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3-rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^ (–1/6)*f* to 2^ (+1/6)*f*). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum up to 6.8 kHz.

9.11.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 9-25.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 9-25 – Limits for the idle channel noise re. reference speech signal level

Performance Class	SINR
1	> 70 dB
2	> 60 dB
3	> 40 dB
4	≤ 40 dB

9.11.4 Design guidance and root-cause analysis

Check shielding, connectors and cabling with respect to potentially induced interfering signals. Check the quality of the amplifiers used to insert signals and the preamplifiers used to receive signals. Verify that preamplifiers do not demodulate RF into NF.

9.12 Bidirectional transport idle channel noise in receive

9.12.1 Parameter description

The bidirectional transport idle channel noise in send N_{BDR} is measured from test point (Rxin) to test point (Rxout).

9.12.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with nominal signal level is applied at test point (Rxin). The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) For the noise measurement, no test signal is used. However, all sources which potentially contribute to noise at the output of the bidirectional transport as GSM noise, electrical noise introduced by the car must be considered. In order to ensure a reliable activation, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of 4 composite source signals according to [ITU-T P.501]. The level of the activation signal is –16 dBm0.

3) The test arrangement is according to clause 9.1.

   The idle channel noise is measured at the output in the frequency range between 100 Hz and 8 kHz. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers shall be taken into account; the time window must be shifted accordingly. The length of the time window is 1 second which is the averaging time for the idle channel noise. The test laboratory has to ensure the correct activation of the bidirectional transport during the measurement. If the bidirectional transport is deactivated during measurement, the measurement window has to be cut to the duration when the transport system remains activated.

   The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hann window is used.

   If it is known that the bidirectional transport stays activated without any activation signal, no activation signal is required. In this case, a simple noise measurement is conducted.

4) The idle channel noise is determined by A-weighting and referring to the reference speech signal level as determined with speech sequence.

5) Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth

of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3-rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^ (–1/6)*f* to 2^ (+1/6)*f*). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum up to 6.8 kHz.

9.12.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 9-26.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 9-26 – Limits for the idle channel noise re. reference speech signal level

Performance Class	SINR
1	> 70 dB
2	> 60 dB
3	> 40 dB
4	≤ 40 dB

9.12.4 Design guidance and root-cause analysis

Check shielding, connectors and cabling with respect to potentially induced interfering signals. Check the quality of the amplifiers used to insert signals and the preamplifiers used to receive signals. Verify that preamplifiers do not demodulate RF into NF.

9.13 Bidirectional transport noise cancellation test in send direction

9.13.1 Parameter description

The send noise cancellation NC_{BDS} is measured from test point (Sxin) to test point (Sxout).

The objective of this test is to check whether any noise cancellation is active in the send direction of the bidirectional transport.

9.13.2 Test

1) The test arrangement is according to clause 9.1.
2) The test signal is the first sentence "The last switch cannot be turned off" from the speech signals described in [ITU T P.501], clause 7.3. The test signal is mixed with a pink noise signal. The level of the pink noise signal is 20 dB below the nominal signal level. The pink noise signal starts 1 s before the speech signal and stops 2 s after the speech signal. The speech signal level is the nominal signal level, the level is averaged over the speech signal. The complete signal (speech plus pink noise) is filtered by the frequency response measured in clause 9.7.
3) The level of the transmitted pink noise signal is measured at test point (Sxout) after the speech signal stops. The measured signal level is referred to the pink noise signal level measured at test point (Sxin) after the speech signal stops and displayed vs time. The levels are calculated from the time domain using an integration time of 5 ms.
4) The attenuation vs time is determined for 1.5 s during the pink noise section.

9.13.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the NC in send has to meet the requirements for the applicable performance classes for the bidirectional transport NC_{BDS} defined in Table 9-27. The attenuation of the simulated background noise test signal (BGN) shall not deviate more than the given value in the relevant performance class for all periods of the test signal.

Table 9-27 – Tolerance for the bidirectional transport background noise attenuation

Performance Class	BGN attenuation
1	<± 1 dB
2	< ± 2 dB
3	< ± 6 dB
4	> ± 6 dB

9.13.4 Design guidance and root-cause analysis

Any activated (tandemed) noise cancellation would interfere with the noise cancellation in the signal processing subsystem and will lead to significant listening speech degradation in the send direction.

If a non-linear or time variant behaviour of the is observed, the tests, as described in [ITU-T P.1100], clause 11.13 (for narrowband systems) or [ITU-T P.1110], clause 11.13 (for wideband systems), can be applied to determine the behaviour of the bidirectional transport in more detail. Instead of inserting the tests signals acoustically, they have to be inserted electrically. Care should be taken since attenuation changes observed might be caused by AGC or companding.

9.14 Bidirectional transport one-way speech quality in send direction

9.14.1 Parameter description

The bidirectional transport listening speech quality in send LQ_{BDS} is measured from test point (Sxin) to test point (Sxout).

The test is intended to determine any impairment of the listening speech quality introduced by the bidirectional transport.

9.14.2 Test

The tests method to be used is [ITU-T P.863].

NOTE – [ITU-T P.863] does not provide a wideband mode. Wideband systems are evaluated on a superwideband scale. Therefore, the MOS requirements are given in MOS-LQOsw. See P.863 for more information.

1) The test signals used are the English test sequences as specified in [ITU-T P.501] (2 male speakers, 2 female speakers, two sentences each). The test signal level is the nominal signal level, measured at test point (Sxin), the test signal level is measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.

   The original speech signal is used as the reference signal for the determination of the speech quality.

2) The test arrangement is according to clause 9.1. For wideband systems MOS-LQOsw is determined. For narrowband systems MOS-LQOn is determined

   The calculation is made using the signal recorded at test point (Sxout).

9.14.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality in send has to meet the requirements for the applicable performance classes given in the table below.

Table 9-28 – Limits for the bidirectional transport listening speech quality in send

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

9.14.4 Design guidance and root-cause analysis

No additional impairment on listening speech quality should be introduced by the network transport. Any degradation of listening speech quality will deteriorate the overall listening speech quality and should be avoided. Therefore, the listening speech quality under quiet conditions is checked.

9.15 Bidirectional transport speech quality stability in send direction

9.15.1 Parameter description

The bidirectional transport listening speech quality stability in send LQS_{BDS} is measured from test point (Sxin) to test point (Sxout).

This test is intended to detect speech quality instabilities due to unreliable connection (speech frame loss...). This test should be repeated in different representative test configurations, e.g., positions of the phone within the vehicle, different SRW links active and different WIFI links active.

9.15.2 Test

Several measurements of MOS-LQOsw (for wideband systems) and MOS-LQOn (for narrowband systems) according to clause 9.14 are performed in series within the same call or for different calls, within the test arrangement defined below. Typically, for the same call, a measurement each 20 s insufficient. The results are reported in terms of statistics.

The procedure is identical for narrowband and wideband systems. In narrowband systems MOS-LQOn is used, in wideband systems MOS-LQOsw.

The assessment of listening speech quality stability is performed in 5 steps:

1) Measure the MOS-LQO periodically over the duration of one communication. **N** measurements provide **N** MOS-LQO values.

2) For each MOS-LQO (i) value i from 2 to N, MOS-LQO_GAP (i) is calculated as the absolute difference with the previous value MOS-LQO (i-1):

   MOS-LQO_GAP (i) = | MOS-LQO (i)- MOS-LQO (i-1)|

3) In order to take into account the subjective perception and measurement accuracy, MOS-LQO_GAP (i) is set to 0 when the difference is equal or lower to THRESHOLD1:

   – if MOS-LQO_GAP (i) > (2*THRESHOLD1), then MOS-LQO_GAP (i) = MOS-LQO_GAP (i)

   – if THRESHOLD1 < MOS-LQO_GAP (i) ≦ 2*THRESHOLD1, then MOS-LQO_GAP (i) = [MOS-LQO_GAP (i) * 2] – (2*THRESHOLD1)

– if MOS-LQO_GAP (i) ≦ THRESHOLD1, then MOS-LQO_GAP (i) = 0

4) The instability (INS_MOS-LQO) associated to the MOS-LQO over the whole N measurements is defined by mean value of MOS-LQO_GAP (i).

INS_MOS-LQO = 1/(N-1)Σ MOS-LQO_GAP (i) with i = [2:N]

5) A linear weighting function is applied in order to express stability ST-MOS-LQO on a 0 to 100 scale.

This formulation is used to determine the listening quality stability (ST-MOS) as

ST-MOS = 100-(250* INS_MOS)

and

ST-MOS = 0 if [100-(250* INS_MOS)] < 0

When ST-MOS is calculated within a single call, the call should be longer than 3 minutes (Recommended duration being between 3 and 5 minutes). The duration of each measurement depends on the length of the speech samples used for the test as described in clause 9.14.

(As an example, if a sample is 15s length and the analysis is done every 20 s a minimum of 10 values will be measured).

9.15.3 Performance level classification based on values of this parameter

For the stability indicator about listening speech quality, THRESHOLD1 = 0.1 and the linear weighting function applies in order to express stability (ST-MOS) on a 0 to 100 scale. By definition stability equals 100 when no variations occur and stability ST-MOS equals 0 when MOS-LQO variation is equal or more than 0.4.

To claim compliance with a certain performance class the listening speech quality stability in send has to meet the requirements for the applicable performance classes given in the table below.

Table 9-29 – Limits for the bidirectional transport listening speech quality stability in send

Performance Class	ST-MOSn	ST-MOSsw
1	> 95	> 95
2	> 80	> 80
3	> 50	> 50
4	≤ 50	≤ 50

9.15.4 Design guidance and root-cause analysis

Listening quality stability during a call (if the position or transmission characteristics change during the call (or during several different calls) takes into account degradations generated on the signal by the transmission link impairment and the phone position.

In the case of systems using SRW transmission, it is the purpose to verify the integration of an SRW radio network by evaluating change of speech quality over time. This will help to detect problems with RF coverage inside the car cabin and verify the error concealment (packet loss, bit errors) caused by weak RF link or interference with other radio.

A guidance to proceed with is:

1 Check RF coverage from an SRW unit to possible mobile positions (Protocol analyser).

2 Identify weak and bad reception areas inside the vehicle. (Protocol analyser)

Check bit error rate, rate of packet loss…

3 Use the speech quality measurement to rate the quality of error concealment in the weak areas identified in step 2 above and compare with measurements from areas with good coverage.

4 Identify possible issues from interference with parallel SRW links or other networks. This will check how the SRW can handle interference and change to undisturbed channels. Speech quality measurements can be used to see the performance of handling these problems.

It is recommended to use to a known link as reference.

9.16 Bidirectional transport one-way speech quality in receive direction

9.16.1 Parameter description

The bidirectional transport listening speech quality in receive LQ_{BDR} is measured from test point (Rxin) to test point (Rxout).

The test is intended to determine any impairment of the listening speech quality introduced by the bidirectional transport.

9.16.2 Test

The tests method to be used is [ITU-T P.863].

NOTE – [ITU-T P.863] does not provide a wideband mode. Wideband systems are evaluated on a superwideband scale. Therefore, the MOS requirements are given in MOS-LQOsw. See P.863 for more information.

1) The test signals used are the English test sequences as specified in [ITU-T P.501] (2 male speakers, 2 female speakers, two sentences each). The test signal level is the nominal signal level measured at test point (Rxin), alternatively (Rxin): the test signal level is measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.

   The original speech signal is used as the reference signal for the determination of the speech quality.

2) The test arrangement is according to clause 9.1. For wideband systems MOS-LQOsw is determined, for narrowband systems MOS-LQOn. The signal measured at test point (Rxout) is used for the calculation

9.16.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality in send has to meet the requirements for the applicable performance classes given in the table below.

Table 9-30 – Limits for the bidirectional transport listening speech quality in receive

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

9.16.4 Design guidance and root-cause analysis

No additional impairment on listening speech quality should be introduced by the network transport. Any degradation of listening speech quality will deteriorate the overall listening speech quality and should be avoided. Therefore, the listening speech quality under quiet conditions is checked.

9.17 Bidirectional transport speech quality stability in receive direction

9.17.1 Parameter description

The bidirectional transport listening speech quality in receive LQ_{BDR} is measured from test point (Rxin) to test point (Rxout).

This test is intended to detect speech quality instabilities due to unreliable SRW connection (speech frame loss...). This test should be repeated in different representative positions of the phone within the vehicle, different SRW links active and different WIFI links active.

9.17.2 Test

The test procedure is described in clause 9.15.2.

9.17.3 Performance level classification based on values of this parameter

For the stability indicator about listening speech quality, THRESHOLD1 = 0.1 and the linear weighting function applies in order to express stability (ST-MOS) on a 0 to 100 scale. By definition stability equals 100 when no variations occur and stability ST-MOS equals 0 when MOS-LQO variation is equal or more than 0.4.

To claim compliance with a certain performance class the listening speech quality stability in send has to meet the requirements for the applicable performance classes given in the table below.

Table 9-31 – Limits for the bidirectional transport listening speech quality stability in receive

Performance Class	ST-MOSn	ST-MOSsw
1	> 95	> 95
2	> 80	> 80
3	> 50	> 50
4	≤ 50	≤ 50

9.17.4 Design guidance and root-cause analysis

Listening quality stability during a call (if the position or transmission characteristics change during the call (or during several different calls) takes into account degradations generated on the signal by the transmission link impairment and the phone position.

In the case of systems using SRW transmission it is the purpose to verify the integration of an SRW radio network by evaluating change of speech quality over time. This will help to detect problems with RF coverage inside the car cabin and verify the error concealment (packet loss, bit errors) caused by weak RF link or interference with other radio.

