International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1150

(01/2020)

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

Communications involving vehicles

In-car communication audio specification

Recommendation ITU-T P.1150

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	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	**P.1100–P.1199**
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	P.1300–P.1399
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	P.1500–P.1599

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

Recommendation ITU-T P.1150

In-car communication audio specification

Summary

Recommendation ITU-T P.1150 describes communication requirements for an in-car communication (ICC) system that utilizes integrated microphones and speakers in the motor vehicle cabin to amplify conversation and which is aimed at providing an improved communication between all occupants in a motor vehicle.

This Recommendation sets base levels of function and quality to ensure that the ICC operates to a quality such that the motor vehicle driver does not feel it necessary to turn their head to amplify their voice when talking to other passengers. This can, for example, aid in preventing driver distraction.

To meet these requirements, the use of audio zones within a motor vehicle cabin for speech communications between audio zones, and tests to ensure good speech intelligibility and quality between those zones are defined. This Recommendation covers requirements and test methods for:

1) system stability;
2) speech intelligibility;
3) speech quality;
4) talker localization accuracy.

The use of wearable headphones (audiophones) by the driver or passengers is outside the scope of this Recommendation.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1150	2020-01-13	12	11.1002/1000/14154

Keywords

Communication, hands-free, motor vehicle, QoS, quality of service.

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

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

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

© ITU 2020

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	2
	3.1 Terms defined elsewhere	2
	3.2 Terms defined in this Recommendation	2
4	Abbreviations and acronyms	3
5	Conventions	3
6	How to use this Recommendation	3
	6.1 Determining compliance with this Recommendation	3
7	Test arrangement	4
	7.1 Zones and measurement paths	4
	7.2 Talker-listener combinations under test	5
	7.3 Test arrangement in a car	5
	7.4 ON/OFF control	6
	7.5 ICC setting	6
	7.6 Side information	7
8	Test signal and signal levels	7
	8.1 Speech signals	7
	8.2 Background noise	9
	8.3 Removal of competing signals by time-domain subtraction	10
	8.4 Other test signals	10
	8.5 Calculation of signal-to-noise ratio, speech and noise level	11
9	Measurement parameters and requirements for ICC	12
	9.1 Stability	12
	9.2 Delay	13
	9.3 Speech Intelligibility	14
	9.4 Listening effort	15
	9.5 Talker localization accuracy	16
	9.6 Adaptive reinforcement	16
	9.7 Signal-to-noise ratio	16
	9.8 Speech quality	18
	9.9 Double talk attenuation	18
	9.10 Media Interaction	19
	9.11 Background noise transmission	20
Annex A – Standard set of user scenarios		21
Annex B – Delay calculation via impulse response measurement		23
Annex C – Delay calculation via cross-correlation		24

		Page
Annex D – Calculations for stability assessment		25
D.1	Introduction	25
D.2	Spectrum vs time	25
D.3	Activity classification	25
D.4	Analysis ranges	26
D.5	Calculation of stability assessment (energy-present above background)	26
Annex E – Calculation of separated speech and noise levels		27
E.1	Introduction	27
E.2	Spectrum vs time	27
E.3	Activity classification	27
E.4	Estimation of average spectra	28
E.5	Speech/noise level and SNR	29
Appendix I – Details on Lombard speech		30
I.1	With ICC reinforcement disabled	30
I.2	With ICC reinforcement enabled	31
Appendix II – Guidance for subjective evaluation of system stability		36
Bibliography		38

Electronic attachment: Concatenated test signals.

Recommendation ITU-T P.1150

In-car communication audio specification

1 Scope

This Recommendation[1] describes communication requirements for an in-car communication (ICC) system that utilizes integrated microphones and speakers in the motor vehicle cabin to amplify conversation and which is aimed at providing an improved communication between all occupants in a motor vehicle.

This Recommendation sets base levels of function and quality to ensure that the ICC operates to a quality such that the motor vehicle driver does not feel it necessary to turn their head to amplify their voice when talking to other passengers. This can, for example, aid in preventing driver distraction.

To meet these requirements, the use of audio zones within a motor vehicle cabin for speech communications between audio zones, and tests to ensure good speech intelligibility and quality between those zones are defined. This Recommendation covers requirements and test methods for:

1) system stability;
2) speech intelligibility;
3) speech quality;
4) talker localization accuracy.

The use of wearable headphones (audiophones) by driver or passengers is outside the scope of 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 G.160] Recommendation ITU-T G.160 (2012), *Voice enhancement devices*.

[ITU-T P.10] Recommendation ITU-T P.10 (2017), *Vocabulary for performance, quality of service and quality of experience*.

[ITU-T P.56] Recommendation ITU-T P.56 (2011), *Objective measurement of active speech level*.

[ITU-T P.57] Recommendation ITU-T P.57 (2011), *Artificial ears*.

[ITU-T P.58] Recommendation ITU-T P.58 (2013), *Head and torso simulator for telephonometry*.

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

[1] This Recommendation includes an electronic attachment containing the concatenated test signals specified in clause 8.4.2.

[ITU-T P.501] Recommendation ITU-T P.501 (2017), *Test signals for use in telephonometry*.

[ITU-T P.581] Recommendation ITU-T P.581 (2014), *Use of head and torso simulator for hands-free and handset terminal testing*.

[ITU-T P.1100] Recommendation ITU-T P.1100 (2019), *Narrowband hands-free communication in motor vehicles*.

[ITU-T P.1110] Recommendation ITU-T P.1110 (2019), *Wideband hands-free communication in motor vehicles*.

[ETSI TS 103 224] ETSI TS 103 224 V1.3.1 (2017), *Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database*.

[ETSI TS 103 558] ETSI TS 103 558 V.1.1.1 (2019), *Speech and multimedia Transmission Quality (STQ); Methods for objective assessment of listening effort*.

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

[IEC 61672-1] IEC 61672-1:2013, *Electroacoustics – Sound Level Meters– Part 1: Specifications*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 fullband signal: [ITU-T P.10].

3.1.2 motor vehicle: [ITU-T P.1100].

3.1.3 narrowband signal: [ITU-T P.10].

3.1.4 super-wideband signal: [ITU-T P.10].

3.1.5 wideband signal: [ITU-T P.10].

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 direct voice delay: The time between the transmission of voice from the talker's mouth and reception by the listener's ear, where talker and listener are physically close enough to converse.

3.2.2 reinforcement: An increase in speech level, normally done in the presence of noise.

3.2.3 reinforced noise: Vehicle interior noise processed through an active in-car communication (ICC) system.

3.2.4 reinforced speech: Speech signal originated from talker zone processed through an active in-car communication (ICC) system.

3.2.5 reinforcement voice delay: The time between the transmission of voice from the talker's mouth through the motor vehicle sound system and received by the listener's ear in the motor vehicle cabin.

3.2.6 speech in receive zone: Speech signal produced by the head and torso simulator (HATS) in the receive zone.

3.2.7 speech in transmit zone: Speech signal produced by the head and torso simulator (HATS) in the transmit zone.

3.2.8 talker localization: The listener's ability to identify the direction and distance of the detected talker.

3.2.9 vehicle interior noise: Actual noise originated from the vehicle inside the cabin, with no in-car communication (ICC) interaction.

3.2.10 zone: An acoustic space inside a motor vehicle, normally at a single seat position.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ASL	Active Speech Level
dBPa	Sound pressure level relative to 1 pascal (Pa), expressed in decibels (dB)
DT	Double-Talk
HATS	Head and Torso Simulator
HFRP	Hands-Free Reference Point
HVAC	Heating Ventilation and Air Conditioning
ICC	In-Car Communication
MOS_{LE}	Mean Listening Effort Opinion Score
MRP	Mouth Reference Point
Rx	Receive direction
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
ST	Single-Talk
Tx	Transmit direction
VIN	Vehicle Interior Noise

5 Conventions

None.

6 How to use this Recommendation

6.1 Determining compliance with this Recommendation

To claim compliance with this Recommendation, the following must be true:

1) the test arrangement shall follow clause 7;
2) the test signals in clause 8 shall be used as specified in the specific tests in clause 9;
3) the standard set of usage scenarios in Annex A shall be followed;
4) the requirements for every specific test in clause 9 where both requirements and test method(s) are defined, shall be fulfilled.

7 Test arrangement

This clause discusses the physical environment and equipment configuration for testing ICC.

7.1 Zones and measurement paths

Figures 7-1 and 7-2 show a typical motor vehicle seating configuration for a three-row passenger vehicle for left-hand and right-hand drive vehicles. Each seating position is a designated zone for a talker or listener to engage with an ICC system.

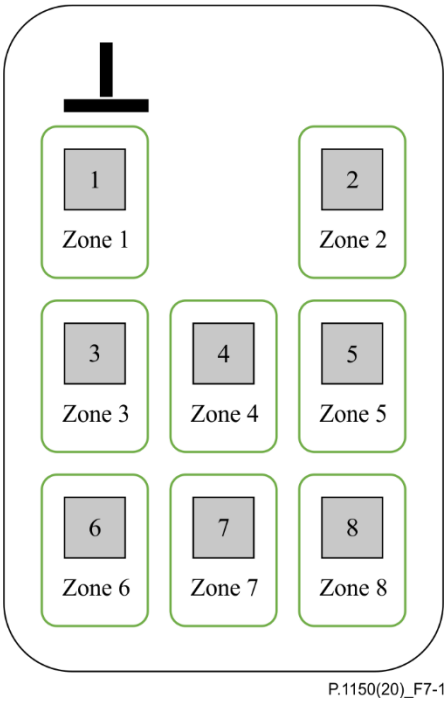


Figure 7-1 – Left-hand drive common passenger motor vehicle seating positions

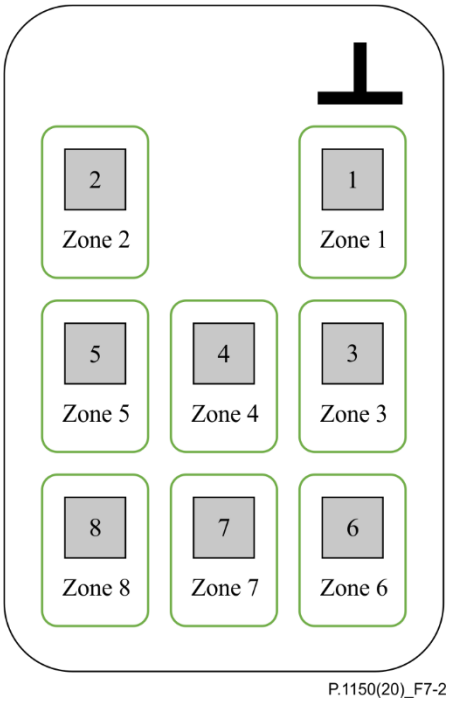


Figure 7-2 – Right-hand drive common passenger motor vehicle seating positions

In reference to Figures 7-1 and 7-2, Table 7-1 shows the ICC paths that may be enabled providing the vehicle is configured with the seating positions. Inter-zone communication between immediately adjacent seats in 3-seat rows are ignored.

Table 7-1 – Zone to Zone communications paths

Zone Interaction	1 Rx	2 Rx	3 Rx	4 Rx	5 Rx	6 Rx	7 Rx	8 Rx
1 Tx	NA	✓	✓	✓	✓	✓	✓	✓
2 Tx	✓	NA	✓	✓	✓	✓	✓	✓
3 Tx	✓	✓	NA	AJ	✓	✓	✓	✓
4 Tx	✓	✓	AJ	NA	AJ	✓	✓	✓
5 Tx	✓	✓	✓	AJ	NA	✓	✓	✓
6 Tx	✓	✓	✓	✓	✓	NA	AJ	✓
7 Tx	✓	✓	✓	✓	✓	AJ	NA	AJ
8 Tx	✓	✓	✓	✓	✓	✓	AJ	NA

AJ – adjacent zones

✓ – ICC operates for this zone pair in the direction indicated

✓ – mandatory inter-zone testing configurations

NA – not applicable

Rx – receive direction

Tx – transmit direction

The above descriptions are a model for test configurations. The manufacturer shall make sure that the system works for all zones and combinations the system is intended to work in. A typical test set-up may be a subset; specifically, it may depend from the vehicle type (e.g., sedan, convertible, van) and symmetries may be exploited. Any seat is considered as a separate zone.

