Final draft ETSI ES 202 396-1 V1.9.1 (2023-03)

**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**

Reference

RES/STQ-308

Keywords

noise, performance, quality, speech

ETSI

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

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

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

If you find a security vulnerability in the present document, please report it through our
Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or
other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

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

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

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This final draft ETSI Standard (ES) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ), and is now submitted for the ETSI standards Membership Approval Procedure.

The present document is part 1 of a multi-part deliverable covering Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise, as identified below:

ETSI ES 202 396-1: "Background noise simulation technique and background noise database";

ETSI EG 202 396-2: "Background noise transmission - Network simulation - Subjective test database and results";

ETSI EG 202 396-3: "Background noise transmission - Objective test methods".

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

Background noise is present in most of the conversations today. Background noise may impact the speech communication performance to terminal and network equipment significantly. Therefore testing and optimization of such equipment is necessary using realistic background noises. Furthermore reproducible conditions for the tests are required which can be guaranteed only under lab type condition.

The present document addresses this issue by describing a methodology for recording and playback of background noises under well-defined and calibratable conditions in a lab-type environment. Furthermore a database with real background noises is included.

1 Scope

The quality of background noise transmission is an important factor, which significantly contributes to the perceived overall quality of speech. Both existing and, even more notably, the new generation of terminals, networks and system configurations, including broadband services, can be greatly improved when designed properly, with consideration and presence of background noise. The present document:

- describes a noise simulation environment using realistic background noise scenarios for laboratory use;

- contains a database including the relevant background noise samples for subjective and objective evaluation.

The present document provides information about the recording techniques needed for background noise recordings and discusses the advantages and drawbacks of existing methods. Additionally, the present document describes the requirements for laboratory conditions. The loudspeaker setup and the loudspeaker calibration and equalization procedure are described. The simulation environment specified can be used for the evaluation and optimization of terminals and of complex configurations including terminals, networks and other configurations. The main application areas should be: office, home and car environment.

The setup and database as described in the present document are applicable for:

- Objective performance evaluation of terminals in different (simulated) background noise environments.

- Speech processing evaluation by using the pre-processed speech signal in the presence of background noise, recorded by a terminal.

- Subjective evaluation of terminals by performing conversational tests, specific double talk tests or talking and listening tests in the presence of background noise.

- Subjective evaluation in third party listening tests by recording the speech samples of terminals in the presence of background noise.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

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

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] Recommendation ITU-T P.57: "Artificial ears".

[2] Recommendation ITU-T P.58: "Head and torso simulator for telephonometry".

[3] ETSI TS 103 224: "Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] Surround Sound Past, Present, and Future: "A history of multichannel audio from mag stripe to Dolby Digital", Joseph Hull - Dolby Laboratories Inc.

[i.2] AES preprint 3332 (1992): "Improved Possibilities of Binaural Recording and Playback Techniques", K. Genuit, H.W. Gierlich; U. Künzli.

[i.3] AES preprint 3732 (1993): "A System for the Reproduction Technique for Playback of Binaural Recordings", N. Xiang, K. Genuit, H.W. Gierlich.

[i.4] NTTAT Database: "Ambient Noise Database CD-ROM".

[i.5] ISO 11904-1: "Acoustics - Determination of sound immission from sound sources placed close to the ear - Part 1: Technique using a microphone in a real ear (MIRE technique)".

[i.6] J. Blauert: "The psychophysics of human sound localization", Spatial Hearing.

[i.7] Void.

[i.8] Void.

[i.9] Recommendation ITU-T P.340: "Transmission characteristics and speech quality parameters of hands-free terminals".

[i.10] Recommendation ITU-T P.64: "Determination of sensitivity/frequency characteristics of local telephone systems".

[i.11] Recommendation ITU-T G.722: "7 kHz audio-coding within 64 kbit/s".

[i.12] Genuit, K.: "A Description of the Human Outer Ear Transfer Function by Elements of Communication Theory (No. B6-8)".

NOTE: Proceedings of the 12th International Congress on Acoustics. Toronto published on behalf of the Technical Program Committee by the Executive Committee of the 12th International Congress on Acoustics.

[i.13] IEC 60050-722: "International Electrotechnical Vocabulary - Chapter 722: Telephony".

[i.14] "Wellenfeldsynthese - Eine neue Dimension der 3D-Audiowiedergabe"; Fernseh- und Kino-Technik, Nr. 11/2002, pp. 735-738.

[i.15] N.Lee: "IOSONO" Computers in Entertainment, volume 2, issue 3 (2004).

[i.16] P. Scherer: "Ein neues Verfahren der raumbezogenen Stereophonie mit verbesserter Übertragung der Rauminformation"; Rundfunktechnische Mitteilungen, 1977, pp. 196-204.

[i.17] Void.

[i.18] ETSI TS 151 010-1: "Digital cellular telecommunications system (Phase 2+) (GSM); Mobile Station (MS) conformance specification; Part 1: Conformance specification (3GPP TS 51.010-1)".

[i.19] Void.

[i.20]   ETSI TR 126 921: "Universal Mobile Telecommunications System (UMTS); LTE; 5G; Investigations on ambient noise reproduction systems for acoustic testing of terminals (3GPP TR 26.921)".

3   Definition of terms, symbols and abbreviations

3.1   Terms

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

cross-talk: appearance of undesired energy in a channel, owing to the presence of a signal in another channel, caused by, for example, induction, conduction or non-linearity

NOTE: See IEC 60050-722 [i.13].

3.2   Symbols

Void.

3.3   Abbreviations

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

CD	Compact Disc
DF	Diffuse Field
EQ	Equalization
FF	Free Field
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
HATS	Head And Torso Simulator
ID	Independent of Direction
IIR	Infinite Impulse Response
MIRE	Microphone In Real Ear
MRP	Mouth Reference Point
NTT	Nippon Telegraph and Telephone corporation
SLR	Send Loudness Rating
VHF	Very High Frequency

4   Overview of existing methods for realistic sound reproduction

4.1   Introduction

In general the existing methods for close to original sound recording and reproduction aimed for different applications:

- Techniques intending to reproduce the actual sound field.

- Techniques providing hearing adequate (ear related) signals in the human ear canal.

- Techniques generating artificial acoustical environments.

Within this clause the different methods are briefly described and their applicability for close to original sound-filed reproduction is discussed. A variety of methods have been studied, in the following a summary of the most important ones relevant to the present document is given. The different methods were analysed on the basis of the following requirements:

- The background noise recording technique should be:

    - easy to use;

    - easy to calibrate;

    - capable of wideband recording;

    - available at reasonable costs;

    - mostly compatible to existing standards and procedures used in telecommunications testing;

    - applicable to different environments (at least office, home and car).

- The background noise simulation arrangement should:

    - be easy to setup;

    - not require any specific acoustical treatment for the simulation requirement;

    - provide a mostly realistic background noise simulation for all typical background noises faced with in telecommunication applications;

    - be easy to calibrate;

    - be mostly insensitive against the positioning of (test)-objects in the simulated sound field;

    - be applicable to all typical terminals used in telecommunication;

    - be available at reasonable costs.

4.2   Surround Sound Techniques

The basics of surround techniques are found in cinema applications. The virtual image provided by stereophonic presentation of sounds seemed not to be sufficient for the large screen display in cinema. In the 1950s 4-channel and 6-channel sound tracks recorded on magnetic stripes associated to the films were developed, 4-channel and 6-channel loudspeaker systems were installed in cinemas to reproduce the multichannel sounds. The newer techniques were mostly developed and marketed by Dolby® [i.1]: Dolby Surround, Dolby Surround Pro Logic, Dolby Digital and Dolby Digital Surround are examples for the techniques introduced more recently. The most common configuration is the "5.1-configuration" used in cinema but in home applications as well. The reproduction system consists of left and right channel, a centre speaker, two surround channels (left and right, arranged in the back of the listener) and a low frequency channel for low frequency effects.

The aim of all surround system is to create an artificial acoustical image in the recording studio rather than recording a real acoustical scenario and providing true to original playback possibilities.

On the recording side, special surround encoders are used allowing the 5-channel signal to be encoded from a special mixing console to the 5.1 digital data stream. The playback system consists of a special decoder allowing to separate the 5 channels again and distribute them on the 5.1 loudspeaker playback system. The systems are mono and stereo compatible and can handle the older 4-channel surround techniques by a specific decoder.

Applications:

Typical applications for surround systems are cinemas and home theatres. The source material is produced by professional recording studios using multi-channel mixing consoles and specific 5.1 decoding techniques. In mostly all cases virtual environments are created which support the visual image by an appropriate acoustical image.

Conclusion:

Surround techniques are designed for creating acoustical images rather than for close to original recording and reproduction. Although the spatial impression provided by surround techniques is sometimes remarkable the acoustical image created is always artificial. Due to the lack of easy to use recording techniques allowing a spatial recording of a sound field surround sound techniques are not suitable for creation of a background noise database with realistic background noises and calibrated background noise simulation in a lab.

4.3   IOSONO®

The IOSONO® sound system (see [i.14], [i.15] and [i.16]) is based on the Wave-Field Synthesis. It employs Huygens principle of wave theory. Applied to acoustics this principle means that it is possible to reproduce any form of wave front with an array of loudspeakers, so that virtual sound sources can be placed anywhere within a listening area. For practical use it is necessary to position loudspeakers all-round the playback room. In order to generate realistic sound fields the input signal for each loudspeaker has to be calculated separately. For this purpose each single sound source (e.g. voices) has to be recorded individually. If the recordings are done in a room, the characteristics (like reverberation) of the recording room also have to be recorded separately. All resulting sound tracks are then mixed and manipulated during the post-editing process and the reproduction.

The natural and realistic spatial sound reproduction is then achieved in a wide area of the play back room. Common 5.1 stereo systems achieve a "realistic" sound reproduction only in a small area of the reproduction room.

Applications:

Typical applications are sound systems for home use, cinemas and other entertainment events. The IOSONO® sound system is also able to play back recordings made in common stereo or 5.1 stereo techniques.

Conclusion:

The drawbacks of this method are the components needed: a sophisticated recording system, a powerful computing unit for real-time mixing the large number of recorded sound tracks and the number of loudspeakers that have to be installed in the listening room. In a common size cinema for example about 200 loudspeakers are needed.

The advantage is that with the IOSONO® sound system a very realistic sound reproduction is possible, but it requires an enormous effort, which is too high for daily use in laboratories.

4.4   Eidophonie

This method was developed for realistic sound reproduction using the VHF transmission technique. The main principle is to separate the base signal from the part of the signal, which contains the information about the direction of sound incidence.

For recording a 1st order gradient microphone with a cardioid directivity is used. During the recordings its directivity rotates with 38 kHz in the recording plane. This "turning microphone" provides an amplitude-modulated signal at its electrical output. The resulting side bands are out of the transmitted frequency range. But these side bands contain the information of the direction of sound incidence. Using the VHF transmission techniques this phase information can be transmitted within the 2nd audio-frequency channel.

The sound reproduction is made by a spatial demodulation: a switch is positioned before each loudspeaker and each switches synchronously with the turning directivity. So a low pass filtered short-term section of the signal containing the information of the direction of sound incidence is played back on each loudspeaker. The loudspeakers are positioned all around the playback room.

Applications:

Eidophonie was developed to provide a realistic sound environment using a signal received from a VHF broadcast station. With this technique the common stereo sound reproduction should be improved. Nevertheless Eidophonie is also compatible to common mono and stereo recordings.

Conclusion:

Benefits of this system are that three loudspeakers are sufficient to produce a realistic sound field. Using more loudspeakers (e.g. 16) the spatial sound reproduction gets more and more independent from the listening position. Moreover the independency of the transmitted sound from the acoustics of the reproduction room increases with the number of loudspeakers used. But there are significant limitations of the method: The microphone directivity is frequency dependent and not ideal. Therefore the interference between the different channels is created. A second problem is the loudspeaker directivity, which does not fit the microphone directivity. This problem could be reduced if the number of channels would be increased. This however is not possible due to the limited directivity of the microphone arrangement used.

Localization of sound sources is hardly possible due to the interference effects of the microphone signals and the loudspeakers. At close to original reproduction depends on the number and distribution of sound sources present. For most of the sound source combinations this goal cannot be achieved.

In general the coding technique needed to record the sound field by a "turning microphone", is complicated and not available commercially. A further drawback of this method is the complicated decoding technique needed on the reproduction side, which is also not commercially available.

4.5   Four-loudspeaker arrangement for playback of binaurally recorded signals

This reproduction procedure was originally investigated to reproduce binaurally signals recorded using artificial head technology. It improves the impressions of direction and distance. Four loudspeakers are typically positioned in a square formation around a central point (listening point) equidistantly, e.g. 2 m. The binaural recordings are reproduced as follows: the two left-hand loudspeakers receive the same free-field equalized artificial head signal of the left-hand channel only. The right-hand side is arranged similarly. For equalization, the transfer function from the two left-hand loudspeakers is measured at the artificial head's left ear channel. With this result, several IIR and FIR filters are designed, with which the input signal of the left-hand loudspeakers during the play back is filtered in such a way that the transfer function then measured at the artificial head's left-hand channel is spectrally flat. The equalization for the right-hand loudspeakers is done similarly.

The equalization procedure does not take into account the correction of cross-talk. This means, the left-hand channel of the artificial head is only equalized for the left-hand loudspeakers, but during the reproduction this left-hand channel will also receive a signal from the right-hand loudspeakers. But despite this simplification, the equalization procedure provides a realistic binaural listening impression.

Investigations [i.2] and [i.3] carried out in different rooms have shown that directional hearing and distance localization by sound reproduction with this four-loudspeaker arrangement are comparable to those with sound reproduction by headphones.

For equalization there are several strategies. The equalization can either be done for each loudspeaker individually or by pairs left - right or by pairs front - rear.

Applications:

A practical application is the sound reproduction for binaural recordings in a typical office-type room, e.g. for listening tests but objective tests as well. The investigations shown in [i.2] and [i.3] indicate that the subjective impression provided by the arrangement corresponds to the headphone reproduction of binaural recordings with respect to the perception of the sound colour, the distance perception and (with some limitations) with respect to the sound source localization. The data provided indicate that the setup could be used for the objective measurements of e.g. telecommunication terminals.

This four-loudspeaker arrangement is also used in advanced driving simulators - which typically provide a visual simulation of the driving situation in addition to the acoustical playback system.

Conclusion:

The advantage of this arrangement on the recording side is the compatibility to standard mono and stereo recordings. Due to this it is easily possible to playback either binaural recordings for subjective and objective experiments or mono/stereo recordings in cases where a less realistic reproduction is sufficient. With slight modifications the geometrical setup can be transferred to other environments like cars for example. Another benefit of this arrangement is the moderate hardware effort.

Concerning the sound reproduction a drawback of this arrangement is that due to the superimposed loudspeaker signals mostly in anechoic rooms interferences appear. This effect obliged test subjects to keep their heads in a mostly fixed position during hearing tests, but also means that an exact sound reproduction is only possible in a small area in anechoic rooms. Another drawback for binaural reproduction is the fact that no exact cross-talk cancellation is possible with this arrangement. However in general this technique seems to be the most promising under the restrictions given (moderate hardware effort on the recording and especially on the reproduction side, close to original reproduction of the scenarios recorded without additional adjustment of the reproduction arrangement).

4.6   NTT Background-Noise Database

The NTT Background Database [i.4], which is commercially available from NTT, is typically used for codec tests. The database contains noise files, which were recorded with a 4-channel recording using 4 directional microphones. The microphones were arranged in an angle of 90 degrees with 70 cm diagonal. Although the original signals were recorded with 20 kHz bandwidth, the signals commercially available are specially coded on a CD providing a bandwidth of 11 kHz/15 bit for each channel. A special decoder is needed if the signals are to be presented acoustically over loudspeakers. The loudspeaker arrangement is suggested in a condition list. For calibration a calibration tone is provided, only level calibration is performed, no equalization procedure is described.

For the electrical evaluation of systems a downmix of the 4-channel recording to a mono channel with 8 kHz sampling rate is available. This signal is mostly used for the evaluation of new speech codecs in order to evaluate the influence of background noise on the coder performance.

Applications:

The NTT-database is mostly used for the evaluation of speech coders using the 8 kHz down sampled signals. When using the acoustical 4-channel playback, the limitation is mostly due to the bandwidth limitation of 11 kHz, which may be not sufficient for future wideband applications.

Conclusion:

The disadvantage of this arrangement is found on the recording and on the playback side. For recording, a special microphone arrangement and a special coding technique is needed. The signal on the reproduction side is band-limited and may not be used in future wideband applications. Currently, it is not clear how the playback system can be calibrated and equalized in order to achieve a close to original sound field and which procedure should be followed during recording in order to get the right calibration and setup for the recording. Furthermore, the chosen microphone arrangement seems to be impractical for recording in smaller enclosures, e.g. in a car.

Another drawback is that the background noise database, including a special decoder, needs to be purchased for each application separately. It is not clear to what extent the recording and coding technique is commercially available.

4.7   General conclusions

Although a variety of reproduction techniques exists, the usability of such methods within the constraints of a laboratory use is limited due to:

- easy and well described recording technique;

- easy to install and easy to use playback technique at reasonable costs;

- applicability of the playback technique in a variety of different rooms with different acoustical conditions;

In consequence, a four-channel loudspeaker setup with associated subwoofer based on binaurally recorded material is selected as the basis for the ETSI background noise simulation arrangement.

5   Recording arrangement

5.1   Binaural recordings

The sound field simulation technique described in the present document is generally based on the binaural recording and reproduction technique as has been known for many years (see [i.6]). The general principle of the recording and playback technique is to provide a recording instrument (artificial head or test subject wearing binaural probe microphones), which allows the recording of the ear signals typically received by the human user in a sound field as close to the original as possible. When using an artificial head (Head And Torso Simulator (HATS)) that represents the "average human", its compliance with Recommendation ITU-T P.58 [2], which describes the head and torso simulator for use in telephonometry, shall be ensured. Since the frequency characteristics of an artificial head strongly dependent on the direction of sound incidence, the output signals of the microphones of an artificial head cannot be directly compared to a standard measurement microphone and an equalization procedure has to be applied in order to get comparable output signals. Various types of equalization procedures for artificial head recordings can be used, as described in detail in clause 5.2.

5.2   Equalization procedure

The artificial head (HATS) used for binaural recordings shall comply with the requirements defined in Recommendation ITU-T P.58 [2]. Although all equalization methods described below can be used, the preferred method is the Diffuse Field (DF) equalization or (if available) the Independent of Direction (ID) equalization. This is because there are the less sharp peaks and dips in the transfer function of the artificial head compared the free-field transfer function (see also Recommendation ITU-T P.58 [2]).

Free-Field Equalization (FF):

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.

The free-field equalization is made for this reference position in an anechoic room. The reference source is placed on the reference axis of the HATS, at a minimum distance of 1,5 m from the HATS lip-ring. When using a binaural microphone this point corresponds to the lip plane of a human user wearing the binaural microphone. The measured free-field response of the HATS is:

$$H_{ff}(0°, 0°, f)$$

From this the free-field equalization is calculated:

$$H_{ff\text{-}EQ}(0°, 0°, f) = 1/H_{ff}(0°, 0°, f)$$

$H_{ff\text{-}EQl}(0°, 0°, f)$ is called free-field equalization of left ear, $H_{ff\text{-}EQr}(0°, 0°, f)$ is called free-field equalization of right ear.

The equalization should be correct within ±0,5 dB within the frequency range from 50 Hz to 10 kHz measured in $1/3^{rd}$ octave bands.

> NOTE: The free-field equalization of a HATS requires a very careful set-up of the measurement. Due to the directivity of the HATS small deviations from the reference position (0°, 0°) may cause incorrectly equalized transfer functions.

Diffuse Field Equalization (DF):

The artificial head is equalized in a diffuse sound field. For random sound incidence the frequency response of the artificial head is flat (see also microphone diffuse field equalization).

The diffuse-field equalization is made for this reference position in a diffuse sound field. Either an echoic room is used or an arrangement of multiple loudspeakers each fed by an uncorrelated noise signal and generating a truly diffusive sound field. The HATS is positioned in the centre of the diffusive sound field.

The measured diffuse-field response of the HATS is:

$$H_{df}(f)$$

From this the diffuse-field equalization is calculated:

$$H_{df-EQ}(f) = 1/H_{df}(f)$$

$H_{df-EQl}(f)$ is called diffuse-field equalization of left ear; $H_{df-Eqr}(f)$ is called diffuse-field equalization of right ear.

The equalization should be correct within $\pm 0,5$ dB within the frequency range from 50 Hz to 10 kHz measured in 1/3rd octave bands.

Independent of Direction Equalization (ID):

This technique is based on the fact that the transmission characteristics of an artificial head is influenced by components independent of direction (cavum concha resonance, ear canal transfer function) and components which contribute depending on the direction of sound incidence to the measured transfer function. The equalization considers only those components of the head related transfer function which are independent of direction.

The independent of direction equalization cannot be measured directly but needs to be calculated by calculating the influence of the non-directional components of the head related transfer functions and generating the equalization function by using the inverse of the ID-transfer function. The influence of the ear canal and the cavum concha is direction independent while the influence of the outer ear, the head, the shoulder and the human body are dependent on the direction of sound incidence. The transfer function of the cavum concha and the ear canal can be modelled by a cavity followed by an acoustical canal which is terminated by the ear drum. The appropriate model and the procedures are described e.g. in [i.12].

The measured free-field response of the HATS can be separated into:

$$H_{ff}(f) = H_{dd}(f) + H_{id}(f)$$

with $H_{dd}(f)$ -   directional dependant part of the head-related transfer function.

          $H_{id}(f)$ -   directional independent part of the head-related transfer function.

From this, the independent of direction equalization is calculated:

$$H_{id-EQ}(f) = 1/H_{id}(f)$$

$H_{id-EQl}(f)$ is called ID-equalization of left ear; $H_{id-EQr}(f)$ is called ID-equalization of right ear.

The equalization should be correct within $\pm 0,5$ dB within the frequency range from 50 Hz to 10 kHz, measured in 1/3rd octave bands.

All these types of equalization can be applied to an artificial head as well as to a binaural probe microphone, which is worn by the user himself instead of using an artificial head. If probe microphone techniques (like e.g. so-called MIRE techniques - microphone in real ear, see ISO 11904-1 [i.5]) are used the same equalization technique as applied for the HATS shall to be used.

The preferred recording methodology is an artificial head, since the characteristics of HATS are well described and problems due to mispositioning, differences for different persons wearing the binaural microphones/headset and equalization errors are unlikely.

To avoid ambiguity, the equalization technique chosen for the recording process should be the same as used in the equalization process for the loudspeakers (see clause 6.3).

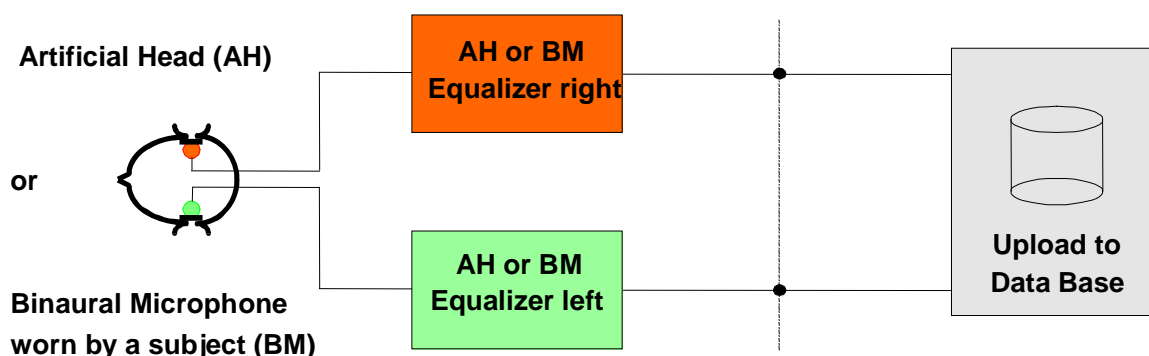The general recording setup is illustrated in Figure 1.

Figure 1: Recording setup

6   Loudspeaker Setup for Background Noise Simulation

6.1   Test Room Requirements

The reproduction technique chosen does not require specific types of rooms such as anechoic rooms. The technique is applicable in typical office rooms as well as in anechoic or semianechoic rooms. The playback room should meet the following requirements:

- **Room Size:**

    The room size should be at least $4,0 \times 4,0$ m to adequately accommodate the speaker placement in clause 6.2. The minimum room size shall be at least 2,5 m $\times$ 3 m. The room height should be made in consideration of the reverberation time. This leads to a typical range between 2,20 m and 2,50 m. Additionally, it is recommended to keep the room dimensions unequal and non-multiple of each other in order to minimize the standing waves problem.

