ETSI TR 126 921 V17.0.0 (2022-05)

TECHNICAL REPORT

Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Investigations on ambient noise reproduction systems for acoustic testing of terminals
(3GPP TR 26.921 version 17.0.0 Release 17)

Reference
RTR/TSGS-0426921vh00

Keywords
5G,LTE,UMTS

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

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.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**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.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document provides a collection of investigations on aspects of UE performance in the presence of ambient noise related to the 3GPP terminal testing specifications TS 26.132 [3] and/or performance requirements per TS 26.131 [2].

The evaluation of terminals in conjunction with speech signals and realistic background noise tests as well as performance requirements were added in Release 11 of these specifications. Increased signal processing capabilities facilitated more sophisticated noise reduction functionality in terminals already before this release. However, an incorrect or mistuned device may substantially impact speech quality, intelligibility/listening effort and user experience in general.

The present document focuses on the evaluation of terminals in handset and hands-free mode, but also other related aspects like, e.g., accuracy of noise field simulations are taken into account.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.131: "Terminal acoustic characteristics for telephony; Requirements".

[3] 3GPP TS 26.132: "Speech and video telephony terminal acoustic test specification".

[4] 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".

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

[6] 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".

[7] Recommendation ITU-T P.56, "Objective measurement of active speech level".

3 Definitions of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

music clarity: ratio of the early sound energy (between 0 and 80 ms) and the late sound energy (that arrives later than 80 ms), typically reported in dB.

reverberation time: time in seconds required for the level of the sound to drop 60 dB after the sound source is turned off.

speech clarity: ratio of the early sound energy (between 0 and 50 ms) and the late sound energy (that arrives later than 50 ms), typically reported in dB.

3.2 Symbols

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

C50	Speech Clarity
C80	Music Clarity
dB	Decibel
RT60	Reverberation Time

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

AbsMax	Absolute Maximum Difference
ASL	Active Speech Level
DUT	Device Under Test
G-MOS	Global MOS (overall quality)
HHHF	Handheld Hands-free
N-MOS	Noise MOS (intrusiveness of noise)
NB	Narrowband telephonometry
MOS	Mean Opinion Score
RMSE	Root-Mean-Square Error
S-MOS	Speech MOS (distortion of speech)
SWB	Super-wideband telephonometry
WB	Wideband telephonometry

4 Ambient Noise Testing in Hands-free Mode

4.1 Introduction

The investigations described in the following clauses refer to ambient noise testing of terminals in handheld and desktop hands-free mode.

4.2 Results of Round Robin Test on Different Background Noise Simulation Techniques

4.2.1 Introduction

The following clauses summarize the results of a Round Robin test conducted between April and August 2015. Besides providing the results, the reproduction accuracy of the different simulation methods in different labs is analysed. Furthermore, the impact of the different analysis methods on the results measured with different mobile terminals is shown. Corresponding meeting temporary documents are provided in Annex A.

The aim of the Round Robin test conducted under the ATeMPO_SPINE work item was mainly to answer the following questions:

- How good is the reproducibility of the different methods?

- Are there differences in the measured performance when using the different devices with different setups?

The following labs were participating in the test:

- Audience Inc. / Knowles Inc.
- HEAD acoustics GmbH
- Orange
- Sony Mobile Communications

In the Round Robin test, seven different devices were used, some labs tested only a subset of six devices as agreed upon in advance. Not all labs conducted the full set of tests in narrowband (NB) and wideband (WB). At HEAD acoustics premises, two rooms were used to conduct the tests; in all other labs, the tests were conducted in one room.

The two noise field simulations according to ETSI ES 202 396-1 [4] and ETSI TS 103 224 [5] were investigated for each lab.

For measurements using the background playback according to ES 202 396-1 [4], occasional playback problems were observed in lab 2.1 (see also clause 4.2.3.2.1), which explains the outliers. The overall tendencies observed in the complete data set are not strongly affected by these issues. On the contrary, it can be noted that for this lab, MOS results were in general consistent between playback methods when using the same noise scenarios, while other labs results showed a slight unexpected offset between the methods. Occasional playback problems were also observed in lab 3.1 (see attached Annex A).

4.2.2 Test Setup

4.2.2.1 Laboratories

The labs according to Table 1 participated in the round robin test. Typical room acoustic parameters are provided as well here.

Table 1: Laboratories of the round robin test

Label	Lab	HATS Rotation [°]	Rev. Time (RT60)	Clarity Index (C80)
Lab 1.1	HEAD acoustics Room 1	0	125 ms	37.1 dB
Lab 1.2	HEAD acoustics Room 2	0	240 ms	20.5 dB
Lab 2.1	Sony	22.5	139 ms	30.0 dB
Lab 3.1	Audience	22.5	117 ms	29.2 dB
Lab 4.1	Orange	0	89 ms	40 dB
Lab 4.2	Orange	22.5	89 ms	40 dB

NOTE 1: For some labs, a rotation of 22.5° of the HATS was chosen in order to avoid obstructions of the sound (at rotation 0°, the HATS is exactly between a certain loudspeaker and the DUT).

NOTE 2: Lab 4.1 and 4.2 represent identical measurement chambers and describe different rotations of the HATS. In some of the following clauses, results may be presented either for both or only one of the setups.

4.2.2.2 Noise field simulations

In order to minimize all variabilities coming from other sources than the different setups and rooms used, the test material and the test procedure was prepared by HEAD acoustics and provided to all test labs. All labs used the HEAD acoustics software ACQUA with the HAE-BGN and 3PASS background noise systems with the identical database. All measurements were collected in a final database containing the results of all labs. Based on this procedure, all types of analyses as needed for the Round Robin experiment could be performed.

The selection of the rooms was up to the test labs. However, the requirements as found in [4] and [5] were followed. The detailed test setup is described in the individual reports of the test labs (see Annex A).

Figure 1 and Figure 2 show the general test setup for the two background noise simulation methods.

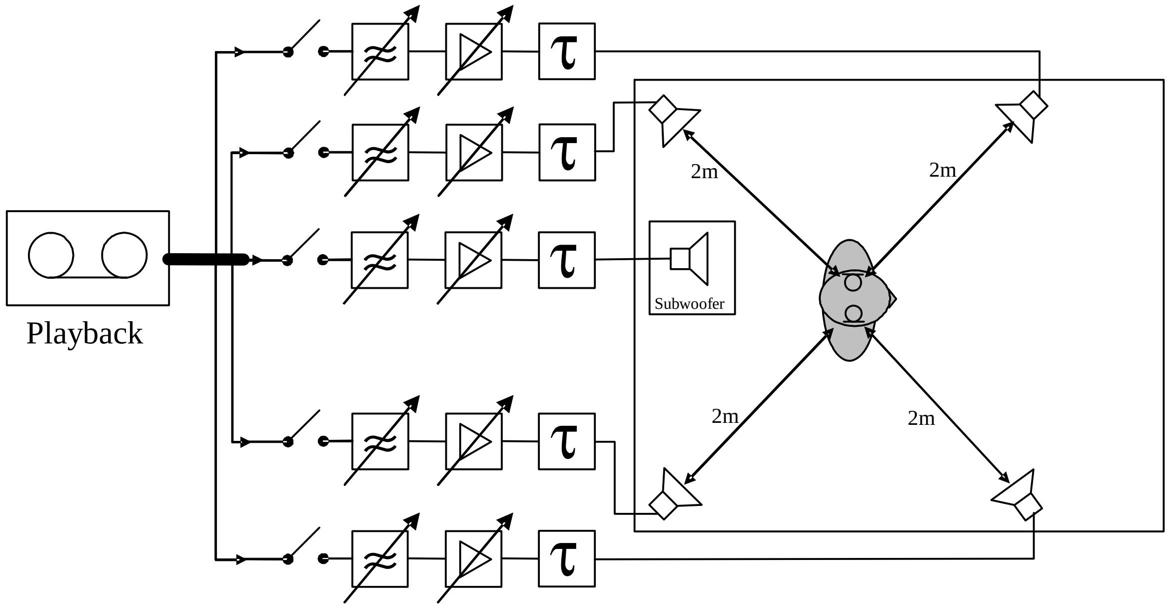


Figure 1: General setup of the 4+1 loudspeaker arrangement according to ES 202 396-1

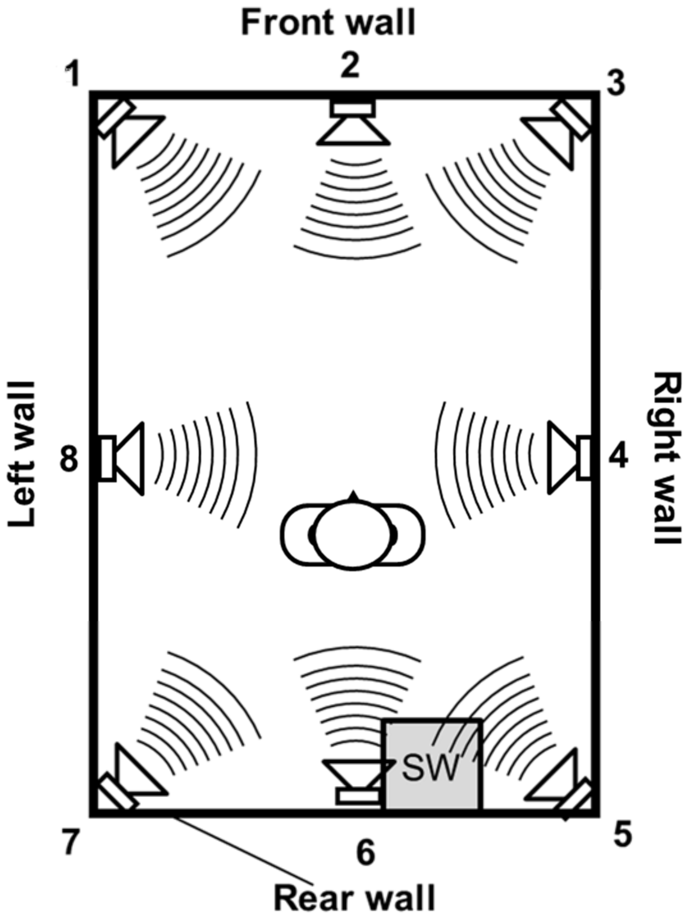


Figure 2: General setup of the 8-loudspeaker arrangement according to TS 103 224

The calibration of the setups was made by each lab individually. The calibration procedures as described in [4] and [5] are implemented in the HAE-BGN and 3PASS software and were followed by each lab. The calibration results can be found in the individual test reports of the test labs (see Annex A)..

4.2.2.3 Noise Types

For testing with the noise field simulation according to ETSI ES 202 396-1 [4], the six binaural noises according to Table 2 were used.

Table 2: Noise types for playback acc. to ES 202 396-1

Alias	Filename in database of [4]
Pub	Pub_Noise_binaural_V2
Crossroad	Outside_Traffic_Crossroads_binaural
Trainstation	Train_Station_binaural
Inside Car	Fullsize_Car1_130Kmh_binaural
Cafeteria	Cafeteria_Noise_binaural
Office	Work_Noise_Office_Callcenter_binaural

For testing with the noise field simulation according to ETSI TS 103 224 [5], the six noises according to Table 3 were used. The corresponding eight-channel recordings are also binaurally available in the noise database, which allows the usage of both systems with the identical noise type. Each of these noises was selected as a counterpart to a binaural recording listed in Table 2. Comparisons between the counterparts can be made based on the alias columns in both tables.

Table 3: Noise types for playback acc. to TS 103 224

Alias	Filename in database of [5]	
	8-channel	Binaural
Pub	Pub_handsfree	Pub_bin
Crossroad	Crossroadnoise_handsfree	Crossroadnoise_bin
Trainstation	Trainstation_handsfree	Trainstation_bin
Inside Car	FullSizeCar_130_handsfree	FullSizeCar_130_bin
Cafeteria	Cafeteria_handsfree	Cafeteria_bin
Office	Callcenter2_handsfree	Callcenter2_bin

4.2.2.4 Devices

Seven devices under test (DUT) according to Table 4 were available for the round robin test. Six phones were tested in all labs.

Table 4: Devices under test

Name	Size
DUT1	138.1 x 67 x 6.9 mm
DUT2	143.4 x 70.5 x 6.8 mm
DUT3	138.5 x 70.9 x 8.9 mm
DUT4	162.8 x 85.4 x 8.7 mm
DUT5	127.3 x 64.9 x 8.6 mm
DUT6	150.1 x 72.7 x 9.6 mm
DUT7	157.7 x 78.7 x 7.7 mm

4.2.2.5 Speech Quality Test Methodology

As source speech material for the evaluation of speech quality in the presence of background noise, 16 American English speech samples (fullband) according to ETSI TS 103 106 [6] were used for both bandwidth modes.

For terminals evaluated in NB mode, the analysis according to ETSI TS 103 106 [6] in NB-mode was carried out for each of the 16 samples and the results were averaged per background noise condition.

For terminals evaluated in WB mode, the analysis according to ETSI TS 103 106 [6] in WB-mode was carried out for each of the 16 samples and the results were averaged per background noise condition.

This usage of the instrumental assessment method provides S-MOS, N-MOS and G-MOS for each lab, bandwidth mode, device, and background noise.

4.2.3 Results for Handheld Hands-free

4.2.3.1 Introduction

In the following clauses, results of several measurements and analyses are presented for terminals in handheld hands-free mode. Without loss of generality, the results of lab 2.1 (one of the labs that conducted all experiments with all devices) are always compared to the results for the other labs. The test setup complies with clause 5.1.3.3 of [3]. Possible differences in results across labs may be influenced and explained by:

- Calibration differences
- Setup differences
- Room differences
- Time variant behaviour of the device under test

4.2.3.2 Comparison of inter-lab accuracy for the different background noise simulation methods

4.2.3.2.1 Outliers

Four outliers could be observed which could clearly be traced back to an error in the measurement setup and therefore do not represent any characteristics of the background noise generation method used. All diagrams in clause 4.2.3.1 (and corresponding sub-clauses) still contain these outliers, while the RMSE-values are calculated without outliers. The outliers are listed in Table 5.

Table 5: Outliers removed from analysis

Bandwidth	System	Noise Type	Lab	DUT
NB	ES 202 396-1	Inside Car	Lab 2.1	DUT2
NB	ES 202 396-1 with noises from TS 103 224	Inside Car	Lab 2.1	DUT2
WB	ES 202 396-1 with noises from TS 103 224	Inside Car	Lab 3.1	DUT2
WB	TS 103 224	Office	Lab 4.2	DUT7

4.2.3.2.2 Wideband

4.2.3.2.2.1 No background noise

In this clause, results under silent conditions in WB mode are presented, as shown in Figure 3 to Figure 5. Basically, the variance to be expected in different labs with no background noise simulation present can be observed.

It seems that the parameters described in clause 4.2.3.1 may impact the results to a certain extent providing the basis for the interpretation of the experiments with the background noise simulation methodologies. The main impact is on S-MOS resulting in somewhat scattered G-MOS results as well.

