International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1120
(03/2017)

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

Communications involving vehicles

Super-wideband and fullband hands-free communication in motor vehicles

Recommendation ITU-T P.1120

ITU-T

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
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
Methods for objective and subjective assessment of speech and video quality	Series	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 speech and video	Series	P.1500

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

Recommendation ITU-T P.1120

Super-wideband and fullband hands-free communication in motor vehicles

Summary

Recommendation ITU-T P.1120 will provide guidance to systems architects, developers, and test engineers of integrated automotive speakerphones that operate around super-wideband (50-14,000Hz) and/or fullband (20-20,000Hz) audio bandwidths.

Guidance is provided on the system level (i.e., mouth/ear-to-network reference point).

Subjective (i.e., perceptual) testing for validating a design under real life condition is provided in this Recommendation.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1120	2017-03-01	12	11.1002/1000/13176

Keywords

Automotive, fullband, FB, hands-free, motor vehicle, quality of service, QoS, speakerphone, speech quality, super-wideband, SWB, voice quality.

* 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

Recommendation ITU-T P.1120

Super-wideband and fullband hands-free communication in motor vehicles

1 Scope

This Recommendation provides speech transmission performance and design guidance for automotive hands-free systems that operate around super-wideband (50-14,000Hz) and/or fullband (20-20,000Hz) audio bandwidths and in the future will support stereo telephony. This Recommendation covers:

– built in hands-free systems,

– aftermarket hands-free car kits,

– corded headsets, and

– wireless headsets,

to be used in motor vehicles for communication.

For testing purposes, the test set-up and the recommended environmental conditions are described.

The methods, the analysis and the performance parameters described in this Recommendation are based on test signals and test procedures, as defined in [ITU-T P.501], [ITU-T P.502], [ITU-T P.340], [ITU-T P.380], and [ITU-T P.581].

This Recommendation takes into account the concept of different acoustic zones targeted to different user positions in the car, other than just the driver's position.

Guidance is provided on the system level (i.e., mouth/ear-to-network reference point). For testing purposes, the test set-up and the recommended environmental conditions are described.

Subjective (i.e., perceptual) testing for validating a design under real life condition is provided in this Recommendation.

2 References

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

[ITU-T P.57] Recommendation ITU-T P.57 (2009), *Artificial ears*.

[ITU-T P.58] Recommendation ITU-T P.58 (1996), *Head and torso simulator for telephonometry*.

[ITU-T P.79] Recommendation ITU-T P.79 (2007), *Calculation of loudness ratings for telephone sets*.

[ITU-T P.340] Recommendation ITU-T P.340 (2000), *Transmission characteristics and speech quality parameters of hands-free terminals*.

[ITU-T P.380] Recommendation ITU-T P.380 (2003), *Electro-acoustic measurements on headsets*.

[ITU-T P.501] Recommendation ITU-T P.501 (2009), *Test signals for use in telephonometry*.

[ITU-T P.502] Recommendation ITU-T P.502 (2000), *Objective test methods for speech communication systems using complex test signals*.

[ITU-T P.581] Recommendation ITU-T P.581 (2000), *Use of head and torso simulator for hands-free terminal testing*.

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

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

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

[ITU-T P.831] Recommendation ITU-T P.831 (1998), *Subjective performance evaluation of network echo cancellers*.

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

[ITU-T P.840] Recommendation ITU-T P.840 (2003), *Subjective listening test method for evaluating circuit multiplication equipment*.

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

[ITU-T P.863.1] Recommendation ITU-T P.863.1 (2013), *Application guide for Recommendation ITU-T P.863*.

[IEC 61260] IEC 61260 (1995), *Electroacoustics – Octave-band and fractional-octave-band filters*.

[ISO 1999] ISO 1999:1990, *Acoustics – Determination of occupational noise exposure and estimation of noise-induced hearing impairment*.

[ISO 3745] ISO 3745:2003, *Acoustics – Determination of sound power levels of noise sources using sound pressure – Precision methods for anechoic and hemi-anechoic rooms*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

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

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

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

3.2.4 diffuse-field equalization: Equalization of the head and torso simulator (HATS) sound pick-up, equalization of the difference, in dB, between the spectrum level of the acoustic pressure at the ear-drum 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].

3.2.5 ear-drum reference point (DRP): A point located at the end of the ear canal, corresponding to the ear-drum position.

3.2.6 free-field 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.

3.2.7 free-field reference point: A point located in the free sound field, at least at 1.5 m distance 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).

3.2.8 fullband speech: Voice service with enhanced quality compared to pulse code modulation (PCM) [b-ITU-T G.711] and allowing the transmission of a vocal frequency range of at least 50 Hz to 16 kHz.

3.2.9 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 [b-ITU-T P.51].

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

3.2.11 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.

3.2.12 headset hands-free terminal: A device that includes telephone receiver and transmitter, which is typically secured to the head or ear of the wearer.

3.2.13 inboard ear: Ear closest to the centreline of the vehicle.

3.2.14 maximum setting of the volume control: When a receive volume control is provided, the maximum setting of the volume control is chosen.

NOTE – The maximum volume should be carefully chosen in order to provide sufficient loudness for typical driving conditions but not to overload the audio system and introduce non-linearities in the echo path.

3.2.15 MOS-LQO (mean opinion score – listening-only quality objective): The score calculated by means of an objective model which aims at predicting 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. For further information about MOS terminology, see Annex A.

3.2.16 MOS-TQO (mean opinion score – talking quality objective): The score 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.

3.2.17 motor vehicle: Any vehicle equipped with a motor where a hands-free system can be installed.

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

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

3.2.20 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.)

3.2.21 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.)

3.2.22 short-range wireless transmission (SRW): Radio transmission link used to connect the hands-free system typically to a mobile phone, which is connected to the mobile network.

3.2.23 speakerphone hands-free terminal: A device that is not attached to the user and does not require the use of hands for holding a telephone receiver and transmitter during communication.

3.2.24 super-wideband speech: Voice service with enhanced quality compared to pulse code modulation (PCM) [b-ITU-T G.711] and allowing the transmission of a vocal frequency range of at least 80 Hz to 12.5 kHz.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR	Absolute Category Rating
A/D	Analogue/Digital
AGC	Automatic Gain Control
$A_{H,R}$	Attenuation Range in Receive direction
$A_{H,R,dt}$	Attenuation Range in Receive direction during Double Talk
$A_{H,S}$	Attenuation Range in Send direction
$A_{H,S,dt}$	Attenuation Range in Send direction during Double Talk
BGN	Background Noise
CDMA	Code Division Multiple Access
CSS	Composite Source Signal
D/A	Digital/Analogue
DRP	ear-Drum Reference Point
DTX	Discontinuous Transmission
DUT	Device Under Test
EC	Echo Cancellation
ERL	Echo Return Loss
ERP	Ear Reference Point
FB	Fullband
FFT	Fast Fourier Transform
HATS	Head and Torso Simulator
HATS-HFRP	Head And Torso Simulator – Hands-Free Reference Point
HF	Hands-Free
HF System	Hands-Free System
HFT	Hands-Free Terminal

HVAC	Heating, Ventilation and Air Conditioning
JLR	Junction Loudness Rating
$L_{R,min}$	minimum activation Level (Receive direction)
$L_{S,min}$	minimum activation Level (Send direction)
MOS	Mean Opinion Score
MOS-LQOs	Mean Opinion Score-Listening-only Quality, super-wideband
MOS-LQOf	Mean Opinion Score-Listening-only Quality, fullband
MPNS	Multi-Point Noise Simulation
MRP	Mouth Reference Point
NC	Noise Criterion
NR	Noise Reduction
PCM	Pulse Code Modulation
POI	Point of Interconnection
QoS	Quality of Service
RF	Radio Frequency
RLR	Receive Loudness Rating
SLR	Send Loudness Rating
SNR	Signal-to-Noise Ratio (also S/N)
SRW	Short-Range Wireless transmission
SWB	Super-Wideband
S/N	Signal-to-Noise ratio (also SNR)
TCL	Terminal Coupling Loss
TCLw	weighted Terminal Coupling Loss
T_{as}	Access-Specific delay
T_{ass}	Access-Specific delay in Send
T_{asr}	Access-Specific delay in Receive
T_{imps}	Implementation dependent send delay HFT
T_{impr}	Implementation dependent receive delay HFT
T_r	Receive delay HF terminal
$T_{r,R}$	built-up Time (Receive direction)
$T_{r,S}$	built-up Time (Send direction)
T_{rtd-}	Round Trip Delay HF terminal
T_{rtdimp}	Implementation dependent round trip delay HFT
Ts	Send delay HF Terminal

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).

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

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; (94 dBSPL=0 dBPa).

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 non-linear signal processing by the human ear.

6 How to use this Recommendation

6.1 Determining compliance with this Recommendation

To claim compliance with a particular performance class in this Recommendation, the following must be true:

- The hands-free terminal (HFT) passes all of the requirements using the test procedures specified.
- The HFT passes all noise related requirements for each of the user scenarios defined in Table D.1.
- If the HFT is intended to be used in multiple vehicles (e.g., after market hands-free car kits), then the HFT must meet the above criteria on a minimum of three vehicles that are representative (e.g., microphone type/placement, noise, etc.) of all vehicles before compliance is claimed.

6.2 Testing at different stages of the development cycle

The different subsystems and their testing will be addressed in a future revision of [b-ITU-T P.1130].

6.3 Application of this Recommendation to acoustic zoning concepts

The idea of acoustic zones in cars is based on the assumption that in specifically designed zones different acoustically separated areas can be created where different services can be provided. For example, different audio/video presentation or different calls can be made within the same car without impairing each other.

This Recommendation does not deal with the performance evaluation of the acoustical zone quality and with the quality or efficiency of zone separation concepts.

In order to apply this Recommendation to zoning concepts the acoustical zone in which hands-free communication is provided is defined by the manufacturer. All requirements of this Recommendation then apply to this zone. If the zone defined is not the drivers position, the head and torso simulator (HATS) is positioned on each seat of the zone defined.

In cases where multiple acoustic zones for hands-free communication are provided this Recommendation applies to each zone individually one at a time. The simultaneous operation of multiple zones is for further study.

Examples for possible zones in different car set-ups are given in Figure 6-1.

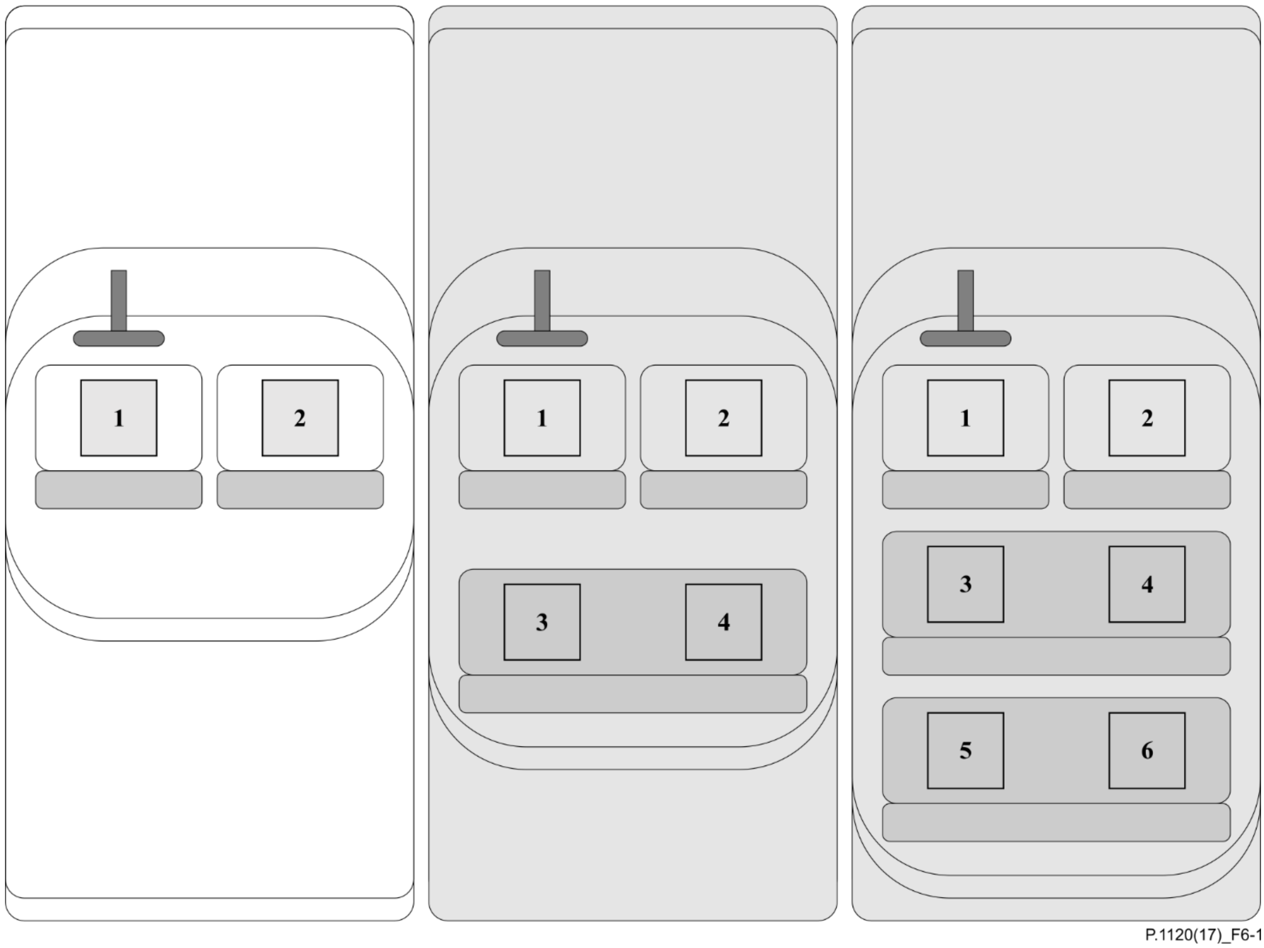


Figure 6-1 – Examples of possible zones in vehicles

7 Test arrangement

The acoustical interface for all hands-free terminals is realized by using an artificial head (HATS – head and torso simulator) according to [ITU-T P.58]. The properties of the artificial head shall conform to [ITU-T P.58] for send as well as for receive acoustical signals.

All hands-free terminals are connected to a system simulator conforming to the required transmission standard with implemented, calibrated audio interface. For some requirements in this Recommendation, the performance limits depend on the transmission system and the speech codec used in this transmission system. The corresponding tables are to be found in each clause. Table 7-1 provides an overview of the speech codecs used for the tests.

NOTE – It is recognized that there are limits on the speech frequencies that current artificial mouths can reproduce, so the capabilities of current test equipment will need to be verified during testing for the bandwidth under investigation.

Table 7-1 – Overview of speech codecs used

System	Codec
VoLTE (IMS)	EVS, 24.4 kbit/s in SWB (32 kHz sampling rate), 13.2 kbit/s CA (32 kHz sampling rate), 24.4 kbit/s in FB (48 kHz sampling rate)
LTE (OTT)	OPUS
LTE (OTT)	ITU-T G.722.1c
LTE (OTT)	ITU-T G.719
LTE (OTT)	SILK

The settings of the system simulator shall be chosen so that the audio signal is not influenced by any signal processing (e.g., DTX, codec bitrate).

The test signals are fed electrically to the system simulator or acoustically to the artificial head. The test arrangement is shown in Figure 7-1.

NOTE 1 –In order to take into account "real life" conditions, bitrates used in the real network should be used for testing and optimization as long as they do not interfere with performance measurements of the hands-free (HF) system in the vehicle.

NOTE 2 – For some mobile phones used in the hands-free set-up, the signal processing cannot be switched off completely. Therefore, care should be taken to use only such phones for the tests which do not introduce additional speech signal processing.

NOTE 3 – The tests are conducted without additional network load such as other data traffic in parallel.

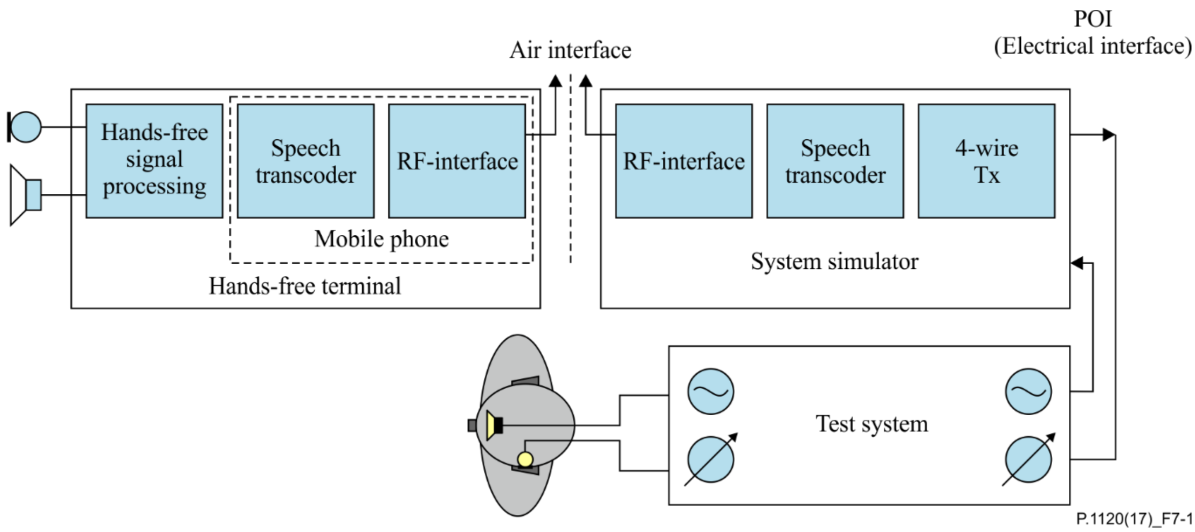


Figure 7-1 – Test arrangement for hands-free terminal

7.1 Test arrangement in a car

7.1.1 Microphone related simulation

The transmission performances of car hands-free terminals are measured in a car cabin. In order to simulate a realistic driving situation, background noise (BGN) is inserted using a 4-loudspeaker arrangement with subwoofer while measurements with background noise are conducted. This method is not a real sound-field reproduction but a simplified method mainly targeted to single microphone solutions. In Figure 7-2, the simulation arrangement is shown. The test arrangement conforms to [b-ETSI ES 202 396-1]. The source signal used is recorded by a measurement microphone positioned close to the hands-free microphone. If possible, the output signal of the hands-free microphone can be used directly. The recordings are conducted in a real car. The loudspeaker arrangement is equalized and calibrated so that the power density spectrum measured at the microphone position is equal to the recorded one. For equalization, either the measurement microphone or the hands-free microphone used for recording is used. The maximum deviation of the A-weighted sound pressure level shall be ±1 dB. The third octave power density spectrum between 50 Hz and 10 kHz shall not deviate more than ±3 dB from the original spectrum. A detailed description of the equalization procedure as well as a database with background noises can be found in [b-ETSI ES 202 396-1].

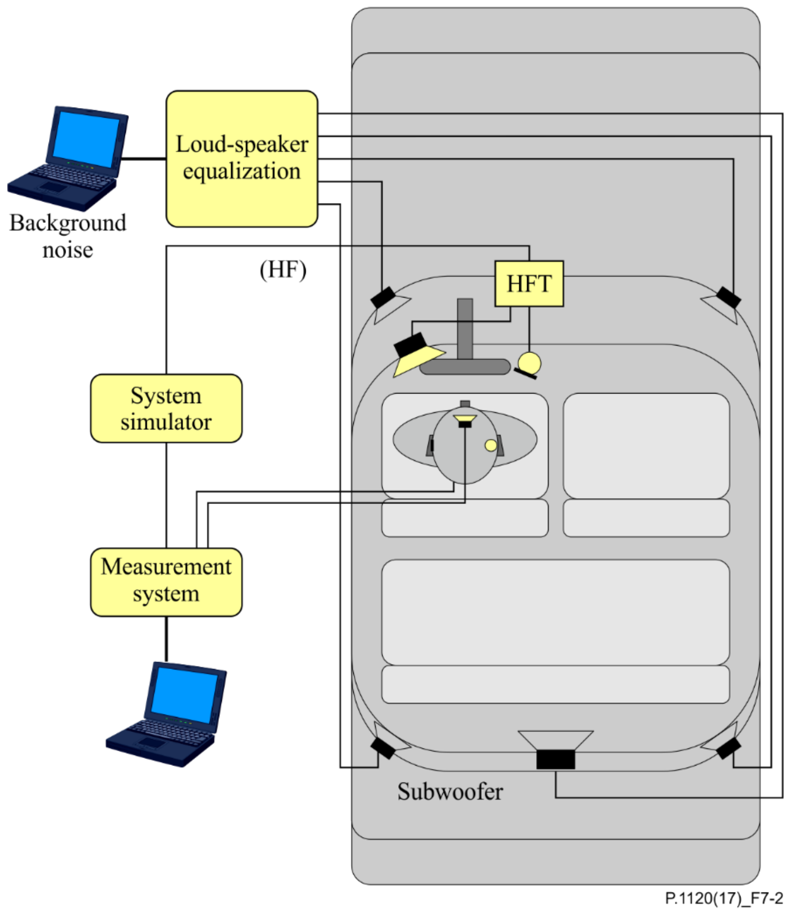


Figure 7-2 – Test arrangement with background noise simulation

For microphone arrays, distributed microphones or directional microphones a more sophisticated background noise simulation technology is recommended. This method is based on a sound-field simulation technique described in [b-ETSI TS 103 224]. Using this technology a more accurate equalization can be performed and the sound-field generated can be reproduced correctly in magnitude and phase up to a frequency of 2-3 kHz. A detailed description on the application of the method can be found in Annex F.

The background noise playback system is time-synchronized to the recording process in the measurement system in order to guarantee reproducibility of recordings in the presence of noise.

7.1.2 Positioning of the hands-free terminals

The speakerphone hands-free terminal is installed according to the requirements of the manufacturers. The positioning of the microphone/microphone array and loudspeaker are given by the manufacturer. If no position requirements are given, the test lab will fix the arrangement. The exact positioning has to be noted. Hands-free terminals installed by the car manufacturer are measured in the original arrangement.

If not stated otherwise, the artificial head (HATS – head and torso simulator, according to [ITU-T P.58]) is positioned in the driver's seat for the measurement. The position has to be in line with the average user's position; therefore, all positions and sizes of users have to be taken into account. Typically, all except the tallest 5% and the shortest 5% of the driving population have to be considered. Detailed information can be found in [b-SAE J941]. The size of these persons can be derived, e.g., from the 'anthropometric data set' for the corresponding year (e.g., based on data used by the car manufacturers). The position of the HATS (mouth/ears) within the positioning arrangement is given individually by each car manufacturer. The position used has to be reported in detail in the test report. If no requirements for positioning are given, the distance from the microphone to the MRP is defined by the test lab.