Mandatory tests and arrangement required are defined in clauses 7.2 through 7.6 of this Recommendation.

7.2 Talker-listener combinations under test

7.2.1 Adjacent zone testing

Testing between adjacent zones indicated in Table 7-1 is optional.

7.2.2 Testing driver and all zones

Inter zone testing of 1-3, 1-5, 1-6, and 1-8 (if applicable) is mandatory. Testing between the driver zone (zone 1) and the co-driver zone (zone 2) is mandatory for convertibles.

7.3 Test arrangement in a car

7.3.1 Introduction

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

All HATS are connected to a test equipment conforming to the required standard with an audio interface possible to handle calibration and equalization.

An example for a test arrangement is shown in Figure 7-3.

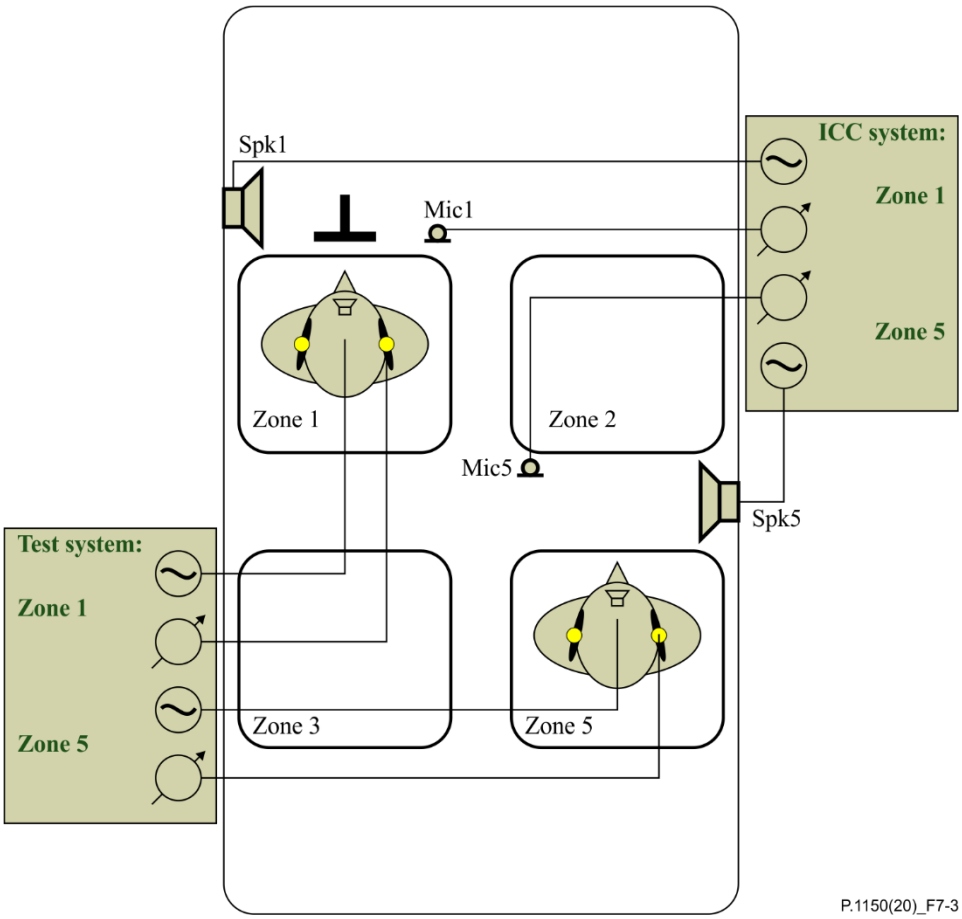


Figure 7-3 – Example of test equipment placement in a car cabin with zones

7.3.2 Artificial mouth

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

With sound pressure level relative to 1 pascal, expressed in decibels (dBPa), the sound pressure level (SPL) is calibrated at the HATS hands-free reference point (HATS-HFRP) to a level of 25.7 dBPa. The SPL at the MRP has to be corrected correspondingly. The detailed description for equalization at the MRP and level correction at the HATS-HFRP can be found in [ITU-T P.581].

7.3.3 Artificial ears

The HATS is equipped with artificial ears of type 3.3 or 3.4 according to [ITU-T P.57]. If not specified otherwise, the ear signals of both ears of the artificial head are used with diffuse-field equalization. More detailed information can be found in [ITU-T P.581].

7.4 ON/OFF control

For testing purposes an ICC ON/OFF control shall be made available.

An end user-accessible ICC ON/OFF control is recommended.

7.5 ICC setting

ICC implementations may allow the users to change the ICC behaviour according to their preferences. For the testing, the ICC system shall be set to a predefined default setting. For test regarding stability, the test shall be done also with the setting which applies the highest gain to the signal.

7.6 Side information

ICC systems may use side information such as vehicle speed to control the ICC system. The ICC vendor shall inform the test house whether it requires use of side information. The method of providing this side information is outside the scope of this recommendation and is solely the responsibility of the ICC vendors and test houses. If such information is provided it shall be used for the relevant tests and its accuracy agreed between the ICC system vendor and the test laboratory.

8 Test signal and signal levels

8.1 Speech signals

8.1.1 Speech material

If not specified otherwise, the British English fullband speech samples according to Annex C of [ITU-T P.501] shall be used as a concatenated sequence. Each sentence shall be centred in a time window of 4.0 s, leading to an overall signal of 32.0 s duration.

Table 8-1 provides a detailed description of the exact crop and merge positions in order to compose the test sequence, which is depicted in Figure 8-1.

Table 8-1 – Composition of test sequences

[ITU-T P.501] Annex C filename	Crop Start [s]	Duration [s]	Merge Position [s]
P501_C_english_m1_FB_48k.wav	0.58	2.15	0.93
P501_C_english_m1_FB_48k.wav	3.73	2.22	4.89
P501_C_english_m2_FB_48k.wav	0.54	2.55	8.73
P501_C_english_m2_FB_48k.wav	4.19	2.63	12.69
P501_C_english_f1_FB_48k.wav	0.52	2.73	16.64
P501_C_english_f1_FB_48k.wav	4.13	3.13	20.44
P501_C_english_f2_FB_48k.wav	0.50	2.64	24.68
P501_C_english_f2_FB_48k.wav	4.07	2.46	28.77

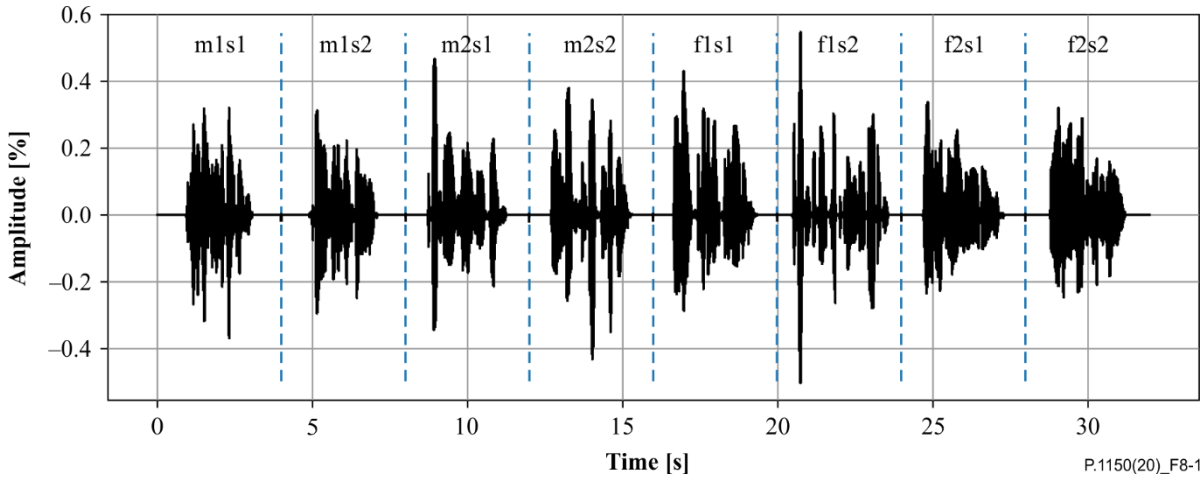


Figure 8-1 – Composed test sequence

The sequence is calibrated to an active speech level (ASL) according to [ITU-T P.56] of –1.7 dBPa at the MRP, which corresponds to –25.7 dBPa at the HFRP.

If not specified otherwise, the Lombard gains as specified in the following clauses (for ICC ON/OFF) shall be applied on the test signal.

NOTE – In general, the widely used speech level of −4.7 dBPa is increased by +3 dB according to [ITU-T P.340]. This gain typically is observed in communication scenarios, where test subjects are forced to talk via hands-free microphone/system. Similar results were found in situations, where the talker is aware that the listener is located further away (e.g., talking from front to back seat), see Appendix I.

8.1.2 Lombard effect with activated ICC

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. To be consistent with the intent that an effective ICC system should significantly reduce the requirement for the talker to raise their voice in the presence of noise, the simulated Lombard effect for ICC is calculated as half the Lombard effect observed when ICC is not present. The level is increased by 1.5 dB for every 10 dB that the long-term A-weighted noise level exceeds 50 dB(A), see Appendix I for more information. This relationship is shown in the following formula:

$$I(N) = \begin{cases} 0 \; for \; N < 50 \\ 0.15 \; (N - 50) for \; N \; \geq 50 \end{cases}$$

where:

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

N = The long-term A-weighted noise level measured close to the talker's head position

As an example, if the vehicle noise measures 70 dB(A), then the output of the mouth would be increased by 3 dB. No gain is applied for noise levels below 50 dB(A). The Lombard level increase shall be limited to 8 dB. Vehicle noise levels are measured using HATS microphones of the talker. The average level between left and right ear signals is used for the Lombard level calculation formula.

8.1.3 Lombard effect with deactivated ICC

Several studies indicated that the Lombard effect is more pronounced in environments without speech reinforcement (see Appendix I) or in noisy hands-free communication scenarios [ITU-T P.1100].

In cases when measurements with deactivated ICC system are conducted and if not specified otherwise, the Lombard level increase shall be twice the amount of the value as specified in clause 8.1.2. The Lombard level increase shall be limited to 8 dB. The average level between left and right ear signals is used for the Lombard level calculation formula.

As an example, if the vehicle noise measures 70 dB(A), then the output of the mouth would be increased by 3 dB with activated ICC and 6 dB without activated ICC.

8.1.4 Convergence sequence

In order to allow the ICC system to adapt its signal processing to the current measurement scenario (speech and background noise level), an appropriate convergence sequence is used in some tests. For this purpose, the British English short conditioning sequence according to clause 7.3.7.1 of [ITU-T P.501] shall be used. The male and female conditioning sequences are added, leading to a condition sequence of 10 s duration.

Unless stated otherwise, the ASL according to [ITU-T P.56] shall be calibrated to the level described in the test, where this sequence is used. Figure 8-2 illustrates the composition of the convergence sequence.

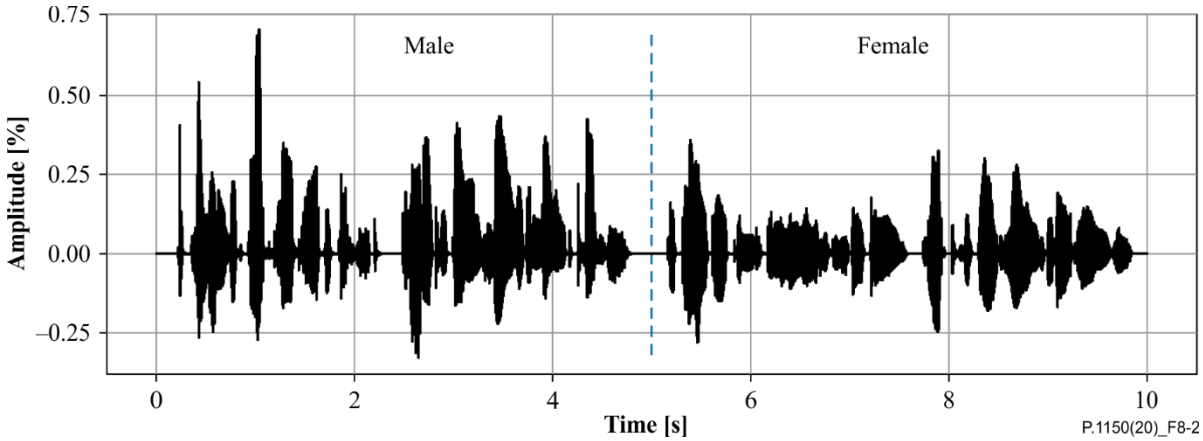


Figure 8-2 – Convergence sequence

8.1.5 Short speech signal

For some measurements, applying a much shorter test signal is required (e.g., determination of delay). In this case, the British English composed speech sample according to Annex D of [ITU-T P.501] in fullband shall be used.