A guidance to proceed with is:

1 Check RF coverage from an SRW unit to possible mobile positions (Protocol analyser).

2 Identify weak and bad reception areas inside the vehicle. (Protocol analyser)

Check bit error rate, rate of packet loss…

3 Use the speech quality measurement to rate the quality of error concealment in the weak areas identified in step 2 above and compare with measurements from areas with good coverage.

4 Identify possible issues from interference with parallel SRW links or other networks. This will check how the SRW can handle interference and change to undisturbed channels. Speech quality measurements can be used to see the performance of handling these problems.

It is recommended to use to a known link as reference.

9.18 Verification of bidirectional transport disabled echo control

9.18.1 Parameter description

The bidirectional transport echo control EC_{BD} is measured from test point (Rxin) or the POI (input of the reference speech coder of the system simulators) (Rxin) to test point (Sxout).

No acoustic echo control shall be active in the mobile phone.

9.18.2 Test

1) For the test, an artificial echo path is inserted at test point (Rxin), alternatively (Rxin). The artificial echo path consists of an attenuation of 20 dB/40 dB and a delay of 20 ms is inserted at test point (Sxin/Rxout) interface. The test set-up is shown in Figures 9-3 and 9-4.

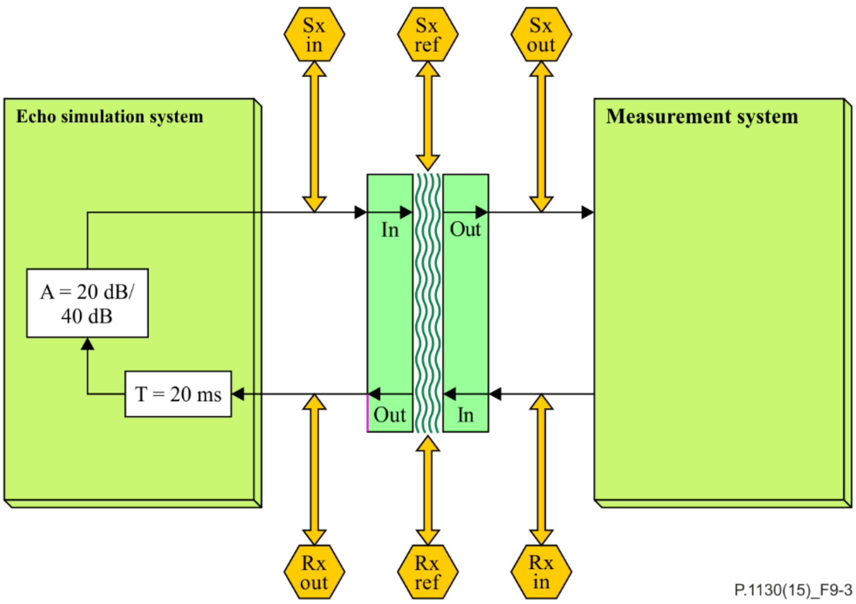


Figure 9-3 – Test set-up with artificial echo loss

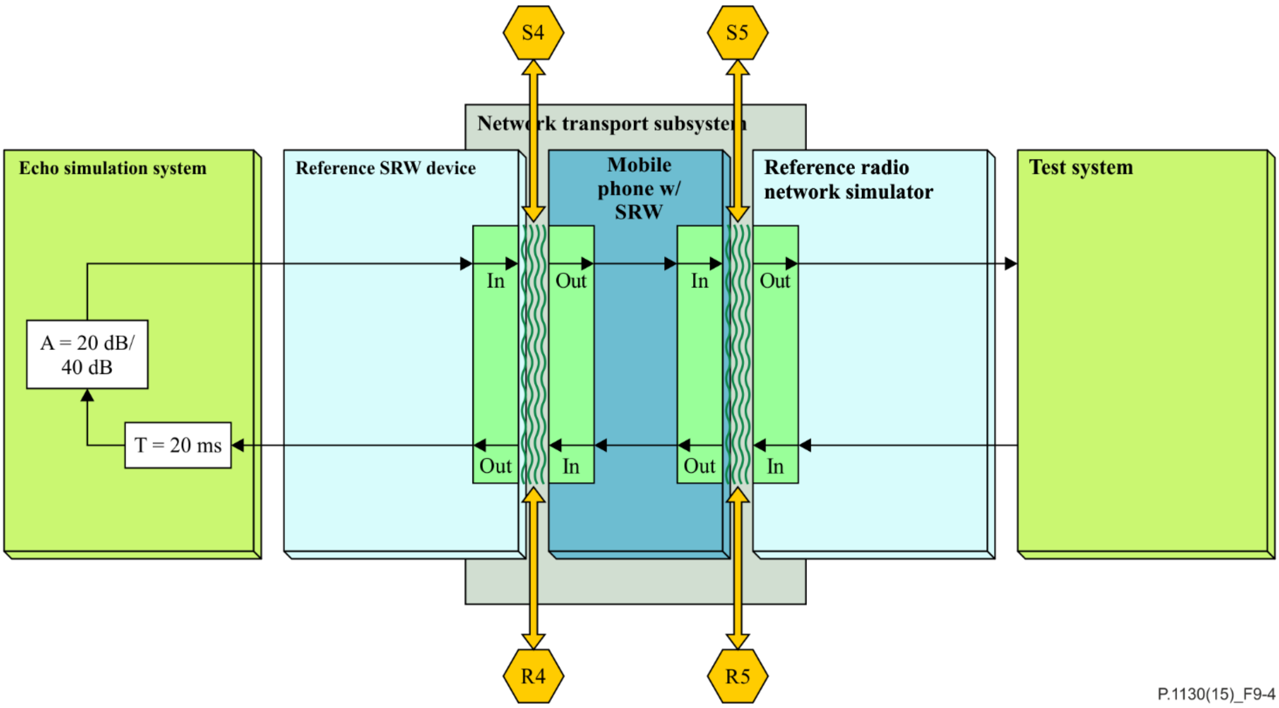


Figure 9-4 – Example of network transport set-up with artificial echo loss

2) The attenuation between the input of test point (Rxin) to the output of test point (Rxin) is measured using a speech-like test signal.

3) Before the actual measurement, a training sequence consisting of the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] is inserted. The training sequence level is the nominal signal level.

4) The test signal is a pn sequence, according to [ITU-T P.501], with a length of 4 096 points (48 kHz sampling rate) and a crest factor of 6 dB. The duration of the test signal is 250 ms, the test signal level is –6 dBov as applicable to the individual interface. The low crest factor is achieved by random alternation of the phase between –180° and +180°.

5) TCL_W is calculated according to clause B.4 of [ITU-T G.122] (trapezoidal pseudo-rule). For the calculation, the average measured echo level at each frequency band is referred to the average level of the test signal measured in each frequency band. For the measurement, a time window has to be applied which is adapted to the duration of the actual test signal (250 ms).

6) The difference between the echo loss measured with 20 dB echo loss and 40 dB echo loss is determined.

9.18.3 Performance level classification based on values of this parameter

The difference between the echo loss ELD_{BD} measured with 20 dB echo loss and 40 dB echo loss is measured. To claim compliance with a certain performance class the listening speech quality stability in send has to meet the requirements for the applicable performance classes given in the table below.

Table 9-32 – Limits for the bidirectional transport echo loss difference

Performance Class	ELD_{BDmin}	ELD_{BDmax}
1	18 dB	22 dB
2	FFS	FFS
3	FFS	FFS
4	≤ 18 dB	≥ 22 dB

9.18.4 Design guidance and root-cause analysis

Any additional echo cancellation would impair the performance of the acoustic echo cancellation integrated in the signal processing subsystem. As a result, additional impairments on speech quality and double talk performance will occur.

9.19 Bidirectional transport subsystem weighted terminal coupling loss ($TCLw_{BD}$) and (TCL_{BD})

9.19.1 Parameter description

This parameter measures acoustic echo loss of the bidirectional transport. Echo loss is measured from the input to the bidirectional transport subsystem in the receive direction (Rxin), to the output of the bidirectional transport subsystem in the send direction (Sxout). In narrowband TCLw is measured in wideband the unweighted TCL is measured.

The test is repeated by reversing the measurement direction and measure the echo loss from the input to the bidirectional transport subsystem in the send direction (Sxin), to the output of the bidirectional transport subsystem in the receive direction (Rxout).

9.19.2 Test

1) The test arrangement is in accordance with clause 8.3.1.
2) The noise level measured at the test point (idle channel noise) shall be less than –63 dBm0. The attenuation between the input to the bidirectional transport subsystem (Rxin) to the output of the bidirectional transport subsystem is measured using the compressed speech test signal as described in [ITU-T P.501], clause 7.3.3, Amendment 1. The test signal level is –10 dBm0.
3) The first 17.0 s of the test signal (6 sentences) are discarded from the analysis to allow for convergence of the acoustic echo canceller. The analysis is performed over the remaining length of the test sequence (last 6 sentences). For the calculation, the average measured echo level at each frequency band is referred to the average level of the test signal measured in each frequency band.
4) In narrowband, TCL_W is calculated according to clause B.4 of [ITU-T G.122], (trapezoidal rule). In wideband, the differences between the averaged echo level and the averaged test signal level in a frequency range from 100 Hz-8 000 Hz is calculated.
5) The test is repeated by reversing the measurement direction and measure the echo loss from the input to the bidirectional transport subsystem in the send direction (Sxin), to the output of the bidirectional transport subsystem in the receive direction (Rxout).

9.19.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the echo loss through the acoustic subsystem has to meet the requirements for the applicable performance classes for the TCL_{BD} (wideband) and $TCLw_{BD}$ defined in Table 9-33.

Table 9-33 – Limits for the bidirectional transport subsystem TCL

Performance Class	$TCLW_{BD(narrowband)}$	$TCL_{BD\ (wideband)}$
1	> 80 dB	> 80 dB
2	> 60 dB	> 60 dB
3	> 40 dB	> 40 dB
4	< 40 dB	< 40 dB

9.19.4 Design guidance and root-cause analysis

Any crosstalk or echo measured might result in poor performance of the overall system due to noticeable echo components.

10 Unidirectional signal transport: Wired and short range wireless transmission

The purpose of this section is to ensure linearity (as far as possible) and time invariance of the signal transport over wired or a short range wireless transmission system. Any unidirectional transport system in conjunction with the devices it is integrated in should not introduce any additional signal processing. Any other functions than pure speech data transmission should be avoided.

10.1 Test set-up

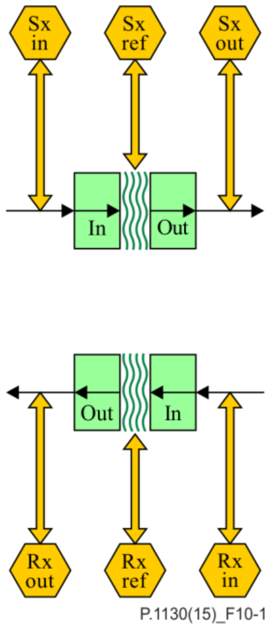


Figure 10-1 – Unidirectional transport subsystem

The DUT is connected to the reference interfaces. The reference interfaces are calibrated as specified in clause 8.1.

The general concept of testing is as follows: Any transport system consists of a transmitter, a transport medium and a receiver. The transport system under consideration can consist of one of the following types:

1 Transmitter, Medium & Receiver (Sxin->Sxout, Rxin->Rxout)

(see clause 8.2)

2 Transmitter & Medium (Sxin->Sxref, Rxin->Rxref)

In this case the receiver at the send side is part of the transport system under test and the transmitter is part of the reference interface of the test equipment used to access Sxref.

3 Medium & Receiver (Sxref->Sxout, Rxref->Rxout)

In this case the transmitter at the receive side is part of the transport system under test and the receiver is part of the reference interface of the test equipment used to access Rxref.

4 Medium only (Sxref, Rxref)

In case of measuring the medium only reference transmitters and reference receivers are used.

10.2 Unidirectional transport delay

10.2.1 Parameter description

The delay T_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout. The delay T_{UD} should be minimized.

The system delay t_{System} depends on the transmission method used, the delay of the reference interface and the network simulator. The delay t_{System} must be known and deducted from the test result.

10.2.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. The pseudorandom noise (pn)-part of the CSS has to be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples (with 48 kHz sampling rate). The test signal is inserted at test point Sxin/Rxin at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 9.1.

2) The delay is determined by cross-correlation analysis between the measured signal at the electrical access point and the original signal. The measurement is corrected by delays which are caused by the test equipment.

3) The delay is measured in ms and the maximum of the cross-correlation function is used for the determination.

$$T_{measure} = T_{UD} + T_{System}$$

10.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the delay in the send direction has to meet the requirements for the applicable performance classes for the bidirectional transport delay T_{UD} defined in Table 9-34.

Table 9-34 – Limits for the unidirectional transport delay for speech communication services

Performance Class	T_{UD}
1	< 2 ms
2	< 10 ms
3	< 30 ms
4	≥ 30 ms

10.2.4 Design guidance and root-cause analysis

Low delay is essential for a seamless conversational performance especially for highly interactive conversations [ITU-T G.114]. Delay also contributes to echo perception: Higher delay low leads to higher impairment resulting even from low level echo components (see [ITU-T G.131]). Therefore, any design providing as low a delay in the connection as possible is preferable.

10.3 Unidirectional clock drift

10.3.1 Parameter description

Due to unsynchronized clocks between subsystems clock drift might occur. The unidirectional clock drift is measured from test point Sxin/Rxin to test point Sxout/Rxout. The clock drift should be minimized.

