ETSI TS 103 224 V1.3.1 (2017-07)

TECHNICAL SPECIFICATION

Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database

Reference
RTS/STQ-250

Keywords
noise, quality, speech, terminal

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 IPR Policy, no investigation, 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.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

The present document describes a sound field recording and reproduction technique which can be applied for all types of terminals but is especially suitable for modern multi-microphone terminals including array techniques. The present document provides an additional simulation technique which can be used instead of the part 1 of ETSI multi-part deliverable ES/EG 202 396 "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" [i.7];

ETSI EG 202 396-2: "Background noise transmission - Network simulation - Subjective test database and results" [i.8];

ETSI EG 202 396-3: "Background noise transmission - Objective test methods" [i.9].

The background noise simulation can be used in conjunction with the objective test methods as described in ETSI EG 202 396-3 [i.9] and ETSI TS 103 106 [i.10].

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 of 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 conditions. Since modern terminals incorporate more advanced noise cancellation techniques, such as multi-microphone based noise cancellation, the use of microphone-array recording techniques and more realistic noise field simulations (compared to the method described in ETSI ES 202 396-1 [i.7]) are required.

The present document addresses this topic by specifying a methodology for recording and playback of realistic background noise fields under conditions that are well-defined and able to be calibrated 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. Terminals, networks, and system configurations including wideband, superwideband, and fullband speech services can be greatly improved with a proper design of terminals and systems in the presence of background noise. The present document:

- describes a sound field simulation technique allowing to simulate the real environment using realistic background noise scenarios for laboratory use;
- contains a database including relevant background noise samples for subjective and objective evaluation.

The present document describes the recording technique used for the sound field simulation, the loudspeaker setup, and the loudspeaker calibration and equalization procedures. Furthermore the present document specifies the test room requirements for laboratory conditions.

The simulation environment specified can be used for the evaluation and optimization of terminals and of complex configurations including terminals, networks and others. The main application areas are: outdoor, 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 signals 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 http://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.

Not applicable.

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] Berkhout A. J., de Vries D., & Vogel, P.: "Acoustic control by wave field synthesis", J. Acoust. Soc. Am., p. 2764-2778, Mai 1993.

[i.2] Gerzon, M. A.: "Periphony: With-Height Sound Production", Journal of the Audio Engineering Society 21, 1973.

[i.3] Ward D. B., Abhayapala T. D.: "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers", IEEE transactions on speech and audio processing, Vol. 9, No.6, p. 697-707, September 2001.

[i.4] Kirkeby O., Nelson P. A., Orduna-Bustamante F., Hamada H.: "Local sound field reproduction using digital signal processing", J. Acoust. Soc. Am. 100(3), p. 1584-1593, September 1996.

[i.5] Kirkeby O., Nelson P. A., Hamada H., Orduna-Bustamante F.: "Fast Deconvolution of Multichannel Systems Using Regularization", IEEE transactions on speech and audio processing, VOL. 6, NO. 2, p. 189-195, March 1998.

[i.6] Recommendation ITU-T P.58: "Head and Torso Simulator for Telephonometry".

[i.7] ETSI ES 202 396-1: "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".

[i.8] ETSI EG 202 396-2: "Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 2: Background noise transmission - Network simulation - Subjective test database and results".

[i.9] ETSI EG 202 396-3: "Speech and multimedia Transmission Quality (STQ); Speech Quality performance in the presence of background noise; Part 3: Background noise transmission - Objective test methods".

[i.10] ETSI TS 103 106: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods".

[i.11] ISO 3382-1: "Measurement of room acoustic parameters -- Part 1: Performance spaces".

3 Symbols and abbreviations

3.1 Symbols

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

c	Sound velocity
C	Matrix of FFT coefficients of Compensation Filters
H	Matrix of FFT coefficients of Impulse Responses

3.2 Abbreviations

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

DUT	Device Under Test
FFT	Fast Fourier Transform
HATS	Head And Torso Simulator
IR	Impulse Response
MLS	Maximum Length Sequence
MOS	Mean Opinion Score
MRP	Mouth Reference Point

SNR	Signal to Noise Ratio
SPL	Sound Pressure Level

4 Methods for realistic sound reproduction

For reproduction of real world sound fields there exists a variety of different methods, two of them are wave field synthesis [i.1] and Ambisonics [i.2]. Both methods, however, require a large number of microphones and loudspeakers to achieve a sound field reproduction which is sufficiently good for testing purposes. The Wave-Field synthesis setup is that complex and expensive that it can be neglected for laboratory purposes. Ambisonics, for example, has to be performed using 43 microphones and 43 loudspeakers to reach a good sound field reproduction up to 2 kHz in a sweet spot with radius 15 cm (using the rule of thumb in [i.3]). It furthermore cannot consider individual room characteristics or insufficiencies, but is only designed for rooms offering pure free field conditions. If, e.g. for testing purposes a HATS is positioned in the artificial noise field, the reproduction quality is reduced by an unknown amount. In summary, the Ambisonics approach is due to its design not feasible for the intended testing scenario.

The present document introduces an alternative least mean squares method [i.4], which requires eight recording channels and eight loudspeakers in order to achieve reasonably good reproduction results. The method is based on eight sweet spots at important testing positions e.g. near the HATS, mainly at the microphone positions of modern phones.

A reasonable reproduction of the recorded sound field at the corresponding eight points in the reproduction situation also yields good reproduction accuracy in between these points. This well-known property of sound fields is limited to an upper cut-off frequency which depends on the distances between the recording microphones (see clause 5.1.1).

In clause 5, the recording technique required for this new method is described, while the setup allowing the reproduction in laboratories and the different steps of the equalization procedure are introduced in clause 6.

5 Recording arrangement

5.0 General

The sound field recording technique (Multi-point sound field recording technique) is based on optimization of the sound field reproduction at different points in space. The optimization criterion is based on minimization of the reproduction error at each microphone position. Based on this principle the microphone locations and as a consequence the points in space for which the sound field reproduction is mostly accurate can be chosen in a wide range. The advantage of the method is that these locations can be adapted to the type of device which is to be tested. E.g. if the device under test (DUT) incorporates a microphone array of the Multi-point sound field recording microphones can be positioned in the area of the microphones of the DUT. If a hands-free device is to be tested the Multi-point sound field recording microphones are positioned in the area of the hands-free device.

The setup described in detail in clause 5 is optimized for the testing of handset or headset terminals using HATS according to Recommendation ITU-T P.58 [i.6] and for hands-free testing. The procedure described here can be followed in the same way for other microphone setups.

In this clause the setups for the microphone arrangements as used in the present document are described. The background noise recordings based on these different recording setups are described in clause 8.

5.1 Microphone array setup

5.1.1 Principle limitations

With a perfect sound field reproduction at two closely spaced points, the cut-off frequency up to which the sound field in between those two points is also correctly reproduced depends on their distance. This upper cut-off frequency can be estimated as:

$$f_{lim} = \frac{c}{2d_{max}} \quad (1)$$

where d_{max} is the maximum distance between two microphones and c is the sound velocity.

EXAMPLE: For the eight microphones in Figure 1, f_{lim} is dependent on the distance of the microphone pair considered and is about *1,7 kHz* in the region of sparsely spaced microphones and approximately *3 kHz* in the region of densely spaced microphones. Note, that at the microphone positions itself the reproduction quality is optimal across the whole frequency range. In between of these positions the accurate spatial reproduction can only be guaranteed up to f_{lim}.

5.1.2 Microphone calibration

In order to yield a good sound field reproduction at the defined positions, the microphone array for recording of the real sound field and the microphone array for equalization and calibration of the reproduction setup have to match. In detail, the frequency/phase response and the directional sensitivity of the corresponding microphones of the two arrays has to be identical. As a consequence, each microphone has to be calibrated individually with regard to frequency response, phase response and level.

The supplier of such devices should provide information regarding the sensitivity of the individual microphones constituting the microphone array for verification purposes. The calibration data provided need to be suitable to ensure a proper phase calibration up to at least 3 kHz, a proper frequency response calibration in the frequency range between 50 Hz and at least 3 kHz with an accuracy of < 0,5 dB in 1/12th octave, between 3 kHz and 10 kHz with an accuracy of < 0,5 dB in 1/3rd octave, between 10 kHz and 20 kHz with an accuracy of < 3 dB in 1/3rd octave and a proper level calibration (at 250 Hz or 1 kHz) with an accuracy of < 0,1 dB.

5.2 Microphone array setup for handset-type and headset terminals

Figure 1 shows the configuration of microphones located around an artificial head. The locations of the microphones define the sweet spots where the reproduction of the recorded signals is optimal for all frequencies. In consequence the majority of these points are at relevant positions where the microphones of the test devices are usually located (see Figure 1, top left). The exact positions for the eight recording microphones are given in Figure 1 (bottom). Eight additional positions are defined by clockwise rotation of the microphone array by 10 degrees. (Figure 1, top right, in dark) around the axis of rotation of the HATS as defined in Recommendation ITU-T P.58 [i.6]. This position is called "fine tuning set" and is used for optimization and verification of the equalization.

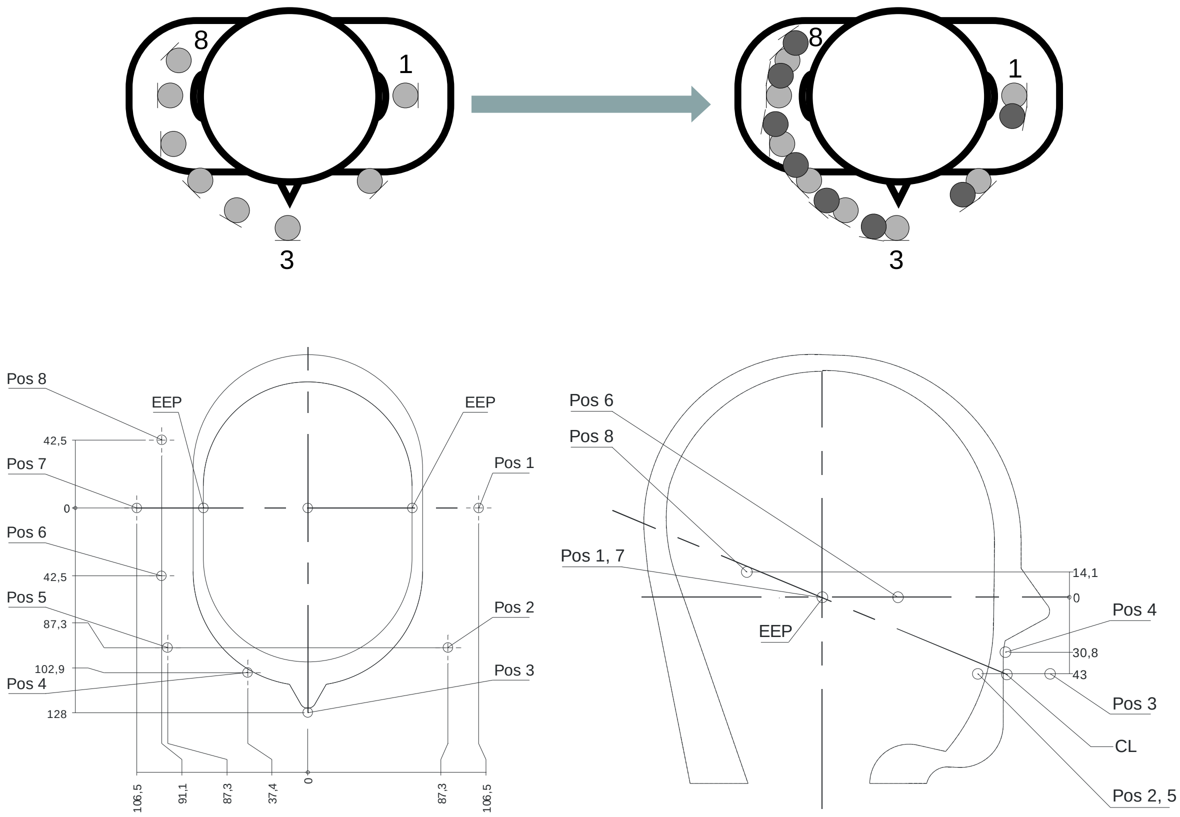


Figure 1: Positions of the recording microphones
Vertical positions are related to the vertical position of the MRP

5.3 Microphone array setup for hands-free terminals

In general, different microphone arrays could be used for hands-free terminals as well as for handsets and headsets. However, to increase reusability and reduce efforts, the same microphone array can be used in both cases. The setup of the array for measuring hands-free terminals is shown in Figure 2.

For the hands-free equalization, the DUT is first positioned at its testing position, which is defined in the relevant standards. Then, the main microphone position of the terminal is determined. In the case of terminals using multi microphone techniques terminals the main microphone is chosen, and in case of array techniques the acoustical centre of the array (typically identical to the centre of the array) is used.

In the setup for hand-held and tablet terminals, the microphone array is positioned such that microphone 5 is in top view right-angled in front of the main microphone position in *25 mm* distance (Figure 2, right) and microphone 6 is at the height of the main microphone position (Figure 2, left).

For desktop operated hands-free terminals, the microphone 5 of the array is positioned right-angled in front of the main microphone position in *25 mm* distance (Figure 3, right) and *25 mm* above the table (Figure 3, left).

Note that the DUT is absent during the equalization procedure itself.

The "fine-tuning set" is realized the same way as described in clause 5.2, rotating the microphone array clockwise by 10 degrees.