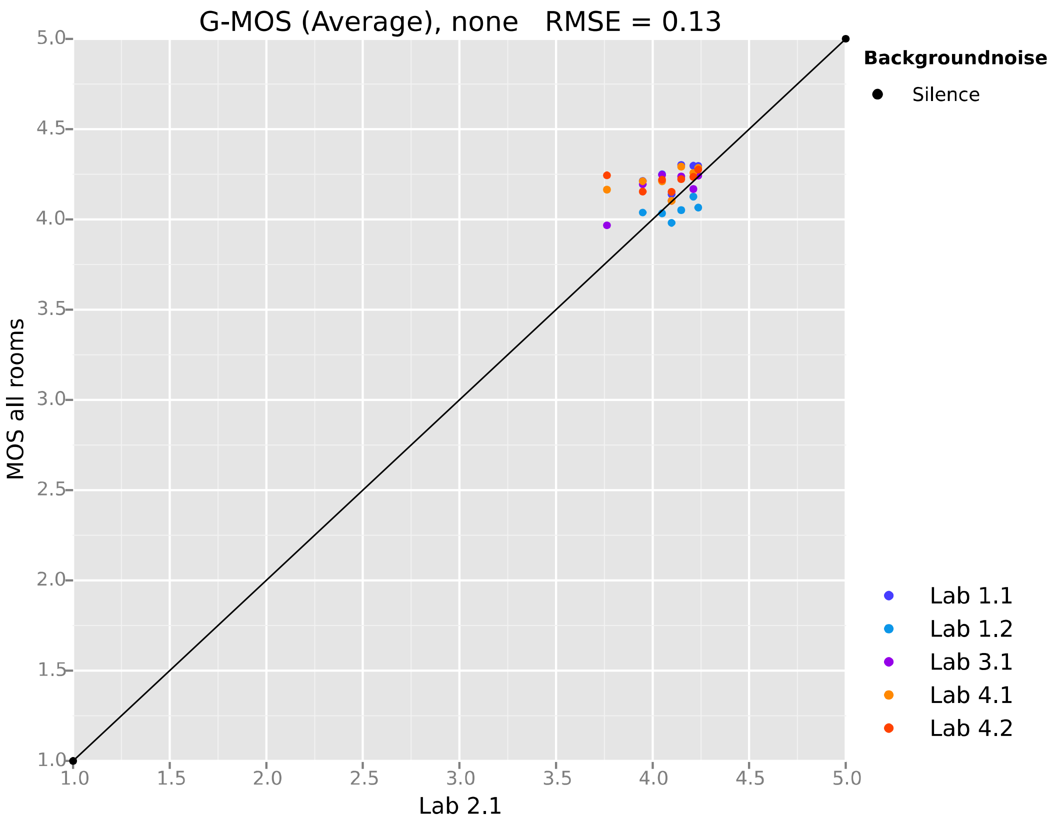


Figure 3: Correlation between G-MOS (WB) results from Lab 2.1 and other labs (Silence)

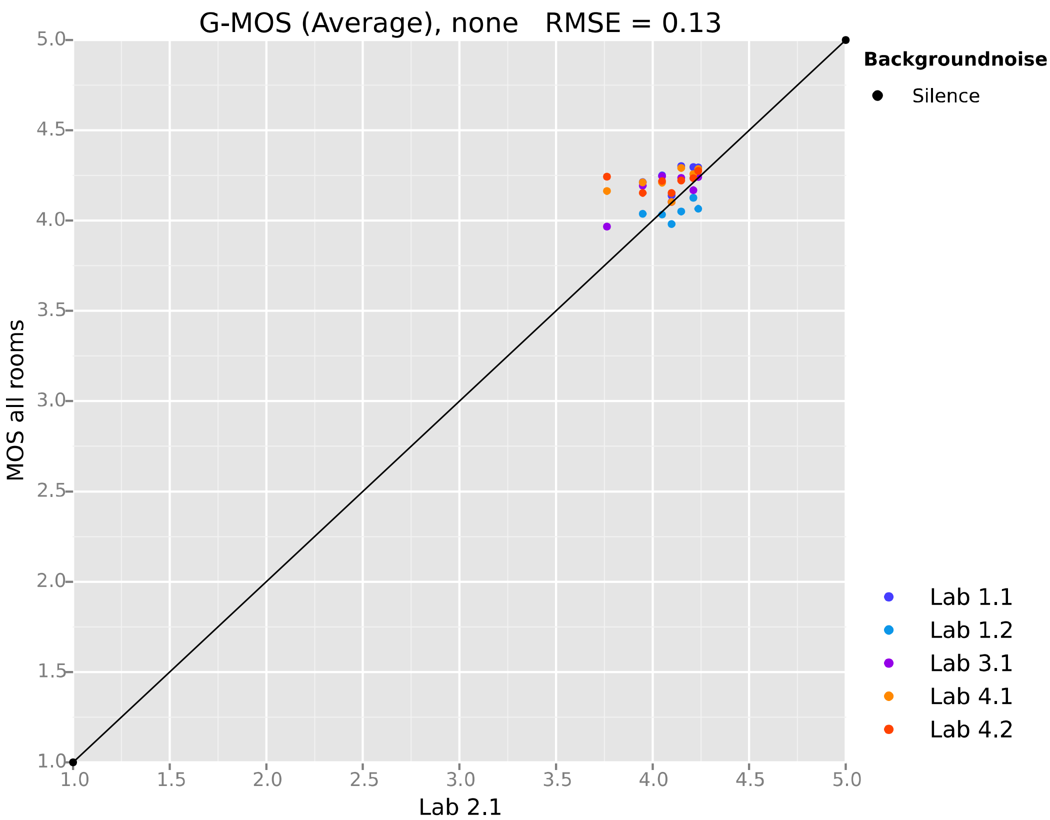


Figure 4: Correlation between S-MOS (WB) results from Lab 2.1 and other labs (Silence)

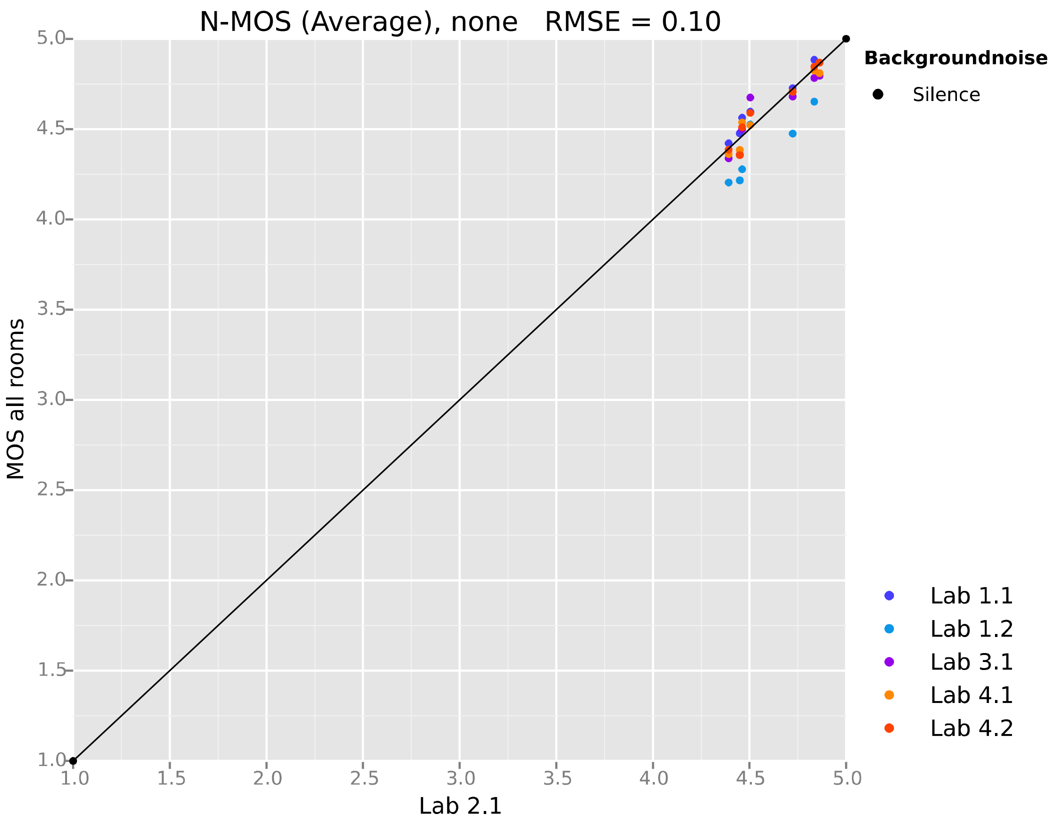


Figure 5: Correlation between N-MOS (WB) results from Lab 2.1 and other labs (Silence)

4.2.3.2.2.2 Background Noises & Simulation acc. to ES 202 396-1

The results shown in Figure 6 to Figure 8 are based on using the ES 202 396-1 noise field simulation, using the binaurally recorded background noises from the associated noise database. The following observations can be made:

- Compared to Lab 2.1 the results measured in the other labs are in general about 0.2 to 0.3 MOS higher for N-MOS and G-MOS
- The results are somewhat scattered leading to RMSE values between 0.13 and 0.22
- N-MOS shows the highest variation in results up to about 0.5 MOS

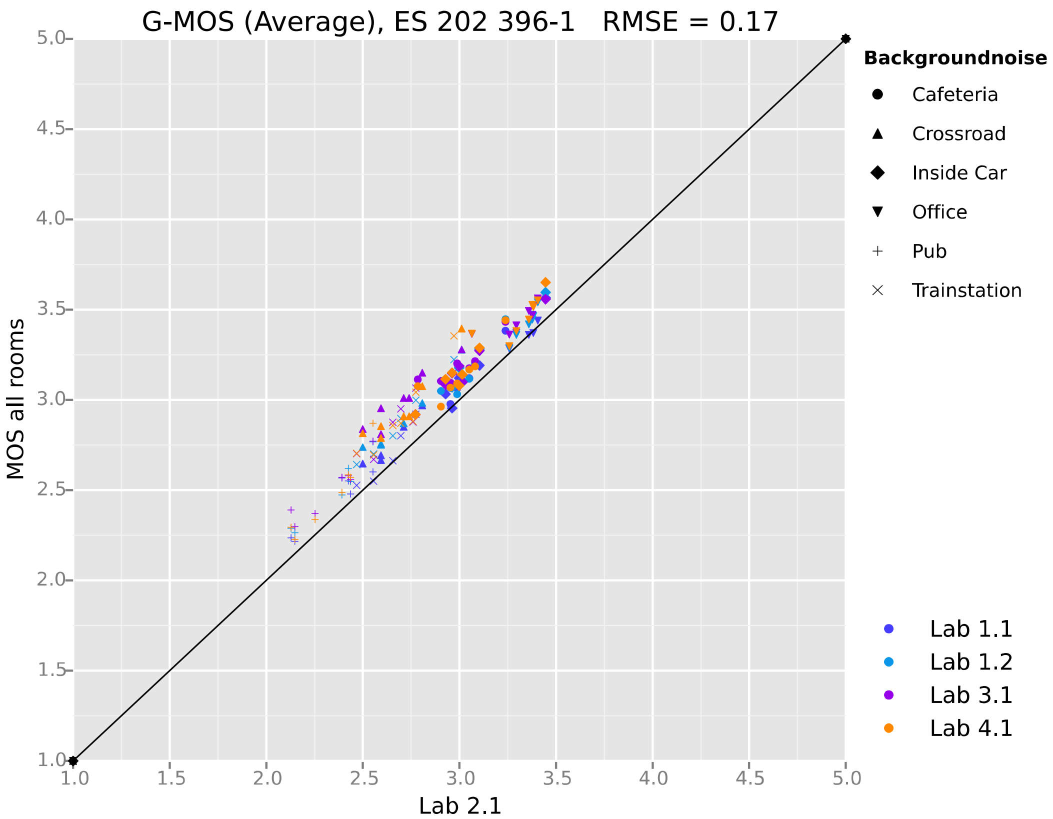


Figure 6: Correlation between G-MOS (WB) between Lab 2.1 and other labs

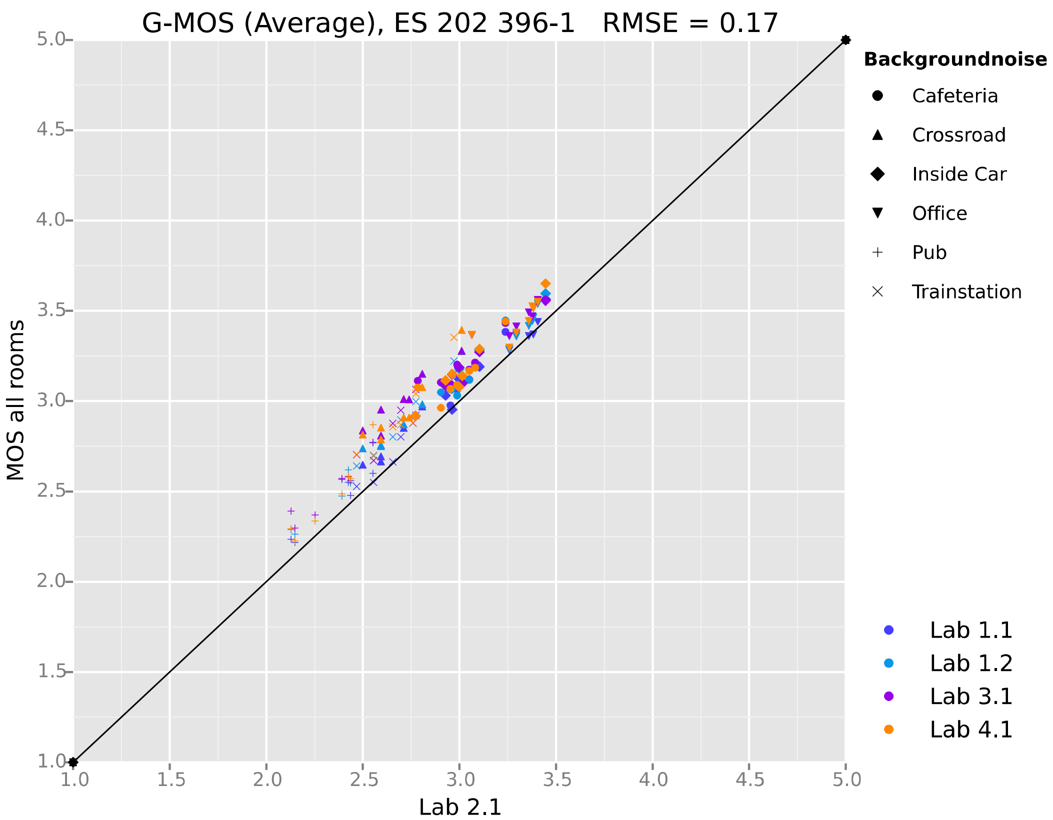


Figure 7: Correlation between S-MOS (WB) between Lab 2.1 and other labs

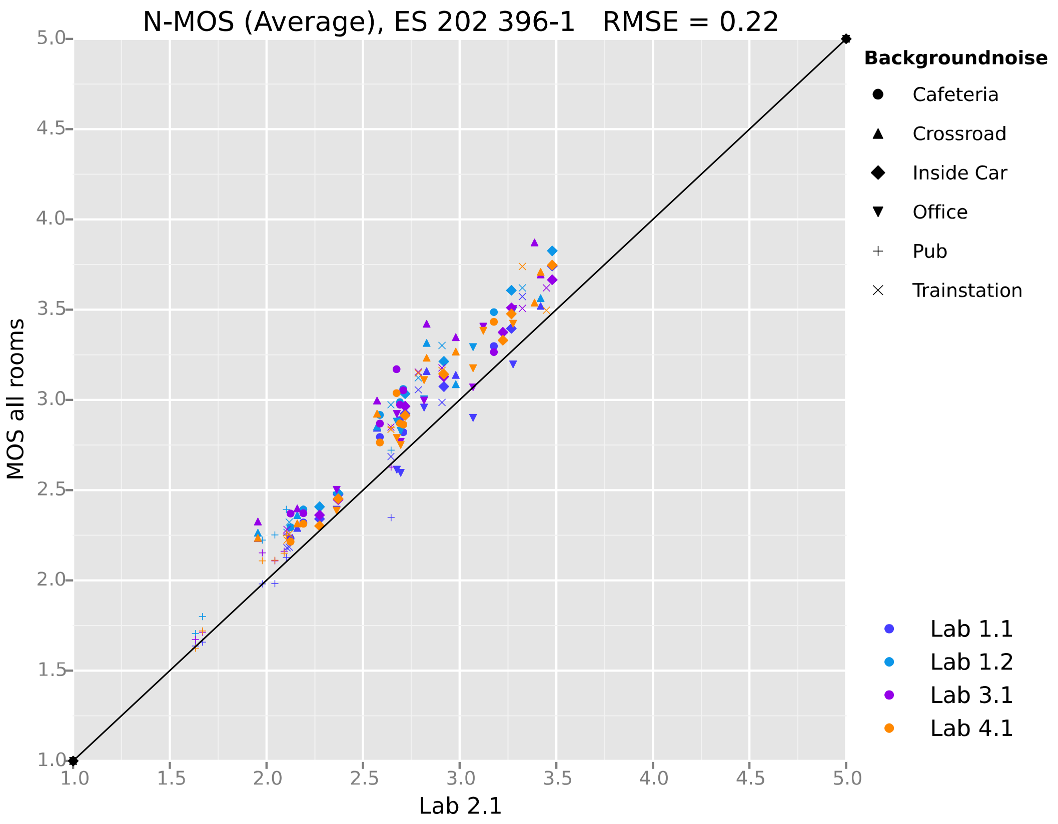


Figure 8: Correlation between N-MOS (WB) between Lab 2.1 and other labs

4.2.3.2.2.3 Simulation acc. to ES 202 396-1 with binaural noises from TS 103 224

The results shown in Figure 9 to Figure 11 are based on using the ES 202 396-1 noise field simulation, but using the binaurally recorded background noises from TS 103 224. The following observations can be made:

- Compared to Lab 2.1, the results measured in the other labs are in general about 0.3 to 0.4 MOS higher for S-MOS, N-MOS and G-MOS
- The results are somewhat scattered leading to RMSE values between 0.17 and 0.24
- N-MOS shows the highest variation in results up to about 0.5 MOS

Please note that the indicated outlier for test lab 3.1 is excluded from the RMSE calculations.