8.1.6 Double talk signal

For concurrent talker scenarios, the British English single-talk (ST) and double-talk (DT) sequences according to clause 7.3.5.1 of [ITU-T P.501] shall be used. Both signals are calibrated and adjusted according to clauses 8.1.1 and 8.1.2.

If not specified otherwise, the single talk speech sequence is used for the HATS simulating the talker (transmit zone) and double talk sequence is used for the HATS simulating the listener (receive zone).

8.2 Background noise

8.2.1 Recordings and test set-ups

Some tests run under constant background noise conditions. For these tests the background noise, which represents a typical driving condition, is inserted. Suitable recordings of background noise are needed for all ICC microphones and listener positions intended to be covered in all noise conditions. The background noise recording shall follow Annex F of [ITU-T P.1100].

When conducting test conditions with background noise, the clauses below describe two alternative options which are recommended to obtain microphone signals with background noise.

Any of these options can be used in all measurements where background noise simulation is required.

In case the inserted noise recording is shorter than required for the measurement (duration including convergence sequence), the noise playback shall be looped with a 20 ms crossfade at the loop point. This applies for all noise reproduction options.

8.2.1.1 Reproduction option 1

The background noises are simulated acoustically using a noise-field simulation for vehicle-type environment as described in clause 7 of [ETSI TS 103 224].

The microphone signals for the noise field simulation shall be simultaneously captured close to all input microphones of the ICC system (at least one) and for each HATS per listener position (binaural). The equalization procedure is carried out for these minimum three microphone signals with at least four loudspeakers. For each evaluated zone, it is recommended to conduct separate noise recordings (at input microphones and HATS ears) and thus, a new equalization of the simulation.

8.2.1.2 Reproduction option 2

The background noises are (digitally) inserted into the ICC system and added (digitally) to the HATS ear signals in order to obtain realistic binaural signals. In this case, the noises of all microphones shall be recorded simultaneously, including the ones of the HATS. This option requires access to the internals of an ICC system, but allows simulation without acoustic noise playback. For details on this approach please see clause 8.2.1 in [ITU-T P.1110].

8.2.2 Recommended test conditions

Test background noise and test scenarios follow test conditions given in Annex A.

Testing with background noise may include scenarios where there is no noise.

8.3 Removal of competing signals by time-domain subtraction

Most of the recording scenarios include several signal components, like e.g., direct path speech, vehicle interior noise (VIN) as well as reinforced speech and noise. For some analyses of this Recommendation, it is necessary to compensate the recording for one or more of these components. Since in many cases the single component has to be activated during a measurement to trigger or activate the ICC system, the procedure described below can be applied to any linear and time-invariant transmitted signals. The method is not applicable for non-linear and/or time-variant signals like e.g., reinforcement speech or noise.

EXAMPLE: In order to analyse solely the reinforcement speech and noise signals received by the HATS without the superimposed vehicle interior noise, the compensation procedure can be applied by conducting a recording without any active speech signals.

The start of the recording of the artificial ear signal(s) and the start of the background noise playback shall be synchronized to ensure consistent timing of repeated recordings and allow cancellation of the component by subtraction in the time domain as described below.

To obtain a compensated signal, the following steps are carried out:

1) set the ICC system and the measurement equipment into the wanted operational mode (e.g., background noise simulation and/or speech signals are applied). Then, a recording of the artificial ear ear-drum reference point (DRP) signal is performed, including all signal components. The playback is stopped, and the recording is saved as *file1*;
2) repeat the recording procedure described in step 1), but this time only with the activated signal component (e.g., only background noise or talker speech). The playback is stopped, and the recording is saved as *file2*;
3) align the files with sample accuracy. A subtraction of the samples is performed (*file1-file2*). The resulting time signal contains a compensated signal, where the targeted component has been essentially eliminated.

The methodology applies for the following components recorded at the HATS in the receive zone:

– speech of talker HATS in transmit zone (direct path);
– speech of talker HATS in receive zone (for e.g., double talk tests);
– vehicle interior noise.

8.4 Other test signals

8.4.1 Maximum-length sequence

Maximum-length sequences (MLS) is the name of a pseudo-random signal consisting of the values +1 and –1. It is periodic with the period $L = 2^N - 1$ and has a flat spectrum (similar to white noise). The advantage over sweep signals is that MLS signals provide good signal-to-noise ration due to the low crest factor. For determining the impulse response from a MLS source and a measured signal, a

simplified deconvolution algorithm (Hadamard transform) can be used. Generation and impulse response calculation can be e.g., found in [b-Cohn].

The number of periods (P) (i.e., repetitions of a base sequence) can be used to average P signals of length (L) to cancel out additive noise during measurement. For P > 1, the first period shall always be omitted.

If not specified otherwise, order N=17 and P ≥ 10 shall be used as a test signal.

8.4.2 Stability test signal

The stability test signal consists of various, concatenated test signals that shall be used for testing and is described in the following. This test signal is provided as an electronic attachment to this Recommendation.

In order to stress the ICC system and try to provoke instability, a modified version of the British English test sequence (see clause 8.1.1) is used as a basis. Timing and gain levels have been modified and some speech signals have been replicated. Additionally, short snippets have been inserted from Annex D of [ITU-T P.501]:

– P501_D_GE_fm_FB_48k.wav (German sample);

– P501_D_CN_fm_FB_48k.wav (Chinese sample);

– P501_D_AM_fm_FB_48k.wav (American English sample).

Since Chinese/Mandarin language contains several phonemes that may trigger instability issues in ICC systems, some snippets are taken from clause 7.4.1 of [ITU-T P.501]:

– FB_male_female_single-talk_seq.wav;

– FB_male_female_double-talk_seq.wav (left channel).

Non-voice sweeps were generated and include:

– a pure tone sweeps from 150 Hz to 8 kHz, duration 0.3 s;

– harmonic tone sweeps with a base frequency of 150 Hz to 8 kHz, durations 0.3 s and 0.5 s. Harmonics at multiples of 1.3, 1.5, 1.7 and 2.0x base frequency with relative amplitudes of 0.5, 0.4, 0.3 and 0.2 base frequency amplitude were included. The overall frequency range of this sweep spans 150 Hz to 16 kHz.

An initial silence period of 8.0 s is included in the test sequence, which allows the ICC system to settle in advance to the test.

The test signal is calibrated to an overall ASL according to [ITU-T P.56] (fullband mode) of −23.2 dBov and an overall activity of 52%. This corresponds to the first utterance of the signal (7.5 s – 10.4 s) calibrated to −26.0 dBov.

For the playback, the sequence is calibrated to an ASL according to [ITU-T P.56] of +1.1 dBPa at the MRP, which corresponds to −22.9 dBPa at the HFRP.

8.5 Calculation of signal-to-noise ratio, speech and noise level

For several tests described in this Recommendation, signal-to-noise ratio (SNR) and/or speech noise level have to be calculated. This clause describes the calculation methods.

A binaural speech signal $s(k)$ recorded with a HATS at the listener position, which is degraded by background noise $n(k)$, can be described for the left (L) and right (R) ear channel according to equations 1 and 2; $d_L(k)\ and\ d_R(k)$ denote the degraded signal.

$$d_L(k) = s_L(k) + n_L(k) \tag{1}$$

$$d_R(k) = s_R(k) + n_R(k) \tag{2}$$

Since most analyses on binaural signals are carried out independently on the left and right ear, the subscript X is used instead of L/R in the following, as shown in equation 3.

$$d_X(k) = s_X(k) + n_X(k) \tag{3}$$

The speech and noise components can further be divided into sub-signals. The speech signal consists of a direct (subscript D) and a reinforcement (subscript R) path. Equally, the noise signal is divided into a vehicle interior (subscript V) and reinforcement (subscript R) portion. The sum of these sub-signals is shown in equation 4.

$$d_X(k) = \left(s_{X,D}(k) + s_{X,R}(k)\right) + \left(n_{X,V}(k) + n_{X,R}(k)\right) \tag{4}$$

Another representation is shown in equation 5. Here the sub-signals in brackets refer to the specific measurement scenario with ICC systems.

$$d_X(k) = \overbrace{s_{X,D}(k)}^{Talker} + \overbrace{n_{X,V}(k)}^{BGN=on} + \overbrace{\left(s_{X,R}(k) + n_{X,R}(k)\right)}^{d_{X,R}(k)} \tag{5}$$

Several of these sub-signals can be measured independently, like the direct path of the talker's signal $s_{X,D}(k)$ or the vehicle interior noise $n_{X,V}(k)$. In case of time-synchronous playback and record, $d_{X,R}(k)$ can be determined by subtraction (see clause 8.3). It contains the speech signal from the talker and noise inserted via input microphone(s) into the ICC system.

The SNR (in dB) of a noisy signal is given as the level ratio between solely speech ($\mathcal{L}(s_X)$) and noise-only ($\mathcal{L}(n_X)$) according to equation 6 (with X=L/R).

$$\mathrm{SNR}_X = 20 \cdot \log_{10} {\mathcal{L}(s_X)}\big/{\mathcal{L}(n_X)} = 20 \cdot \log_{10} {\mathcal{L}(s_D + s_R)}\big/{\mathcal{L}(n_V + n_R)} \tag{6}$$

Since $d_{X,R}(k)$ may be a highly processed version of the speech and noise mixture, it cannot be assumed as reproducible, i.e., the ICC system must not be regarded as a linear, time-invariant system. Thus, the components $s_{X,R}(k)$ and $n_{X,R}(k)$ cannot be determined by measuring separately speech-only and noise-only.

For this purpose, Annex E provides an estimation method in the time-frequency domain to obtain separated average speech and noise spectra. These can be used to calculate ASL, noise-only level and/or SNR. If not specified otherwise, A-weighting according to [IEC 61672-1] shall be applied on the noise spectra.

9 Measurement parameters and requirements for ICC

9.1 Stability

9.1.1 Requirements

No sustained artifacts, instability, or howling, as defined in clause 9.1.2 below, shall be experienced when ICC is active. For each of the analysed octave band, the maximum instability duration determined during signal pauses shall be less than 53 ms (five analysis frames).

9.1.2 Test methodology

If the ICC system has user-adjustable settings it is required that testing be performed at the highest-gain setting per clause 7.5. Due to the serious impact of instability, including possible safety concerns, it is recommended that testing be performed for additional settings if available.

NOTE – It is recommended that subjective human testing as described in Appendix II be performed as much as practical, especially focused into and out of the driver zone.

1) the test set-up according to clause 7.3 is used;

2) the speech material according to clause 8.1.6 shall be used for testing;

3) for each noise scenario, speech and noise shall be played back synchronously. Three binaural measurements per noise condition shall be conducted:

   a) the first measurement run is conducted with deactivated ICC (see clause 7.4). Speech playback is disabled, resulting in a noise-only recording, which includes the vehicle interior noise;

   b) the second measurement run is conducted with activated ICC (see clause 7.4). Speech playback is disabled, resulting in a noise-only recording, which includes vehicle interior and reinforcement noise;

   c) the third measurement run is conducted with activated ICC (see clause 7.4). The ASL of the source sequence is not adjusted according to clause 8.1.2 (no Lombard level is applied);

4) the recording of step 3c) is compensated according to clause 8.3 with the recording of step 3a). The remaining signal contains only the reinforcement speech and noise at the listener's ears;

5) the noise-compensated recording of step 4) and the noise-only recording of step 3b) are analysed and evaluated as described in Annex D;

6) the worst performing channel (left/right) is taken into account for the overall evaluation of the performance requirement;

7) the test is repeated for each combination of noise scenario and investigated zone.