By using suitable measures (marks in the car, relative position to A-, B-pillar, height from the floor, etc.) the exact reproduction of the artificial head position must be possible at any later time.

NOTE – Different positions of the artificial head may highly influence the test results. Depending on the application, different positions of the artificial head may be chosen for the tests. It is recommended to check the worst case position, e.g., those positions where the SNR and/or the speech quality in the send direction may be worse.

7.1.3 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].

In the case of speakerphone hands-free terminals, the sound pressure level is calibrated at the head and torso simulator – hands-free reference point (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].

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]. This relationship is shown in the following formula:

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

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 near the driver's head position. The 3 dB speech level increase according to [ITU-T P.340] and applicable for all hands-free tests in send direction is taken into account independently (see section 9 on test signal levels).

7.1.4 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 clause 7.1.6.1), more detailed information can be found in [ITU-T P.581].

7.1.5 Influence of the transmission system

Measurements may be influenced by signal processing (different speech codecs, DTX, comfort noise insertion), depending on the transmission system and the system simulator used in the test set-up. If requirements cannot be fulfilled due to impairments introduced by the transmission system or the system simulator, reference measurements of the hands-free unit or measurements without acoustical components should be made documenting this behaviour.

7.1.6 Calibration and equalization

The following preparation has to be completed before running the tests:

7.1.6.1 Calibration

- Acoustical calibration of the measurement microphones as well as of HATS microphone.
- Calibration and equalization of the artificial mouth at the MRP.
- HATS-HFRP calibration.
- Diffuse-field equalization of the artificial head is used.

7.1.6.2 Reference measurement

For the compensation of the different power density spectra of the measurement signals, it is required to refer the measured power density spectra to the power density spectra of the test signal. This is denoted as a reference measurement.

- In send direction, the reference spectrum is recorded at the MRP.
- In receive direction, the reference spectrum is recorded at the electrical interface.

7.1.7 System simulator settings

All settings of the system simulator have to ensure that the audio signal is not disturbed by any processing and the transmission of the HF signal is error-free. Discontinuous transmission (DTX) shall be switched-off. For all networks, the RF-level shall be set to maximum. The settings shall be reported in the test protocol.

7.1.8 Environmental conditions

Unless specified otherwise, the background noise level in the vehicle at all measurement locations shall be less than –54 dBPa(A) in conjunction with NC40 [ISO 3745].

For specified tests, it is desirable to have a background noise level of less than –74 dBPa(A) in conjunction with NC20, but the background noise level of –64 dBPa(A) in conjunction with NC30 shall never be exceeded.

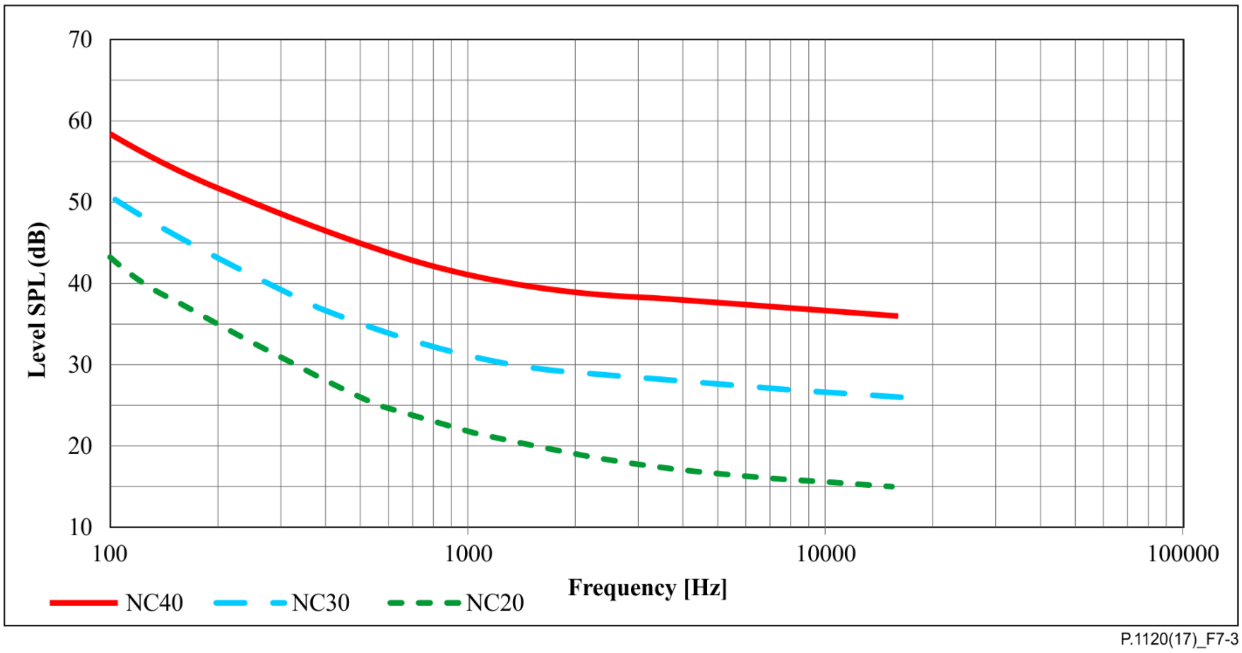


Figure 7-3 – NC-criteria for test environment (values above 8 kHz extrapolated)

7.1.8.1 Verification of environmental conditions

This test is intended to be used in order to verify the environmental conditions as defined in this Recommendation.

1) For the measurements no test signal is used.
2) A free-field measurement microphone is positioned in the test room outside the car.
3) The room noise is measured in the frequency range between 20 Hz and 20 kHz. The measurement duration is 5 seconds which is the averaging time for the idle channel noise.

   The power density spectrum of the noise signal is determined using fast fourier transform (FFT) (8 k samples/48 kHz sampling rate or equivalent). A Hanning window is used.
4) For checking the room noise level the measured spectrum is A-weighted.
5) For checking the NC-criteria the octave levels of the room noise are determined from 63 Hz to 16 kHz.

8 Test signals and test signal levels

8.1.1 Signals

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 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, a shorter sequence consisting of two sentences may be used in cases where it can be shown that the hands-free 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]). In such 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 case composite source signal (CSS), according to [ITU-T P.501], is used, shaping of the wideband CSS spectrum is applied. The shaping response characteristics described in Figure 7-10 of [ITU-T P.501] is applied but with extension of the 5 dB/oct. shaping response characteristics from 4 kHz to 14 kHz, 16kHz or 20 kHz respectively.

For super-wideband hands-free terminals, all test signals – which are used in receive direction – have to be band-limited. The band limitation is achieved by bandpass filtering in the frequency range between 50 Hz and 14 kHz (16 kHz in case of EVS codec in SWB mode) using bandpass filtering providing 24 dB/octave. For fullband hands-free terminals, all test signals which are used in receive direction – are band-limited to 20 kHz using low pass filtering providing 24 dB/octave. In send direction, the test signals are used without band limitation.

NOTE – The ITU-T definition of super-wideband is 50 to 14 kHz. The 3GPP definition of super-wideband is 50 to 16 kHz. Consequently, if the EVS codec is used the test signal bandwidth has to be extended up to 16 kHz.

All test signal levels are referred to the average level of the test signals, averaged over the complete test sequence length, if not described otherwise. In receive direction, the band-limited test signal is measured, in send direction no band-limitation is applied.

The average signal levels for the measurements are as follows:

– –16 dBm0 in receive direction (typical signal level in networks);
– –1.7 dBPa in the send direction at the MRP for **speakerphone hands-free** terminals (typical average speech levels) (equivalent to –25.7 dBPa at the HATS-HFRP).

– –4.7 dBPa in send direction at the MRP (typical average speech levels) (equivalent to –28.7 dBPa at the HATS-HFRP). This level applies for **headset hands-free** terminals.

NOTE – If different networks signal levels are to be used in a test, this is stated in the individual test. The "Lombard Effect" (increased talker speech level due to high background noise) is considered in background noise tests.

Some tests require exact synchronization of test signals in the time domain. Therefore, it is required to take into account the delays of the terminals. When analysing signals, any delay introduced by the test system codecs and terminals have to be taken into account accordingly.

8.1.2 Background noise signals

For background noise tests the background noise scenarios defined in Annex D apply. The background noise signals representing the different scenarios are recorded individually for each target vehicle.

8.1.2.1 Recording of driving noise

Background noise is recorded in the real car. The measurement microphones are positioned close to the hands-free microphone(s). Alternatively, the hands-free microphones can be used for the recording of the background noise if the microphone is easily accessible.

NOTE – In case of microphone arrays, the simple background noise simulation technology based on [b- ETSI ES 202 396-1] is not recommended. For microphone arrays it is recommended to use the background noise recording technology as described in Annex F or to record the electrical output signals of all microphones and insert them electrically as described below. With this methodology, structure-borne noise and wind noise coupled to the microphone can also be included.

Background noise recordings are collected from the vehicle being tested and used in noise related tests. Table D.1 lists the standard set of user scenarios under which noise related requirements must be tested to be considered compliant with this Recommendation. These user scenarios are important because they define what it means to be compliant, ensure that performance is tested for some common usage scenarios, and allow reasonable comparisons across vehicle platforms. If the main goal of testing is to directly compare different hands-free systems, then it is important to more tightly control the experimental variables listed in Table D.1 (e.g., use identical vehicles, identical routes for noise collection, identical noise recordings for testing different algorithms, etc.).

NOTE – The heating, ventilation and air conditioning (HVAC) (see Table D.1) settings may be automatically adjusted by the car during a hands-free call in order to limit the background noise. For the background noise recordings the HVAC active in a hands-free call shall be used.

8.1.2.2 Playback of the recorded background noise

Two ways of background noise playback are recommended:

1) The test lab employs a 4-loudspeaker arrangement for acoustic background noise reproduction in the car cabin. Typically, two loudspeakers are mounted in the front and in the rear (left and right side). The loudspeaker should be carefully positioned in order to minimize disturbances of the transmission paths between loudspeakers and hands-free microphone and the artificial head at the driver's seat. Details can be found in [b-ETSI ES 202 396-1] and Annex F.

2) Background noise can be inserted into the microphone signal and to the reference microphone positioned close to the hands-free microphone. Recording and insertion of the background noise can be analogue or digital. Background noise signals recorded at the output of the hands-free microphone(s) and at the reference microphone are inserted at the electrical access point which was used for the recording. Appropriate electronics allowing the mix of the previously recorded background noise signal(s) with the microphone signal(s) at this access point has to be provided, see Figure 8-1 as an example for analogue mixing. The test lab has to ensure the right calibration of the two signals.

The type of background noise simulation chosen shall be reported in the test report.

NOTE 1 – Both with analogue as well as digital electrical insertion of the noise signal structure-borne noise can be captured as well.

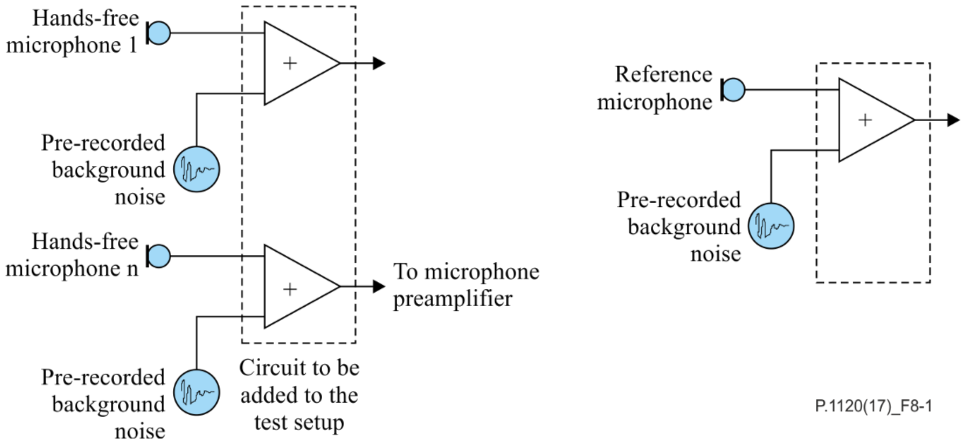


Figure 8-1 – Set-up for insertion of the pre-recorded background noise signal at the hands-free microphone(s) and the reference microphone

9 Measurement parameters and requirements for hands-free terminals

As a general rule for all tests it should be noted that it is always the responsibility of the test lab to verify that measurement results are not influenced by second-order effects such as additional noise falsifying the result of the individual measurement, for example.

Before conducting these tests, proper calibration and equalization of the test system has to be performed.

9.1 Delay

9.1.1 Requirements

In general the delay T_{rtd} consists of an access specific delay, T_{as}, and the implementation dependant delay T_{rtdimp}.

The access-specific specifications define the access specific delays, T_{as}, which have to be taken into account when measuring T_{rtd}.

The HFT implementation dependant delay T_{rtdimp} consists of:

– the HFT signal processing in send and receive

– the air-paths from mouth to microphone and from the loudspeakers to the ear

The HFT delays in the send and receive directions are defined as:

- The HFT delay in the send (uplink) direction Ts is the delay between the first acoustic event at the MRP to the last bit of the corresponding speech frame at the UE antenna.
- The HFT delay in the receive (downlink) direction Tr is the delay between the first bit of a speech frame at the UE antenna and the first acoustic event at the DRP corresponding to that speech frame.

T_{rtd} = Ts + Tr (the delay in the send direction T_s plus the delay in the receive direction T_r) shall be less than 170 ms.

For hands-free implementations using short-range wireless transmission (SRW) in the connection, the round trip delay consist of different components. In such connections T_{as} = $_{TSRW}$. In consequence the roundtrip delay T_{rtd} in such connection consists of the implementation dependent delay and the short range wireless delay: $T_{rtd} = T_{rtdimp} + T_{rtdsrw\ldots}$

In case a SRW connection is used the implementation dependant delay $T_{rtdimp} = T_{imps} + T_{impr}$ shall be less than 70ms.

NOTE 1 – The access specific delays can be found in the relevant specification for the access technologies used and need to be calculated based on the information provided in these specifications. When deriving Tas from these specification it is assumed that any speech signal processing is deactivated and does not introduce any additional delay. For 3GPP UMTS circuit-switched speech and 3GPP LTE MTSI-based speech, definitions, test methods, performance objectives and requirements are found in [b-3GPP TS 26.131] and [b-3GPP TS 26.132].

NOTE 2 – Regarding the user effect of mouth-to-ear delay to the conversational quality in handset mode, guidance is found in ITU-T Recommendation [b-ITU-T G.114].

NOTE 3 – When providing state of the art low delay implementations the delay introduced by the hands-free signal processing T_{rtdimp} should not exceed 70 ms.

9.1.2 Delay in send direction

9.1.2.1 Test

The delay in send direction is measured from the MRP (mouth reference point) to POI (reference speech codec of the system simulator, output). The delay measured in send direction (see Figure 9-1) is:

$$T_s + Tsystem$$

NOTE 1 – The delay should be minimized. This can be accomplished, for example, by designing the speech decoder output, the SRW link, and the hands-free system in a way that sample-based processing and frame-based processing interoperate by using common buffers at their interfaces. Careful matching of frame shift and DFT size for the signal processing in the hands-free system to the SRW link and to the speech coder allows to (partially) embed the delay of one block into the preceding one.

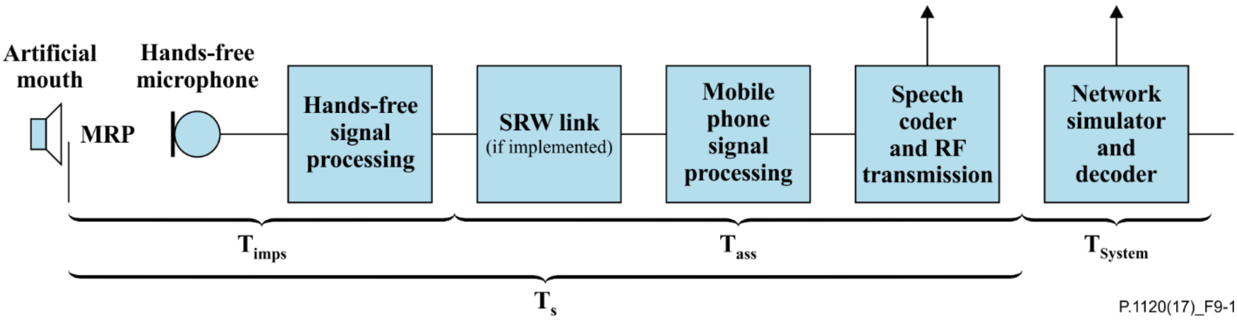


Figure 9-1 – Different blocks contributing to the delay in send direction

The system delay Tsystem depends on the transmission method used and the network simulator. The delay Tsystem must be known.

1) For the measurements, a composite source signal (CSS), according to [ITU-T P.501], is used. The pseudo-random noise (pn)-part of the CSS shall be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples with 48 kHz sampling rate or equivalent. The test signal level is –4.7 dBPa at the MRP, for headset hands-free terminals. For speakerphone hands-free terminals, the test signal level is adjusted to –25.7 dBPa at the HATS-HFRP (see [ITU-T P.581]). The equalization of the artificial mouth is made at the MRP.

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

   The set-up of the hands-free terminal is according to clause 7.1.

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.

9.1.3 Delay in receive direction

9.1.3.1 Test

The delay in receive direction is measured from the point of interconnection (POI) (input of the reference speech coder of the system simulators) to the drum reference point (DRP). The delay measured in receive direction (see Figure 9-2) is:

$$T_r + Tsystem$$

NOTE 1 – The delay should be minimized. This can be accomplished, for example, by designing the speech decoder output, the SRW link, and the hands-free system in a way that sample-based processing and frame-based processing interoperate by using common buffers at their interfaces. Careful matching of frame shift and DFT size for the signal processing in the hands-free system to the SRW link and to the speech coder allows to (partially) embed the delay of one block into the preceding one.

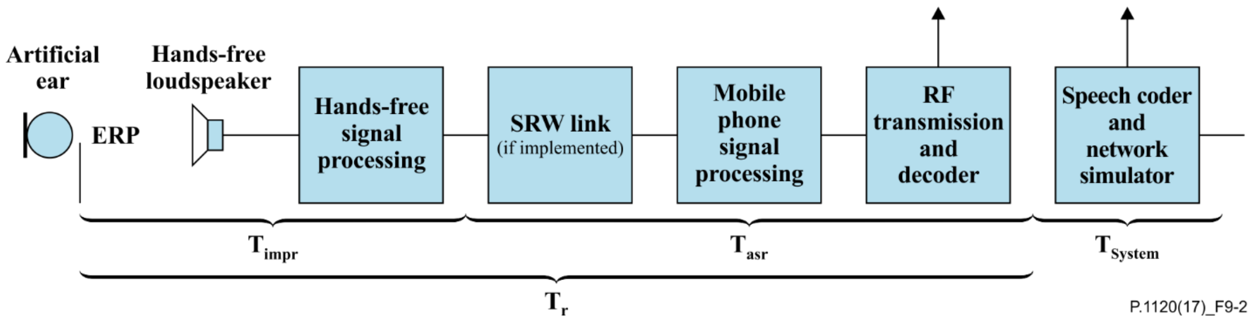


Figure 9-2 – Different blocks contributing to the delay in receive direction

The system delay Tsystem depends on the transmission system and on the network simulator used. The delay Tsystem must be known.

1) For the measurements, a composite source signal (CSS), according to [ITU-T P.501], is used. The pseudo-random noise (pn)-part of the CSS shall be longer than the maximum expected delay. It is recommended to use a pn sequence of 16 k samples with 48 kHz sampling rate or equivalent. The test signal level is –16 dBm0 at the electrical interface (POI).

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

2) The test arrangement is according to clause 7.1. For the measurement of speakerphone hands-free terminals, the artificial head is free-field or diffuse-field equalized according to [ITU-T P.581] (see also clause 7.1.6.1), The equalized output signal of the inboard ear is used for the measurement. For headset hands-free terminals the measured sound pressure is diffuse-field equalized according to [ITU-T P.58].

3) 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.

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

9.2 Loudness ratings

9.2.1 Requirements

The nominal values of send loudness rating/receive loudness rating (SLR/RLR) from/to the electrical reference point (POI) shall be:

For speakerphone hands-free terminals:

SLR = 13 dB ± 4 dB;

RLR = 2 dB ± 4 dB.

For headset hands-free terminals:

SLR = 8 dB ± 4 dB;

RLR = 2 dB ± 4 dB.

For binaural headset hands-free terminals:

SLR = 8 dB ± 4 dB;

RLR (bin) = 8 dB ± 4 dB for each earphone.

If a user-specific volume control is provided, the requirements for RLR given above shall be measured at least for one setting of the volume control. It is recommended to provide a volume control which allows a loudness increase by at least 15 dB referred to the nominal value of RLR. The volume control range shall allow the setting of S/N ≥ 6 dB for all signal and noise conditions. This will allow sufficient loudness of the speech signal in receive direction in the presence of high background noise.

NOTE – It is recognized that the car may be a working place. Therefore, care has to be taken not to exceed the limits for daily noise exposure defined in the different regional standards and directives for working places.

9.2.2 Test loudness rating in send direction

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 equalized at the MRP, the test signal level is –4.7 dBPa at the MRP for headset hands-free terminals. The test signal level is the average level of the complete test signal. For speakerphone hands-free terminals, the level at the HATS-HFRP is adjusted to –25.7 dBPa.

   The measured power density spectrum at the MRP is used as the reference power-density spectrum for determining the send sensitivity.

2) The test arrangement is according to clause 7.1. The 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 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 MRP.

3) The sensitivity is expressed in terms of dBV/Pa, and the SLR shall be calculated according to Annex A of [ITU-T P.79].

9.2.3 Test loudness rating in receive direction

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 –16 dBm0, measured at the electrical reference point and averaged over the complete test signal sequence.

2) The test arrangement is according to clause 7.1. For the measurement of speakerphone hands-free terminals, the artificial head is free-field or diffuse-field equalized (see clause 7.1.6.1) according to [ITU-T P.581]. The equalized output signals of both artificial ears are used for the measurement. The equalized output signal of each artificial ear is power-averaged on the

total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band.

For headset hands-free terminals, the sound pressure is measured at the DRP of the right ear and corrected to the ear reference point (ERP) according to [ITU-T P.57]. The receive sensitivity is determined by the bands 1-20 according to Table A.2 of [ITU-T P.79].

For binaural headset hands-free terminals, the sound pressure is measured separately at the DRP of the right ear and respectively at the left ear and corrected to the ERP, according to [ITU-T P.57]. The receive sensitivity is determined by the bands 1-20 according to Table A.2 of [ITU-T P.79].

In case of headset measurements the tests are repeated five times, in conformance with [ITU-T P.380]. The results are averaged (averaged value in dB, for each frequency).