- **Treatment of the Room:**

    Office type rooms should be equipped with a carpet on the floor and some acoustical damping in the ceiling as typically found in office rooms. A curtain should cover one or two walls in order to avoid strong reflections by hard surfaces in the room. The reverberation time of the room should be less than 0,7 s but higher than 0,2 s for each of the octave bands of 250 Hz and 8 kHz. Explicitly that would be: 250, 500, 1 000, 2 000, 4 000, 8 000 Hz.

    For anechoic or semianechoic rooms, diffuse treatment may be needed to meet reverberation time minimums.

- **Noise Floor:**

    In order to reduce the influence of external noise the noise floor measured in a room should be less than 30 $dB_{SPL}(A)$.

6.2   Loudspeaker Positioning

The loudspeakers should be positioned as indicated in Figure 2. The distance between the centre of the acoustical field (the position which is used for equalization and where the test arrangement is typically placed) should be 2 m. In office-type rooms, the distance between each loudspeaker and the centre of the simulation setup and between two adjacent loudspeakers should be approximately equal. However, depending on room geometry a slightly asymmetrical position of the loudspeakers may lead to better equalization results. Positioning too close to the corners of the rooms, i.e. closer than 50 cm away from the nearest wall surface, should be avoided whenever possible.

A good indicator of a non-optimal positioned loudspeaker is its frequency response measured at the centre of the acoustical field. For high quality loudspeakers, which provide a mostly flat frequency response under free-field conditions, deviations of more than 9 dB (compared to flat response) indicate strong influence from the room due to room modes or strong reflections. These require too strong equalization gains and result typically in a spatially inconsistent sound field.

In anechoic rooms, a slightly asymmetric positioning may help as well to overcome strong comb filter effects due to the symmetrical arrangement. These can again be identified in the frequency response, where narrow peaks and/or notches typically indicates a too strong comb filter effect due to too symmetrical positioning or due to non-optimum delay values (see clause 6.3).

The height of the four loudspeakers should be chosen so that the centre of the acoustical field is of the same height as the loudspeaker positions.

Imperfect frequency response in the low frequency range of the primary four loudspeakers should be compensated with a subwoofer. The positioning of the subwoofer is mostly uncritical; however, placing the subwoofer exactly in the corner of a room should be avoided. The subwoofer is intended to playback signal components lower than the cut-off frequency f_{min} of the four loudspeakers. By default, f_{min} is 120 Hz, i.e. the four loudspeakers should be capable of reproducing frequencies down to (at least) 120 Hz. For setups in smaller rooms and with smaller loudspeakers, this might be difficult to achieve. In these cases, it is possible to increase the cut-off frequency for the subwoofer/loudspeakers. The value of f_{min} shall always be less or equal 200 Hz. The cut-off frequency of high pass filters applied to the four loudspeakers shall match the cut-off frequency of the subwoofer.
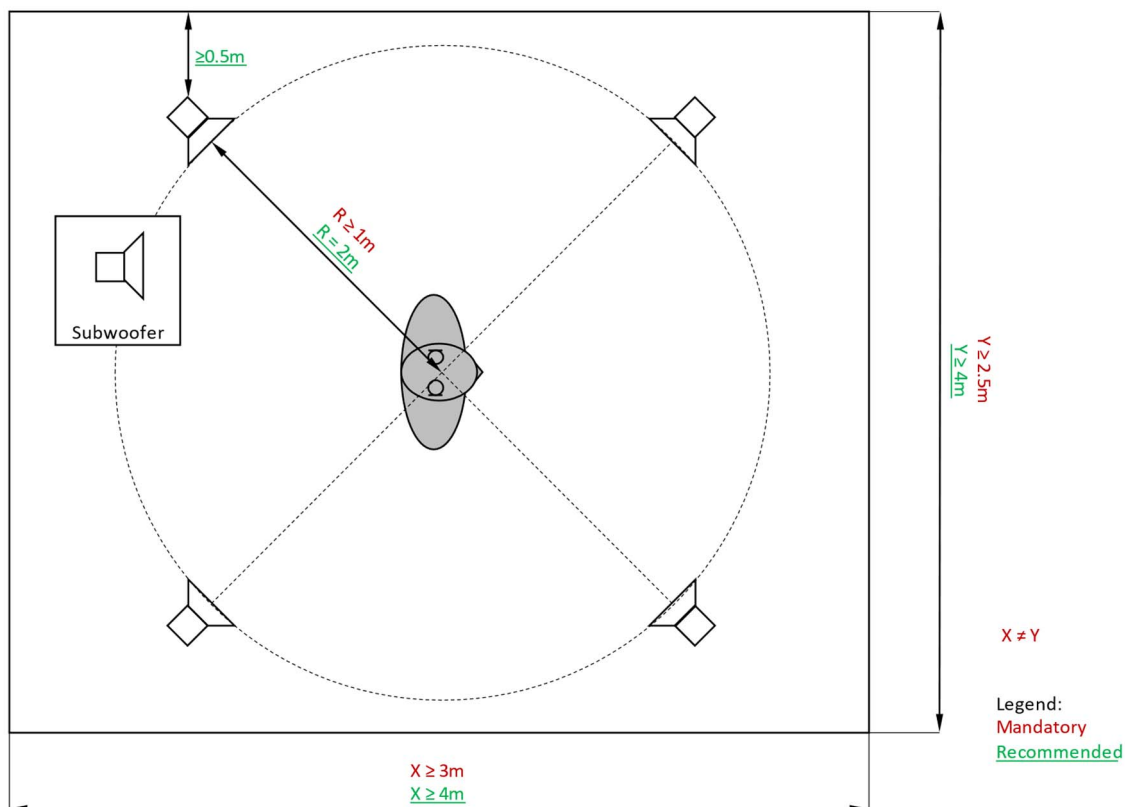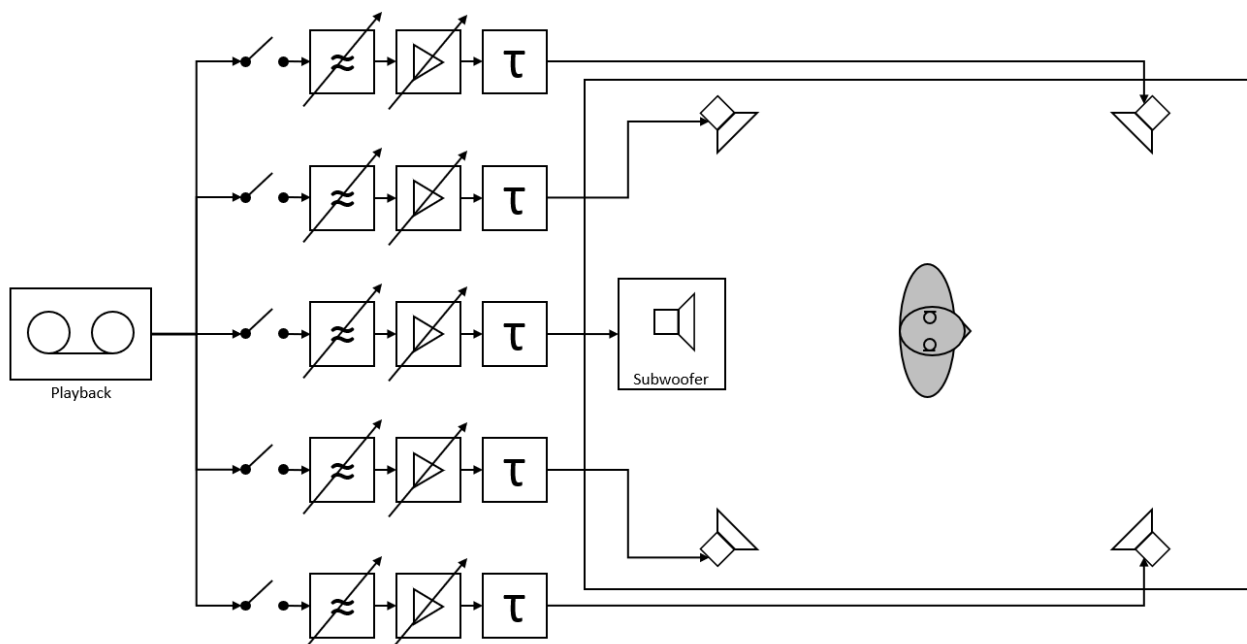

Figure 2: Loudspeaker arrangement in standard office rooms

6.3   Equalization and Calibration

6.3.1   Overview

A four-loudspeaker arrangement as described in clause 6.2 is positioned in a room, as specified in clause 6.1. The complete block diagram of the setup is shown in Figure 3.


Figure 3: General setup of the equalized four loudspeaker arrangement in an office room

For the equalization procedure, an artificial head (HATS) according to Recommendation ITU-T P.58 [2] equipped with two artificial ears according to Recommendation ITU-T P.57 [1] shall be used and its equalization shall correspond to the one used for the sound field recording (FF, DF or ID, see clause 5.2). The artificial head is positioned in the centre of the acoustical field (middle of the four loudspeakers, see Figure 3).

The equalization is conducted in several steps that are described in the following clauses. In case of difficulties with the equalization procedure, please refer to clause 6.3.8.

6.3.2   Separate equalization for each of the four loudspeakers

For the equalization of each loudspeaker, a pink (pseudo random) noise signal is used. It is fed sequentially to the front-left, the rear-left, the front-right and the rear-right loudspeaker with a frequency range of f_{min} to 20 000 Hz. The subwoofer covers the frequency range below f_{min} (see clause 6.2). As an example, an equalization step is illustrated for the front-left loudspeaker in Figure 4. For the other loudspeakers, the process is identical.
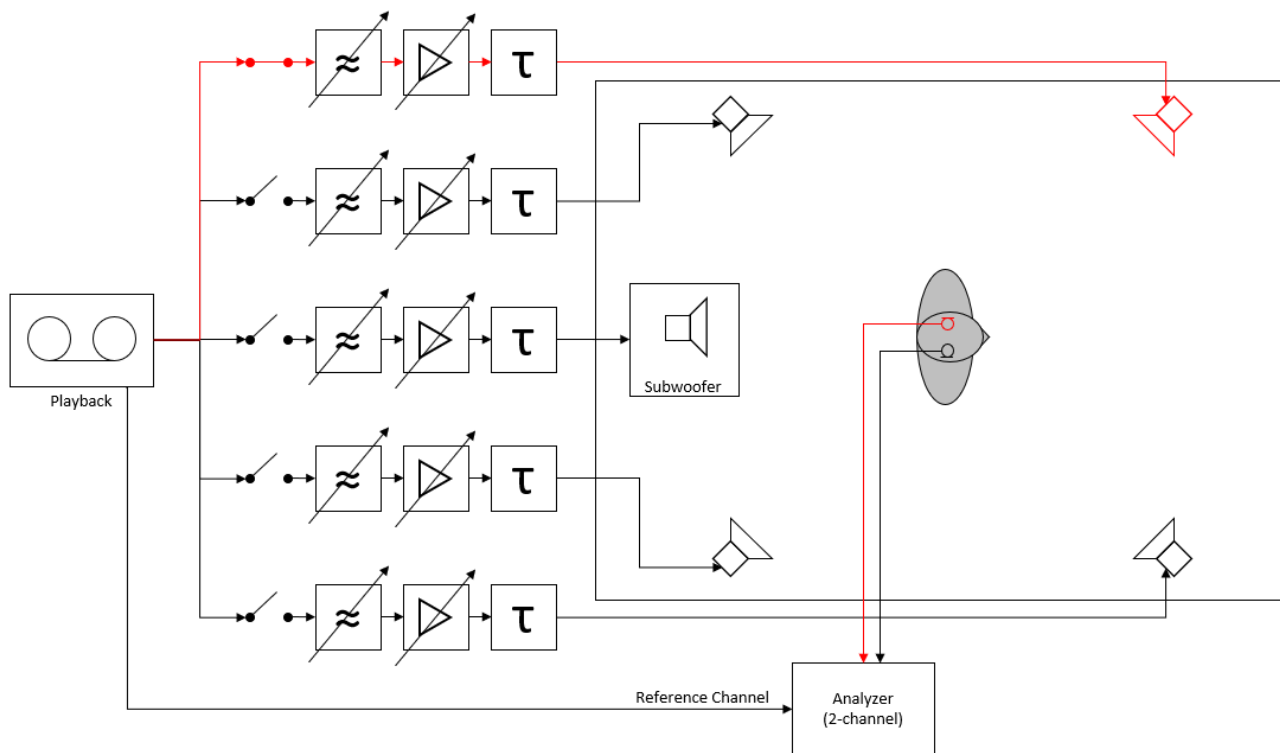
NOTE: Signal path for equalizing the front-left loudspeaker is marked red.

Figure 4: Block diagram of equalization setup

The pink noise signal is filtered with a high pass with cut-off frequency f_{min} and is then fed into the front-left loudspeaker for at least 2 s. Synchronously, the reproduced signal is recorded at the left ear's output of the artificial head. The power density spectra then are calculated for each recorded ear signal and for the original signal. The averaged magnitude of the two power density spectra are subtracted as follows:

$$\left|H_{LS}(f)\right| = \frac{\left|S_{LS}(f)\right|}{\left|S_{pn}(f)\right|}$$

$H_{LS}(f)$ is analysed in 1/3rd octaves. The measured resulting frequency response shall be flat between f_{min} and 10 kHz within a tolerance of ±3 dB.

To fulfil this condition, a filter needs to be designed which compensates the frequency response of the loudspeaker in the room. The compensating filter shall not lead to attenuation and/or amplification higher than 9 dB. If so, it indicates that the set-up arrangement is not optimal and shall be modified. The compensating filter frequency response is calculated from the initially measured loudspeaker frequency response by using the inverse filter function:

$$\left|H_{EQ}(f)\right| = \frac{1}{\left|H_{LS}(f)\right|}$$

The equalized noise signal is applied at the front left loudspeaker. Figure 5 shows an example of the measured loudspeaker frequency response after the equalization process.
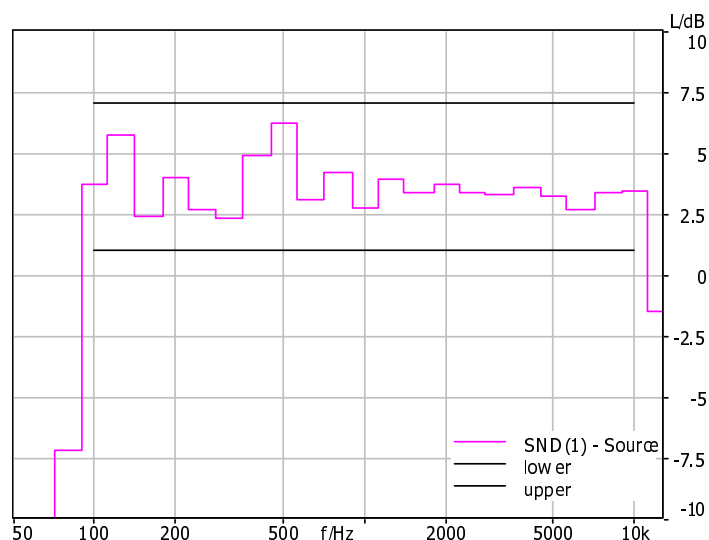
**Figure 5: Difference of the original signal and the equalized signal reproduced
by the front-left loudspeaker measured with the artificial heads left ear (without subwoofer)**

6.3.3 Separate level adjustment for each loudspeaker

After each of the loudspeakers is equalized separately, the sound pressure level of each loudspeaker shall be adjusted. To reproduce the original level of any sound source with this four-loudspeaker arrangement, the level of each of the four loudspeakers shall be reduced by 6 dB compared to the level of the original sound source due to the fact that the sound source is now reproduced by four loudspeakers.

This level adjustment can be achieved by recording and analysing the reproduced signal for each loudspeaker as described in the first step and after reducing the level by 6 dB below the original level.

6.3.4 Equalization for the two left-hand and the two right-hand loudspeakers

In the third step, the equalization is done for the two - separately equalized - left-hand and right-hand loudspeakers in pairs (see Figure 6). A stationary, realistic broadband noise (e.g. broadband office noise or car noise recorded under constant driving condition) is used as the test signal. This test signal is used in order to achieve the same frequency dependant correlation as for typical artificial head recordings and to avoid disturbing influences due to the periodicity of the pink pseudo random noise.
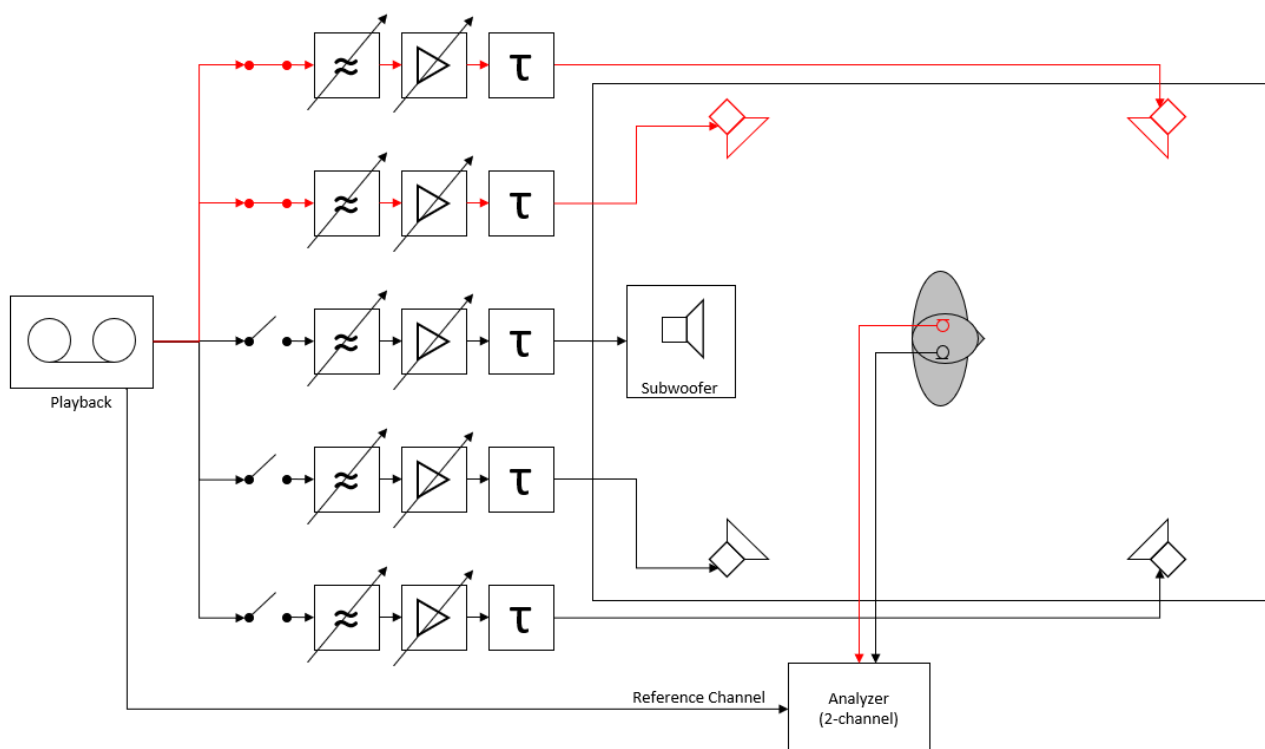
NOTE: Signal path for equalizing the left loudspeakers in combination is marked red.

Figure 6: Block diagram of equalization setup

The test signal is synchronously applied to the left-hand loudspeakers (front and rear) and - recorded with the left ear channel, then spectrally analysed in $1/3^{rd}$ octaves. Again, the measured difference of the recorded and the reproduced signal should meet the ±3 dB tolerance mask as shown in Figure 5. If the recorded signal fails the tolerance mask, the filter for one or both of the left-hand loudspeakers is re-designed to pass the tolerance mask.

   NOTE: The tolerance mask for a single loudspeaker might be violated, if the equalization for the combined left loudspeakers passes the tolerance mask.

The equalization for the right-hand loudspeakers is done similar.

6.3.5   Equalization and level adjustment for the subwoofer

The equalization for the subwoofer is done for the frequency range between 30 Hz and f_{min}. The frequency range above is cut off by a low pass filter providing a roll-off characteristics of at least 18 dB/oct. To avoid too much low frequency content, a highpass filter below 50 Hz with a roll-off characteristics of 12 dB/oct. should be applied. Again, pink noise is used as the test signal. Similar to the equalization for the four loudspeakers, the pink noise is reproduced by the subwoofer and recorded at the right ear. On the basis of the difference of the recorded and the original pink noise spectra, a filter is designed, which allows for the difference (analysed in $1/3^{rd}$ octaves) to be spectrally flat within a ±3 dB tolerance mask.

The level is adjusted to the level of the original sound source.

6.3.6   Delay compensation

For the four loudspeaker arrangement including the subwoofer, a compensation of the individual loudspeaker delay to the centre of the acoustical field is required in order to keep the correlation of the recorded noise material with respect to the centre of the acoustical field. After this alignment of delays, additional delays are introduced into the four playback paths of the loudspeakers to increase the diffusivity of the sound field and to decrease the cross-talk.

Wrong delay adjustment might lead to comb filter effects, which leads to unnatural sounding of the background noise signals. The following default delay values for the four loudspeakers arrangement are recommended:

-        0 ms (front-left)

- 11 ms (front-right)

- 17 ms (rear-left)

- 29 ms (rear-right)

However, this may not be the optimum delay setting for all types of rooms. Since room and loudspeakers properties, dimensions and positions are not always the same, these delay values may be replaced by better ones proven to lead to a better equalization results. No delays higher than 50 ms shall be used.

Due to the low frequencies produced by the subwoofer, a delay compensation for the subwoofer is not needed.

6.3.7   Overall equalization for all loudspeakers

The last step is to verify the equalization process including the delay compensation. Therefore a stationary, realistic pre-recorded broadband background noise signal is fed to the four-loudspeaker arrangement. All equalizers, level and delay adjustments are active, the signal is reproduced by all loudspeakers including the subwoofer (see Figure 7). The reproduced sound field is recorded at both ears of the artificial head. For each ear-channel, the recording is analysed in $1/3^{rd}$ octave band spectrum, which is referenced to the spectrum of the original signal. The measured frequency response for the left ear signal and the right ear signal shall be within a tolerance of ±3 dB for each channel in the frequency range from 50 Hz to 10 kHz.
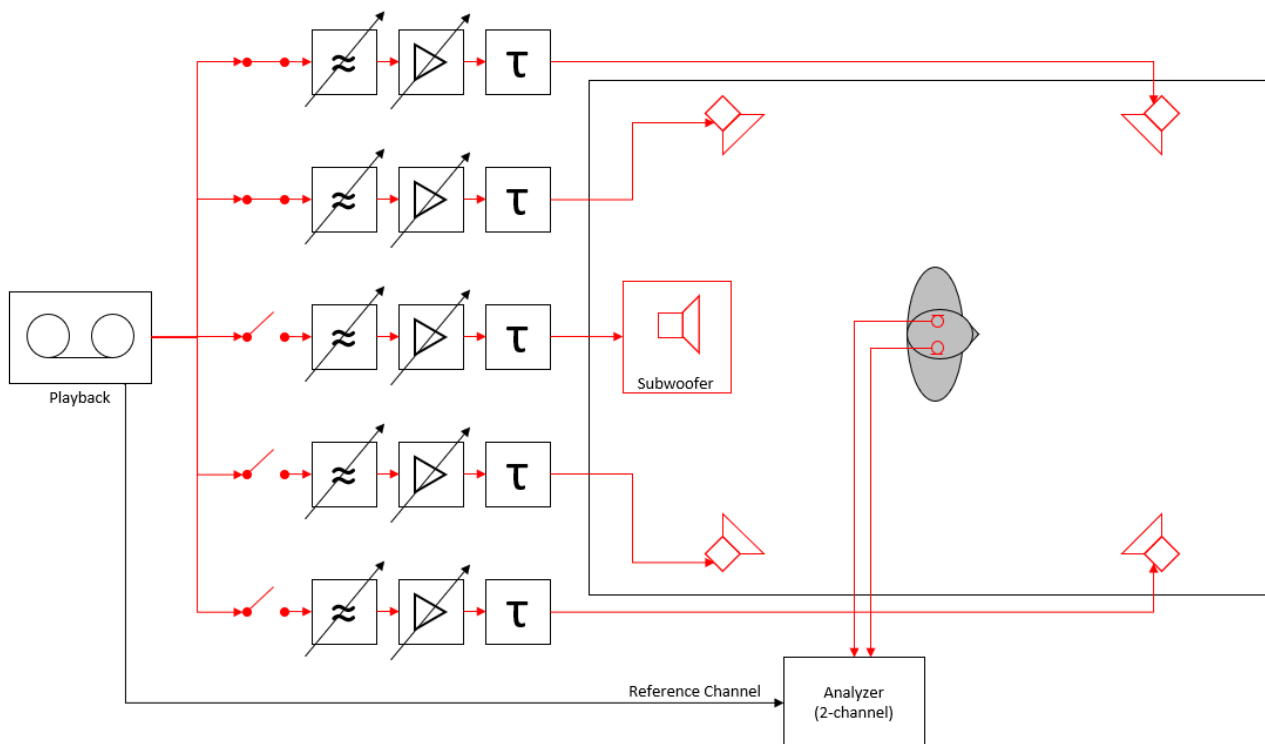
NOTE: Signal paths for overall equalization is marked red.