10.3.2 Test

1) For the measurements, a composite source signal (CSS) according to [ITU-T P.501] is used. A sequence of CS signals (active signal length = 250 ms) is repeated for 120 s in order to analyse clock drift and any other time-variant delay. The pause length between two CS bursts is 100 ms and 1.2 s after every fourth burst in order to simulate a speech pause, which may lead to buffer adjustments. The test signal is inserted at test point Sxin/Rxin at a level according to clause 8.1.5.

   The reference signal is the original signal (test signal).

   The test set-up is according to clause 9.1.

2) A cross correlation analysis versus time is carried out over the whole 120 s sequence between the received and the original test signal. The duration of the measurement (120 s) is indicated on the x-axis, the result of the cross correlation analysis (delay) is plotted on the y-axis.

3) The resulting clock drift within an analysis time range of at least 60 s is calculated as follows:

$$clock\ drift\ [ppm] = \frac{delay\ drift\ [s]}{analysis\ duration\ [s]} \cdot 1 \cdot 10^6$$

10.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the clock drift in the send direction has to meet the requirements for the applicable performance classes for the bidirectional clock defined in Table 9-35.

Table 9-35 – Limits for the bidirectional transport clock drift in send direction for speech communication services

Performance Class	Clock drift in ppm
1	0
2	< 10
3	< 50
4	≥ 50

10.3.4 Design guidance and root-cause analysis

No or at least low clock drift is essential for a seamless conversational performance especially for highly interactive conversations. Unsynchronized clocks and resulting delay drift lead to additional buffering increasing the overall delay and to potential loss of speech frames impacting the listening speech quality. Therefore, all synchronized designs are preferable.

10.4 Unidirectional transport junction loudness ratings

10.4.1 Parameter description

The junction loudness rating JLR_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout.

The JLR describes any amplification or attenuation on the transport layer. The junction loudness rating JLR_{UD} should be 0 dB.

10.4.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signal level is the nominal signal level, the level is averaged over the complete test signal.

   The measured power density spectrum at test point interface Sxin/Rxin is used as the reference power-density spectrum for determining the unidirectional transport send sensitivity.

2) The test arrangement is according to clause 9.1. For wideband the unidirectional sensitivity is calculated from each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1-20. For narrowband bands 4-17 are used for calculation.

   For the calculation, the average measured level at the electrical reference point for each frequency band is referred to the average test signal level measured in each frequency band at the reference interface.

3) The sensitivity is expressed in dBV/V, the junction loudness rating JLR_{UD} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors W_J for JLR according to Table A.2 of [ITU-T P.79]. For narrowband bands 4-17 are used for the calculation.

10.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR has to meet the requirements for the applicable performance classes for the bidirectional transport JLR_{UD} defined in Table 9-36.

Table 9-36 – Limits for the unidirectional transport JLR

Performance Class	JLR_{UD}
1	0 ± 0.5 dB
2	0 ± 3 dB
3	0 ± 6 dB
4	> ± 6 dB

10.4.4 Design guidance and root-cause analysis

Any deviation of the JLR_{BDS} from 0 dB may result either in insufficient or too high speech levels and thus may deteriorate the performance of the interconnected devices and should be avoided.

However, it is recognized that different SRW implementations deviate significantly (±6 dB) from this level. In order to maintain a good system performance, appropriate corrections have to be integrated in the hands-free implementation to ensure a good system performance. Based on these tests, appropriate level corrections can be inserted.

10.5 Unidirectional transport linearity

10.5.1 Parameter description

The linearity of the junction loudness rating JLR_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout

The test is aimed to detect any amplitude non-linearities including AGC or companding.

10.5.2 Test

1) The test signal used for the measurements shall be CSS according to [ITU-T P.501]. The test signals are in the range of –40 dBV to 5 dBV in steps of 5 dB relative to the nominal signal level, measured at test point Sxin/Rxin. The test signal level is the average level of the complete test signal.

   The measured power density spectrum at test point Sxin/Rxin is used as the reference power-density spectrum for determining the send sensitivity.

2) The test arrangement is according to clause 9.1. In wideband the sensitivity is calculated from each band of the 20 frequencies given in Table A.2 of [ITU-T P.79], bands 1-20. For narrowband systems bands 4-17 are used.

   For the calculation, the average measured level at the electrical reference point for each frequency band is referred to the average test signal level measured in each frequency band at test point Sxin/Rxin reference interface.

3) In wideband the sensitivity is expressed in dBV/V, the bidirectional junction loudness rating JLR_{UD} shall be calculated according to [ITU-T P.79], Formula A-23d, bands 1-20, M = 0.175, and the weighting factors in the send direction according to Table A.2 of [ITU-T P.79]. For narrowband systems bands 4-17 are used for the calculation.

10.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the JLR has to meet the requirements for the applicable performance classes defined in Table 9-37. The bidirectional transport JLR_{UD} deviation is measured relative to the measured JLR_{UDnom} measured with nominal signal level:

Table 9-37 – Limits for the bidirectional transport linearity

Performance Class	Input signal level range	JLR_{UD} re JLR_{UDnom}
1	–40 dB/+5 dB	0 ± 0.5 dB
2	–40 dB/+5 dB	0 ± 3 dB
3	–40 dB/+5 dB	0 ± 6 dB
4	–40 dB/+5 dB	> ± 6 dB

10.5.4 Design guidance and root-cause analysis

Any AGC or companding on the network transport should be avoided. Additional non-linearities of this type may counteract algorithms implemented in the signal processing subsystem and deteriorate its performance.

10.6 Unidirectional transport sensitivity frequency response

10.6.1 Parameter description

The frequency response FR_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout.

The sensitivity response on the network transport should be mostly flat in the entire frequency range in order not to interfere with any (wanted) response characteristic in the signal processing subsystem or the acoustical frontend.

10.6.2 Test

1) The test signal used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The test signal level is the nominal signal level, the level is averaged over the complete test signal.

   The measured power density spectrum at test point Sxin/Rxin interface is used as the reference power density spectrum for determining the unidirectional transport send sensitivity.

2) The test arrangement is according to clause 9.1. In wideband the unidirectional send sensitivity is determined in third octave intervals, as given by [IEC 61260] for frequencies of 100 Hz and 8 kHz, inclusive. In narrowband it is determined for frequencies from 200 Hz to 4 kHz. In each third octave band, the level of the measured signal is referred to the level of the reference signal averaged over the complete test sequence length.

3) The sensitivity is determined in dBV/V.

10.6.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response has to meet the requirements for the applicable performance classes for the unidirectional transport FR_{UD} as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Table 9-38 to 9-45 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 9-38 – Tolerance mask for the unidirectional transport wideband sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
6 200	0.5	–0.5
7 000	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-39 – Tolerance mask for the unidirectional transport wideband sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
6 200	2	–2
7 000	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-40 – Tolerance mask for the unidirectional transport wideband sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
6 200	4	–4
7 000	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-41 – Tolerance mask for the unidirectional transport wideband sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	N/A	N/A
6 200	N/A	N/A
7 000	N/A	N/A
NOTE – There is no frequency tolerance mask requirement for class 4.		

Table 9-42 – Tolerance mask for the unidirectional transport narrowband sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
50	0.5	–0.5
3 100	0.5	–0.5
3 500	0.5	–3
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-43 – Tolerance mask for the unidirectional transport narrowband sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	2	–2
3 100	2	–2
3 500	2	–5
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-44 – Tolerance mask for the unidirectional transport narrowband sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	4	–4
3 100	4	–4
3 500	4	–9
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table 9-45 – Tolerance mask for the unidirectional transport narrowband sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	N/A	N/A
3 100	N/A	N/A
3 500	N/A	N/A
NOTE – There is no frequency tolerance mask requirement for class 4.		

10.6.4 Design guidance and root-cause analysis

Deviations from a flat frequency response characteristic may cause degradation of the listening speech quality in the send direction or may result in insufficient listening speech quality in the presence of background noise which affects the listening speech quality perceived by the far-end subscriber.

10.7 Unidirectional transport idle channel noise

10.7.1 Parameter description

The unidirectional transport idle channel noise in N_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout.

10.7.2 Test

1) The British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501] with nominal signal level is applied at test point Sxin/Rxin. The test signal level is the average level of the complete test signal. The output level is measured using a speech level voltmeter according to [ITU-T P.56]. This level is the reference speech signal level.

2) For the noise measurement, no test signal is used. However, all sources which potentially contribute to noise at the output of the unidirectional transport as GSM noise, electrical noise introduced by the car must be considered. In order to ensure a reliable activation, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of 4 composite source signals according to [ITU-T P.501]. The level of the activation signal is –16 dBm0.

3) The test arrangement is according to clause 9.1.

   The idle channel noise is measured at the output in the frequency range between 100 Hz and 8 kHz. The measurement requires a time window which starts exactly at the end of the activation signal. Any "ringing" of filters or receivers shall be taken into account; the time window must be shifted accordingly. The length of the time window is 1 second which is the averaging time for the idle channel noise. The test laboratory has to ensure the correct activation of the unidirectional transport during the measurement. If the unidirectional transport is deactivated during measurement, the measurement window has to be cut to the duration when the transport system remains activated.

The power density spectrum of the noise signal is determined using FFT (8 k samples/48 kHz sampling rate or equivalent). A Hann window is used.

If it is known that the unidirectional transport stays activated without any activation signal, no activation signal is required. In this case, a simple noise measurement is conducted.

4) The idle channel noise is determined by A-weighting and referring to the reference speech signal level as determined with speech sequence.

5) Spectral peaks are measured in the frequency domain. The frequency spectrum of the A-weighted idle channel noise is measured by a spectral analysis having a noise bandwidth of 8.79 Hz (determined using FFT 8 k samples/48 kHz sampling rate with Hanning window or equivalent). The idle channel noise spectrum is stated in dB. A smoothed average idle channel noise spectrum is calculated by a moving average (arithmetic mean) 1/3rd octave wide across the idle noise channel spectrum stated in dB (linear average in dB of all FFT bins in the range from 2^ (–1/6)*f* to 2^ (+1/6)*f*). Peaks in the idle channel noise spectrum are compared against a smoothed average idle channel noise spectrum up to 6.8 kHz.

10.7.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the speech to idle noise ratio SINR has to meet the requirements for the applicable performance classes for the SINR defined in Table 9-46.

Spectral peaks in the frequency domain shall not exceed the averaged spectrum by more than 10 dB.

Table 9-46 – Limits for the idle channel noise re. reference speech signal level

Performance Class	SINR
1	> 70 dB
2	> 60 dB
3	> 40 dB
4	≤ 40 dB

10.7.4 Design guidance and root-cause analysis

Check shielding, connectors and cabling with respect to potentially induced interfering signals. Check the quality of the amplifiers used to insert signals and the preamplifiers used to receive signals. Verify that preamplifiers do not demodulate RF into NF.

10.8 Unidirectional transport noise cancellation test

10.8.1 Parameter description

The send noise cancellation NC_{UD} is measured from test point Sxin/Rxin to test point (xy).

The objective of this test is to check whether any noise cancellation is active in the unidirectional transport.

10.8.2 Test

1) The test arrangement is according to clause 9.1.

2) The test signal is the first sentence "The last switch cannot be turned off" from the speech signals described in [ITU T P.501], clause 7.3. The test signal is mixed with a pink noise signal. The level of the pink noise signal is 20 dB below the nominal signal level. The pink noise signal starts 1 s before the speech signal and stops 2 s after the speech signal. The speech signal level is the nominal signal level, the level is averaged over the speech signal. The complete signal (speech plus pink noise) is filtered by the frequency response measured in clause 10.5.

3) The level of the transmitted pink noise signal is measured at test point Sxout/Rxout after the speech signal stops. The measured signal level is referred to the pink noise signal level measured at test point Sxin/Rxin after the speech signal stops and displayed vs time. The levels are calculated from the time domain using an integration time of 5 ms.

4) The attenuation vs time is determined for 1.5 s during the pink noise section.

10.8.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the NC in send has to meet the requirements for the applicable performance classes for the unidirectional transport NC_{UD} defined in Table 9-47. The attenuation of the simulated background noise test signal (BGN) shall not deviate more than the given value in the relevant performance class for all periods of the test signal.

Table 9-47 – Tolerance for the unidirectional transport background noise attenuation

Performance Class	BGN attenuation
1	<± 1 dB
2	< ± 2 dB
3	< ± 6 dB

10.8.4 Design guidance and root-cause analysis

Any activated (tandemed) noise cancellation would interfere with the noise cancellation in the signal processing subsystem and will lead to significant listening speech degradation in the send direction.

If a non-linear or time variant behaviour of the is observed, the tests, as described in [ITU-T P.1100], clause 11.13 (for narrowband systems) or [ITU-T P.1110], clause 11.13 (for wideband systems), can be applied to determine the behaviour of the unidirectional transport in more detail. Instead of inserting the tests signals acoustically, they have to be inserted electrically. Care should be taken since attenuation changes observed might be caused by AGC or companding.

10.9 Unidirectional transport speech quality

10.9.1 Parameter description

The unidirectional transport listening speech quality LQ_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout.

The test is intended to determine any impairment of the listening speech quality introduced by the unidirectional transport.

10.9.2 Test

The tests method to be used is [ITU-T P.863].

NOTE – [ITU-T P.863] does not provide a wideband mode. Wideband systems are evaluated on a superwideband scale. Therefore, the MOS requirements are given in MOS-LQOsw. See P.863 for more information.

1) The test signals used are the English test sequences as specified in [ITU-T P.501] (2 male speakers, 2 female speakers, two sentences each). The test signal level is the nominal signal level, measured at test point Sxin/Rxin; the test signal level is measured as "active speech level" according to [ITU-T P.56]. The speech activity should be between 30% and 70%.

   The original speech signal is used as the reference signal for the determination of the speech quality.