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 terms of dBPa/V and the RLR shall be calculated according to Annex A of [ITU-T P.79], without the L_E factor.

4) For speakerphone hands-free terminals, the correction 8 dB, according to [ITU-T P.340], is used for the correction of the measurement results.

5) The test is repeated for maximum volume control setting.

9.3 Loudness

Besides loudness ratings the verification of the subjectively perceived loudness is required especially in systems employing automatic gain control (AGC) or companding techniques.

The nominal values of loudness in send/loudness in receive from/to the electrical reference point (POI) shall be:

For further study.

9.3.1 Test loudness in send direction

For further study.

9.3.2 Test loudness in receive direction

For further study.

9.4 Sensitivity frequency responses

9.4.1 Send sensitivity frequency response

9.4.1.1 Requirements

The send sensitivity frequency response is measured from the MRP to the output of the speech codec at the electrical point (output of the system simulators, POI).

The tolerance mask for the send sensitivity frequency responses are given in Tables 9-1 and 9-2, and Figures 9-3 and 9-4 respectively. The mask is drawn by straight lines between the breaking points in tables on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 9-1 – Tolerance mask for the send sensitivity frequency response for SWB

Frequency	Upper limit	Lower limit
80 Hz	4 dB	–10 dB
125 Hz	4 dB	–4 dB
8 000 Hz	4 dB	–4 dB
12 500 Hz	4 dB	–7 dB
15 000 Hz*	4 dB	–7 dB
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. *applicable only in case of EVS		

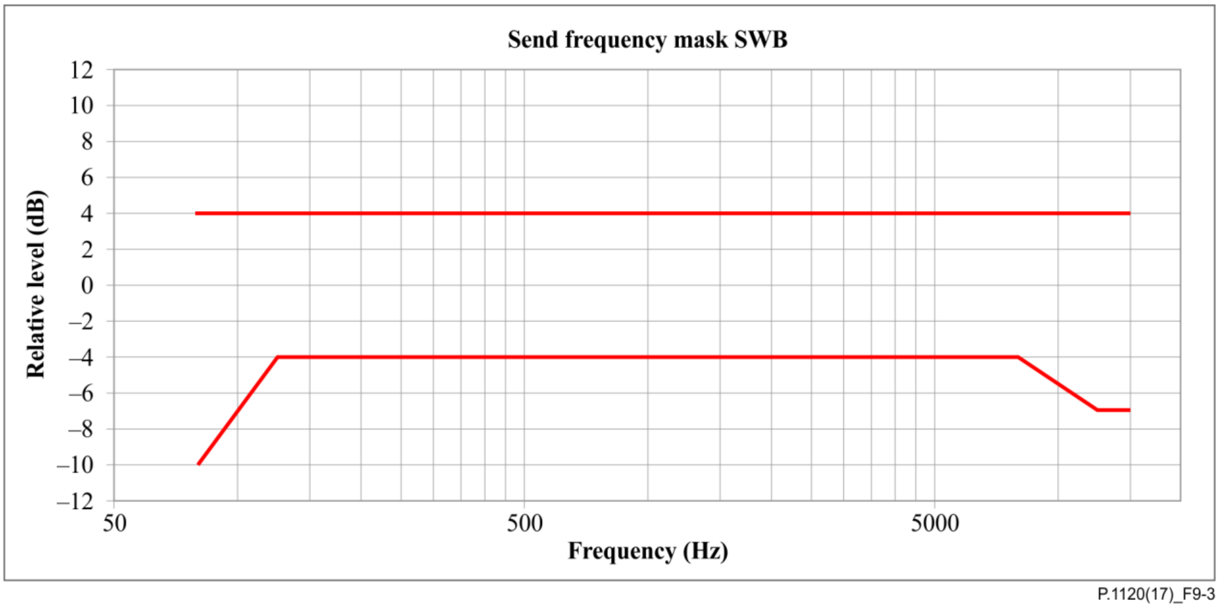


Figure 9-3 – Tolerance mask for the SWB send sensitivity frequency response (Figure is informative)

Table 9-2 – Tolerance mask for the send sensitivity frequency response for FB

Frequency	Upper limit	Lower limit
50 Hz	4 dB	–10 dB
100 Hz	4 dB	–4 dB
12 500 Hz	4 dB	–4 dB
16 000 Hz	4 dB	–7 dB
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 – Ideally, the response characteristics in send should be flat in the frequency range of transmission (80 Hz – 16 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.

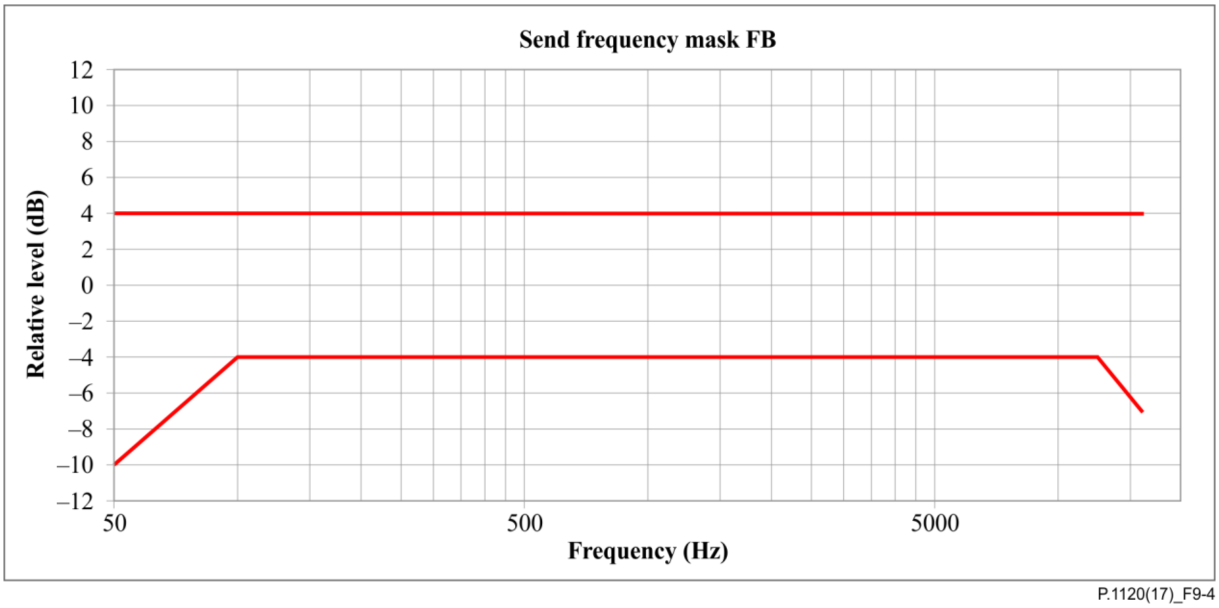


Figure 9-4 – Tolerance mask for the FB send sensitivity frequency response (Figure is informative)

9.4.1.2 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 equalized at the MRP, the test signal level is –4.7 dBPa at the MRP for headset hands-free terminals. The test signal level is the average level of the complete test signal. For speakerphone hands-free terminals, the level at the HATS-HFRP is adjusted to –25.7 dBPa.

   The measured power density spectrum at the MRP is used as the reference power density spectrum for determining the send sensitivity.

2) The test arrangement is according to clause 7.1. The sensitivity frequency response is determined in third octave intervals as given by [IEC 61260] for frequencies of 50 Hz and 14 kHz inclusive for super-wideband (SWB) and 20 kHz for fullband (FB). 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.

9.4.2 Receive sensitivity frequency response

9.4.2.1 Requirements

The receive sensitivity frequency response is measured from the electrical reference point (input of the system simulators, POI) to the DRP when headset hands-free terminals are measured. The HATS is diffuse-field equalized. For speakerphone hands-free terminals, the sound pressure of the free-field or diffuse-field equalized (see clause 7.1.6.1) HATS is measured.

The tolerance mask for the receive sensitivity frequency responses are given in Tables 9-3 to 9-6 and Figures 9-5 to 9-8, the mask is drawn by straight lines between the breaking points in the tables on a logarithmic (frequency) – linear (dB sensitivity) scale.

Table 9-3 – Tolerance mask for the receive sensitivity frequency response for SWB speakerphone hands-free

Frequency	Upper limit	Lower limit
80 Hz	4 dB	–10 dB
125 Hz	4 dB	–4 dB
8 000 Hz	4 dB	–4 dB
12 500 Hz	4 dB	–7 dB
15 000 Hz*	4 dB	–7 dB
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. *applicable only in case of EVS		

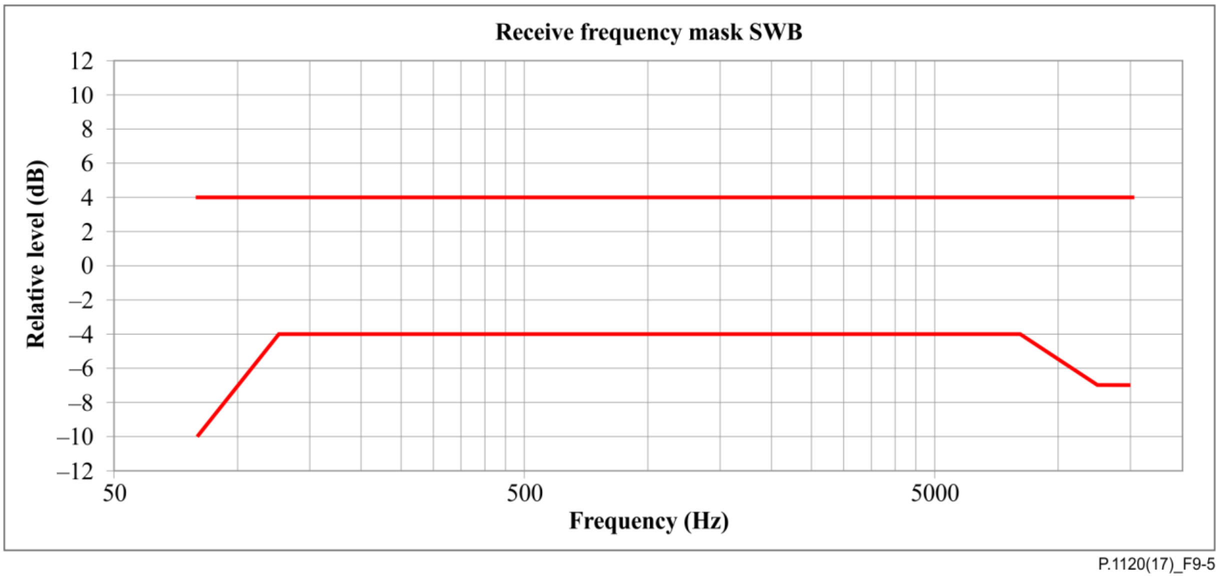


Figure 9-5 – Tolerance mask for the SWB receive sensitivity frequency response (Figure is informative)

Table 9-4 – Tolerance mask for the receive sensitivity frequency response for FB speakerphone hands-free

Frequency	Upper limit	Lower limit
50 Hz	4 dB	–10 dB
100 Hz	4 dB	–4 dB
12 500 Hz	4 dB	–4 dB
16 000 Hz	4 dB	–7 dB
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.		

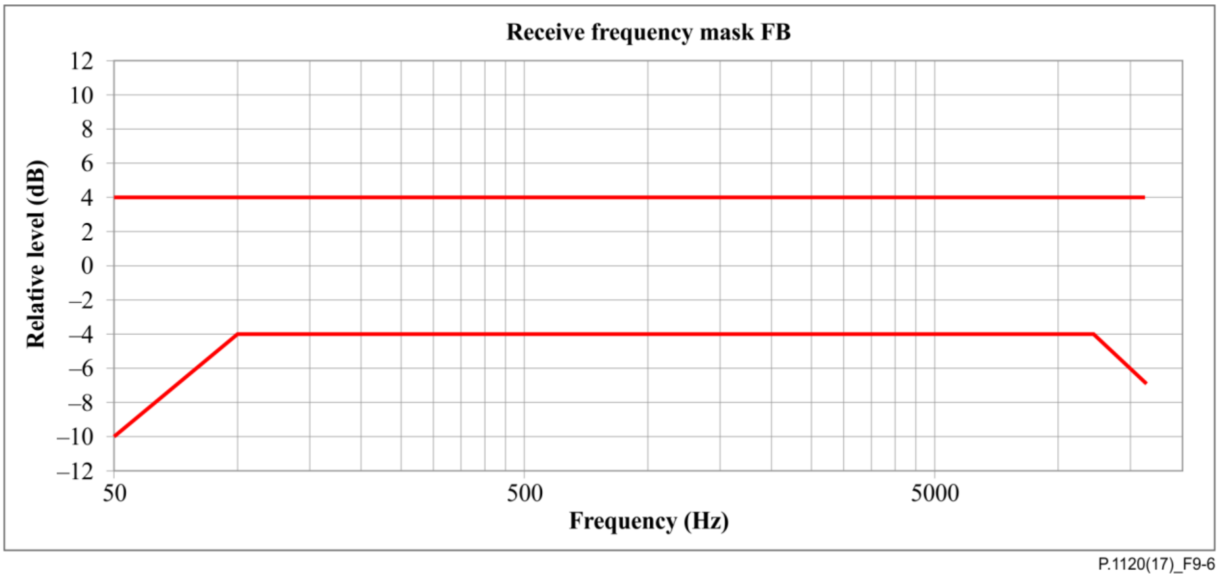


Figure 9-6 – Tolerance mask for the FB receive sensitivity frequency response (Figure is informative)

Table 9-5 – Tolerance mask for the receive sensitivity frequency response for SWB headset hands-free

Frequency	Upper limit	Lower limit
100 Hz	3 dB	
120 Hz	3 dB	–5 dB
200 Hz	3 dB	–5 dB
400 Hz	3 dB	–5 dB
1 010 Hz	(Note 1)	–5 dB
1 200 Hz	(Note 1)	–8 dB
1 500 Hz	(Note 1)	–8 dB
2 000 Hz	9 dB	–3 dB
3 200 Hz	9 dB	–3 dB
12 5000 Hz	9 dB	–12.5 dB
15 000 Hz*	9 dB	–13.5 dB

NOTE 1 – The limit curves shall be determined by straight lines joining successive coordinates given in the table, where frequency response is plotted on a linear dB scale against frequency on a logarithmic scale. The mask is a floating or "best fit" mask.

NOTE 2 – The basis for the target frequency responses in send and receive is the orthotelefonic reference response which is measured between 2 subjects in 1 m distance under free field conditions and is assuming an ideal receive characteristic. Under these conditions, the overall frequency response shows a rising slope. In contrast to other standards, this Recommendation no longer uses the ERP as the reference point for receive but the free-field. With the concept of diffuse-field based receive measurements, a rising slope for the overall frequency response is achieved by a flat target frequency response in send and a diffuse-field based receive frequency response.

*applicable only in case of EVS

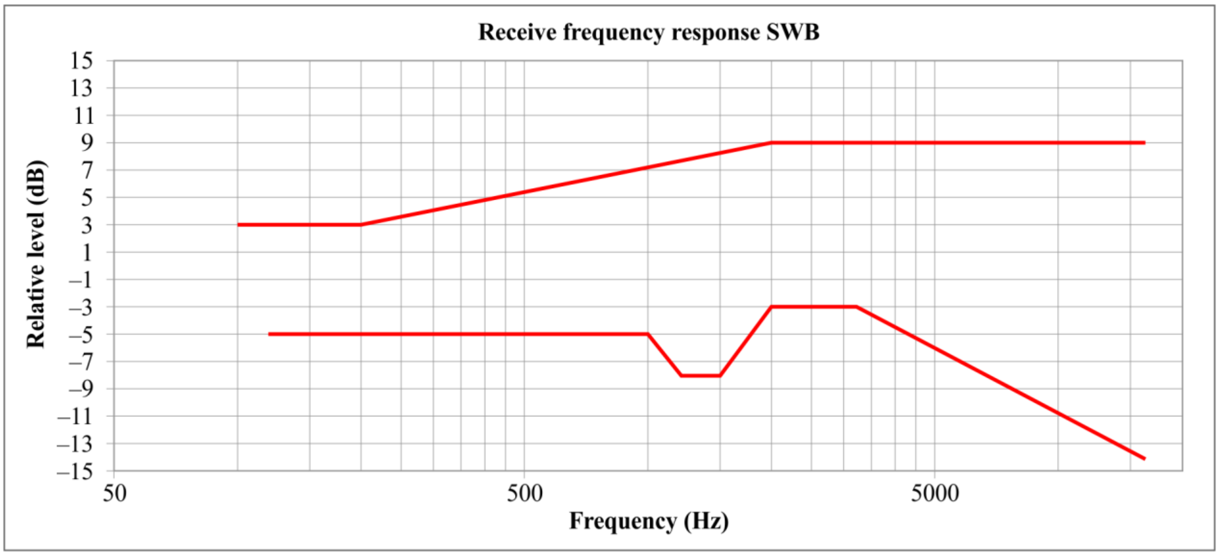


Figure 9-7 – Tolerance mask for the SWB headset receive sensitivity frequency response (Figure is informative)

Table 9-6 – Tolerance mask for the receive sensitivity frequency response for FB headset hands-free

Frequency	Upper limit	Lower limit
100 Hz	3 dB	
120 Hz	3 dB	–5 dB
200 Hz	3 dB	–5 dB
400 Hz	3 dB	–5 dB
1 010 Hz	(Note 1)	–5 dB
1 200 Hz	(Note 1)	–8 dB
1 500 Hz	(Note 1)	–8 dB
2 000 Hz	9 dB	–3 dB
3 200 Hz	9 dB	–3 dB
12 5000 Hz	9 dB	–12.5 dB
16 000 Hz	9 dB	–14 dB
NOTE 1 – The limit curves shall be determined by straight lines joining successive coordinates given in the table, where frequency response is plotted on a linear dB scale against frequency on a logarithmic scale. The mask is a floating or "best fit" mask. NOTE 2 – The basis for the target frequency responses in send and receive is the orthotelefonic reference response which is measured between 2 subjects in 1 m distance under free field conditions and is assuming an ideal receive characteristic. Under these conditions, the overall frequency response shows a rising slope. In contrast to other standards, this Recommendation no longer uses the ERP as the reference point for receive but the free-field. With the concept of diffuse-field based receive measurements, a rising slope for the overall frequency response is achieved by a flat target frequency response in send and a diffuse-field based receive frequency response.		

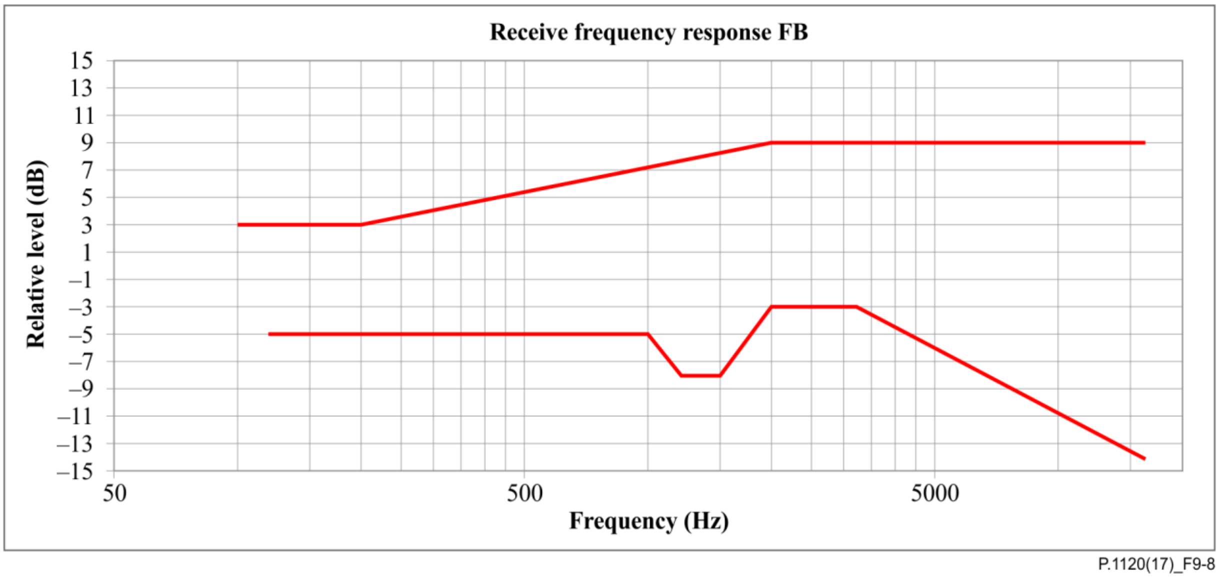


Figure 9-8 – Tolerance mask for the FB headset receive sensitivity frequency response (Figure is informative)

9.4.2.2 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 is –16 dBm0, measured at the electrical reference point and averaged over the complete test signal sequence.

2) The test arrangement is according to clause 7.1. For the measurement of speakerphone hands-free terminals, the artificial head is free-field or diffuse-field equalized (see clause 7.1.6.1) according to [ITU-T P.581]. The equalized output signals of both artificial ears are used for the measurement. The equalized output signal of each artificial ear is power-averaged on the total time of analysis; the "right" and "left" signals are voltage-summed for each 1/3 octave band frequency band.

   For headset hands-free terminals, the sound pressure is measured at the DRP. The HATS is diffuse-field equalized, as described in [ITU-T P.581]. The diffuse-field correction, as defined in [ITU-T P.58], is applied. The receive sensitivity frequency response for SWB is determined in third octaves as given in [IEC 61260] for frequencies from 50 Hz to 14 kHz, inclusive. For FB the upper frequency is 20 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.

   For binaural headset hands-free terminals, the sound pressure is measured separately at the DRP of the right ear and respectively at the left ear. The HATS is diffuse-field equalized, as described in [ITU-T P.581]. The diffuse-field correction, as defined in [ITU-T P.58], is applied. The receive sensitivity frequency response in SWB is determined separately for each earphone in third octaves as given in [IEC 61260] for frequencies from 50 Hz to 14 kHz, inclusive. For FB the upper frequency is 20 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.

   In case of headset measurements, the tests are repeated 5 times, in conformance with [ITU-T P.380]. The results are averaged (averaged value in dB, for each frequency).

3) The sensitivity is determined in dBPa/V.

NOTE – Different listener positions should be taken into account. Therefore, the measurements should be repeated by moving the seat with the artificial head in different, typical positions.

9.5 Listening speech quality during single talk

9.5.1 One-way listening speech quality in send direction

9.5.1.1 Requirement

The nominal values for the speech quality measured from/to the electrical reference point (POI) shall be:

$$\text{MOS-LQOsw} \geq 3.6$$

$$\text{MOS-LQOfb} \geq 3.6$$

NOTE – This requirement is non-normative at this point in time because there is currently no normative test method.