Figure 7: Block diagram of equalization setup

6.3.8   Troubleshooting failed equalizations

If the result is outside the tolerance, it needs to be checked which of the four loudspeakers is mostly contributing to the equalization error. Typically the tolerance mask is violated only in a small frequency range.

In the first step, it is checked whether the equalization error is due to the cross-talk signal. Therefore the loudspeakers producing the cross-talk signal are switched off. If now the equalization error disappears, first the front loudspeaker (producing the cross-talk signal) and then the back loudspeaker (producing the cross-talk signal) is switched on again. By this, the loudspeaker mostly contributing to the equalization error is found. The equalization of this loudspeaker then is corrected and overall the equalization is checked again for the left and the right channel. This procedure is repeated until the equalization error is within the required tolerance for both channels. Sometimes, the equalizer for the front and the rear loudspeaker needs to be corrected.

If the equalization error is mostly due to the loudspeakers producing the direct sound for the channel under consideration, then the equalizers for these loudspeakers need to be corrected. The procedure is the same as described for the cross-talk loudspeakers as described above.

NOTE 1:   In exceptional cases, an equalization correction needs to be applied to both the direct sound loudspeakers and the cross-talk loudspeakers. In such cases, it is recommended to carefully adjust the equalization of the different loudspeakers contributing to the error and not try to minimize the equalization error by correcting the equalization of a single loudspeaker by a strong filter change.

NOTE 2:   If the optimal geometry of the setup cannot be achieved, alterations may be applied:

- Loudspeakers can be set up along a circle down to 1m in radius

- Loudspeakers and/or HATS can be set up asymmetrically within the room

- Loudspeakers can be set up closer the wall, down to what is physically possible

These modifications are only allowed so long as system equalization is successfully performed.

6.3.9   Automated equalization and calibration procedure

The steps described in clauses 6.3.2 to 6.3.7 may in some cases require some manual tuning of the equalization filters to obtain the required flat frequency response within the ±3 dB tolerance (for each channel, in the frequency range from 50 Hz to 10 kHz). Thus, an equalization repeated in the same room and with same test equipment could lead to different sets of filters, which depend on e.g. subjective preferences or experience of the test operator. To avoid such variability caused by manual tuning of the noise simulation, an alternative and fully automated equalization procedure for binaural recordings is described in clause 7.7 of ETSI TS 103 224 [3]. It is considered equivalent, i.e. can alternatively be used as a replacement that complies with the equalization procedure described in clauses 6.3.2 to 6.3.7 of the present document.

A comprehensive study on both equalization methods can be found in ETSI TR 126 921 [i.20].

6.4   Accuracy of the reproduction arrangement

6.4.0   Introduction

During the validation process, a variety of experiments were conducted in order to get an estimate of the reproduction accuracy of the four-loudspeaker playback arrangement. This validation was conducted in various steps. All validations were performed by comparing the 1/12th octave spectra and 1/3rd octave spectra. Since the results in the following are very similar, only the 1/3rd octave results are displayed.

6.4.1   Comparison between original sound field and simulated sound field

The data reported below were derived from the following experiment.

In an office room, a background noise sound field was generated. Two loudspeakers and a computer, which produce typical fan noise, were positioned. The sound field produced by this arrangement was recorded using an artificial head positioned in the middle of the room. The arrangement is shown in Figure 8.
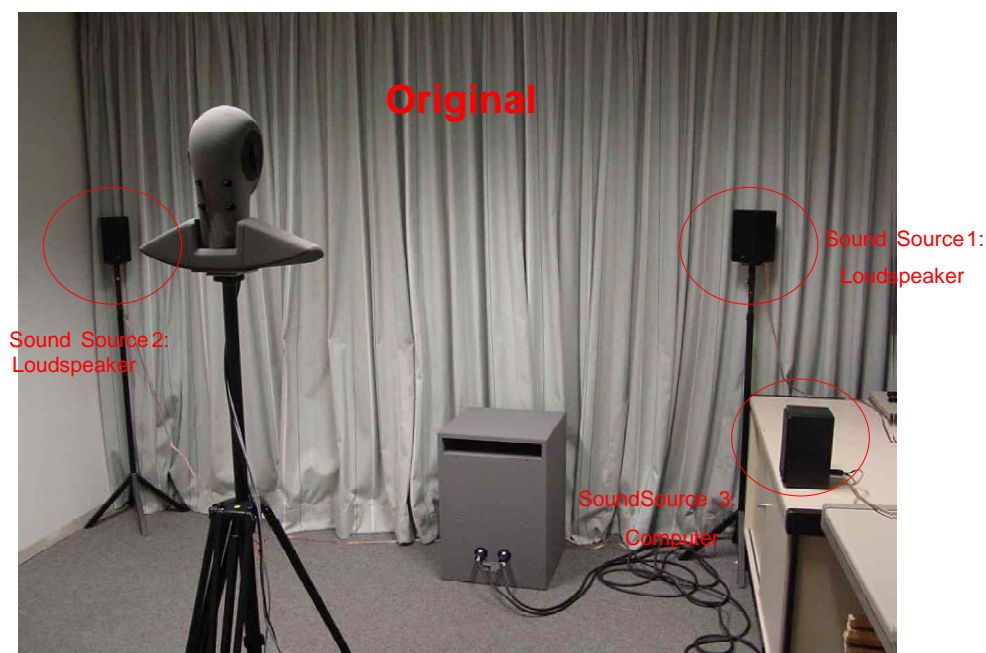
Figure 8: Simulation arrangement in a standard office room

In addition, the four-loudspeaker arrangement was installed in the same room. The installation was made so that the artificial head was positioned in the centre of the acoustical field of the simulation arrangement.

The comparison was made by recording the original sound field and comparing the recorded signals to the signals recorded when using the four-loudspeaker simulation system.
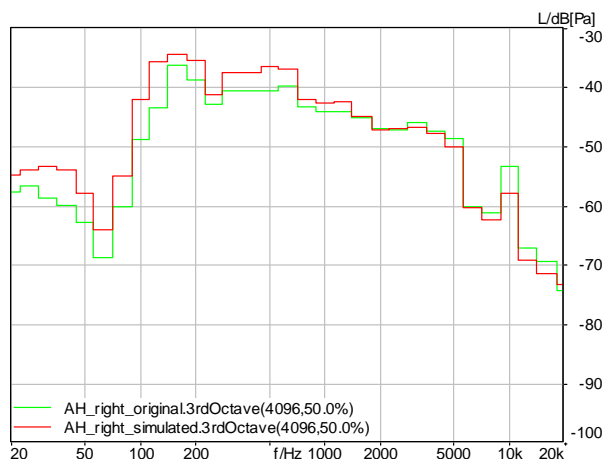
The spectra recorded at the left and the right ear are shown in Figures 9 and 10. It can be seen that the spectra of the original and the simulated sound field match quite well.
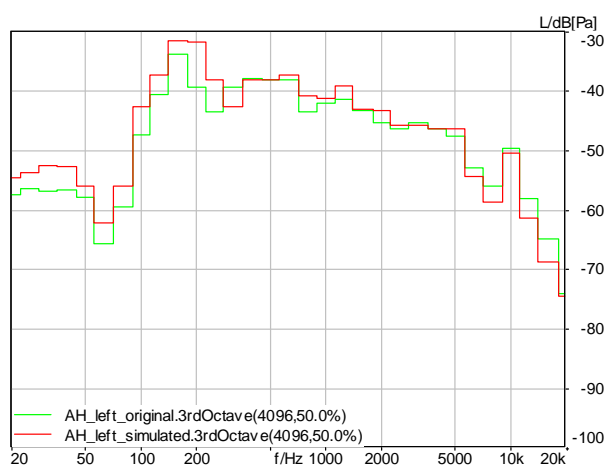
NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 9: Power density spectra measured at the artificial heads left ear

NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 10: Power density spectra measured at the artificial heads right ear

6.4.2 Displacement of the test arrangement in the simulated sound field

For these experiments the head and torso simulator used for the measurements was displaced by 10 cm to the frontal right and 10 cm backwards left. For these positions in comparison to the centre position of the sound field where the artificial head was equalized to, the spectral differences are displayed in Figures 11 and 12. It can be seen that despite of the displacement the recorded spectra still match comparably well the maximum error measured in third octaves, which can be expected due to misplacing of about 5 dB.
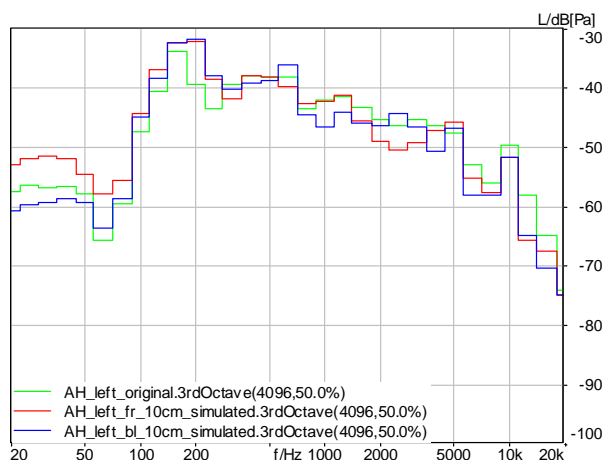
NOTE: Displaced 10 cm to the frontal right (red curve), displaced 10 cm backwards left (blue curve).

**Figure 11: Original (green curve) and simulated power density spectra measured
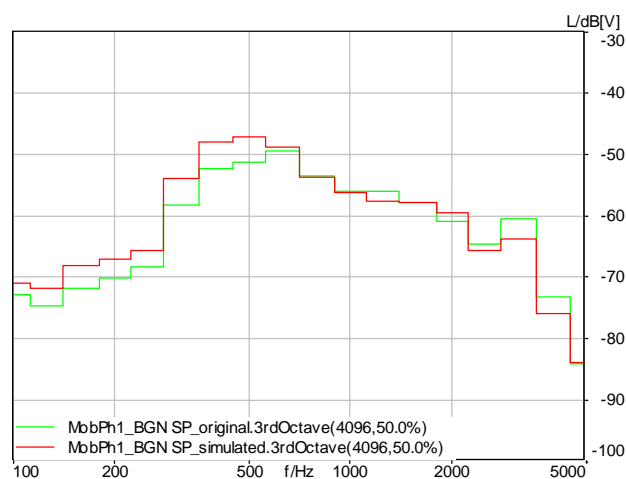at the left ear of the artificial head**
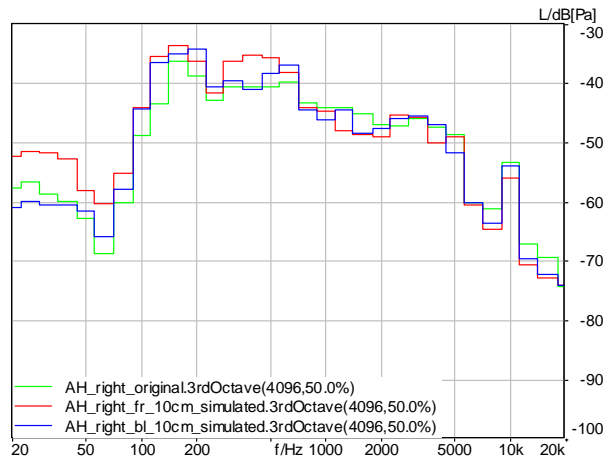
NOTE: Displaced 10 cm to the frontal right (red curve), displaced 10 cm backwards left (blue curve).

**Figure 12: Original (green curve) and simulated power density spectra measured
at the right ear of the artificial head**

A further test was conducted by moving the artificial head by 30 cm away from the centre of the four loudspeaker arrangement to the frontal right position. Again the difference of the spectra recorded in this position is displayed for the right and the left output signal of the artificial head, see Figures 13 and 14. It can be seen that for the left ear of the artificial head now the maximum difference measured in third octave is increase up to about 7 dB compared to the original position. For the right artificial ear the spectra still match quite well.
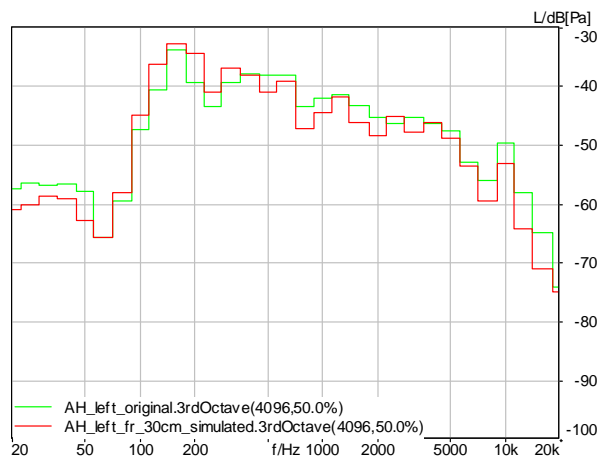
NOTE: Displaced 30 cm to the frontal right (red curve).

**Figure 13: Original (green curve) and simulated power density spectra measured
at the left ear of the artificial head**
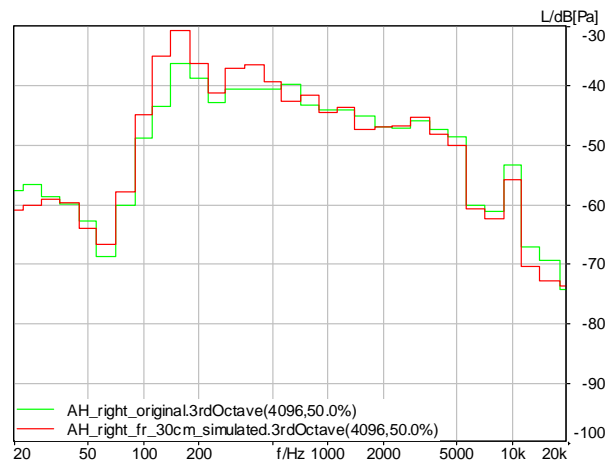
NOTE: Displaced 30 cm to the frontal right (red curve).

**Figure 14: Original (green curve) and simulated power density spectra measured
at the right ear of the artificial head**

6.4.3 Transmission of background noise: Comparison of terminal performance in the original sound field and the simulated sound field

Handset terminals:

Further experiments were aimed to validate the performance of the four-loudspeaker arrangement with typical terminals applied to the test arrangement. For terminal tests a head and torso simulator according to Recommendation ITU-T P.58 [2] equipped with an artificial ear according to Recommendation ITU-T P.57 [1], type 3.4 was placed in the centre of the four-loudspeaker arrangement. Different terminals were applied to the HATS, positioned as described in Recommendation ITU-T P.64 [i.10]. For comparison the output signals of the terminals in sending direction were recorded. These recordings were conducted when applying the original background noise field and when applying the simulated background noise field. The typical differences measured in sending direction for a terminal with background noise signal processing is shown in Figure 15. The result for a linear terminal without background noise signal processing is shown in Figure 16.

NOTE: Original sound field (green curve), simulated sound field (red curve).
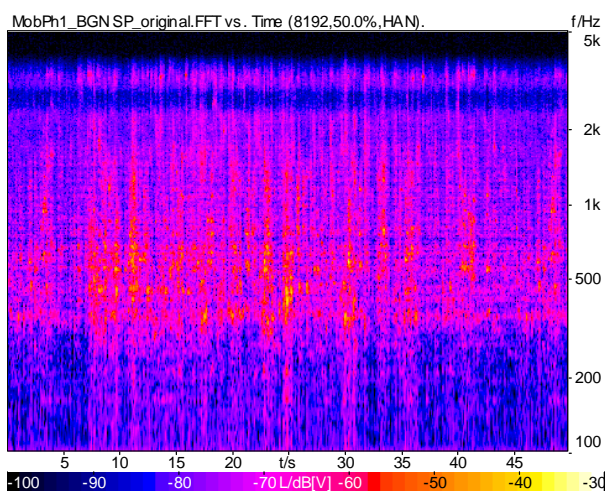
**Figure 15: Power density spectra measured with a common mobile phone
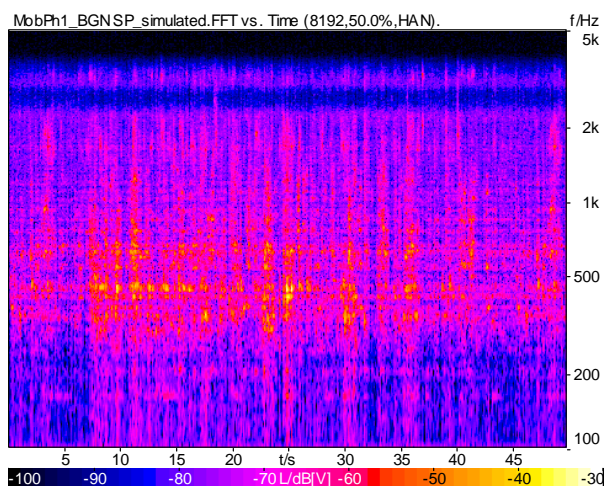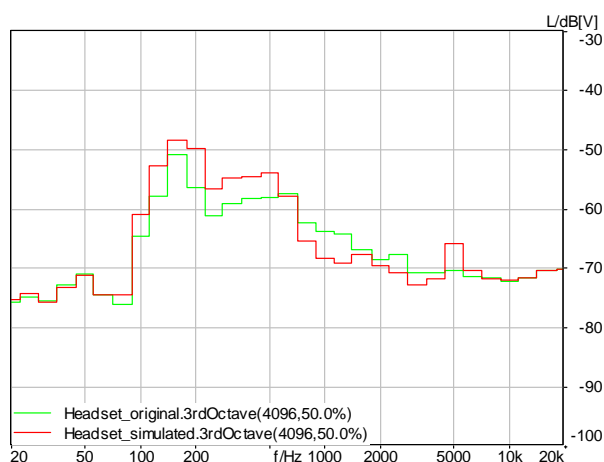(with background noise signal processing) positioned at the right ear of the artificial head**
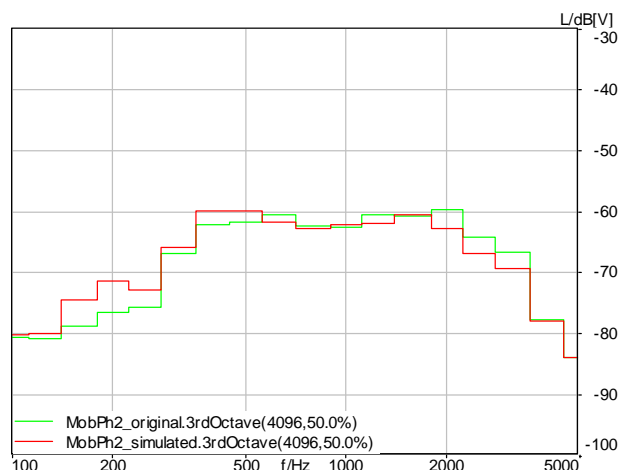
NOTE: Original sound field (green curve), simulated sound field (red curve).

**Figure 16: Power density spectra measured with a common mobile phone
(without background noise signal processing) positioned at the right ear of the artificial head**

Furthermore, the temporal structure of the transmitted background noise was evaluated. The comparison of the temporal structure when applying the original background noise field to the situation with simulated background noise field is shown in Figures 17 and 18. In both figures the spectral presentation of output signal is shown. The characteristics of the temporal structure of the background noise are very comparable. It can be expected that also for other types of terminals with integrated signal processing that the simulated noise field will produce similar behaviour than the terminal would show in the original noise field.

**Figure 17: 3D power density spectrum of background noise measured with a common mobile phone
(with background noise signal processing): original sound field**

Figure 18: 3D power density spectrum of background noise measured with a common mobile phone (with background noise signal processing): simulated sound field

Wideband headset:

The same types of experiments as described above were conducted using a wideband headset equipped with a directional microphone. The results - the transmitted noise spectra for the original sound field compared to the simulated sound field - are shown below in Figure 19. Since the orientation of the microphone was in such a way that the lowest sensitivity of the microphone was exactly facing one of the noise generating loudspeakers on the left side, the transmitted signal energy of the background noise is lower in the original sound field as compared to the simulated sound field where the sound field is more diffusive. The measured difference of about 5 dB can be regarded as a kind of worst case situation, as the more diffusive the recorded sound field is the less errors are expected in the simulation.

Desktop hands-free terminal:

Since the background noise simulation should be suitable for handset terminals and hands-free terminals another experiment was conducted using a desktop hands-free terminal in the same test room. A table was positioned in the test room in front of the artificial head, which was still positioned in the centre of the simulated acoustical field. The hands-free terminal was positioned as described in Recommendation ITU-T P.340 [i.9]. Again the transmitted background noise signal between the original sound field and the simulated sound field are compared. The resulting power density spectra for the transmitted background noise are shown in Figure 20. Although the four-loudspeaker arrangement was not reequalized after the table was placed in the room, the spectra of the transmitted background noise match rather well. It can be expected that even for desktop hands-free phones placed in the simulation environment, the measured performance of background noise transmission is comparable to the one measured in typical office environments.

NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 19: Power density spectra measured with a common (wideband) headset with directional microphone worn by the artificial head
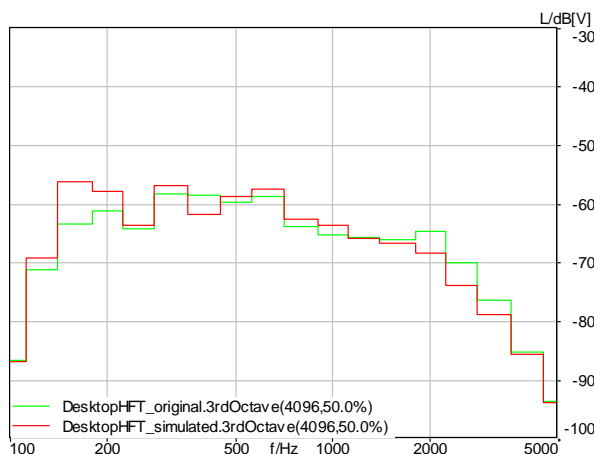
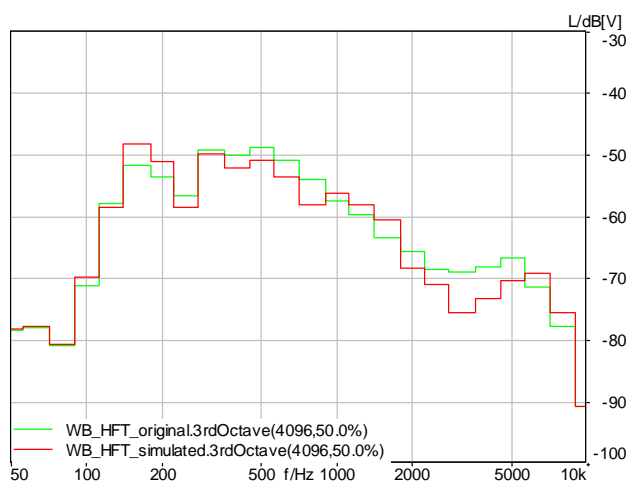NOTE: Original sound field (green curve), simulated sound field (red curve).

**Figure 20: Power density spectra measured with a common desktop
hands-free telephone positioned in front of the artificial head**

Wideband conference system:

The conference system was installed like a typical desktop hands-free system. The setup is according to Recommendation ITU-T P.340 [i.9], the setup used is identical to the hands-free setup as described above. The wideband conference system was equipped with a directional microphone (on the table), Recommendation ITU-T G.722 [i.11] coding was used.

The difference between the transmitted background noise of the original sound field and the simulated sound field is shown in Figure 21.

As seen already for the narrow band hands-free system, the measured spectra are rather close; the maximum difference is about 6 dB at 3 kHz.

NOTE: Original sound field (green curve), simulated sound field (red curve).