9.2 Delay

9.2.1 Introduction

Delay in amplified systems is very annoying and humans are very perceptive to this effect. A talker hearing their own voice excessively delayed perceives this as an echo and will quickly stop talking, while a listener hearing the talker's direct path speech and later the same speech through the amplified system will notice the difference in arrival time and be confused or annoyed.

9.2.2 Reinforcement delay

9.2.2.1 Requirements

The delay of the direct sound is defined as the first wave front of the talker's speech signal captured by the artificial ears of the listening HATS.

The delay of the reinforcement sound is defined as the first wave front of the speech signal, which is processed by the ICC system, transmitted via the loudspeakers of the car cabin and then captured by the artificial ears of the listening HATS.

The difference in delay between direct and reinforced voice received by the listener (ΔRx) shall be less than 15 ms, with a preference for lower values. The purpose of this delay requirement is to ensure overall system quality hence the value of the delay shall be minimized in implementations, as far as possible.

This requirement is equally applicable for all combinations of zone to zone communications as shown in Table 7-1.

NOTE – The masking of the direct sound at the listener position caused by the background noise (compensated during measurement) and consequences on the performance requirement is for further study. A suitable criterion in case the direct path is not audible is for further study.

9.2.2.2 Test methodology

As described in clause 7.5, this test should be performed with the ICC set for maximal gain, if available.

Testing shall be carried out in the second quietest and second loudest driving condition (see clause 8.2.2 and/or Annex A).

The measurement is conducted for both ears. The larger value for ΔRx of L/R is considered for the evaluation of the performance requirement.

1) the ICC System in the vehicle and the noise playback are disabled;
   a) an MLS signal (see clause 8.4.1) is played back from the HATS in the talker zone and is recorded by the HATS in the listener zone. The recorded signal is used to determine the impulse response (cf. clause 8.4.1);
   b) the delay k_D between talker and listener is calculated from the impulse response with the method described in Annex B;
2) the noise playback is enabled for the next measurement;
   a) a short speech signal according to clause 8.1.5 is played back from the HATS in the talker zone and is recorded from the HATS in the listener zone;
3) the ICC system and the noise playback are enabled for the next measurement;
   a) a short speech signal according to clause 8.1.5 is played back from the HATS in the talker zone and is recorded from the HATS in the listener zone;
4) the recording of step 3a) is compensated according to clause 8.3 with the recording of step 2a). The remaining signal contains only the reinforcement signal at the listener's ears;
5) the delay k_R of the reinforcement path is calculated via normalized cross-correlation function according to Annex C;
6) the difference between the delays of step 1.2 and step 5 represents the reinforcement delay: $\Delta Rx = k_R - k_D$.

9.2.3 Self-hearing delay

9.2.3.1 Requirements

For further study.

9.2.3.2 Test

For further study.

9.3 Speech Intelligibility

9.3.1 Introduction

The main objective of ICC is to hold a conversation between driver and passengers without turning the head and being distracted from operating the vehicle. Speech Intelligibility is crucial to maintain a clear conversation where none of the participants feels the need to change their normal speaking level.

9.3.2 Requirements

For further study.

9.3.3 Test methodology

For further study.

9.4 Listening effort

9.4.1 Requirements

An acceptable minimum listening effort shall be maintained by the ICC system. On the other hand, the system shall provide a certain degree of improvement regarding listening effort compared to the deactivated system.

For all considered noise scenarios, the following requirements should be met:

1) with ICC activated, a minimum listening effort of 1.75 mean listening effort opinion score (MOS_{LE}) should be achieved;
2) for MOS_{LE} determined with deactivated ICC between 1.0 and 3.0, an improvement between 0.5 MOS_{LE} and 0.25 MOS_{LE} should be obtained (linearly interpolated between these two points);
3) for MOS_{LE} determined with deactivated ICC between 3.0 and 4.0, an improvement between 0.25 MOS_{LE} and 0.0 MOS_{LE} should be obtained (linearly interpolated between these two points).

For MOS_{LE} determined with deactivated ICC larger than 4.0, no improvement has to be achieved.

Figure 9-1 illustrates the requirements as a tolerance graph.

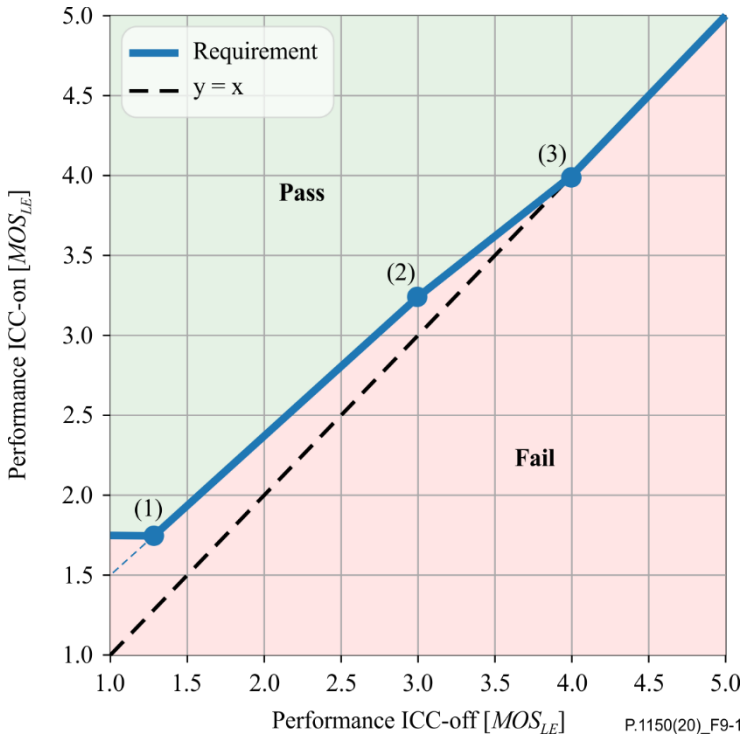


Figure 9-1 – Tolerance graph for listening effort

9.4.2 Test methodology

1) the test set-up according to clause 7.3 is used;
2) the speech material according to clause 8.1.1 shall be used for testing;
3) for each noise scenario according to clause 8.2.2, speech and noise shall be played back synchronously. Two measurements per noise condition shall be conducted:
   a) the first measurement run is conducted with deactivated ICC (see clause 7.4). The ASL of the source sequence is adjusted according to clause 8.1.3;

b) the second measurement run is conducted with activated ICC (see clause 7.4). The ASL of the source sequence is adjusted according to clause 8.1.2. The convergence sequence according to clause 8.1.4 is applied prior to the test sequence;

4) each noisy speech sample of the two binaural recordings are individually analysed. For this, a possible objective method under investigation is the listening effort prediction model according to clause 6 of [ETSI 103 558]. Running a subjective testing is an alternative to an objective method, a possible subjective method is described in clause 5 of [ETSI 103 558]. The MOS_{LE} value per condition ("ICC on", "ICC off") is determined as the average across the eight sentences;

5) the requirement is evaluated for each combination of noise scenario and investigated zone.

9.5 Talker localization accuracy

9.5.1 Requirements

Talker localization accuracy should be maintained in moderate noise conditions when ICC is activated.

Further details are for further study.

9.5.2 Test methodology

For further study.

9.6 Adaptive reinforcement

While in a low noise environment the ICC should reinforce the talker's voice such that the listener can hear the talker as in normal conversation. If there is an increase in noise at the listener position, such as opening a window or increase in vehicle noise, ICC should reinforce the talker's voice such that the listener continues to have an acceptable experience.

9.6.1 Requirements

The ICC system is expected to change reinforcement during changing noise regardless of the presence or absence of talker speech being present. The new reinforcement level should be achieved in a timely manner, i.e., in a range of 0.5 s to 3 s of attaining a new steady-state noise level and should be consistent for the noise level irrespective of the nature of the noise change – rising or falling – and as long as the rate of change is realistic for the vehicle being tested.

9.6.2 Test methodology

For further study.

9.7 Signal-to-noise ratio

9.7.1 Introduction

From telephony it is known that there is a range of speech levels preferred by users. Neither too high nor too low speech levels are appreciated by users. Experiments published in [b-Müsch] show that the preferred speech levels for ICC systems depend on the level of background noise, as shown in Figure 9-2.

Hence a preferred SNR range can be defined depending on the absolute background noise level.

NOTE – The minimum speech level ASL_{min} = 55 dBSPL and the maximum speech level ASL_{max} = 85 dBSPL were arbitrarily chosen.

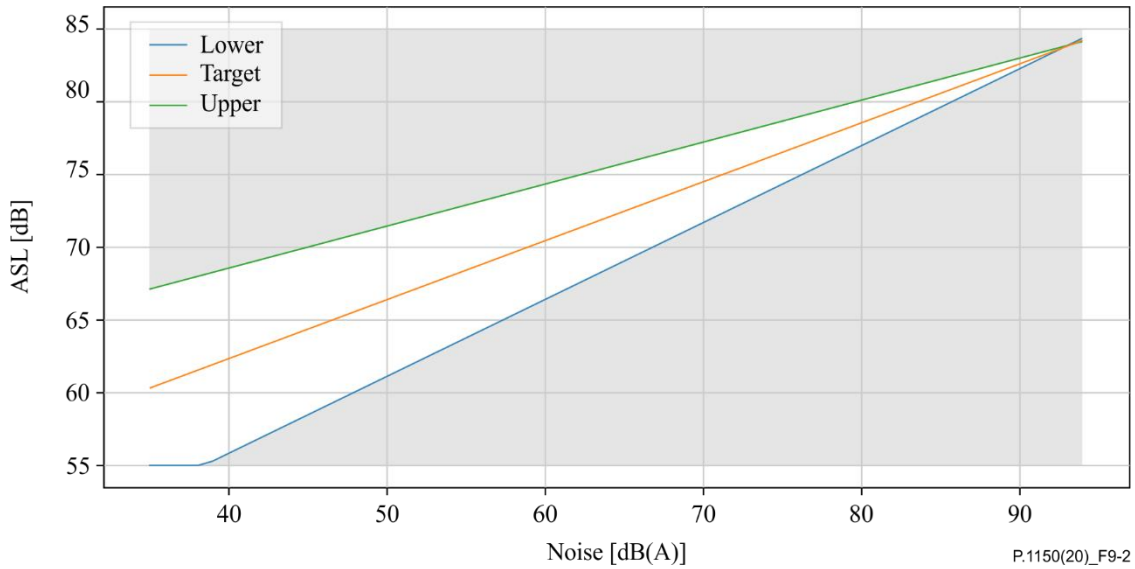


Figure 9-2 – Minimum/preferred/maximum speech level, as a function of background noise level, from experimental data

Following [b-Müsch], the boundary SNR (in dB) as a function of background noise N can be described as:

$$SNR_{lower/upper}(N) = SNR_0 + c \cdot N \tag{7}$$

SNR_0 and c are constants derived from listening tests in reference [b-Müsch] and highlighted in Table 9-1. This simplification is applicable in the Speech range 55 dBSPL < ASL < 85 dBSPL.

Table 9-1 – Lower and upper limit bounds for the constants SNR0 and c

	$\boldsymbol{SNR_0}$**[dB]**	$\boldsymbol{c}$ **[dB / dB(A)]**
Lower	34.71	−0.47
Upper	57.03	−0.71

The corresponding boundary ASL (in dBSPL) can be defined in a similar way as a function of background noise level N (in dBSPL):

$$ASL_{lower/upper}(N) = N + SNR_{lower/upper}(N) = SNR_0 + (1 + c) \cdot N \tag{8}$$

The clipping with minimum/maximum speech level is applied in the next step:

$$\widetilde{ASL}_{lower/upper}(N) = \max(ASL_{min}\,; \min(SNR_0 + (1 + c) \cdot N; ASL_{max})) \tag{9}$$

The resulting modified upper/lower SNR boundary then equals:

$$\widetilde{SNR}_{lower/upper}(N) = \widetilde{ASL}_{lower/upper}(N) - N \tag{10}$$

$$\widetilde{SNR}_{lower/upper}(N) = \max(ASL_{min}\,, \min(SNR_0 + (1 + c) \cdot N, ASL_{max})) - N \tag{11}$$

9.7.2 Requirements

The measured SNR shall be between the noise-dependent thresholds SNR_{lower} and SNR_{upper}, as described in clause 9.7.1. This ensures sufficient, but not too high speech levels in the range of 55 dBSPL to 85 dBSPL. The lower SNR bound is only evaluated up to 78 dBSPL(A) noise level at the listening position.