2) The test arrangement is according to clause 9.1. For wideband systems MOS-LQOsw is determined. For narrowband systems MOS-LQOn is determined.

   The calculation is made using the signal recorded at test point Sxout/Rxout.

10.9.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the listening speech quality has to meet the requirements for the applicable performance classes given in the table below.

Table 9-48 – Limits for the unidirectional transport listening speech quality

Performance Class	MOS-LQOn	MOS-LQOsw
1	> 4.2	> 4.5
2	> 3.8	> 4.1
3	> 3.4	> 3.7
4	≤ 3.4	≤ 3.7

10.9.4 Design guidance and root-cause analysis

No additional impairment on listening speech quality should be introduced by the signal transport. Any degradation of listening speech quality will deteriorate the overall listening speech quality and should be avoided. Therefore, the listening speech quality under quiet conditions is checked.

10.10 Unidirectional transport speech quality stability

10.10.1 Parameter description

The unidirectional transport listening speech quality stability LQS_{UD} is measured from test point Sxin/Rxin to test point Sxout/Rxout.

This test is intended to detect speech quality instabilities due to unreliable connection (speech frame loss...). This test should be repeated in different representative test configurations, e.g., positions of the phone within the vehicle, different SRW links active and different WIFI links active.

10.10.2 Test

Several measurements of MOS-LQOsw (for wideband systems) and MOS-LQOn (for narrowband systems) according to clause 10.9 are performed in series within the same call or for different calls, within the test arrangement defined below. Typically, for the same call, a measurement each 20 s is insufficient. The results are reported in terms of statistics.

The procedure is identical for narrowband and wideband systems. In narrowband systems MOS-LQOn is used, in wideband systems MOS-LQOsw.

The assessment of listening speech quality stability is performed in 5 steps:

1) To measure the MOS-LQO periodically over the duration of one communication. **N** measurements provide **N** MOS-LQO values.

2) For each MOS-LQO (i) value i from 2 to N, MOS-LQO_GAP(i) is calculated as the absolute difference with the previous value MOS-LQO (i-1):

   MOS-LQO_GAP (i) =| MOS-LQO (i)- MOS-LQO (i-1)|

3) In order to take into account the subjective perception and measurement accuracy, MOS-LQO_GAP (i) is set to 0 when the difference is equal or lower to THRESHOLD1:

   – if MOS-LQO_GAP (i) > (2*THRESHOLD1), then MOS-LQO_GAP (i) = MOS-LQO_GAP (i)

   – if THRESHOLD1 < MOS-LQO_GAP (i) ≦ 2*THRESHOLD1, then MOS-LQO_GAP (i) = [MOS-LQO_GAP (i) * 2] – (2*THRESHOLD1)

   – if MOS-LQO_GAP (i) ≦ THRESHOLD1, then MOS-LQO_GAP (i) = 0

4) The instability (INS_MOS-LQO) associated to the MOS-LQO over the whole N measurements is defined by mean value of MOS-LQO_GAP (i).

   INS_MOS-LQO = 1/(N-1)Σ MOS-LQO_GAP (i) with i = [2:N]

5) A linear weighting function is applied in order to express stability ST-MOS-LQO on a 0 to 100 scale.

This formulation used to determine the listening quality stability (ST-MOS) as

ST-MOS = 100-(250* INS_MOS)

and

ST-MOS = 0 if [100-(250* INS_MOS)] < 0

When ST-MOS is calculated within a single call, the call should be longer than 3 minutes (Recommended duration being between 3 and 5 minutes). The duration of each measurement depends on the length of the speech samples used for the test as described in clause 10.9.

(As an example, if a sample is 15 s length and the analysis is done every 20 s a minimum of 10 values will be measured).

10.10.3 Performance level classification based on values of this parameter

For the stability indicator about listening speech quality, THRESHOLD1 = 0.1 and the linear weighting function applies in order to express stability (ST-MOS) on a 0 to 100 scale. By definition stability equals 100 when no variations occur and stability ST-MOS equals 0 when MOS-LQO variation is equal or more than 0.4.

To claim compliance with a certain performance class the listening speech quality stability has to meet the requirements for the applicable performance classes given in the table below.

Table 9-49 – Limits for the unidirectional transport listening speech quality stability

Performance Class	ST-MOSn	ST-MOSsw
1	> 95	> 95
2	> 80	> 80
3	> 50	> 50
4	≤ 50	≤ 50

10.10.4 Design guidance and root-cause analysis

Listening quality stability during a call (if the position or transmission characteristics change during the call (or during several different calls) takes into account degradations generated on the signal by the transmission link impairment and the phone position.

In the case of systems using SRW transmission it is the purpose to verify the integration of an SRW radio network by evaluating change of speech quality over time. This will help to detect problems with RF coverage inside the car cabin and verify the error concealment (packet loss, bit errors) caused by weak RF link or interference with other radio.

A guidance to proceed with is:

1 Check RF coverage from an SRW unit to possible mobile positions (Protocol analyser).

2 Identify weak and bad reception areas inside the vehicle. (Protocol analyser)

Check bit error rate, rate of packet loss…

3 Use the speech quality measurement to rate the quality of error concealment in the weak areas identified in step 2 above and compare with measurements from areas with good coverage.

4 Identify possible issues from interference with parallel SRW links or other networks. This will check how the SRW can handle interference and change to undisturbed channels. Speech quality measurements can be used to see the performance of handling these problems.

It is recommended to use to a known link as reference.

Annex A

Microphone measurements in anechoic conditions

(This annex forms an integral part of this Recommendation.)

The scope of these measurements is the verification of microphone parameters in a defined acoustic environment. The influence of integration such as mounting, orientation and in-car acoustics is quite important for the overall performance and therefore it is recommended to integrate the microphone in a surrogate environment as close as possible to the production environment. In case wind noise may have a major influence on the microphone performance, it is recommended to also simulate the wind noise in the anechoic environment.

A.1 Microphone sensitivity (in anechoic conditions)

A.1.1 Parameter description

Microphone sensitivity has to be measured in the free sound field. The sensitivity refers to the sound pressure of the undisturbed free sound field (in the absence of the microphone). The sensitivity is measured at the microphone output connected to the appropriate test circuit (see clause 8.2).

Microphone sensitivity at 1 kHz shall be measured in the direction of its maximum sensitivity.

A.1.2 Test

1) The test signal is a sine wave of 1 kHz at a level of 0 dBPa at the microphone position in the undisturbed free sound field.
2) The microphone is positioned at a distance of 1 m in the acoustic centre line of the loudspeaker.
3) The microphone is oriented towards the loudspeaker with its direction of maximum sensitivity.
4) The sensitivity is determined in mV/Pa.

Further information can be found in [IEC 60268-4].

A.1.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone sensitivity has to meet the requirements for the applicable performance classes for the microphone sensitivity defined in Table A.1.

Table A.1 – Limits for the microphone sensitivity (in anechoic conditions)

Performance Class	Sensitivity
1	nominal sensitivity in mV/Pa +/–1 dB
2	nominal sensitivity in mV/Pa +/–2 dB
3	nominal sensitivity in mV/Pa +/–3 dB
4	nominal sensitivity in mV/Pa > +/–3 dB

A.1.4 Design guidance and root-cause analysis

The nominal sensitivity should be chosen so that there is a good balance between detecting speech accurately without causing clipping due to high audio level in the car (e.g., 300 mV/Pa with a supply voltage of 8V.

A.2 Microphone frequency response (in anechoic conditions)

A.2.1 Parameter description

The microphone frequency response has to be measured in the free sound field. The frequency response refers to the sound pressure of the undisturbed free sound field (in the absence of the microphone). The frequency response is measured at the output of the test circuit.

A.2.2 Test

1) The test signals are sine waves at a level of 0 dBPa at the microphone position in the undisturbed free sound field covering at least the defined frequency range.
2) The microphone is positioned at a distance of 1 m in the acoustic centre line of the loudspeaker.
3) The microphone is oriented towards the loudspeaker with its direction of maximum sensitivity.
4) The sensitivity for each frequency is determined in mV/Pa.

Further information can be found in [IEC 60268-4].

A.2.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the frequency response has to meet the requirements for the applicable performance classes for the microphone send sensitivity as defined in the tables below.

The masks are drawn by straight lines between the breaking points in Tables A.2 to A.5 on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table A.2 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
100	N	
125	0	
200	0	–14
315	0	–13
400	0	–12
500	0	–11
630	0	–10
1 000	0	–8
1 300	2	–8
1 600	3	–8

Table A.2 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 1

Frequency (Hz)	Upper limit	Lower limit
2 000	4	–8
3 100	4	–8
4 000	4	–8
8 000	4	
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table A.3 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 2

Frequency (Hz)	Upper limit	Lower limit
100	FFS	FFS
125	FFS	FFS
200	FFS	FFS
315	FFS	FFS
400	FFS	FFS
500	FFS	FFS
630	FFS	FFS
1 000	FFS	FFS
1 300	FFS	FFS
1 600	FFS	FFS
2 000	FFS	FFS
3 100	FFS	FFS
4 000	FFS	FFS
8 000	FFS	FFS
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table A.4 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
100	FFS	FFS
125	FFS	FFS
200	FFS	FFS
315	FFS	FFS
400	FFS	FFS
500	FFS	FFS

Table A.4 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 3

Frequency (Hz)	Upper limit	Lower limit
630	FFS	FFS
1 000	FFS	FFS
1 300	FFS	FFS
1 600	FFS	FFS
2 000	FFS	FFS
3 100	FFS	FFS
4 000	FFS	FFS
8 000	FFS	FFS
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

Table A.5 – Tolerance mask for the microphone send sensitivity frequency response for Performance Class 4

Frequency (Hz)	Upper limit	Lower limit
100	NA	NA
125	NA	NA
200	NA	NA
315	NA	NA
400	NA	NA
500	NA	NA
630	NA	NA
1 000	NA	NA
1 300	NA	NA
1 600	NA	NA
2 000	NA	NA
3 100	NA	NA
4 000	NA	NA
8 000	NA	NA
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

A.2.4 Design guidance and root-cause analysis

NOTE 1 – Depending on customer demands, other tolerance schemes than those described in Table A.2 may be applied and have to be defined in an equivalent format.

NOTE 2 – Ideally, the response characteristics of the microphone should be flat in the frequency range of wideband transmission (100 Hz-7 kHz). However, especially in the presence of background noise, a bandwidth limitation may be desirable. No explicit recommendation can be given here since such limitation would depend on level and spectral content of the background noise and ideally should be adaptive. If, however, a bandwidth limitation is introduced, it should be made at both the high and low frequencies.

NOTE 3 – Table A.2 applies wider tolerances than Table 8-6 to achieve the freedom of adapting the microphones frequency response to the needs in the car.

A.3 Microphone directional characteristics (in anechoic conditions)

A.3.1 Parameter description

The directional characteristic of a microphone is described by different sensitivities at different angels of sound incidence.

The front to back ratio is the ratio between the sensitivity in the direction of highest sensitivity and the sensitivity at the angle of lowest sensitivity expressed in dB at 1 kHz. The front to back ratio is measured at the output of the test circuit.

A.3.2 Test

1) The test signal is a sine wave of 1 kHz at a level of 0 dBPa at the microphone position in the undisturbed free sound field.
2) The microphone is positioned at a distance of 1 m in the acoustic centre line of the loudspeaker.
3) The first measurement is done with the microphone oriented towards the loudspeaker with its direction of maximum sensitivity. The second measurement is done with the microphone oriented towards the loudspeaker with its direction of minimum sensitivity. If the direction of minimum sensitivity is not known, it has to be determined by rotating the microphone until the minimum is found.
4) The front to back ratio is determined in dB.

Further information can be found in [b-IEC 60268-4].

A.3.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone directionality has to meet the requirements for the applicable performance classes for the microphone directional characteristics defined in Table A.6.

Table A.6 – Limits for the microphone directional characteristics

Performance Class	Front to back ratio
1	>= 10 dB
2	FFS
3	FFS
4	< FFS

A.3.4 Design guidance and root-cause analysis

To achieve appropriate noise reduction a high front to back ratio is desired.

NOTE – Depending on mounting and orientation, also lower front to back ratios can be an advantage.

A.4 Microphone distortion (in anechoic conditions)

A.4.1 Parameter description

The microphone distortion refers to the sound pressure of the undisturbed free field. The distortion is measured at the output of the test circuit.

The harmonic distortion with a sound pressure level of 0 dBPa (94 dBSPL) at the position of the microphone shall be measured in the narrowband frequency range.

A.4.2 Test

1) The test signal is a sine wave with a frequency of 300 Hz, 500 Hz, 1 kHz, 2 kHz at a level of 0 dBPa.
2) The microphone is positioned in an acoustic centre line of the loudspeaker.
3) The microphone is oriented towards the loudspeaker with its direction of maximum sensitivity.
4) The total harmonic distortion is expressed in %.

Care has to be taken that the loudspeaker is able to produce the required sound pressure level with a lower distortion than the microphone under test.

Further information can be found in [b-IEC 60268-4].

A.4.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone distortion has to meet the requirements for the applicable performance classes for the microphone distortion defined in Table A.7.

Table A.7 – Limits for the microphone distortion

Performance Class	SDN
1	< 1%
2	FFS
3	FFS
4	> FFS

A.4.4 Design guidance and root-cause analysis

TBD

A.5 Microphone maximum sound pressure level (in anechoic conditions)

A.5.1 Parameter description

The maximum sound pressure is defined by the sound pressure level where the total harmonic distortion of the microphone at 1 kHz is 3% in the narrowband frequency range. The total harmonic distortion is measured at the output of the test circuit.