**Figure 21: Power density spectra measured with a wideband conference system
positioned in front of the artificial head**

6.5   Simulation of additional acoustic conditions

Some acoustic conditions cannot be simulated with sufficient accuracy using the binaural capture and reproduction approach described above. One specific example is the simulation of a person talking nearby to the test position, whose speech is considered as interfering with the desired speech at the test point. In this case, while binaural recording preserves the signal and interaural information, reproduction over four loudspeakers positioned as described in clause 6.2 does not accurately simulate the generation of sound by the nearby interfering talker at the test point.

For this case, it is recommended to place an artificial mouth at the location of the desired simulated interfering talker(s). The artificial mouth shall be properly equalized. The speech level at MRP of the interfering talker(s) can be set as is needed to simulate the desired scenario. The selection of speech material is outside the scope of the present document, but should be appropriate for the desired scenarios. Figure 22 illustrates one possible arrangement using a second HATS to simulate an interfering nearby talker, but the specific arrangement and location should be as needed to simulate the desired scenarios.
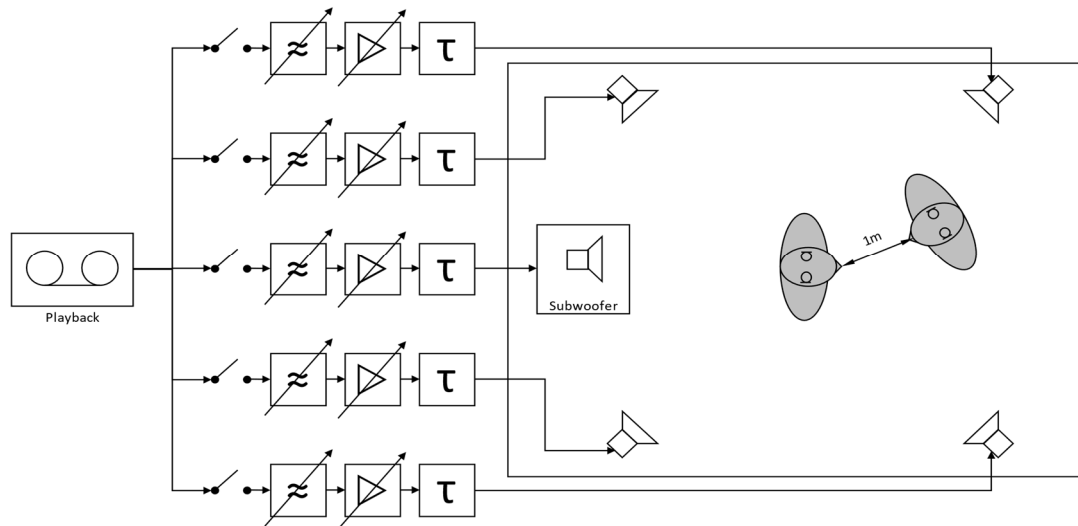
Figure 22: An example of acoustic simulation of single talker

7   Background Noise Simulation in cars

7.1   General setup

Some types of terminals are mostly used in a car environment, e.g. portable hands-free units with mobile phones or hands-free terminal systems, which are integrated in cars. For portable hands-free phones worn by the user in the car, the setup and equalization process is similar to the one described in the office room except that the artificial head is positioned at the driver's seat instead of positioning it in the middle of the room.

For hands-free terminals installed in the car, in principle the same setup is used for background noise simulation in a car and in office rooms. Four loudspeakers and a subwoofer are used. The subwoofer is positioned in the trunk of the car. The two rear loudspeakers are positioned in the back of the car between the headrest and the window respectively the C-pillar. The two front loudspeakers are positioned either above the instrument panel or at the bottom in the footwell (only in cases where the instrument panel position is impossible to realize) - depending on the tested car setup. The block diagram of the setup is shown in Figure 23.

It should be noted that the use of the four loudspeaker arrangement in a car is more sensitive concerning the positioning of the microphone than in the setup inside a test room. Due to the smaller car cabin room dimensions and the surrounding reverberant windows standing waves, interferences can easily occur.

Therefore, it is highly recommended to use the same microphone for equalization and terminal testing. This requires that the microphone can be separated from the hands-free terminal and can be used for recordings and system equalization as well. The equalization procedure with the terminal's microphone is described in clause 7.3. This ensures that the sound field is exactly reproduced for the relevant position.

If the terminal's microphone signal is not accessible for recording, equalization/calibration and playback, a pair of cardioid microphones should be used instead. They should be located as close as possible to the expected terminal microphone position to obtain best possible sound field simulation. The equalization with a pair of cardioid microphones is described in clause 7.4.

In general, the recording and equalization methods described can be applied for other microphone positions in a car, like e.g. at microphone positions intended for the co-driver. The stereo cardioid microphone technique at other locations in the car may be applicable but its accuracy may be reduced.
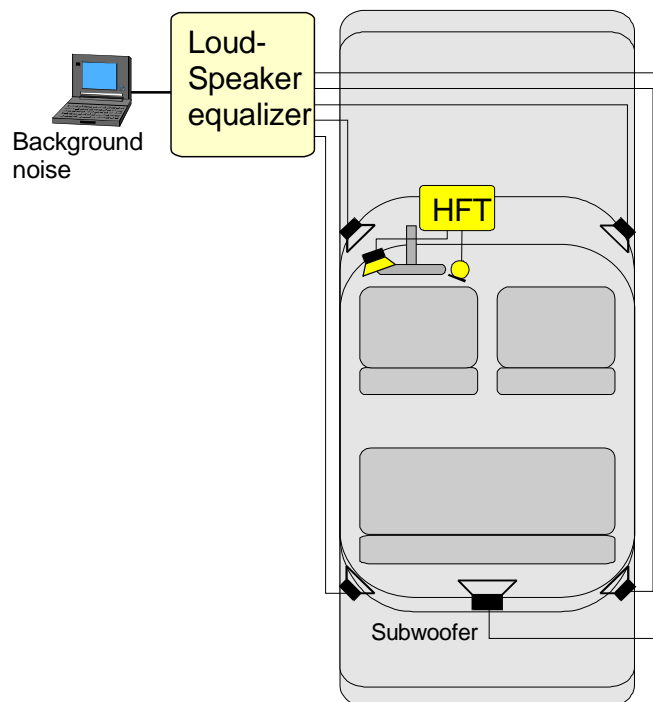
**Figure 23: General setup of four loudspeaker arrangement
and hands-free terminal in a car**

7.2   Recording arrangement

7.2.0   Introduction

The recording arrangement depends on the microphone setup that is used for the equalization/calibration and playback. The same microphones and positions shall be used for recording and equalization/calibration.

7.2.1   Recording setup with the terminal's microphone

The terminal's microphone is positioned at the location intended to be used with the hands-free terminal, like e.g. under the ceiling or in or on top of the rear view mirror. The position shall be the same for the recordings, the calibration/equalization and for the playback with the simulation arrangement. A block diagram of the recording arrangement is shown in Figure 24.

With this configuration, recordings should be made at different driving speeds as required to simulate real use conditions. Pass-by effects and road bumps should be avoided.

> NOTE: The microphone should be attached firmly to the ceiling to avoid structure-borne noise induced by accidental movements of the microphone.
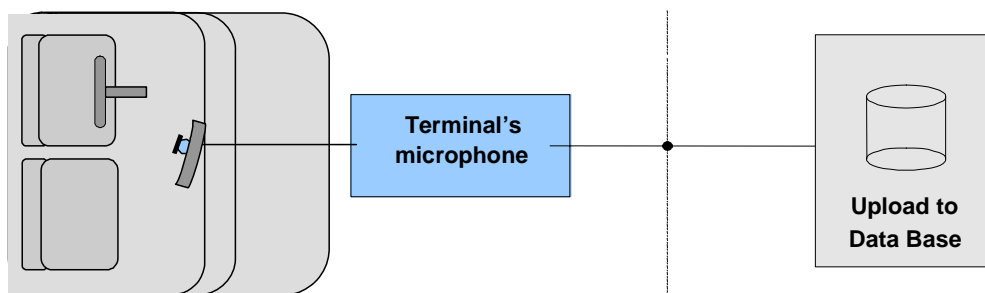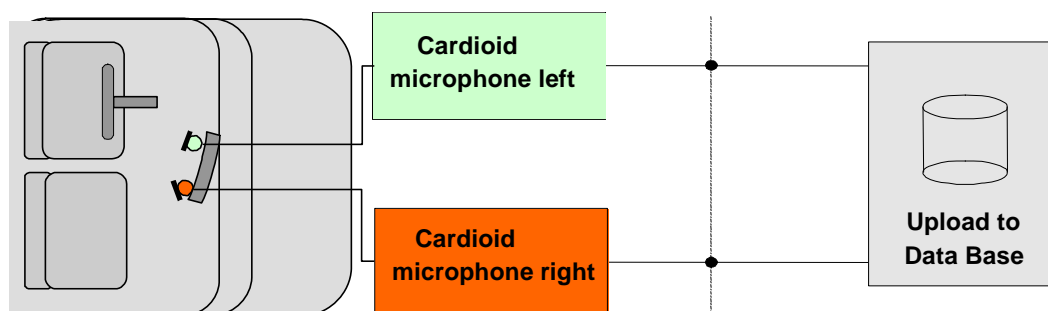
Figure 24: Setup for recordings with terminal microphones in cars

7.2.2   Recording setup with a pair of cardioid microphones

If a pair of cardioid microphones is used for equalization and playback, the recordings shall also be done with these microphones. They should be positioned as close as possible to the expected hands-free microphone position, see Figure 25. When a microphone array is used as the hands-free microphone, the centre of the array should be used as the reference point for positioning the cardioid microphones.

The microphones should be fastened with shock mount to avoid recording structure-borne noise.

Figure 25: Setup for recordings with a pair of cardioid microphones in cars

With this configuration recordings should be made for different driving speeds as required to simulate real use conditions. Pass-by effects and road bumps should be avoided.

7.3   Equalization and calibration with the terminal's microphone

7.3.1   Overview

A four loudspeaker arrangement as described in clause 7.1 is positioned in a car in which the hands-free terminal is installed. A complete block diagram of the setup is shown in Figure 26.
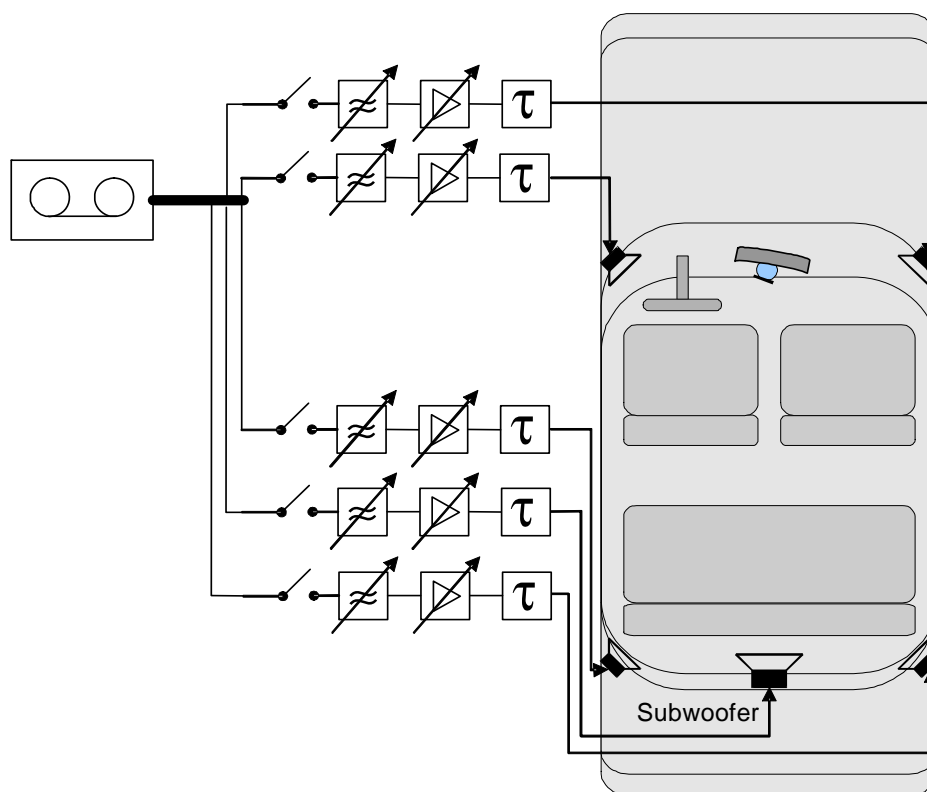
**Figure 26: General setup of the equalized four loudspeaker arrangement in a car**

The equalization is conducted in several steps that are described in the following clauses.

7.3.2   Separate equalization for each of the four loudspeakers

For the equalization of the loudspeakers, a stationary, realistic broadband noise (e.g. broadband car noise recorded under constant driving conditions) is used. It is sequentially fed to the front-left, the rear-left, the front-right and the rear-right loudspeaker. The feeding should be done for a frequency range of f_{min} to 20 000 Hz (for the four loudspeakers only). The subwoofer covers the frequency range below f_{min} (see clause 6.2).

As an example, the equalization is described for the front-left loudspeaker (see Figure 27). For the other loudspeakers the process is applied accordingly.

The stationary, realistic broadband noise is filtered with a high-pass with cut-off frequency f_{min} and then fed to the front-left loudspeaker for at least 10 s. Synchronously, the reproduced signal is recorded with the terminal's microphone. The power density spectra then are calculated for the recorded microphone signal (S_{LS}) and for the original signal (S_{bbn}). The averaged magnitude of the two power density spectra are subtracted as follows.
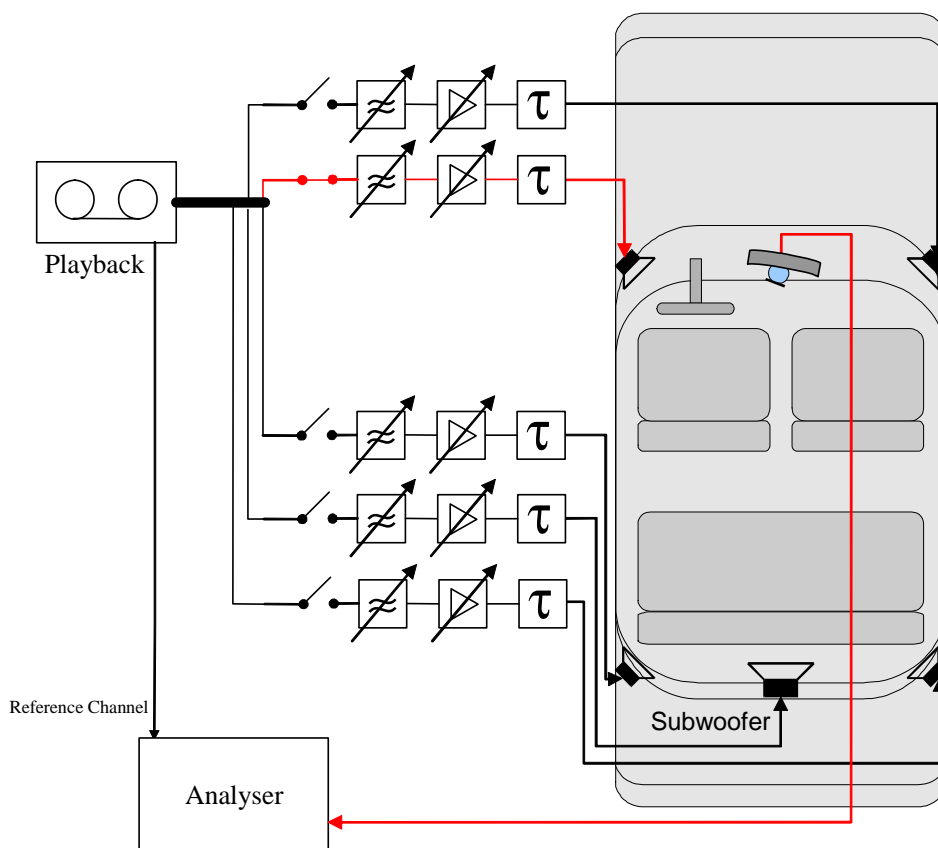
NOTE: Signal path for equalizing the front-left loudspeaker is marked red.

**Figure 27: Block diagram of equalization setup**

$$\left|H_{LS}(f)\right| = \frac{\left|S_{LS}(f)\right|}{\left|S_{bbn}(f)\right|}$$

$H_{LS}(f)$ is analysed in $1/3^{rd}$ octaves. The measured resulting frequency response should be flat between f_{min} and 10 kHz within a tolerance of ± 3 dB.

To fulfil this condition, a filter $H_{EQ}(f)$ needs to be designed that compensates the frequency response of the loudspeaker and the air path in the car. Its frequency response is calculated from the initially measured loudspeaker frequency response by using the inverse filter function:

$$\left|H_{EQ}(f)\right| = \frac{1}{\left|H_{LS}(f)\right|}$$

The equalized broadband noise is applied at the front-left loudspeaker again. Figure 28 shows an example of the measured loudspeaker frequency response after the equalization process.
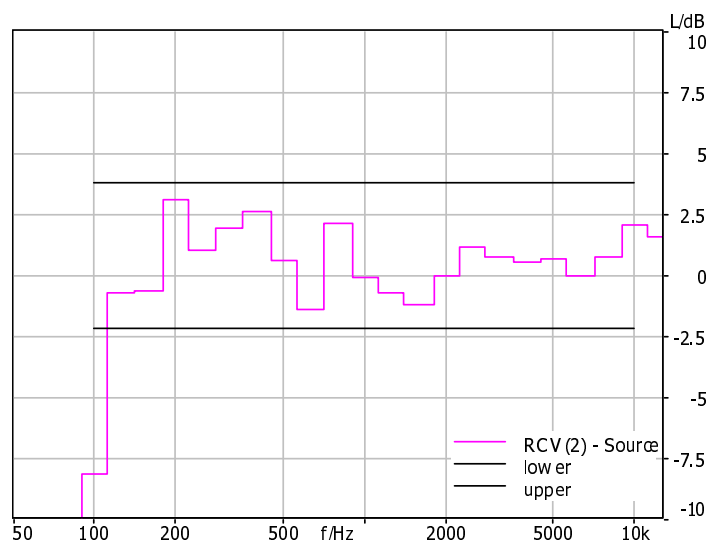
Figure 28: Difference of the original signal and the equalized signal reproduced by the front-left loudspeaker measured with the terminal's microphone (without subwoofer)

7.3.3 Separate level adjustment for each loudspeaker

After each loudspeaker is equalized separately, the sound pressure level of each loudspeaker shall be adjusted. To reproduce the original level of any sound source with this four-loudspeaker arrangement, the level of each of the four loudspeakers shall be reduced by 6 dB compared to the level of the original sound source due to the fact that the sound source is now reproduced by four loudspeakers.

This level adjustment can be achieved by recording and analysing the reproduced signal for each loudspeaker as described in the first step and after reducing the level by 6 dB below the original level.

NOTE: For microphones with a strong directivity to the back of the car, an improved simulation can be achieved by adjusting the sound pressure level of the rear loudspeakers to an approximately 5 dB higher level and the sound pressure level of the front loudspeakers to an approximately 3 dB lower level. Be aware that the overall sound pressure level is still the same as for the original sound.

7.3.4 Equalization for the two front and the two rear loudspeakers

In the third step, the equalization is done for the two - separately equalized - front respectively rear loudspeakers in combination (see Figure 29). The test signal is synchronously applied to the front loudspeakers and - similar to step 1 - recorded with the terminal's microphone and spectrally analysed in $1/3^{rd}$ octaves. Again the measured difference of the recorded and the reproduced signal shall be within the ±3 dB tolerance mask as shown in Figure 28. If the recorded signal does not pass the tolerance mask, the filter for one or both of the front loudspeakers should be re-designed to pass the tolerance mask.

NOTE: The tolerance mask for a single loudspeaker might be violated, if the equalization for the combined front loudspeakers passes the tolerance mask.

The equalization for the rear loudspeakers is applied accordingly.
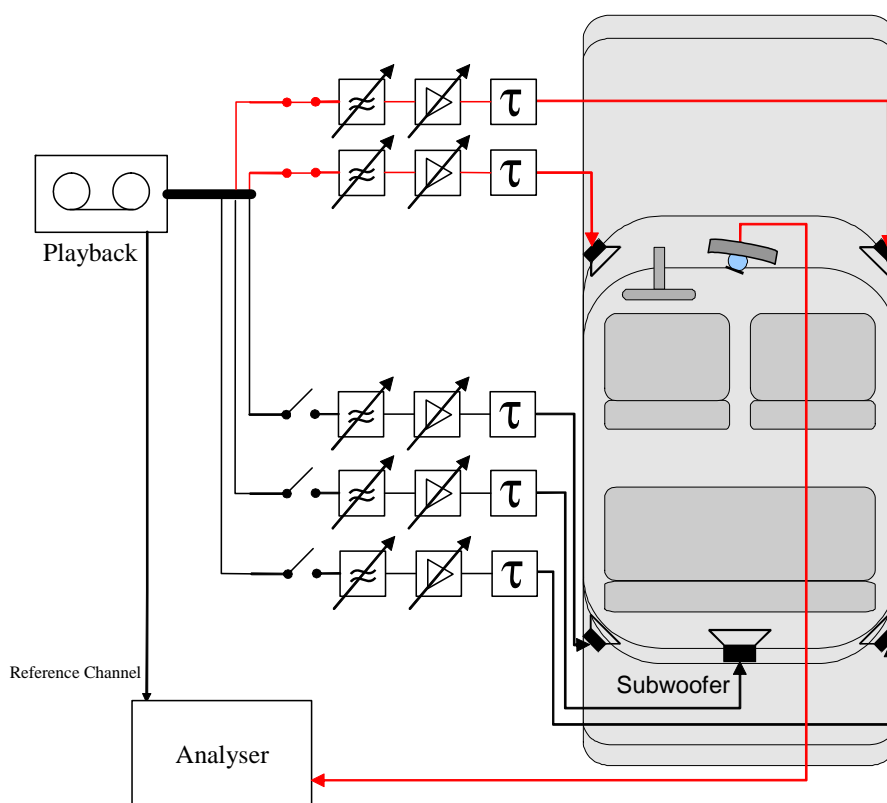
NOTE: Signal paths for equalizing the front loudspeakers in combination are marked red.

Figure 29: Block diagram of equalization setup

7.3.5   Equalization and level adjustment for the subwoofer

The equalization for the subwoofer is applied for the frequency range between 30 Hz and f_{min}. A low-pass filter providing a roll off characteristics of at least 18 dB/octave shall be used for the subwoofer channel. Again a stationary, realistic broadband noise is used as test signal. Similar to the equalization for the four loudspeakers, the broadband noise is reproduced by the subwoofer and recorded with the terminal's microphone. The difference between the recorded and the original broadband noise spectra defines an equalization filter, which shall provide a flat frequency response for the subwoofer (analysed in $1/3^{rd}$ octaves) within a ± 3 dB tolerance mask.

The level is adjusted to the level of the original sound source in this frequency range.

7.3.6   Delay adjustment

The delay from each loudspeaker to the microphone's position shall be measured and compensated individually. Compensating with a single common delay for all loudspeakers would produce the highest possible signal correlation at the microphone's position. Similar as in clause 6.3.6, delays are artificially increased after compensation, which will provide a low correlation at the microphone's position, i.e. leading to a mostly a diffuse sound field. With this method, the sound directivity is not correctly reproduced, but an equalization with only one microphone does in any not include any directivity information. The advantage of providing a diffuse sound field at the microphone's position is that the spectral equalization can typically be realized without further efforts.

For determining the arbitrary delay values per channel, the following rules should be followed:

-      The delay for front loudspeakers should be shorter than for the rear loudspeakers.

-      The minimum delay difference should be greater than the FFT-length used (typically > 20 ms).

-      The delays including the acoustical delay should be chosen such that they are not multiple of each other.