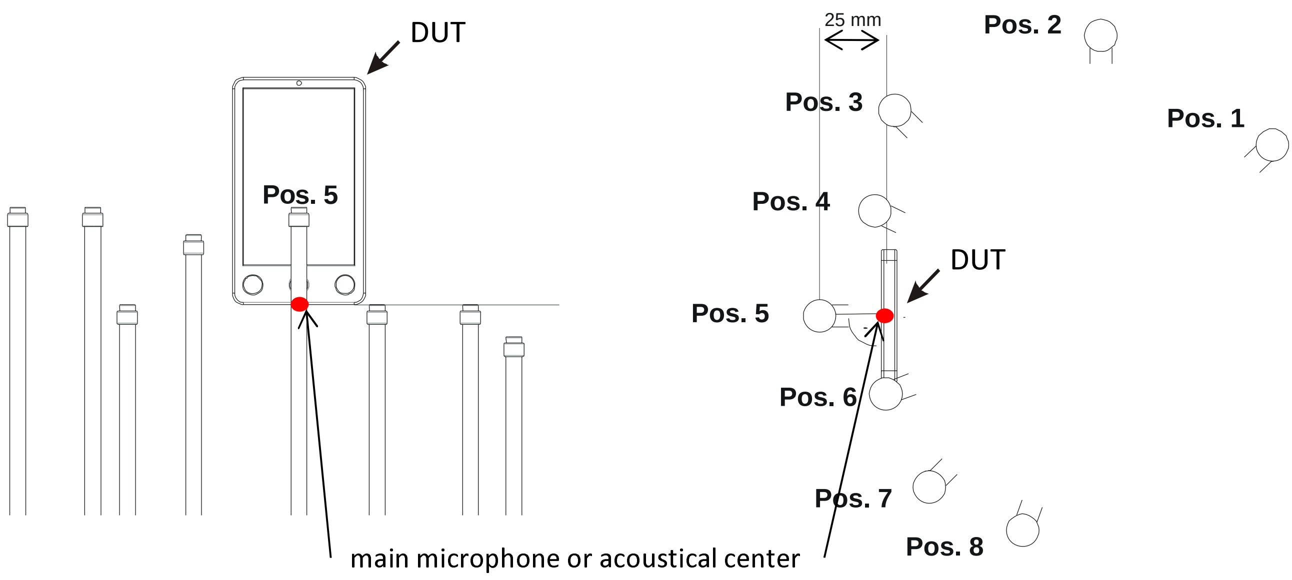


Figure 2: Positions of the recording microphones in a hands-free setup for hand-held and tablet terminals

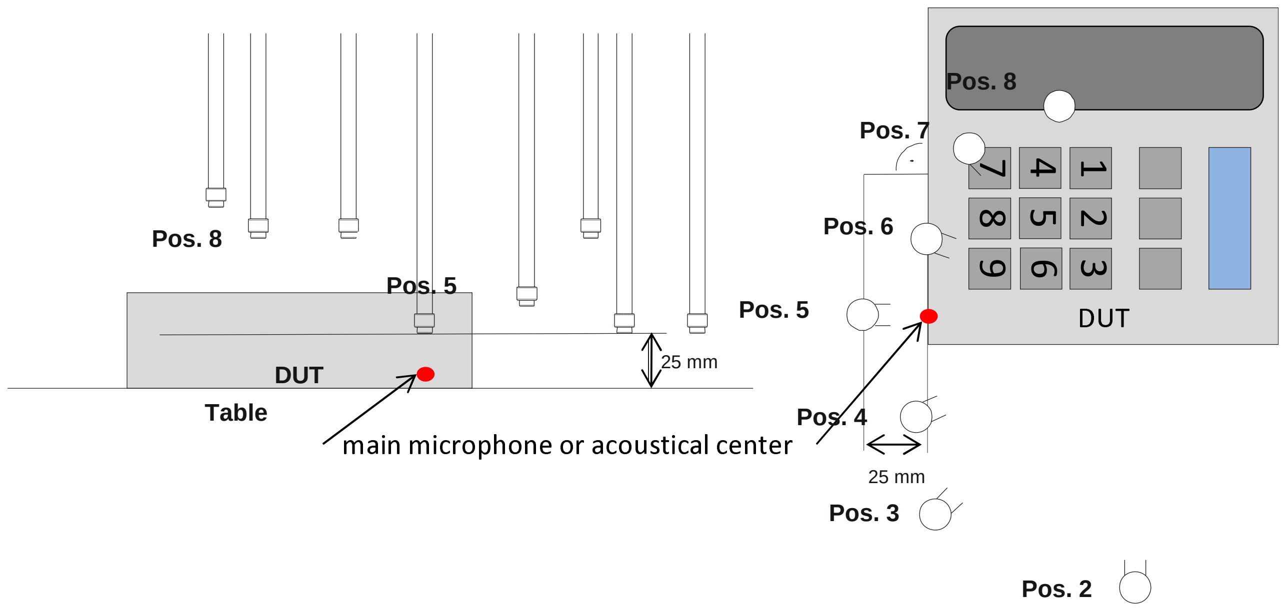


Figure 3: Positions of the recording microphones in a hands-free setup for desktop operated hands-free terminals

6 Loudspeaker setup for background noise simulation

6.0 General setup

It should be noted that the position height of the loudspeakers as well as the exact spacing between them in general is not critical since the equalization procedure described below accounts for the individual loudspeaker positions. The difference which might be observed between different loudspeaker positions is a different deviation from the original sound field at the intermediate positions of the microphone array. In order to allow better inter-lab accuracy of the sound field reproduction the following positioning arrangement should be followed if the room allows.

Figure 4 shows the setup of the eight loudspeakers for the desired sound field reproduction. The vertical position of the loudspeakers is adjusted so that the centre of every other loudspeaker (e.g. 1, 3, 5 and 7) is about 15 cm above the HATS reference plane [i.6] and the centre of the remaining four loudspeakers (e.g. 2, 4, 6 and 8) is about 15 cm below the HATS reference plane. The distance between the loudspeakers to the HATS as well as the horizontal distribution of the loudspeakers can be selected depending on the room, hence the spacing between the loudspeakers does not have to be exactly equal. The setup may be a square or a circle around the HATS or a setup in between depending on what fits the room best.

The distance between the surface of the artificial head and the loudspeaker fronts should be at least 50 cm and should not exceed 2,5 m. Note, that the maximum distance is also limited by the maximum sound pressure level which can be produced by the loudspeakers. For the application of reproducing realistic background noises the reproduction of a maximum sound pressure level of 105 dB SPL in the frequency range from 50 Hz to 5 kHz is considered to be sufficient. Due to the typically much lower signal energy from 5 kHz to 20 kHz the sound pressure level produced at such frequencies may be lower. In general it is advisable to select high quality loudspeakers with a mostly flat free-field response characteristics and low distortion at maximum desired sound pressure.

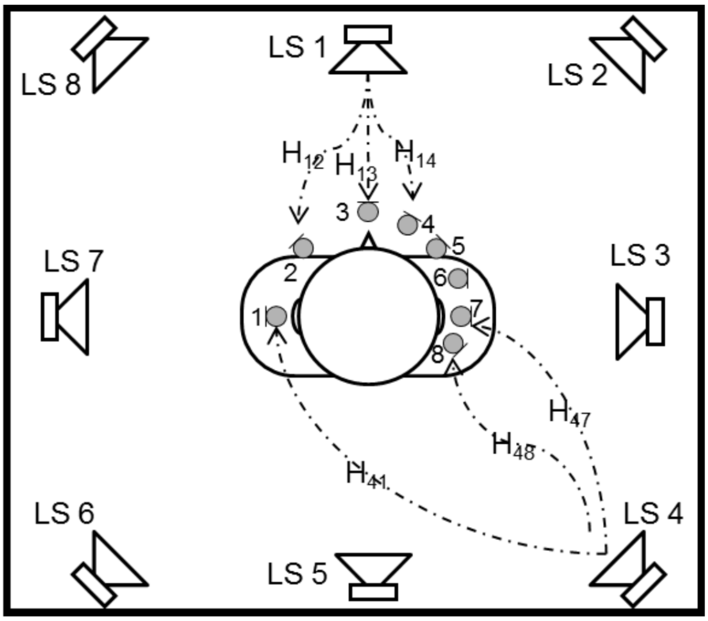


Figure 4: General loudspeaker setup and principle of the equalization paths for the handset and headset measurement setup

6.1 Test room requirements

The room required by the reproduction technique may vary from acoustically treated office rooms to anechoic rooms. The playback room should meet the following requirements:

- **Room size:**

  The room size should be in a range between 1,8 m × 2,4 m × 2,1 m to 8 m × 9 m × 4,5 m (L x W x H).

- **Room acoustic parameter clarity 80:**

  The most important criterion a room has to fulfil depends on the clarity 80 (C_{80}) [i.11]. This parameter is defined as the signal energy of the first *80 ms* of the impulse response (IR) in relation to the remaining energy of the impulse response expressed in dB:

$$C_{80} = 10\log\left(\frac{\int_{0ms}^{80ms} p^2(t)\,\mathrm{d}t}{\int_{80ms}^{\infty} p^2(t)\,\mathrm{d}t}\right) \quad (2)$$

  Suitable rooms shall have $C_{80} > 20\ dB$.

- **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. Additional damping materials may need to be applied in order to reach the C_{80} value given above.

  For anechoic or semi-anechoic chambers no additional treatment is needed.

- **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 Equalization and calibration

6.2.0 Overview of the equalization procedure

For equalization the same microphone array setup with the same microphone position has to be used as for the recording setup described in clause 5.2. Accordingly, the microphone array has to be calibrated as described in clause 5.1.2. The equalization itself can then be performed completely automated - independent of the microphone array setup.

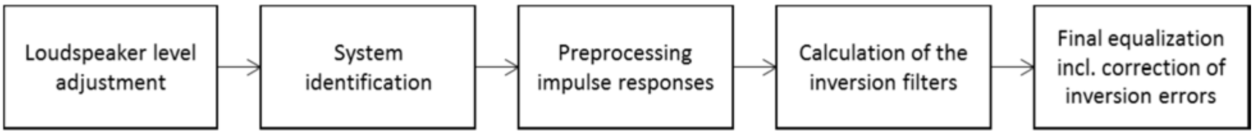


Figure 5: Blocks of the equalization procedure

Figure 5 shows the overview of the complete equalization procedure, which consists of the following steps:

1) separate level adjustment for every loudspeaker;
2) system identification;
3) pre-processing of the impulse responses for equalization;
4) calculation of the inversion filters;
5) first test of inversion with recorded noise and adjusting the filters to compensate possible inversion errors.

Each of these steps is described in clause 6.2.

6.2.1 Separate level adjustment for each loudspeaker

First of all, the sound pressure level of each loudspeaker is adjusted to be the same for all loudspeakers. To achieve that the average sound pressure level is measured and calculated across the whole frequency range and across the eight microphones. The average sound pressure level for every loudspeaker should be at least 70 dB SPL, which is necessary to ensure a sufficient SNR for measuring the impulse responses in the next step. Care should be taken not to overload the loudspeakers, a headroom of at least 30 dB should be left for the equalization procedure. The test signal used for the loudspeaker level adjustment is a logarithmic sweep signal (or MLS signal) with a constant amplitude. The signal level is determined by averaging the signal level over the entire sweep (resp. MLS) signal.

6.2.2 System identification

In this step the impulse responses between all combinations of loudspeakers and microphones are measured. Figure 4 shows the eight-microphone/eight-loudspeaker setup. H_{ri} represents the impulse response (frequency response) from loudspeaker *r* to microphone *i*.

There exist different possibilities for measuring impulse responses, e.g. using maximum length sequences (MLS) or using swept-sines (sweeps). The advantage of sweeps is that non-linearities can easily be observed and that the SNR in lower frequencies is higher than with MLS. Using sweeps is therefore recommended for system identification. Using the sweep *S(f)*, which is played back with the loudspeaker, and the recorded microphone signal *Y(f)*, the frequency response $H_{ri}(f)$ can be calculated as:

$$H_{ri}(f) = \frac{Y(f)}{S(f)}. \tag{3}$$

For sufficient system identification, the response should be calculated in the frequency range 20 Hz to 20 kHz.

6.2.3 Pre-processing of the impulse responses

As motivated in clause 5.1.1, the sound field is only correctly reproduced up to a cut-off frequency, which is in the range between *2 kHz* and *3 kHz* for the given microphone setup. For higher frequencies, the tail of the impulse response degrades the quality of the inversion. To cope with that, a low pass filter with a time-variant cut-off frequency is applied to the measured impulse response.

1) Filter the impulse response in the frequency domain by multiplying a window function which has the value 1 for frequencies smaller than f_{pass} and 0 for frequencies greater than f_{stop}. Between those two frequencies the window function has a cosine (hanning) characteristic.

2) The pre-processed impulse response is calculated in the time-domain as the weighted sum of the original impulse response and the lowpass-filtered impulse response from step 1).
The time-variant weighting of the original impulse response is 1 up to t_{min} after the start of the impulse response and falls down to 0 with a cosine (hanning) characteristic until t_{max} after the start of the impulse response. The time-variant weighting of the lowpass-filtered impulse response from step 1) is 1 minus the weighting of the original impulse response.

The values for these parameters are given in Table 1.

Table 1: Parameters for pre-processing the impulse responses between loudspeakers and microphones

f_{pass}	*2 kHz*
f_{stop}	*4 kHz*

t_{min}	*1 ms*
t_{max}	*4 ms*

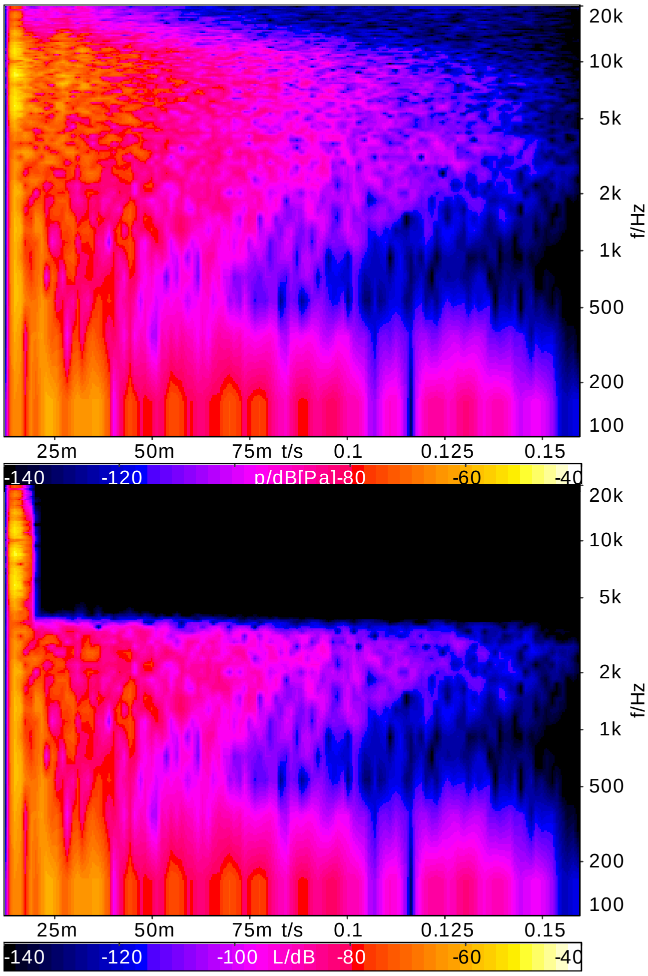


Figure 6: Comparison of spectrum vs. time analysis: original (measured) impulse response (upper) and pre-processed impulse response (lower)

The result of this procedure can be seen in Figure 7 as an example a typical test room.

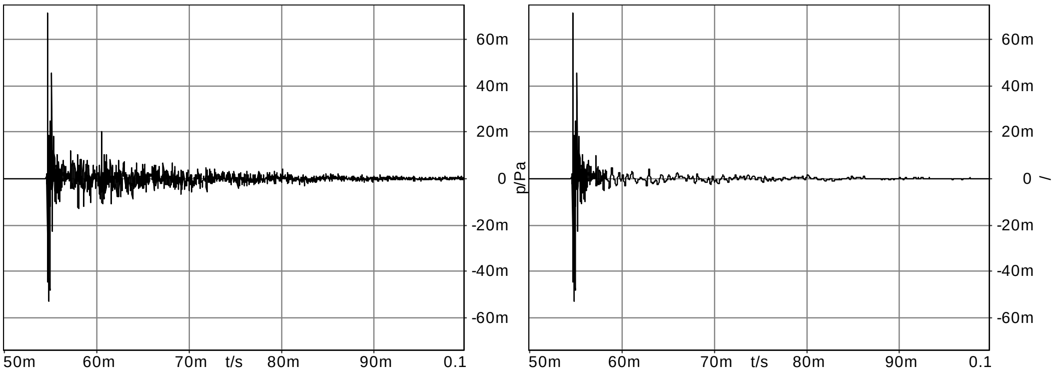


Figure 7: Original (measured) impulse response (left) and processed impulse response (right)

6.2.4 Calculation of the inversion filters

6.2.4.0 Overview

Figure 8 provides a block diagram of the inversion process of the impulse responses.