A.5.2 Test

1) The test signal is a sine wave with a frequency of 1 kHz and an increasing level to determine the level of 3% total harmonic distortion.
2) The microphone is positioned in an acoustic centre line of the loudspeaker.
3) The microphone is oriented towards the loudspeaker with its direction of maximum sensitivity.
4) The maximum sound pressure level is expressed in dBSPL or dBPa.

Care has to be taken that the loudspeaker is able to produce the required sound pressure level with a lower distortion as the microphone under test.

A.5.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone sound pressure level has to meet the requirements for the applicable performance classes for the microphone maximum sound pressure level defined in Table A.8.

Table A.8 – Limits for the microphone maximum sound pressure level

Performance Class	
1	≥115 dB SPL
2	≥ 105 dB SPL
3	≥ 100 dB SPL
4	< 100 dB SPL

A.5.4 Design guidance and root-cause analysis

With a good microphone design, the maximum sound pressure level is electrically limited by the supply circuit. A microphone with higher sensitivity will reach the electrical output limits at a lower sound pressure level compared to another microphone with lower sensitivity.

A.6 Microphone dynamic range (in anechoic conditions)

A.6.1 Parameter description

The dynamic range shall be measured at the output of the test circuit according to Figure 7-2 in quiet conditions.

A.6.2 Test

1) For the measurement, no test signal is used.
2) The microphone has to be powered with a low noise voltage supply.
3) The self-noise is measured at the output of the test circuit in the frequency range between 100 Hz and 8 kHz, A-weighting has to applied.
4) The dynamic range is measured by referring the self-noise is expressed in dBV(A) to the maximum sound pressure level measured for the microphone.

Care has to be taken that the environmental noise is below the equivalent self-noise of the microphone.

A.6.3 Performance level classification based on values of this parameter

To claim compliance with a certain performance class the microphone has to meet the requirements for the applicable performance classes for the microphone dynamic range defined in Table A.9.

Table A.9 – Dynamic Range of the microphone

Performance Class	dB rel. to MaxSPL
1	>70 dB
2	> 65 dB
3	>60 dB
4	≤ 60 dB

A.6.4 Design guidance and root-cause analysis

TBD

Annex B

Standard set of user scenarios

(This annex forms an integral part of this Recommendation.)

Table B.1 – Standard set of user scenarios used to collect noise recordings

		Vehicle settings						**Environmental conditions**			
User scenario	**Description**	**Vehicle speed**	**HVAC settings**	**Windows**	**Wipers**	**Turn signal**	**Back-ground talkers**	**Road surface (see Note 3)**	**Wind speed**	**Precipi-tation**	**Temp.**
1	Stationary vehicle with low HVAC noise	0 km/h (at idle)	**FAN** = Lowest setting	Up	Off	Off	None	N/A	< 5 m/s (12 mph)	None	> –20C and < 40C
2	City driving with high HVAC noise	60 km/h (37 mph)	**FAN** = Setting closest to 6 dB(A) above driving noise with FAN = lowest setting; **AIRFLOW** = Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C
3	Highway driving with low HVAC noise	120 km/h (75 mph)	**FAN** = Lowest setting	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C
4	Highway driving with high HVAC noise	120 km/h (75 mph)	**FAN** = Setting closest to 6 dB(A) above driving noise as in condition 3; **AIRFLOW** = Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C
5 see Note 1	Highway driving with high HVAC noise	≥160 km/h (≥100 mph)	**FAN** = Lowest Setting **AIRFLOW** = Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C

		Vehicle settings						Environmental conditions			
User scenario	**Description**	**Vehicle speed**	**HVAC settings**	**Windows**	**Wipers**	**Turn signal**	**Back-ground talkers**	**Road surface (see Note 3)**	**Wind speed**	**Precipi-tation**	**Temp.**
6 see Note 2	Stationary vehicle with high HVAC noise	0 km/h (at idle)	**FAN** = Highest setting, **AIRFLOW** = Directed to microphone if possible	Up	Off	Off	None	N/A	< 5 m/s (12 mph)	None	> –20C and < 40C
7 see Note 3	Highway driving with low HVAC noise; windows down and open roof (if possible)	120 km/h (75 mph)	**FAN** = Lowest settings	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C

NOTE 1 – Optional test: If the hands-free system is to be deployed in countries where higher driving speeds are allowed, then testing at 160 k.p.h. or more should be performed in addition to the standard set of user scenarios.

NOTE 2 – Required test if the airflow can be directed towards the microphone. (This may result in wind buffeting.)

NOTE 3 – Optional test: This test is applicable if the hands-free system is required to work with windows down, roof open or convertibles.

NOTE 4 – Smooth road surfaces that generate very little tyre noise shall not be used. Also, road surfaces with bumps that cause significant impulse noises shall not be used either. If available, concrete surfaces are preferred because they often result in worst case conditions that cause impairments not seen on other road surfaces. The road surface used should not introduce changes in the noise structure in such a way that users would change their voice level.

NOTE 5 – AIRFLOW refers to the HVAC mode settings related to how air is directed into the cabin. For example, in North American vehicles there is typically a "Defrost" setting that will direct the flow of air onto the windows.

Annex C

Frame process and delay

(This annex forms an integral part of this Recommendation.)

C.1 General

This annex provides definition and handling of delay for frame process subsystems.

Typical speech enhancement subsystems have spectral analysis/synthesis processes to realize good noise reduction or echo cancellation. Time domain input signal in a fixed length of window is applied, analysed, enhanced, then synthesized by overlap-add method. The windowing period is often called "frame shift".

Typical data transport subsystems also take frame processes for efficient data compression or error correction. A unit of the data transferred is called a packet.

When two subsystems, sender and receiver, which have different frame data interval interfaces to each other, a special data buffer must be inserted to arbitrate the difference of intervals. Several definitions of delay and framing process should be given to such processing. This annex provides the definitions for such parameters.

C.2 Frame interval

Frame interval is defined as a logical time interval of input or output frame data. Frame interval of time domain data is regarded as one sample period. A subsystem which has framed input or output signals shall specify the frame interval.

Frame rate is the reciprocal of frame interval.

C.3 Delay

In general, the subsystem delay is measured as **input-to-output delay**. Subsystems which have framed input or output signals may have three types of delays: **buffer delay, algorithmic delay** and **computational delay**. If the subsystem is composed of software, **algorithmic delay** is defined, and **computational delay** is also defined after the software implementation. Table C.1 provides the definition of these delays. A subsystem supplier shall specify the correspondent delay time.

Table C.1 – Definition of delays

<table>
<tr><th colspan="2">Category</th><th>Definition</th></tr>
<tr><td colspan="2">Subsystem input-to-output delay (Note)</td><td>A subsystem's physical delay time from data input to output.
Depending on implementation the input-to-output delay may consist of:
• buffer delay (if framing is used)
• algorithmic delay (if signal processing is used)
• computational delay (delay caused by any processing).</td></tr>
<tr><td colspan="2">Buffer delay</td><td>Additional delay time when a buffer is introduced at interface to arbitrate frame size difference.</td></tr>
<tr><td rowspan="2">Software</td><td>Algorithmic delay</td><td>A subsystem's logical delay time in case the subsystem is realized by software.</td></tr>
<tr><td>Computational delay</td><td>Additional computational delay time from data input to output caused by a processor. It is applied in case the software subsystem is implemented onto the processor.</td></tr>
<tr><td colspan="3">NOTE – Input-to-output delay is measured with other subsystems disconnected. If framing is used the test system has to provide appropriate frame sizes.</td></tr>
</table>

Figure C.1 shows a system integrating two framed software subsystems on a processor. Input buffer A is placed before subsystem A to convert input sampled data to framed data. If subsystems A and B have the same frame interval, a buffer between these subsystems should not be necessary. After subsystem B, the output framed data is converted to sampled data by output buffer B which can be regarded as no delay.

The overall delay of this system is calculated by summing buffer delay A, algorithmic delay A, algorithmic delay B and computational delay.

Overall system delay =

(Buffer delay A) + (Algorithmic delay A) + (Algorithmic delay B) + (Computational delay)

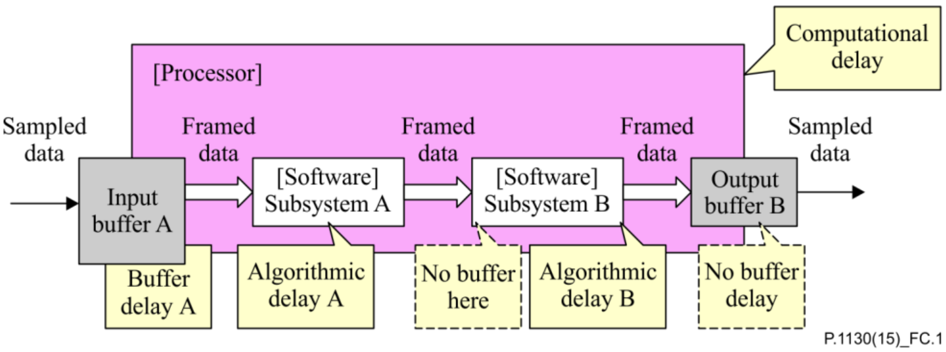


Figure C.1 – Framing and delays

C.4 Arbitration between two subsystems with different frame intervals

When a receiver subsystem interfaces with a sender subsystem which has a different frame interval, a data buffer shall be categorized into the following cases.

Case 1) Sender's and receiver's frame intervals are identical

No additional delay is to be inserted.

Case 2) Sender's frame interval is a multiple of receiver's frame interval

No delay will be added at the receiver. The total delay equals to the sender's delay. Figure C.2 shows an example where the sender's frame interval is 15 ms and the receiver's is 5 ms. Algorithmic delay and computational delay are omitted. In Figure C.1, subsystem A can be regarded as the sender, subsystem B as the receiver, buffer A as the sender's input buffer, and buffer B as the receiver's output buffer.

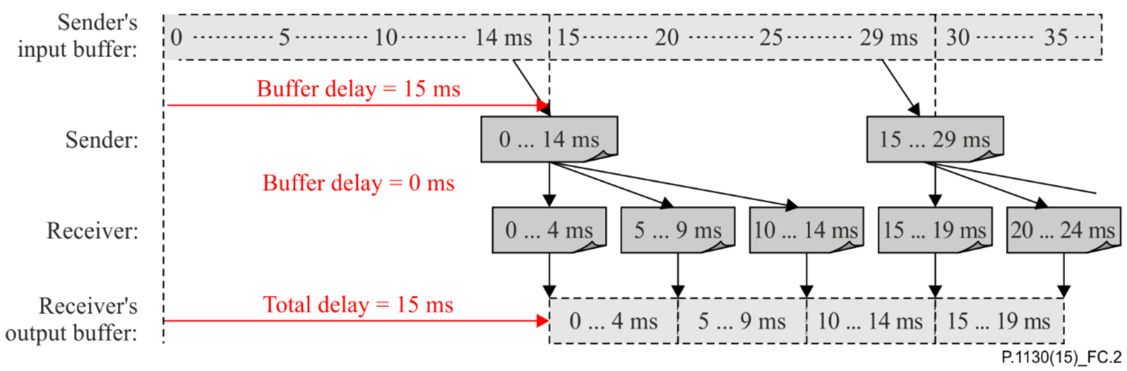


Figure C.2 – Example of Case 2

Case 3) Receiver's frame interval is a multiple of sender's frame interval

Receiver's delay buffer size is equal to its frame interval minus sender's frame interval. However, the total buffer delay is equal to the receiver's frame interval. Figure C.3 shows an example where the sender's frame interval is 5 ms and the receiver's frame interval is 15 ms.

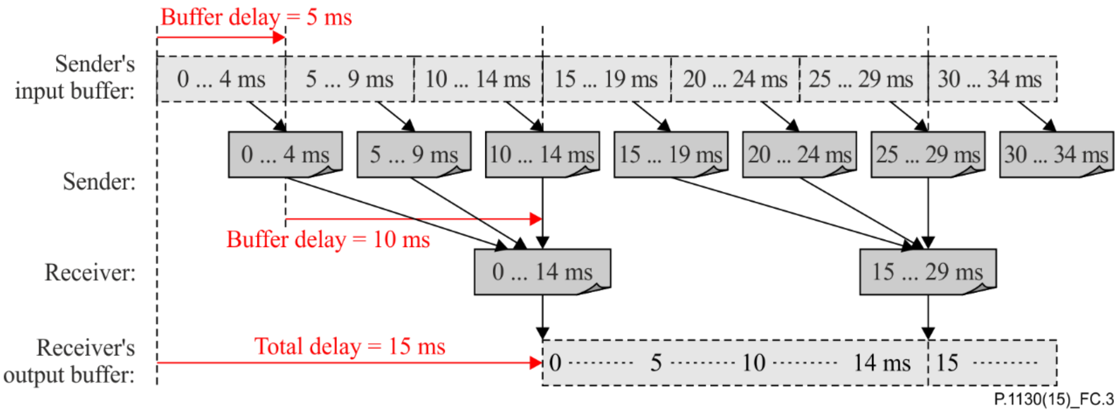


Figure C.3 – Example of Case 3

Case 4) Other cases

In a general case, an arbitration buffer is inserted between two subsystems. The receiver has to read the buffer with the delay equal to its frame interval. It coincides with the case that the data is transferred with the time domain signal. The resultant buffering delay is equal to the sum of frame intervals of two subsystems. Figure C.4 shows an example where the sender's frame interval is 5 ms and the receiver's is 8 ms.