9.5.1.2 Test

A test method for measuring the one-way speech quality via the acoustic interface is currently under study.

9.5.2 One-way listening speech quality in receive direction

9.5.2.1 Requirement

The nominal values for the speech quality measured from/to the electrical reference point (POI) shall be:

$$\text{MOS-LQOsw} \geq 3.6$$

$$\text{MOS-LQOfb} \geq 3.6$$

NOTE – This requirement is non-normative at this point in time because there is currently no normative test method.

9.5.2.2 Test

A test method for measuring the one-way speech quality via the acoustic interface is currently under study.

9.6 Listening speech quality stability

9.6.1 Listening speech quality stability in send direction

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 due to the phone position and location.

In case of systems using SRW transmission, it is the purpose to verify the integration of SRW radio network by evaluating change of speech quality over time. This will help detect problems with the radio frequency (RF) coverage inside the car cabin and verify the error concealment (packet loss, bit errors) caused by the weak RF link or interference with other radios.

A guidance to proceed is:

1) Check RF coverage from the 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, etc.
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 a known phone as a reference.

NOTE – This measurement could be used as well to objectively investigate the system performance under time-varying channel conditions e.g. when driving the car under varying network conditions.

9.6.1.1 Requirement

For 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-LQOsw variation is equal to or more than 0.4.

$$\text{ST-MOS shall be} \geq 90$$

Currently MOS-LQO requirements for super-wideband and fullband systems are under study. In the interim period the requirement is only considering deviations between measurements derived from the undisturbed link and measurements derived from the disturbed link.

For super-wideband and fullband systems [ITU-T P.863] is used to determine MOS-LQOsw.

9.6.1.2 Test

Several measurements of the MOS-LQOsw score performed according to clause 9.5.1.1 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 enough. The results are reported in terms of statistics.

The assessment of listening speech quality stability is performed in 5 steps:

1) To measure the MOS-LQOsw periodically over the duration of one communication. **N** measurements provide **N** MOS-LQOsw values.
2) For each MOS-LQOsw (i) value i from 2 to N, MOS-LQOsw_GAP (i) is calculated as the absolute difference with the previous value MOS-LQOsw (i–1):

   $$\text{MOS-LQOsw\_GAP (i)} = |\text{MOS-LQOsw (i)} - \text{MOS-LQOsw (i–1)}|$$

3) In order to take into account the subjective perception and measurement accuracy, MOS-LQOsw_GAP (i) is set to 0 when the difference is equal to or lower than THRESHOLD1:
   - if MOS-LQOsw_GAP (i)> (2*THRESHOLD1), then MOS-LQOsw_GAP (i)= MOS-LQOsw_GAP (i)
   - if THRESHOLD1 < MOS-LQOsw_GAP (i) ≦ 2*THRESHOLD1, then MOS-LQOsw_GAP (i)= [MOS-LQOsw_GAP (i) * 2] – (2*THRESHOLD1)
   - if MOS-LQOsw_GAP (i) ≦ THRESHOLD1, then MOS-LQOsw_GAP (i)=0
4) The instability (INS_MOS-LQOsw) associated with the MOS-LQOsw over the whole N measurements is defined by mean value of MOS-LQOsw_GAP (i).

   $$\text{INS\_MOS-LQOsw} = 1/(N-1)\Sigma\ \text{MOS-LQOsw\_GAP (i) with i=[2:N]}$$

5) A linear weighting function is applied in order to express Stability ST-MOS-LQOsw on a 0 to 100 scale.

This formulation is used to determine the listening quality stability (ST-MOS) as:

$$\text{ST-MOS} = 100 - (250 * \text{INS\_MOS})$$

and

$$\text{ST-MOS} = 0 \text{ if } [100 - (250 * \text{INS\_MOS})] < 0$$

When ST-MOS is calculated within a single call, the call shall 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 this clause.

(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.)

9.6.2 Listening speech quality stability in receive direction

9.6.2.1 Requirement

For 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-LQOsw variation is equal to or more than 0.4.

ST-MOS shall be $\geq$ 90

Currently MOS-LQO requirements for super-wideband and fullband systems are under study. In the interim period the requirement is only considering deviations between measurements derived from the undisturbed link and measurements derived from the disturbed link.

For super-wideband and fullband systems [ITU-T P.863] is used to determine MOS-LQOsw.

9.6.2.2 Test

The test procedure is described in clause 9.6.1.2.

9.7 Idle channel noise

All tests are conducted with average RF-signal power settings. It is recommended to check the requirement, in addition, with different RF-power settings. The requirement should be fulfilled for all RF-power settings.

For the measurements, it is desirable to have a background noise level of less than –74 dBPa(A), a background noise level of –64 dBPa(A) shall not be exceeded.

9.7.1 Idle channel noise in send direction

9.7.1.1 Requirements

The maximum idle channel noise in send direction, measured at the electrical reference point (POI) in quiet conditions, shall be less than $\leq$ –64 dBm0(A).

No peaks in the frequency domain higher than 10 dB above the average noise spectrum shall occur.

NOTE 1 – Care should be taken that the measured spectral peaks result from the hands-free terminal under test and not from other sources e.g. noise produced by the car in quiet conditions.

NOTE 2 – In case spectral peaks higher than 10 dB above the average noise floor are produced by the hands-free terminal, but which are considered to be inaudible due to the very low noise floor produced by the hands-free terminal on average, it is the responsibility of the test lab to demonstrate the desired performance subjectively.

9.7.1.2 Test

1) For the measurement no test signal is used. In order to ensure a reliable activation of the terminal, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of four composite source signals according to [ITU-T P.501]. The spectrum of the test signal at the MRP is equalized under free-field conditions. The level of the activation signal is –4.7 dBPa for headset hands-free terminals, measured at the MRP and –25.7 dBPa for speakerphone hands-free terminals, measured at the HATS-HFRP.

2) The test arrangement is described in clause 7.1.

   The idle channel noise is measured at the electrical reference point in the frequency range between 50 Hz and 16 kHz in super-wideband and 20 kHz in fullband. 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 1 second which is the averaging time for the idle channel noise. The test lab has to ensure that the terminal is activated during the measurement. If the terminal is deactivated during the measurement, the measurement window has to be cut to the duration while the terminal 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.

3) The idle channel noise is determined by A-weighting.

4) 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 8.79Hz (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. In super-wideband the analysis is limited to 13 kHz, and in fullband the analysis is limited to 18 kHz.

9.7.2 Idle channel noise in receive direction

9.7.2.1 Requirements

The requirements for the maximum noise produced by the hands-free terminal, in case no signal is applied to the receive direction, is as follows:

If a user-specific volume control is provided, it is adjusted to the RLR value close to the nominal value. Hands-free terminals, without user-specific volume controls, are measured in normal operating conditions. The idle channel noise level measured at the DRP shall be less than –53 dBPa (A).

No peaks in the frequency domain higher than 10 dB above the average noise spectrum shall occur.

NOTE 1 – Care should be taken that the measured spectral peaks result from the hands-free terminal under test and not from other sources e.g. noise produced by the car in quiet conditions.

NOTE 2 – In case spectral peaks higher than 10 dB above the average noise floor are produced by the hands-free terminal but considered to be inaudible due to the very low noise floor produced by the hands-free terminal in average, it is the responsibility of the test lab to demonstrate the desired performance subjectively.

9.7.2.2 Test

1) For the measurements no test signal is used. In order to ensure a reliable activation of the terminal, an activation signal is inserted before the actual measurement. The activation signal consists of a sequence of four composite source signals according to [ITU-T P.501]. The level of the activation level is adjusted to –16 dBm0, measured at the electrical reference point. The level of the activation signal is averaged over the complete duration of the activation signal.

2) The test arrangement is according to clause 7.1. For the measurement of speakerphone hands-free terminals, the artificial head diffuse-field equalized (see clause 7.1.6.1), according to [ITU-T P.581]. The equalized output signal of the right ear is used for the measurement. For headset hands-free terminals, the sound pressure is measured at the DRP and diffuse-field equalized, according to [ITU-T P.57].

3) The idle channel noise is measured at the DRP in the frequency range between 50 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 or reverberance influence 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 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.

4) The idle channel noise is A-weighted.

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 8.79Hz (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. In super-wideband the analysis is limited to 13 kHz, and in fullband the analysis is limited to 18 kHz.

6) In case of headset measurements the tests are repeated 5 times, in conformance with [ITU-T P.380]. The results are averaged (averaged value in dB, for each frequency).

9.8 Distortion in send

The distortion in send up to 16 kHz in SWB and up to 20 kHz in FB is measured from the MRP to the electrical reference point (input of the system simulator, POI).

NOTE – It is recognized that for some systems, including AGC or companding techniques, the distortion test does not lead to valid results.

For the measurements, it is desirable to have a background noise level of less than –74 dBPa(A), a background noise level of –64 dBPa(A) should not be exceeded.

9.8.1 Requirements

The harmonic distortion in send shall not be higher than 3%.

9.8.2 Test

1) The test signal is a sinusoidal signal with a frequency of 300 Hz, 500 Hz, 1 kHz, 2 kHz and 5 kHz. The test signal level is –4.7 dBPa. In order to guarantee a reliable activation of the hands-free terminal, a sequence of four composite source signals, according to [ITU-T P.501], is sent to the terminal before the actual test signal. The activation signal level is –4.7 dBPa, measured at the MRP. The activation signal level is averaged over the total length of the activation signal.

2) The test signal is inserted immediately after the activation sequence, after the voiced sound of the last CSS-burst (instead of the pn-sequence). The test signal duration is 200 ms.

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

4) The harmonic distortion produced by the hands-free terminal is measured at the electrical reference point.

9.9 Distortion in receive

The distortion in receive is measured from the POI to the artificial ear up to 20 kHz.

NOTE – It is recognized that for some systems including AGC, companding techniques or bandwidth extension, the distortion test does not lead to valid results.

For the measurements, it is desirable to have a background noise level of less than –74 dBPa(A), a background noise level of –64 dBPa(A) shall not be exceeded.

9.9.1 Requirements

The distortion in receive is measured from a linear access point (access point at the hands-free system where no non-linear and time-variant signal processing except the speech coder is present) to the ERP.

The harmonic distortion shall be less than 3% when producing a sound pressure level needed to achieve an S/N ≥ 6 dB (see clause 7, preparation measurements) and for the maximum volume control setting.

This test is applicable if a linear access point without any non-linear signal processing to the loudspeaker amplifier is available. If this access point is not available, the measurement may be conducted with some care since non-linear processing may influence the test result.

9.9.2 Test

1) The test signal is a sinusoidal signal with a frequency of 100 Hz, 300 Hz, 500 Hz, 1 kHz, 2 kHz, 3 kHz and 5 kHz. The test signal level is the level measured at the linear access point when inserting a test signal with –16 dBm0 at the POI. In order to guarantee a reliable activation of the hands-free terminal, a sequence of 4 composite source signals, according to [ITU-T P.501], is sent to the terminal before the actual test signal. The activation signal level is the level equivalent to the level when inserting a test signal at the POI with –16 dBm0, measured at the linear access point. The activation signal level is averaged over the total length of the activation signal.
2) The test signal is inserted immediately after the activation sequence, after the voiced sound of the last CSS-burst (instead of the pn-sequence). The test signal duration is 200 ms.
3) For the analysis, a Hanning window is used which is adapted to the duration of the test signal (200 ms).
4) The harmonic distortion is measured for each test signal frequency.

NOTE – This measurement method does not apply to systems including artificial bandwidth extension.

9.10 Echo performance

Due to the expected delay in networks, the echo loss presented at the electrical reference point (POI) should be at least 50 dB during single talk. This echo loss, weighted terminal coupling loss (TCL_W), should be achieved for a wide range of acoustical environments and delays.

NOTE – When realizing echo loss by speech-activated attenuation/gain control, "comfort noise" should be inserted in case the signal is completely suppressed.

For the measurements, it is desirable to have a background noise level of less than –74 dBPa(A), a background noise level of –64 dBPa(A) shall not be exceeded.

9.10.1 Terminal coupling loss

9.10.1.1 Requirements

Terminal coupling loss (TCL) in quiet environments is measured as unweighted echo loss and shall be at least 46 dB for nominal setting of the volume control. For maximum setting of the volume control, TCL shall be higher than 46 dB. The implemented echo control mechanism should provide a sufficient echo loss for all typical environments and typical impulse responses.

NOTE – A TCL of ≥ 50 dB is recommended as a performance objective when measured under free-field conditions at the nominal setting of the volume control. Depending on the idle channel noise in the sending direction, it may not always be possible to measure an echo loss ≥ 50 dB.

When conducting the tests, it should be checked whether the signal measured is an echo signal and not a comfort noise inserted in send direction in order to mask an echo signal or noise emitted by the loudspeakers. This could be checked, for example, by conducting the idle channel noise measurement with the maximum volume control setting.

NOTE – There may be implementations where echo problems may be observed although the TCL test gives a high number. In such cases, it is recommended to verify the echo performance by subjective tests including different situations which are not addressed in this test.

9.10.1.2 Test

1) All tests are conducted in the car cabin, the test arrangement is described in clause 7.1. The noise level measured at the electrical access point (idle channel noise) shall be less than –63 dBm0. The attenuation between the input of the electrical reference point to the output of the electrical reference point is measured using a speech-like test signal.

2) The test signal is the compressed real speech signal described in clause 7.3.3 of [ITU-T P.501]. The signal level shall be –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).

4) TCL is calculated as unweighted echo loss from 100 Hz to 8 kHz. For the calculation, the averaged test signal level at each frequency band is referred to the averaged measured echo signal level 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. The echo loss is calculated by the equations

$$L_e = C - 10\log_{10}\sum_{i=1}^{N}(A_i + A_{i-1})(\log_{10} f_i - \log_{10} f_{i-1})$$

and

$$C = 10\log_{10}(2\,(\log_{10} f_N - \log_{10} f_0)$$

where

A_0 is the output/input power ratio at frequency $f_0 = 100\ Hz$;

A_1 the ratio at frequency f_i; and

A_N the ratio at frequency $f_N = 8000\ Hz$

The above equation is a generalized form of the equation defined in clause B.4 [b-ITU-T G.122] for calculating echo loss based on tabulated data, which allows the calculation of echo loss within any frequency range between f_0 and f_N.

9.10.2 Echo level vs time

9.10.2.1 Requirements

This test is intended to verify that the system will maintain sufficient echo attenuation during single talk. When measuring using the CS-signal the measured echo attenuation during single talk should not decrease by more than 6 dB from the maximum measured during the test. When measuring using the British-English single talk sequence the echo level variation shall be less than 6 dB.

NOTE 1 – The echo path is kept constant during this test, and the test should begin 5 seconds after the initial application of a reference signal such that a steady state converged condition is achieved.

NOTE 2 – The analysis is conducted only during the active signal part.

9.10.2.2 Test

1) The test arrangement is according to clause 7.1.

2) The test signal consists of 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 eight 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. In addition, the test is repeated with the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. One sequence of male and one sequence of female voice is used. 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 level analysis of the reference signal.

3) When using the CS-signal the measurement result is displayed as attenuation vs time. The exact synchronization between input and output signal has to be guaranteed.

4) When using the speech signal the measurement is displayed as level versus time.

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).

9.10.3 Spectral echo attenuation

9.10.3.1 Requirements

The echo attenuation vs frequency shall be below the tolerance mask given in Table 9-7.

Table 9-7 – Spectral echo attenuation mask

Frequency (Hz)	Upper limit
100 Hz	–41 dB
1 300 Hz	–41 dB
3 450 Hz	–46 dB
5 200 Hz	–46 dB
7 500 Hz	–37 dB
8 000 Hz	–37 dB
12 500 Hz	–37 dB
NOTE 1 – All sensitivity values are expressed in dB. NOTE 2 – The limit at intermediate frequencies lies on a straight line drawn between the given values on a log (frequency) – linear (dB) scale.	

During the measurement, it should be ensured that the measured signal is really the echo signal and not the comfort noise which possibly may be inserted in send direction in order to mask the echo signal.

NOTE – This requirement should be fulfilled at any point in time. Therefore, it should be verified at different time intervals of the test sequence.

9.10.3.2 Test

1) The test arrangement is according to clause 7.1.

2) Before the actual measurement, a training sequence is fed in consisting of 10 seconds CS-signal according to [ITU-T P.501]. The level of the training sequence is –16 dBm0.

3) The test signal consists of a periodically repeated composite source signal. The measurement is carried out under steady-state conditions. The average test signal level is –16 dBm0, averaged over the complete test signal. Four CS-signals, including the pauses, are used for the measurement which results in a test sequence length of 1.4 s. 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, rectangular window).

4) The spectral echo attenuation is analysed in the frequency domain in dB.

9.10.4 Initial convergence without background noise

9.10.4.1 Requirements

The initial convergence (echo attenuation vs time) during single talk immediately after activating the hands-free terminal with maximum volume control setting shall conform to the requirement in Figure 9-9.

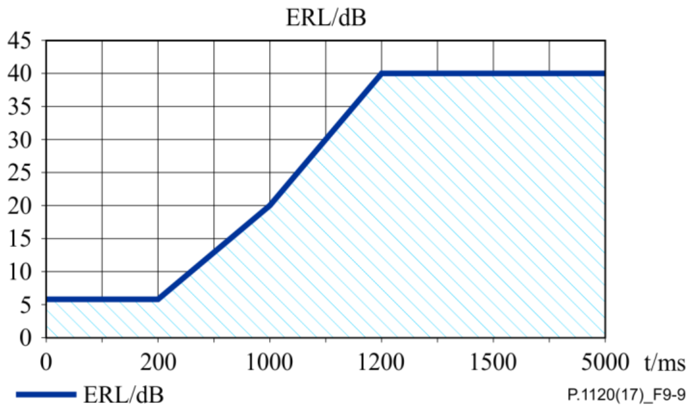


Figure 9-9 – Initial convergence, ERL vs time

9.10.4.2 Test

1) The test arrangement is described in clause 7.1. 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) 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 should be achieved for different speech signals.

9.10.5 Initial convergence with background noise

9.10.5.1 Requirements

The initial convergence (echo attenuation vs time) during single talk immediately after activating the hands-free terminal with background noise and with maximum volume control setting shall conform to the requirement in Figure 9-10.

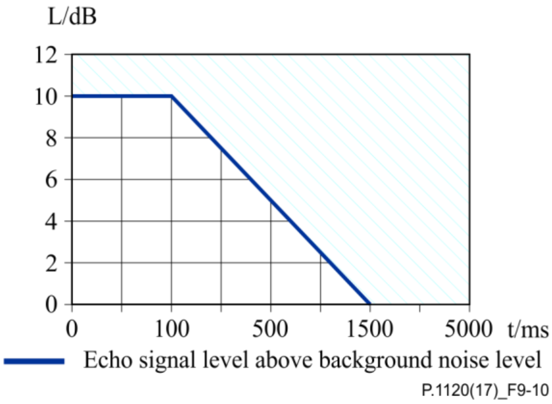


Figure 9-10 – Initial convergence with background noise, requirement on echo signal level vs time

9.10.5.2 Test

1) The test arrangement is described in clause 7.1.
2) The background noise defined by the manufacturer/test house is played back at least 5 s before the start of the actual measurement. This allows time for some adaptive algorithms in the hands-free unit 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. 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 first male sentence and the first female sentence are used. The starting point of each signal is as defined by the start of the male sentence. 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.
5) The measurement is displayed as echo level vs 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 signals should be achieved for different starting points of the speech signal.

9.10.6 Echo performance with time variant echo path

9.10.6.1 Requirements

This test is intended to verify that the system will maintain sufficient echo attenuation during time variant echo path. The echo level measured with a time varying echo path shall not increase by more than 6 dB compared to the echo level observed under steady state conditions. The time-variant echo-path may be realized as follows:

- A rotating 30 cm × 40 cm reflecting surface (e.g., piece of cardboard, wood, or plastic) is positioned on the seat in the same row closest to the HATS. The initial state of the reflecting surface (i.e., 0° position) is such that it is in the median plane (perpendicular to the front of the vehicle) with a bottom-to-top height of 40 cm, a front-to-back length of 30 cm; and the centre of the reflecting surface is at a point in the vehicle that is symmetric with the centre of the HATS in the seat. The reflecting surface then pivots 90° such that the most forward edge of the reflecting surface rotates out towards the seat in the same row closest to the HATS side window; the centre of the reflecting surface serves as the axis point and stays in the same location during this rotation. At the 90° position, the reflecting surface is in the frontal plane (parallel with the front of the vehicle). The reflecting surface continuously rotates between the 0° and 90° positions during the measurements at a rate of 90°/second. The rotation of the reflecting plane is time-synchronized with the test signals by means of a control channel.

- Alternatively the time variant echo path is realized by opening and closing the door closest to the zone under test during the measurement. Care has to be taken to quietly open and close the door in order not to impair the measurement by noises produced when opening and closing the door or by warning signals produced by the car.
- As a third alternative, the test is conducted by positioning a person on the seat in the same row closest to the HATS, which is quietly moving one arm close to the microphone relevant for the zone under test during the measurement.

The type of echo path variation chosen shall be reported.

9.10.6.2 Test

1) Before conducting the test, the echo canceller should be fully converged.
2) The test arrangement is according to clause 7.l.
3) The measurement is started by introducing the time-variant echopath.
4) The test signal consists of 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 eight 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) The measurement result is displayed as attenuation vs 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).

9.10.7 Echo performance with time variant echo path and speech

9.10.7.1 Requirements

This test is intended to verify that the system will maintain sufficient echo attenuation during time variant echo path when applying speech. The echo level measured with a time varying echo path shall not increase by more than 6 dB compared to the echo level observed under steady-state conditions. The time-variant echopath is realized as described in clause 9.10.6.1.

The type of echo path variation chosen shall be reported.

9.10.7.2 Test

1) Before conducting the test, the echo canceller shall be fully converged.
2) The test arrangement is according to clause 7.1. The test signal used is the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. The first male sentence and the first female sentence are used. 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 a reference.
3) Now a second measurement is started by introducing the time-variant echopath.
4) The test is repeated simulating the time-variant echo path using the British-English single talk sequence described in clause 7.3.2 of [ITU-T P.501]. Again, the first male sentence and

the first female sentence are used. 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.

5) The difference of the echo level between the reference and the measured echo loss with the time-variant echopath is determined.

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

9.11 Switching characteristics

9.11.1 Activation in send direction

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

The activation level described in the following is always referred to the test signal level at the mouth reference point (MRP).

9.11.1.1 Requirements

The minimum activation level $L_{S,min}$ shall be ≤ –20 dBPa.

The built-up time $T_{r,S,min}$ (measured with minimum activation level) shall be ≤ 50 ms.

9.11.1.2 Test

The structure of the test signal is shown in Figure 9-11. The test signal consists of CSS components according to [ITU-T P.501] with increasing level for each CSS burst.