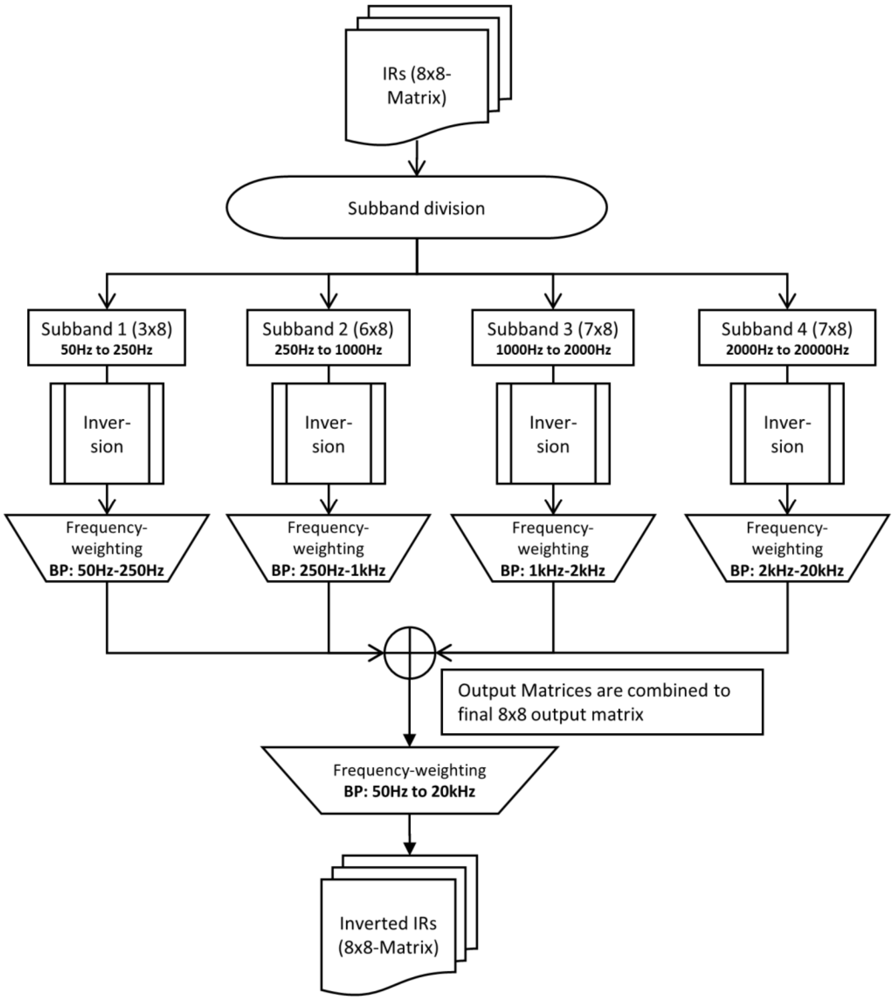


Figure 8: Block diagram of the inversion process of the impulse responses

NOTE: Every subband is processed with a subset of impulse responses according to the microphones used in each subband as defined in Table 2. The matrix-size of the matrix which has to be inverted is given in brackets, e.g. "Subband 1 (3×8)".

All impulse responses for all combinations of loudspeakers and microphones are inverted individually. Each impulse response is first segmented in four subbands as specified in Figure 8. As described in Table 2, only a subset of all microphones is used in each subband.

After finalizing the regularization process the resulting inverted impulse responses are combined in the different frequency bands chosen by applying the appropriate frequency weighting (bandpass filtering) shown in the filter blocks after the inversion blocks in Figure 8 are available.

6.2.4.1 Inversion procedure

The starting point for developing the inversion procedure is writing the signal $P_i(f)$ arriving at microphone i as a linear combination of the signals $X_l(f)$ played back with loudspeaker l multiplied with a transfer function $H_{li}(f)$, which models the acoustic path between microphone and loudspeaker. This can easily be seen in Figure 4.

$$P_i(f) = \sum_{l=1}^{L} H_{li}(f) X_l(f). \tag{4}$$

This linear combination can be written as a linear system (for the sake of simplicity the frequency index f is omitted in the following equations):

$$\underbrace{\begin{pmatrix} H_{11} & H_{21} & \dots & H_{L1} \\ H_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ H_{1N} & \dots & \dots & H_{LN} \end{pmatrix}}_{\mathbf{H}} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_L \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \quad (5)$$

As stated above the aim of this procedure is to find the signals $X_l(f)$ to be reproduced by the loudspeakers for a given set of spectra $P_i(f)$ recorded at the microphones. Obviously, the number of loudspeakers have be at least the number of microphones otherwise the linear system is under-determined.

The goal of the equalization is finding compensating filters C. These filters are then used for generating loudspeaker signals which in combination reproduce a given sound pressure at the microphone positions.

$$\underbrace{\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1N} \\ C_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ C_{L1} & \dots & \dots & C_{LN} \end{pmatrix}}_{\mathbf{C}} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_L \end{pmatrix} \quad (6)$$

The exact solution of this linear system is given by:

$$\mathbf{C} = (\mathbf{H}^*\mathbf{H})^{-\mathbf{1}}\mathbf{H}^*. \quad (7)$$

The direct solution of (4) may introduce problems: at lower frequencies the values of H_{ij} are very similar to each other which means that the matrix becomes ill-conditioned and the condition number $(A)=(\sigma_{max}(A))/(\sigma_{min}(A))$ increases, where $\sigma_{max}(A)$ and $\sigma_{min}(A)$ are the maximum and the minimum singular value of A, respectively. A big condition number may result in practically unfeasible high gains.

Therefore the Thikonov-equalization [i.5] is used.

$$\mathbf{C} = (\mathbf{H}^*\mathbf{H} + \beta\mathbf{I})^{-\mathbf{1}}\mathbf{H}^*. \quad (8)$$

Here a regularization factor β is introduced which reduces poles and improves the condition of the matrix. The regularization factor β is investigated in clause 6.2.4.3 in more detail.

In case a mixture of different types of loudspeakers is used for reproduction, their different characteristics, e.g. their frequency response can be considered with a weighted version of (8):

$$\mathbf{C} = (\mathbf{H}^*\mathbf{W}\mathbf{H} + \beta\mathbf{I})^{-\mathbf{1}}\mathbf{W}\mathbf{H}^*. \quad (9)$$

The loudspeaker weighting matrix **W** is a frequency-dependent diagonal matrix. It is used as a high-pass for small loudspeakers which prevents damaging the speakers at low frequencies. **W** shall be 1 for the supported frequency range of the loudspeaker, i.e. above the -3 dB point of the frequency response, and faded to 0 with cosine characteristic within a 1/3rd octave transition.

6.2.4.2 Different microphones for different frequency bands

On the one hand, a rising number of microphones leads to a worse condition of the matrix which has to be inverted and therefore the number of microphones should be limited. On the other hand, the quality of the spatial reproduction increases with a higher number of microphones. In order to further optimize the inversion quality, different numbers of microphones can be used for different frequency bands. For lower frequencies, the wavelengths are rather long and therefore the distance between the used microphones can be higher, whereas the used microphones should be closer to each other for higher frequencies.

Measurements have shown that the following microphone combinations should be used.

Table 2: Subband definitions and used microphones per subband

Subband	Frequency range	Microphone positions
1	50 Hz to 250 Hz	1,3,7
2	250 Hz to 1 000 Hz	1,2,3,5,7,8
3	1 000 Hz to 2 000 Hz	1,2,3,4,5,7,8
4	2 000 Hz to 20 000 Hz	1,3,4,5,6,7,8

6.2.4.3 Search for the optimum regularization factor

6.2.4.3.0 Introduction

In order to find the optimum regularization factor, a (basic) iterative procedure is applied in all subbands individually. In the last subband with frequencies above 2 kHz, the wavelengths are shorter and a second iterative procedure refines the result of the first one.

6.2.4.3.1 Basic methodology to find the optimum regularization factor

The first iterative process used for the calculation of the inversion filters in each block of the equalization procedure (see Figure 8) is shown in Figure 9.

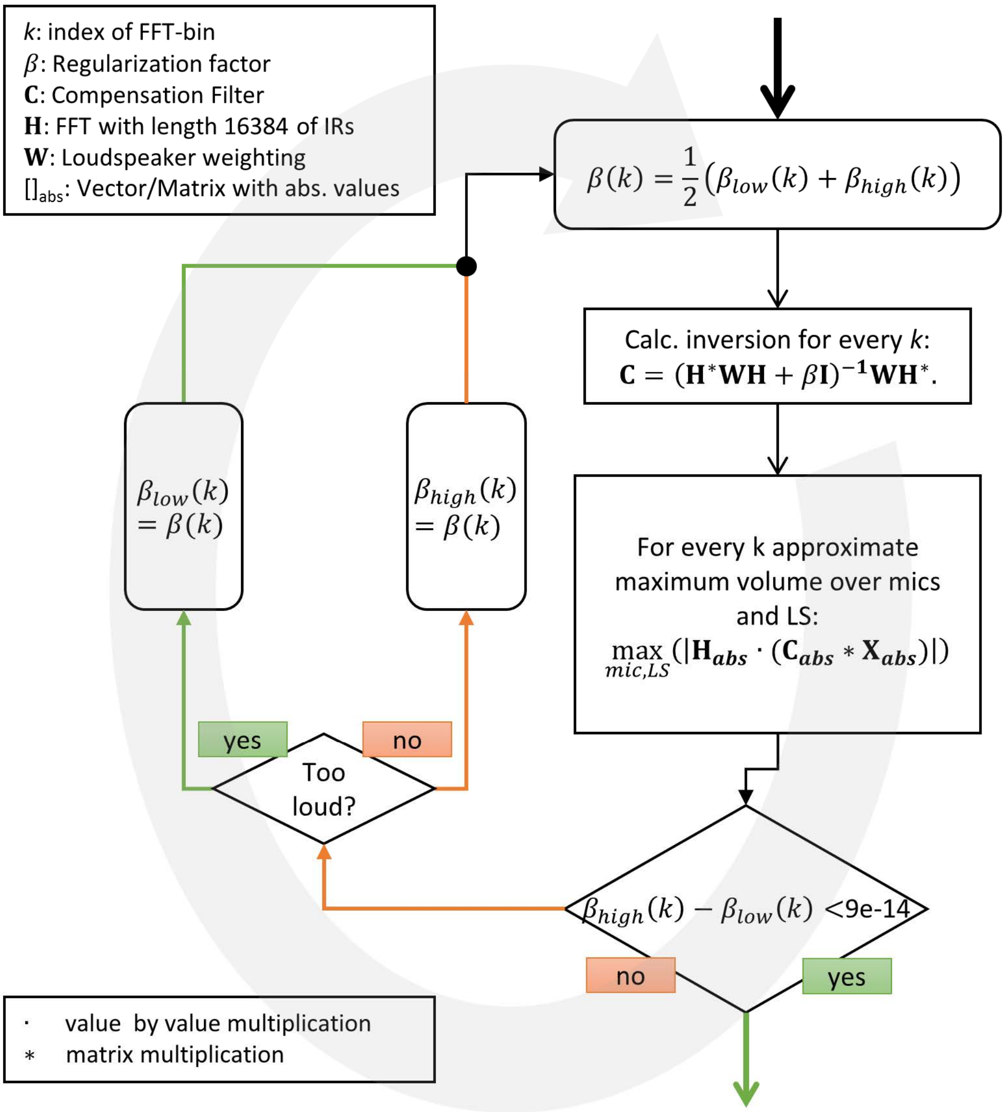


Figure 9: Basic iterative process for calculating the regularization factors

To find a more exact solution of the equation system, an iterative procedure is used to find the lowest possible regularization factor β. This process is needed to avoid unstable solutions which in consequence may cause high loudspeaker output levels and leading to distortions of the loudspeaker signals.

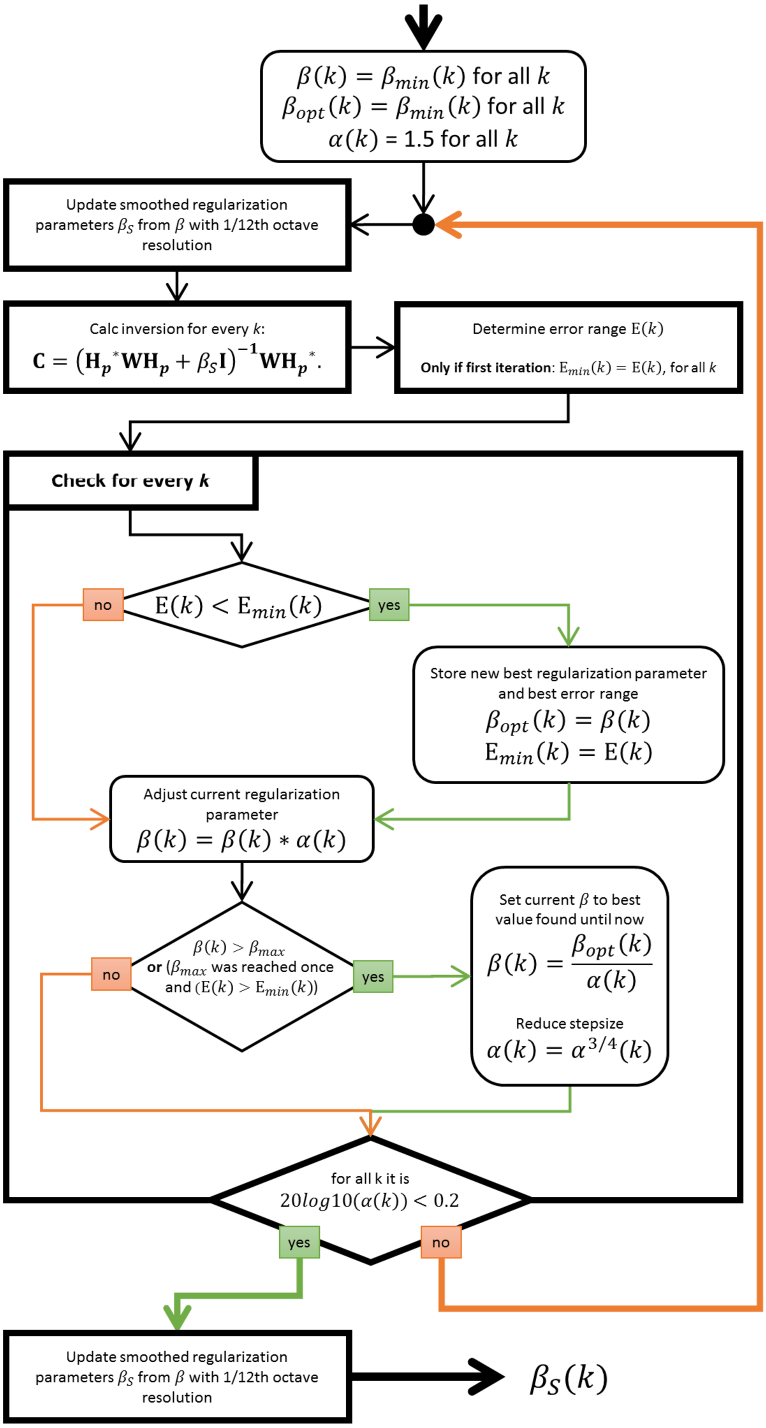
What follows is a summary of the search for the best regularization factor as shown in Figure 9. First, the variables β_{low} and β_{high} are initialized with a very small value for β_{low} (e.g. 1^{-12}) and a large value for β_{high} (e.g. *10*). Then, the regularization factor is chosen to the mean of β_{low} and β_{high}, and the compensating filters are calculated. Using these filters, the maximum output level of the loudspeakers is approximated. This is accomplished by simulating a playback of a typical (loud) sound recording. The spectrum of this recording is represented by $\mathbf{X}$ or $\mathbf{X_{abs}}$ meaning its absolute version.