7.3.7   Overall equalization

The last step is to verify the equalization process including the delay compensation. Therefore the stationary, realistic broadband background noise signal is fed to the four-loudspeaker arrangement. All equalizers, level and delay adjustments are active, the signal is reproduced by all loudspeakers including the subwoofer (see Figure 30). The reproduced sound field is recorded with the terminal's microphone. The measured frequency response at the microphones position shall be within a tolerance of ±3 dB in the frequency range from 50 Hz to 10 kHz.
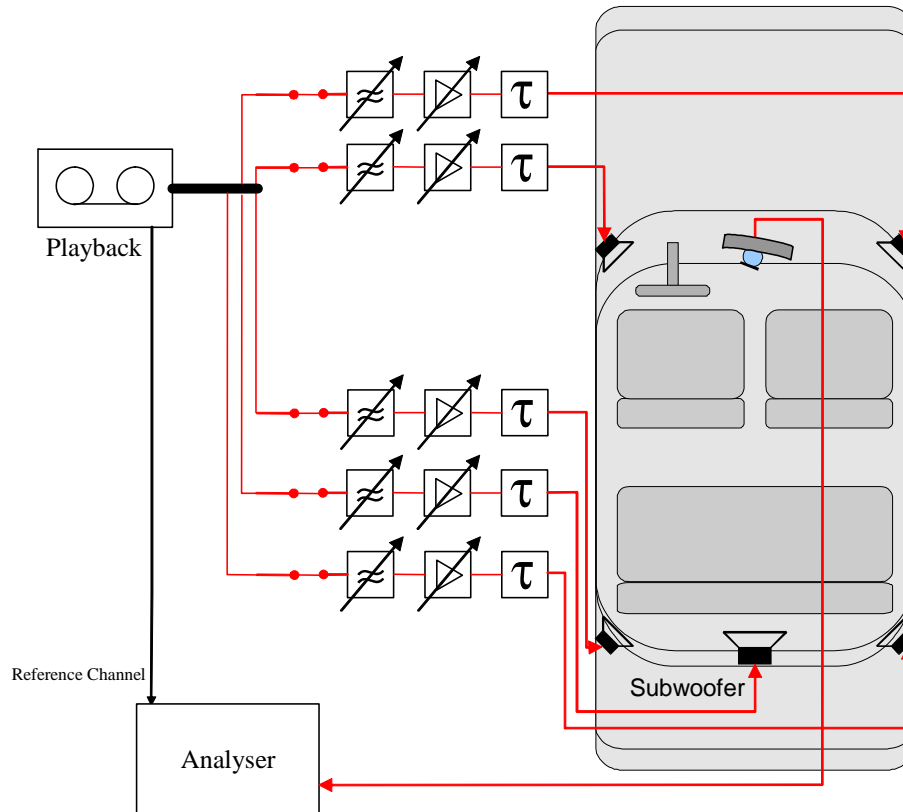
NOTE: Signal paths for overall equalization is marked red.

Figure 30: Block diagram of equalization setup

7.3.8   Troubleshooting failed equalizations

If the result is outside the tolerance, it needs to be checked which of the four loudspeakers is mostly contributing to equalization error. Typically this occurs only in a small frequency range. See also clause 6.3.8 for further information.

7.4   Equalization and Calibration with a pair of cardioid microphones

7.4.1   Overview

When using a pair of cardioid microphones for the equalization care should be taken that their frequency response characteristics is sufficiently flat (±1 dB) within the desired frequency range. If this cannot be achieved by the microphones, additional appropriate equalization may be applied.

A four loudspeaker arrangement as described in clause 7.1 is positioned in a car for which the terminal is to be tested. A complete block diagram of the setup is shown in Figure 31.
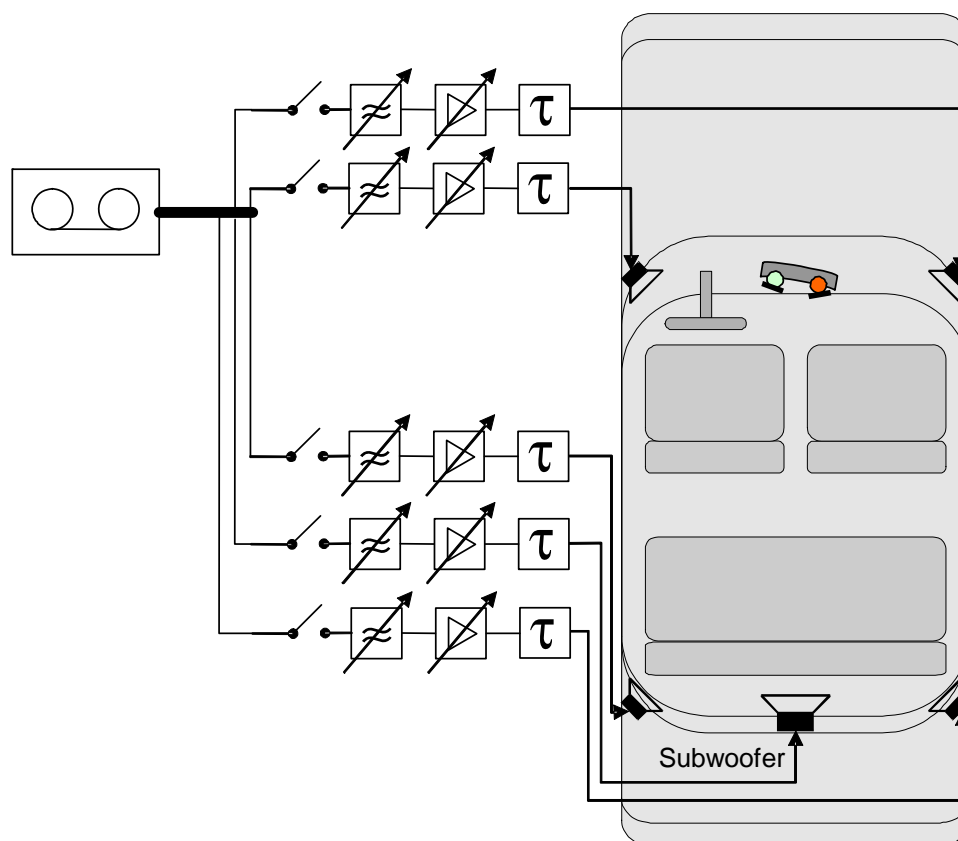
Figure 31: General setup of the equalized four loudspeaker arrangement in a car

For the equalization two cardioid microphones should be located as close as possible to the expected position of the terminal's microphone later used by customers, e.g. under the ceiling or in or on top of the rear view mirror. The left microphone should be directed to the left back seat and the right one to the right back seat.

The equalization is conducted in several steps that are described in the following clauses.

7.4.2 Pair-wise equalization for the left-hand loudspeakers

At first, the levels of the front and rear left loudspeaker should be adjusted to approximately the same level.

For the equalization of the loudspeakers a stationary, realistic broadband noise (e.g. broadband car noise recorded under constant driving conditions) is used. The feeding should be done for a frequency range of f_{min} to 20 000 Hz (for the four loudspeakers only). The subwoofer covers the frequency range below f_{min} (see clause 6.2).

A block diagram of the setup is shown in Figure 32.
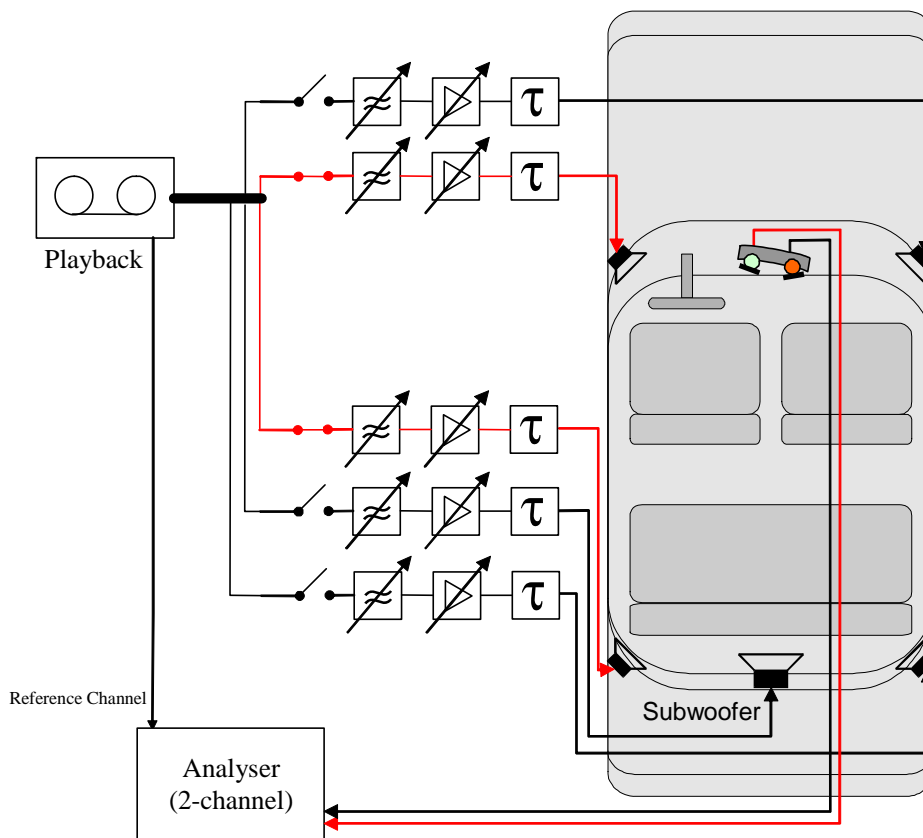
NOTE: Signal paths for equalizing the left-hand loudspeakers in combination are marked red.

Figure 32: Block diagram of equalization setup

The broadband noise signal is filtered with a high-pass with cut-off frequency f_{min} and then synchronously fed to the left-hand loudspeakers for at least 10 s. Synchronously the reproduced signal is recorded with the left cardioid microphone. The power density spectra then are calculated for the recorded microphone signal (S_{LS}) and for the original signal (S_{bbn}). The averaged magnitude of the two power density spectra are subtracted as follows:

$$\left|H_{LS}(f)\right| = \frac{\left|S_{LS}(f)\right|}{\left|S_{bbn}(f)\right|}$$

$H_{LS}(f)$ is analysed in $1/3^{rd}$ octaves. The measured resulting frequency response should be flat between f_{min} and 10 kHz within a tolerance of ±3 dB.

To fulfil this condition, a filter $H_{EQ}(f)$ needs to be designed that compensates the frequency response of the loudspeaker and the air path in the car. Its frequency response is calculated from the initially measured loudspeaker frequency response by using the inverse filter function:

$$\left|H_{EQ}(f)\right| = \frac{1}{\left|H_{LS}(f)\right|}$$

The equalized broadband noise signal is applied at the left-hand loudspeakers. Figure 33 shows an example of the measured loudspeaker frequency response after the equalization process.
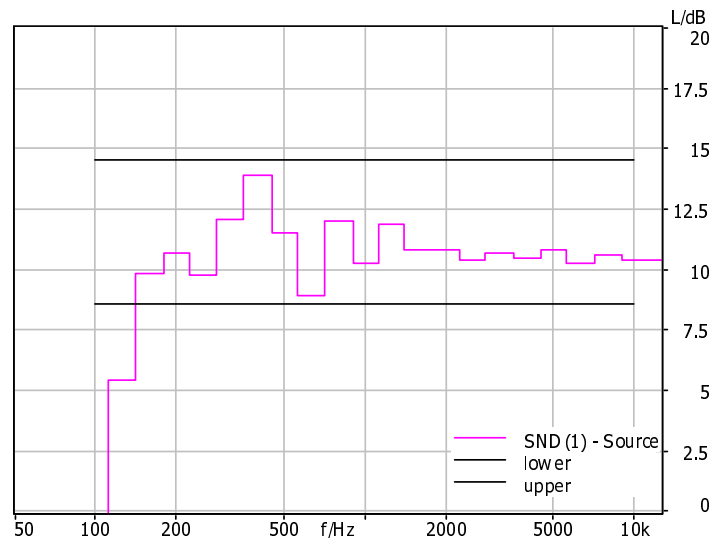
Figure 33: Difference of the original signal and the equalized signal reproduced by the front-left loudspeaker measured with the left cardioid microphone (without subwoofer)

7.4.3　Separate level adjustment for the left-hand loudspeakers

After the left-hand loudspeakers are equalized jointly, the sound pressure level of each loudspeaker shall be adjusted. To reproduce the original level of any sound source with this four-loudspeaker arrangement, the level of each of the four loudspeakers shall be reduced by 6 dB compared to the level of the original sound source due to the fact that the sound source is now reproduced by four loudspeakers.

This level adjustment can be achieved by recording and analysing the reproduced signal for each single loudspeaker as described in the first step and after reducing the level by 6 dB below the original level.

It should also be verified that - while playing back with both left-hand loudspeakers - the level at the left microphone is 3 dB below the original level. If not, the level of the front-left or the rear-left loudspeaker should be re-adjusted.

7.4.4　Pair-wise equalization and separate level adjustment for the right-hand loudspeakers

The pair-wise equalization and the separate level adjustment for the right-hand loudspeakers are applied accordingly to the equalization of the left-hand loudspeakers, as described in the steps of clause 7.4.3.

7.4.5　Equalization and level adjustment for the subwoofer

The equalization for the subwoofer is applied for the frequency range between 30 Hz and f_{min}. A low-pass filter providing a roll of characteristics of at least 18 dB/octave shall be used for the subwoofer channel. Again the realistic broadband noise is used as test signal. Similar to the equalization for the four loudspeakers, the broadband noise is reproduced by the subwoofer and recorded with the right microphone. The difference between the recorded and the original broadband noise spectra defines an equalization filter, which shall provide a flat frequency response for the subwoofer (analysed in $1/3^{rd}$ octaves) within a ±3 dB tolerance mask.

The level is adjusted to the level of the original sound source in this frequency range.

7.4.6　Delay compensation

For the four loudspeaker arrangement including the subwoofer, a compensation of the individual loudspeaker delay to the position of the cardioid microphones is required in order to keep the correlation of the recorded noise signals with respect to the microphone's position in the acoustical field. Wrong delay adjustment might lead to comb filter effects, which leads to unnatural sounding of the background noise signals. Therefore it is necessary to compensate the delays of all loudspeakers to one common delay for all.

The following delays are measured and the corresponding signal paths shall be compensated towards the longest delay (mostly caused by the subwoofer):

- delay from the left loudspeakers to the left microphone (front and rear separately);

- delay from the right loudspeakers to the right microphone (front and rear separately);

- delay from the subwoofer to the right microphone.

7.4.7   Overall equalization

The last step is to verify the equalization process including the delay compensation. Therefore a stationary, realistic broadband background noise signal is fed to the four-loudspeaker arrangement. All equalizers, level and delay adjustments are active, the signal is reproduced by all loudspeakers including the subwoofer (see Figure 34). The reproduced sound field is recorded with both cardioid microphones. For each microphone-channel the recording is spectrally analysed in $1/3^{rd}$ octaves and referenced to the spectrum of the original signal. The measured frequency response for the left ear signal and the right ear signal again shall be within a tolerance of ± 3 dB for each channel in the frequency range from 50 Hz to 10 kHz.
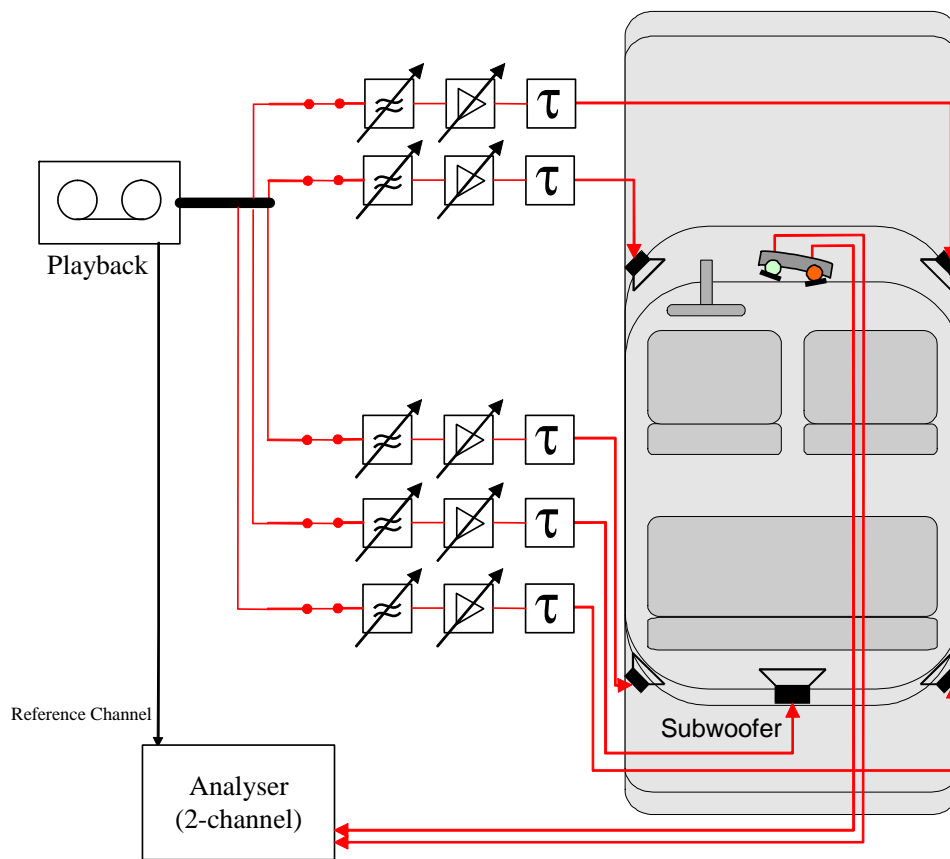
NOTE: Signal paths for overall equalization is marked red.

Figure 34: Block diagram of equalization setup

7.4.8   Troubleshooting failed equalizations

If the result is outside the tolerance it needs to be checked which of the four loudspeakers is mostly contributing to the equalization error. Typically this occurs only in a small frequency range. See also clause 6.3.8 for further information.

7.5   Accuracy of the reproduction arrangement

7.5.1   Comparison between original sound field and simulated sound field

The data reported below were derived from the following experiment:

A pair of cardioid microphones was attached near the rear view mirror of a car. With these microphones recordings were made during a ride on a highway at 130 km/h. In the following those recordings are labelled as "originalCar".

For the simulation, a recording of a speed of 130 km/h was reproduced via the equalized four + one loudspeaker arrangement in the car and again recorded with the two cardioid microphones. In the following the recordings during the simulated sound field are labelled as "simulatedCar".

The comparison of the spectra - recorded with the pair of cardioid microphones - of the original and the simulated sound field is shown in Figures 35 and 36. It can be seen the spectra of the original and the simulation match quite well. The maximum deviation is about 4 dB.
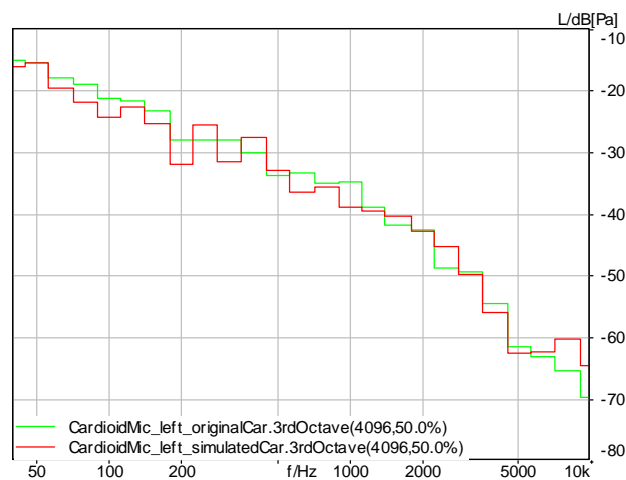
NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 35: Power density spectra measured with the left cardioid microphone
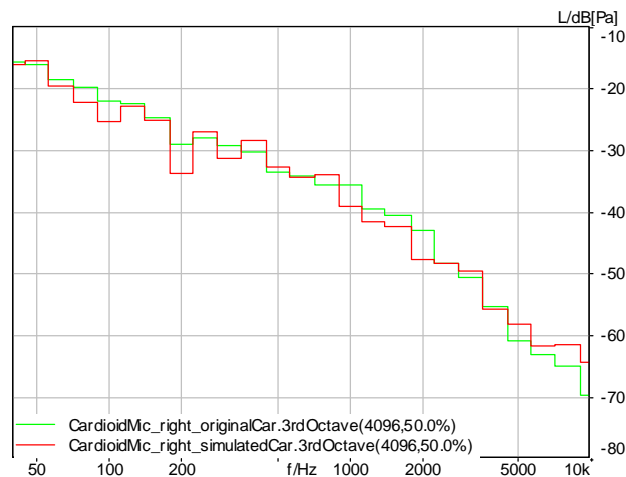
NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 36: Power density spectra measured with the right cardioid microphone

7.5.2   Transmission of background noise: Comparison of terminal performance in the original sound field and the simulated sound field

Further experiments aimed at validating the performance of the four-loudspeaker arrangement with typical terminals applied to the test arrangement. The tests were conducted with two car hands-free terminals and a microphone array with integrated background noise signal processing.

For each terminal, a recording was made during a drive via the terminal's microphone and the terminal itself (labelled as "originalCar") and synchronously with the two cardioid microphones.

For the recording with the terminal in the simulated sound field, the recording of the cardioid microphone during the drive was reproduced via the equalized four+one loudspeaker arrangement in the car and transmitted via the terminal (labelled as "simulatedCar").

NOTE: The equalization was done with a pair of cardioid microphones positioned close to the rear view mirror, but all terminal microphones were attached to the ceiling. Due to the size of the cardioid microphones, the equalization could not be done at the exact hands-free microphone position. This causes differences between the spectra of the original sound field and the simulated sound field.

Car hands-free terminal I

Figure 37 shows the spectra of the original (green curve) and the simulated sound field (red curve) for hands-free terminal I. It can be seen that the curves do not match very well in the frequency range below 1 400 Hz. Further investigations lead to the conclusions that this effect occurs due to the fact that the microphone of this hands-free terminal is very sensitive to structure-borne noise. During driving this structure-borne sound is added to the airborne noise. The structure-borne component is missing in the simulation, since only airborne noise is simulated. Another possible reason for the differences might be the unknown directivity of the microphone. If the directivity contains strong lobes facing towards the back of the car, the terminal might have recorded mainly the noise from the back during the drive.

The results with two modified test setups are shown in Figures 37 and 38. For the transmission with the simulated sound field the sound pressure level of the rear loudspeakers was increased about 5 dB and the sound pressure level of the front loudspeakers was decreased about 3 dB. The result is shown in Figure 37 (blue curve). Compared to the previous result (red curve), the spectrum is still close to the original, but also deviations for frequencies below 350 Hz can be observed.

Figure 38 shows the spectrum of the simulated sound with the modified sound pressure levels (-3 dB front, +5 dB rear) and simulated structure-borne noise. These two actions in combination obviously result in an improvement, the maximum deviation from the original spectrum is approximately 6 dB.
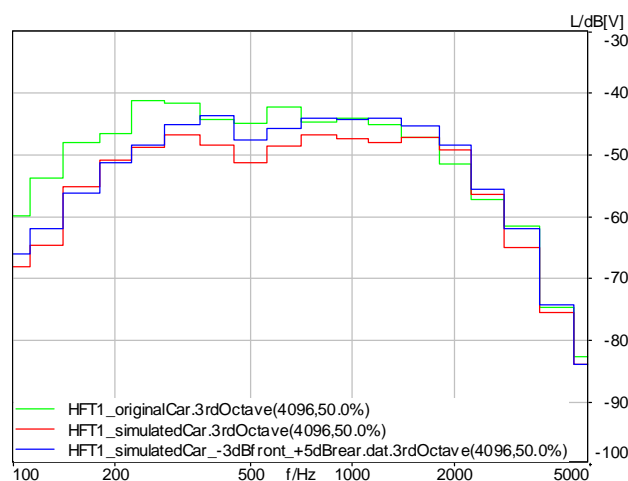
NOTE: Original sound field (green curve), simulated sound field (red curve), simulated sound field with 3 dB reduced sound pressure level at the front loudspeakers and 5 dB increase sound pressure level at the rear loudspeakers (blue curve).

**Figure 37: Power density spectra measured with a common hands-free terminal
(with background noise signal processing) positioned at the ceiling near the rear-view mirror**
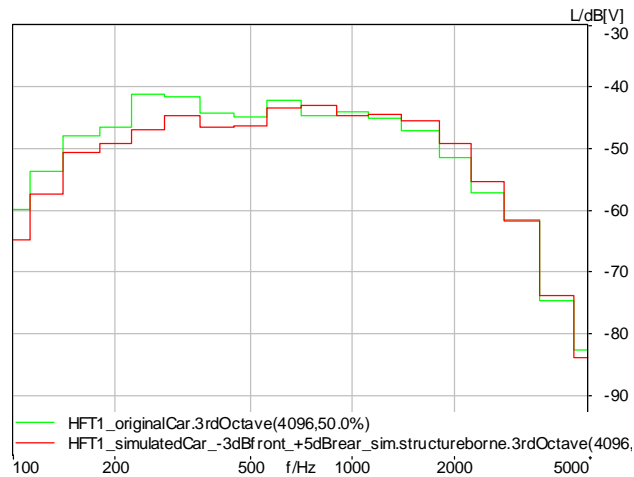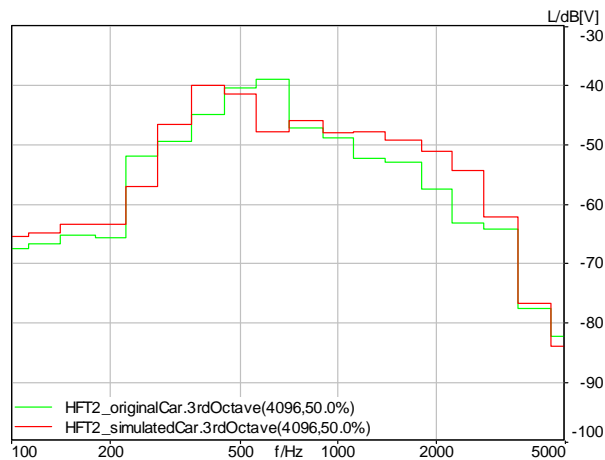
NOTE: Original sound field (green curve), simulated sound field with 3 dB reduced sound pressure level at the front loudspeakers and 5 dB increase sound pressure level at the rear loudspeakers and simulated structure-borne (red curve).