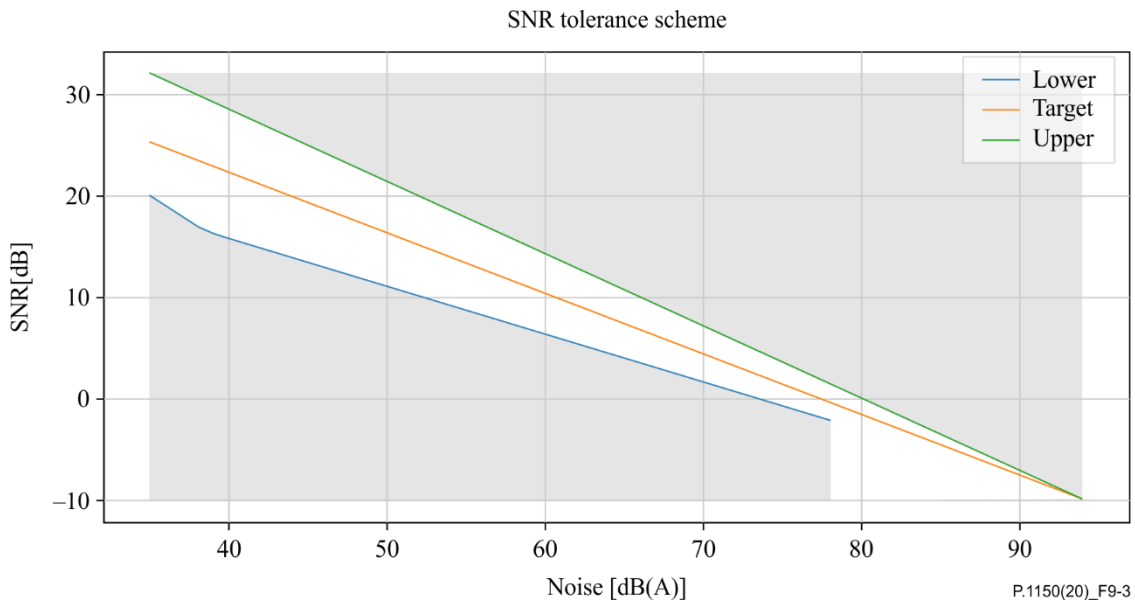


Figure 9-3 – Visualization of the upper and lower SNR level as a function of noise level

9.7.3 Test methodology

1) the ICC System in the vehicle is disabled and the noise playback is enabled:
   a) noise is recorded binaurally by the HATS in the listener zone (*noise-only*);
2) the ICC System in the vehicle as well as the noise playback are enabled:
   a) speech according to clause 8.1.1 is played back from the HATS in the talker zone and is recorded by the HATS in the listener zone (*speech+reinforcement*);
3) the A-weighted noise level of the recording of step 1a) (*noise-only*) is calculated. This value (in dB) is used to obtain the values for SNR_{lower} and SNR_{upper};
4) the SNR is calculated according to clause 8.5 and Annex D. The analysis as well as the performance requirement evaluation is conducted for both ear signals.

9.8 Speech quality

The speech quality between talker and receiver position should not be deteriorated below a threshold by the ICC System, neither in vehicle silence/non-moving state nor in vehicle driving/idle noise condition.

NOTE – Speech quality evaluation is partly covered by listening effort evaluation in clause 9.4.

9.8.1 Requirements

For further study.

9.8.2 Test methodology

For further study.

9.9 Double talk attenuation

9.9.1 Requirement

When Listener decides to reply or talk back during a conversation, the system must not mute or attenuate the talker's voice significantly (e.g., to simplify the work for noise/echo canceller or for ease stability control).

The speech level compared to the single-talk case shall not change by more than +/-3 dB when speech is concurrently applied.

9.9.2 Test methodology

The speech sequence according to clause 8.1.6. is used for double talk testing.

1) the ICC system in the vehicle and noise playback are disabled:
   a) speech is played back from the HATS in the listener zone and is recorded by the HATS in the listener zone (*self-speech*);
2) the ICC system in the vehicle is disabled and the noise playback is enabled:
   a) noise is recorded binaurally by the HATS in the listener zone (*noise-only*);
3) the ICC system in the vehicle and noise playback are enabled:
   a) speech is played back from the HATS in the talker zone and is recorded by the HATS in the listener zone (*single talk*);
   b) speech from both HATS in the talker and in the listener zones is played back and recorded by the HATS in the listener zone (*double talk*);
4) the recording of step 3a) is compensated according to clause 8.3 with the recording of step 2a) (*single talk w/o ambient noise*);
5) the recording of step 4 is compensated according to clause 8.3 with the recording of step 1a) and step 2a) (*double talk w/o ambient noise, w/o self-speech*);
6) level vs. time according to [IEC 61672-1] with time constant of 35 ms is calculated for the signals of step 4 (DT) and 5 (ST). Both curves are reported in dB;
7) the performance requirement is carried out on the difference DT-ST (in dB).

9.10 Media Interaction

The ICC module will operate in a system, and it is imperative that the ICC module does not impair any other functionalities. When the vehicle is powered on, ICC is expected to be initiated after the media system, and without negative impact on the passengers in the vehicle. Interaction between the ICC module and other modules is envisioned without negative effects; e.g., if the driver initiates or receives a wireless phone call, ICC functionalities to/from the driver are recommended to be discontinued. Three different behaviours are expected from the ICC system based on the following functionalities:

1) ICC talk-through: ICC is ON but does not boost the speech signal from the talker side in the presence of media. It is anticipated that this could be for alert-type media;
2) ICC talk-over: ICC is ON and boosts speech from the talker side to compensate for the media level at the listener. It is anticipated that this could be for music type media;
3) ICC disabled: ICC is deactivated by the system when the media is presented. It is anticipated this could be for voice-media such as speech recognition or hands-free calling.

9.10.1 Requirements

The ICC system shall be transparent to any media in operation.

ICC should not adversely affect other audio or hands-free systems utilizing the motor vehicle speakers.

9.10.1.1 Media transparency in the presence of ICC

The ICC system shall be transparent to any media in operation. This requirement shall be tested for all available ICC modes.

The ICC system shall not increase the media level at the listener position by more than [*value is to be defined*] dB, or significantly shape the characteristics of the media at the listener position [*score is to be defined*].

9.10.1.2 ICC reinforcement in the presence of media

The ICC reinforcement level in the presence of media shall be consistent with the expectation of the media functional category being presented. This requirement shall be tested for both talk-through and talk-over ICC media if implemented in the system.

1) for talk-through media, ICC shall not increase the perceived talker level by more than [*value is to be defined*] dB when media is present, regardless of the media level;
2) for talk-over media, ICC shall increase the perceived talker level when media is present, consistent with the media level and within the constraints defined;
3) For ICC-disabled media, ICC shall not increase the perceived talker level by more than [*value is to be defined*] dB when media is present, above that of ICC off.

9.10.2 Test methodology

9.10.2.1 Media transparency in the presence of ICC

To be defined.

9.10.2.2 ICC reinforcement in the presence of media

To be defined.

9.11 Background noise transmission

9.11.1 Requirements

The background noise at the listener's position is compared with enabled and disabled ICC system. The following requirements shall be met:

– the increase in A-weighted noise level shall be less than 2 dB;
– the increase in A-weighted level vs. time shall be less than 3 dB;
– changes in the average spectrum shall be less than 2 dB.

9.11.2 Test methodology

Beside the convergence sequence (clause 8.1.4), no speech signal is used for this measurement and analysis. The test shall be conducted with the ICC setting providing highest gain (if applicable).

1) the ICC system in the vehicle is disabled and the noise playback is enabled:
   a) noise is recorded binaurally by the HATS in the listener zone;
2) the ICC system in the vehicle and the noise playback is enabled. The convergence sequence as described in clause 8.1.4 shall be used to activate the ICC system:
   a) noise is recorded binaurally by the HATS in the listener zone;
3) the time range between 1.0 s and 5.0 s of the recordings obtained in step 1a) and step 2a) is analysed further;
4) based on the signals extracted in step 3, the following analyses are calculated:
   a) A-weighted noise level;
   b) A-weighted level vs. time according to [IEC 61672-1] with a time constant of 125 ms;
   c) average spectrum in 1/3rd octave bands according to [IEC 61260-1] (row B). The average spectrum shall be calculated by utilizing the fast Fourier transform (FFT) method (transformation size 4096, 50 percent overlap, Hann window). For the performance evaluation, frequency bands between 20 Hz and 20 kHz are considered.

Annex A

Standard set of user scenarios

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

Table A.1 – Standard set of user scenarios used to collect noise recordings

		Vehicle settings						Environmental conditions			
User scenario	**Description**	**Vehicle speed**	**Heating ventilation and air conditioning (HVAC) settings**	**Windows**	**Wipers**	**Turn signal**	**Back-ground talkers**	**Road surface (see 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	> –20C and < 40C
2	City driving	60 km/h (37 mph)	FAN = Medium setting (see Note 5); AIRFLOW = Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C
3	Highway	120 km/h (75 mph)	FAN = Lowest setting	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C
4	Highway driving	120 km/h (75 mph)	FAN = Medium setting (see Note 5); AIRFLOW = Directed to windows	Up	Off	Off	None	Dry; rough road	< 5 m/s (12 mph)	None	> –20C and < 40C

Table A.1 – Standard set of user scenarios used to collect noise recordings

		Vehicle settings						Environmental conditions			
User scenario	**Description**	**Vehicle speed**	**Heating ventilation and air conditioning (HVAC) settings**	**Windows**	**Wipers**	**Turn signal**	**Back-ground talkers**	**Road surface (see Note 3)**	**Wind speed**	**Precipitation**	**Temp.**
5 see 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	> –20C and < 40C

NOTE 1 – Optional test: If the ICC 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 not any HVAC vent positions that result in severely degraded performance due to wind buffeting.

NOTE 3 – Smooth road surfaces that generate very little tyre noise shall not be used. Also, road surfaces with bumps that cause significant impulse noises shall not be used either. If available, concrete surfaces are preferred because they often result in worst-case conditions that cause impairments not seen on other road surfaces.

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 B

Delay calculation via impulse response measurement

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

One of the issues of measuring the delay of the direct path in vehicle cabins is that the influences of room acoustics may cause measurement errors: the reflections of the talker's voice may have more energy than the direct path sound. In consequence, the commonly used cross-correlation analysis for delay calculation (abscissa argument of maximum absolute peak) may not be accurate.

Thus, the determination of direct path delay has to be more robust against reflections. Since the direct path only depends on talker/listener positions and the acoustics of the vehicle cabin, it is the impulse response between mouth simulator of the talker and left/right ear of the listener is measured.

The determined impulse response $h(k)$ of duration L is then transformed to an energy decay curve $S(k)$ by the Schroeder backward integration:

$$S(k) = \sum_{j=k}^{L} 10 \cdot \log_{10} h^2(j) \tag{B.1}$$

In order to detect the first appearance of direct sound k_D (=delay of direct path), a threshold T = 0.01 dB on the logarithmic decay curve is applied:

$$k_D = \underset{k}{\operatorname{argmin}} |S(k) - T| \tag{B.2}$$

NOTE – The impulse response approach cannot be used for measuring the delay of the reinforced sound, since the signal processing through an ICC system cannot be regarded as linear and time-invariant.