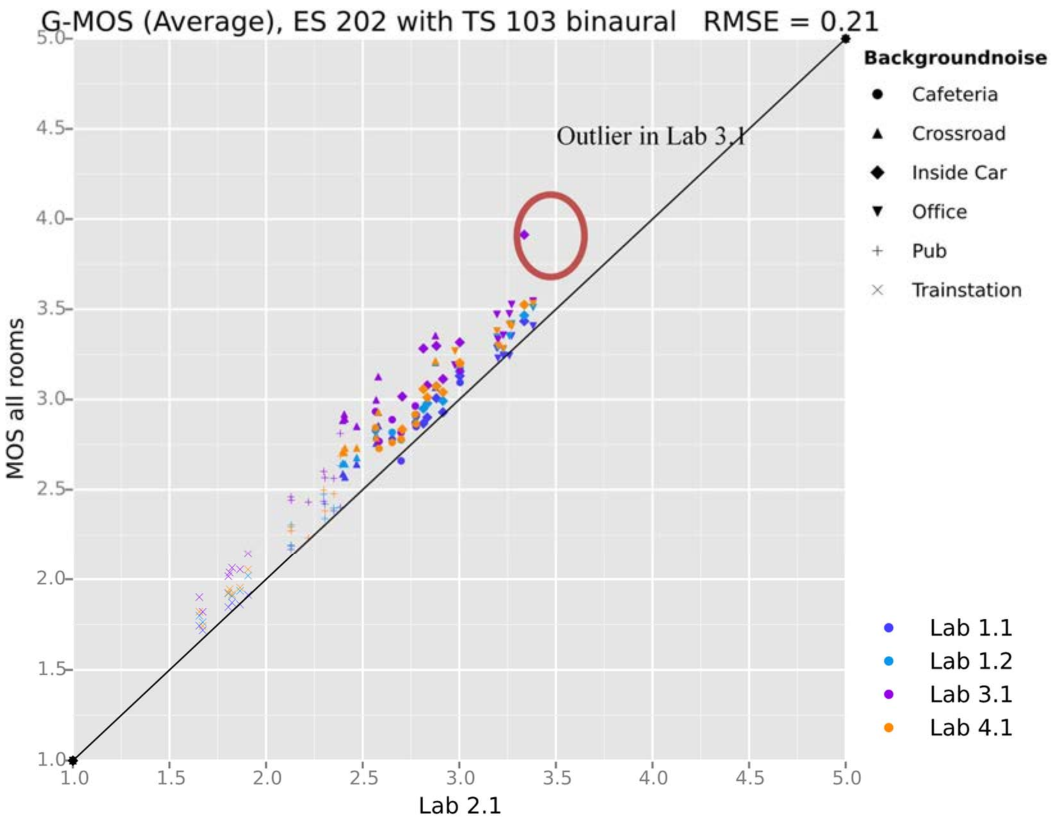


Figure 9: Correlation between G-MOS (WB) between Lab 2.1 and other labs

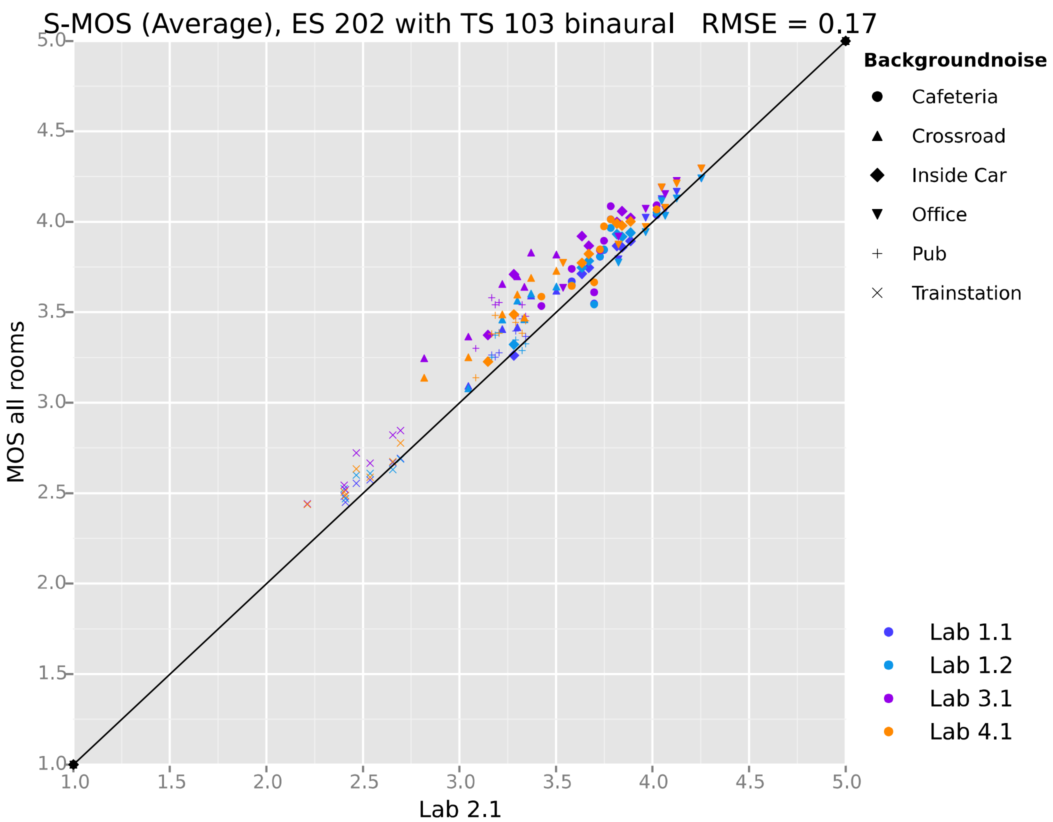


Figure 10: Correlation between S-MOS (WB) between Lab 2.1 and other labs

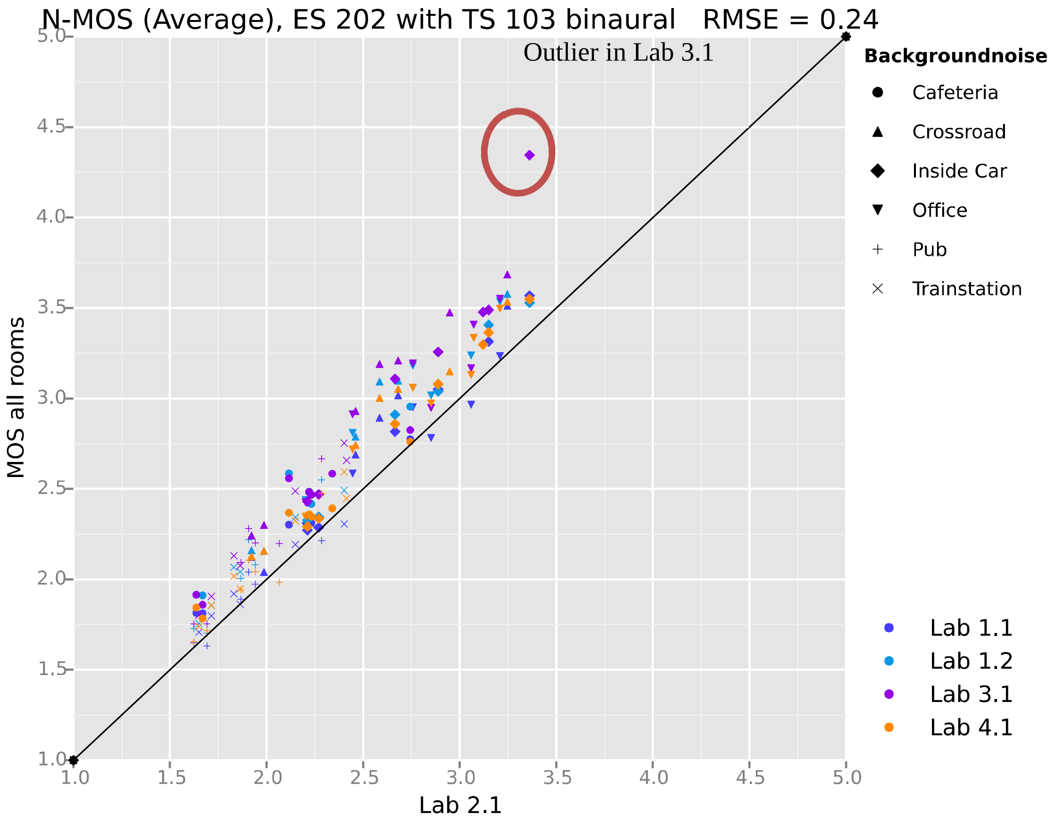


Figure 11: Correlation between N-MOS (WB) between Lab 2.1 and other labs

4.2.3.2.2.4 Background Noises & Simulation acc. to TS 103 224

The results shown in Figure 12 to Figure 14 are based on using the TS 103 224 noise field simulation, using the 8-channel noise recording from the associated noise database. For this setup, the following observations can be made:

- All results line up fairly well, no offset between the labs can be observed
- The results are less scattered leading to low RMSE values of 0.1
- Especially N-MOS is measured very consistent between labs

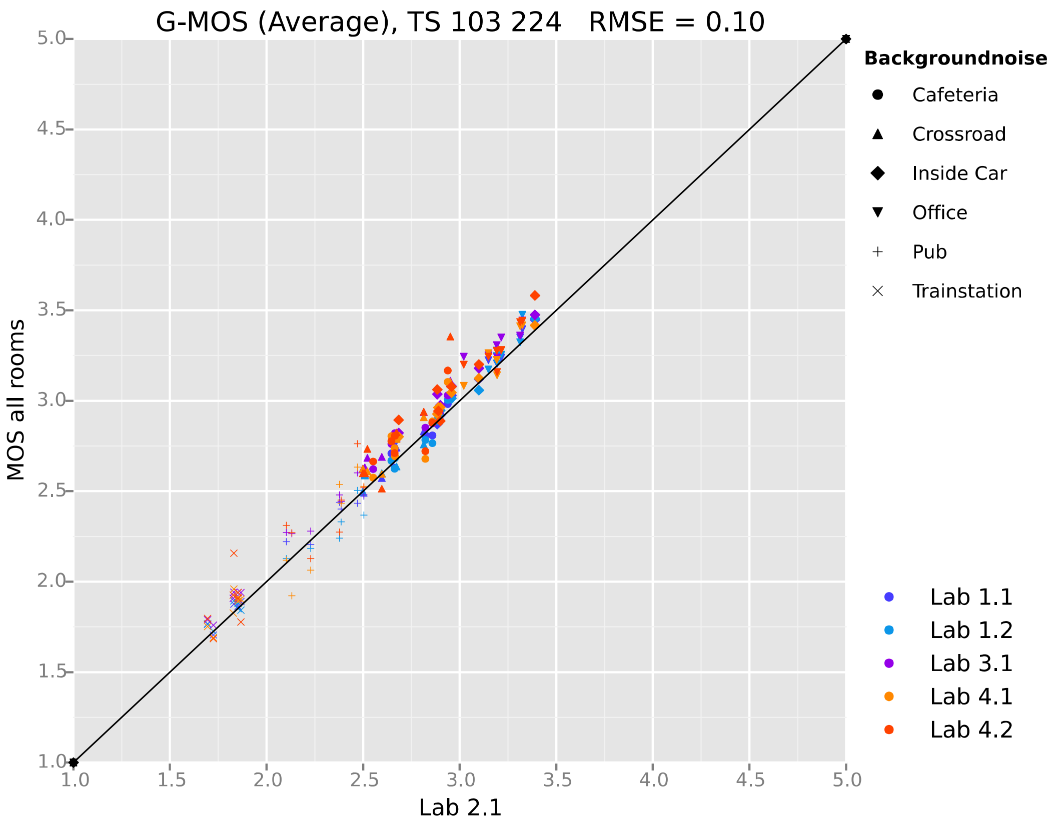


Figure 12: Correlation between G-MOS (WB) between Lab 2.1 and other labs

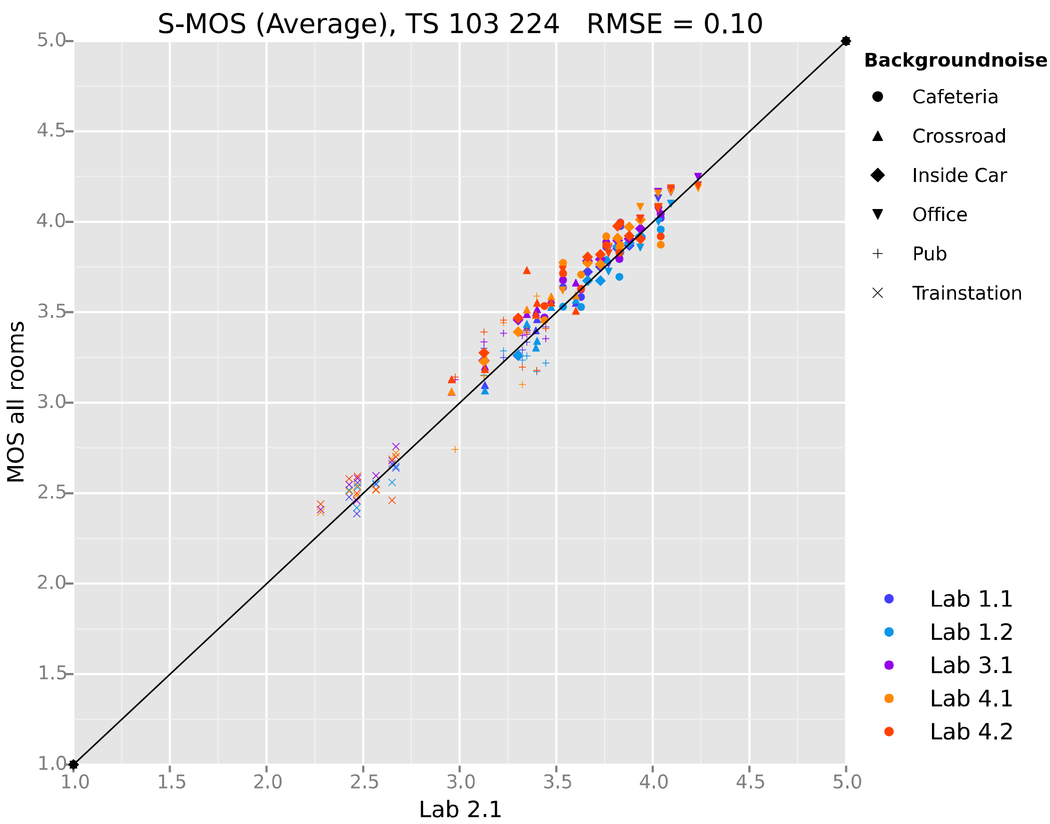


Figure 13: Correlation between S-MOS (WB) between Lab 2.1 and other labs

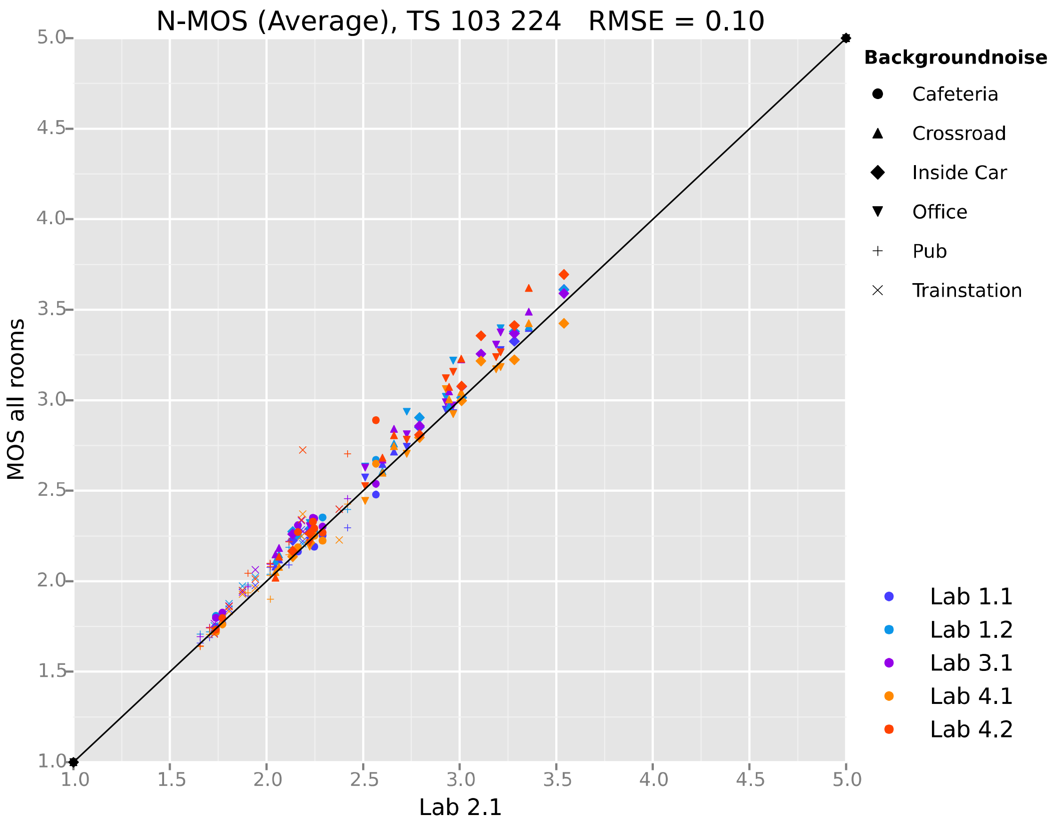


Figure 14: Correlation between N-MOS (WB) between Lab 2.1 and other labs

4.2.3.2.3 Narrowband

4.2.3.2.3.1 No background noise

In this clause, results under silent conditions in NB mode are presented, as shown in Figure 15 to Figure 17. Basically, the variance to be expected in different labs with no background noise simulation present can be observed.