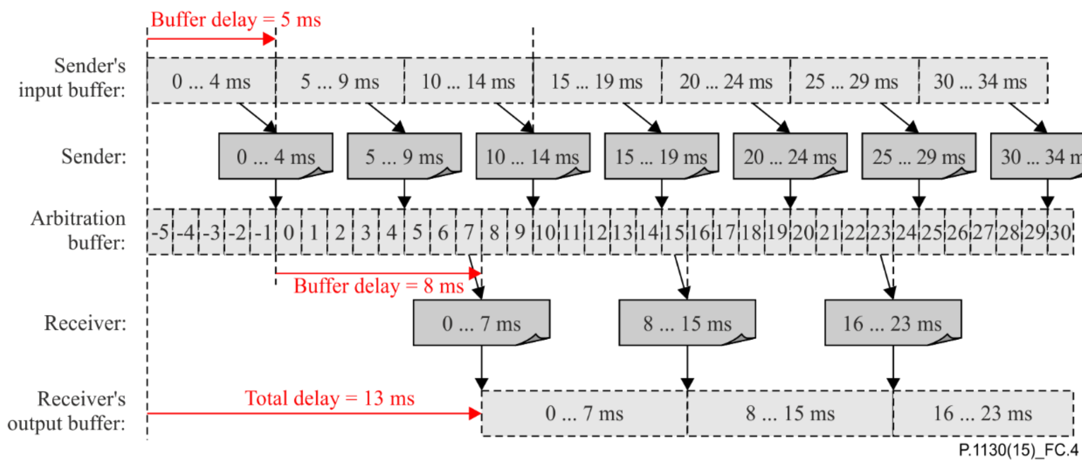


Figure C.4 – Example of Case 4

In special cases where two subsystems are operating on asynchronous clocks, a drop/insert buffer with a unit or the multiples of frame size should be used to arbitrate them.

C.5 Subsystem delay and relationship to overall delay

It is recognized that for frame-based implementations the sum of the input-to-output delays of the individual subsystems may be higher than the total delay of the implementation because frame buffers could be shared by different subsystems.

In this specification, the subsystem delay is defined as the physical input-to-output delay. For example, if a subsystem processes 10 ms frame interval data with 20 ms algorithmic delay, the input-to-output delay is calculated as 20 ms + 10 ms = 30 ms.

C.6 Example

An example of wideband mode SRW/hands-free system is shown. Examples of frame process are shown in Table C.2.

Table C.2 – Frame intervals of subsystems

Subsystem	Frame interval
Signal enhancement	15 ms (to meet SRW frame interval)
SRW	1.25, 2.5, 3.75 ms (CVSD) for narrowband 7.5, 15 ms (mSBC) for wideband
Cell phone	20 ms (Speech codec)

An example of subsystems composing of an automotive hands-free send system is shown in Figure C.5. The upper boxes specify frame intervals and delays of each subsystem. The total send delay is calculated by summing all delays surrounded by red circles excepted for cell phone factors which are out of send delay specified by [ITU-T P.1100] or [ITU-T P.1110]. According to Case 2 in clause C.4, buffer delay at input of SRW is omitted.

1.5 ms + 0.5 ms + 15 ms + 25 ms + 7.5ms + 15 ms = 64.5 ms

Here the delay of the send signal enhancement subsystem can be measured as follows:

15 ms (buffer delay) + 25 ms (algorithmic delay) + 7.5 ms (computational delay) = 47.5 ms

In the same way, the delay of the send SRW subsystem is:

7.5 ms (buffer delay) + 15 ms (algorithmic delay) + 7.5 ms (computational delay) = 30 ms

As it is seen, the total system delay calculated above is much shorter than just the sum of these two subsystems. The reason comes from the subsystem integration sharing part of data buffering (−7.5 ms) and computational delay (–7.5 ms).

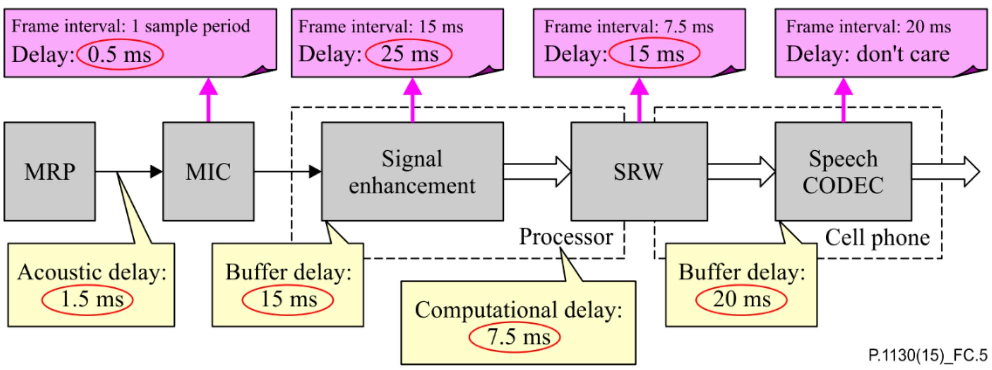


Figure C.5 – An example of automotive hands-free send system

In the same way, the subsystems composing of the receive system is shown in Figure C.6. The total receive delay is calculated by summing all delays surrounded by red circles excepted for cell phone factors. According to Case 3 in clause C.4, buffer delay at input of SRW is omitted.

15 ms + 15 ms +15 ms + 7.5 ms + 0.5 ms + 2 ms = 55 ms

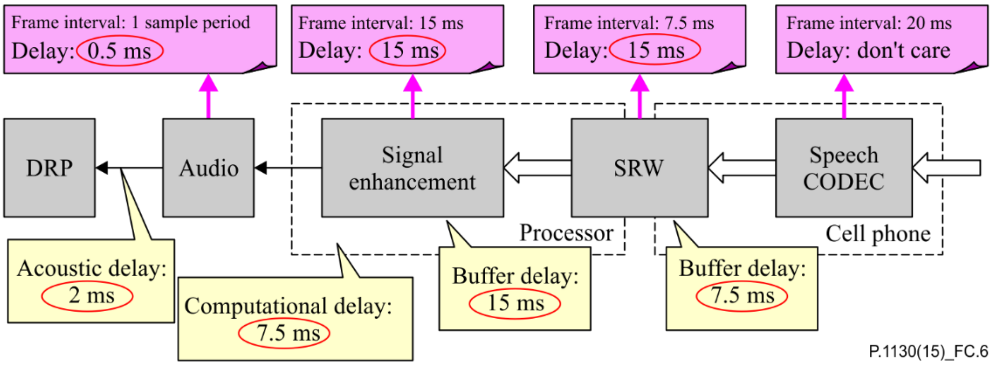


Figure C.6 – An example of automotive hands-free receive system

In this scenario, the round trip delay is equal to 119.5 ms, which meets [ITU-T P.1110].

The next examples are the typical cases that the signal enhancement subsystems of an automotive hands-free send system and receive system are implemented without special care for delay or realized as an independent chip from SRW and having 16 ms frame interval. Since the interval is not a simple multiple of SRW frame interval, these processes are implemented as asynchronous tasks. Figure C.7 shows the send system. The total delay of the send system is calculated as below. According to Case 4 in clause C.4, buffer delay at input of SRW is also added.

1.5 ms + 0.5 ms + 16 ms + 16 ms + 26 ms + <u>7.5</u> ms + 7.5 ms + 15 ms = 90 ms

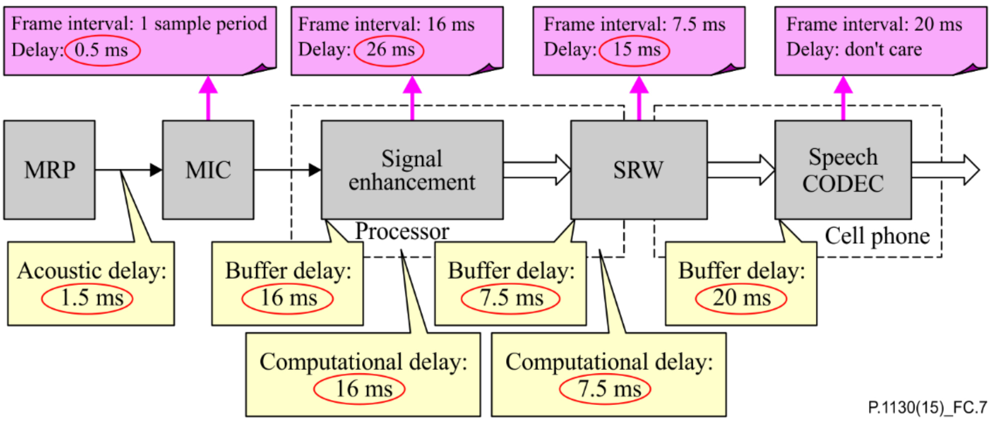


Figure C.7 – A bad example of automotive hands-free send system

The receive system is shown in Figure C.8. The total receive delay is calculated as below. According to Case 4 in clause C.4, buffer delay at input of SRW is also added.

<u>7.5 ms</u> + 15 ms + 7.5 ms + 16 ms + 16 ms + 16 ms + 0.5 ms + 2 ms = 80.5 ms

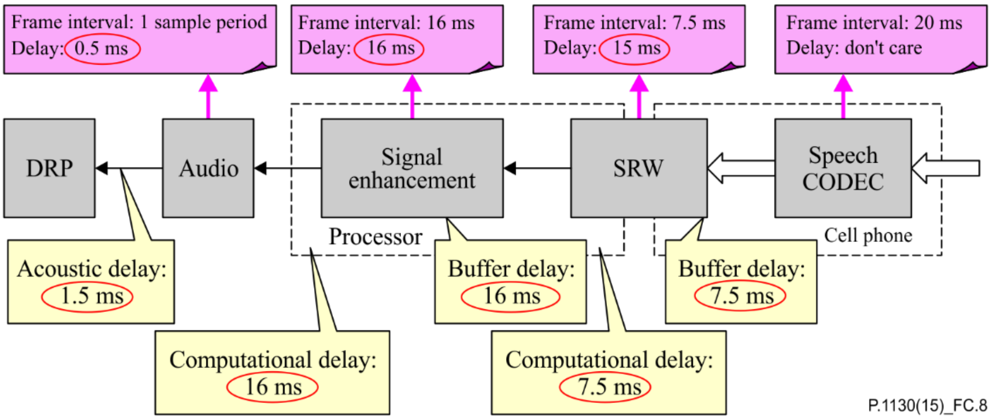


Figure C.8 – A bad example of automotive hands-free receive system

In this scenario, the round trip delay is equal to 170.5 ms, which cannot meet [ITU-T P.1110].

Appendix I

Testing for Noise Distortion

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

This appendix (see [b-Yu_1], [b-Yu-2]) describes a test using an objective noise distortion measure specifically designed to quantify the amount of musical tones. The system under test is the noise reduction functionality in the signal enhancement subsystem, as described in clause 8.4. If comfort noise generation is part of the system, it should be switched off. The purpose of the measure described in this appendix is to help to optimize the signal enhancement subsystem during the development and optimization process. The described test is recommended, although not mandatory.

I.1 Parameter description

The noise distortion is measured in the send direction with the weighted log-average kurtosis ratio WLAKR of the signals at test point (S2) and at test point (S3) of the signal enhancement subsystem as shown in Figure 8-13. The weighted log-average kurtosis ratio WLAKR describes the amount of musical tones by measuring the average change of the weighted kurtosis of the unprocessed background noise signals and the processed background noise signals using the signal enhancement subsystem.

I.2 Test

In principle, the test could be carried out using noise signals acquired by the hands-free microphone in the car. A number of not less than 18 noise samples of length 8 seconds taken from typical stationary driving conditions should then be recorded and taken for the subsequent measurement. An even more relevant problem to solve may be to allow comparison to tests carried out for other signal enhancement subsystems in other laboratories. Therefore, as the recommended alternative, the noise data shall be taken from the inside-car background noise signals in section 8.2 ("Stereophonic Signals") of [b-ETSI ES 202 396-1]. The left channel of all 9 signals including fullsize car1, midsize car1 and midsize car2 with 80 km/h, 100 km/h and 130 km/h are selected, respectively. Each signal is downsampled from 48 kHz to 16 kHz, the first 16 seconds are taken and segmented into 2 noise signals with a length of 8 seconds each. This leads to altogether 18 reference background noise signals. For the test of a wideband subsystem [ITU-T P.341] weighting filter [ITU-T P.341] for wideband signals (50 … 7 000 Hz) is applied. For the test of a narrowband subsystem the modified IRS (MIRS) weighting filter according to [ITU-T P.830] (300 … 3 400 Hz) is applied and subsequently downsampled to 8 kHz. The mean weighted log-average kurtosis ratio WLAKR is then computed from all 18 weighted log-average kurtosis ratios. *For each of the 18 background noise signals*, one weighted log-average kurtosis ratio WLAKR is calculated as follows:

1) The long-term rms level of the 16 kHz/8 kHz sampled background noise signal of length 8 seconds is adjusted to –26 dBov according to [ITU-T P.56]. Then the signal is digitally stored as the reference signal, and fed into test point (S2) as an input to the signal enhancement subsystem.
2) The background noise signal processed by the signal enhancement subsystem is taken from test point (S3) and also digitally stored.
3) Now the signals are transformed into the Fourier domain. To accomplish this, a square root Hann window of length 512 (or 256 for narrowband signals) is applied, followed by a discrete Fourier transform (DFT) of the same length and a frame shift of 50%. These operations are separately performed for the stored reference signal $n(\mathrm{n})$ and the processed noise signal $\tilde{n}(\mathrm{n})$ with n being the time index, respectively.