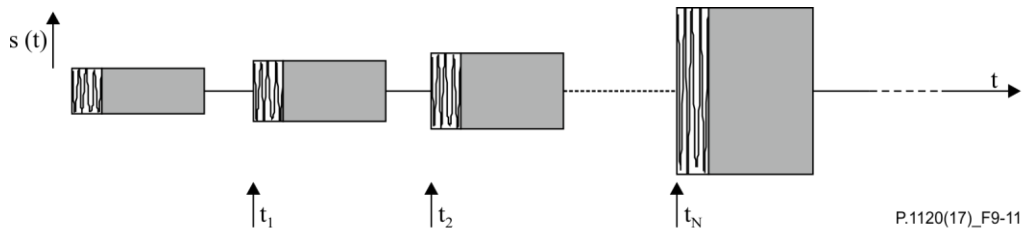


Figure 9-11 – Test signal to determine the minimum activation level and the built-up time

The settings of the test signal are given in Table 9-9 and the text that follows.

Table 9-9 – Settings of the CSS in send direction

	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	–23 dBPa (Note 1)	1 dB
NOTE 1 – The level of the active signal part corresponds to an average level of –24.7 dBPa at the MRP for the CSS according to [ITU-T P.501], assuming a pause of 101.38 ms. NOTE 2 – When testing speakerphone hands-free system, the signal level 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 7.1.

2) The level of the transmitted signal is measured at the electrical reference point. The measured signal level is referred to the test signal level and displayed vs time. The levels are calculated from the time domain using an integration time of 5 ms.

3) 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 a one-syllable word instead of the CS-signal. The word used should be of similar duration, the average level of the word must be adapted to the CS-signal level of the corresponding CSS-burst.

9.11.2 Activation in receive direction

The activation in receive direction is mainly determined by the built-up time $T_{r,R,min}$ and the minimum activation level ($L_{R,min}$). The minimum activation level is the level required to completely remove 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.

The activation level described below is always referred to the test signal level at the electrical reference point (POI).

In order to guarantee a higher accuracy when recording the transmitted signal in receive direction, a measurement microphone is used for this test and positioned close to the loudspeaker of the hands-free terminal.

9.11.2.1 Requirements

The minimum activation level $L_{R,min}$ shall be ≤ –35.7 dBm0 (measured during the active signal part).

The built-up time $T_{r,R,min}$ (measured with minimum activation level) shall be ≤ 50 ms.

9.11.2.2 Test

The signal construction is shown in Figure 9-11. The test signal settings are shown in Table 9-10 and the text that follows.

Table 9-10 – Settings of the CSS in receive direction

	CSS duration/ Pause duration	**Level of the first CS signal (active signal part at the POI)**	**Level difference between two periods of the test signals**
CSS to determine switching characteristics in receive direction	248.62 ms/ 451.38 ms	–38.7 dBm0 (Note)	1 dB
NOTE – The level of the active signal part corresponds to an average level of –40 dBm0 at the POI for the CSS according to [ITU-T P.501], assuming a pause of 101.38 ms.			

1) The test arrangement is according to clause 7.1.

2) The transmitted signal is recorded by a microphone positioned close to the loudspeaker. For headset hands-free terminals, the sound pressure is measured at the DRP and corrected to the ERP according to [ITU-T P.57]. The measured signal level is referred to the test signal level and displayed vs time. The integration time of the level analysis used shall be 5 ms.

3) 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 instead of the CS-signal. The word used should be of similar duration, the average level of the word must be adapted to the CS-signal level of the corresponding CSS-burst.

9.11.3 Attenuation range in send direction

The attenuation range in send direction is determined by applying the test signal in send direction after the terminal was activated in receive direction. During the measurement, the attenuation range in send direction ($A_{H,S}$) and the built-up time in send direction ($T_{r,S}$) are determined.

9.11.3.1 Requirements

The attenuation range $A_{H,S}$ shall be less than 20 dB.

The built-up time $T_{r,S}$ shall be less than 50 ms. It is recommended to reduce the attenuation within 15 ms to at least 13 dB below the final value.

9.11.3.2 Test

The structure of the test signals is shown in Figure 9-12. 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.

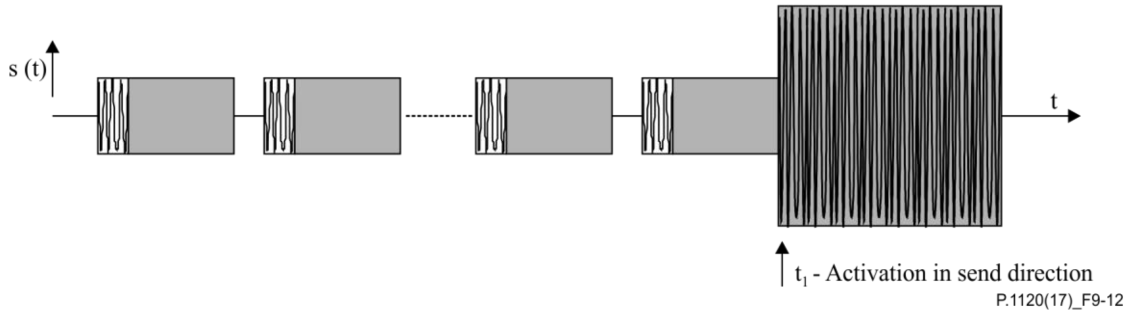


Figure 9-12 – Structure of the test signal for measuring the attenuation range

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

1) The test arrangement is according to clause 7.1.
2) The test signal used is according to Figure 9-12, the receive direction is activated first. The measurement parameters are shown in Table 9-11.

Table 9-11 – Signal levels for double talk tests in send and receive direction

	Receive direction	Send direction (at the MRP)
Average signal level (including 101.38 ms pauses)	–16 dBm0	–
Active signal part for speakerphone HFT	–14.7 dBm0	0 dBPa
Active signal part for headset HFT	–14.7 dBm0	–3 dBPa

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

3) The level is determined as level vs 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 in send direction (t_1 in Figure 11-6) until complete activation in send direction.

NOTE – In case this measurement fails e.g. due to a noise suppressor wrongly detecting the voiced sound as a background noise, the measurement may be repeated by using a real voiced sound of speech in order to demonstrate the desired performance.

9.11.4 Attenuation range in receive direction

The attenuation range in the receive direction is determined after the terminal was previously activated in the send direction. During the measurement, the attenuation range in receive direction ($A_{H,R}$) as well as the built-up time in receive direction ($T_{r,R}$) are determined.

9.11.4.1 Requirements

The attenuation $A_{H,R}$ shall be less than 15 dB.

The built-up time $T_{r,R}$ shall be less than 50 ms. It is recommended to reduce the attenuation within 15 ms to less than 9 dB.

9.11.4.2 Test

The structure of the test signal is shown in Figure 9-12. Again, CSS bursts are used for activating the opposite direction (now send direction) and the voiced sound is used to measure the receive direction. The test signals are synchronized in time at the acoustical interface. The delay of the test arrangement shall be constant during the measurement.

1) The test arrangement is according to clause 7.1.

2) The test signal shown in Figure 9-12 is used, the send direction is activated first.

The measurement parameters are as shown in Table 9-12.

Table 9-12 – Signal levels for double talk tests in send and receive direction

	Receive direction	Send direction (at the MRP)
Average level speakerphone HFT(including 101.38 ms pauses)	–	–1.7 dBPa
Average level headset HFT (including 101.38 ms pauses)	–	-4.7 dBPa
Active signal part for speakerphone HFT	–14.7 dBm0	0 dBPa
Active signal part for headset HFT	–14.7 dBm0	-3 dBPa

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

3) The level is determined as level vs 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 receive direction (t_1 in Figure 11-6) and the complete activation in receive direction.

9.12 Double talk performance

NOTE – Before starting the double talk tests, the test lab should ensure that the echo canceller is fully converged. This can be done by an appropriate training sequence.

During double talk, the speech is mainly determined by two parameters: impairment caused by echo during double talk and level variation between single and double talk (attenuation range).

In order to guarantee sufficient quality under double talk conditions, the talker echo loudness rating should be high and the attenuation inserted should be as low as possible. Terminals which do not allow double talk in any case should provide a good echo attenuation which is realized by a high attenuation range in this case.

The most important parameters determining the speech quality during double talk are (see [ITU-T P.340] and [ITU-T P.502]):

– Attenuation range in send direction during double talk ($A_{H,S,dt}$)
– Attenuation range in receive direction during double talk ($A_{H,R,dt}$)
– Echo attenuation during double talk

9.12.1 Attenuation range in send direction during double talk: AH,S,dt

9.12.1.1 Requirements

Based on the variation of the level in send direction during double talk $A_{H,S,dt}$, the behaviour of hands-free terminals can be classified according to Table 9-13.

Table 9-13 – 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*
$A_{H,S,dt}$ [dB]	≤3	≤6	≤9	≤12	>12

The requirements apply for nominal and maximum setting of the receive volume control.

The requirements apply for 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.

NOTE – If the maximum setting of the volume control is chosen such that non-linearities occur in the echo path, the double talk performance will decrease.

In general, Table 9-13 provides a quality classification of terminals regarding double talk performance. However, concerning the overall quality, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality as well.

9.12.1.2 Test

The test signal to determine the attenuation range during double talk is shown in Figure 9-13. 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 send and is used for analysis.

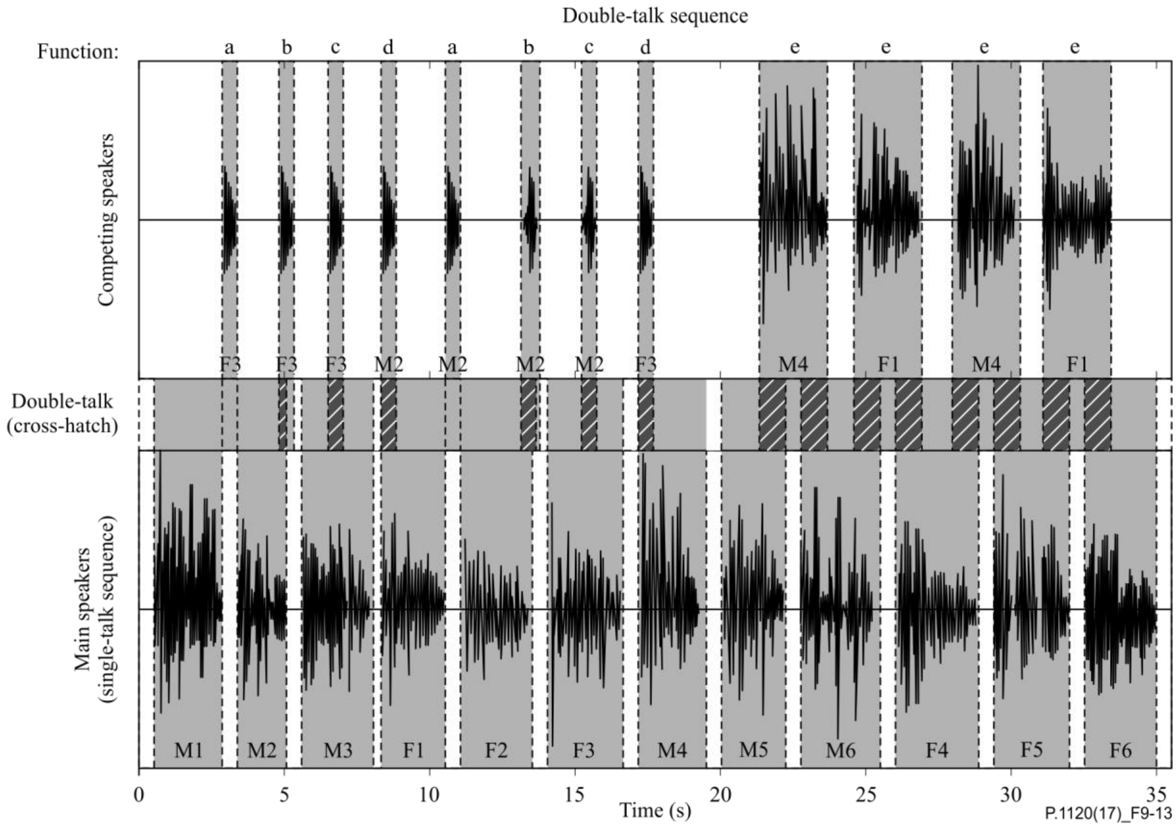


Figure 9-13 – Double talk test sequence with overlapping speech sequences in send and receive directions

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

The settings for the test signals are given in Table 9-14:

Table 9-14 – Signal levels of the double talk sequences

	Receive direction	**Send direction**
Average signal level for speakerphone HFT	–16 dBm0	–1.7 dBPa
Average signal level for headset HFT	–16 dBm0	–4.7 dBPa

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

1) The test arrangement is according to clause 7.1. 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 a level of –16 dBm0 is applied to the electrical reference point.
2) When determining the attenuation range in send direction, the signal measured at the electrical reference point is referred to the test signal inserted.
3) The attenuation range during double talk is determined as described in [ITU-T P.502]. The double talk performance is analysed for the sequence of words and the sequence of sentences produced by the competing speaker. The requirement has to be met for the sequence of words and the sequence of sentences produced by the competing speaker.
4) The test is repeated for all level combinations as defined in the requirements.

9.12.2 Attenuation range in receive direction during double talk: AH,R,dt

To ensure a higher accuracy measuring the transmitted signal in receive direction, a measurement microphone is used which is positioned as close as possible to the loudspeaker of the hands-free terminal.

9.12.2.1 Requirements

Based on the level variation in receive direction during double talk $A_{H,R,dt}$, the behaviour of the hands-free terminal can be classified according to Table 9-15.

Table 9-15 – 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*
$A_{H,R,dt}$ [dB]	≤3	≤5	≤8	≤10	>10

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

In general, Table 9-15 provides a quality classification of terminals regarding double talk performance. However, concerning the overall quality, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality as well.

9.12.2.2 Test

The test signal to determine the attenuation range during double talk is shown in Figure 9-14. 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 acoustical interface. The delay of the test arrangement shall be constant during the measurement.

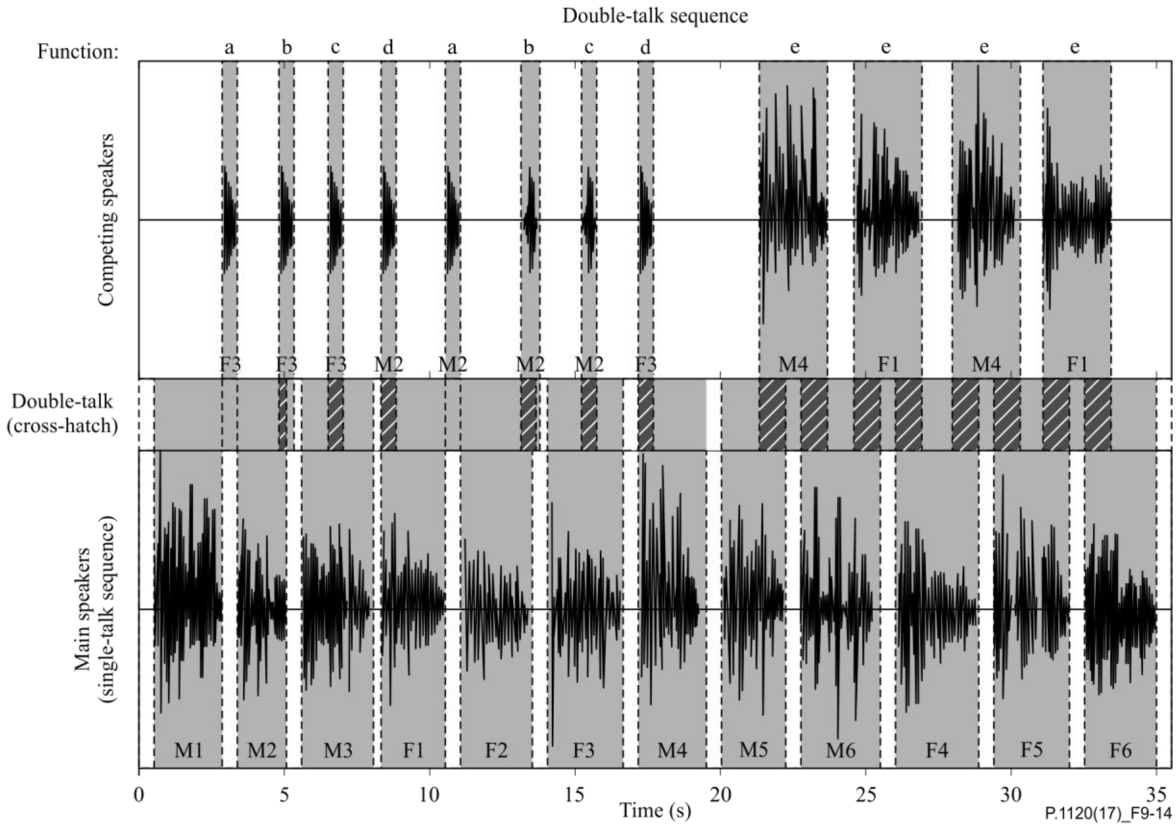


Figure 9-14 – Double talk test sequence with overlapping speech sequences in receive and send directions

The settings for the test signals are given in Table 9-16:

Table 9-16 – Signal levels of the double talk sequences

	Receive direction	**Send direction**
Average signal level for speakerphone HFT	–16 dBm0	–1.7 dBPa
Average signal level for headset HFT	–16 dBm0	–4.7 dBPa

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

1) The test arrangement is according to clause 7.1.

2) When determining the attenuation range in receive direction, the signal measured at the loudspeaker of the hands-free terminal is referred to the test signal inserted. For headset hands-free terminals, the sound pressure is measured at the DRP and corrected to the ERP according to [ITU-T P.57].

3) The attenuation range during double talk is determined as described in [ITU-T P.502]. The double talk performance is analysed for the sequence of words and the sequence of sentences produced by the competing speaker. The requirement has to be met for the sequence of words and the sequence of sentences produced by the competing speaker.

4) The test is repeated for all level combinations as defined in the requirements.

9.12.3 Detection of echo components during double talk

9.12.3.1 Requirements

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 requirements given in Table 9-17 are applicable (more information can be found in Annex A of [ITU-T P.340]).

Table 9-17 – 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

9.12.3.2 Test

1) The test arrangement is according to clause 7.1.

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 signal is described in [ITU-T P.501]. The signal settings used are shown in Figure 9-15. A detailed description can be found in [ITU-T P.501].

   The signals are fed simultaneously in send and receive directions. The level in send direction is –4.7 dBPa at the MRP (nominal level) for headset hands-free terminals, –25,7 dBPa at the HATSHFRP for speakerphone hands-free terminals, the level in receive direction is –16 dBm0 at the electrical reference point (nominal level).

Send direction		**Receive direction**	
$f_0^{(1)}$ **(Hz)**	$\pm\Delta f^{(1)}$ **(Hz)**	$f_0^{(2)}$**(Hz)**	$\pm\Delta f^{(2)}$ **(Hz)**
125	±2.5	180	±2.5
250	±5	270	±5
500	±10	540	±10
750	±15	810	±15
1 000	±20	1 080	±20
1 250	±25	1 350	±25
1 500	±30	1 620	±30
1 750	±35	1 890	±35
2 000	±40	2 160	±35
2 250	±40	2 400	±35
2 500	±40	2 650	±35
2 750	±40	2 900	±35
3 000	±40	3 150	±35

Send direction		Receive direction	
$f_0^{(1)}$ (Hz)	$\pm\Delta f^{(1)}$ (Hz)	$f_0^{(2)}$(Hz)	$\pm\Delta f^{(2)}$ (Hz)
3 250	±40	3 400	±35
3 500	±40	3 650	±35
3 750	±40	3 900	±35
4 000	±40	4 150	±35
4 250	±40	4 400	±35
4 500	±40	4 650	±35
4 750	±40	4 900	±35
5 000	±40	5 150	±35
5 250	±40	5 400	±35
5 500	±40	5 650	±35
5 750	±40	5 900	±35
6 000	±40	6 150	±35
6 250	±40	6 400	±35
6 500	±40	6 650	±35
6 750	±40	6 900	±35
7 000	±40		

The signal generation is according to [ITU-T P.501].

Figure 9-15 – Parameters of the two test signals for double talk measurement based on AM-FM modulated sine waves

3) The test signal is measured at the electrical reference point (send direction). The measured signal consists of the double talk signal which was fed in by the artificial mouth and the echo signal. The echo signal is filtered by comb filter using mid-frequencies and bandwidth according to the signal components of the signal in 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 receive direction, the echo attenuation can be measured separately. The requirement for category 1 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 9-17. The echo attenuation is to be achieved for each individual frequency band from 200 Hz to 6 950 Hz according to the different categories.

NOTE – Some HFTs 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 DUT is considered compliant with this requirement.

9.12.4 Sent speech attenuation during double talk

9.12.4.1 Requirements

The sent speech attenuation during double talk is based on the parameter $A_{H,S,dt}$.

Based on the level variation in send direction during double talk $A_{H,S,dt}$, the behaviour of hands-free terminals can be classified according to Table 9-18.

Table 9-18 – 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*
$A_{H,S,dt}$ [dB]	≤3	≤6	≤9	≤12	>12

The requirements apply for nominal and maximum setting of the receive volume control.

The requirements apply for 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.

In general, Table 9-18 provides a quality classification of terminals regarding double talk performance. However, this does not mean that a terminal which is category 1 based on the double talk performance is of high quality concerning the overall quality as well.

The test is conducted in addition to the test described in clause 9.12.1 in order to guarantee that no switching device with short reaction time is classified as a duplex or partially duplex system.

9.12.4.2 Test

1) The test arrangement is according to clause 7.1.

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 signal used is described in [ITU-T P.501]. The signal settings are shown in Figure 9-15. A detailed description can be found in [ITU-T P.501].

   The signals are fed simultaneously in send and receive directions. The level in send direction is –4.7 dBPa at the MRP (nominal level) for headset hands-free terminals, -25,7 dBPa at the HATSHFRP for speakerphone hands-free terminals, the level in receive direction is –16 dBm0 at the electrical reference point (nominal level). The test signal is given in clause 9.12.3.2.

3) The test signal is measured at the electrical reference point (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 by comb filter using mid-frequencies and bandwidth according to the signal components of the signal in 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 send direction, the sent speech attenuation $A_{H,S,dt}$ can be measured separately. The requirement for category 1 is fulfilled if, in each frequency band, the attenuation of the signal in send direction is below the required limit. If attenuation is detectable, the classification is based on Table 9-18. The sent speech attenuation $A_{H,S,dt}$ is to be achieved for each individual frequency band from 200 Hz to 6900 Hz according to the different categories.

5) The test is repeated for all level combinations as defined in the requirements.

9.13 Background noise transmission

9.13.1 SNR improvement provided by the HFT algorithm

For further study.