NOTE: In general different spectra can be used for **X**. As a general rule a spectrum representative for the maximum level and spectral content to be reproduced by the simulation system should be chosen. This can be achieved for instance by measuring with a single measurement microphone in typical noise situations which the system will be used for. After that the spectra of these measurements are averaged.

As $\mathbf{X_{abs}}$ is an n-element vector, the matrix multiplication $\mathbf{C_{abs}} * \mathbf{X_{abs}}$ results in a virtual source signal which is used for simulating a playback by the loudspeakers. As the maximum level recorded at one microphone resulting from one loudspeaker shall be determined (in contrast to the maximum at one microphone resulting from all loudspeakers playing at the same time), the value by value result of $\mathbf{H_{abs}} \cdot (\mathbf{C_{abs}} * \mathbf{X_{abs}})$ is calculated. After that the maximum of the resulting matrix can be calculated to determine the maximum output level.

In cases where the loudspeakers reach their performance limit, i.e. for levels greater than 0 dB [Pa] at the microphone positions, the regularization factor is increased and the next iteration cycle is performed until β_{low} and β_{high} converged and an optimal value for β is found.

6.2.4.3.2 Extended methodology to find the optimum regularization factor for frequencies above 2 kHz



β_{max}	Maximum regularization factor (constant = 1)		
All of the following values are in FFT-resolution (16384) and depend on k which is used to index single frequency bins. k is only written if needed.			
β_{min}	Regularization factors from basic regularization iteration method	β	Current regularization factor
β_{opt}	Optimum regularization factor	β_S	Smoothed regularization factor
$\mathbf{C}$	Compensation Filter (matrix)	E	Error Range
$\mathbf{H_p}$	FFT of preprocessed IRs (matrix)	E_{min}	Minimum error range
$\mathbf{W}$	Loudspeaker weighting (matrix)	α	Multiplier to increase β in each iteration

Figure 10: Extended iterative process for calculating the regularization factors

The error at microphones in between the calibration set positions may become larger for frequencies higher than *2 kHz* since the wavelengths are quite short in this frequency range. This might become a problem if a microphone of the DUT is located at such a position.

To overcome this, the regularization factor can be increased which on the one hand decreases the quality of reproduction at the calibration set positions but on the other hand improves the error range over all microphones of the calibration as well of the fine-tuning set.

Therefore, a second iterative procedure is used to find the regularization factors with the lowest error range starting with the best β values found during the first procedure. In each iteration, the β values are smoothed over frequency and the filters are calculated. Then a playback is simulated by filtering the reference noise recording (see clause 8.1) with those calculated filters C and convolving the result with the original impulse responses H before pre-processing for the calibration as well as the fine-tuning set positions.

After that, the range of the error, both at calibration and fine-tuning position, is calculated by determining the spectral deviation between the reference noise recording and the simulated playback and by calculating the difference between minimum and maximum.

1) A playback should be simulated using the non-preprocessed IRs and the filters calculated with current regularization factors.

2) The simulated signals from 1 as well as the expected/source signals shall be smoothed with $1/24^{th}$ octave resolution.

3) The differences between expected and simulated spectra shall be calculated.

4) Minimum and maximum over all microphones shall be determined for every k.

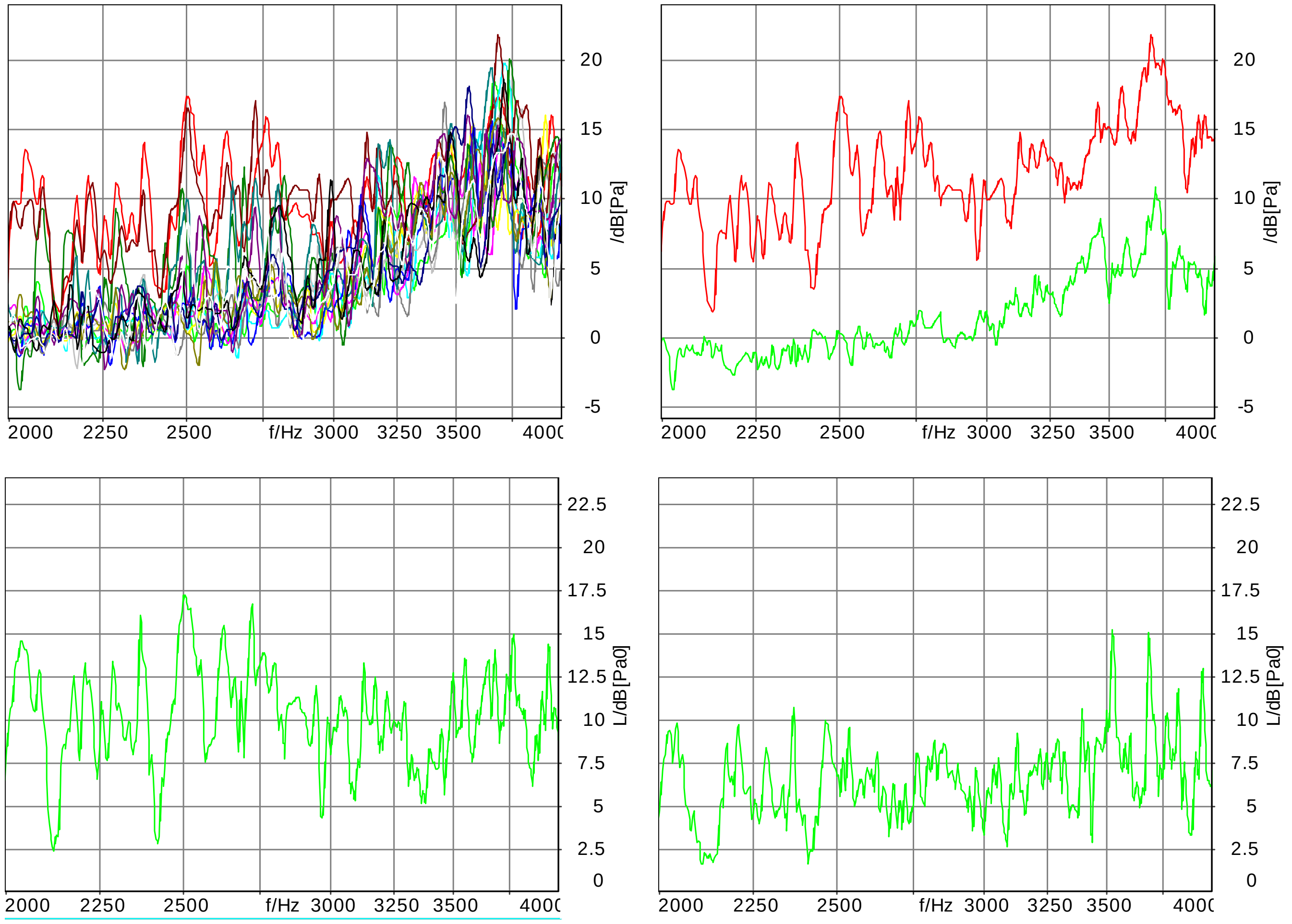
5) Then the error range can be calculated for all k: $\mathrm{E(k)} = \mathrm{MAX(k)} - \mathrm{MIN(k)}$

The smoothing which is applied during the iterative process shown in Figure 10 is calculated as follows:

1) Calculate $1/n^{th}$ octave representation of given vector.

2) Perform cubic interpolation between the bins of the 1/nth octave representation.

3) Read values at FFT-bins from interpolated curve.



NOTE: Top left: absolute error before optimization at 16 microphone positions;
top right: minimum and maximum of error;
bottom left: error range before optimization (difference between minimum and maximum);
bottom right: final error range after optimization.

Figure 11: Example of the calculation of the error range and its enhancement due to optimization

If the error range became smaller, the regularization factor β is stored as new best value. Then β is increased by multiplication with α, which is initialized with 1,5. If β exceeds the upper threshold $\beta_{max} = 1$, it is set back to a little less than the found best values and the multiplier is decreased. It is also set back if the error range did not become smaller and β has already been reset once. The iteration ends if the multiplier becomes very small, i.e. $20log10(\alpha)<0,2$.

6.2.5 First test of equalization and filter adjustment for inversion error compensation

The quality of the equalization is assessed by playing back the reference noise recording (see clause 8.1). The reproduced sound field is recorded with the microphones of the calibration as well as with those of the fine-tuning set and compared to the reference noise recordings. The error *E(f)* is calculated by first determining the amplitude error for each used microphone in 96^{th} octave bands, finding the minimum and the maximum error over frequency and calculating the mean of those.

As a last step, a correction filter D(f) compensates for the impact of the impulse response pre-processing on the inversion quality as well as arithmetic uncertainties (e.g. the regularization factor). The frequency response of this filter is calculated as the reciprocal of the previously calculated mean reproduction error *E(f)*:

$$D(f) = \frac{1}{E(f)}. \tag{10}$$

For all frequencies below 1 800 Hz, the level is compensated with the average of D(f) in the frequency range between 50 Hz and 1 800 Hz. Above 1 800 Hz, a minimum phase version of D(f) is applied to achieve causality and minimum additional delay.

This correction function is applied to all filters.

6.2.6 Accuracy of the equalization

After a successful equalization the following criteria shall be met:

1) Level accuracy

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

2) Magnitude and phase of the cross correlation between broadband noise and simulated broadband noise at the fine-tuning position

- The magnitude of the cross correlation between the alternative reference noise recording (see clause 8.1) and the reproduced signals at the fine-tuning position averaged over the individual microphones shall fulfil the following requirements:
    - In the frequency range from 100 Hz to 1 kHz the magnitude of the complex coherence (normalized cross correlation spectrum) for the microphones 3 to 8, which are used for the equalization in this frequency range, shall be larger than 0,9, measured in 1/3rd octaves.
    - In the frequency range from 100 Hz to 1 kHz the phase of the complex coherence for the microphones 3 to 8, which are used for the equalization in this frequency range, shall be accurate within ±10 degrees and within ±30 degrees in the range from 1 kHz to 1,5 kHz, both measured in 1/3rd octaves.

NOTE: Microphones 1 & 2 are required for the equalization process in the low frequency domain however, due to their position far away from the device under test as well as due to their use in the low frequency domain only no requirements are set for these microphones.

3) Spectrum reproduction accuracy

- The difference of the amplitude spectrum of the original alternative reference noise recording and the amplitude spectrum of the reproduced broadband noise (both measured in dB) shall be for the microphones 3 to 8 within ±3 dB, measured in 1/3rd octaves from 50 Hz to 10 kHz and ±6 dB from 10 kHz to 16 kHz. The average spectrum accuracy, averaged over all microphones shall be within ±3 dB from 50 Hz to 20 kHz.

6.3 Accuracy of the reproduction arrangement

6.3.0 Introduction

During the validation process a variety of experiments was conducted in order to get an estimate of the reproduction accuracy of the multi-point sound field recording and reproduction arrangement. This validation was conducted in various steps.

6.3.1 Comparison between original sound field and simulated sound field

The comparison to the original sound field is an inherent part of the equalization technique. Figure 12 shows the differences of the microphone levels as well as the complex coherence between the real sound field and the simulated sound field after equalization at the relevant microphone positions of the calibration set for four different types of rooms.

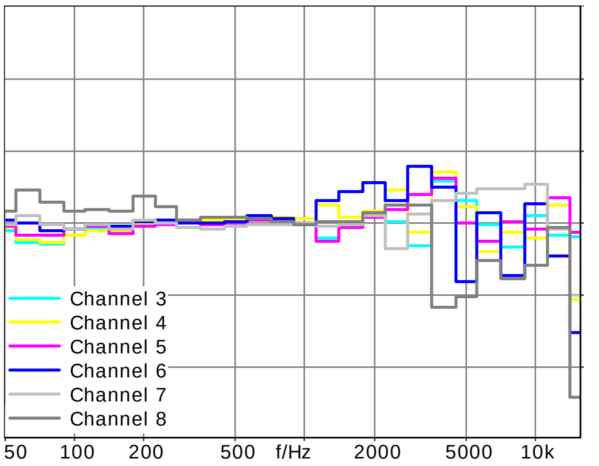
Table 3: Types of rooms the setup was validated in

Name	RT60	C80
Room1	27 ms	50,25 dB
Room2	123 ms	36,84 dB
Room3	98 ms	45,6 dB
Room4	264 ms	20,3 dB

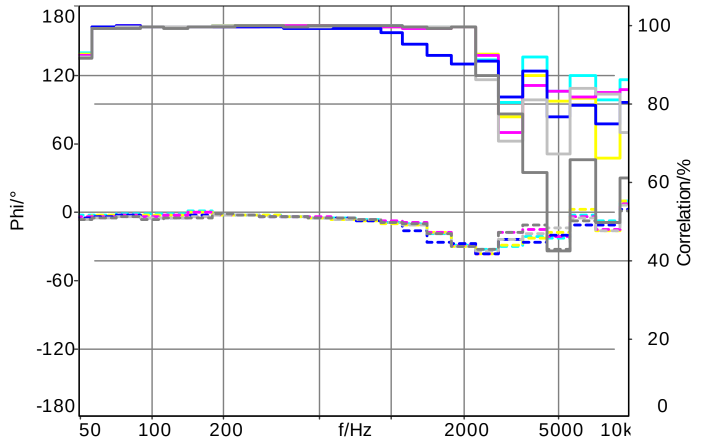
In all of the rooms the equalization was successful because the level difference between reference and simulated sound field is within the range from -3 to +3 dB and also the magnitude of the complex coherence is greater than 90 % in the frequency range 100 Hz to 1 000 Hz.