4) Based on the DFT coefficients of the reference signal and the processed background noise signal, the weighted log-average kurtosis ratio WLAKR = ln $\left(\frac{\psi_{\tilde{n}}^{\mathrm{w}}}{\psi_{n}^{\mathrm{w}}}\right)$ is computed, where $\ln(\cdot)$ is the natural logarithm, ψ_{n}^{w} and $\psi_{\tilde{n}}^{\mathrm{w}}$ are the average weighted kurtosis of the reference signal $n(\mathrm{n})$ and the processed noise signal $\tilde{n}(\mathrm{n})$, respectively. ψ_{n}^{w} is calculated by averaging the instantaneous kurtosis of the weighted squared amplitude reference signal DFT coefficients $|N(\ell,k)|^2$over all frames:

$$\psi_{n}^{\mathrm{w}}(\ell) = \frac{\frac{1}{K}\sum_{k=0}^{K-1}\left[\alpha_n(k)\cdot|N(\ell,k)|^2 - \overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2}\right]^4}{\left(\frac{1}{K}\sum_{k=0}^{K-1}\left[\alpha_n(k)\cdot|N(\ell,k)|^2 - \overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2}\right]^2\right)^2}, \quad \psi_{n}^{\mathrm{w}} = \frac{1}{L}\sum_{\ell=1}^{L}\psi_{n}^{\mathrm{w}}(\ell), \tag{I.1}$$

with $\overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2} = \frac{1}{K}\sum_{\kappa=0}^{K-1}\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2$, ℓ and k (or κ) being the frame index and frequency bin, respectively, and K being the DFT length. The weighting factor $\alpha_n(k) = \left(\frac{1}{L}\sum_{\ell=1}^{L}|N(\ell,k)|^2\right)^{-1}$ is being calculated for each frequency bin as the inverse mean value of $|N(\ell,k)|^2$ across all L frames. The average weighted kurtosis $\psi_{\tilde{n}}^{\mathrm{w}}$ can be calculated in the same way by using $\left|\tilde{N}(\ell,k)\right|^2$ and $\alpha_{\tilde{n}}(k)$ instead of $|N(\ell,k)|^2$ and $\alpha_n(k)$ in equation (I.1), respectively. The noise distortion, i.e., the amount of musical tones produced by the signal enhancement subsystem, shall be evaluated in terms of the mean of the weighted log-average kurtosis ratio WLAKR. More details are described in [YU-KURTOSIS-WB] and [YU-KURTOSIS-NB].

I.3 Classification of performance level based on values of this parameter

To claim compliance with a certain performance class the WLAKR of the signal enhancement subsystem in the send direction has to meet the requirement for the applicable performance classes defined in Table I.1 for wideband subsystems, and in Table I.2 for narrowband subsystems. The audibility of musical tones is rated with seven categories on a mean opinion score (MOS) scale: (1) intolerably audible, (2) loudly audible, (3) rather loudly audible, (4) moderately audible, (5) slightly audible, (6) just audible, and (7) inaudible.

Table I.1 – Limits for the mean of the weighted log-average kurtosis ratio WLAKR of the signal enhancement subsystem in send direction (wideband)

Performance Class	WLAKR	Interpretation	
		Mean opinion score (MOS)	**Audibility of musical tones**
1	WLAKR < 0.72	5.5 < MOS	Just audible or inaudible
2	0.72 ≤ WLAKR < 1.16	4.5 < MOS ≤ 5.5	Slightly audible
3	1.16 ≤ WLAKR < 1.56	3.5 < MOS ≤ 4.5	Moderately audible
4	1.56 ≤ WLAKR	MOS ≤ 3.5	Intolerably audible or loudly audible or rather loudly audible

Table I.2 – Limits for the mean of the weighted log-average kurtosis ratio WLAKR of the signal enhancement subsystem in send direction (narrowband)

Performance Class	WLAKR	Interpretation	
		Mean opinion score (MOS)	Audibility of musical tones
1	WLAKR < 0.30	5.5 < MOS	Just audible or inaudible
2	0.30 ≤ WLAKR < 0.67	4.5 < MOS ≤ 5.5	Slightly audible
3	0.67 ≤ WLAKR < 1.00	3.5 < MOS ≤ 4.5	Moderately audible
4	1.00 ≤ WLAKR	MOS ≤ 3.5	Intolerably audible or loudly audible or rather loudly audible

I.4 Design guidance and root-cause analysis

Noise distortion introduced by the signal enhancement subsystem may lead to a high degradation of the overall perceived speech quality. Therefore, the described noise distortion measure should be evaluated during the optimization of the signal enhancement subsystem, and judged along with speech signal-related quality measures. A high noise distortion may indicate a too low spectral gain floor in spectral noise reduction, or a too low noise overestimation factor in spectral noise power estimation, or a too low SNR floor and/or a too weak temporal smoothing in SNR estimation.

Appendix II

Reference-free SNR measurement

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

II.1 Introduction

This appendix describes an approach to measure the SNR of a speech signal distorted by car noise as close as possible to the reference SNR from [ITU-T P.56], but without using any reference signals based on [b-Berger]. This "reference-free" nature [b-Fodor] allows for a wide flexibility; the proposed method just has to be applied to the noisy speech signal. Moreover, it offers low complexity, and can be applied both to narrowband and wideband signals.

To measure the SNR of the noisy speech signal in this approach, first the speech and noise power have to be estimated. Because of the additive nature of the noise, the speech power is approximated by the difference between the noisy speech power and the noise power. The noise power is estimated by a noise variance tracking algorithm. To estimate the ITU-T P.56 active speech level only speech active frames are taken for the speech power estimation. Therefore, a voice activity detection (VAD) is needed. The power of the noise component is estimated by means of the noise tracking algorithm and in order to ensure a robust estimation a speech pause detection (SPD). Then, the measured SNR is the ratio of the speech and noise power estimates. Since the effectiveness of the employed algorithms (such as noise variance tracking, VAD, and SPD) decreases at low reference SNR values, the raw SNR measurement leads to systematically measurement errors in this region. Therefore, the resulting raw SNR values will be corrected by a mapping curve.

II.2 Signal-to-noise ratio measurement

The input signal $y(n)$ of the measurement system is assumed to consist of the clean speech signal $s(n)$ and an additive noise signal $n(n)$, with n being the discrete time index. After segmentation, windowing, and the discrete Fourier transform (DFT), the input signal can be rewritten as $Y(\ell,k) = S(\ell,k) + N(\ell,k)$ with Y, S, N, as well as ℓ, and k being the short-time spectra of the noisy speech, the clean speech, the noise component, as well as the analysis frame index, and the frequency bin index, respectively. We define the signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = 10\log_{10}\frac{P_S}{P_N}\ [\mathrm{dB}], \tag{II.1}$$

with $\log_{10}(.)$, P_S, P_N being the logarithm to the base 10, the speech and the noise power, respectively.

Since P_S and P_N are not accessible in practice, they have to be estimated. Exploiting the additive nature of the noise, the numerator is estimated with contributions $\widehat{P_S}(\ell,k) = \max\{|Y(\ell,k)|^2 - \widehat{\sigma_N^2}(\ell,k)\}$, where $|Y(\ell,k)|^2$ and $\widehat{\sigma_N^2}(\ell,k)$ are the power spectral density (PSD) of the noisy speech signal and the estimated noise variance (see clause II.3), respectively. The maximum operator ensures a non-negative speech power estimate. Conforming to the definition of the active speech level [ITU-T P.56], for the estimation of P_S in (1) only speech active frames are taken as

$$\widehat{P_S} = \frac{1}{|\Lambda_1|K}\sum_{\ell\epsilon\Lambda_1}\sum_k \max\{|Y(\ell,k)|^2 - \widehat{\sigma_N^2}(\ell,k), 0\}, \tag{II.2}$$

with Λ_1, $|\Lambda_1|$, and K being the set of active speech frames detected by a VAD as a part of the noise tracking algorithm (see clause II.3), the number of elements in Λ_1, and the number of all frequency bins, respectively.

The precise estimation of the noise power P_N is essential for the SNR estimation. However, the accuracy of the noise variance tracking algorithm is affected especially at low SNRs. In order to increase the robustness of the approach, for the estimation of P_N in (II.1), the estimated noise variance $\widehat{\sigma_N^2}(\ell,k)$ is averaged over all frequency bins k and those frames ℓ belonging to speech pause:

$$\widehat{P_N} = \frac{1}{|\Lambda_0|K} \sum_{\ell \epsilon \Lambda_0} \sum_k \widehat{\sigma_N^2}(\ell,k), \tag{II.3}$$

with Λ_0 and $|\Lambda_0|$ being the set of speech pause frames detected by a separate conservative SPD described in clause II.4, and the number of elements in Λ_0, respectively.

II.3 Noise variance tracking and voice activity detection

The estimation of the noise variance $\widehat{\sigma_N^2}(\ell,k)$ in clause II.2 is performed utilizing a 3-state classifier specific to this estimator. Based on the smoothed periodogram of the noisy speech signal:

$$\overline{|Y(\ell,k)|^2} = 0.5\overline{|Y(\ell-1,k)|^2} + 0.5|Y(\ell,k)|^2 \tag{II.4}$$

and a dynamic threshold $\Theta(\ell,k)$, one hypothesis $H(\ell,k)$ out of three is chosen:

H_{sp}: Speech *presence* is assumed if $\overline{|Y(\ell,k)|^2} > 2\Theta(\ell,k)$.

H_{sa}: Speech *absence* is assumed if $\overline{|Y(\ell,k)|^2} \leq 2\Theta(\ell,k) \ \wedge \ \overline{|Y(\ell,k)|^2} < \widehat{\sigma_N^2}(\ell-1,k)$.

H_{st}: Speech *transition* is assumed if $\overline{|Y(\ell,k)|^2} \leq 2\Theta(\ell,k) \ \wedge \ \overline{|Y(\ell,k)|^2} \geq \widehat{\sigma_N^2}(\ell-1,k)$.

The noise variance estimate $\widehat{\sigma_N^2}(\ell,k)$ is updated as:

$$\widehat{\sigma_N^2}(\ell,k) = \varepsilon(\ell,k)\widehat{\sigma_N^2}(\ell-1,k) + [1-\varepsilon(\ell,k)]\overline{|Y(\ell,k)|^2} \tag{II.5}$$

where the initial value is $\widehat{\sigma_N^2}(\ell=0,k) = 0$ and the time-varying smoothing factor $\varepsilon(\ell,k)$ depends on the hypothesis of the current frame $H(\ell,k)$:

$$\varepsilon(\ell,k) = \begin{cases} 1 & \text{if} \quad H(\ell,k) = H_{\mathrm{sp}}, \\ 0.5 & \text{if} \quad H(\ell,k) = H_{\mathrm{sa}}, \\ 0.875 & \text{if} \quad H(\ell,k) = H_{\mathrm{st}}. \end{cases} \tag{II.6}$$

The update of the dynamic threshold $\Theta(\ell,k)$ is performed as follows:

$$\Theta(\ell,k) = \begin{cases} \Delta\Theta(\ell-1,k) & \text{if} \quad \overline{|Y(\ell,k)|^2} > 2\Theta(\ell-1,k), \\ \overline{|Y(\ell-1,k)|^2} & \text{if} \quad \overline{|Y(\ell,k)|^2} < \Theta(\ell-1,k), \\ \Theta(\ell-1,k) & \text{else}, \end{cases} \tag{II.7}$$

where the multiplication by the step-size constant $\Delta = 1.07$ for narrowband signals ($\Delta = 1.2$ for wideband signals) slowly but steadily decreases the probability of selecting the speech presence hypothesis. The unsmoothed update to the last periodogram value $\overline{|Y(\ell-1,k)|^2}$ allows for rapid threshold adaptation in speech pauses. The control parameter is initialized as $\Theta(\ell=0,k) \to \infty$.

The frame-wise VAD is based on the hypotheses of the noise variance tracking algorithm as introduced above. Therefore, it was integrated into the noise variance tracking algorithm. Frame ℓ is detected as voice active and becomes an element of set Λ_1, if at least 90 % of its frequency bins from the range [500 Hz, 2 500 Hz] containing relevant speech information are classified as speech active H_{sp} or transient H_{st} by the 3–state classifier.

II.4 Speech pause detection

This clause describes the algorithm which is employed to detect frames with speech pause Λ_0. Please note that different to the VAD, this is a separate algorithm which is independent from the noise power tracking. The SPD is based on the frame energy of the noisy speech spectrum calculated by means of (II.4) as follows:

$$\overline{P_Y}(\ell) = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\overline{|Y(\ell,k)|^2} \tag{II.8}$$

with $\mathcal{K}$ and $|\mathcal{K}|$ being the set of frequency bins between 500 Hz and 2500 Hz and the number of elements in this set, respectively. This frequency range excludes low frequencies where most of the car noise power is concentrated. Based on an adaptive threshold $\Xi(\ell)$, the SPD delivers a hypothesis $H^{\mathrm{SPD}}(\ell)$ based on the following 3 states:

H_{SP}: Speech *presence* is assumed if $\overline{P_Y}(\ell) > \Xi(\ell)$.

H_{ST}: Speech *transition* state follows every H_{SP} decision with a duration of L_{trans} frames, unless threshold $\Xi(\ell)$ has been exceeded again.

H_{SA}: Speech *pause* is assumed in all other situations.