Figure 38: Power density spectra measured with a common hands-free terminal (with background noise signal processing) positioned at the ceiling near the rear-view mirror

Car hands-free terminal II

Figure 39 shows the spectra of the original and the simulated sound field for hands-free terminal II. For the microphone of this terminal, a structure-borne sensitivity was not observed.

It can be seen that there are differences of approximately 9 dB. They are also caused by the equalization to the cardioid microphone's position.

NOTE: Original sound field (green curve), simulated sound field (red curve).

Figure 39: Power density spectra measured with a common hands-free terminal (with background noise signal processing) positioned at the ceiling near the rear-view mirror
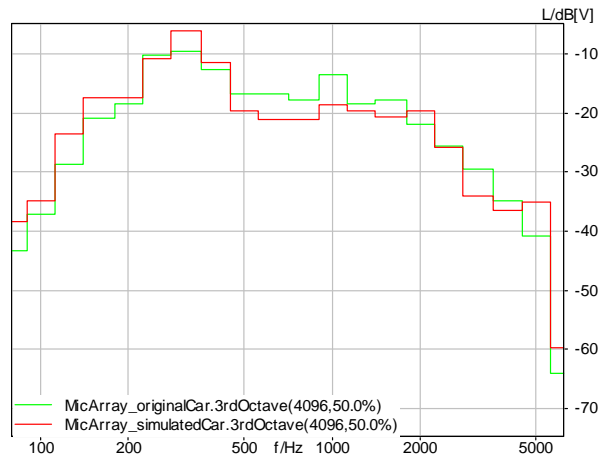
NOTE: Original sound field (green curve), simulated sound field (red curve).

**Figure 40: Power density spectra measured with a microphone array
(with background noise signal processing) positioned at the ceiling near the rear-view mirror**

Microphone array

The microphone array was tested without a terminal. For the test, the single-channel signal provided by the array was used, which includes signal processing that are typical for beamforming microphone arrays.

In Figure 40 the spectra of the original and the simulated sound field transmitted with the microphone array are shown. The maximum deviation from the original spectrum is about 5 dB.

7.6 Automated equalization and calibration procedure

The steps described in clauses 7.3 and 7.4 may in some cases require some manual tuning of the equalization filters to obtain the required flat frequency response within the ±3 dB tolerance (for each channel, in the frequency range from 50 Hz to 10 kHz). Thus, an equalization repeated in the same room and with same test equipment could lead to different sets of filters, which depend on e.g. subjective preferences or experience of the test operator. To avoid such variability caused by manual tuning of the noise simulation, an alternative and fully automated equalization procedure for binaural recordings is described in clause 7.6 of ETSI TS 103 224 [3]. It is considered equivalent, i.e. can alternatively be used as a replacement that complies with the equalization procedure described in clauses 7.3 and 7.4 of the present document.

8 Background Noise Database

8.0 Introduction

The background noise database is available separately from the standard. All files are two-channel wave files at 48 kHz sampling rate. If not specified otherwise, the background noise recordings include ID equalization (see clause 5.2).

The background noises, the file names and the levels of the original signals are found in the following lists below. The information about the levels can be used to calibrate the wave files to the reproduction arrangement used.

The files have been truncated to a maximum length of 30 s each. If longer background noise presentation is needed, the files can be played back periodically using smooth crossover fading.

NOTE: These files can be found at:
http://portal.etsi.org/docbox/stq/Open/EG%20202%20396-1%20Background%20noise%20database.

8.1   Binaural signals

Inside Car Noise, Fullsize Car1

Description	File name	Duration	Level	Type
Recording at the drivers position	Fullsize_Car1_80Kmh_binaural	30 s	L: 63,4 dB(A) R: 61,8 dB(A)	binaural
Recording at the drivers position	Fullsize_Car1_100Kmh_binaural	30 s	L: 65,4 dB(A) R: 63,2 dB(A)	binaural
Recording at the drivers position	Fullsize_Car1_130Kmh_binaural	30 s	L: 69,1 dB(A) R: 68,1 dB(A)	binaural

Inside Car Noise, Midsize_Car1

Description	File name	Duration	Level	Type
Recording at the drivers position	Midsize_Car1_80Kmh_binaural	30 s	L: 62,3 dB(A) R: 61,4 dB(A)	binaural
Recording at the drivers position	Midsize_Car1_100Kmh_binaural	30 s	L: 63,4 dB(A) R: 62,6 dB(A)	binaural
Recording at the drivers position	Midsize_Car1_130Kmh_binaural	30 s	L: 67,0 dB(A) R: 65,9 dB(A)	binaural

Inside Car Noise, Midsize_Car2

Description	File name	Duration	Level	Type
Recording at the drivers position	Midsize_Car2_Car_80Kmh_binaural	30 s	L: 68,3 dB(A) R: 67,8 dB(A)	binaural
Recording at the drivers position	Midsize_Car2_100Kmh_binaural	30 s	L: 70,6 dB(A) R: 70,4 dB(A)	binaural
Recording at the drivers position	Midsize_Car2_130Kmh_binaural	30 s	L: 74,4 dB(A) R: 74,5 dB(A)	binaural

Inside Bus Noise

Description	File name	Duration	Level	Type
Recording in passenger cabin	Inside_Bus_Noise_binaural	30 s	L: 75,8 dB(A) R: 77,6 dB(A)	binaural

Inside Train Noise

Description	File name	Duration	Level	Type
Recording in passenger cabin	Inside_Train_Noise1_binaural	30 s	L: 73,0 dB(A) R: 72,9 dB(A)	binaural
Recording in passenger cabin	Inside_Train_Noise2_binaural	30 s	L: 62,1 dB(A) R: 64,8 dB(A)	binaural
Recording in passenger cabin	Inside_Train_Noise3_binaural	30 s	L: 79,17 dB(A) R: 80,52 dB(A)	binaural

Inside Airplane Noise

Description	File name	Duration	Level	Type
Recording in passenger cabin	Inside_Aircraft_Speech_binaural	30 s	L: 80,4 dB(A) R: 79,9 dB(A)	binaural
Recording in passenger cabin	Inside_Aircraft_Noise1_binaural	30 s	L: 74,6 dB(A) R: 75,1 dB(A)	binaural
Recording in passenger cabin	Inside_Aircraft_Noise2_binaural	30 s	L: 81,4 dB(A) R: 81,1 dB(A)	binaural

Outside Traffic Street Noise

Description	File name	Duration	Level	Type
Recording at pavement	Outside_Traffic_Crossroads_binaural	20 s	L: 69,1 dB(A) R: 69,6 dB(A)	binaural
Recording at pavement	Outside_Traffic_Road_binaural	30 s	L: 74,9 dB(A) R: 73,9 dB(A)	binaural

Public Places Noise

Description	File name	Duration	Level	Type
Recording in pub	Pub_Noise_binaural_V2	30 s	L: 75,0 dB(A) R: 73,0 dB(A)	binaural
Recording in a cafeteria	Mensa_binaural	22 s	L: 63,4 dB(A) R: 61,9 dB(A)	binaural
Recording while walking	Shopping_Center_binaural	30 s	L: 68,3 dB(A) R: 67,8 dB(A)	binaural
Recording in gym at marginal pos.	Sports_Noise1_Indoor_Soccer_binaural	30 s	L: 68,0 dB(A) R: 68,7 dB(A)	binaural
Recording in gym at marginal pos.	Sports_Noise2_Badminton_binaural	30 s	L: 68,0 dB(A) R: 68,7 dB(A)	binaural
Recording at departure platform	Train_Station_binaural	30 s	L: 68,2 dB(A) R: 69,8 dB(A)	binaural
Recording at sales counter	Cafeteria_Noise_binaural	30 s	L: 68,4 dB(A) R: 67,3 dB(A)	binaural

Workplace Noise

Description	File name	Duration	Level	Type
Recording in business office	Work_Noise_Office_Callcener_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	binaural
Recording besides road works	Work_Noise_Jackhammer_binaural	30 s	L: 76,2 dB(A) R: 78,4 dB(A)	binaural
Recording in machine shop	Work_Noise_Machine_Shop_binaural	30 s	L: 78,1 dB(A) R: 78,8 dB(A)	binaural

Living Room Noise

Description	File name	Duration	Level	Type
Recording in private living room	Living_Room_Noise_binaural	30 s	L: 56,6 dB(A) R: 57,8 dB(A)	binaural

Children Noise

Description	File name	Duration	Level	Type
Recording in playroom	Kindergarten_Noise1_binaural	30 s	L: 73,20 dB(A) R: 71,94 dB(A)	binaural
Recording in playroom	Kindergarten_Noise2_binaural	30 s	L: 80,59 dB(A) R: 80,38 dB(A)	binaural
Recording besides schoolyard	Schoolyard_Noise2_binaural	30 s	L: 67,89 dB(A) R: 68,07 dB(A)	binaural

Nature Noise

Description	File name	Duration	Level	Type
Recording in the forest	Nature1_Forest_Noise_binaural	30 s	L: 41,85 dB(A) R: 41,24 dB(A)	binaural
Recording in the forest	Nature2_Creek_Noise_binaural	30 s	L: 65,93 dB(A) R: 56,51 dB(A)	binaural
Recording in open cornfield	Nature3_Open_Field_Noise_binaural	30 s	L: 42,63 dB(A) R: 42,90 dB(A)	binaural

Single Voice Distractor, Male

Description	File name	Duration	Level	Type
Recording in a quiet room	Male_Single_Voice_Distractor_binaural	36 s	L: 68,4 dB(A) R: 68,2 dB(A)	binaural

Conference Noise

Description	File name	Duration	Level	Type
HATS and microphone arrays at a table in a conference room with stationary noises	Conference1	70 s	L: 44,08 dB(A) R: 45,87 dB(A)	binaural
HATS and microphone arrays at a table in a conference room with stationary and transient noises	Conference2	70 s	L: 45,40 dB(A) R: 47,00 dB(A)	binaural
HATS and microphone arrays at a table in a conference room with stationary and transient noises	Conference3	70 s	L: 49,50 dB(A) R: 50,80 dB(A)	binaural

Music

Description	File name	Duration	Level	Type
Recording of music in listening room	RockMusic01m48k_ETSI_3m18s_1s Ramp.wav	198 s	L: 73,9 dB(A) R: 73,8 dB(A)	binaural
Recording of orchestra warm up	Orchestra_noise_binaural.wav	149 s	L: 74.1 dB(A) R: 73.5 dB(A)	binaural

8.2   Binaural signals identical to the background noise recordings provided in ETSI TS 103 224

Name	Description	Length	Binaural Levels (Left)	Binaural Levels (Right)
Inside Car Noise				
Mid-size car 80 km/h (*MidSizeCar_80*)	HATS and microphone array at co-drivers position	30 s	66,8 dB(A)	67,7 dB(A)
Mid-size car 100 km/h (*MidSizeCar_100*)	HATS and microphone array at co-drivers position	30 s	66,0 dB(A)	66,7 dB(A)
Mid-size car 130 km/h (*MidSizeCar_130*)	HATS and microphone array at co-drivers position	30 s	70,5 dB(A)	71,5 dB(A)
Full-size car 80 km/h (*FullSizeCar_80*)	HATS and microphone array at co-drivers position	30 s	60,9 dB(A)	62,9 dB(A)
Full-size car 100 km/h (*FullSizeCar_100*)	HATS and microphone array at co-drivers position	30 s	63,8 dB(A)	66,0 dB(A)
Full-size car 130 km/h (*FullSizeCar_130*)	HATS and microphone array at co-drivers position	30 s	68,7 dB(A)	70,7 dB(A)
Inside Train Noise				
Inside Train (*Inside_Train*)	HATS and microphone array in passenger cabin of a train	30 s	68,2 dB(A)	68,4 dB(A)
Inside Bus Noise				
Inside Bus (*Inside_Bus*)	HATS and microphone array in passenger cabin of a bus	30 s	70,9 dB(A)	73,4 dB(A)
Outside Traffic Street Noise				
Roadnoise (*Roadnoise*)	HATS and microphone array standing outside near a road	30 s	74,0 dB(A)	74,1 dB(A)
Crossroadnoise (*Crossroadnoise*)	HATS and microphone array standing outside near a crossroad	30 s	70,8 dB(A)	71,6 dB(A)
Public Places Noise				
Cafeteria (*Cafeteria*)	HATS and microphone array inside a cafeteria	30 s	69,8 dB(A)	70,3 dB(A)
Departure platform (*TrainStation*)	HATS and microphone array on the departure platform of a train station	30 s	78,1 dB(A)	78,8 dB(A)
Pub Noise (*Pub*)	HATS and microphone array in a pub	30 s	77,7 dB(A)	76,2 dB(A)

Name	Description	Length	Binaural Levels (Left)	Binaural Levels (Right)
Sales Counter (*SalesCounter*)	HATS and microphone array in a supermarket	30 s	66,7 dB(A)	66,6 dB(A)
Workplace Noise				
Callcenter 1 (*Callcenter*)	HATS and microphone array in business office	30 s	72,7 dB(A)	70,2 dB(A)
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	60,2 dB(A)	60,0 dB(A)
Office room	Recording of an office room scenario, including projector, writing by hand and keyboard, phone ringing, phone call, outside noise	85 s	55,4 dB(A)	58,0 dB(A)
Home Environment Noise				
Bathroom	Recording of a bathroom scenario, including shower, razor, sink, toilet flushing, hairdryer	85 s	72,9 dB(A)	73,1 dB(A)
Bathroom with music	Same as "bathroom", but with additional playback of radio broadcast	85 s	73,8 dB(A)	74,2 dB(A)
Kitchen	Recording of a kitchen scenario, including range hood, frying, tableware rattle, mixer, sink, knife on cutting board	85 s	67,1 dB(A)	65,8 dB(A)
Living room	Recording of a living room scenario, including vacuum cleaner, clink of drinking glass, coughing, TV, cleaning up	85 s	64,1 dB(A)	63,4 dB(A)

8.3   Stereophonic signals

Inside Car Noise, Fullsize Car1				
Description	**File name**	**Duration**	**Level**	**Type**
Recording at the mirror position	Fullsize_Car1_80Kmh_2cardiod_mics	30 s	L: 61,5 dB(A) R: 61,1 dB(A)	2 cardiod microphones
Recording at the mirror position	Fullsize_Car1_100Kmh_2cardiod_mics	30 s	L: 62,7 dB(A) R: 62,4 dB(A)	2 cardiod microphones
Recording at the mirror position	Fullsize_Car1_130Kmh_2cardiod_mics	30 s	L: 68,4 dB(A) R: 67,4 dB(A)	2 cardiod microphones

Inside Car Noise, Midsize_Car1				
Description	**File name**	**Duration**	**Level**	**Type**
Recording at the mirror position	Midsize_Car1_80Kmh_2cardiod_mics	30 s	L: 60,3 dB(A) R: 58,8 dB(A)	2 cardiod microphones
Recording at the mirror position	Midsize_Car1_100Kmh_2cardiod_mics	30 s	L: 61,9 dB(A) R: 60,2 dB(A)	2 cardiod microphones
Recording at the mirror position	Midsize_Car1_130Kmh_2cardiod_mics	30 s	L: 65,9 dB(A) R: 64,4 dB(A)	2 cardiod microphones

Inside Car Noise, Midsize_Car2				
Description	**File name**	**Duration**	**Level**	**Type**
Recording at the mirror position	Midsize_Car2_80Kmh_2cardiod_mics	30 s	L: 68,3 dB(A) R: 67,8 dB(A)	2 cardiod microphones
Recording at the mirror position	Midsize_Car2_100Kmh_2cardiod_mics	30 s	L: 70,6 dB(A) R: 70,4 dB(A)	2 cardiod microphones
Recording at the mirror position	Midsize_Car2_130Kmh_2cardiod_mics	30 s	L: 74,4 dB(A) R: 74,5 dB(A)	2 cardiod microphones

Annex A (informative):
Comparison of Tests in Sending Direction and D-Values Conducted in Different Rooms

A.0   Introduction

Annex A describes the results of comparison test conducted in an office room, a semianechoic chamber and an anechoic chamber based on clauses 5 and 6 of the present document.

A.1   Test Setup

All tests reported below were made with different mobile phones. Mobile phones were used since they show a big variety of different designs and most of them use integrated noise cancellation techniques. So both, the influence of the acoustical/mechanical design as well as the influence of noise cancellation techniques could be evaluated.

The following types of phones from different manufacturers were used:

Phone 1:	Flip Phone with Noise Cancellation
Phone 2:	Short Phone with Noise Cancellation
Phone 3:	Short Phone with moderate Noise Cancellation
Phone 4:	Large Size Phone with Directional Microphone, no Noise Cancellation
Phone 5:	Medium Size Phone with strong Noise Cancellation
Phone 6:	Medium Size classical "Business Phone" with Noise Cancellation
Phone 7:	Short Phone with Noise Cancellation
Phone 8:	Flip Phone with Noise Cancellation

All tests were conducted using a HATS conforming to Recommendation ITU-T P.58 [2] equipped with an artificial ear type 3.4 (Recommendation ITU-T P.57 [1]). The phones were positioned as described in Recommendation ITU-T P.64 [i.10].

The tests were conducted as described in ETSI TS 151 010-1 [i.18]. Two types of background noises were used: In one experiment a diffuse pink noise field was generated using 4 uncorrelated noise sources presented over 4 loudspeakers equalized and calibrated to the different test rooms.

The background noise simulation in the office type room was equalized and calibrated as described in clauses 5 and 6 of of the present document. When installing the background noise simulation in the anechoic and the semianechoic room the same setup was used. However, instead of time aligning the 4 loudspeakers in each path of the playback arrangement a delay was introduced in order to increase the diffusivity of the binaurally recorded background noise from the Database in clause 8 of the present document and to avoid standing waves in the anechoic test rooms. The delays inserted in each path were 0 ms, 11 ms, 17 ms, 29 ms. After inserting the delays the equalization and calibration procedure as described in clauses 5 and 6 of the present document was applied. For the tests the cafeteria noise from clause 8 of the present document database was used.

All tests were conducted in the following rooms:

Room R1:	fully anechoic chamber
Room R2:	office type room
Room R3:	semianechoic room (anechoic above 300 Hz)

A.2   Results of the Tests

A.2.0   Introduction

All measured responses which are the basis for the different calculations shown below are found in the Annex B.

A.2.1   Sending Frequency Response Characteristics and SLR

In a first step the frequency response characteristics in sending was measured for all phones in the different rooms in order to find the general differences which can be expected when conducting these most simple and reproducible tests (from the setup point of view). Furthermore the SLRs were determined.

Table A.1: SLRs measured for the different phones in the different rooms

SLR Phone Nr.	1	2	3	4	5	6	7	8
R1 (fully anechoic)	12,7	10,0	11,6	13,3	8,5	5,2	10,9	13,6
R2 (office type)	12,1	9,1	11,4	11,8	7,0	4,9	9,6	13,1
R3 (semi anechoic)	12,9	12,4	10,0	11,6	9,1	5,1	10,1	13,7

From the measured frequency responses as well as from the SLR results it can be seen that, although positioning and calibration was made very carefully, differences can be expected. This is due to limited positioning accuracy as well as due to the fact that most of the phones use non-linear and time variant signal processing. This signal processing may react on slightly different signal levels (second order effect of slightly different processing). Furthermore there may be memory effects present which lead to different results when measuring the phone at different times during a connection. Table A.1 shows that even SLRs may deviate up to about 3 dB depending on the type of phone.

A.2.2   D-Value with Pink Noise

In a second experiment the "traditional" background noise simulation setup consisting of 4 loudspeakers producing an uncorrelated noise field was used in order to determine the D-value of the phones. The results are shown in Table A.2.

Table A.2: D-value with pink noise measured for the different phones in the different rooms

SLR Phone Nr.	1	2	3	4	5	6	7	8
R1 (fully anechoic)	13,7	5,9	3,3	-0,4	-3,0	0,4	0,5	10,7
R2 (office type)	13,2	5,5	3,3	-0,6	-12,9	1,5	-0,1	8,4
R3 (semi anechoic)	14,1	4,6	0,6	-1,4	-5,6	-2,9	-1,6	9,3

As seen already for the frequency response tests the differences measured for the D-value are in the range of up to 4 dB except mobile phone no. 5 which deserves some more detailed discussion. Mobile phone no. 5 shows a behaviour which can be found more and more in modern designs. Depending on type, level and duration of a background noise signal the phone changes its behaviour in sending drastically. Certainly this is mostly due to the simple background noise signal which can be identified easily by most phones and typically leads to a stronger background noise cancellation than realistic background noises. As a consequence the results measured may differ quite a bit depending on the test condition. The behaviour of this phone shows that the test conditions including the preconditioning for the tests need to be defined more thorough than in the past. Furthermore it can be stated that all different D-values may occur in real life depending on the use condition.

A.2.3   D-Value with Cafeteria Noise

In the third experiment the simulation of a cafeteria noise taken from the ETSI background noise database was made. The setup used in the office type room was according to clause 8 of the present document. In the anechoic room as well as in the semianechoic room the modified setup as described above was used. Again, the measured response characteristics which are the basis for the D-value calculation are found in the Annex B.

Table A.3: D-value with cafeteria noise measured for the different phones in the different rooms

SLR Phone Nr.	1	2	3	4	5	6	7	8
R1 (fully anechoic)	-1,8	-10,4	-9,6	-3,4	-10,3	-9,3	-9,2	2,7
R2 (office type)	-1,8	-10,4	-9,7	-3,9	-12,9	-10,1	-9,5	0,0
R3 (semi anechoic)	-3,0	-12,2	-9,9	-6,4	-14,7	-11,7	-10,5	-0,1

The differences in the results are in a similar range as already observed before. However, the big differences observed for mobile phone no. 5 were not observed with cafeteria noise. The results document, that the setup in the present document, when used in anechoic or semianechoic rooms lead to similar results as found when using the original setup in office type rooms. The differences observed are in the same range which can be expected when measuring the same phones different times. The measured diffuse field sensitivity with cafeteria is typically even closer between the different rooms than the one measured with pink noise. Sometimes the different D-value is mainly caused by a different direct sound sensitivity.

A.3   Conclusions

The experiments conducted lead to the following conclusions:

- The repeatability of mobile phone tests in sending direction is a bit more critical than expected, frequency response characteristics as well as SLR may differ up to 3 dB.

- The main reason for the differences observed is the highly non-linear and time variant signal processing techniques used in modern phones. Slight positioning inaccuracies may further contribute to these differences (sometimes this is just a second order effect due to a slightly different signal level of the test signal at the mobile phone's microphone).

- The D-value differences measured are in a similar range as found for the SLRs.

- The pink noise setup may lead to higher differences than the use of a realistic cafeteria background noise.

- The use of the setup described in the present document in different rooms provides very comparable results for the different phones tested.