It seems that the parameters described in clause 4.2.3.1 may impact the results to a certain extent providing the basis for the interpretation of the experiments with the background noise simulation methodologies. The impact on S-MOS and G-MOS is less pronounced than in wideband mode.

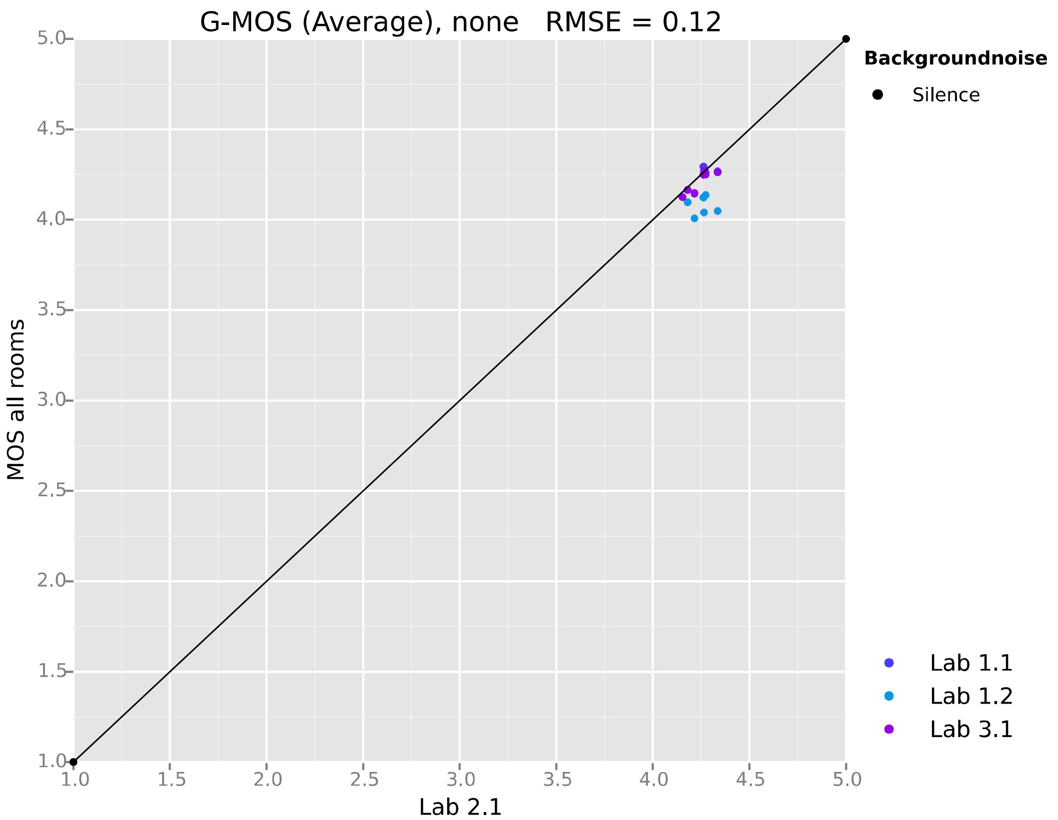


Figure 15: Correlation between G-MOS (NB) between Lab 2.1 and other labs (Silence)

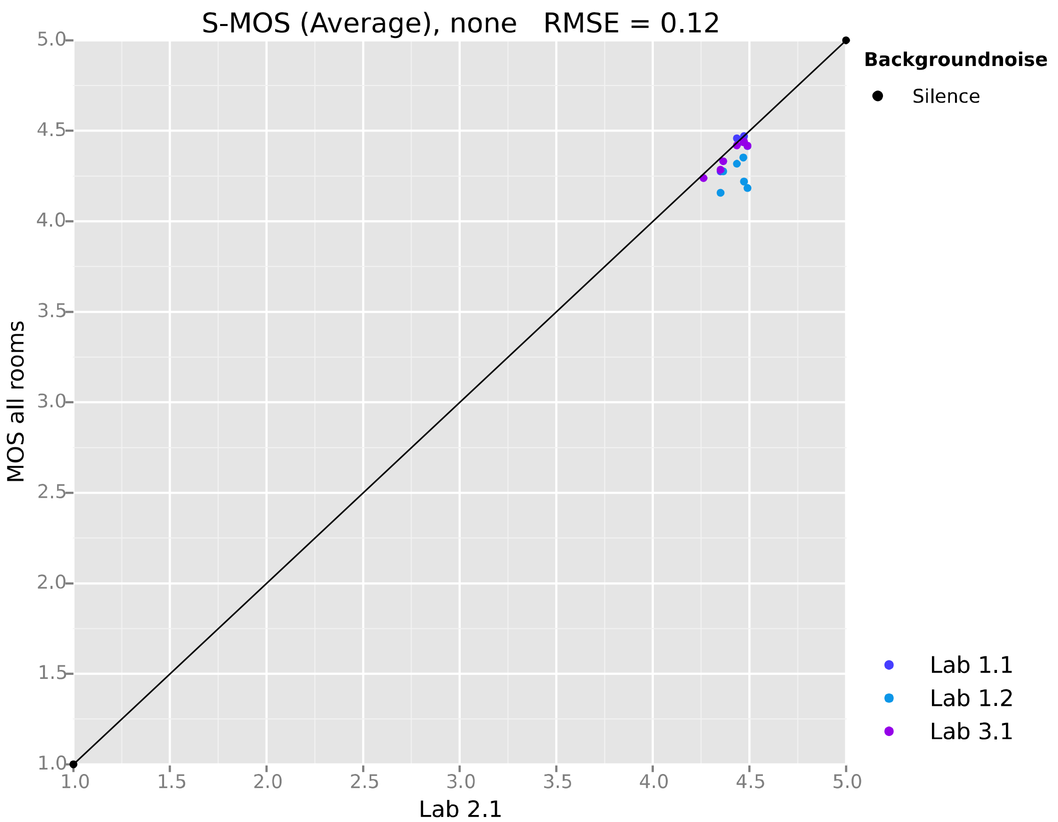


Figure 16: Correlation between S-MOS (NB) between Lab 2.1 and other labs (Silence)

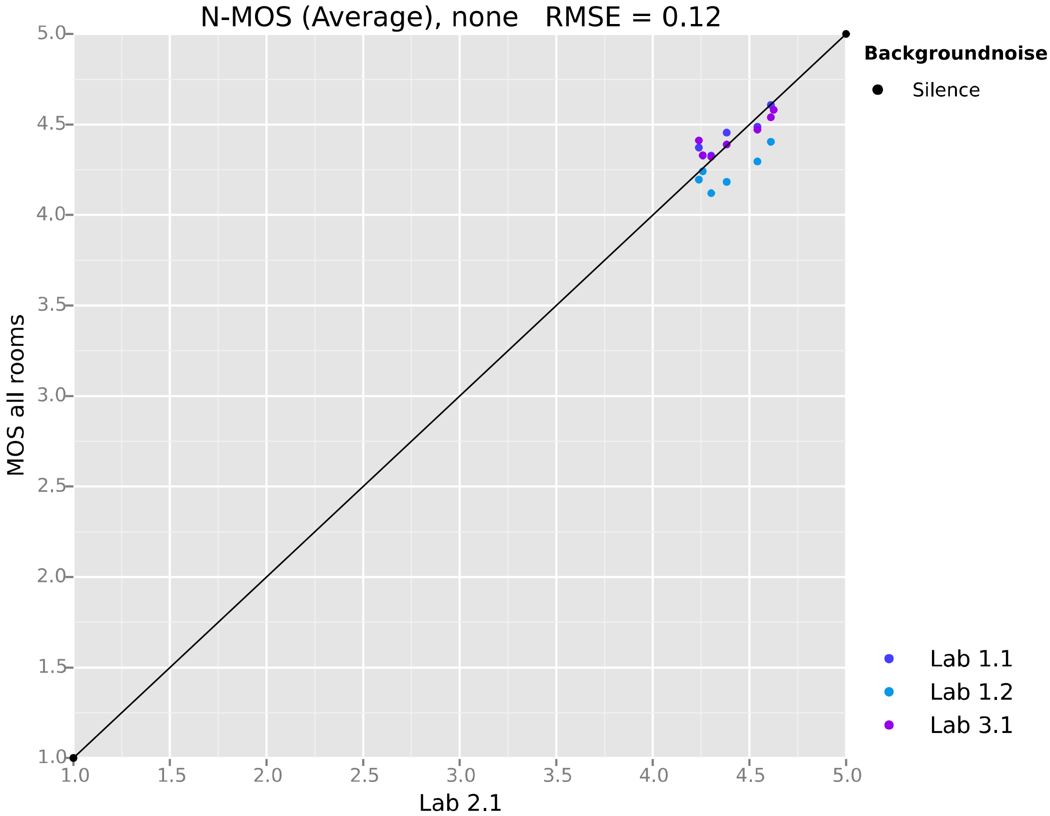


Figure 17: Correlation between N-MOS (NB) between Lab 2.1 and other labs (Silence)

4.2.3.2.3.2 Background Noises & Simulation acc. to ES 202 396-1

The results shown in Figure 18 to Figure 20 are based on using the ES 202 396-1 noise field simulation, using the binaurally recorded background noises from the associated noise database. The following observations can be made:

- Compared to Lab 2.1 the results measured in the other labs are in general about 0.2 to 0.3 MOS higher for N-MOS and G-MOS
- The results are somewhat scattered leading to RMSE values between 0.15 and 0.21
- N-MOS shows the highest variation in results up to about 0.5 MOS

Note that the indicated outlier for test lab 2.1 is excluded from the RMSE calculations.