Level difference (3rd octave)

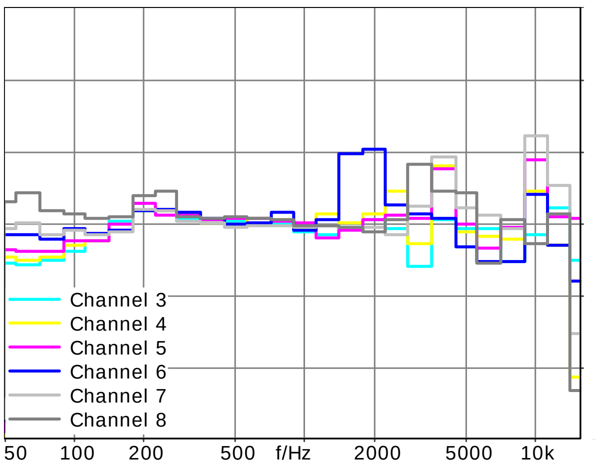
Room1 (RT60=27 ms)

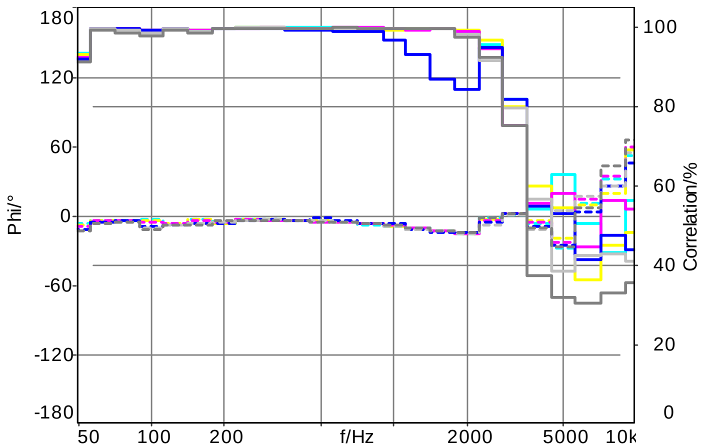


Complex Coherence (3rd octave)



Room2 (RT60=123ms)





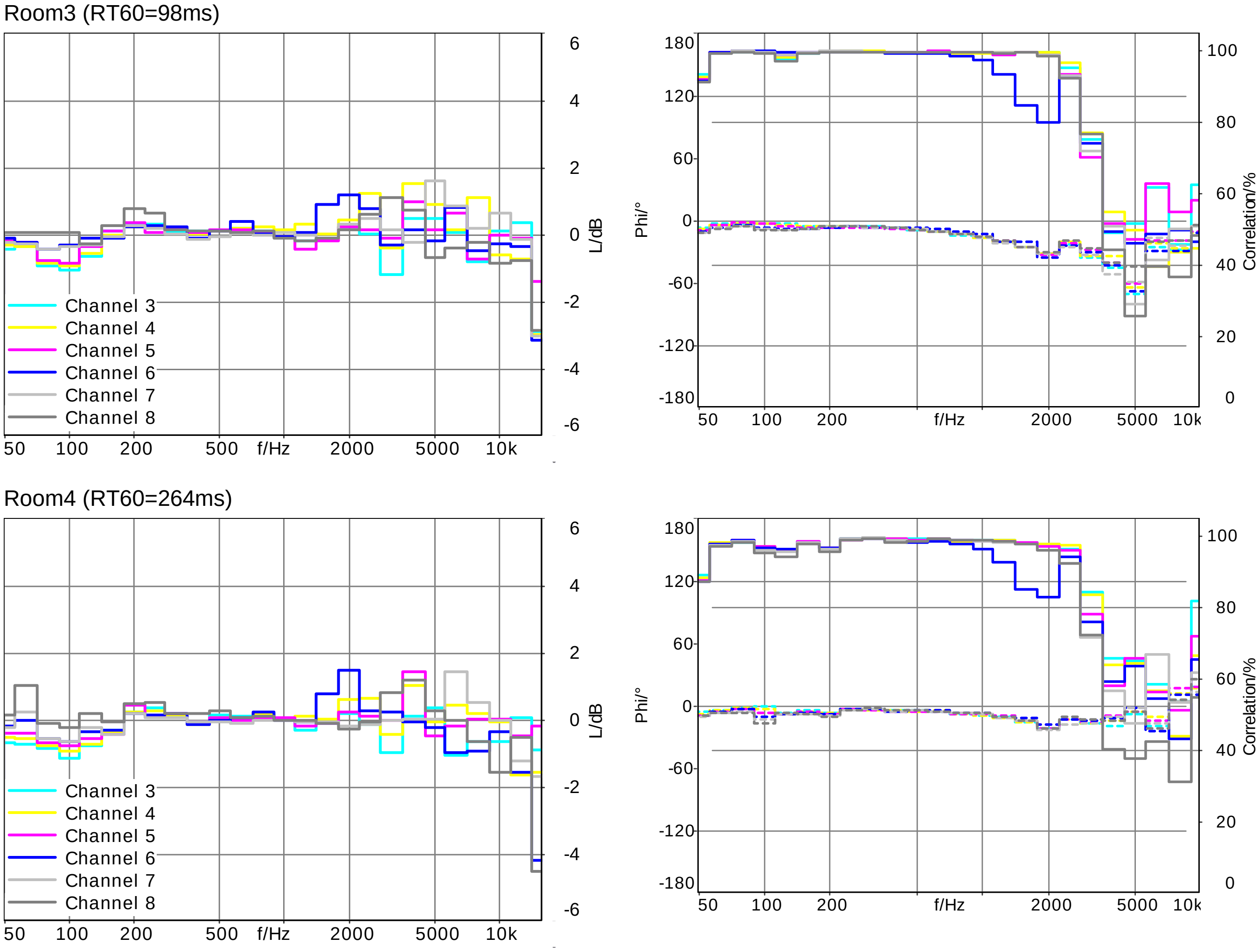


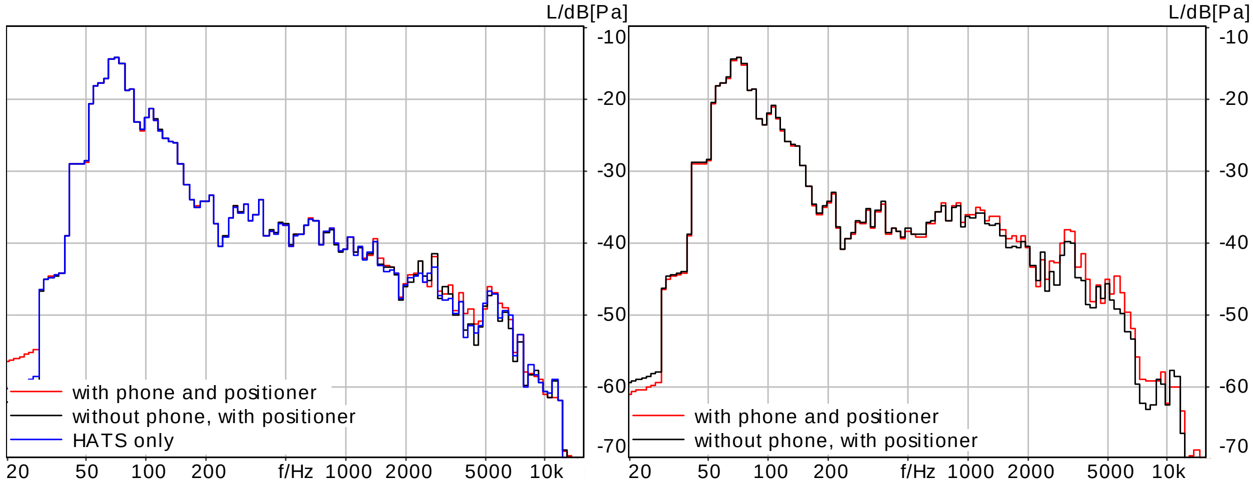
Figure 12: Level difference between reference and simulated sound field as well as the complex coherence in magnitude (solid) and phase(dashed)

6.3.2 Impact of handset positioner and phone on the simulated sound field

The impact of the handset-positioner, and the impact of the terminal itself, on the sound pressure produced by the multi-microphone based simulation were examined in more detail. Therefore a measurement microphone was positioned closely to the main microphone as well as closely to the rear microphone of a mobile phone using two-microphone based noise cancelling technique. A typical background noise was generated and the spectrum at these positions was measured. The same measurement was repeated with no mobile phone in place but with the microphones left at the same positions. For the main microphone position this result was repeated a second time without the handset positioned and without the mobile phone present. The results are given in the note below.

Figure 13 with a 12th octave resolution.

It can be seen that the impact of the handset positioned as well as the influence of the handset positioned and the device under test is very small and that it can be neglected in the setup procedure.



NOTE: The left diagram shows the deviations at the main microphone location whereas the right diagram shows the deviations at the rear microphone location.

Figure 13: Example for level deviations at the microphones of a typical two-microphone device

6.3.3 Comparison of terminal performance in the original sound field and the simulated sound field

6.3.3.1 Introduction

A reference sound field was generated in a reverberant room by eight loudspeakers located in a reverberant room not used for the sound field simulation. The loudspeakers were positioned at different locations as the loudspeaker used later for the sound field simulation. The only adjustment was a level adjustment so that all loudspeakers had the same level.

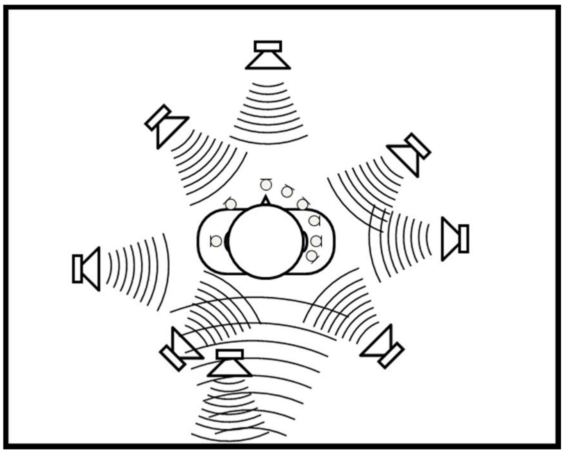


Figure 14: Setup for generating a reference sound field

Two different noise sources were used for sound field generation:

- Eight different car noise recordings on every loudspeaker (stationary), this recording is referred to as "carnoise".
- A combination of stationary noises like car noise with instationary noises like cafeteria noise, this recording is referred to as "mix".

First, those sound fields were recorded for later playback with the microphone array described in clause 5.2 for playback with the system presented in the present document.

After that, three different mobile phones with different properties were positioned in handset-mode. These are referred to as listed in table 4.

Table 4: Different phones which were used in the experiments

Name	Network-Mode	Size
Phone1	3G	4,8"
Phone2	4G	5,5"
Phone3	3G	4,3"

Two different measurements were conducted:

- the measurement of the transmitted background noise (see clause 6.3.3.2);
- S-/N-/G-MOS Analysis in accordance to [i.10] (see clause 6.3.3.3).

The MOS Analysis was carried out in three different positions of the Xe-position of the handset positioner, namely Xe = -20, 0, +25.

Two of those phones were also setup in hands-free-mode.

After those measurements had been carried out the playback system was equalized in four rooms as described in clause 6.2. In those rooms the same measurements as in the reference situation were conducted with a simulated background noise using the microphone array recording done previously. These results are discussed in clauses 6.3.3.2 and 6.3.3.3.

6.3.3.2 Background noise transmission

6.3.3.2.0 Validation Procedure

In a first step the different spectra of the transferred background noise are compared. First 30 seconds of speech were played back in addition to the background noise. Then the speech was switched off and 30 seconds of background noise were played back.

The simulation of background-noise was then recorded with a measurement microphone at the position of the main microphone of a mobile phone as it was also done in the reference-recording. This signal is called "unprocessed". In addition to that the sending signal of the mobile phone was also recorded. This signal is called "processed".

The parts where only background noise was recorded was then analysed.

The wide curve represents the original spectrum whereas the other curves represent the simulations in different rooms.

6.3.3.2.1 Handset

The equalization setup was as described in clause 5.2.

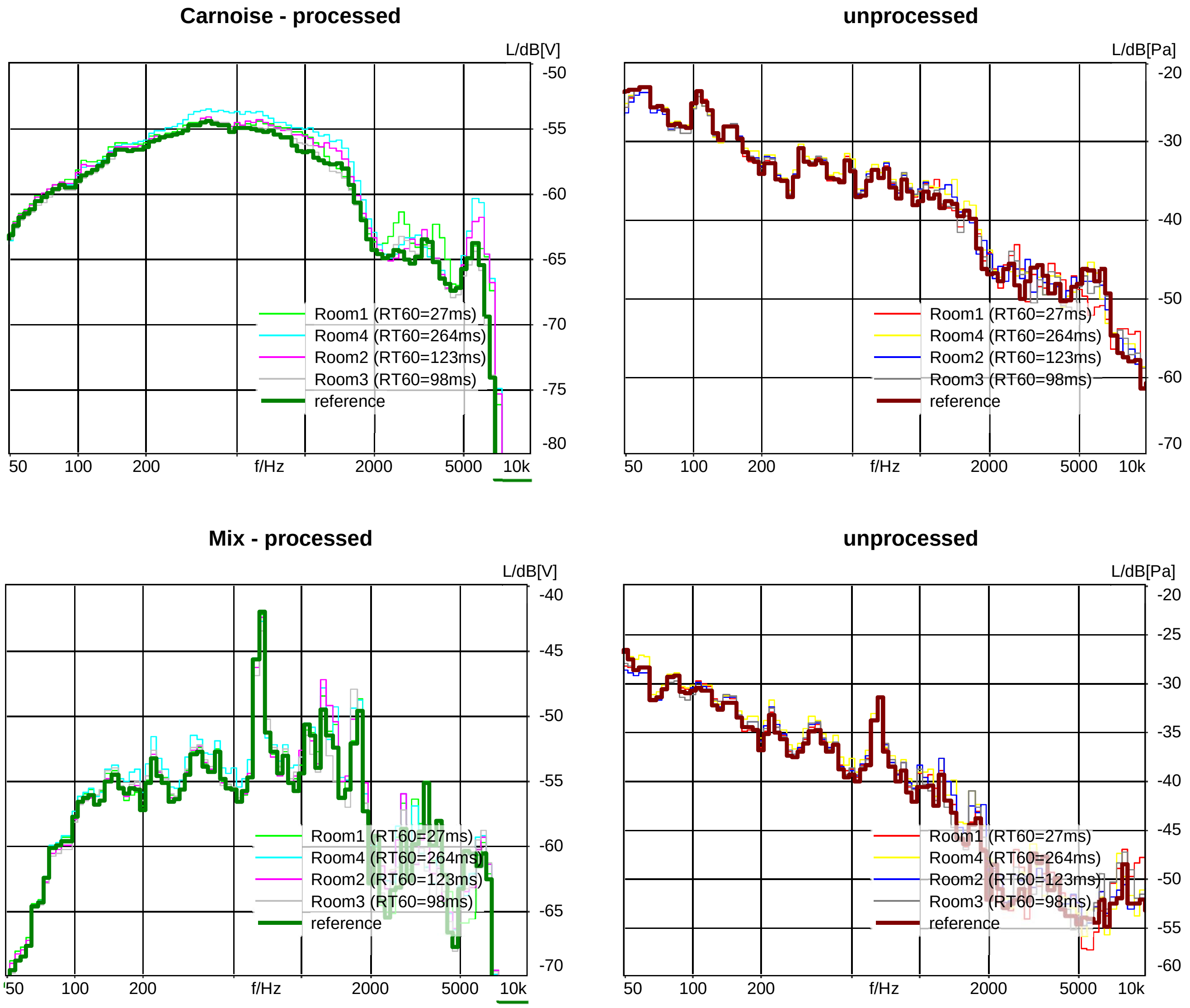


Figure 15: Phone1, handset mode, Comparison of 12th octave spectra for carnoise and mix-signal in different rooms (processed on the left, unprocessed on the right)