Annex B (informative):
Graphs of test results in Annex A

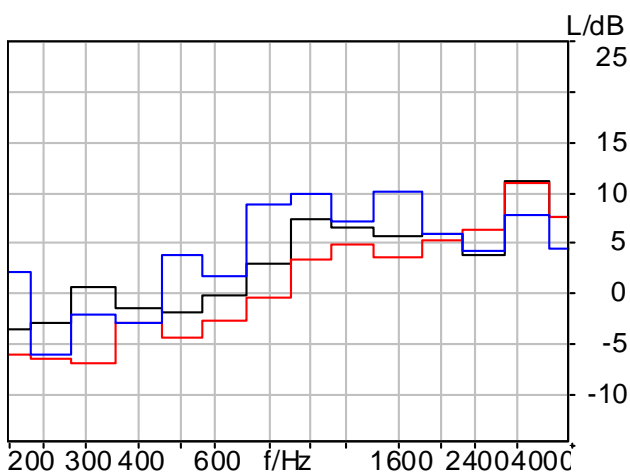**Sending Frequency Response**
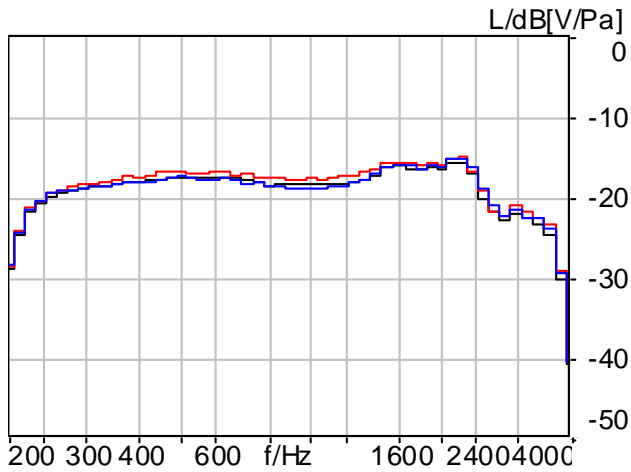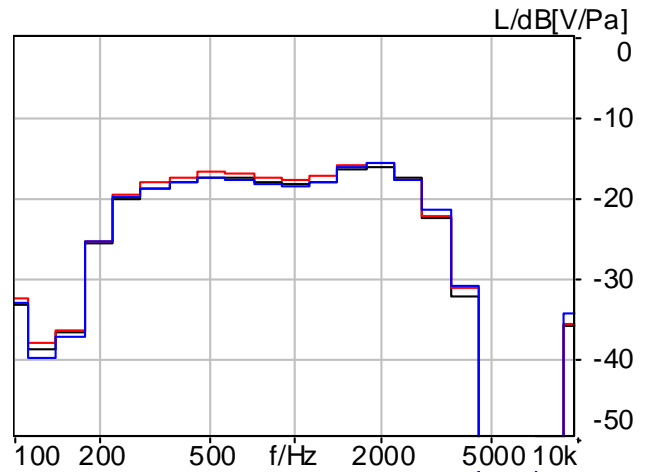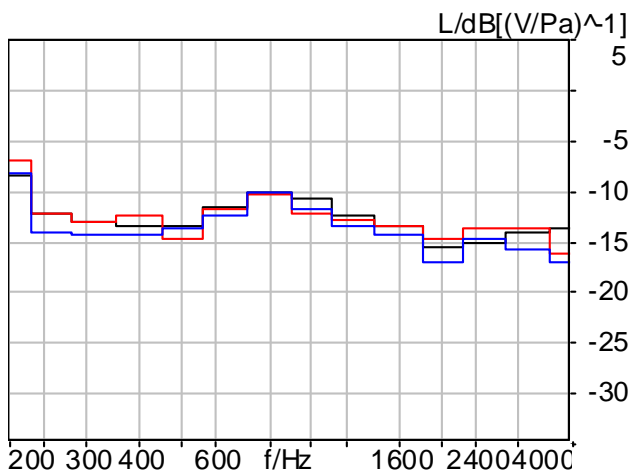
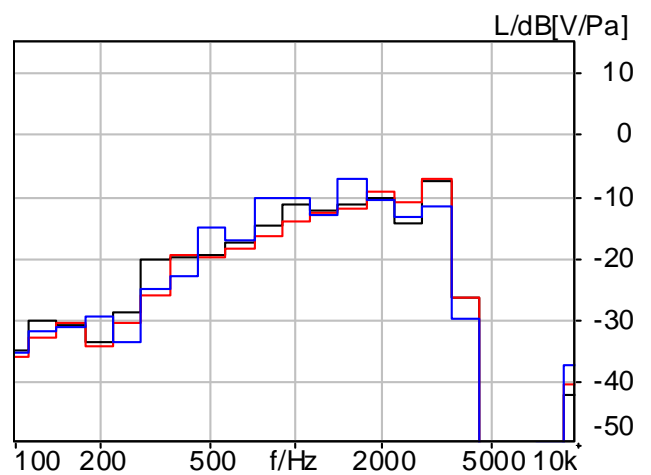**Direct Sound Sensitivity**

D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise
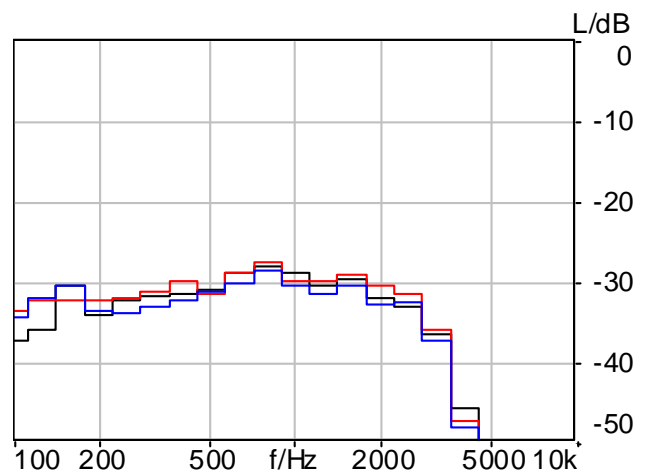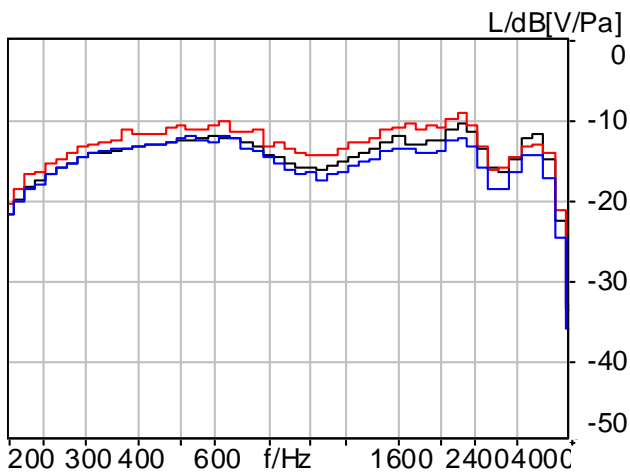
D-Value, Cafeteria Noise
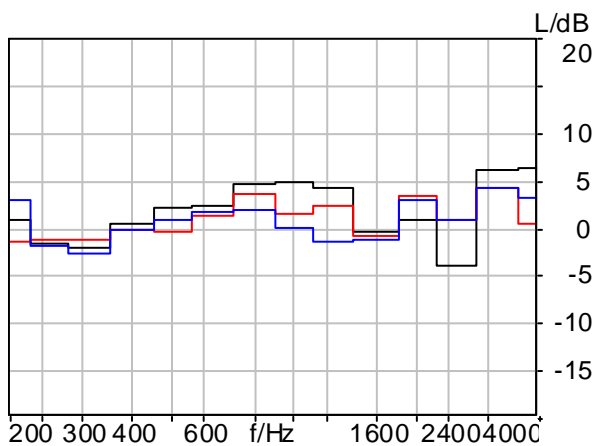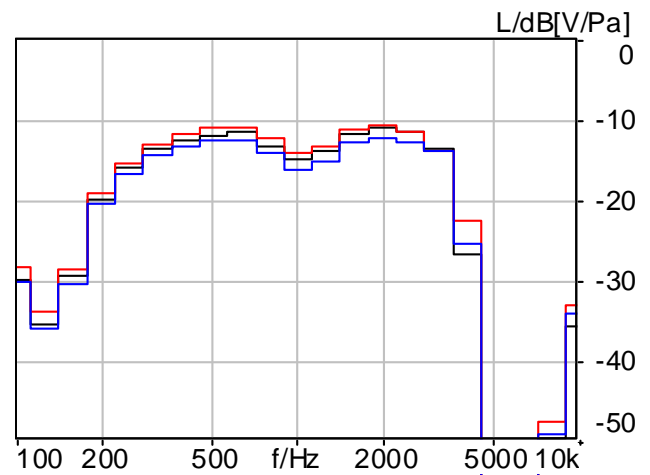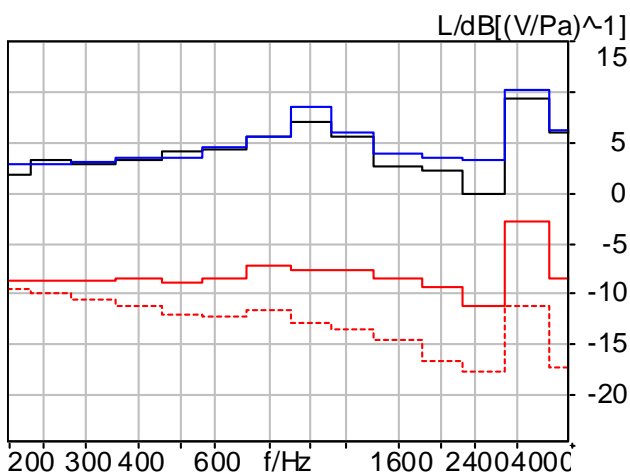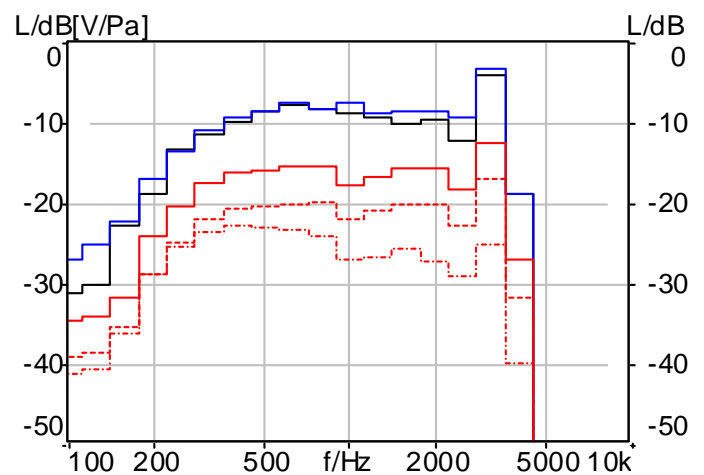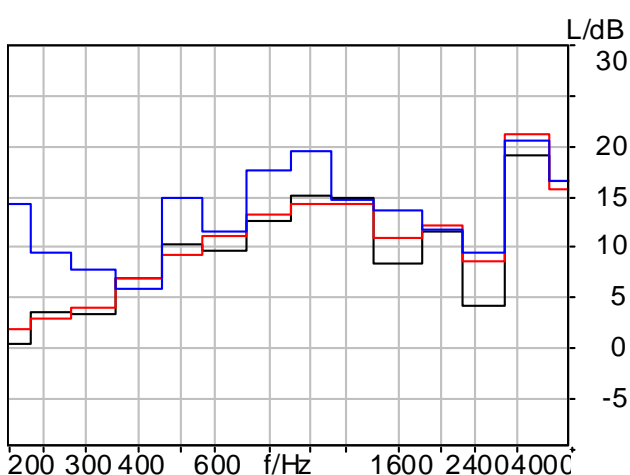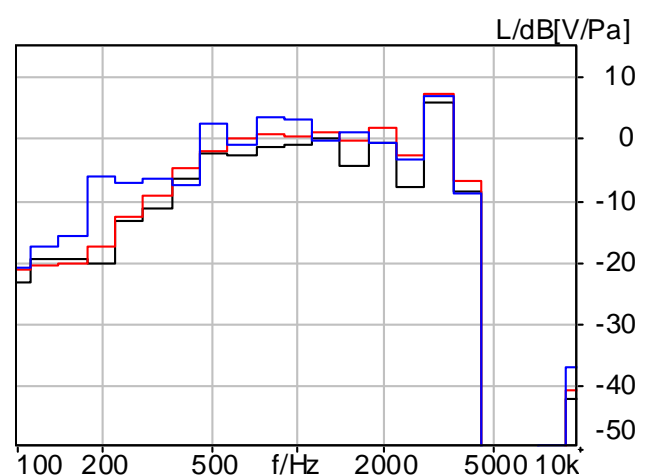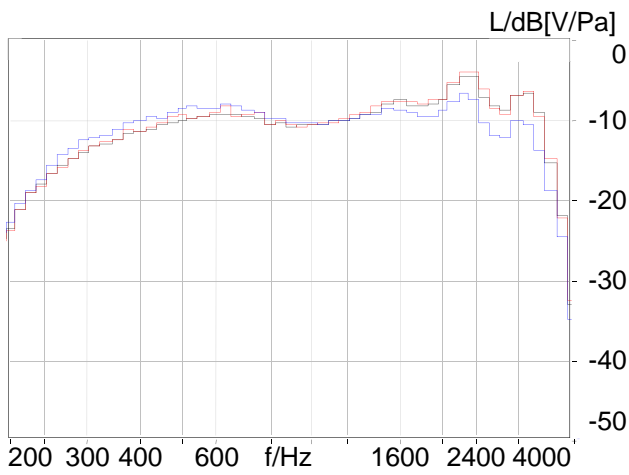
Diffuse Sound Sensitivity, Cafeteria Noise
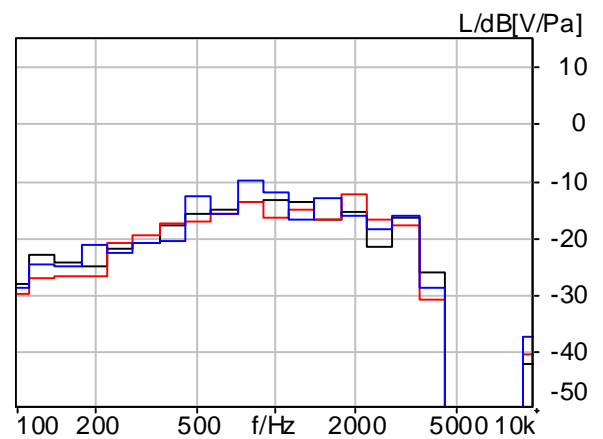
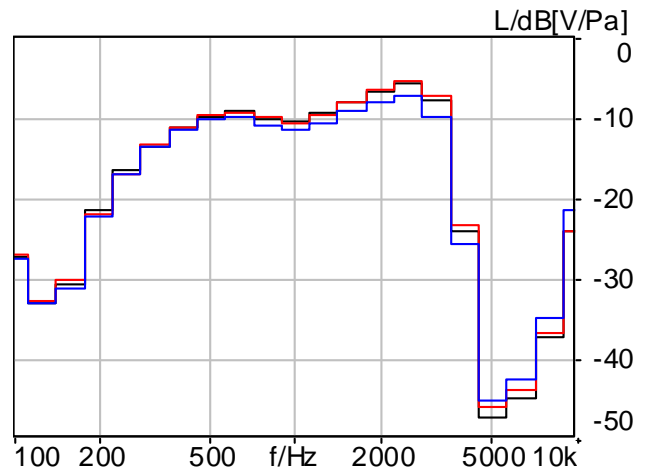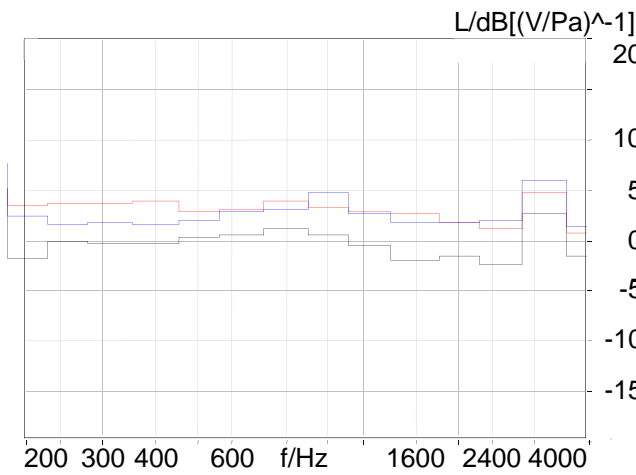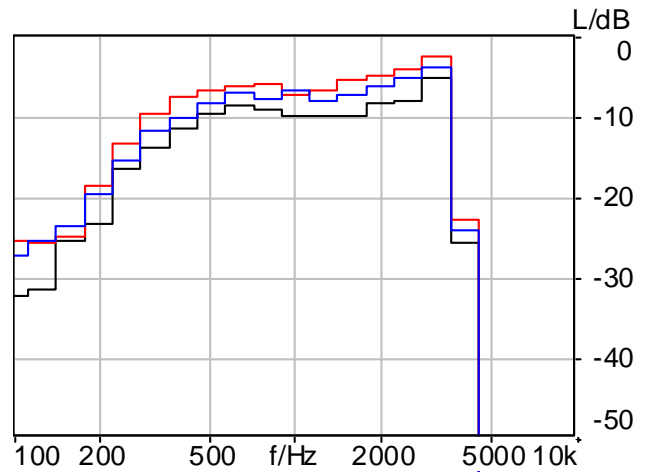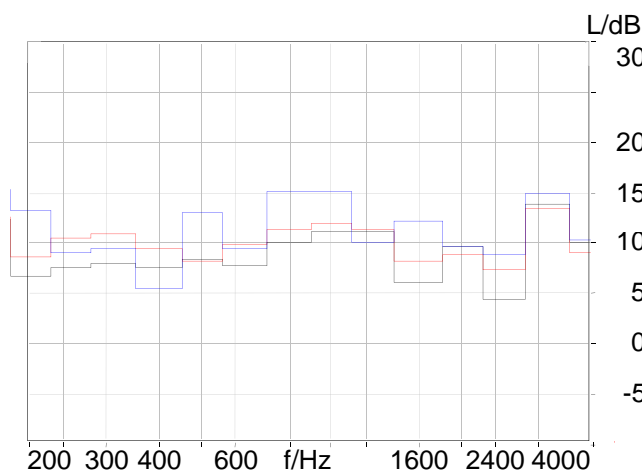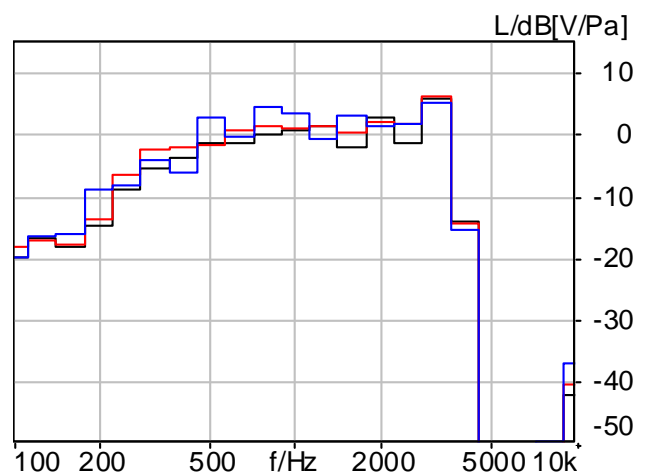**Figure B.1: Measurements of phone 1: black - fully anechoic;
red - office type room; blue - semi anechoic**
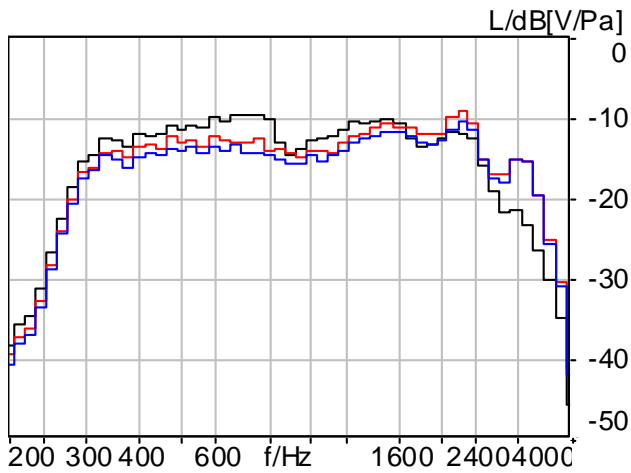
Sending Frequency Response

Direct Sound Sensitivity
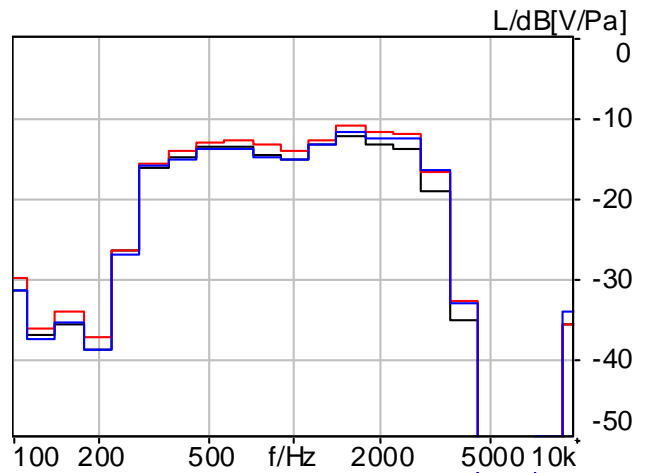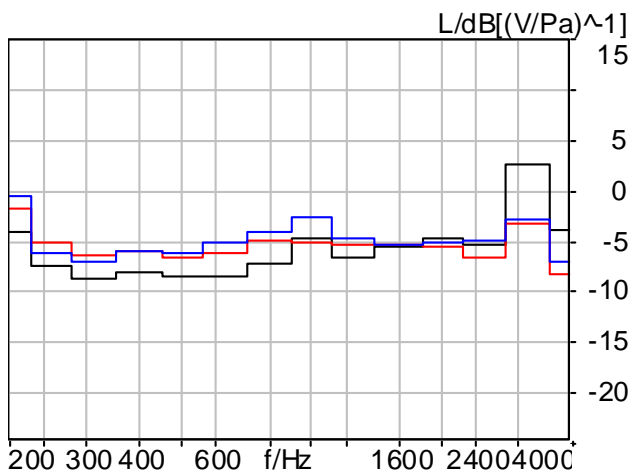
D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise
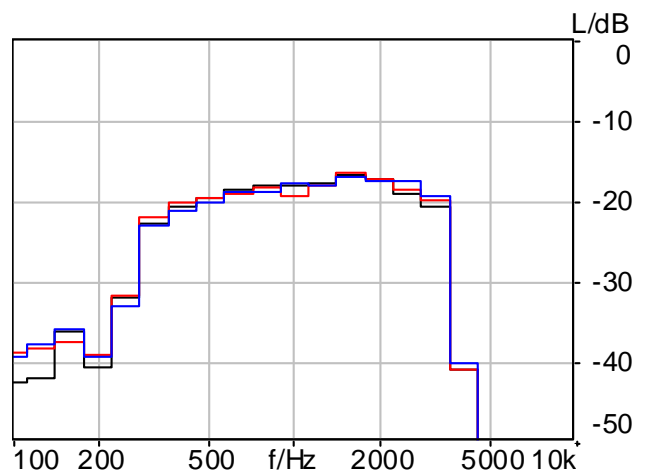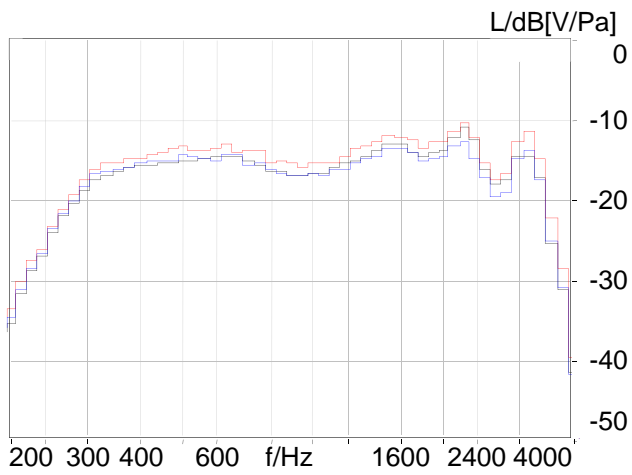
D-Value, Cafeteria Noise
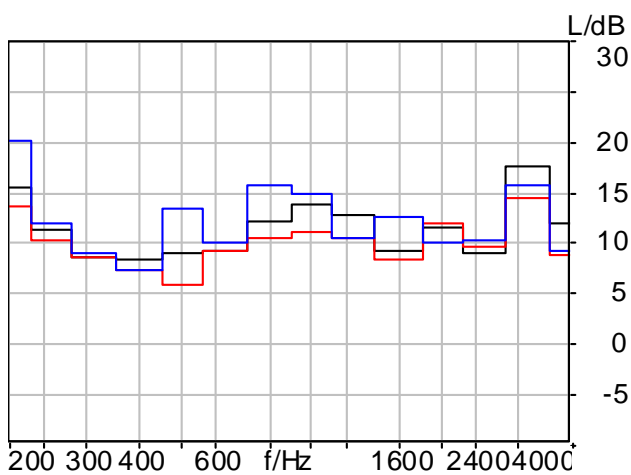
Diffuse Sound Sensitivity, Cafeteria Noise
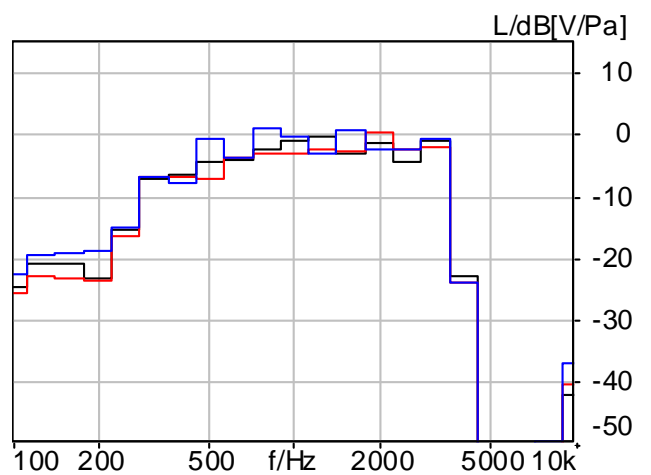
**Figure B.2: Measurements of phone 2: black - fully anechoic;
red - office type room; blue - semi anechoic**
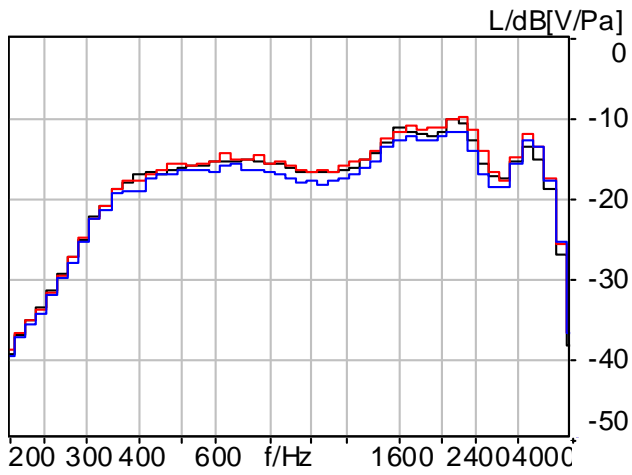
Sending Frequency Response