Annex C

Delay calculation via cross-correlation

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

The short-time cross-correlation functions $\Phi_{xy}(i, \tau)$ between the source signal $S_x(k)$ (played back from the HATS at the talker zone) and the output signal $S_y(k)$ (recorded at an artificial ear in the listener zone) is calculated in the time domain according to equation (C.1).

$$\Phi_{xy}(i,\tau) = \frac{1}{T}\int_{t=\frac{-T}{2}}^{t=\frac{+T}{2}} S_x(\mathrm{t_S}(i)+t)\cdot S_y(\mathrm{t_S}(i)+t+\tau)\cdot dt \qquad \text{(C.1)}$$

The analysis window T shall correspond to 4096 samples, which allows a faster implementation of the cross-correlation calculation. The start time $\mathrm{t_S}(i)$ is provided in equation (C.2) for each frame and corresponds to 90% overlapping analysis windows ($ovl = 90\%$), i.e., a step size of 410 samples.

$$\mathrm{t_S}(i) = i\cdot(1-ovl)\cdot N \qquad \text{(C.2)}$$

Then, the envelope $E(i,\tau)$ according to equation (C.4) of the cross-correlation function $\Phi_{xy}(i,\tau)$ is calculated by the Hilbert transformation $H\{\Phi_{xy}(i,\tau)\}$ of the cross-correlation, as shown in equation (C.3).

$$\mathrm{H}\{\Phi_{xy}(i,\tau)\} = \frac{1}{\pi}\sum_{u=-\infty}^{+\infty}\frac{\Phi_{xy}(i,u)}{\tau-u} \qquad \text{(C.3)}$$

$$E(i,\tau) = \sqrt{\left[\Phi_{xy}(i,\tau)\right]^2 + \left[\mathrm{H}\{\Phi_{xy}(i,\tau)\}\right]^2} \qquad \text{(C.4)}$$

The delay $k_C(i)$ is calculated for each frame as the maximum of the envelope $E(i,\tau)$ according to equation (C.5). The maximum $E_{max}(i)$ for each frame provides a possible proof of reliability and shall be stored as well, as per equation (C.6).

$$k_C(i) = \underset{\tau}{\operatorname{argmax}}|E(i,\tau)| \qquad \text{(C.5)}$$

$$E_{\max}(i) = \max_{\tau}|E(i,\tau)| \qquad \text{(C.6)}$$

A threshold of P_{min} is determined according to equation (C.7) with possible reliable frames.

$$P_{\min} = \max(10\%, \max\left(E_{\max}(i)\right) - 20\%) \qquad \text{(C.7)}$$

Only frames indices i' providing $E_{\max}(i) > P_{\min}$are regarded in the following. Finally, the delay k_C is determined as the 90th-percentile of the remaining reliable frames, as shown in equation (C.8).

$$k_C = \mathrm{percentile}_{90}\ (k_C(i')) \qquad \text{(C.8)}$$

The delay calculation method works for cases when the measured delay is less than T/2; it is the responsibility of the user to ensure a proper application of the method. For a sampling rate of 48 kHz, T/2 corresponds to 42.7 ms, i.e., the delay of the reinforcement sound (compared to the source) has to be lower than this value in order to produce a meaningful result.

Annex D

Calculations for stability assessment

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

D.1 Introduction

The following clauses provide a calculation method to analyse possible instability artefacts versus frequency. Signal components in the reinforcement signal above the vehicle interior and reinforcement noise may in general be perceivable, in case they exceed a certain level threshold and duration.

The method expects three input signals per analysis:

1) the reference signal r(k) (e.g., as defined in clause 8.1.6), scaled to $ASL_{Ref} = -26.0$ dBov;
2) noise-only recording n(k), containing vehicle interior and reinforcement noise;
3) degraded, but noise-compensated speech recording p(k), containing direct path reinforcement speech plus noise.

The recordings n(k) and p(k) must be time-aligned with the reference signal r(k). This can be done based on the measured direct delay (talker to listener), as described in clause 9.2.2.2.

D.2 Spectrum vs time

First, octave-band spectra versus time according to [IEC 61260-1] (row B) are calculated for the p(k), n(k) and r(k). The spectrogram shall be calculated via FFT, with a transformation size of 2048 samples, and 75 percent overlapping Hann windows (step / frame size of 10.6 ms). This results in the time-frequency representations $P(i,j)$, $N(i,j)$ and $R(i,j)$.

D.3 Activity classification

As a second analysis, the reference signal r(k) is scaled to an ASL according to [ITU-T P.56] of $ASL_{Ref} = -26.0$ dBov.

In order to detect active speech parts in the reference signal (and thus, in the degraded signal as well), a level versus time of the reference is calculated according to equations (D.1) and (D.2).

$$\mathrm{L'_{Ref}(i)} = \sum_j R(i,j)^2 \tag{D.1}$$

$$\mathrm{L_{Ref}(i)\ [dB]} = 10 \cdot \log_{10}(\max(-90, \mathrm{L'_{Ref}(i)})) \tag{D.2}$$

The level thresholds according to Appendix II of [ITU-T G.160] are applied on $\mathrm{L_{Ref}(i)}$ and the per-frame classifications according to Table D.1.

Table D.1 – Activity classification

Class No.	Description	Lower Level Threshold	Upper Level Threshold
0	Silence	n/a	ASL_{Ref} – 35 dB
1	Short pauses	n/a	ASL_{Ref} – 35 dB
4	Uncertain	ASL_{Ref} – 35 dB	ASL_{Ref} – 16 dB
5	Low activity	ASL_{Ref} – 16 dB	ASL_{Ref} – 10 dB
6	Medium activity	ASL_{Ref} – 10 dB	ASL_{Ref} – 1 dB
7	High activity	ASL_{Ref} – 1 dB	n/a
NOTE – Short pauses (Class 1) are defined as Silence periods (Class 0), which provide a duration less than 400 ms.			

The resulting activity versus time $a(i)$ is then used to determine signal parts, where no speech is active, i.e., where only noise is present. These frames are stored in the set C_N as specified in equation (D.3). Only frames containing silence (0) are taken into account.

$$C_N = a(i) \in [0] \tag{D.3}$$

In a similar way, a set C_S is defined according to equation (D.4). Here frames providing at least uncertain activity (4) as well as short pauses (1) are included.

$$C_S = a(i) \in [1,4,5,6,7] \tag{D.4}$$

D.4 Analysis ranges

For the evaluation of frames above background noise, only transitions from C_S to C_N are taken into account, i.e., endings of speech parts / beginnings of pauses are analysed. Only octave bands with centre frequencies in the range starting at 500 Hz, up to 8 kHz, are taken into account for this and the following analyses.

To account for typical vehicle reverberation and allowable ICC delay (clause 9.2.2), each analysis start is extended by 137 ms (13 analysis frames), encompassing both direct, reverberant, and ICC-reinforced energy.

D.5 Calculation of stability assessment (energy-present above background)

For each frame within a located analysis range, $P(i,j)$ is compared to $N(i,j)$. Frames where $P(i,j)$ exceeds the background $N(i,j)$ by more than 6 dB are counted as *instable*. The instability duration is given as the number of instable frames per analysis range. If no instability frames are counted, the instability duration is zero in this case. All instability durations per octave band shall be reported.

Annex E

Calculation of separated speech and noise levels

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

E.1 Introduction

As already introduced in clause 8.5, the SNR in dB is given according to equation (E.1) as the ratio between speech ($\mathcal{L}(s)$) and noise ($\mathcal{L}(n)$) level. In case of an ICC system, the speech signal consists of a direct path (D) and a reinforced (R) component. The noise may also consist of two components, the vehicle interior (A) and the reinforcement (R) noise.

$$\mathrm{SNR} = 20 \cdot \log_{10} {\mathcal{L}(\mathrm{s})}/{\mathcal{L}(n)} = 20 \cdot \log_{10} {\mathcal{L}(s_D + s_R)}/{\mathcal{L}(n_A + n_R)} \tag{E.1}$$

However, independent which components are present in a noisy signal, for the calculation of SNR only the superposed speech and noise energies are crucial. The levels can be determined according to equation (E.2) in the frequency domain from average spectra S(j) and N(j), respectively.

$$\mathcal{L}(s) = \mathcal{L}(S) = \sqrt{\sum_{j=J_L}^{j=J_H} S(j)^2} \quad , \quad \mathcal{L}(n) = \mathcal{L}(N) = \sqrt{\sum_{j=J_L}^{j=J_H} N(j)^2} \tag{E.2}$$

Here J_L corresponds to the frequency index of the lower integration limit f_L and $J_H$to the index of the upper integration limit f_H. If not specified otherwise, $f_L = 50$ Hz and $f_H = 20\ k$Hz are used.

Since $S(j)$ and $N(j)$ are not directly accessible, estimations $\hat{\mathrm{S}}$(j) and $\widehat{\mathrm{N}}$(j) have to be determined. For this purpose, the reference signal r(k) and the degraded signal d(k) are used and shall be time-aligned regarding possible delay (see Annex D).

E.2 Spectrum vs time

First, 1/3rd octave-band spectra versus time according to [IEC 61260-1] (row B) are calculated for the d(k) and r(k). The spectrogram shall be calculated via FFT, with a transformation size of 8192 samples, and 90% overlapping Hann windows (step / frame size of 17.1 ms). This results in the time-frequency representations $D(i,j)$ and $R(i,j)$.

E.3 Activity classification

As a second analysis, the reference signal r(k) is scaled to an ASL according to [ITU-T P.56] of $\mathrm{ASL_{Ref}} = -26.0$ dBov.

In order to detect active speech parts in the reference signal (and thus, in the degraded signal as well), a level versus time of the reference is calculated according to equations (E.3) and (E.4).

$$\mathrm{L'_{Ref}(i)} = \sum_j R(i,j)^2 \tag{E.3}$$

$$\mathrm{L_{Ref}(i)\ [dB]} = 10 \cdot \log_{10}(\max(-90, \mathrm{L'_{Ref}(i)})) \tag{E.4}$$

The level thresholds according to Appendix II of [ITU-T G.160] are applied on $\mathrm{L_{Ref}(i)}$ and the per-frame classifications according to Table E.1. Figure E.1 visualizes the activity detection method with an example.

Table E.1 – Activity classification

Class No.	Description	Lower level threshold	Upper level threshold
0	Silence	n/a	$\mathrm{ASL_{Ref}}$ – 35 dB
1	Short pauses	n/a	$\mathrm{ASL_{Ref}}$ – 35 dB
4	Uncertain	$\mathrm{ASL_{Ref}}$ – 35 dB	$\mathrm{ASL_{Ref}}$ – 16 dB

Table E.1 – Activity classification

Class No.	Description	Lower level threshold	Upper level threshold
5	Low activity	ASL_{Ref} – 16 dB	ASL_{Ref} – 10 dB
6	Medium activity	ASL_{Ref} – 10 dB	ASL_{Ref} – 1 dB
7	High activity	ASL_{Ref} – 1 dB	n/a
NOTE – Short pauses (Class 1) are defined as Silence periods (Class 0), which provide a duration less than 400 ms.			

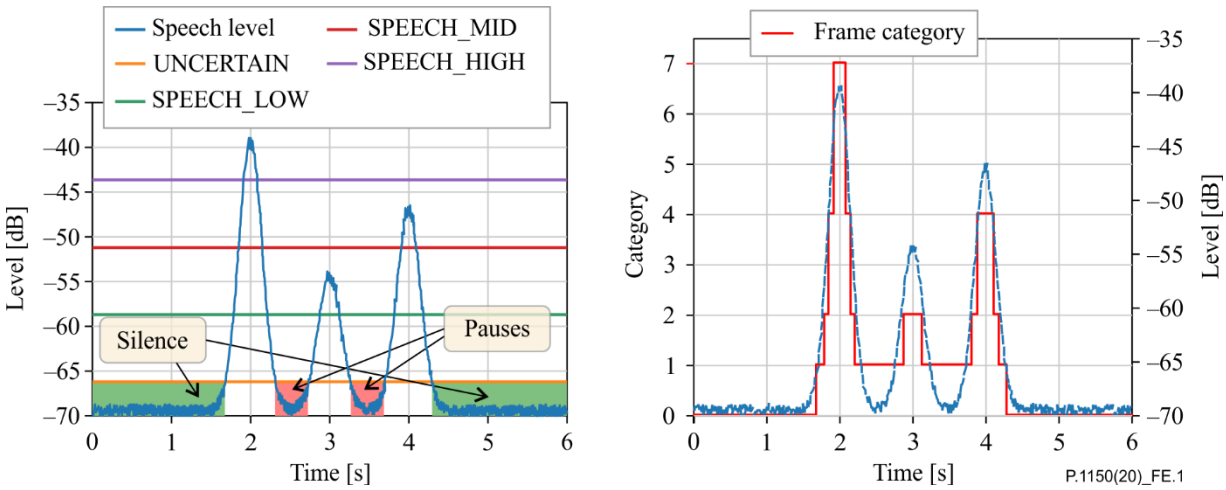


Figure E.1 – Example of activity detection method according to [ITU-T G.160] Appendix II

The resulting activity versus time $a(i)$ is then used to determine signal parts, where no speech is active, i.e., where only noise is present. These frames are stored in the set C_N as specified in equation (E.5). Frames containing silence (0) or short pauses (1) are taken into account.

$$C_N = a(i) \in [0,1] \tag{E.5}$$

In a similar way, a set C_S is defined according to equation (E.6). Here frames providing at least uncertain activity (4) are included.

$$C_S = a(i) \in [4,5,6,7] \tag{E.6}$$

E.4 Estimation of average spectra

Based on these frame classifications, an estimate of the noise-only average spectrum can be determined with the spectrum versus time of the degraded signal $D(i,j)$ according to equation (E.7).

$$\hat{N}(j) = \sqrt{\frac{1}{[C_N]} \sum_{i \in C_N} D(i,j)^2} \tag{E.7}$$

It is assumed that in average the speech signal is uncorrelated to the noise signal. In this case, the approximation as shown in equation (E.8) is valid.

$$D(i,j)^2 = N(i,j)^2 + S(i,j)^2 \approx \hat{N}^2(j) + \hat{S}^2(j) \tag{E.8}$$

Based on this simplification, the estimate for the average speech spectrum can be determined according to equation (E.9).

$$\hat{S}(j) = \sqrt{\frac{1}{[C_S]} \sum_{i \in C_S} \max(D(i,j)^2 - \hat{N}(j)^2, 10^{-90/20})} \tag{E.9}$$

For equation (E.7), $[C_N]$ denotes the number of frames classified as noise/silence periods, in equation (E.9), $[C_S]$ the frames classified as active speech.