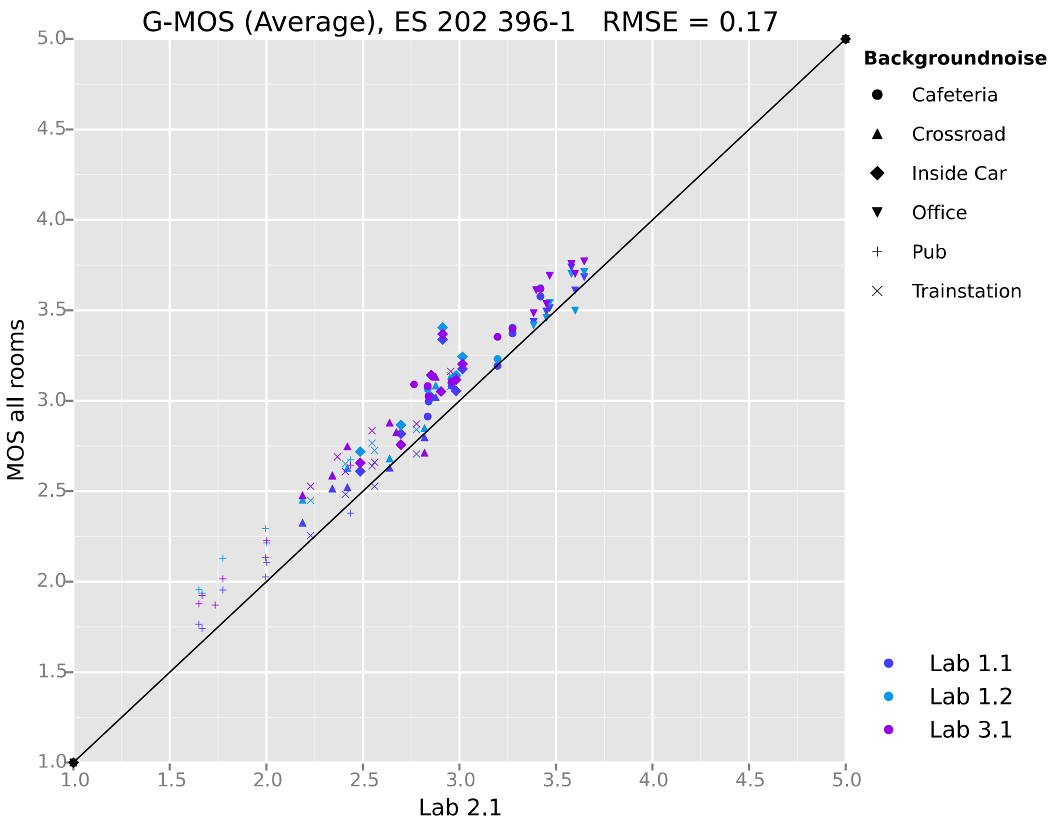


Figure 18: Correlation between G-MOS (NB) results from Lab 2.1 and other labs

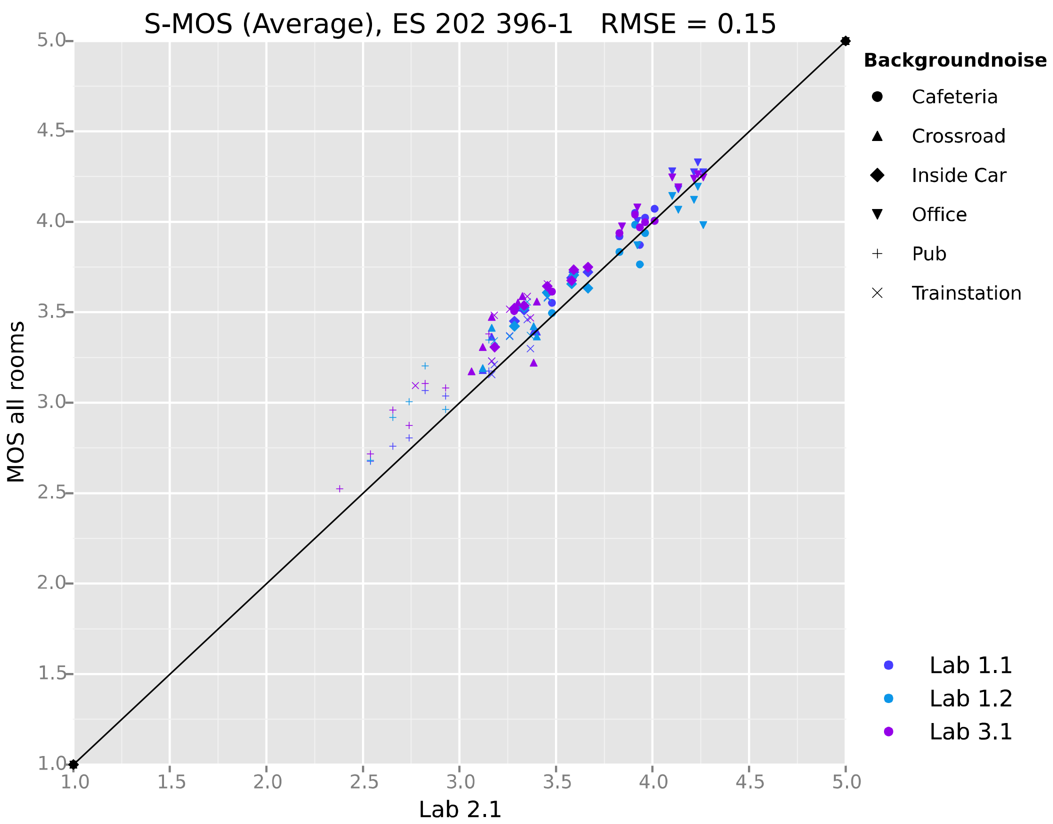


Figure 19: Correlation between S-MOS (NB) results from Lab 2.1 and other labs

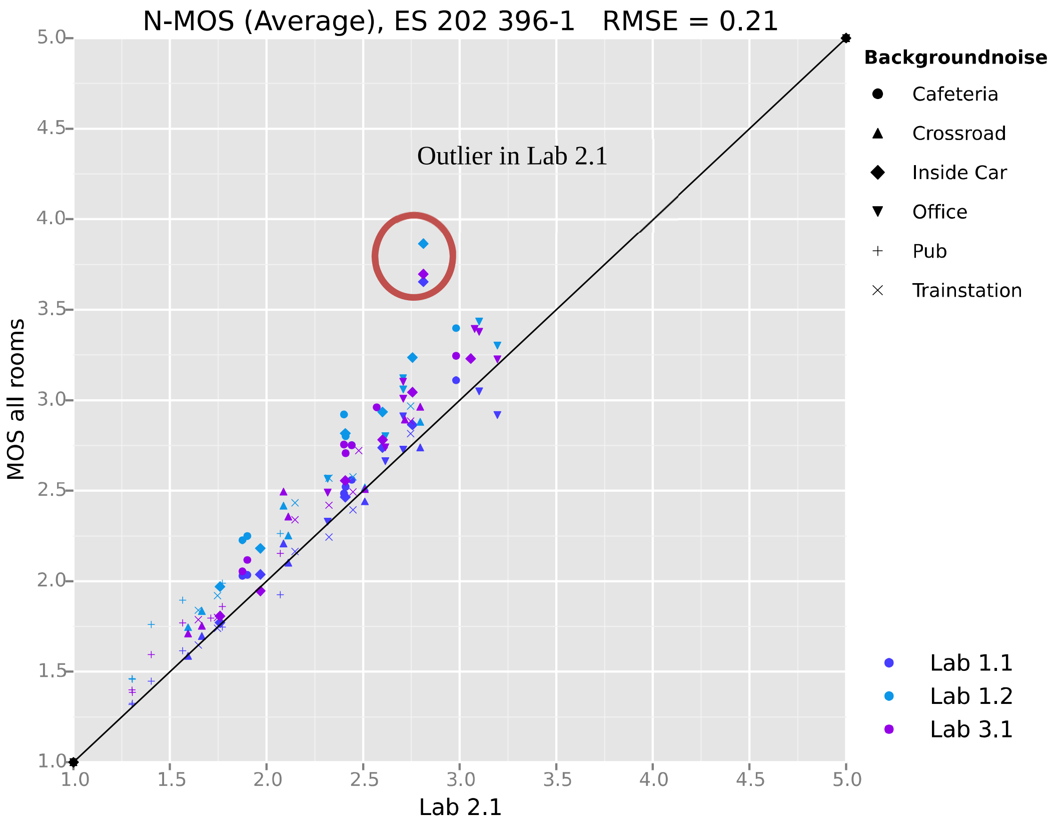


Figure 20: Correlation between N-MOS (NB) results from Lab 2.1 and other labs

4.2.3.2.3.3 Simulation acc. to ES 202 396-1 with binaural noises from TS 103 224

The results shown Figure 21 to Figure 23 are based on using the ES 202 396-1 noise field simulation, but using the binaurally recorded background noises from TS 103 224. The following observations can be made:

- Compared to Lab 2.1, the results measured in the other labs are in general about 0.3 to 0.4 MOS higher for S-MOS, N-MOS and G-MOS
- The results are somewhat scattered leading to RMSE values between 0.22 and 0.25
- N-MOS shows the highest variation in results up to about 0.5 MOS
- S-MOS shows a high variation of the results for Pub-noise and Train station noise

Note that the indicated outlier for test lab 2.1 is excluded from the RMSE calculations.

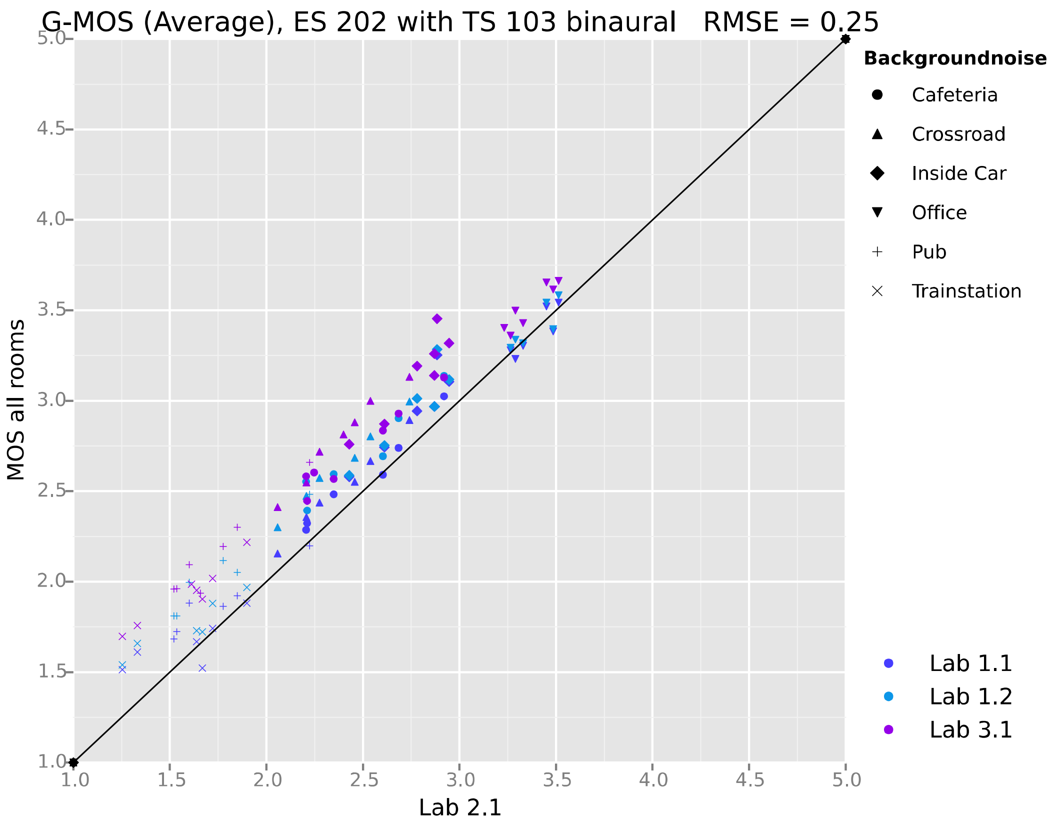


Figure 21: Correlation between G-MOS (NB) between Lab 2.1 and other labs

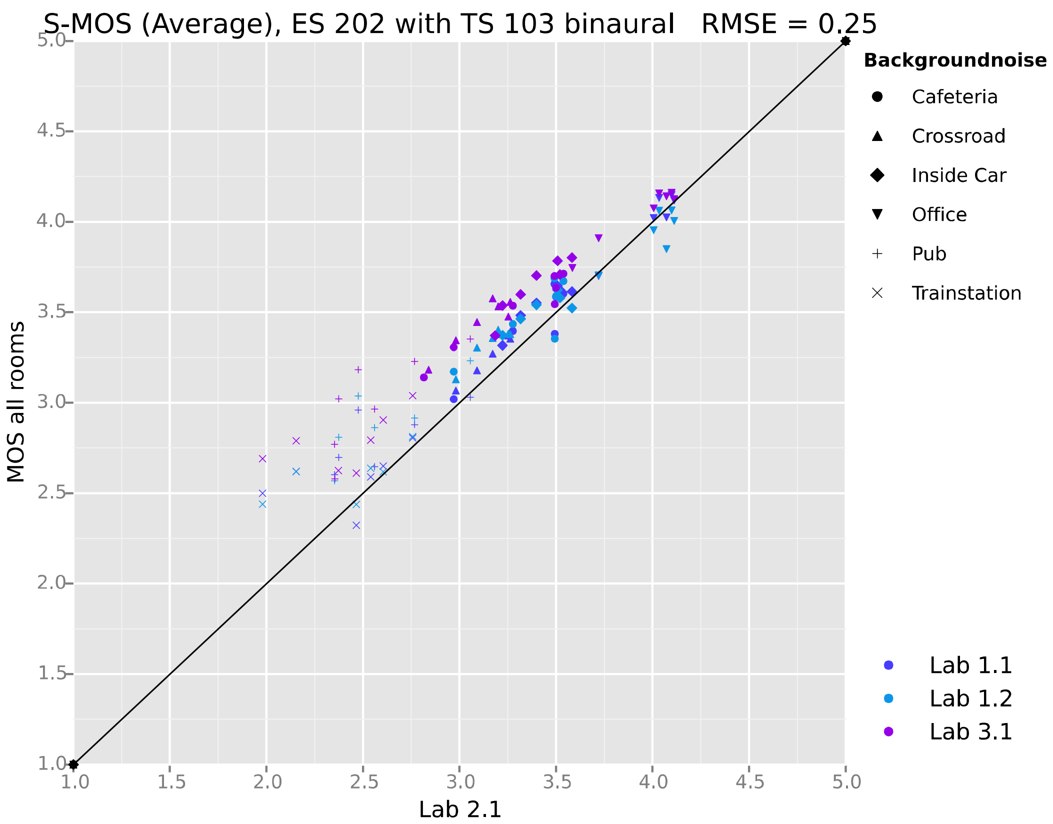


Figure 22: Correlation between S-MOS (NB) between Lab 2.1 and other labs

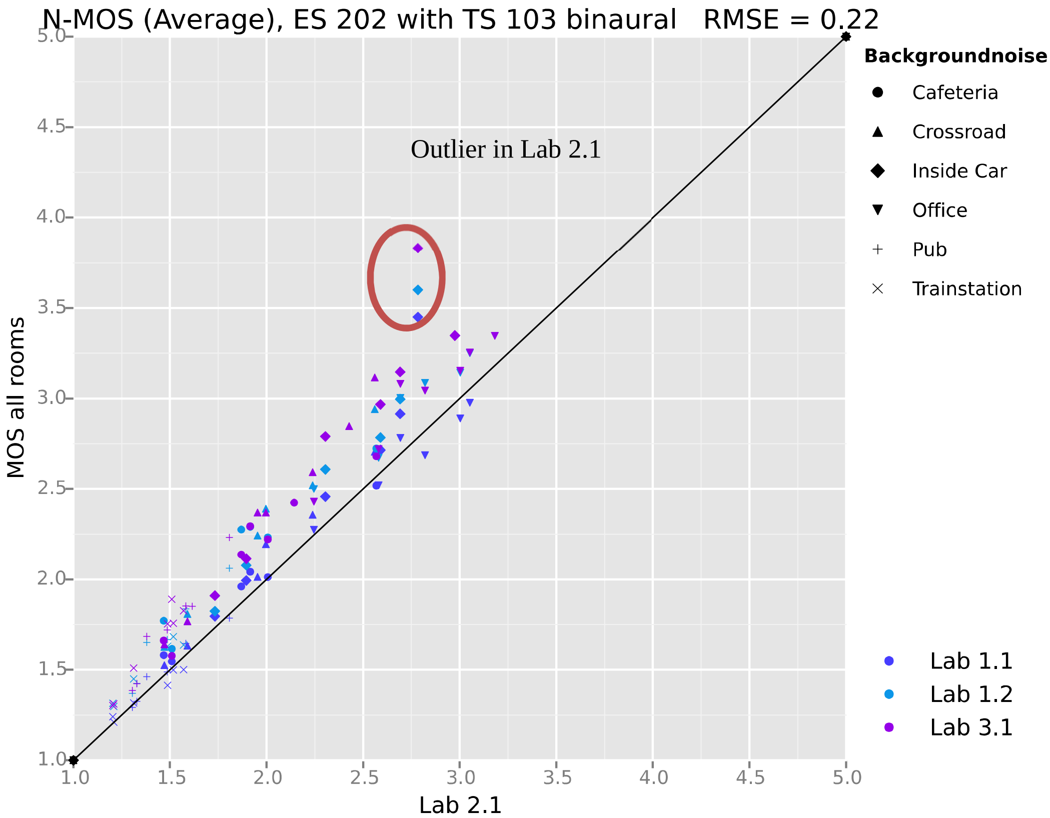


Figure 23: Correlation between N-MOS (NB) between Lab 2.1 and other labs

4.2.3.2.3.4 Background Noises & Simulation acc. to TS 103 224

The results shown in this clause are based on using the TS 103 224 noise field simulation, using the 8-channel noise recording from the associated noise database. For this setup, the following observations can be made:

- All results line up fairly well, no offset between the labs can be observed
- The results are less scattered leading to low RMSE values between 0.11 and 0.14
- Especially the measured N-MOS values are very consistent