9.13.2 Background noise transmission after call set-up

9.13.2.1 Requirements

The analysis based on the Relative Approach [b-Sottek] (see Annex B) shall not indicate remarkable characteristics exceeding 6 cp/cPa. The first transmitted signal peak in send direction shall not cause higher excitation than 15 cp/cPa between 200 Hz and 16 kHz.

9.13.2.2 Test

1) The test arrangement is given in clause 7.1.

2) According to the specification of the manufacturer/test lab, the background noise is played back. The test shall be carried out during a constant driving situation.

3) The terminal is switched off and on again (to provide a reset) and a call is established by the system simulation. The incoming call is answered at the terminal. Special care shall be taken not to produce any disturbances or unwanted noise by touching the terminals housing while answering the incoming call.

4) The transmitted signal in send direction is recorded at the POI starting at least 1 s before the call is answered and for at least 7 s after the call is established. The analysis range is chosen to 8 s including an initial pause of 1 s before the call was established.

5) The recorded signal is analysed using the Relative Approach [b-Sottek].

9.13.3 Speech quality in the presence of background noise

9.13.3.1 Requirements

Currently a test method for super-wideband and fullband is under study in ETSI TC STQ. The test method will lead to three MOS-LQO quality numbers:

N-MOS-LQOs/f:	Transmission quality of the background noise (super-wideband/fullband)
S-MOS-LQOs/f:	Transmission quality of the speech noise (super-wideband/fullband)
G-MOS-LQOs/f:	Overall transmission quality noise (super-wideband/fullband)

According to specifications of manufacturer/test lab, a realistic background noise is played back. For the background noises chosen, the following requirements apply:

Background noises in quiet, low fan (see Annex D):

N-MOS-LQOs ≥ 4.0

S-MOS-LQOs ≥ 4.0

G-MOS-LQOs ≥ 4.0

Background noises in quiet, low fan (see Annex D):

N-MOS-LQOf ≥ 4.0

S-MOS-LQOf ≥ 4.5

G-MOS-LQOf ≥ 4.0

Background noises for driving speed ≤ 80 km/h, super-wideband:

N-MOS-LQOs ≥ 3.0

S-MOS-LQOs ≥ 3.0

G-MOS-LQOs ≥ 3.0

Background noises for driving speed ≤ 80 km/h, fullband:

N-MOS-LQOf ≥ 3.0

S-MOS-LQOf ≥ 3.0

G-MOS-LQOf ≥ 3.0

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 7.1.3.

Background noises for driving speed ≤ 130 km/h, super-wideband:

N-MOS-LQOs ≥ 2.5

S-MOS-LQOs ≥ 2.5

G-MOS-LQOs ≥ 2.5

Background noises for driving speed ≤ 130 km/h, fullband:

N-MOS-LQOf ≥ 2.5

S-MOS-LQOf ≥ 2.5

G-MOS-LQOf ≥ 2.5

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 7.1.3.

NOTE – It is recommended to test the terminal performance with different types of background, e.g., open window, different types of road surfaces and other relevant conditions Time variant conditions should specially be taken into account.

9.13.3.2 Test

For further study.

9.13.4 Quality of background noise transmission (with far-end speech)

9.13.4.1 Requirements

The test is carried out applying the composite source signal in receive direction. During and after the end of composite source signal bursts (representing the end of far-end speech simulation), the signal level in send direction shall not vary more than 10 dB (during transition to transmission of background noise without far-end speech).

9.13.4.2 Test

1) The test arrangement is according to clause 7.1.
2) According to the specification of the manufacturer/test lab, the background noise is played back. The test should be carried out during a constant driving situation.
3) First, the measurement is conducted without inserting the signal at the far end. At least 10 seconds 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 inserting the CS-signal at the far end. The exact 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 shall be applied for at least 5 seconds in order to allow adaptation of the noise reduction (NR) algorithms. After at least 5 seconds, a composite source signal according to [ITU-T P.501] is applied in receive direction with a duration of ≥ 2 CSS periods. The test signal level is –16 dBm0 at the electrical reference point.
5) The send signal is recorded at the electrical reference point. The test signal level versus time is calculated using a time constant of 35 ms.

6) The level variation in 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 the reference signal and the signal measured with the far-end signal.

9.13.5 Quality of background noise transmission (with near-end speech)

9.13.5.1 Requirements

The test is carried out applying a simulated speech signal in send direction. During and after the end of the simulated speech signal (composite source signal bursts), the signal level in send direction shall not vary more than 10 dB.

9.13.5.2 Test

1) The test arrangement is according to clause 7.1.
2) According to the specification of the manufacturer/test lab, the background noise is played back. The test should be carried out during a constant driving situation. The background noise shall be applied for at least 5 s in order to allow adaptation of the noise reduction algorithms.
3) The near-end speech is simulated using the composite source signal according to [ITU-T P.501] with a duration of ≥ 2 CSS periods. The test signal level is –4.7 dBPa at the MRP for headset hands-free terminals, –25,7 dBPa for hands-free. For speakerphone hands-free terminals, the HATS-HFRP correction has to be applied.
4) The send signal is recorded at the electrical reference point. The test signal level versus time is calculated using a time constant of 35 ms.
5) First, the measurement is conducted without inserting the signal at the near end. The signal level is analysed vs time. In a second step, the same measurement is conducted but with inserting the CS-signal at the near end. The level variation is determined by the difference between the background noise signal level without inserting the CS-signal and the maximum level of the noise signal during and after the CSS-bursts in send direction.

9.13.6 "Comfort Noise" injection

This clause is applicable only if comfort noise is inserted by the hands-free unit.

9.13.6.1 Requirements

1) The level of comfort noise is adjusted in a range of +2 and –5 dB to the original (transmitted) background noise. The noise level is calculated with A-weighting.
2) The spectral difference between comfort noise and original (transmitted) background noise shall be in between the mask given through straight lines between the breaking points on a logarithmic (frequency) – linear (dB sensitivity) scale as given in Table 9-19.

Table 9-19 – Requirements for spectral adjustment of comfort noise (mask)

Frequency (Hz)	Upper limit	Lower limit
200	12	–12
800	12	–12
800	10	–10
2000	10	–10
2000	6	–6
4000 14000	6 6	–6 –6
NOTE – All sensitivity values are expressed in dB on an arbitrary scale.		

9.13.6.2 Test

1) The test arrangement is according to clause 7.1. Background noise is played back.

2) The test signal is applied in receive direction consisting of an initial pause of 10 s and a periodical repetition of the composite source signal in receive direction (duration ≥ 10 s) with nominal level to enable comfort noise injection.

3) The transmitted signal is recorded in send direction at the electrical reference point.

4) The power density spectrum measured in send direction during the initial pause of the test signal (8k 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 receive direction (8k FFT/48 kHz sampling rate or equivalent, averaged over ≥ 5 s). Spectral differences between both power density spectra are analysed and compared to the requirements given in Table 9-19.

5) The level of the transmitted signal in send direction is determined during the initial pause of the test signal in receive direction and referred to the level of the transmitted signal in send direction determined during the application of the test signal in receive direction. Both levels are calculated using A-weighting.

10 Guidance on subjective testing

Besides objective testing of hands-free telephones, a subjective performance evaluation is also necessary. The tests described here – in addition to the tests as described in the P.800-series of ITU-T Recommendations – are targeted mainly to "in situ" hands-free tests for optimizing hands-free systems in the target car and under conditions which are not covered by the objective test specification.

For conducting the tests, the hands-free system under test has to be installed in the target car, which will be referenced here as near-end, as far-end either serves a landline phone (car-to-land test) or an observing car equipped also with the hands-free system under test (car-to-car test). It is recommended to not only test the hands-free system in a landline connection but also in a car-to-car connection because the latter case can be regarded as a worst-case scenario resulting in worse hands-free quality compared to landline connections.

The evaluation of the hands-free performance should be done for different background noise scenarios, such as different driving speeds, fan/defrost settings, etc.

For the main part of the subjective tests, a native language should be used. For the recordings, additional languages can be selected.

Since conversational tests are rather time-consuming, most of the hands-free tests are conducted as single-talk and double-talk tests following a clear given structure.

The evaluation is done at the far-end and/or the near-end, depending on the type of the test category.

To conduct the tests as effective as possible, it is advantageous to use a tool providing the test persons on both sides of the telephone connection with the detailed test procedure and the possibility to easily do the rating.

The performance evaluation of the hands-free system covers categories like:

- echo cancellation (echo intensity, speed of convergence, etc.)
- double-talk performance (echo during double-talk, speech level variation, etc.)
- speech and background noise quality in send direction (level, level variation, speech distortion, etc.)
- speech quality in receive direction (level, level variation, speech distortion)

- stability of the echo canceller for "closed loop" connection when doing car-to-car hands-free communication.

For the evaluation, some ITU-T Recommendations can serve as a guideline, such as [ITU-T P.800], [ITU-T P.800.1] and others from the P.800-series of ITU-T Recommendations. The judgment is done according to the rating scales given for each test case. The offered rating scales are of MOS type having grades 1 to 5 to be chosen, from where 5 denotes "best" and 1 denotes "worst" quality. Some of the rating scales are designed for a more diagnostic purpose (e.g., "echo duration").

The evaluation has to be done by experts who are experienced with subjective testing of hands-free systems. However, some of the tests described here could be conducted with naive subjects when following the procedures described in the relevant P.800-series of ITU-T Recommendations.

During the tests, the signals on near-end and far-end may be recorded to be used later on for third-party listening evaluation.

Tables 10-1 and 10-2 show possible test scenarios and rating categories.

Table 10-1 – Overview of car-to-landline tests

Echo canceller		**Rating**
Single talk	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h) Fan/defrost (off or "worst case") Receive volume: nominal, maximum, varying Enclosure dislocation	Disturbance caused by echo Echo characteristics: • Intensity • Duration • Frequency of occurrence Background noise variation
Double talk	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h) Receive volume: nominal, maximum, varying Enclosure dislocation	Disturbance caused by echo Echo characteristics: • Intensity • Duration • Frequency of occurrence Speech level variation Speech intelligibility/listening effort
Conversation	0 km/h	Disturbance caused by echo Echo characteristics: • Intensity • Duration • Frequency of occurrence Speech level variation Speech intelligibility/listening effort
Speech and background noise quality (send direction)		**Rating**
Stationary noise	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h) Driver's window closed/open Several fan settings	Speech level Speech level fluctuation Speech sound quality Intelligibility/listening effort speech naturalness Background noise quality Signal-to-noise ratio
Transient noise	Fan: switching on/off/change setting Indicator noises, wiper noise	Background noise quality Adaptation to background noise

Table 10-1 – Overview of car-to-landline tests

Echo canceller		**Rating**
	Pass by vehicles	
Speech quality (receive direction)		**Rating**
Single talk far-end	Maximum receive volume New call	Speech level Speech sound quality Speech intelligibility/listening effort

Table 10-2 – Overview of car-to-car tests

Echo canceller		**Rating**
Single talk	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h) Fan (off or "worst case") Receive volume: nominal, maximum, varying Enclosure dislocation	Disturbance caused by echo Echo characteristics: • Intensity • Duration • Frequency of occurrence Background noise variation
Double talk	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h) Receive volume: nominal, maximum, varying Enclosure dislocation	Disturbance caused by echo Echo characteristics: • Intensity • Duration • Frequency of occurrence Speech level variation Speech intelligibility/listening effort
System stability (car-to-car)		**Rating**
Stability	Receive volume: nominal, maximum Test signal: speech or impulse-like excitation	Stability Echo characteristics: • Intensity • Duration • Frequency of occurrence
Speech and background noise quality (send direction)		**Rating**
Stationary noise	3 typical driving scenarios (Germany: e.g., 0 km/h, 100 km/h, 130 km/h, 160 km/h)	Speech level Speech level fluctuation Speech sound quality Speech intelligibility/listening effort Background noise quality Signal-to-noise ratio
NOTE – The driving condition selected should reflect the type of car and the typical range of driving speeds in the country where the car hands-free system is intended to be installed.		

Test environment and equipment

Figures 10-1 and 10-2 outline the test environment for the test scenarios of car-to-landline and car-to-car. In both test scenarios, the supervisor is located at the far end, i.e., landline or the observing car, respectively. He is guiding through the test procedure and doing most of the test evaluation.

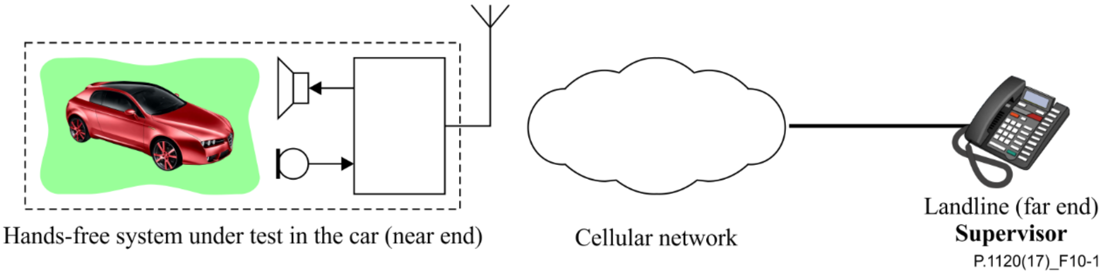


Figure 10-1 – Hands-free test car-to-landline

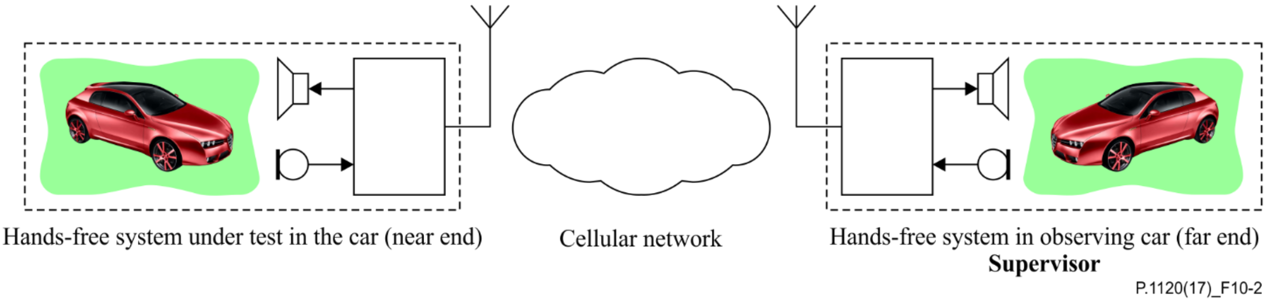


Figure 10-2 – Hands-free test car-to-car

Required test equipment:

- Car equipped with the hands-free telephone system under test.
- For car-to-car tests, both vehicles should be identical in terms of brand, model, model year, and hands-free system.
- Appropriate recording equipment is needed at the landline station and the car for documentation purposes and if third-party listening tests are required later on.

Further requirements:

- The tests have to be conducted by expert listeners.
- Male and female test persons should be in the car, quiet and loud talkers should be considered; different positions of the talkers with respect to the microphone should be considered.
- The hands-free tests have to be conducted in areas with good cellular coverage.
- In case of a SRW connection between a phone and a hands-free system, the phone should be placed in an area with a good SRW and cellular coverage.
- Some cell-phone networks have an influence on the hands-free performance (e.g., echo, AGC, noise reduction, etc.). These networks should not be chosen for the tests, if possible.
- Both the far-end and the near-end participants should be familiar with each other's voices.
- For the duration of the tests, there should be no change in the driver, front-passenger, or landline speaker.

NOTE – The network quality of service (QoS) parameters can be monitored by specific systems during the tests (based on 3GPP).

A description of the testing environment should be provided and should contain the following information:

- Car brand, model, model year, tyres, type of road, engine type, interior trim (cloth or leather, sunroof, etc.).
- Mobile telephone type and SW version.
- Telematics microphone – description, location, orientation, distance to driver.

- Network provider.
- HW and SW version of the hands-free device under test (DUT).
- HW and SW version of all devices which are part of the sound system.

Background noise and driving situations

For the tests, several different driving and background noise scenarios should be considered, which correspond to the main operating conditions of the hands-free system. Additionally, some worst-case scenarios might be regarded.

For the driving scenarios, this implies that the scenarios might differ for different countries depending, e.g., on the national speed limit for cars.

As an example, the following scenarios for different driving noise levels could be used (Germany):

- "low": 0 km/h
- "medium": 100 km/h
- "high": 160 km/h

For countries with speed limit, the maximum allowed speed might be used additionally with the climate control switched on to an appropriate level.

In addition to the driving scenarios, some different settings of the fan/defrost/climate control/recirculation are also useful background noise cases. As "worst case" scenario, a setting of the fan might serve where the air stream directly flows to the hands-free microphone in the car.

Test documentation

- After each test, a rating of the performance is done by referring to the given rating scales.
- At the beginning of every test, one of the test participants announces the test number (this is done for recording purposes).
- Audio signals should be recorded for documentation purposes:
  - On landline: Telephone audio recording of uplink and downlink signal
  - In the observing car: Binaural recording
  - In the car under test: Binaural recording

Notes on performance rating

For the ratings, the following items should be considered:

- The listeners who do the ratings have to be experienced with hands-free telephone systems.
- Limitations of the network have to be taken into account, e.g., for rating the hands-free system when the network's voice codec quality depends on the telephone connection traffic (e.g., CDMA).
- The offered rating scales are of MOS type having grades 1 to 5 to be chosen from, where 5 denotes "best" and 1 denotes "worst" quality.
- Some of the rating scales are designed for a more diagnostic purpose (e.g., the "echo duration").

Annex A

Speech quality measurements

(This annex forms an integral part of this Recommendation.)

In the following, testing methods and terminologies for determining the speech quality in send and receive direction and its relation to existing ITU-T terminology according to [ITU-T P.800.1] is described.

MOS related to listening-only situations

These MOS scores are applicable to a listening-only situation. Three different cases have to be distinguished.

MOS-LQS

The score has been collected in a laboratory test by calculating the arithmetic mean value of subjective judgments on a 5-point absolute category rating (ACR) quality scale, as it is defined in [ITU-T P.800].

Subjective tests carried out according to [ITU-T P.830], [ITU-T P.835], and [ITU-T P.840] give results in terms of MOS-LQS.

MOS-LQO

The score is calculated by means of an objective model which aims at predicting 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.

MOS-LQO (acoustical)

This kind of measurement is performed at acoustical interfaces. In order to predict the listening quality as perceived by the user, this measurement includes the actual telephone set products provided by the manufacturer or vendor. In combination with the choice of the acoustical receiver in the lab test ("artificial ear"), there will be a more or less leaky condition between the handset's receiver and the artificial ear. The same constraints apply for hands-free telephony. Consequently, for more realistic test scenarios, there may be a degradation of the measured MOS value, while for more artificial test scenarios there may be a negligible difference.

	Listening-only	**Conversational**	**Talking**
Subjective	MOS-LQSy	MOS-CQSy	MOS-TQSy
Objective	MOS-LQOy	MOS-CQOy	MOS-TQOy
Estimated	MOS-LQEy	MOS-CQEy	MOS-TQEy

NOTE – The letter "y" at the end of the above acronyms is a placeholder for the descriptor of the respective audio bandwidth, see the following provisional instructions:

– N for MOS scores obtained for narrowband (300 – 3400 Hz) speech relative to a narrowband high quality reference. This is applicable for instance to narrowband only subjective tests or to ITU-T P.863 scores.

– W for MOS scores obtained for wideband (50 – 7000 Hz) speech relative to a wideband high quality reference. This is applicable for instance to wideband-only subjective tests or to ITU-T P.863 scores.

– S for MOS scores obtained for super-wideband (20 – 14000 Hz) speech relative to a wideband high quality reference. This is applicable for instance to wideband-only subjective tests or to ITU-T P.863 scores.

– F for MOS scores scores obtained for audio signals up to fullband (10-20000 Hz) relative to a fullband high quality reference.

Further information can be found in [ITU-T P.800.1].

Annex B

Principles of Relative Approach

(This annex forms an integral part of this Recommendation.)

The Relative Approach [b-Sottek] is an analysis method developed to model a major characteristic of human hearing. This characteristic is the much stronger subjective response to distinct patterns (tones and/or relatively rapid time-varying structure) than to slowly changing levels and loudnesses.

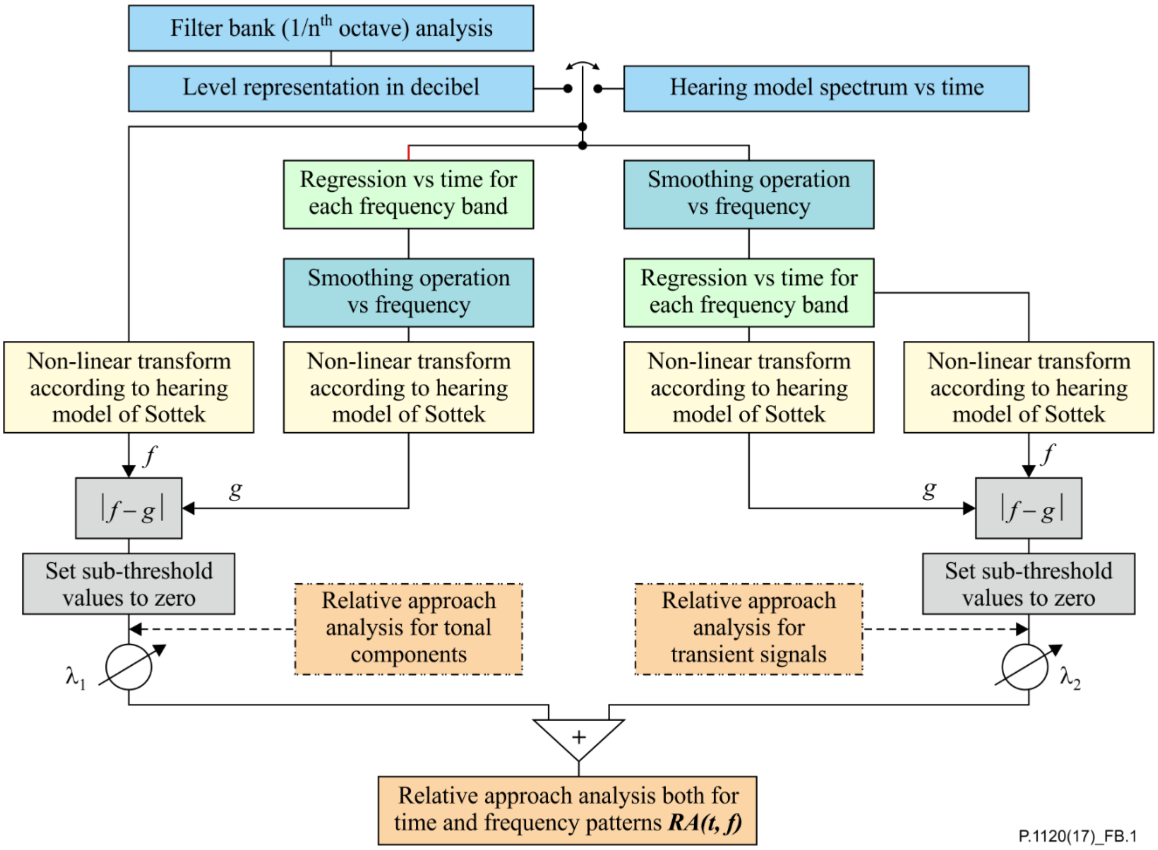


Figure B.1 – Block diagram of Relative Approach

The idea behind the Relative Approach analysis is based on the assumption that human hearing creates a running reference sound (an "anchor signal") for its automatic recognition process against which it classifies tonal or temporal pattern information moment-by-moment. It evaluates the difference between the instantaneous pattern in both time and frequency and the "smooth" or less-structured content in similar time and frequency ranges. In evaluating the acoustic quality of a complex "patterned" signal, the absolute level or loudness is almost without any significance. Temporal structures and spectral patterns are important factors in deciding whether a sound is judged as annoying or disturbing [b-Sottek].