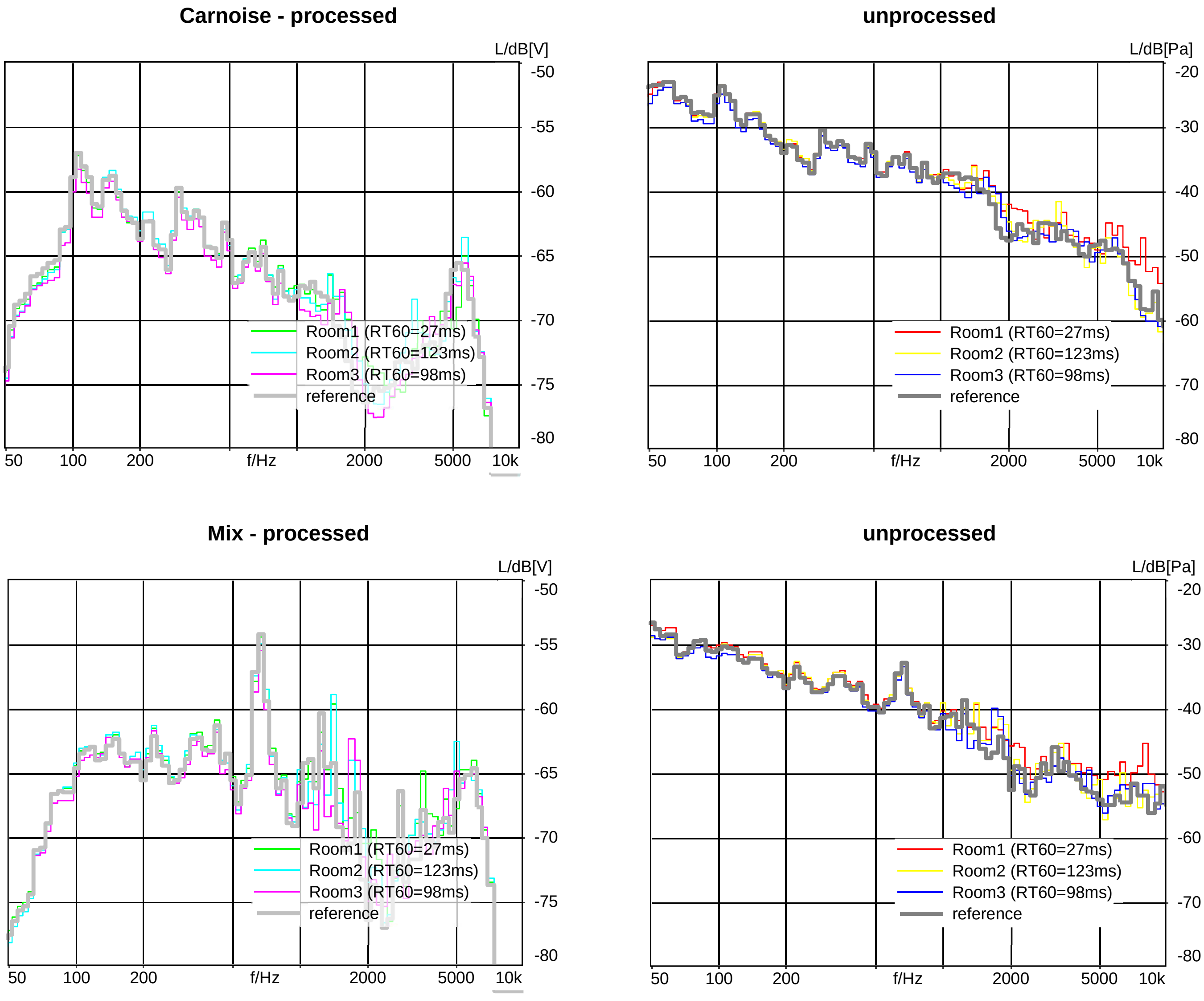


Figure 16: Phone2, handset mode, Comparison of 12th octave spectra for carnoise and mix-signal in different rooms (processed on the left, unprocessed on the right)

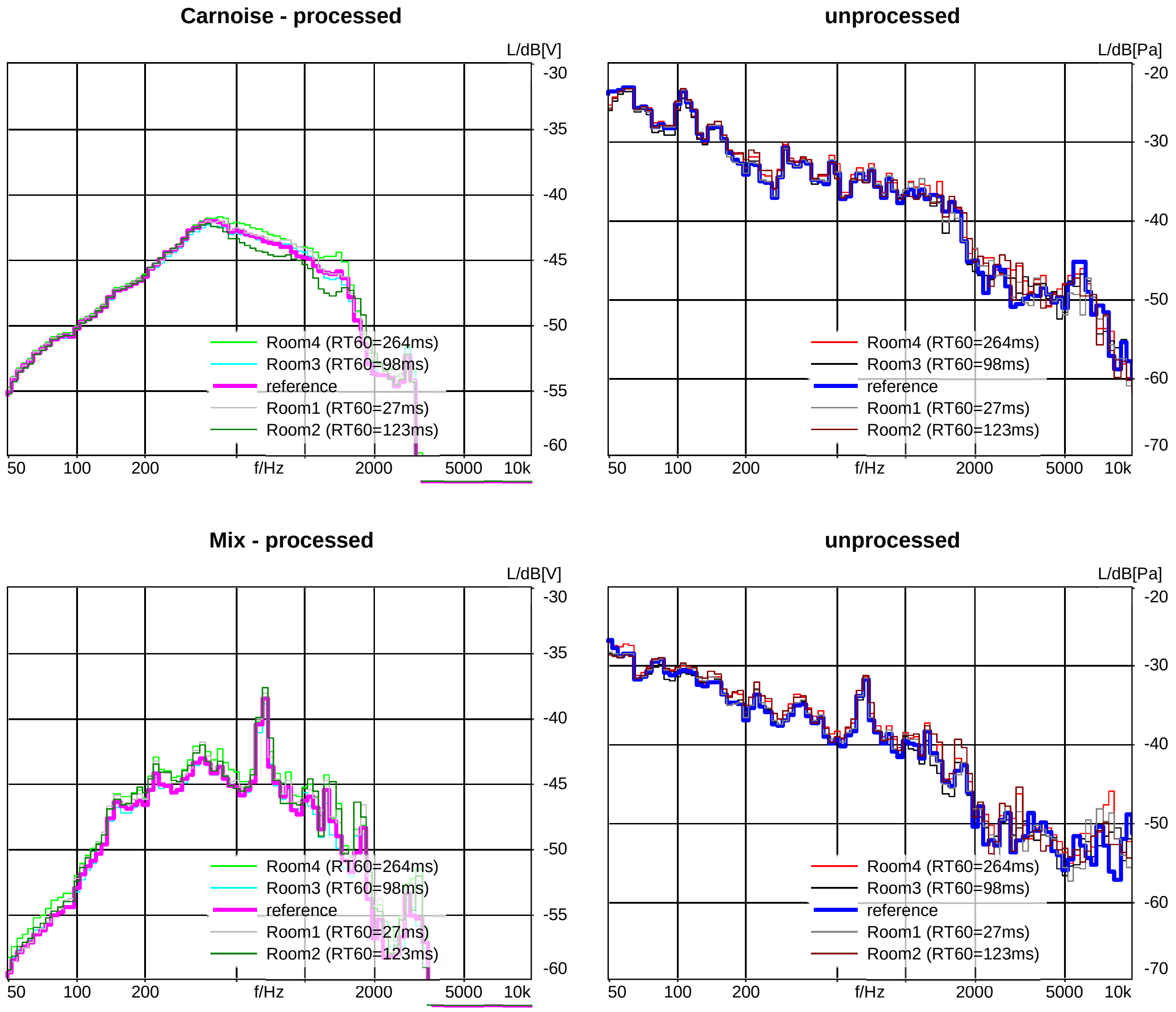
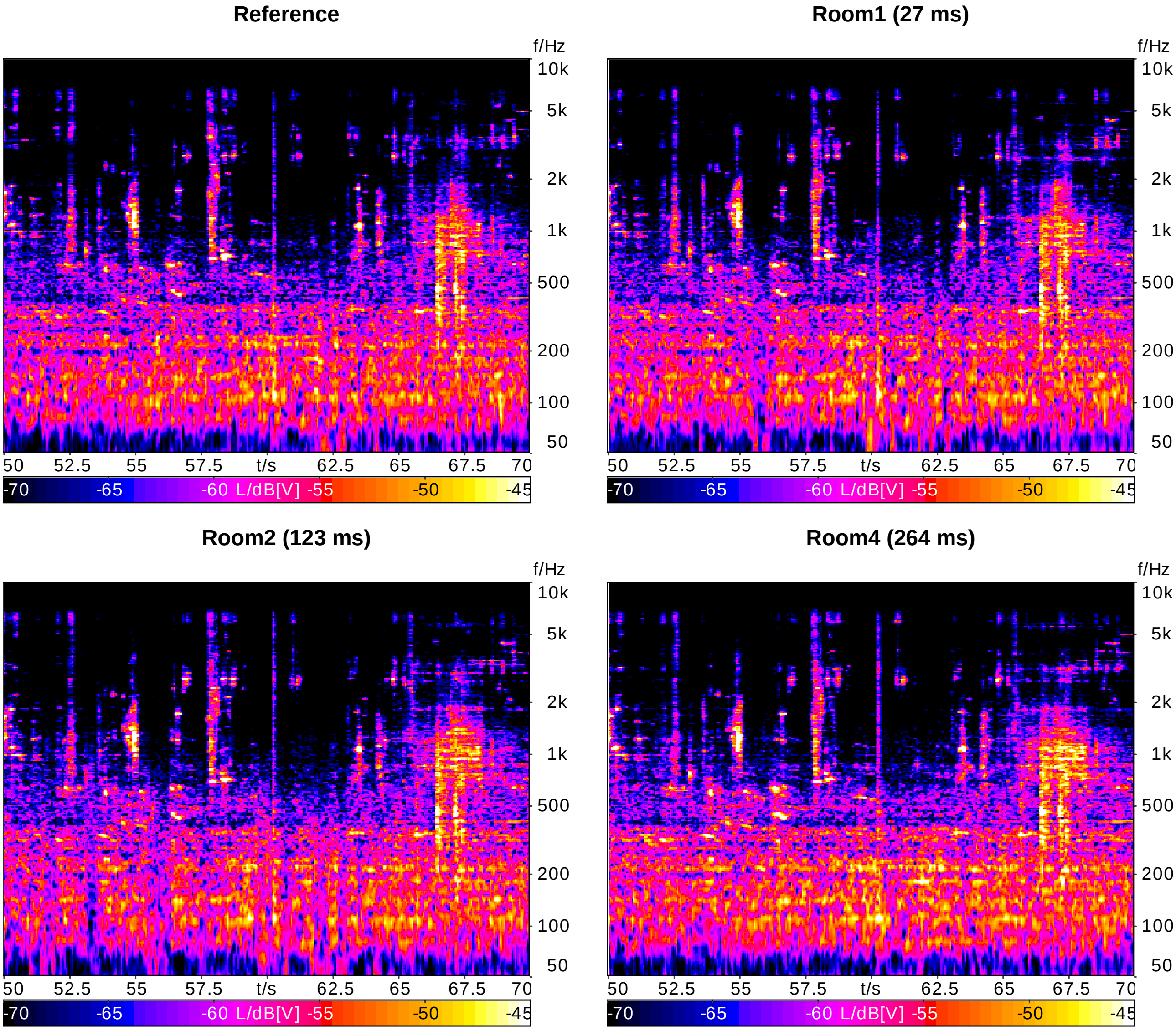


Figure 17: Phone3, handset mode, Comparison of 12th octave spectra for carnoise and mix-signal in different rooms (processed on the left, unprocessed on the right)

A further validation of the accuracy of the reproduction system when applied to a modern mobile phone is shown in Figure 18. The spectrograms of the same device tested in different rooms show a high degree of correlation not only in the spectral but in the time domain as well. The different transient sounds being part of the "mix" background noise are well reproduced independent of the test room used. It has to be noted that the phone incorporates certainly time-variant processing techniques which in general may lead to small deviations of the result between tests.



NOTE: The top left diagram shows the spectrum vs. time analysis of the reference recording.

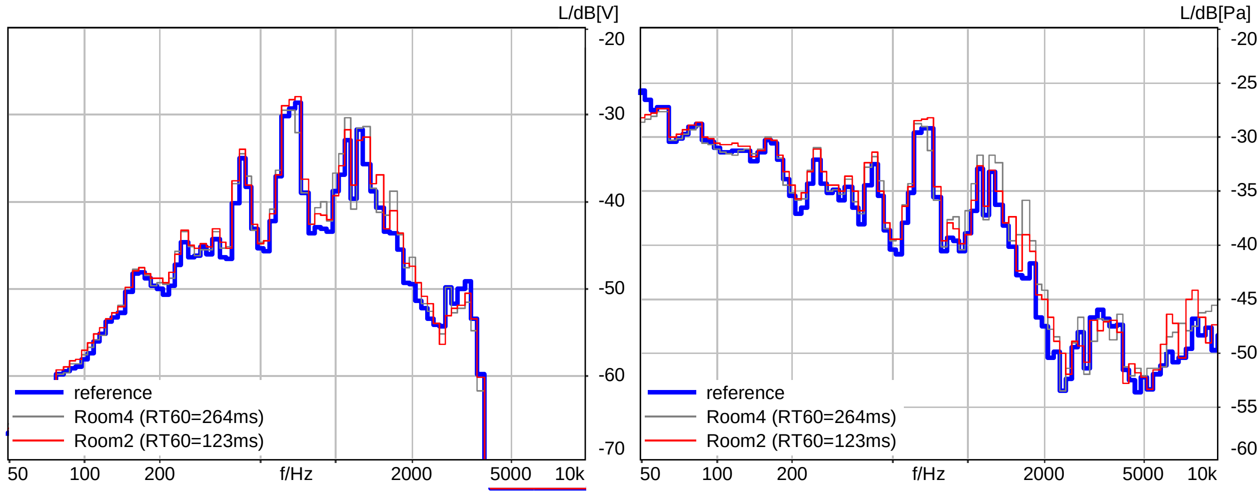
Figure 18: Spectrum vs. time analysis of "Mix"-signal for three rooms for Phone1

6.3.3.2.2 Handheld Hands-free

A hands-free equalization was conducted in two different rooms. The simulation of the "mix"-signal was then recorded with a measurement microphone at the position of the main microphone of a mobile phone as described in clause 5.3. Simultaneously the sending signal of the mobile phone which was limited to narrowband-mode was recorded. The results of this experiment can be seen in Figure 19.

It can be observed that the spectra measured in the different rooms are very close to the reference recording.

This confirms a good reproduction quality of the simulation arrangement for hand-held hands-free terminals as well.



NOTE: The left diagram shows the sending signal of the mobile phone whereas the right diagram shows the measured signal using a measurement microphone at the hands-free main microphone position.

Figure 19: Hands-free equalization - Comparison of 12th octave spectra of reference recording vs. simulated recording for the "mix"-signal

6.3.3.2.3 Desktop Hands-Free

An equalization for a desktop hands-free terminal was conducted in the reverberant room. The equalization arrangement was again setup as described in clause 5.3. The results can be seen in Figure 20. No big deviations can be observed.