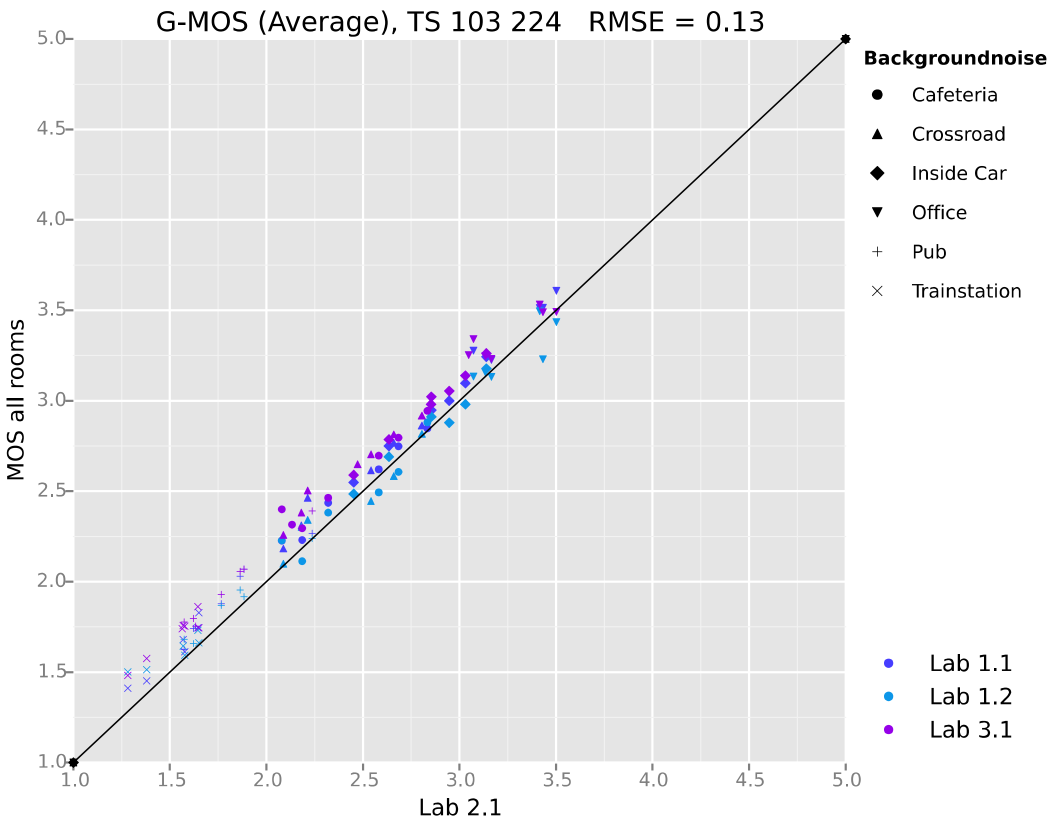


Figure 24: Correlation between G-MOS (NB) results from Lab 2.1 and other labs

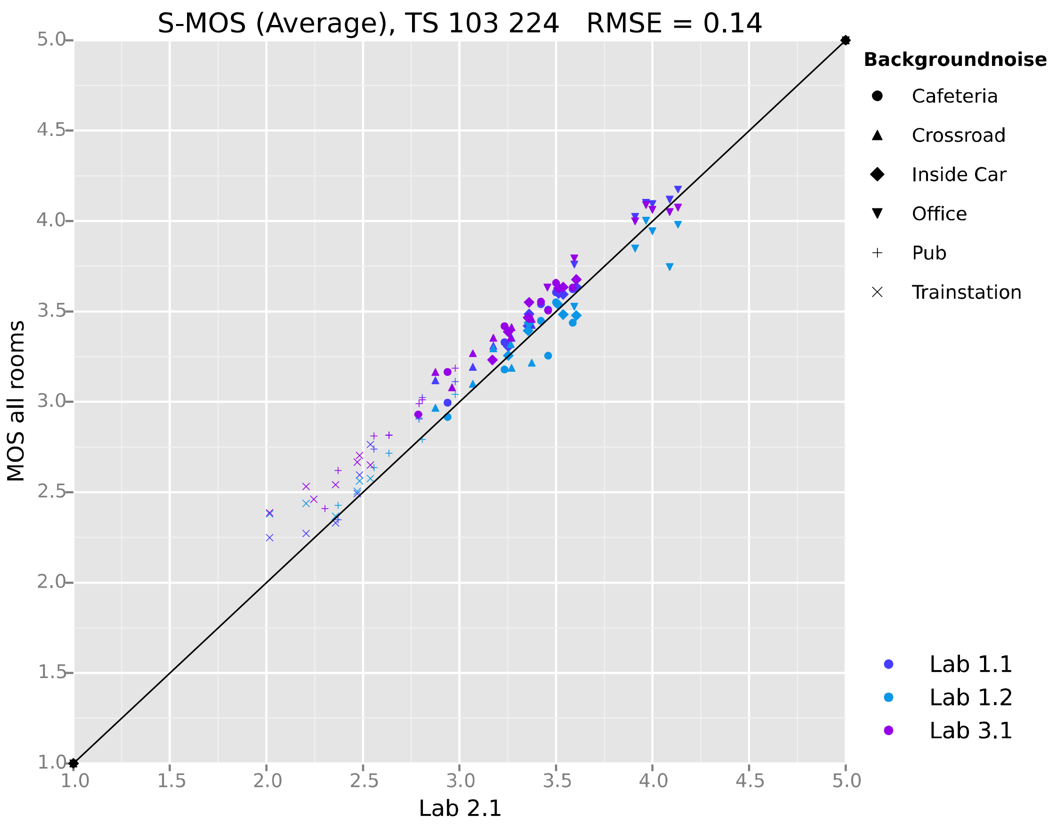


Figure 25: Correlation between S-MOS (NB) results from Lab 2.1 and other labs

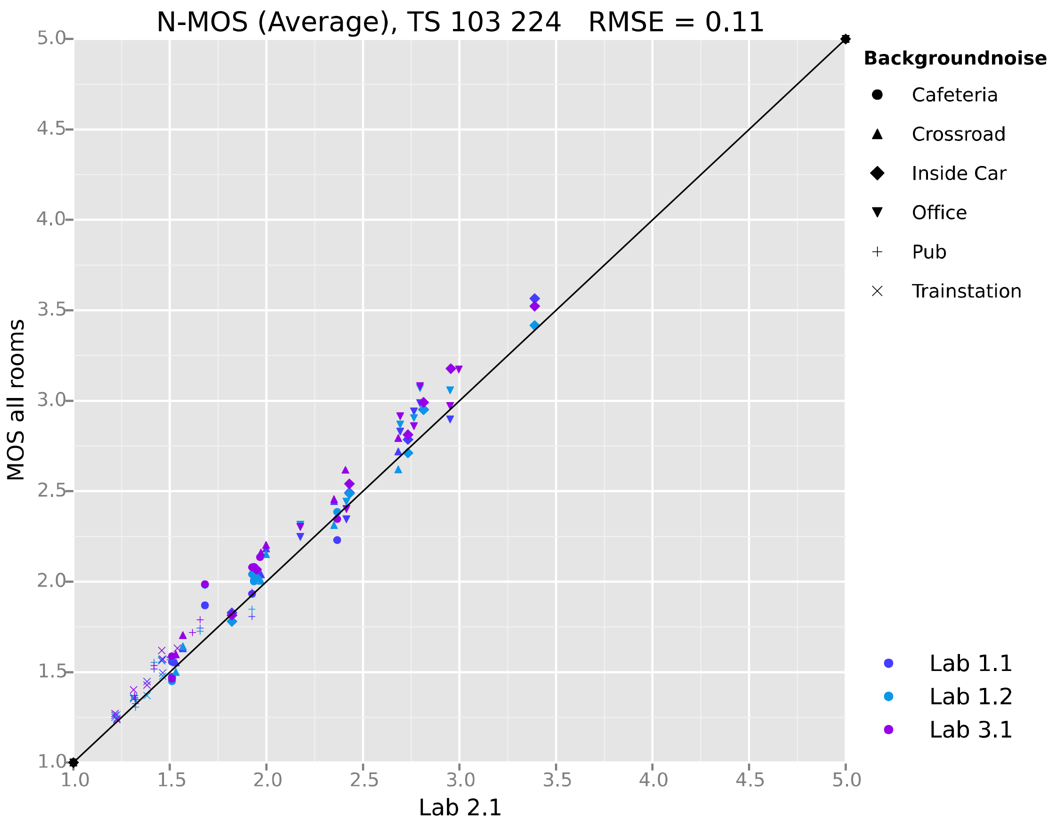


Figure 26: Correlation between N-MOS (NB) results from Lab 2.1 and other labs

4.2.3.3 Comparison of measurement results between ES 202 396-1 and TS 103 224 background noise simulation technique

4.2.3.3.1 Introduction

In this clause, S-, N-, and G-MOS as measured in the different rooms for different terminals are analysed. The comparison is focused on the differences between background noise simulation systems. The comparison is carried out between each method and the different sets of background noises. The relation between the results measured using the ETSI ES 202 396-1 background noise simulation system and the results obtained using the ETSI TS 103 224 simulation system are presented for all noises, all labs, and all terminals.

The results of the different devices are colour-coded as indicated in the following figures.

4.2.3.3.2 Wideband

4.2.3.3.2.1 Background Noises & Simulation acc. to ES 202 396-1 compared to TS 103 224 simulation

While for crossroad, inside car office and pub noise the measurements lead to quite comparable results, deviations are visible for the cafeteria and train station noise. The reason for these differences is the different nature of the noises in the different noise databases. E.g. the cafeteria noise is in average about 6 dB lower in level for the recordings in ES 202 396-1, the train station noise is about 7 dB lower in level.

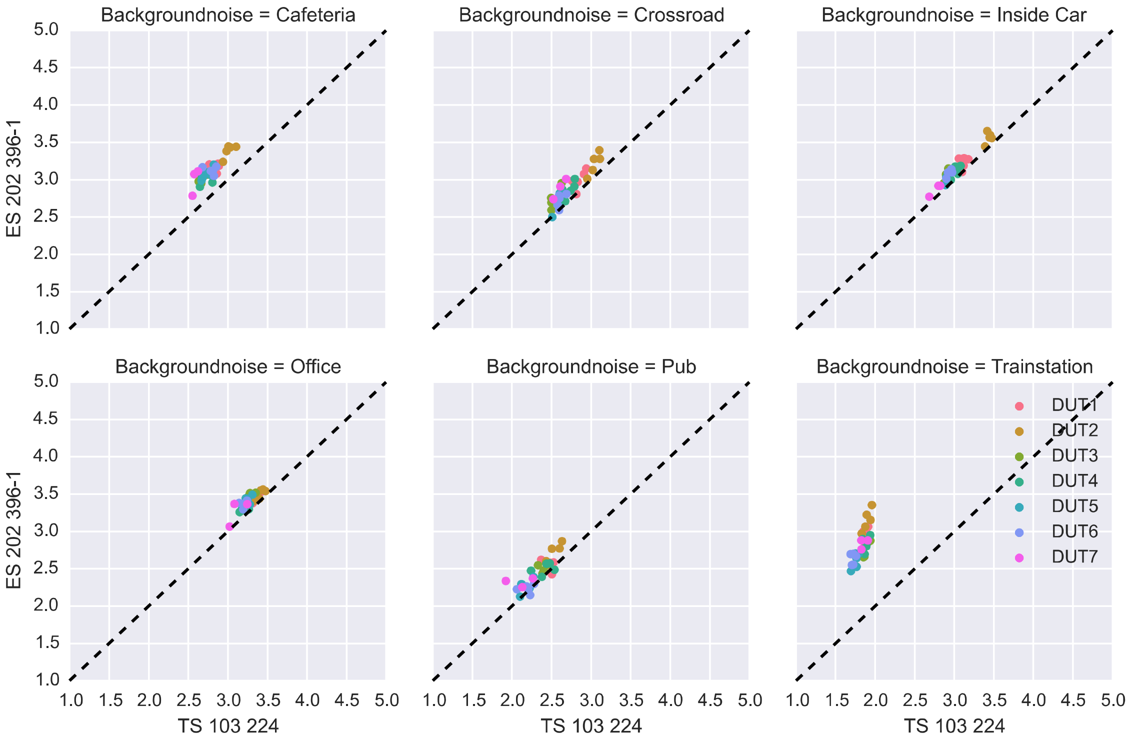


Figure 27: Correlation of G-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology

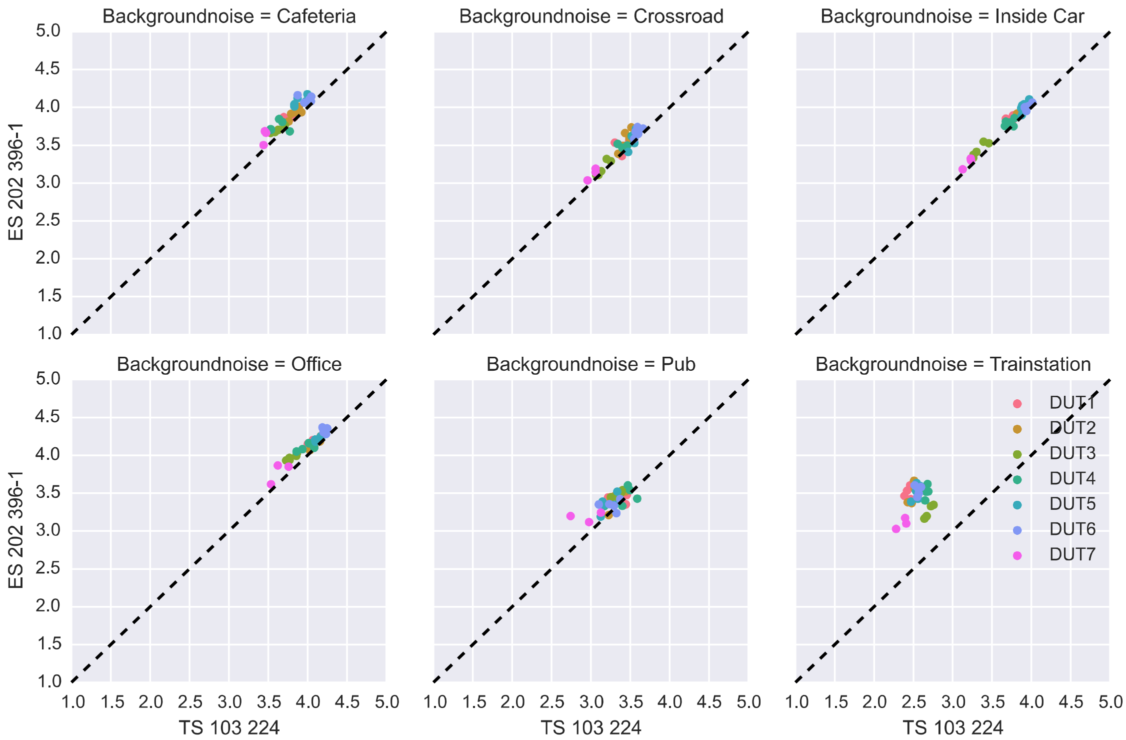


Figure 28: Correlation of S-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology

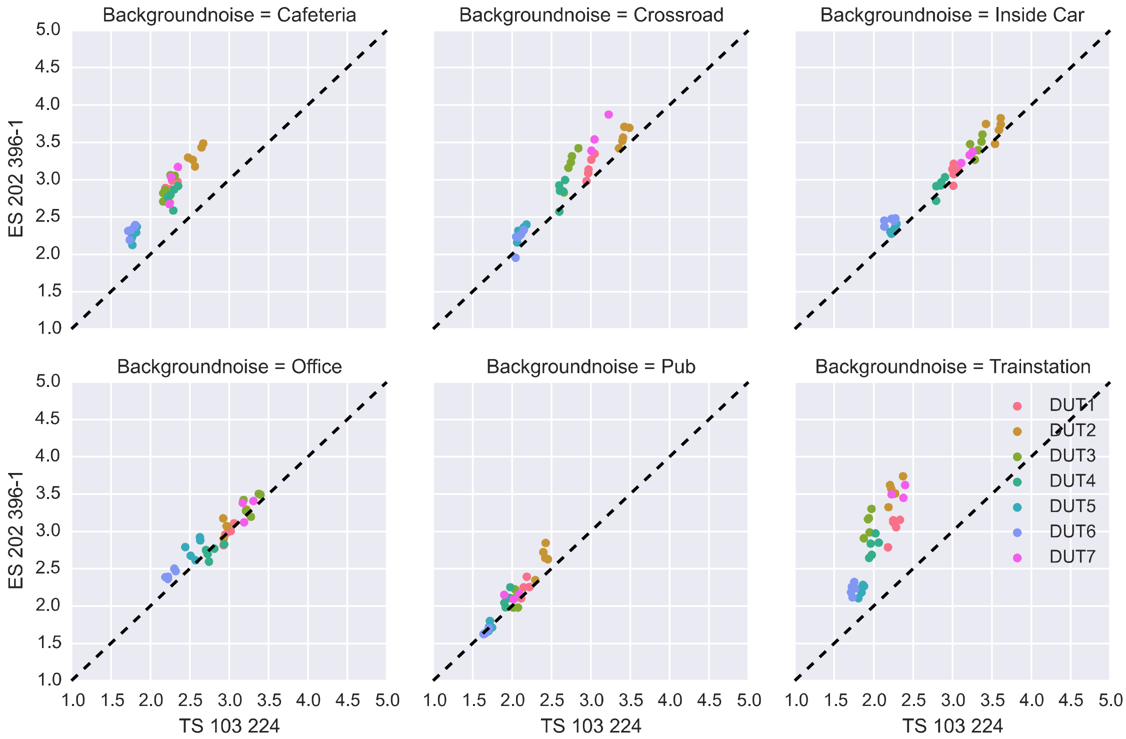


Figure 29: Correlation of N-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology

4.2.3.3.2.2 Simulation acc. to ES 202 396-1 with noises from TS 103 224 (binaural) compared to TS 103 224 simulation

When choosing the same type of recordings by taking the binaurally recorded noises from TS 103 224, the differences between the methods become smaller and results line up in general. More variation can be seen for the N-MOS results than for the S-MOS results. This is clearly indicating the difference in the background noise simulation method. The S-MOS seems to be mainly determined by the speech signal which is unchanged in the experiments.