Similar to human hearing and in contrast to other analysis methods, the Relative Approach algorithm does *not* require any reference signal for the calculation. Only the signal under test is analysed. Comparable to the human experience and expectation, the algorithm generates an "internal reference" which can be best described as a forward estimation. The Relative Approach algorithm objectifies pattern(s) in accordance with human perception by resolving or extracting them while largely rejecting pseudostationary energy. At the same time, it considers the context of the relative difference of the "patterned" and "non-patterned" magnitudes.

Figure B.1 shows a block diagram of the Relative Approach. The time-dependent spectral pre-processing can either be done by a filter bank analysis (1/nth octave, typically 1/12th octave) or a hearing model spectrum versus time according to the hearing model of Sottek (see [b-Sottek]). Both of them result in a spectral representation versus time. Both calculate the spectrograph using only linear operation and their outputs are therefore directly comparable. The hearing model analysis parameters are fixed and based on the processing in human ears, whereas the input parameters for the filter bank analysis can vary. The filter bank pre-processing approximates the hearing model version.

Two different variants of the Relative Approach can be applied to the pre-processed signal. The first one applies a regression versus time for each frequency band in order to cover human expectation for each band within the next short period of time. Afterwards, for each time slot, a smoothing versus frequency is performed. The next step is a non-linear transformation according to the Hearing Model of Sottek (see [b-Sottek]). This output is compared to the source signal which is also the hearing model transformed. Non-relevant components for human hearing are finally set to zero. This approach focuses on the detection of tonal components. The second version first smoothes versus frequency within a time slot and then applies the regression versus time. This output signal is transformed non-linearly to the hearing model. It is compared to the output of the smoothing versus frequency which is also non-linearly transformed according to the hearing model. Finally, non-relevant components for human hearing are again set to zero. Thus more transient structures are detected.

Via the factors λ_1 and λ_2, the weighting of the Relative Approach for tonal and transient signals can be set. Typically $\lambda_1 = 0$ and $\lambda_2 = 1$ are chosen. Thus, the model is tuned to detect time-variant transient structures.

The result of the Relative Approach analysis is a 3D spectrograph displaying the deviation from the "close to the human expectation" between the estimated and the current signal is displayed versus time and frequency. The Relative Approach uses a time resolution of $\Delta t = 6.66$ ms. The frequency range from 15 Hz to 24 kHz is divided into 128 frequency bands Δf_m which corresponds to a 1/12th octave resolution. Due to the non-linearity in the relationship between sound pressure and perceived loudness, the term "compressed pressure" in compressed Pascal (cPa) is used to describe the result of applying the non-linear transform. The Relative Approach can determine how "close to the human expectation" a signal is, but not if this expectation is of a high or a low quality origin.

Annex C

Example for a questionnaire for subjective testing

(This annex forms an integral part of this Recommendation.)

C Performance rating

C.1 Overview

The performance evaluation of the hands-free system covers categories such as:

- echo cancellation (EC) performance during single talk and double talk (echo during single talk, convergence after enclosure dislocation, echo during double talk, speech level variation during double talk, etc.),
- speech and background noise quality in send direction (level, level variation, speech naturalness, etc.),
- speech quality in receive direction (level, level variation, speech naturalness),
- stability of the echo canceller for "closed loop" operation when doing car-to-car hands-free communication.

For the evaluation, some ITU-T Recommendations have served as guidelines, such as [ITU-T P.800], [ITU-T P.800.1] and others from the ITU-T P.800 series. The judgement is done according to rating scales given for each test case. The offered rating scales are of MOS type having grades 1 to 5 to be chosen from where 5 denotes "best" and 1 denotes "worst" quality. Some of the rating scales are designed for a more diagnostic purpose (e.g., "echo duration").

The following items should be further considered for performance rating:

- the evaluation has to be done by experts who are experienced with subjective testing of hands-free systems,
- limitations of the network have to be taken into account and should be documented. For example, for rating the hands-free system when using code division multiple access (CDMA), the rating could say "not better than 2 because of CDMA network",
- the offered rating scales are of MOS type having grades 1 to 5 to be chosen from where 5 denotes "best" and 1 denotes "worst" quality,
- some of the rating scales are designed for a more diagnostic purpose (e.g., the "echo duration").

C.2 Test categories and rating types

Table C.1 gives an overview of the test categories and the related rating scales.

Table C.1 – Test categories and rating types overview

Test category	Test subcategory	Conversation type	Rating side	Rating type	Rating scales	Test condition/ variation
Speech and background noise quality in send direction	Speech level	Single talk near-end	Far-end	Speech level	Loudness preference (office)	Stationary background noise scenario: low/medium/high, fan/defrost, window
	Speech quality	Single talk near-end	Far-end	Speech quality	Speech level fluctuation Speech sound quality Speech naturalness Intelligibility/Listening effort	Stationary background noise scenario: low/medium/high, fan/defrost, window
	Background noise quality during near-end single talk	Single talk near-end	Far-end	Background noise quality	Signal-to-noise ratio Noise quality	Stationary background noise scenario: low/medium/ high, fan/defrost
	Transient background noise quality	Idle and single talk near-end	Far-end	Transient background noise quality	Transient noise quality	Transient background noise scenario: fan/defrost start up, wiper, indicator
		Idle	Far-end	Adaptation to background noise	Adaptation to background noise	Transient background noise scenario: noise jump, e.g., fan/defrost start up
Speech quality in receive direction	Speech level	Single talk far-end	Near-end	Speech level (car, max. vol.)	Loudness preference (car) for max. vol.	Only high background noise scenario
		Single talk far-end	Near-end	Speech level (car, nominal. Vol., new call)	Loudness preference (car) for nominal volume, new call	Low background noise scenario
	Speech quality	Single talk far-end	Near-end	Speech quality	Speech sound quality Intelligibility/Listening effort	Low background noise scenario
Echo cancellation (EC)	Echo during single talk	Single talk far-end	Far-end	Echo	Disturbance caused by echo Echo characteristics (only to be rated if echo occurs): • Intensity • Duration • Frequency of occurrence • Echo intelligibility	Stationary background noise scenario: low/medium/high, fan/defrost; volume (car), movement of driver (enclosure dislocation)

Table C.1 – Test categories and rating types overview

Test category	Test subcategory	Conversation type	Rating side	Rating type	Rating scales	Test condition/ variation
	Background noise quality during far-end single talk	Single talk far-end	Far-end	Background noise quality	Comfort noise quality (EC test)	Stationary background noise scenario: medium/high, fan/defrost
	Echo during double talk	Double talk	Far-end	Echo	Disturbance caused by echo Echo characteristics (only to be rated if echo occurs): • Intensity • Duration • Frequency of occurrence • Echo intelligibility	Stationary background noise scenario: low/medium/high, fan/defrost; volume (car), movement of driver (enclosure dislocation)
	Speech quality at near-end during double talk	Double talk	Near-end	Speech quality	Speech level variation during double talk Intelligibility/Listening effort during double talk	Low background noise scenario
	Speech quality at far-end during double talk	Double talk	Far-end	Speech quality	Speech level variation during double talk Intelligibility/Listening effort during double talk	Stationary background noise scenario: low/medium/high, fan/defrost, window
System stability	System stability	Special test	Far-end	Echo convergence and stability	System stability: Speed of convergence of echo cancellation and robustness against echo back coupling	Car-to-car, low background noise scenario, EC not adapted at start of test, max. volume

In addition to the tests given in Table C.1, some conversational tests can also be performed as described in [ITU-T P.800], and [ITU-T P.831] (conversation opinion scale).

C.3 Speech and background noise quality in send direction

C.3.1 Speech level (office)

Description:

The rating scale is applied in test cases which evaluate the preferred speech level of the received signal at the far-end (office).

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end (office)
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings, window open, etc.
Scale type:	Loudness-preference scale, [ITU-T P.800], ACR

The rating scale is given in Table C.2.

Table C.2 – Loudness preference (office)

Rating description	Grade
Much louder than preferred	1
Louder than preferred	3
Preferred	5
Quieter than preferred	3
Much quieter than preferred	1

C.3.2 Speech level fluctuations

Description:

The rating scale is applied in test cases which evaluate level variations in speech in single-talk situations.

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings, window open, etc.
Scale type:	Special scale for HF system diagnostic evaluation

Level fluctuations to be examined for this evaluation are characterized by:

- level fading,
- short drop-outs, e.g., due to missing data packets,
- cut-offs (missing word-ends or syllables),
- chopped voice.

The rating scale is given in Table C.3.

Table C.3 – Speech level fluctuations

Rating description	Grade
No speech level variation audible	5
Slight level variations, just audible or very rare occurring	4
Moderate speech level variations, may occur frequently	3
Sometimes words or syllables are attenuated or missing	2
Many drop outs, cut offs, missing words or syllables, heavily chopped voice	1

C.3.3 Speech quality/speech naturalness

Description:

The rating scale is applied in test cases that evaluate speech naturalness received on the far-end for different background noise scenarios on the near-end. This evaluation includes possible impairments caused by signal distortion and band limiting effects which also degrade the speech naturalness. The best quality case for this evaluation would be a hand-set comparable voice quality.

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings, window open, etc.
Scale type:	Special scale for HF system diagnostic evaluation (considering also the degradation scale, [ITU-T P.800], degradation category rating)

Properties to be examined for this evaluation are

- synthetic/robotic sound,
- speech signal distortion characterized by a scratchy sound,
- band limitation or filtering effects characterized by:
  - shrill, sharp, thin, tinny or muffled sounding speech,
  - emphasis on high frequencies or low frequencies.

The rating scale is given in Table C.4.

Table C.4 – Speech quality/speech naturalness

Rating description		Grade
Speech sound is comparable to hand-set voice quality; Speech sounds clear and transparent	Natural	5
Minor degradation to hand-set, still natural voice; Possibly slight distortions and/or slight band limitation effects	–	4
Maybe slight synthetic voice at times and/or low level distortion and/or moderate band limitation effects	–	3
Very noticeable synthetic voice and/or heavy distortion and/or higher degree of band-limitation	–	2
Signal barely recognizable as voice	Unnatural	1

C.3.4 Intelligibility/listening-effort

Description:

The rating table is applied to evaluate the effort required to understand the meaning of words and sentences. The applicable test cases are single talk and different background noise scenarios at the near-end and evaluation on the far-end.

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings, window open, etc.
Scale type:	Listening-effort scale, [ITU-T P.800], ACR

The rating scale is given in Table C.5.

The question heading this scale could be for example:

How would you judge the effort required to understand words and sentences of your remote partner?

Table C.5 – Intelligibility/listening-effort

Rating description	Grade
Every word was clearly understood with no effort required.	5
Speech of the other side is understood with no appreciable effort required.	4
Some words were hard to understand, moderate effort is required.	3
Many words were hard to understand, considerable effort required.	2
No meaning understood with any feasible effort.	1

C.3.5 Signal-to-noise ratio for near-end single talk

Description:

The following rating scale is intended to evaluate the noise level compared to the speech level.

The evaluation is usually done in a high environmental noise test condition. The signal-to-noise ratio directly depends on the background noise level of the test scenario. The judgment, therefore, will get to a worse grade with the test scenario changing to a higher environmental noise. Therefore, it is problematic for this test category to be used as absolute evaluation, it is more applicable for comparison of different systems for the same environmental noise scenario.

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings.
Scale type:	Scale is adapted from "Degradation category scale" (Annex D of [ITU-T P.800]) and the extended "detectability scale" of the "Quantal-Response Detectability Test" (Annex C of [ITU-T P.800])

The rating scale is given in Table C.6.

Table C.6 – Signal-to-noise ratio

Rating description	Grade
Noise very low, just audible	5
Noise audible, noise level clearly lower than speech level, noise is not disturbing	4
Medium noise level, lower than speech level, noise slightly disturbing	3
High noise level, almost same level as speech, clearly disturbing, but call would be continued	2
Noise louder than speech, intolerably disturbing, call would be abandoned	1

C.3.6 Background noise quality

Description:

This scale is for evaluation of the sound quality of the near-end background noise examined at the far-end. For explicit evaluation of transient noise sources, another rating scale is defined in clause C.3.7.

Test category:	Speech and background noise quality in send direction
Conversation type:	Single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings.
Scale type:	Special scale for HF system diagnostic evaluation

The following sound characteristics have to be taken into account:

- are there changes of noise-sound and level over time?
- are there artefacts audible (clicks, pops, rattle) which cannot be matched to a natural source (e.g., road bumps) or do not sound like their natural source?
- naturalness:
  - does the noise sound like it is part of the natural background?
  - does the noise sound synthetic (musical tones, watery sound)?
  - does the noise sound distorted?

Artefacts, synthetic sound, level and sound variation over time, and noise sounding unrelated to the natural background result in lower (worse) grades.

The rating scale is given in Table C.7.

Table C.7 – Background noise quality

Rating description		**Grade**
Comfortable, natural sound, constant in sound and level, no artefacts	Natural	5
Slight distortion/synthetic sound, almost no artefacts, almost constant in sound and level	–	4
Moderate distortion/synthetic sound, or some artefacts/clicks/plops audible, or some moderate variation in sound and level	–	3
Clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level	–	2
Completely unnatural/distorted/synthetic sound, or permanent artefacts/clicks/plops, or permanent variations in sound and level, very uncomfortable to listen to	Unnatural	1

C.3.7 Noise quality for transient noise sources

Description:

This scale is for evaluation of the sound quality of transient near-end background noise examined at the far-end. Transient noise sources can be, for example, the car's activated wiper, indicator, etc.

Test category:	Speech and background noise quality in send direction
Conversation type:	Idle and single talk at near-end
Rating side:	Rated at far-end
Test conditions:	Transient noise sources active (indicator, wiper, etc.)
Scale type:	Special scale for HF System diagnostic evaluation

For the evaluation, the naturalness of the noise sound has to be taken into account, for example:

- does the noise sound like related to the natural background?
- does the noise sound synthetic?
- does the noise sound distorted?

The rating scale is given in Table C.8.

Table C.8 – Transient noise quality

Rating description		**Grade**
Comfortable/natural sound	Natural	5
Almost natural sounding; Slight distortion/synthetic sound	–	4
Moderate unnatural sounding; Moderate distortion/synthetic sound	–	3
Clearly unnatural/distorted/synthetic sounding	–	2
Completely unnatural/distorted/synthetic sound	Unnatural	1

C.3.8 Adaptation to background noise

Description:

This scale is for evaluation of the speed of adaptation of the noise suppression to the near-end background noise after a noise level jump. The evaluation is done on the far-end.

Test category:	Speech and background noise quality in send direction
Conversation type:	Idle
Rating side:	Rated at far-end
Test conditions:	Transient background noise: noise jump, e.g., fan/defrost start up
Scale type:	Special scale for HF system diagnostic evaluation

The test can be conducted, for example, by turning on the defrost/fan in the car to a high setting. When conducting such a test, the start-up time of the defrost/fan has to be taken into account.

The rating scale is given in Table C.9.

Table C.9 – Adaptation to background noise

Rating description		Grade
Immediate adaptation	Very fast	5
Adaptation time ≤ 1 second		4
Adaptation time 2 … 3 seconds		3
Adaptation time 3 … 10 seconds		2
Adaptation time ≥ 10 seconds	Very slow	1

C.4 Speech quality in receive direction (in the car under test)

C.4.1 Speech sound quality/speech naturalness (receive)

Description:

The rating scale is applied in test cases which evaluate speech naturalness received in the car under test (near-end). This evaluation includes possible impairments caused by signal distortion and band limiting effects which also degrade the speech naturalness. The best quality case for this evaluation would be a hand-set comparable voice quality. The evaluation is done in low background noise condition using nominal volume setting in the car.

Test category:	Speech quality in receive direction
Conversation type:	Single talk at far-end
Rating side:	Rated at near-end
Test conditions:	Low background noise condition
Scale type:	Special scale for HF system diagnostic evaluation (considering also the degradation scale, [ITU-T P.800], degradation category rating)

Properties to be examined for this evaluation are:

- synthetic/robotic sound,
- speech signal distortion characterized by a scratchy sound,
- band limitation or filtering effects characterized by:
  - shrill, sharp, thin, tinny or muffled sounding speech,
  - emphasis on high frequencies or low frequencies.

For the rating scale, see Table C.4.

C.4.2 Intelligibility/listening-effort (Receive)

Description:

The rating table is applied to evaluate the effort required to understand the meaning of words and sentences. The applicable test cases are single talk at the far-end examined in the car under test (near-end) in a low background noise condition using nominal volume setting in the car.

Test category:	Speech quality in receive direction
Conversation type:	Single talk at far-end
Rating side:	Rated at near-end
Test conditions:	Low background noise condition
Scale type:	Listening-effort scale, [ITU-T P.800], ACR

The question heading this scale could be for example:

How would you judge the effort required to understand words and sentences of your remote partner?

For the rating scale, see Table C.5.

C.4.3 Speech level (receive, maximum volume)

Description:

The rating table is applied to evaluate the speech level heard from the loudspeakers in the car (near-end) when being in a high background noise condition and having the telephone volume set to maximum.

Test category:	Speech quality in receive direction
Conversation type:	Single talk at far-end
Rating side:	Rated at near-end
Test conditions:	High background noise condition
Scale type:	Loudness-preference scale, [ITU-T P.800], ACR

The rating scale is given in Table C.10.

Table C.10 – Loudness preference (car) for maximum volume setting

Rating description	Grade
Much louder than preferred	1
Louder than preferred	3
Preferred (for maximum volume setting)	5
Quieter than preferred	3
Much quieter than preferred	1

C.4.4 Speech level for nominal volume and new call

Description:

The rating table is applied to evaluate the speech level heard from the loudspeakers in the car (near-end) when engaged in a new call, after having set the volume to nominal in the prior call.

Test category:	Speech quality in receive direction
Conversation type:	Single talk at far-end
Rating side:	Rated at near-end
Test conditions:	High background noise condition
Scale type:	Loudness-preference scale, [ITU-T P.800], ACR

The rating scale is given in Table C.11.

Table C.11 – Loudness preference for maximum volume and new call

Rating description	Grade
Much louder than preferred	1
Louder than preferred	3
Preferred	5
Quieter than preferred	3
Much quieter than preferred	1

C.5 Echo cancellation performance

For the evaluation of echo, as a first step, the disturbance caused by echo is judged. If echo is perceived, then an evaluation of additional echo characteristics is performed. The evaluation of these echo characteristics is intended for diagnostic purposes.

The tests are intended to rate the perceived quality according to:

- amount and nature of echo during single talk,
- amount and nature of echo during double talk,
- convergence characteristics of the EC to handle variation of the echo path (e.g., when the driver in the car is moving),
- speech quality during double talk situations (e.g., intelligibility and speech level variation), this judgement is done on the near-end and the far-end,
- stability of the EC in car-to-car communication.

The tests are performed for different background noise scenarios and driving conditions, to get some information about the EC robustness in high background noise conditions.

The scales given in this clause can be applied for steady-state conditions and initial convergence tests.

C.5.1 Disturbance caused by echo

Description:

The rating scale is applied to evaluate the disturbance caused by echo examined at the far-end. The test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end Double talk
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings; enclosure dislocation due to movement of the driver; volume setting in the car: nominal, maximum.
Scale type:	Conversation impairment scale, [ITU-T P.800], and [ITU-T P.831]

The other echo rating scales (echo intensity, duration, frequency of occurrence and intelligibility) following hereafter are intended for diagnostic purposes. They are only used if echo is perceived.

The rating scale is given in the following table.

For the evaluation, the participant has to answer a question such as:

How would you judge the degradation/impairment/disturbance from echo of your own voice during the test?

Table C.12 – Disturbance caused by echo

Rating description	Grade
Imperceptible	5
Perceptible but not annoying	4
Slightly annoying	3
Annoying	2
Very annoying	1

C.5.2 Echo intensity

For diagnostic purposes only.

Description:

This rating is intended for diagnostic purposes and needs only to be done if echo is perceived.

The scale is applied for evaluation of the echo level occurring in far-end single-talk and in double-talk test cases. The evaluation is done on the far-end. The test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end Double talk
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings; enclosure dislocation due to movement of the driver; volume setting in the car: nominal, maximum.
Scale type:	Special scale for HF system diagnostic evaluation

The rating scale is given in Table C.13.

Table C.13 – Echo intensity

Rating description	Grade
–	–
Slight	4
Moderate	3
Loud	2
Very loud	1

C.5.3 Echo duration

For diagnostic purposes only.

Description:

This rating is intended for diagnostic purposes and needs only to be done if echo is perceived.

The scale is applied for evaluation of the echo duration occurring in far-end single-talk and in double-talk test cases. The evaluation is done on the far-end. The test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end Double talk
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings; enclosure dislocation due to movement of the driver; volume setting in the car: nominal, maximum.
Scale type:	Special scale for HF system diagnostic evaluation

The rating scale is given in Table C.14.

Table C.14 – Echo duration

Rating description	Grade
Very short	
Short	
Moderate	
Long	
Very long/permanent	

C.5.4 Frequency of echo occurrence

For diagnostic purposes only.

Description:

This rating is intended for diagnostic purposes and needs only to be done if echo is perceived.

The scale characterizes the number of echo events occurring during the echo test, the test cases include far-end single-talk and double-talk. The evaluation is done on the far-end. The test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end Double talk
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings; enclosure dislocation due to movement of the driver; volume setting in the car: nominal, maximum.
Scale type:	Special scale for HF system diagnostic evaluation

The rating scale is given in Table C.15.

Table C.15 – Frequency of echo occurrence

Rating description	Grade
Only once during the test	
Only twice during the test	
Infrequently several times	
Echo occurs more often than not	
Permanent	

C.5.5 Echo intelligibility

For diagnostic purposes only.