NOTE – In contrast to Appendix II of [ITU-T G.160], frames of type "uncertain" are taken into account for the speech level calculation. This was chosen to align estimated speech levels as close as possible to levels calculated according to [ITU-T P.56]. Here lower energy frames are used for the integration versus time as well.

E.5 Speech/noise level and SNR

Finally, equation (E.2) is given with average speech spectrum and noise-only average spectrum as (E.10).

$$\mathcal{L}(s) \approx \mathcal{L}(\hat{S}) = \sqrt{\sum_{j=J_L}^{j=J_H} \hat{S}(j)^2} \quad , \quad \mathcal{L}(n) \approx \mathcal{L}(\hat{N}) = \sqrt{\sum_{j=J_L}^{j=J_H} \hat{N}(j)^2} \tag{E.10}$$

For the noise level, it is common to calculate and report the A-weighted spectrum. The frequency-dependent weighting $W_A(j)$ according to [IEC 61672-1] is applied according to equation (E.11) to the estimated noise spectrum.

$$\mathcal{L}(\hat{N})(\mathrm{A}) = \sqrt{\sum_{j=J_L}^{j=J_H} W_A(j) \cdot \hat{N}(j)^2} \tag{E.11}$$

Appendix I

Details on Lombard speech

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

I.1 With ICC reinforcement disabled

I.1.1 Recommendation

When ICC reinforcement is disabled during testing, and the goal is to simulate Lombard speech changes in the presence of noise, the recommendation is to use the same level and methodology as described in [ITU-T P.1100] and [ITU-T P.1110], with the 3dB increase for non-handset communication [ITU-T P.340] applied in all noise conditions.

I.1.2 Supporting evidence

I.1.2.1 Overview

Seven pairs of subjects were placed in a three-row sports utility vehicle (SUV). Subjects were paired and seated front-to-back. Communication was between zones (2 ←→5) and zones (2←→8), with approximate distances of 96 cm and 170 cm respectively.

Subjects were fitted with directional headset microphones and the task was an informal game of Trivial Pursuit™, where trivia questions were asked and answered turn-wise. The game rules were left to the participants to modify – the goal was to elicit constant back-and-forth conversation with a high level of turn-taking, with a requirement of constant bi-directional information exchange.

Noise playback was calibrated to an approximate mid-head position for both locations using a freeway recording at approximately 110 kph. This produced a SPL of approximately 71 dBA at the front seat and 73 dBA at the rear seating positions.

Talkers were exposed to five noise conditions, as shown in Table I.1. The same headset equipment was used to record the [ITU-T P.501] target speech using a HEAD acoustics ACQUA system with a calibrated B&K HATS in a quiet environment. No noise-related Lombard was applied but the +3 dB for non-handset was included.

Table I.1 – Noise conditions

Nominal noise	Front seat	Rear seat
Quiet	~30 dBA	~30 dBA
−20 dB	51 dBA	53.5 dBA
−10 dB	61 dBA	63.5 dBA
Freeway 110 kph	71 dBA	73.5 dBA
Maximum (+6 dB)	77 dBA	79.5 dBA

I.1.2.2 Data analysis

For paired in-vehicle recordings, the last 60 s of each interval was extracted and each segment was manually trimmed to remove long gaps, laughter, and non-conversational components (humming, "um", "aah"), but thinking-talk ("I'm not sure, but…") was left for analysis. The ASL ([ITU-T P.56]) was calculated for each segment.

For the HATS recordings no additional trimming was done, and [ITU-T P.56] active speech was calculated across all utterances. This quiet, [ITU-T P.501] plus 3 dB ([ITU-T P.340]) level was used as a zero reference – all values are expressed as a difference from this nominal level.

The two rear seat data sets were combined into a single set for analysis purposes.

I.1.2.3 Results

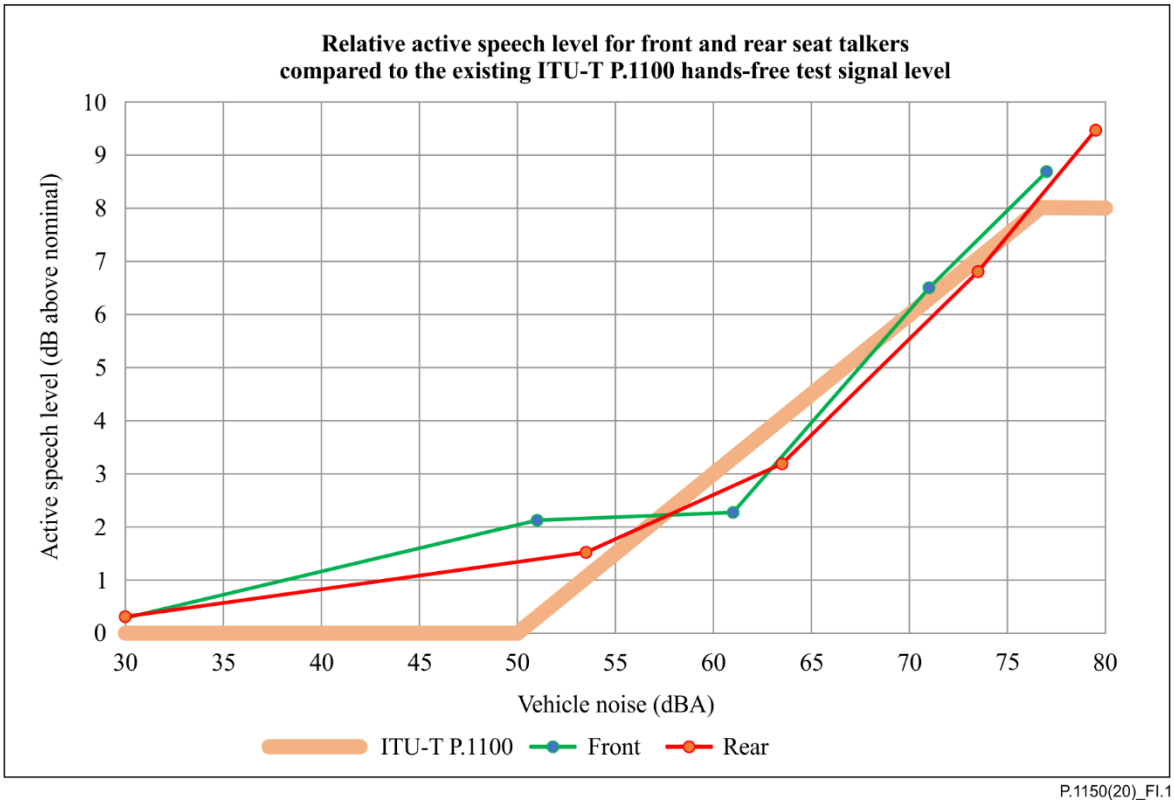


Figure I.1 – Talker speech level, by seating position as a function of the vehicle noise level, in comparison to the ITU-T P.1100 test signal level requirement

I.1.2.4 Conclusion

Data collected supported the assertion that the [ITU-T P.1100] Lombard model was a good approximation of both front-to-back and back-to-front communication when no ICC reinforcement was present.

No data was presented regarding side-to-side communication in any location in the vehicle.

I.2 With ICC reinforcement enabled

I.2.1 Recommendation

When ICC reinforcement is enabled during testing, and the goal is to simulate Lombard speech changes in the presence of noise, the recommendation is to use a level of speech that represents a 50 percent reduction of Lombard effect relative to the increase without ICC reinforcement.

This relationship is shown in the following formula:

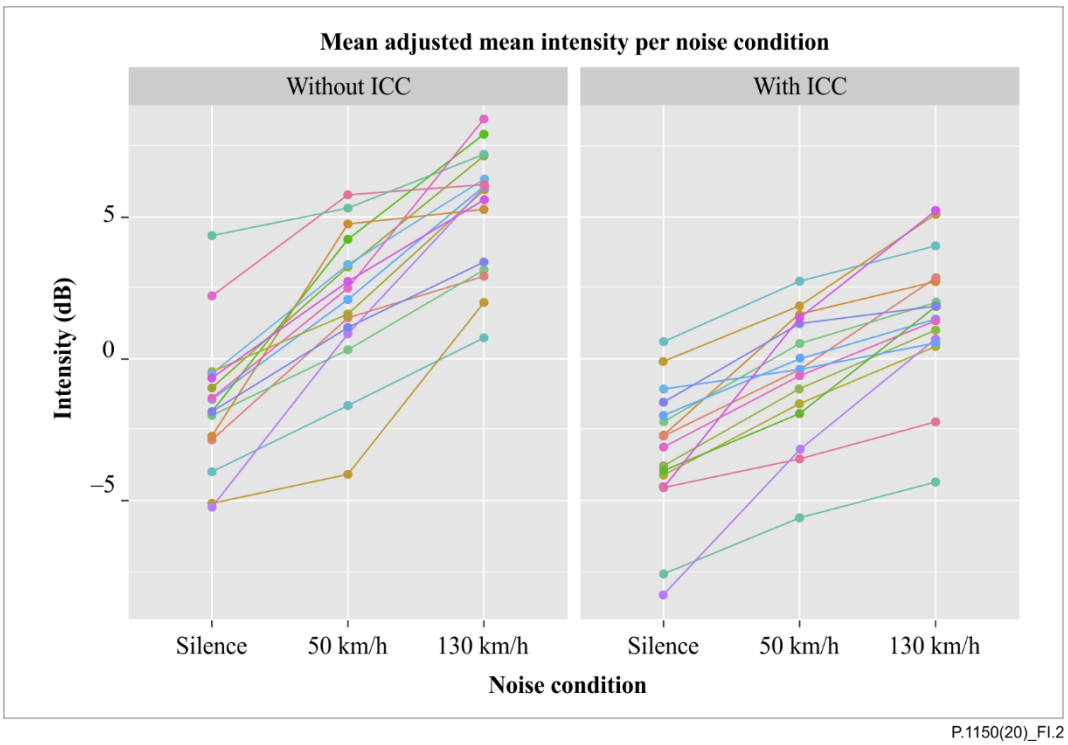
$$I(N) = \begin{cases} 0 \; for \; N < 50 \\ 0.15\,(N - 50) for \; N \geq 50 \end{cases}$$

where:

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

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

The 3 dB increase for non-handset communication [ITU-T P.340] is in addition to the relationship above and is applied in all noise conditions.



NOTE – Different lines represent different subjects.

Figure I.2 – Normalized mean intensity per noise condition with and without the use of ICC

I.2.2 Supporting evidence

I.2.2.1 Existing data

[b-Landgraf] showed that when in vehicle with an acceptably tuned ICC system, there was an overall reduction of talker intensity for a given condition and that the increase in speech intensity between noise conditions was consistent with an estimate of "about one half of the increase without ICC".