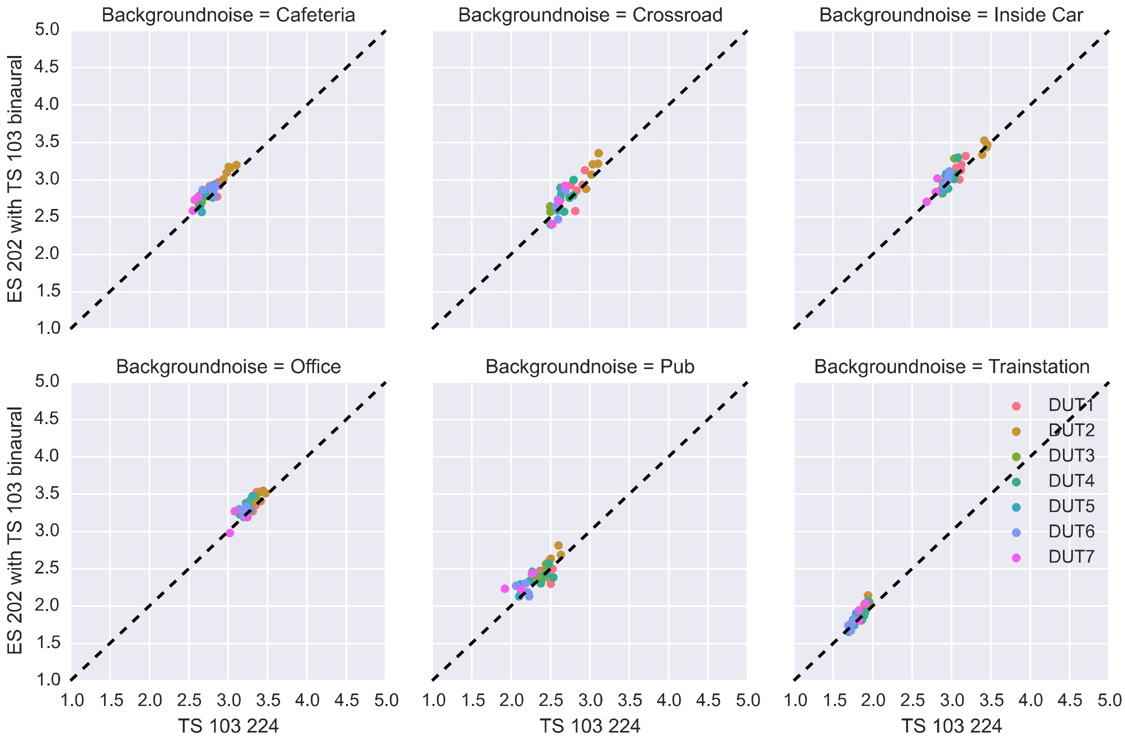


Figure 30: Correlation of G-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

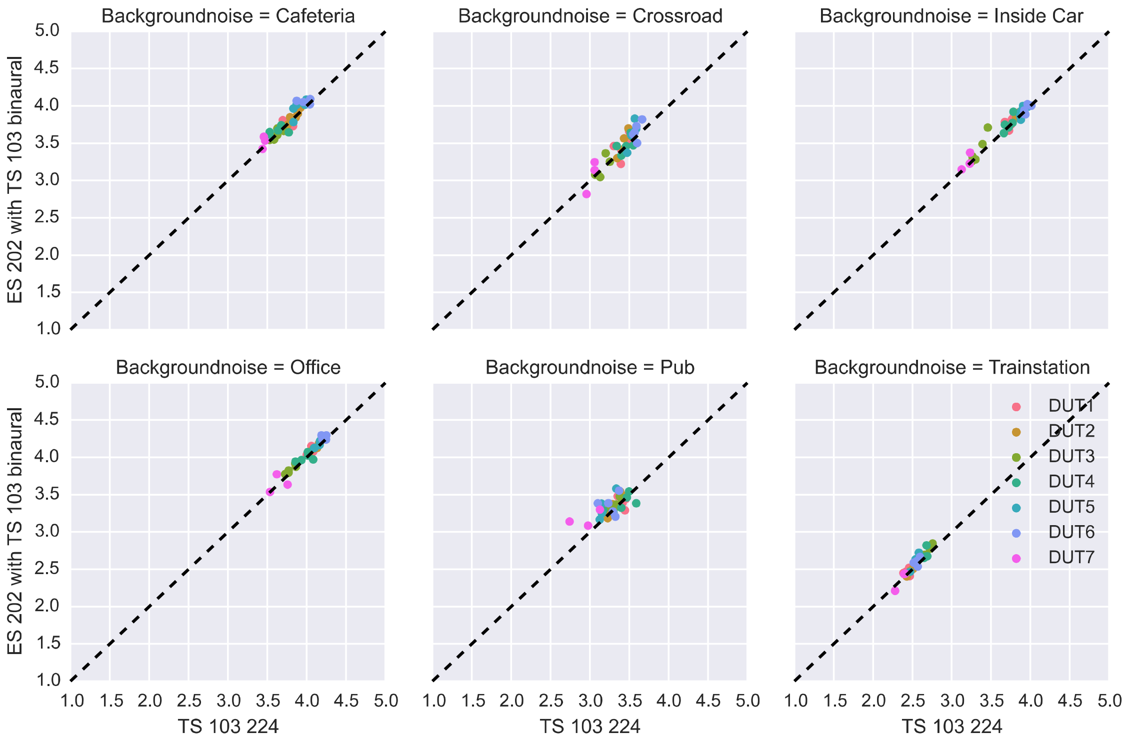


Figure 31: Correlation of S-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

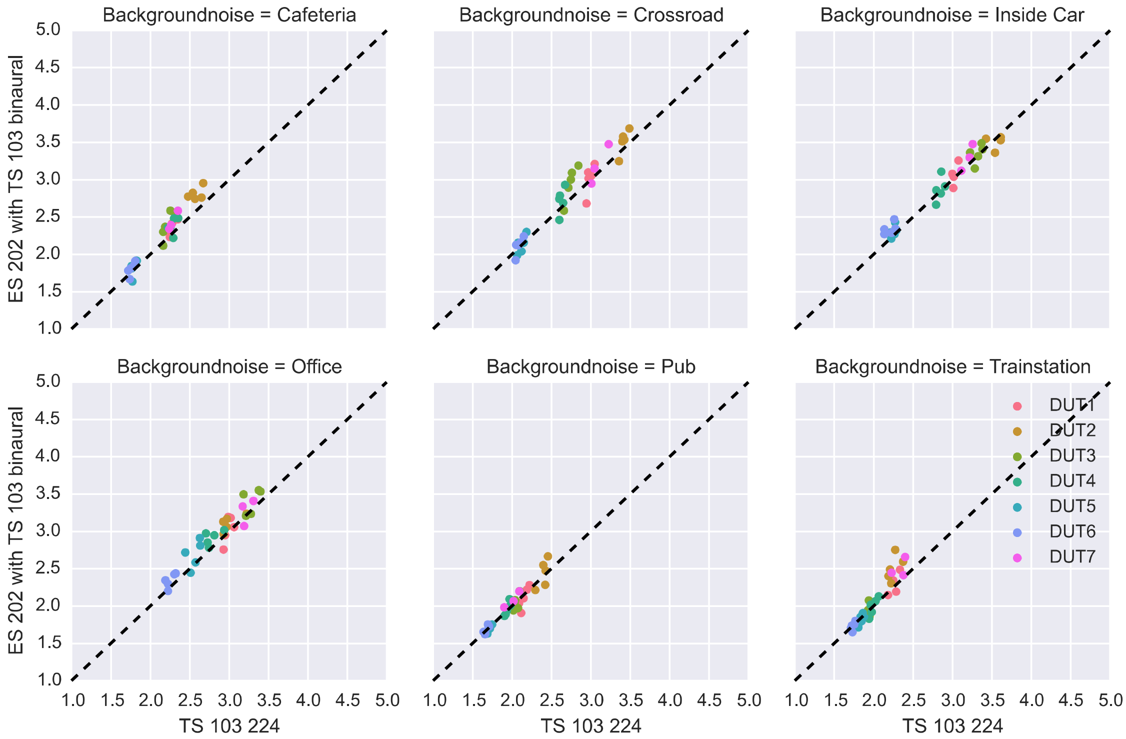


Figure 32: Correlation of N-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

4.2.3.3.3 Narrowband

4.2.3.3.3.1 Background Noises & Simulation acc. to ES 202 396-1 compared to TS 103 224 simulation

While for crossroad, inside car office and pub noise the measurements lead to quite comparable results, deviations are visible for the cafeteria and train station noise. The reason for these differences is the different nature of the noises in the different noise databases. E.g. the cafeteria noise is in average about 6 dB lower in level for the recordings in ES 202 396-1, the train station noise is about 7 dB lower in level.

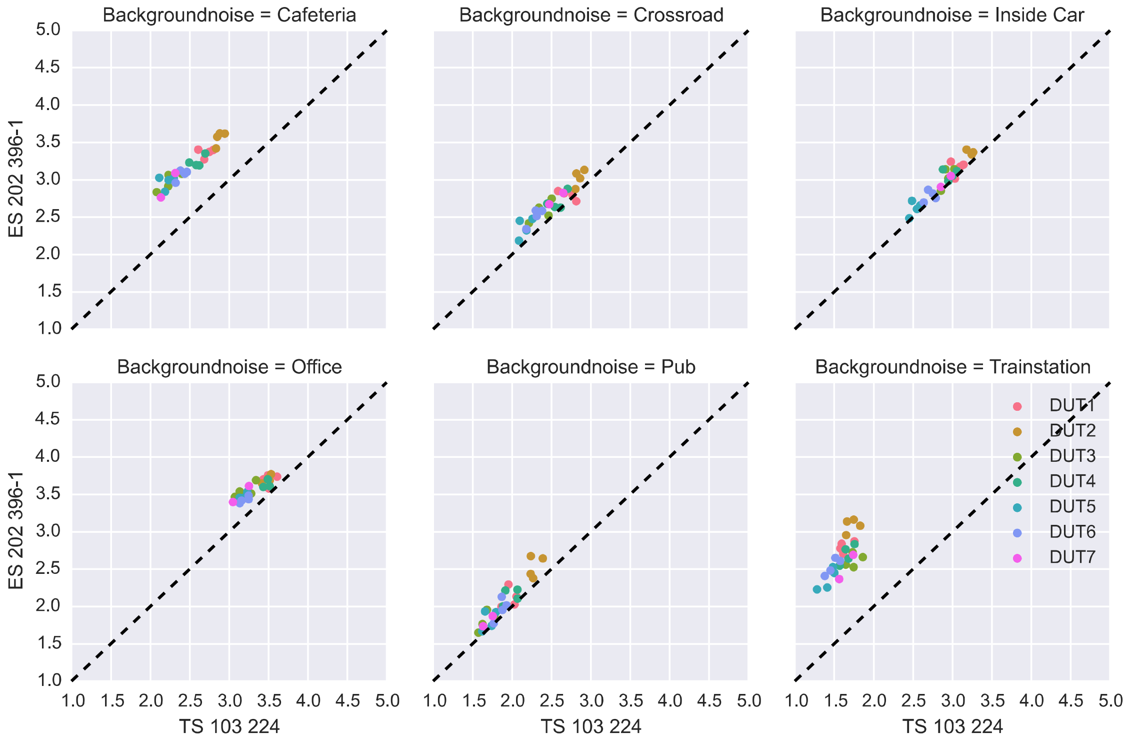


Figure 33: Correlation of G-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology

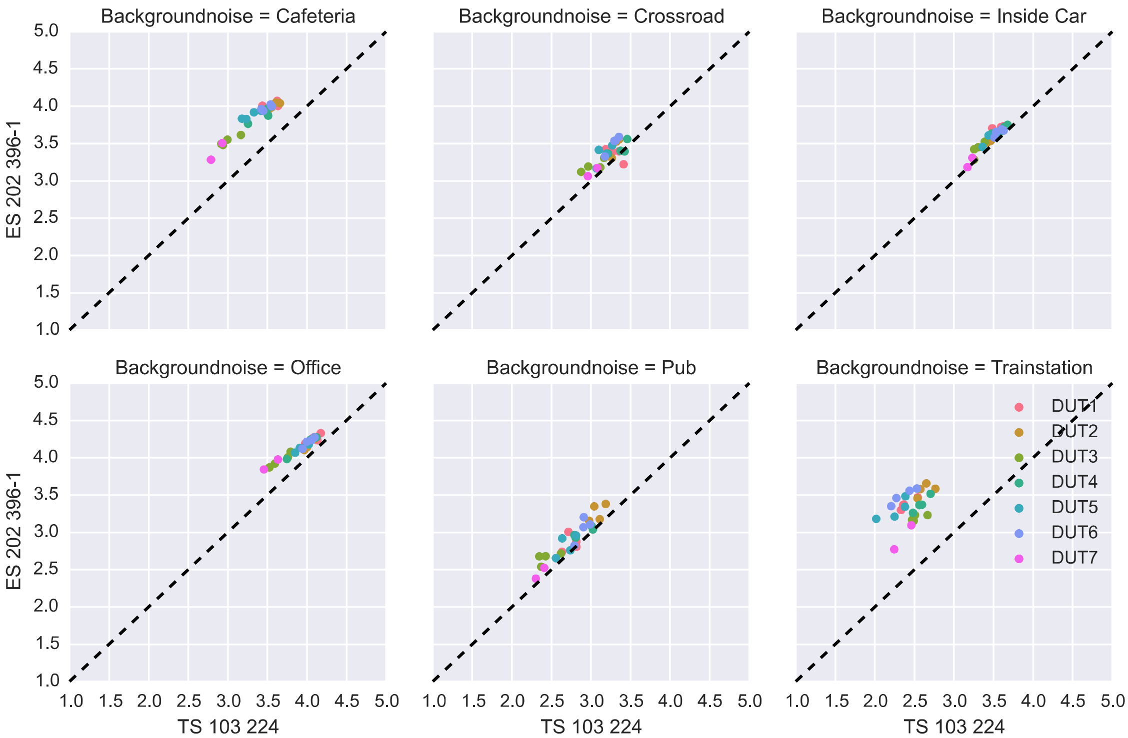


Figure 34: Correlation of S-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology

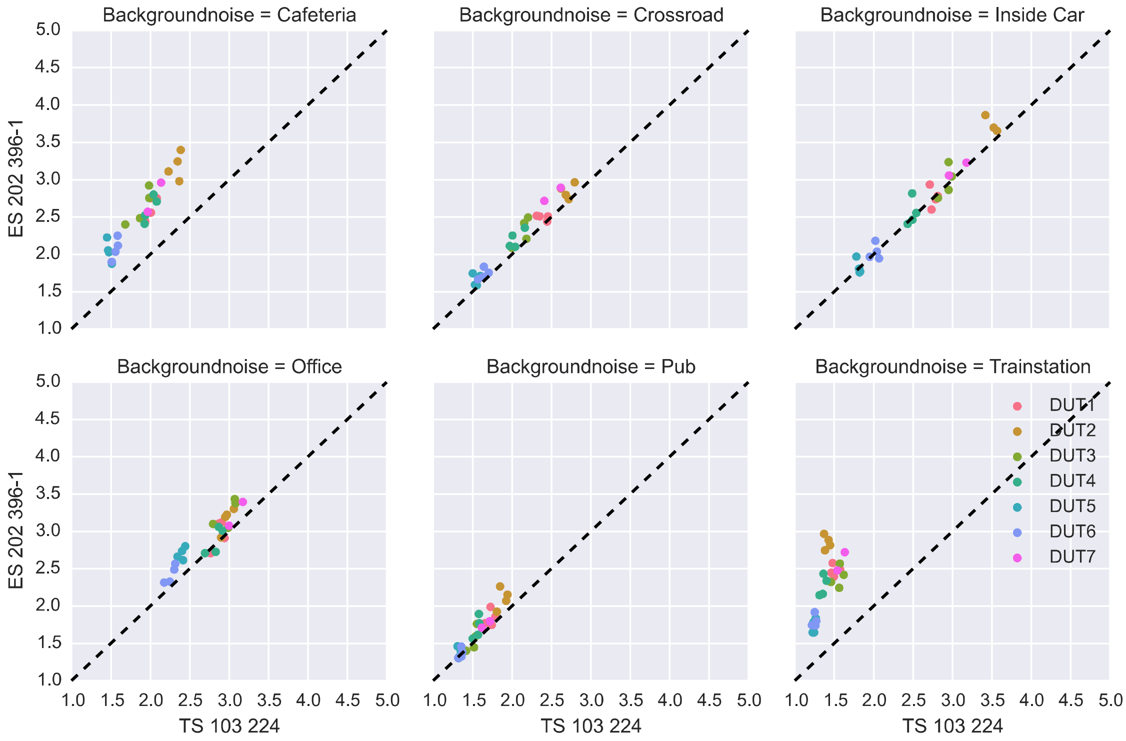


Figure 35: Correlation of N-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology

4.2.3.3.3.2 Simulation acc. to ES 202 396-1 with noises from TS 103 224 (binaural) compared to TS 103 224 simulation

When choosing the same type of recordings by taking the binaurally recorded noises from TS 103 224, the differences between the methods become smaller and results line up in general. More variation can be seen for the N-MOS results than for the S-MOS results. This is clearly indicating the difference in the background noise simulation method. The S-MOS seems to be mainly determined by the speech signal which is unchanged in the experiments.