Frame ℓ is considered as speech absent and becomes element of Λ_0, if $H^{\mathrm{SPD}}(\ell) = H_{\mathrm{SA}}$. The adaptive threshold $\Xi(\ell)$ is calculated as:

$$\Xi(\ell) = 5\Phi(\ell-1) + \alpha, \tag{II.9}$$

with the additive term α. The SPD floor signal $\Phi(\ell)$ is updated recursively as follows (cf. (II.5)):

$$\Phi(\ell) = \beta(\ell)\Phi(\ell-1) + [1-\beta(\ell)]\overline{P_Y}(\ell), \tag{II.10}$$

where the initial value is $\Phi(\ell = 0) = 0$ and the time-varying smoothing factor $\beta(\ell)$ is defined as follows (cf. (II.6)):

$$\beta(\ell) = \begin{cases} 0.875 & \text{if} \quad \mathcal{A} \wedge \overline{P_Y}(\ell) > \Phi(\ell-1), \\ 0.5 & \text{if} \quad \mathcal{A} \wedge \overline{P_Y}(\ell) \le \Phi(\ell-1), \\ 1 & \text{else,} \end{cases} \tag{II.11}$$

with $\mathcal{A} = \{\overline{P_Y}(\ell) \le 2\Upsilon(\ell-1) \wedge H^{\mathrm{SPD}}(\ell) \ne H_{\mathrm{ST}}\}$ ensuring a more conservative floor signal update in speech pauses than just conditioning on H_{SA}.

The SPD control parameter $\Upsilon(\ell)$ is increased slightly in speech presence and updated strongly during speech absence as (cf. (II.7))

$$\Upsilon(\ell) = \begin{cases} \delta\Upsilon(\ell-1) & \text{if} \quad \overline{P_Y}(\ell) > 2\Upsilon(\ell-1), \\ \overline{P_Y}(\ell) & \text{if} \quad \overline{P_Y}(\ell) < \Upsilon(\ell-1), \\ \Upsilon(\ell-1) & \text{else,} \end{cases} \tag{II.12}$$

with the control update constant δ. The control parameter is initialized as $\Upsilon(\ell = 0) \to \infty$.

The SPD parameters which differ for the narrowband and the wideband implementation are summarized in Table II.1.

Table II.1 – Parameters of the SPD

Parameter	L_{trans}	α	δ
$f_s = 8\,\mathrm{kHz}$	6	10^7	1.085
$f_s = 16\,\mathrm{kHz}$	7	10^8	1.055

II.5 Measurement set-up

At a sampling frequency of 8 kHz (16 kHz), the noisy speech signal $y(n)$ has to be segmented by a Hann window, an analysis frame length of 256 (512) samples, and an analysis frame shift of 50% (50%) to obtain $Y(\ell, k)$. The *raw* SNR value can be calculated as follows:

$$\widehat{\mathrm{SNR}}_{\mathrm{raw}} = 10\log_{10}\frac{\widehat{P_S}}{\widehat{P_N}}\ [\mathrm{dB}], \tag{II.13}$$

using (II.2), the noise variance tracker, and the VAD as described in clause II.3, as well as (II.3), the noise variance tracker from clause II.2, and the SPD as introduced in clause II.4. This *raw* SNR value has to be corrected by the following mapping function in order to obtain an unbiased SNR measurement

$$\widehat{\mathrm{SNR}} = \max\left\{\sum_{j=0}^{10} p_j \left(\frac{\widehat{\mathrm{SNR}}_{\mathrm{raw}} - m}{s}\right)^j, -11\right\}\ [\mathrm{dB}], \tag{II.14}$$

with the scaling parameters m and s, as well as the polynomial coefficients p_j as shown in Table II.2.

Table II.2 – Scaling parameters m and s as well as the polynomial coefficients p_j of the mapping function

f_s	m	s	p_j										
			p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
8 kHz	16.043	11.252	–2.1312	6.4129	2.0957	–19.199	5.0992	19.709	–7.7268	–6.6348	2.0857	13.066	11.555
16 kHz	15.461	11.798	–0.8082	2.8537	–0.3609	–7.5337	4.3304	7.2828	–5.0623	–1.8734	0.8842	13.06	11.48

The approach is summarized by means of a pseudocode in Table II.3.

Table II.3 – Overview of the approach by means of pseudocode

```
... % parameter initialization for signal segmentation, FFT, VAD, SPD, etc.
for l=1:L, % main loop over all frames
    ... % signal analysis: segmentation, windowing, FFT, absolute square
    (4), (7), (6), (5) % 3-state noise variance tracking, VAD
    (4), (8), (9), (12), (11), (10) % SPD
end
(2)  % speech power estimation
(3)  % noise power estimation
(13) % raw SNR
(14) % corrected SNR
```

Appendix III

Measuring and applying impulse response traces for dynamic test conditions

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

III.1 Introduction

This appendix describes a procedure for acquiring and applying impulse response traces to clean speech signals for the generation of dynamic test conditions. This real-time capable procedure depicts an addition or alternative to the generation of speech material for automotive hands-free systems testing in time-variant conditions according to [ITU-T P.1110], clauses 11.11.6.1, 11.11.7.1 or [ITU-T P.1100], clauses 11.11.6.1, 11.11.7.1, with a high amount of reproducibility, flexibility, and time efficiency. It detaches the room characteristics from the speech signal's content by normalized least mean squares (NLMS) system identification with perfect sweep (PS) input, thus offering a high degree of abstraction. These dynamic room characteristics can then be interactively applied to clean speech test data for time-variant evaluation of subsystems.

Interactive generation of dynamic speech data via convolution is presented here exemplarily for the echo path between REF and S2. In case other transmission paths shall be simulated by this interactive convolution method, signal identifiers have to be adopted accordingly.

Further information about reproducible evaluation of signal enhancement subsystems in dynamic conditions can be found in [b-Jung].

III.2 Excitation signal

The type of excitation signal for a given system identification task is crucial to achieve a high signal-to-noise ratio (SNR) and especially for time-variant scenarios good tracking abilities to trace even highly dynamic processes. Perfect sweeps, as well as all perfect sequences in general, are *perfect* in the sense, that they have an impulse-like autocorrelation function, thus leading to fast convergence.

Design of a PS sequence of length M in the discrete Fourier transform domain is as simple as follows:

$$P(k) = \begin{cases} \exp\left(-j4m\pi k^2 / M^2\right), & 0 \le k \le \frac{M}{2} \\ P^*(M-k), & \frac{M}{2} < k < M \end{cases}, \qquad \text{(III.1)}$$

with stretch factor $m = M/2$, here set to equal energy distribution. For highly time-variant set-ups, filter length and PS sequence length are chosen to $N = M = 256$ (for wideband) and $N = M = 128$ (for narrowband) to ensure fast filter convergence whilst still providing good frequency resolution and guaranteeing periodicity with $N = M$. For stationary set-ups, filter length N and PS sequence length M should be chosen equal to the appropriately chosen impulse response length, indicated by the reverberation time. However, $N = M$ has to be maintained.

Due to the sweeping character of the PS signal, its perfectness, and low crest factor, relatively high amounts of energy can be fed into the system without severe nonlinear distortions, thus leading to a high SNR. Additionally, periodic repetitions are possible without transition artefacts, which further allows to increase the excitation energy. If the system to be identified is undermodelled in terms of impulse response length, the tail of the estimated impulse response cut off after N samples will be projected as *systematic* error at the beginning of the estimated impulse response. This systematic error appears to be more forgiving in terms of audio degradation as opposed to an *unsystematic*, as it is observed for noise-like excitation signals.

However, the necessity of a short filter length N and PS sequence length M to ensure a good convergence behaviour of the identification algorithm also brings along the situation, that in an undermodelled set-up, care has to be taken to avoid overfitting of the device under test (DUT) in terms of filter length. This means, that the DUT's filter length $\widetilde{N}$ should not necessarily be chosen to the same value as N (which may give overrated results in this matched state). The choice of $\widetilde{N}$ should rather be driven by the reverberation time of the acoustic environment of the DUT. A guiding number for the filter length of the DUT in an automotive environment could be $32\ \mathrm{ms}$, corresponding to $\widetilde{N} = 512$ (wideband) or $\widetilde{N} = 256$ (narrowband).

III.3 Measurement Set-up

The time-domain PS excitation signal $p(n)$ with sample index n is normalized to 0 $\mathrm{dB_{ov}}$, fed into test point R2, and played back at a preferably high loudness over the loudspeaker (LSP) to achieve a good SNR at the receiving microphone (MIC). In accordance to [ITU-T P.1100] and [ITU-T P.1110], a reflecting surface of size 0.3 m × 0.4 m is placed in the centred position at the co-driver's seat and rotated with $\omega \approx {}^{360°}/_{4\ \mathrm{s}}$ to generate a time-varying echo path between LSP and MIC. The initial position ($\Phi = 0°$) of the board hereby corresponds to the set-up, where its surface is orthogonal to the vehicles windshield. The driver's seat has been occupied. The microphone signal is accessed at test point S2 and stored for later usage.

III.4 System model

The system model is based on the well-known set-up of a system identification process, where the excitation signal $p(n)$ is radiated over the loudspeaker(s) LSP into the acoustical environment to be identified represented as linear time-variant impulse response $\boldsymbol{h}(n) = [h_0(n), h_1(n), \dots, h_{N-1}(n)]^T$ with transpose operator $[\cdot]^T$ thus forming a system output signal $d(n)$. Superimposed at the microphone position with the observation noise signal $n(n)$ the resulting microphone signal $y(n)$ is subject to subtraction by an estimated system output signal $\hat{d}(n) = \widehat{\boldsymbol{h}}^H(n)\boldsymbol{x}(n)$ with $\widehat{\boldsymbol{h}}(n)$ being an estimated replica of the linear system $\boldsymbol{h}(n)$, $(\cdot)^H$being the Hermitian operator, and $\boldsymbol{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$. The resulting error signal $e(n) = y(n) - \hat{d}(n)$ then is to be minimized by the adaptive filter $\widehat{\boldsymbol{h}}(n)$, in our case by making use of the NLMS algorithm, which iteratively updates the replica system's impulse response according to

$$\widehat{\boldsymbol{h}}(n+1) = \widehat{\boldsymbol{h}}(n) + \mu\ \frac{\boldsymbol{x}(n)e^*(n)}{\|\boldsymbol{x}(n)\|^2}, \qquad \text{(III.2)}$$

with complex conjugate operator $(\cdot)^*$ and step size $\mu = 1$, thus being in the range of $0 < \mu < 2$ for stable operation.

III.5 Signal convolution

The time-variant speech test data is generated by convolution of $x(n)$, from test point REF, with the dynamic impulse response $\widehat{\boldsymbol{h}}(n)$ from the previous step[2], mixed with (convolved) near-end speech $s(n)$ and (convolved) noise $n(n)$ to yield $y(n)$. Care has to be taken to adequately choose the levels of the different signal components to guarantee the desired SNR and speech level ratio.

The mixed signal $y(n)$ is then fed into the subsystem via S2. In addition, the far-end signal has to be fed into the subsystem via R3.

---

[1] If the convolved signal shall be explicitly set to a specific loudness level later on, the impulse response vector may be normalized by its Euclidean norm to prevent a loudness change due to the convolution with the impulse response.

Bibliography

[b-ITU-T P.51] Recommendation ITU-T P.51 (1996), *Artificial mouth*.

[b-ETSI ES 202 396-1] ETSI ES 202 396-1 V1.5.1 (2014), *Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database*.

[b-ETSI ES 202 396-3] ETSI ES 202 396-3 V1.3.2 (2010), *Speech and multimedia Transmission Quality (STQ); Speech Quality performance in the presence of background noise; Part 3: Background noise transmission – Objective test methods*.

[b-ETSI ES 202 739] ETSI ES 202 739 V1.3.2 (2010), *Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user*.

[b-ETSI ES 202 740] ETSI ES 202 740 V1.3.1 (2009), *Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user*.

[b-IEC 60268-4] IEC 60268-4 (2004), *Sound system equipment – Part 4: Microphones*.

[b-Berger] Berger, J. (12-2000), *Results of objective speech quality assessment including receiving terminals using the advanced TOSQA2001*, ITU-T Contribution, COM 12-20-E.

[b-Fingscheidt] Fingscheidt, T., and Suhadi, S. (2007), *Quality Assessment of Speech Enhancement Systems by Separation of Enhanced Speech, Noise, and Echo*, INTERSPEECH 2007, Antwerpen, Belgium.

[b-Fodor] Fodor, B.; Fingscheidt, T. (2012), *Reference-free SNR Measurement for Narrowband and Wideband Speech Signals in Car Noise*, in Speech Communication; 10. ITG Symposium; Proceedings of, pp.1-4, 26-28 Sept.

[b-Jung] Jung M.-A., Richter L., Fingscheidt T. (2013), *Towards Reproducible Evaluation of Automotive Hands-free Systems in Dynamic Conditions*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, Canada, May.

[b-Kettler/Gierlich] Kettler, F., Gierlich, H.W. (2008), *Evaluation of Hands-Free Terminals*, in Speech and Audio Processing in Adverse Environments, edited by E. Hänsler, G. Schmidt, Springer, ISBN: 978-3-540-70601-4.

[b-Sottek] Sottek, R., Genuit, K. (2005), *Models of signal processing in human hearing*, International Journal of Electronics and Communications, pp. 157-165.

[b-Steinert] Steinert, K., Suhadi, S., and Fingscheidt, T. (2009), *A Comparison of Instrumental Measures for Wideband Speech Quality Assessment of Hands-free Systems in Echoic Condition*, DAGA 2009, Rotterdam, The Netherlands.

[b-Yu_1] Yu H. and Fingscheidt T. (2012), *A Weighted Log Kurtosis Ratio Measure for Instrumental Musical Tones Assessment in Wideband Speech*, 10th ITG Fachtagung Sprachkommunikation, accepted for publication, Braunschweig, Germany, Sep.

[b-Yu_2] Yu H. and Fingscheidt T. (2012), *Instrumental Musical Tones Measurement of Arbitrary Noise Reduction Systems*, in Proc. of DAGA'12, pp. 255-256, Darmstadt, Germany, Mar.

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	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2015