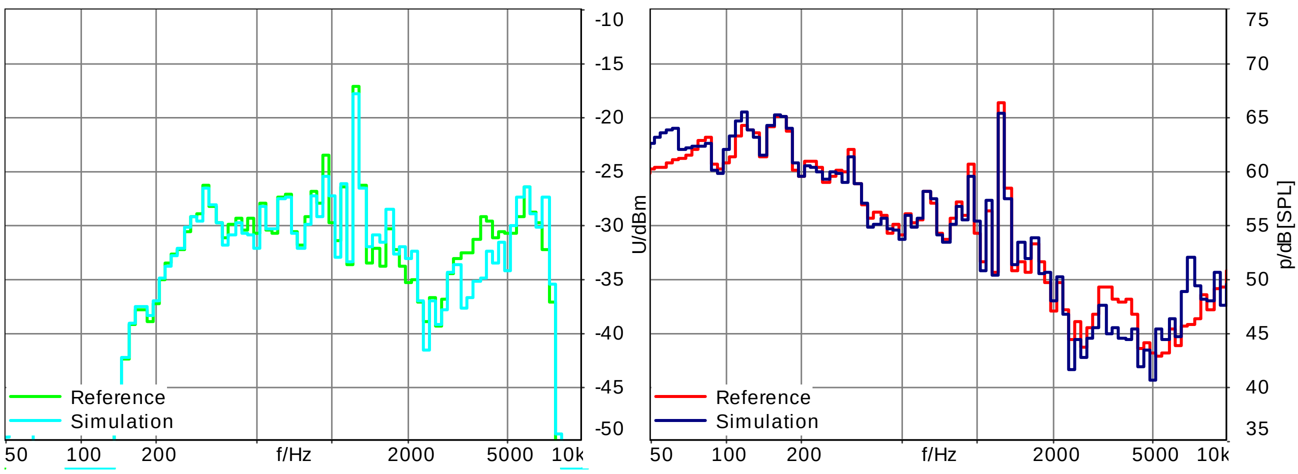


Figure 20: Desktop phone equalization - Comparison of 12th octave spectra of reference recording vs. simulated recording for the "mix"-signal

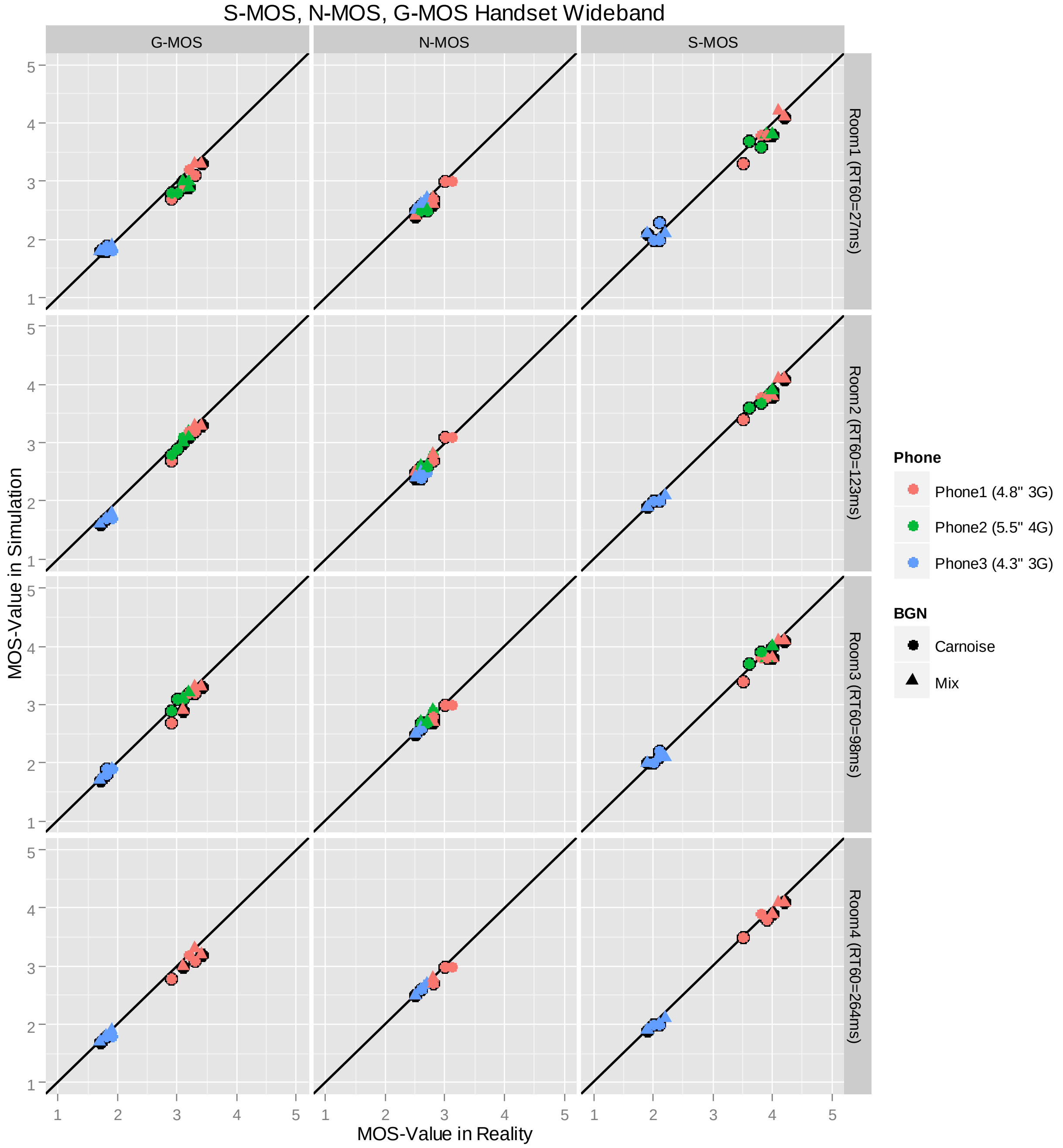
6.3.3.3 S-/N-/G-MOS Analysis according to ETSI TS 103 106

6.3.3.3.1 Handset

Figure 21 shows the MOS-Values measured in the simulation in relation to the MOS-Values measured in reality for the handset case. An excellent correlation between simulation and reality can be observed. The correlation factors are given in Table 5.

Table 5: Correlations for S-/N-/G-MOS for handset mode of different phones in different rooms

Name	G-MOS	N-MOS	S-MOS
Room1	98,9 %	90,4 %	98,9 %
Room2	99,6 %	93,3 %	99,7 %
Room3	99,2 %	90,3 %	99,5 %
Room4	99,5 %	97,4 %	99,7 %



NOTE: The phone was located in handset position. The results where Xe≠0 are marked black.

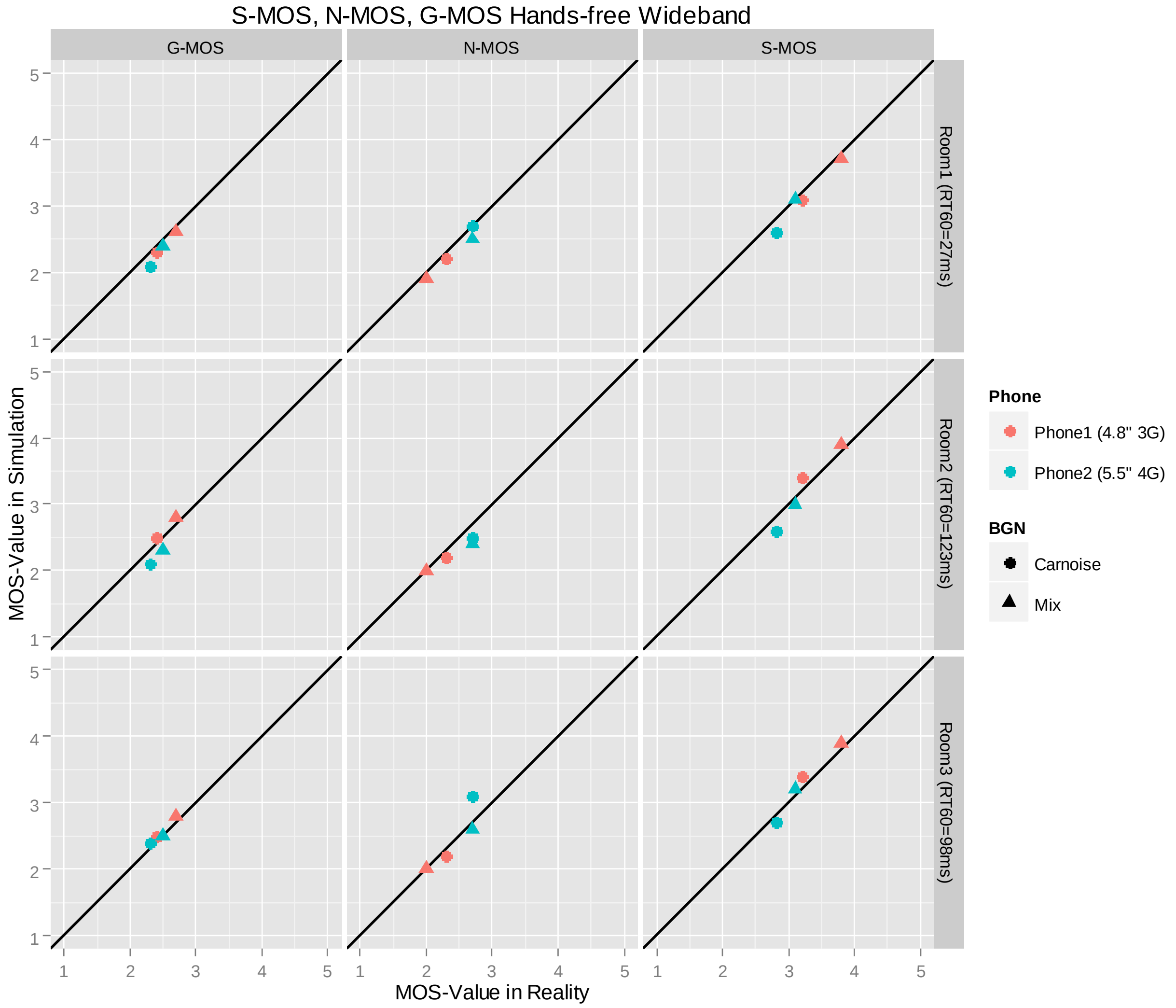
Figure 21: Results of the S-/N-/G-MOS measurements for different phones in different rooms

6.3.3.3.2 Hands-free

Figure 22 shows the MOS-Values measured in the simulation in relation to the MOS-Values measured in reality for the hands-free case. A good correlation between simulation and reality can be observed. The correlation factors are given in Table 6.

Table 6: Correlations for S-/N-/G-MOS for hands-free mode of different phones in different rooms

Name	G-MOS	N-MOS	S-MOS
Room1	98,4 %	97,2 %	98,4 %
Room2	86,6 %	98,2 %	96,8 %
Room3	95,7 %	89,2 %	97,6 %



NOTE: The phone was located in hands-free position.

Figure 22: Results of the S-/N-/G-MOS measurements for different phones in different rooms

7 Generalization of the method for a more flexible loudspeaker and microphone arrangement

7.0 Introduction

Clause 5 defines the loudspeaker equalization for a special microphone arrangement. In this clause a generalization of the above described method is given which extends the underlying equalization method for other, more flexible microphone or loudspeaker configurations.

7.1 Loudspeaker configuration

The generalized Multi-Point Noise Simulation Method requires at least 2 loudspeakers, preferably 4 loudspeakers or more. The used loudspeakers should not be placed too close to each other but should be as distributed as possible. In any case, the loudspeakers should not be located in the space between the microphones.

Dependent on the presentation room there might be problems to reproduce frequencies down to 50 Hz (i.e. the -3 dB point of the loudspeakers frequency response is much higher than 50 Hz). In this case, an additional subwoofer is needed and the weighted version of the Thikonov-equalization (Equation 9 in clause 6.2.4.1) is used during the inversion procedure, where the diagonal weighting matrix **W** accounts for the different frequency ranges of the loudspeakers. It is suggested to categorize the used loudspeakers into three groups as described in Table 7.

Table 7: Definitions of loudspeaker categories

Loudspeaker category	Frequency range
Subwoofer	40 Hz to 120 Hz
Large Loudspeaker	50 Hz to 22 000 Hz
Small Loudspeaker	120 Hz to 22 000 Hz

The values for the weighting matrix **W** are then 1 for the given frequency range and faded to 0 with cosine characteristic within a $1/3^{rd}$ octave transition.

7.2 Microphone setup

No fixed microphone arrangement is required, hence the setup can be adapted individually for every application. If possible, the output signal of the microphones of the device under test (DUT) can be used directly for equalization or otherwise additional measurement microphones are placed near the DUT microphones.

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

The number of microphones in the calibration set, i.e. the microphones which are used for the matrix inversion in clause 6.2.4, shall not exceed the number of loudspeakers in the setup excluding potential subwoofer. In general, it is recommended to have more full-range loudspeakers in the setup than microphones in the calibration set, to improve the accuracy of the equalization.

If it is not possible to use the DUT microphones directly, it is recommended to place one measurement microphone near to each DUT microphone. If special consideration of the directivity characteristic of the DUT microphone is necessary, at least two measurement microphones should be placed in the surrounding of the DUT microphone in order to more accurately simulate the soundfield characteristics in the area around the DUT microphone.

7.3 Background noise recordings and reference noise

Since the arrangement of the microphones is not fixed but differs between different applications and (possibly) different DUTs, no default background noise scenarios can be provided. Instead, individual background noise recordings shall be done for every application and microphone arrangement.

The equalization step of clause 6.2 requires a reference noise for the calculation of the inversion filters. The reference noise should be 5 to 10 s long and representative for the individual background noise recordings because the reference noise will be used for testing the equalization as well as for determining the maximum output volume of the loudspeakers. In consequence, one appropriate portion of the individual background noise recordings shall be selected as reference noise.

7.4 Equalization and calibration

The method described in clause 6.2 defines the following steps to be performed:

1) Separate level adjustment for every loudspeaker
2) System identification
3) Pre-processing of the impulse responses
4) Calculation of the inversion filters:
    a) Inversion procedure
    b) Different microphones for different frequency bands
    c) Search for the optimum regularization factor
5) First test of equalization and filter adjustment for inversion error compensation

However, for the generalized Multi-Point Noise Simulation Method some modifications are needed.

The matrix size (used to identify the system and to calculate the matrix inversion for equalization) depends on the number of loudspeakers and microphones chosen for the background noise simulation.

In clause 6.2.3 the impulse responses are pre-processed with a low-pass filter with a time-variant cut-off frequency to avoid a degradation of the quality of the inversion by high-frequency components of the tails of the impulse responses (step 3). This step reduces the accuracy of the equalization for higher frequencies. In this variant step 3 is considered as optional: As long as the tails of the impulse responses are rather short it is suggested to skip the step.

In clause 6.2.4.2 different microphones were selected to calculate the inversion filters for each frequency band based on the corresponding wavelength (step 4) b)). For the proposed generalized Multi-Point Noise Simulation Method, the microphone distances are unknown and therefore this type of optimization is not applied. All microphones are used across the whole frequency range to calculate the inversion filters.

In clause 6.2.4.3 the optimum regularization factor is searched subject to a limited loudspeaker output level to avoid distortion of the loudspeaker signals (step 4) c)). *"A spectrum representative for the maximum level and spectral content"* (see clause 6.2.4.3.1) can be calculated from the reference noise which was selected above. For the generalized method, the level of the filtered reference noise radiated by each loudspeaker to each microphone position shall not exceed the level which a pink noise of 0 dB [Pa] would generate in the same FFT-bin.

The extended methodology which aims to reduce the error at the so called fine-tuning positions is not applicable because the locations of the DUT microphones are well-known and therefore the calibration microphones can be located close to the DUT microphones.

In clause 6.2.5 a correction filter D(f) is applied to compensate i.e. for the arithmetic uncertainties (e.g. the regularization factor) of the inversion. For the generalized method, this correction filter consists of a constant factor below 1,8 kHz and a minimum phase filter above. For the generalized method, the microphone distances are unknown and therefore the cut-off frequency is reduced to 50 Hz. In order to retain the phase in the previously constant frequency range below 1,8 kHz, a zero-phase-filter is used instead of a minimum-phase filter for the whole frequency range above 50 Hz.

7.5 Accuracy of the equalization

After a successful equalization the following criteria, which are comparable to those of clause 6.2.6 shall be met:

1) **Level accuracy**

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

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

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

   - In the frequency range from 100 Hz to 1 kHz the magnitude of the complex coherence (normalized cross correlation spectrum) shall be larger than 0,9, measured in $1/3^{rd}$ octaves.
   - In the frequency range from 100 Hz to 1 kHz the phase of the complex coherence shall be accurate within ±10 degrees and within ±30 degrees in the range from 1 kHz to 1,5 kHz, both measured in $1/3^{rd}$ octaves.