Description:

This rating is intended for diagnostic purposes and needs only to be done if echo is perceived.

The scale is applied for characterizing the type of sound of the echo occurring in far-end single-talk and in double-talk test cases (pure artefacts or the echoed voice of the talker). The evaluation is done on the far-end. The test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end Double talk
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings; enclosure dislocation due to movement of the driver; volume setting in the car: nominal, maximum.
Scale type:	Special scale for HF system diagnostic evaluation

The rating scale is given in Table C.16.

Table C.16 – Echo intelligibility

Rating description	Grade
Pure artefacts	
Hardly recognizable as voice	
Distorted voice	
Slightly distorted voice	
Clear voice	

C.5.6 Comfort noise quality (EC test)

Description:

The scale is for evaluation of the near-end background noise sound quality received at the far-end during far-end single talk. The evaluation gives information about the quality of comfort noise injection. Transient noise should be avoided at the near-end during the test.

Test category:	Echo cancellation
Conversation type:	Single talk at far-end
Rating side:	Rated at far-end
Test conditions:	Stationary background noise scenarios of low/medium/high level; additional background noise scenarios: different fan/defrost settings.
Scale type:	Special scale for HF system diagnostic evaluation

The following sound characteristics have to be taken into account:

- are there changes of noise sound and level over time (e.g., when changing from natural near-end background noise to comfort noise injection and vice versa)?
- are artefacts audible (clicks, pops, rattle)?
- naturalness:
  - does the noise sound like it is part of the natural background?
  - does the noise sound synthetic (musical tones, watery sound)?
  - does the noise sound distorted?

Artefacts, synthetic sound, level and sound variation over time, and noise sounding unrelated to the natural background will result in lower (worse) rating grades.

The rating scale is given in Table C.17.

Table C.17 – Comfort noise quality (EC test)

Rating description		Grade
No difference between comfort noise and natural background noise perceivable; Comfortable, natural sound, constant in sound and level, no artefacts	Natural	5
Slight difference between comfort noise and natural background noise perceivable; Slight distorted/synthetic sound, almost no artefacts, almost constant in sound and level	–	4
Moderate difference between comfort noise and natural background noise perceivable; Moderate distorted/synthetic sound, or some artefacts/clicks/plops audible, or some moderate variation in sound and level	–	3
Clear difference between comfort noise and natural background noise perceivable; Clearly unnatural/distorted/synthetic sounding, or many artefacts/clicks/plops, or often variation in sound and level	–	2
Comfort noise does not sound like the natural background noise at all; Very unnatural/distorted/synthetic sound, or permanent artefacts/clicks/plops, or permanent variations in sound and level, very uncomfortable to listen to	Unnatural	1

C.5.7 Speech level variation during double talk

Description:

The scale is applicable to evaluate speech level variations during double-talk. The evaluation is done on both sides, near-end and far-end. When the evaluation is done on the far-end, then the test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Double talk
Rating side:	Rated at near-end Rated at far-end
Test conditions:	For rating at the near-end: • low background noise scenario For rating at the far-end: • stationary background noise scenarios of low/medium/high level; • additional background noise scenarios: different fan/defrost settings; • enclosure dislocation due to movement of the driver; • volume setting in the car: nominal, maximum.
Scale type:	Special scale for HF system diagnostic evaluation

Level variations are characterized by:

- switching of an attenuation exactly during the double talk phases,
- level fading,
- short drop-outs or cut-offs (missing word-ends or syllables),
- chopped voice.

The rating scale is given in Table C.18.

Table C.18 – Speech level variation during double talk

Rating description	Grade
No speech level variation audible	5
Slight level variations that are barely audible or that occur very rarely	4
Moderate speech level variations may occur frequently, sometimes words or syllables might be attenuated or missing, or moderate constant attenuation being switched during the double talk phases	3
Many drop outs, cut-offs, missing words or syllables, heavily chopped voice, or high constant attenuation being switched during the double talk phases	2
Not possible to hear the other side at all during double talk	1

C.5.8 Intelligibility/listening-effort during double talk

Description:

The rating scale is applied to evaluate the effort required to understand the meaning of words and sentences during double talk. The evaluation is done on both sides, near-end and far-end. When the evaluation is done on the far-end, then the test scenarios should include different background noise conditions at the near-end.

Test category:	Echo cancellation
Conversation type:	Double talk
Rating side:	Rated at near-end Rated at far-end
Test conditions:	For rating at the near-end: • low background noise scenario For rating at the far-end: • stationary background noise scenarios of low/medium/high level; • additional background noise scenarios: different fan/defrost settings; • enclosure dislocation due to movement of the driver; • volume setting in the car: nominal, maximum.
Scale type:	Listening-effort scale, [ITU-T P.800], ACR

The rating scale is given in Table C.19.

Table C.19 – Intelligibility/listening-effort during double talk

Rating description	Grade
Every word was clearly understood during double talk with no effort required	5
Speech of the other side is understood during double talk with no appreciable effort required	4
Some words were hard to understand during double talk, moderate listening-effort is required	3
Many words were hard to understand during double talk, considerable listening-effort is required	2
No meaning understood with any feasible effort during double talk	1

C.6 Hands-free system stability tests (car-to-car)

The evaluation for system stability is intended to examine the convergence characteristic of the echo cancellation for "closed loop" operation when doing car-to-car hands-free communication. For the relevant tests, the hands-free system under test is installed in both cars. Moreover, both systems do not have the echo cancellation filter adapted when starting the test.

System stability

Description:

The scale is applied for evaluation of the convergence of the echo cancellation and the robustness against back coupling of echo in car-to-car communication. In one test case, single talk at the far-end is performed, in another test case an impulse-like noise signal is generated close to the microphone on the far end.

In both cars, the EC-filter is not adapted at the start of the test. The evaluation is done on the far-end.

As a suggestion, an appropriate test procedure could be as follows:

Both cars are in a standstill position and have the doors open and the volume set to nominal. Then single talk is performed in both cars, one after the other, to give the EC-filters the chance to adapt to this situation (or not to adapt when the doors are closed afterwards). The volume is then set to maximum in both cars. To generate an impulse-like noise, for example, the doors of the cars could be slammed. Another possibility would be to quietly close the doors and generate impulse-like noise by clapping the hands close to the microphone.

Test category:	System stability
Conversation type:	Single talk at far-end Impulse-like noise at far-end
Rating side:	Rated at far-end
Test conditions:	Initial state of EC-filter: not adapted Volume setting in the car: maximum
Scale type:	Special scale for HF System diagnostic evaluation

The rating scale is given in Table C.20.

Table C.20 – System stability

Rating description	Grade
No echo is audible	5
Some echo can be heard, but disappears very quickly	4
The echo disappears slowly, the recurrences are audible for a few seconds	3
The echo disappears very slowly, the recurrences are audible for more than 10 seconds	2
The echo builds up like in an unstable feedback system	1

Annex D

Standard set of user scenarios

(This annex forms an integral part of this Recommendation.)

Table D.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 (Note 3)**	**Wind speed**	**Precipitation**	**Temp.**
1	Stationary vehicle	0 km/h (at idle)	FAN=Lowest setting	Up	Off	Off	None	N/A	< 5 m/s (12 mph)	None	> –20°C and < 40C
2	City driving	60 km/h (37 mph)	FAN= Medium setting; AIRFLOW=Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20°C and < 40°C
3	Highway driving	120 km/h (75 mph)	FAN=Lowest setting	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20°C and < 40°C
4	Highway driving	120 km/h (75 mph)	FAN=Medium setting; AIRFLOW=Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20°C and < 40°C

Table D.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 (Note 3)**	**Wind speed**	**Precipitation**	**Temp.**
5 (Note 1)	Highway driving	≥ 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	> –20°C and < 40°C

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 – Additional testing should be performed to verify that there are no HVAC vent positions that result in severely degraded performance due to wind buffeting.

NOTE 3 – Smooth road surfaces that generate very little tire 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.

NOTE 4 – 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.

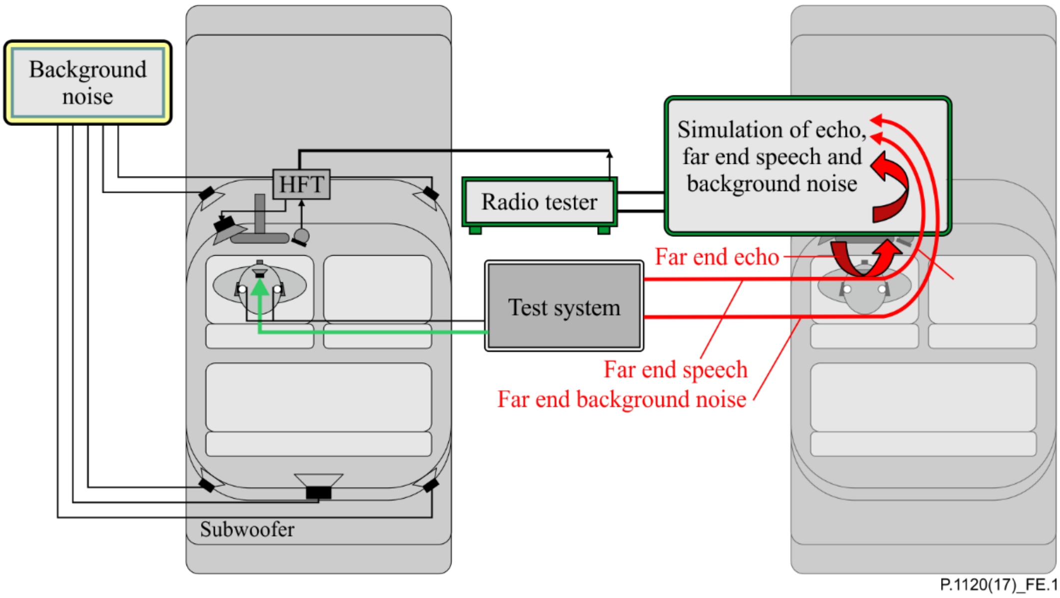
NOTE 5 – The setting of the HVAC shall be documented. Some cars automatically adjust the HVAC settings during a call. In such cases the background noise recording shall be made with the HVAC setting active during the call.

Annex E

System stability with insufficient far-end echo loss

(This annex forms an integral part of this Recommendation.)

Car-to-car communication bears the risk of a closed loop feedback channel, especially when near-end and/or far-end echo loss is low. This may happen, e.g., at the beginning of a call, when ECs are not fully adapted or during echo path changes. In order to simulate this situation, the following laboratory test set-up is used:



Device under test (DUT, laboratory test set-up) Simulated vehicle at far end

Figure E.1 – Test set-up for simulating insufficient far-end echo loss

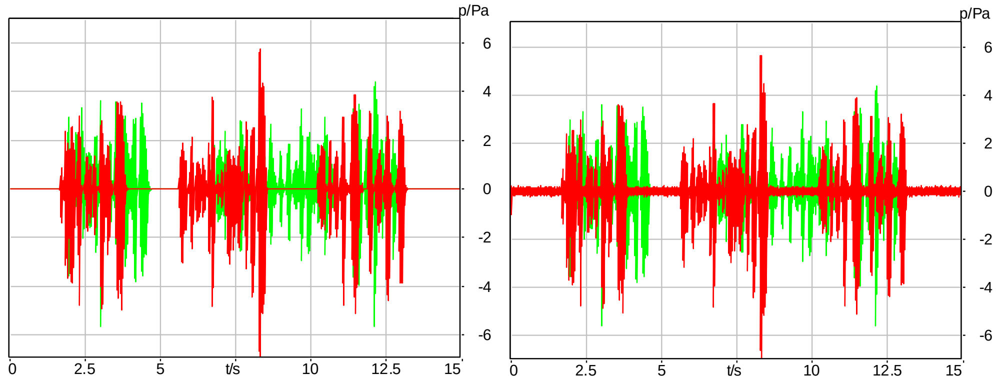
The test set-up used to test car-to-car scenarios in a laboratory environment is shown in Figure E.1. The far-end side is simulated by an echo attenuation (echo return loss (ERL)), echo path delay or measured impulse response in existing cars), the driver's voice (designated as "far end speech" and recorded from an HFT microphone in a test vehicle) and the background noise (recorded from an HFT microphone in a test vehicle).

Real speech, applied under single and double talk conditions for the DUT, is recommended in order to most realistically reproduce car-to-car communication. Two examples are shown above.

- The red coloured signal represents the simulated far-end signal, the green signal is applied via the artificial mouth in the test vehicle (DUT, laboratory set-up).
- The left-hand example represents the test signal without background noise in the far-end vehicle. The right-hand example shows a similar speech sequence but with added background noise at the far-end side.
- The Lombard effect should be considered on both sides in order to simulate realistic scenarios and realistic speech and noise levels. This can be achieved by:
  - i) using real Lombard speech recorded in the test lab; or
  - ii) choosing [ITU-T P.501] test sentences and considering the Lombard effect by the corresponding level adjustment. Guidelines for level adjustments as a function of background noise levels are given in clause 7.1.3;

iii) the different driving conditions described in Annex D shall be considered including a no background noise condition.

- The sequences start with a double talk sequence, followed by a short single talk sequence. In the following graphs, the sending and receiving directions are represented by green and red, respectively.



The transmitted speech signals are recorded in receive direction of the DUT, i.e., in the vehicle in the laboratory set-up. No howling or feedback shall be detected.

Requirements

The minimum far-end ERL still leading to a stable performance is documented.

This test shall be conducted for an echo path realization by a pure attenuation or by a simulated impulse response measured from a real car.

If different implementations are to be compared, it is recommended to simulate the far-end echo path by an attenuation.

Test

1) The test arrangement is according to Figure E.1.

2) The tests are carried out under the following test conditions at the far-end side:
   - Test case 1 (mandatory): Simulation of ERL by pure attenuation, echo path delay 0 ms, variation of ERL from 50 dB to 0 dB (under study) in steps of 5 dB.
   - Test case 2 (customized): Simulation of ERL and echo path delay by measured impulse response in a real vehicle, variation of ERL from 50 dB to 0 dB (under study) in steps of 5 dB.

3) Each test shall be carried out with the EC coefficients initially cleared (reset, new call set-up) or starting from any artificial state (e.g., by converging EC with open drivers' door).

4) The recorded signals from the HATS (free-field or diffuse-field equalized (see clause 7.1.6.1)) are judged subjectively by the test engineers.

Annex F

Microphone related sound field reproduction technique – multi-point noise simulation (MPNS) method

(This annex forms an integral part of this Recommendation.)

[b-ETSI TS 103 224] describes a background noise simulation method for handset, headset, and hands-free terminals, which are evaluated in an acoustically treated test room of a certain size (between 1,8 m × 2,4 m × 2,1 m to 8 m × 9 m × 4,5 m). It requires eight loudspeakers for the desired sound field reproduction, which are distributed around the HATS in a certain way to cover the most important directions of sound incidence in a typical background noise scenario. The necessary recordings require the same special microphone arrangement for all the use cases, where the eight microphones are arranged approximately in a circle with special spacings and vertical positions.

In motor vehicles, most of these requirements cannot be fulfilled: the vehicle interior is usually much smaller and the possible loudspeaker positions are much more limited. Furthermore, the required points of equalization are much more wide spread, which makes the circular microphone arrangement inapplicable. This also raises the need for another definition of a "fine-tuning set" of microphones which in [b-ETSI TS 103 224] are defined by a 10° rotation of the original ("calibration") set of microphones. The required adaptations of [b-ETSI TS 103 224] in order to use the multi-point noise simulation (MPNS) method successfully are described in the following.

F.1 Loudspeaker set-up in the car

The multi-point noise simulation method requires at least four loudspeakers, which are arranged as described in [b-ETSI ES 202 396-1]. If more than four loudspeakers are used, the loudspeakers should not be placed close to each other but distributed in space as much as possible. The loudspeakers should not be placed in the "microphone space", i.e., between the microphones.

If the loudspeakers cannot reproduce frequencies down to 50 Hz, an additional subwoofer is needed. In this case, the weighted version of the Thikonov-equalization (Equation 9 in clause 6.2.4.1 of [b-ETSI TS 103 224] is used during the inversion procedure, where the diagonal weighting matrix **W** accounts for the different frequency ranges of the loudspeakers.

Accordingly, the MPNSmethod is compatible with the loudspeaker arrangement of [b- ETSI ES 202 396-1] (four loudspeakers plus subwoofer).

F.2 Microphone set-up in the car

No fixed microphone arrangement is required as it has to be arranged individually for each car's interior and for each application. Either the output signal of the microphones of the device under test (DUT) can be used directly for equalization or additional measurement microphones are placed in the vehicle interior close to the DUT microphones.

In any case, the same microphone arrangement is used for the recording of the background noise (see below) and the equalization of the system. The microphone characteristics like frequency or phase response and directionality shall be the same for both use cases.

[b-ETSI TS 103 224] differentiates between the "calibration set" of microphone positions which are used in all steps of the equalization procedure and the "fine-tuning set" of microphone positions which are only used in clauses 6.2.4.3.2 and 6.2.5 of [b-ETSI TS 103 224]. As the microphone arrangement is flexible the microphones have to be assigned to the calibration set and to the fine-tuning set. The fine-tuning set may be empty but, if "fine-tuning microphones" are used they shall be placed near the "calibration set" microphones, i.e. with a maximum distance of 5 cm.

If the DUT microphones can be used directly, the number of microphones in the "calibration set" must not exceed the number of loudspeakers in the set up excluding a potential subwoofer.

If it is not possible to use the DUT microphones directly, the number of loudspeakers shall be at least twice the number of microphones in the "calibration set". In general, it is recommended to place at least two measurement microphones near each location of DUT microphone(s) to create an "equalized zone" around the DUT microphone location. If it is not possible to use twice as many loudspeakers than microphones with this microphone arrangement, one of the two measurement microphones is assigned to the fine-tuning set of microphone positions.

F.3 Background noise recordings and reference noise

Individual background noise recordings are made for each vehicle and microphone arrangement.

No "default" reference noise as in [b-ETSI TS 103 224] can be given for the proposed multi-point noise simulation method. Instead, one portion of the vehicle individual background noise recordings is selected as reference noise. The selected reference noise shall be 5 to 10 s long and representative for the background noise recordings in the individual car.

F.4 Equalization

In general the equalization process is described in [b-ETSI TS 103 224]. The matrix size (used to identify the system and to calculate the matrix inversion for equalization) depends on the number of loudspeakers and microphones chosen for the background noise simulation. The following adaptations are required for the car:

In clause 6.2.3 of [b-ETSI TS 103 224], the impulse responses are pre-processed with a low-pass filter with a time-variant cut-off frequency to avoid a degradation of the quality of the inversion by high-frequency components of the tails of the impulse responses. Since the reverberation time in vehicle interiors is negligible compared to treated office rooms, the low-pass filtering is omitted for the proposed multi-point noise simulation method.

In clause 6.2.4.2 of [b-ETSI TS 103 224], different microphones are selected to calculate the inversion filters for each frequency band based on the corresponding wavelength. For the MPNS-method, the microphone distances are unknown and therefore this type of optimization is not applied. All microphones are used across the whole frequency range to calculate the inversion filters.

In clause 6.2.4.3 [b-ETSI TS 103 224], the optimum regularization factor is searched subject to a limited loudspeaker output level to avoid distortion of the loudspeaker signals. "*A spectrum representative for the maximum level and spectral content*" ([b-ETSI TS 103 224], clause 6.2.4.3.1) can be calculated from the reference noise which was selected in F.3. For the MPNS method, the level of the filtered reference noise radiated by each loudspeaker to each microphone position shall not exceed the level which a pink noise of 0 dB [Pa] would generate in the same FFT-bin.

In clause 6.2.5 of [b-ETSI TS 103 224], a correction filter D(f) is applied to compensate, *inter alia*, for the arithmetic uncertainties (e.g. the regularization factor) of the inversion. This correction filter consists of a constant factor below 1.8 kHz and a minimum phase filter above. For the MPNS-method, the microphone distances are unknown and therefore the lower cut-off frequency is reduced to 50 Hz. In order to retain the phase in the previously constant frequency range below 1.8 kHz, a zero-phase-filter is used instead of a minimum-phase filter for the frequency range above 50 Hz.

F.5 Accuracy of the equalization

After a successful equalization the following criteria, which are comparable to those of [b-ETSI TS 103 224], shall be met:

1) Level accuracy

The level of the reproduced sound field at each microphone of the calibration set shall be accurate within ±1 dB.

2) Magnitude and phase of the cross correlation between reference noise recording and simulated reference noise recording at the calibration position

The magnitude of the cross correlation between the reference noise recording (see clause F.3) and the reproduced signals at the calibration set positions averaged over the individual microphones shall fulfil the following requirements:

i) In the frequency range from *100 Hz* to *1 kHz* the magnitude of the complex coherence (normalized cross correlation spectrum) shall be larger than 0.9, measured in 1/3rd octaves.

ii) In the frequency range from *100 Hz* to *1 kHz* the phase of the complex coherence shall be accurate within ±10 degrees and within ±30 degrees in the range from *1 kHz* to *1.5 kHz*, both measured in 1/3rd octaves.

3) Spectrum reproduction accuracy

The difference between the amplitude spectrum of the original reference noise recording and the amplitude spectrum of the simulated reference noise recording (both measured in dB) for the individual microphones shall be within ±3 dB, measured in 1/3rd octaves from 50 Hz to 10 kHz and ±6 dB from 10 kHz to 16 kHz. The average spectrum accuracy, averaged over all microphones shall be within ±3 dB from 50 Hz to 20 kHz.

Bibliography

[b-ITU-T G.114] Recommendation ITU-T G.114 (2003), *One-way transmission time*.

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

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

[b-ITU-T P.51]. Recommendation ITU-T P.51 (1996), *Artificial mouth*.

[b-ITU-T P.1130] Recommendation ITU-T P.1130 (2015), *Subsystem requirements for automotive speech services*.

[b-3GPP TS 26.131] 3GPP TS 26.131 v12.3.0 (2014-12-22), *Terminal acoustic characteristics for telephony; Requirements*.

[b-3GPP TS 26.132] 3GPP TS 26.132 v12.5.0 (2015-03-21), *Speech and video telephony terminal acoustic test specification*.

[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 TS 103 224] ETSI TS 103 224 (2015), *Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database*. <http://pda.etsi.org/pda/queryform.asp>

[b-Kettler] Kettler, F., and Gierlich, H.W. (2008), *Evaluation of Hands-Free Terminals*, Topics in Speech and Audio Processing in Adverse Environments, edited by E. Hänsler, G. Schmidt, Springer, ISBN: 978-3-540-70601-4.

[b-SAE J941] SAE J941 *201003: Surface vehicle recommended practice.*

[b-Sottek] Sottek, R., Genuit, K. (2005), *Models of signal processing in human hearing*, International Journal of Electronics and Communications, pp. 157-165.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
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	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling, and associated measurements and tests
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, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2017