NOTE – the actual noise and speech levels were not reported in the document, nor were they available for post-analysis. Conversation with three of the authors did, however, support the conclusion drawn from the figures in the document that "about half the Lombard was observed."

I.2.2.2 New study overview

A B&K HATS was placed in zone 2, representing the simulated talker for all recordings.

Recordings of four speakers saying diagnostic rhyming test (DRT) words and Harvard phonetically rich sentences were made in a quiet room, with subjects wearing open-backed headsets playing a loud vehicle noise to elicit a Lombard effect.

For playback, the average ASL across a subset of the sentences for all four talkers was used as a measurement for scaling the playback level. Specifically, scaling was performed with a fixed level across all talkers and utterances, not on a per-talker or per-utterance basis. The goal was to reach a targeted ASL, but to maintain a level of realistic variability within the data.

A HATS, outfitted with a calibrated binaural recording headset was positioned in either zone 5 (middle) or zone 8 (rear) representing the listener; this represents that data analysed for the experiment described.

In-vehicle noise was simulated at a 110 kph level.

The simulated speech was scaled as described in this appendix for with-ICC-reinforcement and without-ICC-reinforcement recordings.

A simplistic ICC system was implemented in the vehicle and minimally tuned for the test noise condition. This system was not intended to represent an optimal ICC system, rather it was to represent a fairly simplistic system that could reasonably be expected to be achievable in commercial production environments.

Data were recorded as follows:

1) listener position:
   a) middle seat;
   b) rear seat.
2) talker loudness:
   a) nominal/loud: "no ICC level";

   This level was only recorded with ICC disabled.

   b) reduced: "proposed ICC test level".
3) amount of ICC applied (only applied to reduced talker level):
   a) none;
   b) moderate;
   c) high.

I.2.2.3 Data collection – intelligibility

12 listeners were each tasked to identify 60 words in each test condition. Presentation of data was randomized for order and data selection within condition (not all listeners heard the same words within a condition).

I.2.2.4 Data collection – preference/speech quality

Twelve listeners were each presented 20 pairs of phonetically rich sentences for each comparison option test. Listeners were directed to focus on "speech quality", but no direction was given to listen for particular qualities such as "naturalness" or "harshness."

For each pair of utterances, the listener could either choose a preferred utterance or select "no preference."

I.2.2.5 Results on intelligibility

Table I.2 – DRT word recognition accuracy for baseline and three ITU-T P.1150 talker-level conditions

	Listener position		
	Middle	**Rear**	**Average**
Baseline (actual talker level, no ICC)	79.9%	76.8%	78.3%
Moderate ICC (ITU-T P.1150 speech level)	75.8%	73.6%	74.7%
Higher ICC (ITU-T P.1150 speech level)	79.2%	78.6%	78.9%
No ICC (ITU-T P.1150 speech level)	74.9%	71.1%	73.0%
NOTE – For the average data, Baseline versus Higher ICC is not significantly different; all other comparisons were significantly different ($p<0.05$).			

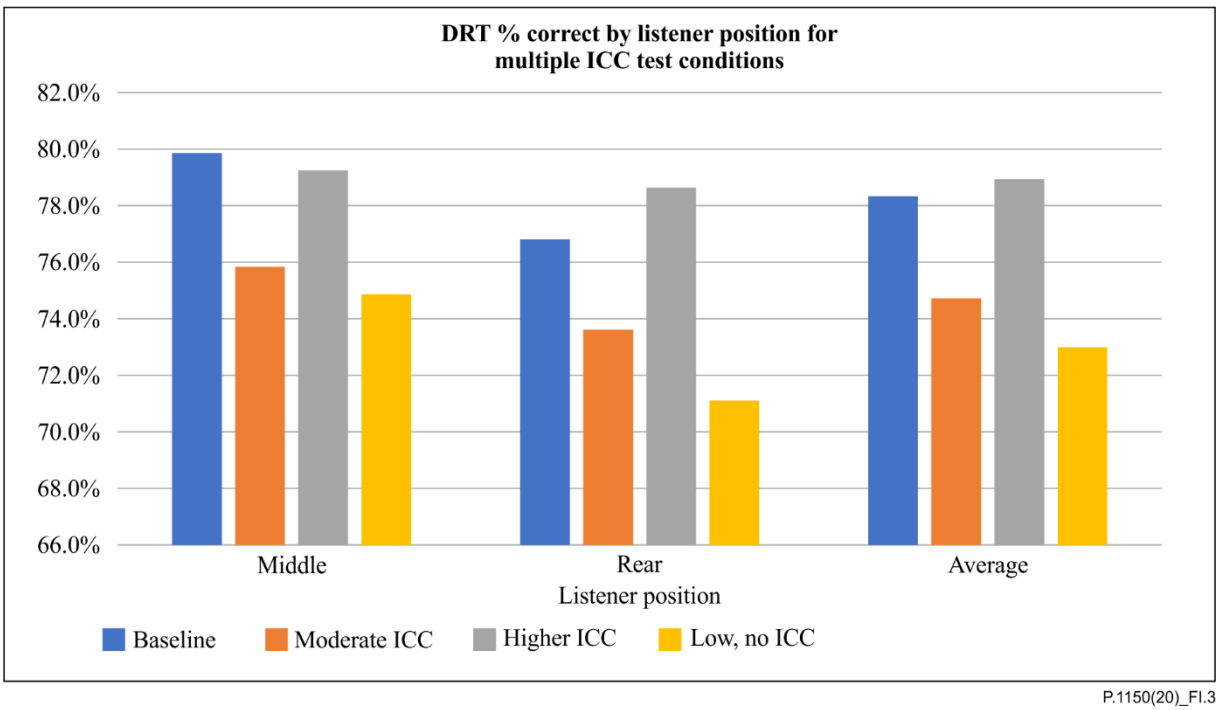


Figure I.3 – DRT word recognition accuracy for Baseline and three ITU-T P.1150 talker-level conditions

I.2.2.6 Results speech quality/preference

Because of the relatively small subset of data that was presented to the listeners and the variability present in the results between listening subjects on this objective task no quantitative results are being reported.

An example of some results is presented below for discussion purposes only.

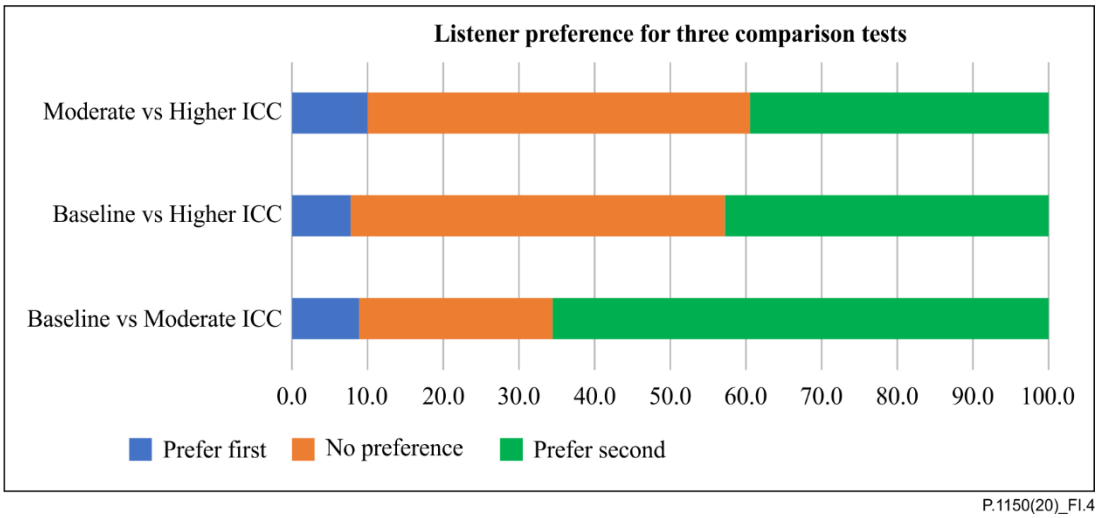


Figure I.4 – Comparison MOS (CMOS) results for rear seat listening data. Comparisons between Baseline (no ICC realistic talker level) and Moderate and Higher ICC reinforcement (with ITU-T P.1150 reduced talker level)

I.2.2.7 Conclusion

Data collected supports the assertion that a minimally implemented ICC system stimulated by a speech level consistent with the reduced Lombard of this Recommendation can result in performance,

as measured by intelligibility and listener preference, that is as good or better than the non-ICC outcome with the higher level of Lombard.

A minimally implemented ICC system can allow talkers to reduce their talking effort to that suggested for testing while still performing adequately.

Appendix II

Guidance for subjective evaluation of system stability

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

Besides objective testing of the ICC system, a subjective performance evaluation is also recommended. The tests described here are targeted mainly to "in situ" ICC evaluation in the target car and under conditions which are not necessarily covered by the objective test specification.

The primary performance evaluation of the ICC system is to ensure stability:

– the system should never produce undamped ringing or howling, regardless of the input;

– audio artifacts that could result in driver-distraction or surprise should not occur during reasonable use, and if they occur in extreme testing should disappear quickly;

– the perceived level of reinforcement should never be excessively loud.

Recognizing that "stability" and "quality" are not the same metric, attention should be paid to artifacts that could be considered startling, distracting, or extreme objectionable. Sustained or extremely loud artifacts should be treated as Failures even if they occur infrequently.

Beyond stability, subjective evaluation of system quality can occur, but is beyond the scope of this description.

It is recommended that testing be performed between all zones as described in clause 7.2.2. If time constraints are present, it is recommended that a different subset of tests be performed between different zones, resulting in all tests being performed between at least one zone combination.

If the ICC system allows multiple settings (clause 7.5) it is recommended that testing be performed in as many settings as possible.

If a user accessible on/off control is available for ICC it should be toggled on and off as part of testing even if it is not accessible from the talker zone.

Minimally, it is recommended that testing be performed:

– from the driver zone to at least one listener zone;

– from at least one non-driver zone into the driver zone (if the ICC implementation supports reinforcement to the driver zone from any other zones);

– with at least the most-reinforcing (highest) setting and one additional setting if multiple settings are available for the ICC system.

Testing should be performed in as many noise conditions as practical, including low, moderate, and very high noise conditions.

The following examples should be considered suggestive rather than prescriptive in that the goal is to expose the ICC system to conditions that are well beyond the typical usage to try and maximize the likelihood of triggering instability. Recommendations include:

– the tester should actively change their talking volume, well beyond what might be considered normal conversation. Loud short sounds should be included with normal speech content, focusing on high-frequency, short duration sounds;

– the talker should move dynamically during testing, including getting very close to the ICC microphone(s) to maximize coupling, shouting, intermixed with low-level taking or whispering;

– movement near the microphone(s) such as waving hands and including physical contact with the microphone(s), should be part of the testing;

– if available, a portable media player such as a phone can be placed near the ICC microphones and dynamic music or other audio can be toggled on an off. This can be done with or without additional speech present. If not available, paper rustling, finger snapping, or other dynamic sources can be employed;
– double talk, where testers sit in both the designated talker and listener zones and simultaneously interact with the system as suggested above. This testing is strongly recommended if the ICC system supports more than one talker zone;
– varying the noise level quickly (via background simulation or by turning fan/defrost on and off) while talking;
– if there is a user-accessible switch to toggle the system on/off or to change operating modes, this should be tested in the presence of speech, and with varying noise if possible.

Bibliography

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

[b-ITU-T P.1120] Recommendation ITU-T P.1120 (2017), *Super-wideband and fullband hands-free communication in motor vehicles*.

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

[b-ITU-T P.1140] Recommendation ITU-T P.1140 (2017), *Speech communication requirements for emergency calls originating from vehicles*.

[b-Cohn] Cohn, Lempel (1977), *On Fast M-Sequence Transform*, IEEE Transactions on Information Theory, Vol. 23, pp 135-137.

[b-Müsch] Müsch, Kettler (2015), *Instrumental Testing of In-Car Communication Systems*, DAGA 2015 Nürnberg, pp 406-409.

[b-Landgraf] Landgraf, Kohler-Kaeß, et al. (2016); *Can you hear me now? Reducing the Lombard effect in a driving car using an In-Car Communication system*, In: Proceedings of 8th International Conference of Speech Prosody, International Speech Communication Association, pp. 479-483.

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, 2020