3) **Spectrum reproduction accuracy**

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

7.6 Example use case: equalization inside a vehicle

7.6.0 Introduction

As an example for the application of the generalized method, this clause provides instructions to apply the method to loudspeaker equalization inside a car.

7.6.1 Loudspeaker configuration

As the generalized Multi-Point Noise Simulation Method is compatible with the loudspeaker arrangement of [i.9] (four loudspeakers plus subwoofer) it is suggested to use the same loudspeaker arrangement. As described above the required number of loudspeakers depends on the number of microphones. Accordingly the number of loudspeakers has to be increased if more microphones shall be used in the "calibration set".

7.6.2 Microphone setup

The microphones have to be arranged individually for each car's interior and for each application. Often microphones can be found at the rear view mirror or at the ceiling above the driver or the co-driver. As it was described previously, the measurement microphones should be positioned close to the vehicle's microphones. If the DUT microphone signals are accessible directly, those signals can be used instead.

7.6.3 Equalization

Since the reverberation time in vehicle interiors is negligible compared to treated office rooms, the low-pass filtering (step 3) is omitted.

8 Background noise database

8.0 Introduction

The background noise database is available separately from the present document. All files are 8 channel *.wav files with 48 kHz sampling rate and 24 bit resolution.

The background noises, the file names and the levels of the original signals are found in the list below. The information about the levels can be used to adjust the scaling of the *.wav 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://docbox.etsi.org/STQ/Open/TS%20103%20224%20Background%20Noise%20Database/.

8.1 Reference noise recording

For the inversion procedure described in clause 6.2.4.3.2 and for the validation of the inversion in clauses 6.2.5 and 6.2.6, a reference background noise is provided. For this background noise the sound pressure has to be known at the calibration set and at the fine-tuning set of microphones. Therefore a well-defined sound field was generated by using eight uncorrelated broadband pink noise sources. The total signal level is about 70 dB [SPL]. This sound field is then recorded with microphones in the calibration set and at the fine-tuning set. For hands-free/handset mode different recordings have been made. The frequency range of the noise signal covers all frequencies for which the system has to be equalized and is provided in the range from 20 Hz to 20 kHz.

8.2 Background noise signals for terminal testing

Table 8: Background noise signals for playback with the multi microphone method (levels in dBSPL(A))

Name	Description	Length	Handset Levels	Hands-free Levels
Inside Car Noise				
Mid-size car 80 km/h (*MidSizeCar_80*)	HATS and microphone array at co-drivers position	30 s	1: 66,5 dB 2: 66,4 dB 3: 66,8 dB 4: 67,3 dB 5: 67,4 dB 6: 67,7 dB 7: 67,7 dB 8: 68,4 dB	1: 67,1 dB 2: 67,7 dB 3: 67,4 dB 4: 67,6 dB 5: 67,4 dB 6: 67,5 dB 7: 67,7 dB 8: 68,2 dB
Mid-size car 100 km/h (*MidSizeCar_100*)	HATS and microphone array at co-drivers position	30 s	1: 65,7 dB 2: 65,2 dB 3: 65,4 dB 4: 66,0 dB 5: 66,2 dB 6: 66,6 dB 7: 66,6 dB 8: 67,5 dB	1: 65,1 dB 2: 65,9 dB 3: 65,6 dB 4: 65,9 dB 5: 65,9 dB 6: 65,8 dB 7: 65,9 dB 8: 66,3 dB
Mid-size car 130 km/h (*MidSizeCar_130*)	HATS and microphone array at co-drivers position	30 s	1: 70,3 dB 2: 70,1 dB 3: 70.3 dB 4: 70,9 dB 5: 71,0 dB 6: 71,6 dB 7: 71,5 dB 8: 72,3 dB	1: 70,1 dB 2: 70,9 dB 3: 70,9 dB 4: 71,3 dB 5: 71,2 dB 6: 71,2 dB 7: 71,2 dB 8: 71,7 dB
Full-size car 80 km/h (*FullSizeCar_80*)	HATS and microphone array at co-drivers position	30 s	1: 60,5 dB 2: 60,3 dB 3: 60,8 dB 4: 61,4 dB 5: 61,8 dB 6: 62,4 dB 7: 62,7 dB 8: 64,0 dB	1: 61,5 dB 2: 60,5 dB 3: 60,5 dB 4: 60,5 dB 5: 60,5 dB 6: 60,7 dB 7: 61,1 dB 8: 61,4 dB
Full-size car 100 km/h (*FullSizeCar_100*)	HATS and microphone array at co-drivers position	30 s	1: 63,5 dB 2: 63,0 dB 3: 63,6 dB 4: 64,3 dB 5: 64,8 dB 6: 65,4 dB 7: 65,7 dB 8: 67,1 dB	1: 64,0 dB 2: 63,3 dB 3: 63,4 dB 4: 63,4 dB 5: 63,4 dB 6: 63,3 dB 7: 63,6 dB 8: 64,0 dB
Full-size car 130 km/h (*FullSizeCar_130*)	HATS and microphone array at co-drivers position	30 s	1: 68,5 dB 2: 68,3 dB 3: 68,8 dB 4: 69,5 dB 5: 69,9 dB 6: 70,5 dB 7: 70,8 dB 8: 71,9 dB	1: 69,5 dB 2: 68,6 dB 3: 68,6 dB 4: 68,7 dB 5: 68,8 dB 6: 68,8 dB 7: 69,2 dB 8: 69,7 dB
Inside Train Noise				
Inside Train (*Inside_Train*)	HATS and microphone array in passenger cabin of a train	30 s	1: 68,5 dB 2: 68,4 dB 3: 68,2 dB 4: 68,4 dB 5: 68,4 dB 6: 68,7 dB 7: 68,7 dB 8: 69,6 dB	1: 68,2 dB 2: 67,8 dB 3: 67,9 dB 4: 68,0 dB 5: 67,8 dB 6: 67,6 dB 7: 67,6 dB 8: 67,8 dB
Inside Bus Noise				
Inside Bus (*Inside_Bus*	HATS and microphone array in passenger cabin of a bus	30 s	1: 70,5 dB 2: 71,9 dB 3: 71,7 dB 4: 72,1 dB 5: 72,5 dB 6: 72,9 dB 7: 73,1 dB 8: 73,2 dB	1: 72,8 dB 2: 72,5 dB 3: 72,2 dB 4: 72,3 dB 5: 72,2 dB 6: 72,3 dB 7: 72,7 dB 8: 73,4 dB
Outside Traffic Street Noise				
Roadnoise (*Roadnoise*)	HATS and microphone array standing outside near a road	30 s	1: 72,8 dB 2: 71,6 dB 3: 72,0 dB 4: 72,9 dB 5: 72,2 dB 6: 73,1 dB 7: 73,0 dB 8: 73,8 dB	1: 69,9 dB 2: 70,7 dB 3: 70,9 dB 4: 71,0 dB 5: 70,8 dB 6: 70,8 dB 7: 70,9 dB 8: 71,0 dB
Crossroadnoise (*Crossroadnoise*)	HATS and microphone array standing outside near a crossroad	30 s	1: 70,6 dB 2: 70,2 dB 3: 70,5 dB 4: 71,4 dB 5: 70,6 dB 6: 71,4 dB 7: 71,2 dB 8: 71,6 dB	1: 69,9 dB 2: 69,6 dB 3: 69,6 dB 4: 69,9 dB 5: 69,6 dB 6: 69,5 dB 7: 69,6 dB 8: 69,7 dB

Public Places Noise				
Cafeteria (*Cafeteria*)	HATS and microphone array inside a cafeteria	30 s	1: 70,0 dB 2: 70,0 dB 3: 70,1 dB 4: 70,7 dB 5: 70,5 dB 6: 70,8 dB 7: 70,6 dB 8: 71,0 dB	1: 69,0 dB 2: 69,7 dB 3: 69,6 dB 4: 69,8 dB 5: 69,5 dB 6: 69,5 dB 7: 69,7 dB 8: 70,0 dB
Departure platform (*TrainStation*)	HATS and microphone array on the departure platform of a train station	30 s	1: 78,9 dB 2: 78,8 dB 3: 79,1 dB 4: 80,0 dB 5: 79,4 dB 6: 79,6 dB 7: 78,8 dB 8: 80,1 dB	1: 78,7 dB 2: 78,6 dB 3: 78,5 dB 4: 78,6 dB 5: 78,5 dB 6: 78,4 dB 7: 78,5 dB 8: 78,5 dB
Pub Noise (*Pub*)	HATS and microphone array in a pub	30 s	1: 77,2 dB 2: 76,6 dB 3: 75,7 dB 4: 76,0 dB 5: 76,0 dB 6: 76,3 dB 7: 76,0 dB 8: 76,4 dB	1: 75,2 dB 2: 75,1 dB 3: 74,9 dB 4: 75,1 dB 5: 74,8 dB 6: 74,8 dB 7: 74,8 dB 8: 75,0 dB
Sales Counter (*SalesCounter*)	HATS and microphone array in a supermarket	30 s	1: 66,6 dB 2: 66,1 dB 3: 65,7 dB 4: 66,5 dB 5: 66,3 dB 6: 66,8 dB 7: 66,6 dB 8: 67,1 dB	1: 65,5 dB 2: 65,3 dB 3: 65,2 dB 4: 65,5 dB 5: 65,6 dB 6: 65,3 dB 7: 65,2 dB 8: 65,3 dB
Recording in airport hallway (*Airport*)	Airport hallway with overhead public address announcement	50 s	1: 77,5 dB 2: 78,3 dB 3: 78,7 dB 4: 78,7 dB 5: 78,4 dB 6: 78,8 dB 7: 78,1 dB 8: 78,1 dB	1: 77,2 dB 2: 77,4 dB 3: 77,6 dB 4: 77,7 dB 5: 78,1 dB 6: 77,9 dB 7: 77,8 dB 8: 77,9 dB

Workplace Noise				
Callcenter 1 (*Callcenter*)	HATS and microphone array in business office	30 s	1: 72,6 dB 2: 71,8 dB 3: 71,3 dB 4: 71,4 dB 5: 70,5 dB 6: 70,6 dB 7: 70,5 dB 8: 70,9 dB	1: 70,4 dB 2: 69,9 dB 3: 69,8 dB 4: 69,9 dB 5: 69,7 dB 6: 69,6 dB 7: 69,6 dB 8: 69,8 dB
Callcenter 2 (Callcenter)	HATS and microphone array in business office	30 s	1: 60,2 dB 2: 60,0 dB 3: 60,1 dB 4: 60,8 dB 5: 60,2 dB 6: 60,6 dB 7: 60,2 dB 8: 60,7 dB	1: 59,3 dB 2: 59,3 dB 3: 59,5 dB 4: 59,6 dB 5: 59,4 dB 6: 59,3 dB 7: 59,3 dB 8: 59,5 dB

Name	**Description**	**Length**	**Handset Levels**	**Handheld Hands-free Levels**	**Desktop Hands-free Levels**
Conference Noise					
Conference1	HATS and microphone arrays at a table in a conference room with stationary noises	70 s	1: 45,5 dB 2: 45,3 dB 3: 45,3 dB 4: 45,5 dB 5: 45,6 dB 6: 45,6 dB 7: 45,6 dB 8: 45,4 dB	1: 44,7 dB 2: 44,6 dB 3: 44,5 dB 4: 44,6 dB 5: 44,6 dB 6: 44,6 dB 7: 44,6 dB 8: 44,6 dB	1: 52,1 dB 2: 52,1 dB 3: 50,2 dB 4: 49,4 dB 5: 49,0 dB 6: 48,2 dB 7: 48,7 dB 8: 48,6 dB
Conference2	HATS and microphone arrays at a table in a conference room with stationary and transient noises	70 s	1: 46,5 dB 2: 46,4 dB 3: 46,4 dB 4: 46,9 dB 5: 46,5 dB 6: 46,7 dB 7: 46,6 dB 8: 46,6 dB	1: 45,7 dB 2: 45,6 dB 3: 45,7 dB 4: 45,6 dB 5: 45,6 dB 6: 45,6 dB 7: 45,6 dB 8: 45,8 dB	1: 51.6 dB 2: 51.2 dB 3: 50.3 dB 4: 50.2 dB 5: 49.8 dB 6: 49.8 dB 7: 50.1 dB 8: 49.6 dB
Conference3	HATS and microphone arrays at a table in a conference room with stationary and transient noises	70 s	1: 50,6 dB 2: 50,4 dB 3: 50,5 dB 4: 51,1 dB 5: 50,5 dB 6: 50,7 dB 7: 50,5 dB 8: 50,9 dB	1: 50,1 dB 2: 50,0 dB 3: 50,0 dB 4: 50,1 dB 5: 50,0 dB 6: 50,0 dB 7: 50,0 dB 8: 50,1 dB	1: 54,3 dB 2: 55,0 dB 3: 54,2 dB 4: 53,7 dB 5: 53,8 dB 6: 52,5 dB 7: 52,8 dB 8: 52,5 dB

Name	Description	Length	Handset Levels	Hands-free Levels
Music				
Recording of Orchestra warmup in Opera House	HATS and microphone array at listener's position	30 s	1: 76,1 dB 2: 77,3 dB 3: 77,7 dB 4: 78,1 dB 5: 76,8 dB 6: 76,2 dB 7: 75,5 dB 8: 74,5 dB	1: 76,0 dB 2: 75,6 dB 3: 75,8 dB 4: 75,7 dB 5: 75,6 dB 6: 75,5 dB 7: 75,3 dB 8: 75,4 dB
Recording of music in a listening room (*RockMusic*)	HATS and microphone array at listener's position	198 s	1: 75,0 dB 2: 75,4 dB 3: 74,5 dB 4: 76,5 dB 5: 76,2 dB 6: 75,9 dB 7: 75,2 dB 8: 74,8 dB	1: 73,9 dB 2: 74,5 dB 3: 73,6 dB 4: 74,2 dB 5: 73,9 dB 6: 74,2 dB 7: 74,2 dB 8: 75,4 dB

Single Talker Recordings				
Recording of single talker, 0 degree (*SingleTalker 0deg*)	Single talker at distance of 1m from microphone array.	90 s	1: 61,2 dB 2: 62,9 dB 3: 63,5 dB 4: 63,5 dB 5: 62,9 dB 6: 62,0 dB 7: 61,0 dB 8: 60,1 dB	1: 61,3 dB 2: 60,9 dB 3: 60,6 dB 4: 60,7 dB 5: 60,5 dB 6: 60,5 dB 7: 60,4 dB 8: 60,7 dB
Recording of single talker, -22 degrees (*SingleTalker -22deg*)	Single talker at distance of 1m from microphone array.	90 s	N/A	1: 61,2 dB 2: 61,0 dB 3: 60,9 dB 4: 61,0 dB 5: 60,7 dB 6: 60,7 dB 7: 60,6 dB 8: 60,8 dB
Recording of single talker, +22 degrees (*SingleTalker +22deg*)	Single talker at distance of 1m from microphone array.	90 s	1: 58,9 dB 2: 61,1 dB 3: 62,6 dB 4: 63,1 dB 5: 62,6 dB 6: 62,1 dB 7: 61,2 dB 8: 60,0 dB	N/A

History

Document history		
V1.1.1	August 2014	Publication
V1.2.1	August 2015	Publication
V1.3.1	July 2017	Publication