Direct Sound Sensitivity
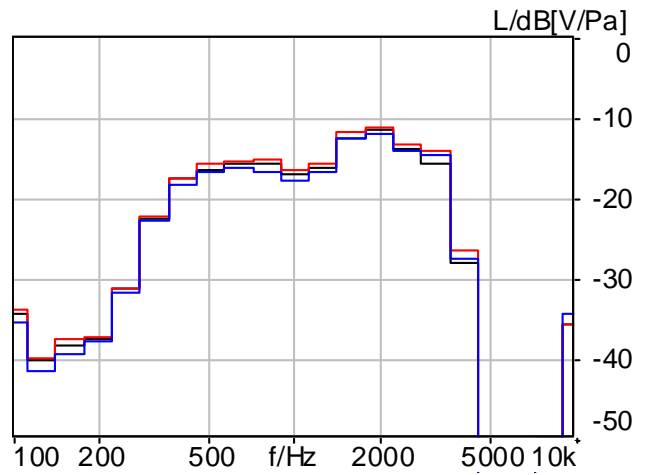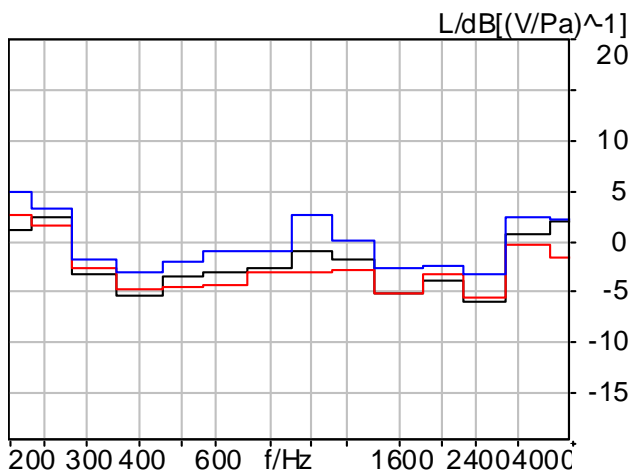
D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise
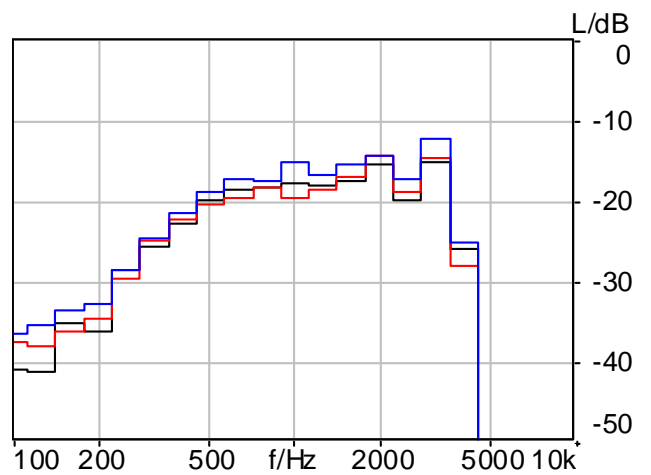
D-Value, Cafeteria Noise
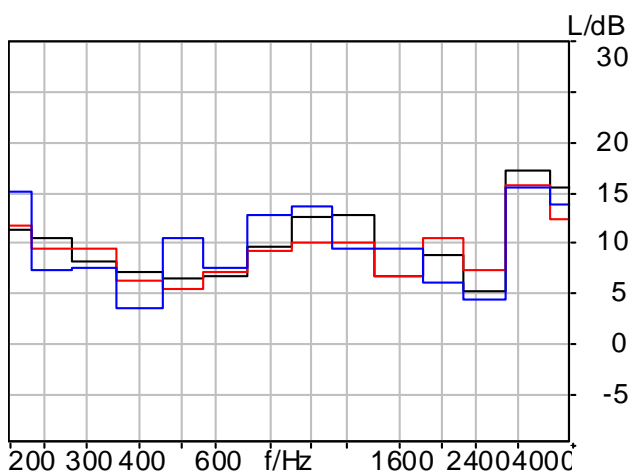
Diffuse Sound Sensitivity, Cafeteria Noise
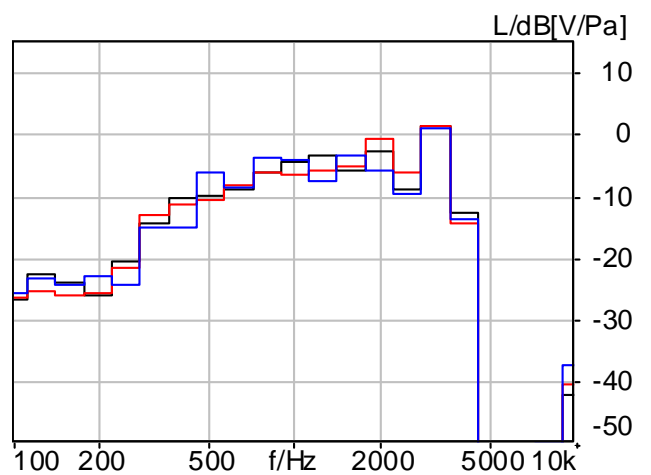
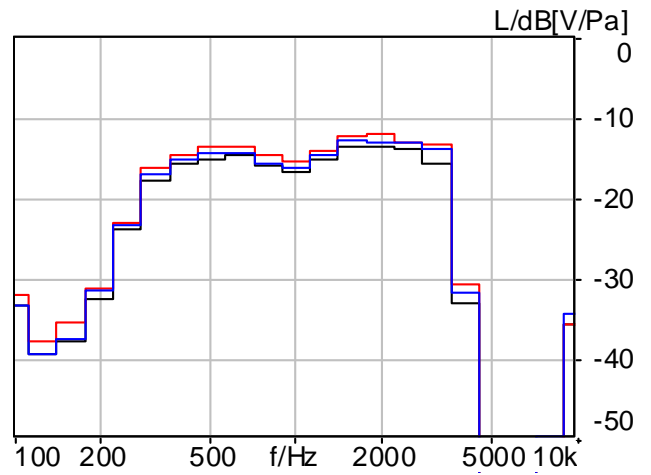**Figure B.3: Measurements of phone 3: black - fully anechoic;
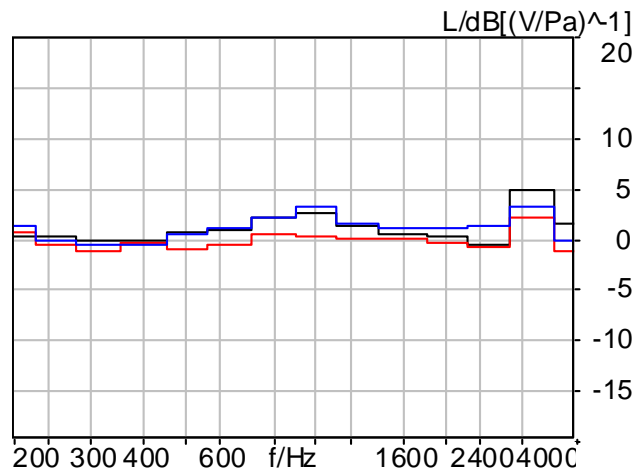red - office type room; blue - semi anechoic**
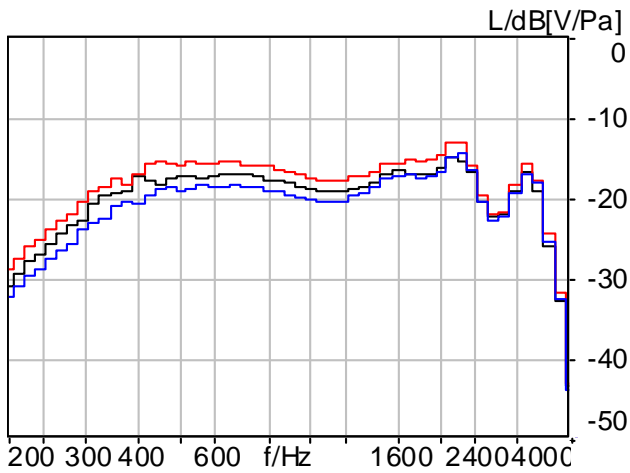
Sending Frequency Response

Direct Sound Sensitivity
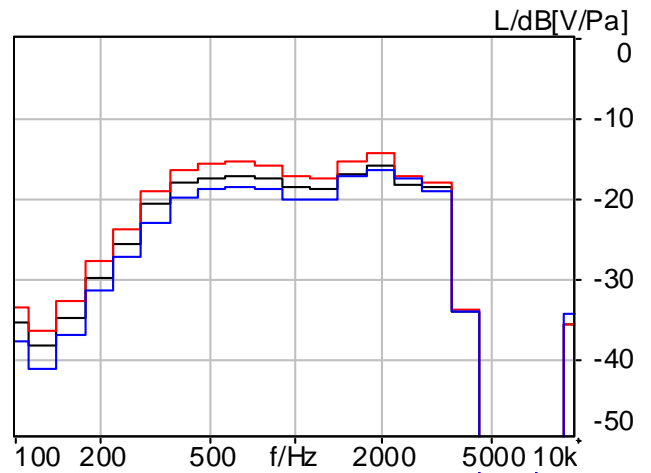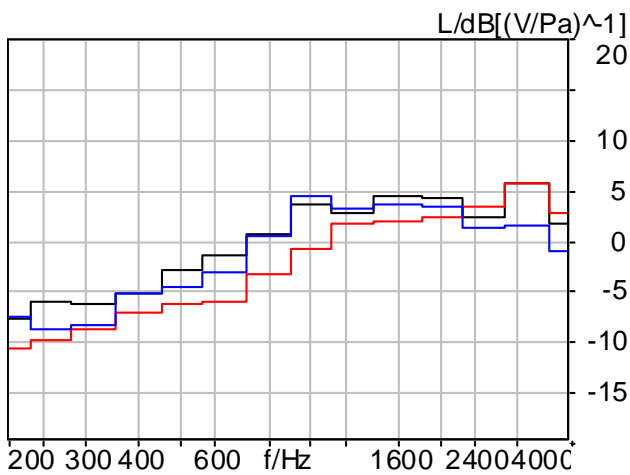
D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise
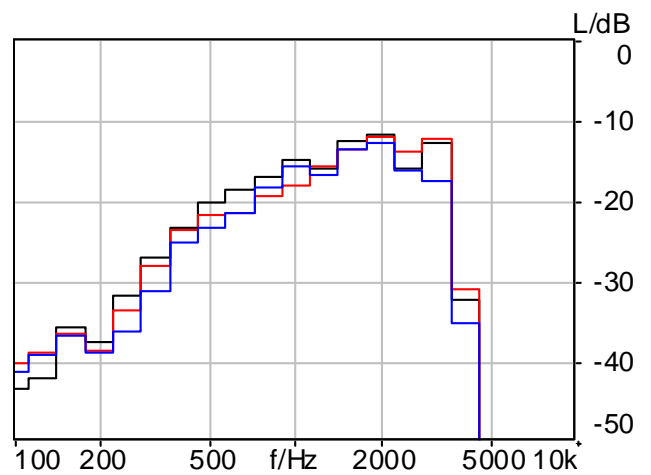
D-Value, Cafeteria Noise

Diffuse Sound Sensitivity, Cafeteria Noise

**Figure B.4: Measurements of phone 4: black - fully anechoic;
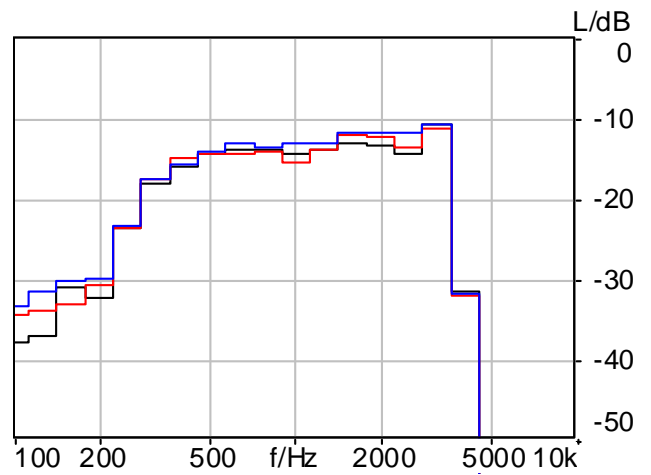red - office type room; blue - semi anechoic**

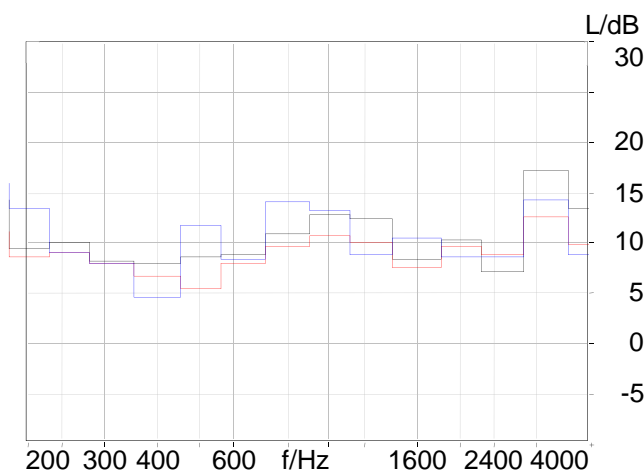**Sending Frequency Response**

Direct Sound Sensitivity

D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise

D-Value, Cafeteria Noise

Diffuse Sound Sensitivity, Cafeteria Noise

NOTE: Different line types = same test setup but test conducted different times with pink noise field.
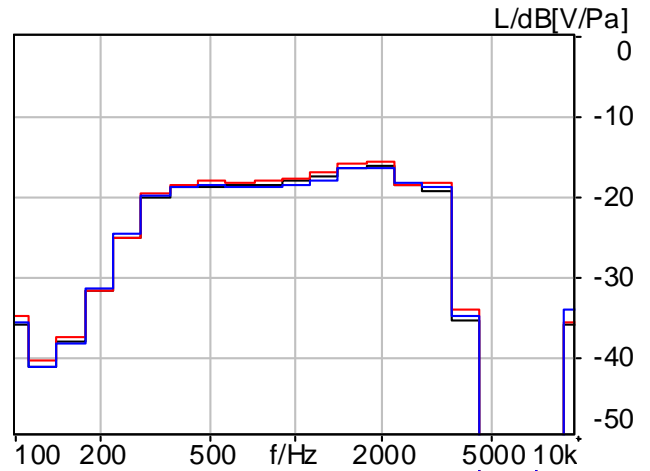
**Figure B.5: Measurements of phone 5: black - fully anechoic;
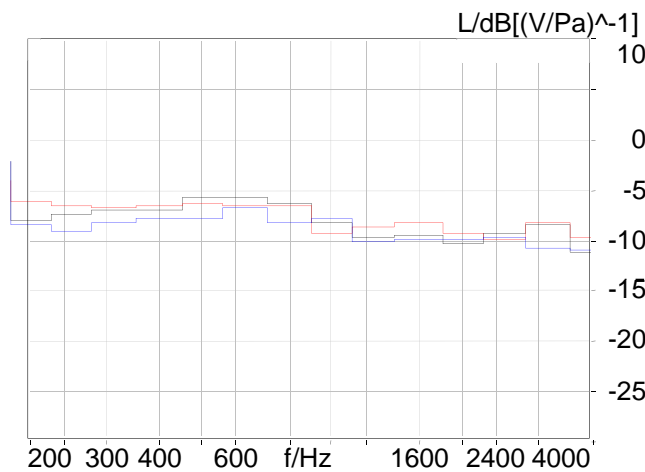red - office type room; blue - semi anechoic**

Sending Frequency Response

Direct Sound Sensitivity

D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise

D-Value, Cafeteria Noise

Diffuse Sound Sensitivity, Cafeteria Noise

**Figure B.6: Measurements of phone 6: black - fully anechoic;
red - office type room; blue - semi anechoic**

Sending Frequency Response

L/dB[V/Pa]

Direct Sound Sensitivity

L/dB[V/Pa]

D-Value, Pink Noise

L/dB[(V/Pa)^1]

Diffuse Sound Sensitivity, Pink Noise

L/dB

D-Value, Cafeteria Noise

L/dB

Diffuse Sound Sensitivity, Cafeteria Noise

L/dB[V/Pa]
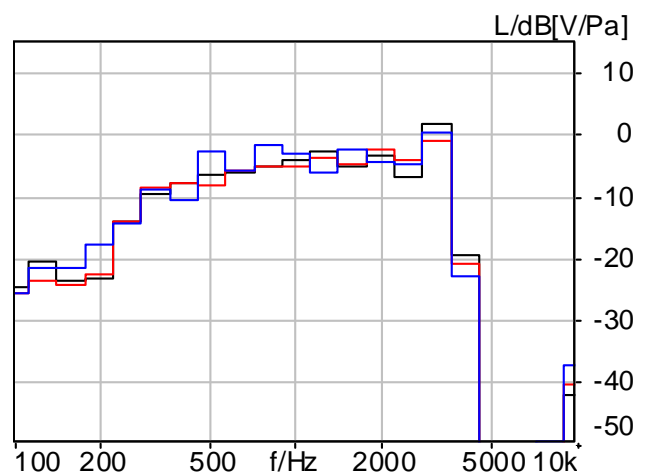
**Figure B.7: Measurements of phone 7: black - fully anechoic;
red - office type room; blue - semi anechoic**
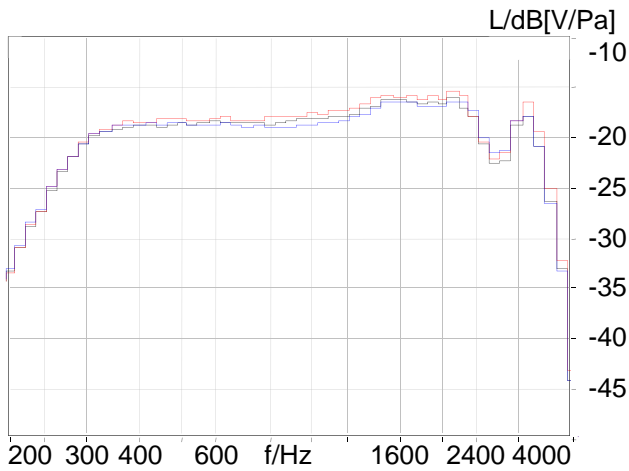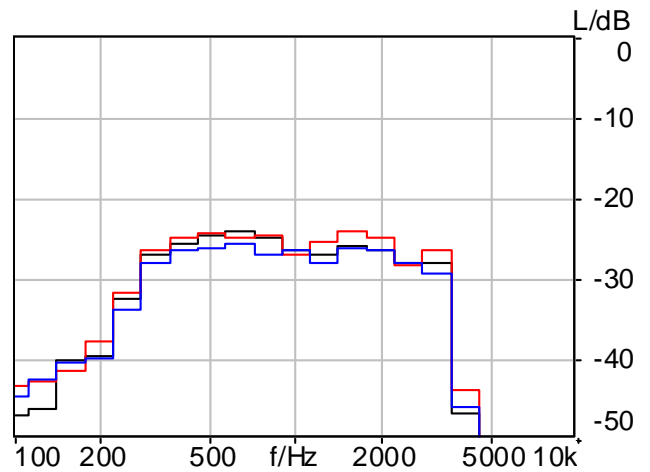
Sending Frequency Response
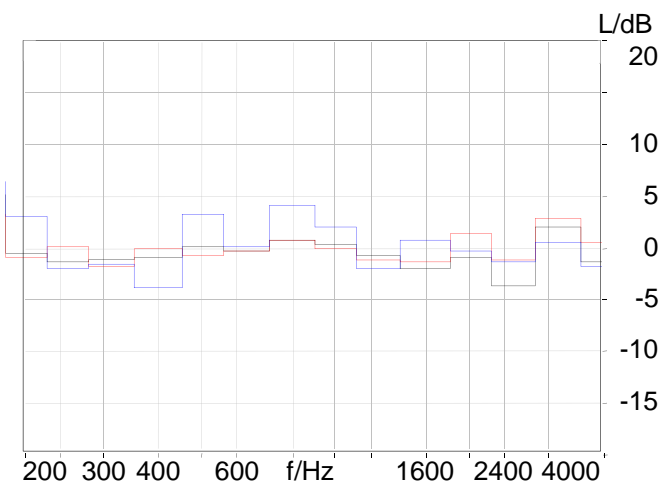
Direct Sound Sensitivity
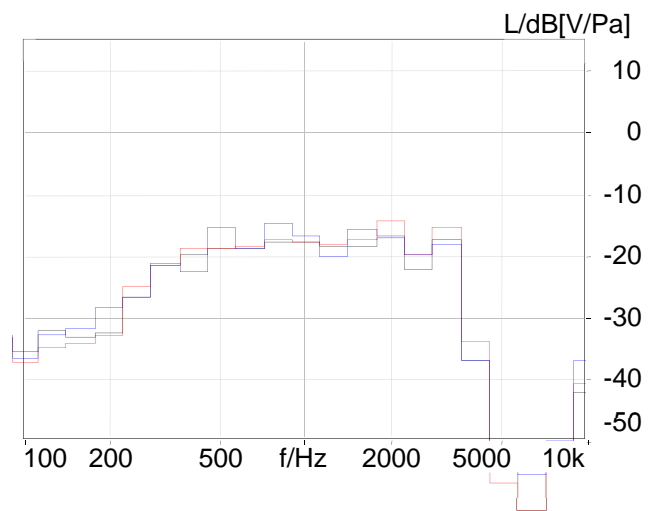
D-Value, Pink Noise

Diffuse Sound Sensitivity, Pink Noise

D-Value, Cafeteria Noise

Diffuse Sound Sensitivity, Cafeteria Noise

**Figure B.8: Measurements of phone 8: black - fully anechoic;
red - office type room; blue - semi anechoic**

History

Document history		
V1.1.1	November 2005	Publication as ETSI EG 202 396-1
V1.1.2	January 2006	Publication as ETSI EG 202 396-1
V1.2.2	September 2008	Publication as ETSI EG 202 396-1
V1.2.3	March 2009	Publication as ETSI EG 202 396-1
V1.2.4	February 2011	Publication as ETSI EG 202 396-1
V1.4.1	October 2012	Publication
V1.5.1	January 2014	Publication
V1.6.1	June 2015	Publication
V1.7.1	October 2017	Publication
V1.8.1	May 2022	Publication
V1.9.1	March 2023	Membership Approval Procedure MV 20230509: 2023-03-10 to 2023-05-09