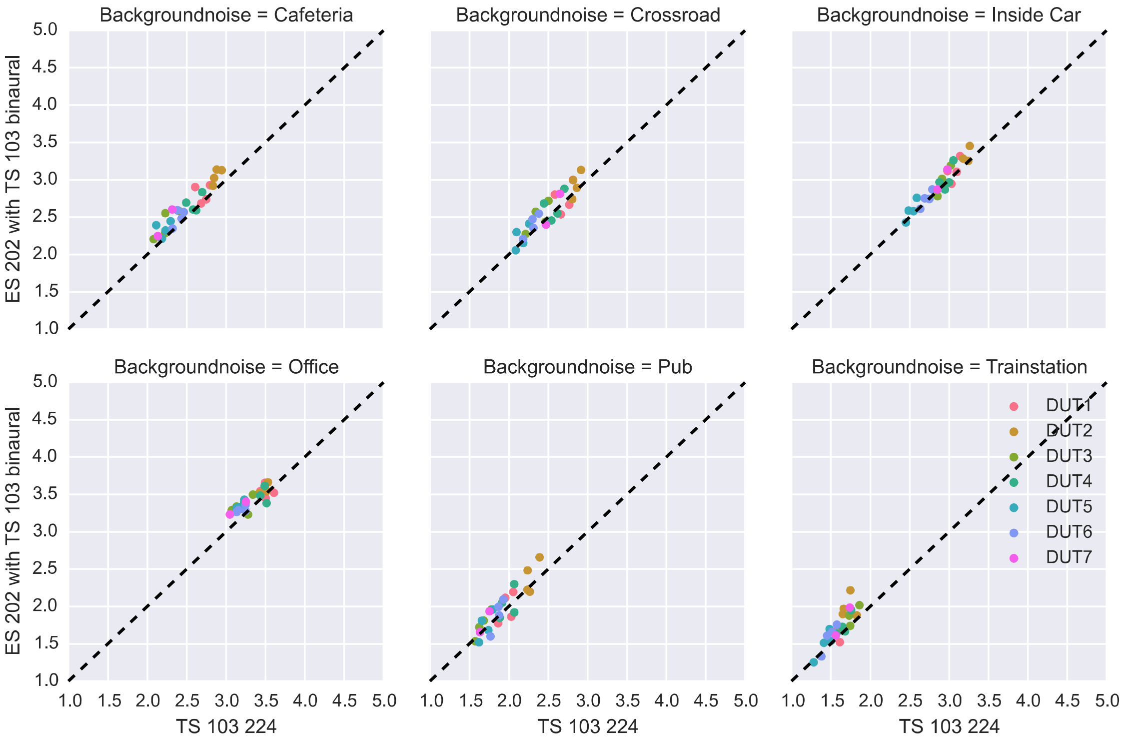


Figure 36: Correlation of G-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

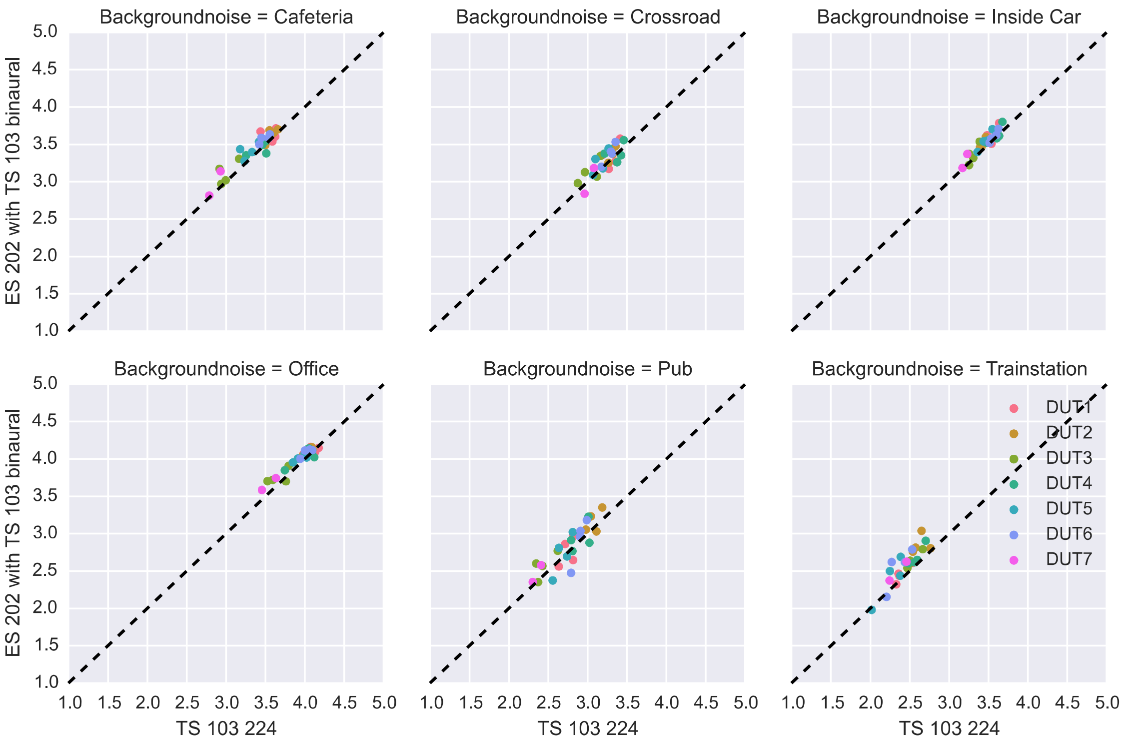


Figure 37: Correlation of S-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

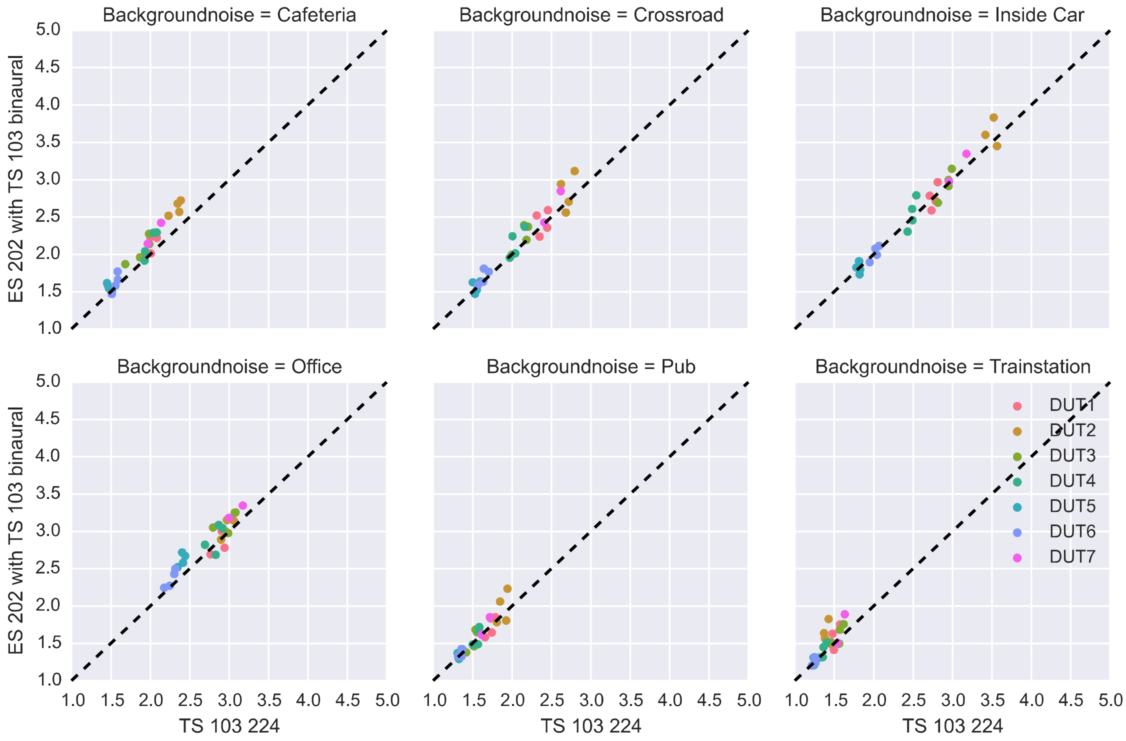


Figure 38: Correlation of N-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology (noises from TS 103 224)

4.2.3.4 Analysis results according to TS 26.132

4.2.3.4.1 Introduction

In this clause, the speech quality analysis as specified in [3] and [2] is carried out. Averaging the results across all noise conditions provides one single S-MOS, N-MOS and G-MOS for each device.

4.2.3.4.2 Wideband

4.2.3.4.2.1 Comparison of absolute Results

The results in Figure 39 show the same tendency for all phones when measured using the different background noise simulation setups. The rank order in performance remains unchanged.

Figure 39: S-/N-/G-MOS (WB) averaged across background noises and all labs

4.2.3.4.2.2 Differences between TS 103 224 and ES 202 396-1

Differences in the absolute values when comparing ES 202 396-1 with TS 103 224 can be seen for all terminals (see Figure 40). The offset is not constant but depending on the type of terminal. The main difference is found in the N-MOS values indication the different sensitivity of individual terminals to the different types of noise. The performance for all terminals is better when using the ES 202 396-1. Again, this is due to the less stressing noises in this standard.

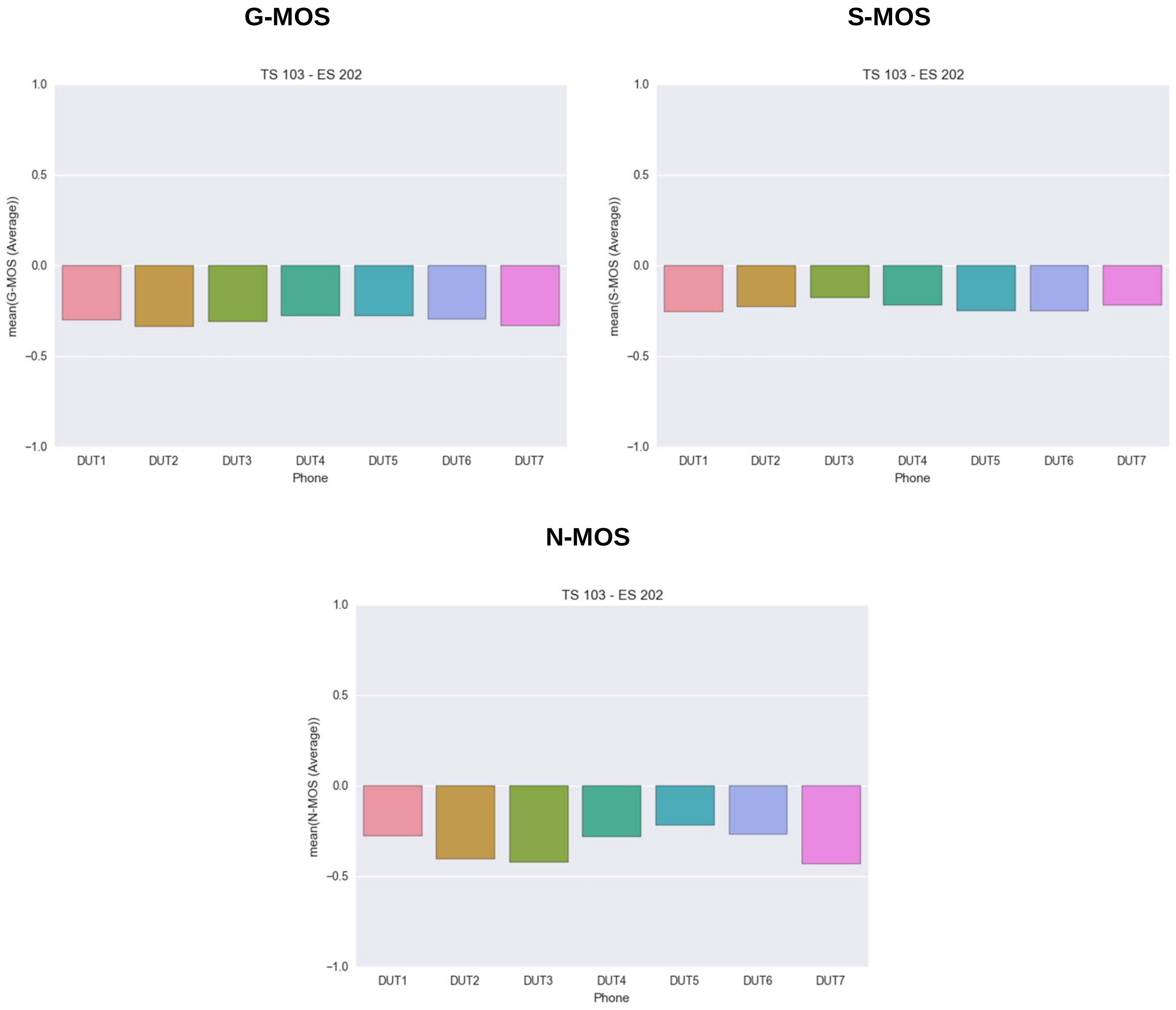


Figure 40: Differences (WB) between TS 103 224 and ES 202 396-1 methodology

4.2.3.4.2.3 Differences between TS 103 224 and ES 202 396-1 with noises from TS 103 224

When using the binaurally recorded background noises from TS 103 224 for testing, the difference between the two systems gets much smaller (see Figure 41). Again, the offset is not constant, but depends on the type of terminal. The performance for all terminals is slightly better when using the ES 202 396-1 simulation method compared to the sound field simulation technique described in TS 103 224.

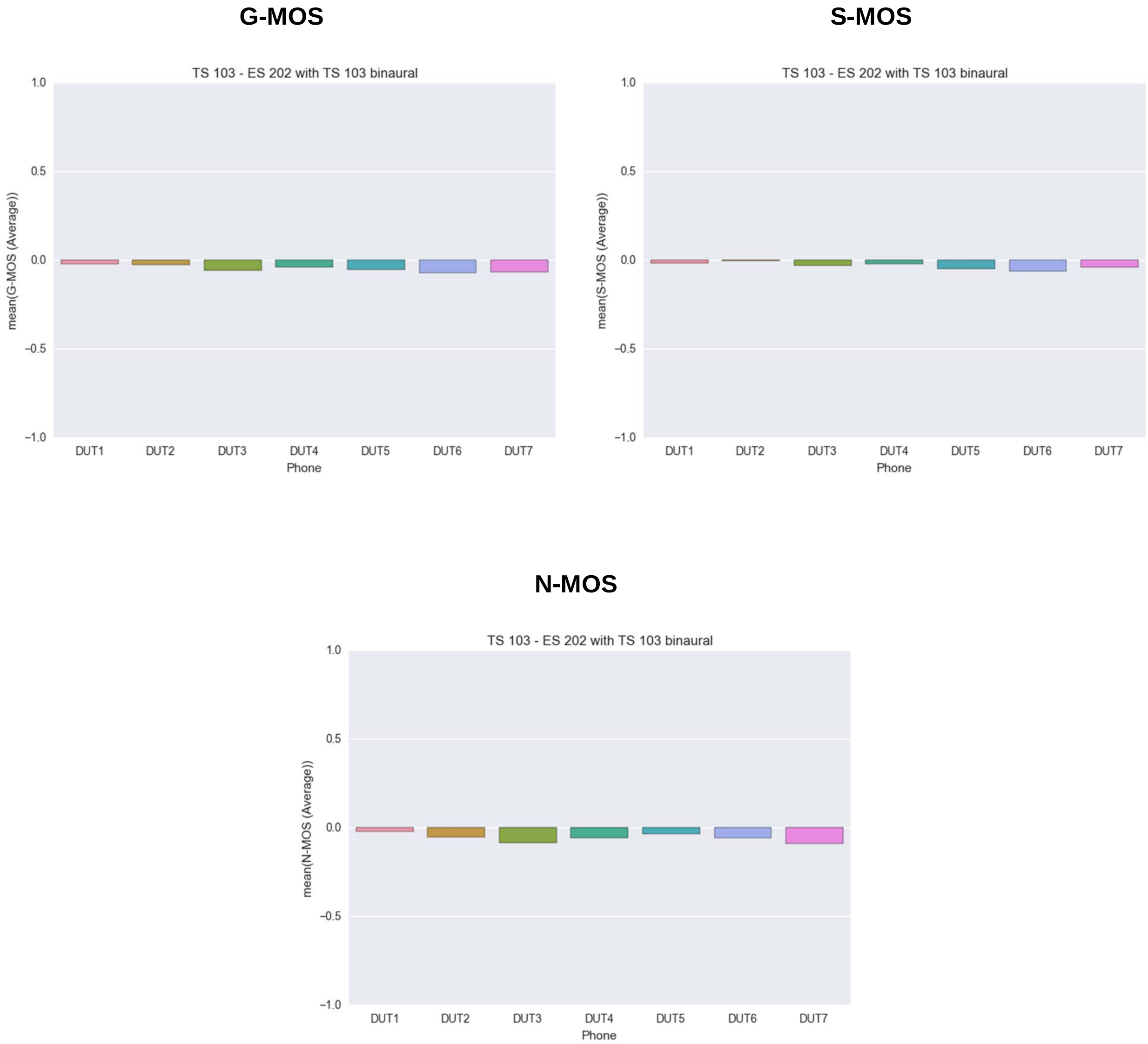


Figure 41: Differences (WB) between TS 103 224 and ES 202 396-1 methodology (binaural noises from TS 103 224 database)

4.2.3.4.3 Narrowband

4.2.3.4.3.1 Comparison of absolute Results

The results in Figure 42 show the same tendency for all phones when measured using the different background noise simulation setups. The rank order in performance remains unchanged.

Figure 42: S-/N-/G-MOS (NB) averaged across background noises and all labs

4.2.3.4.3.2 Differences between TS 103 224 and ES 202 396-1

Differences in the absolute values when comparing ES 202 396-1 with TS 103 224 can be seen for all terminals (see Figure 43). The offset is not constant but depending on the type of terminal. The differences are found in S-, N-, and G-MOS values. As in wideband the performance for all terminals is better when using the ES 202 396-1. Again, this is due to the less stressing noises in this standard.

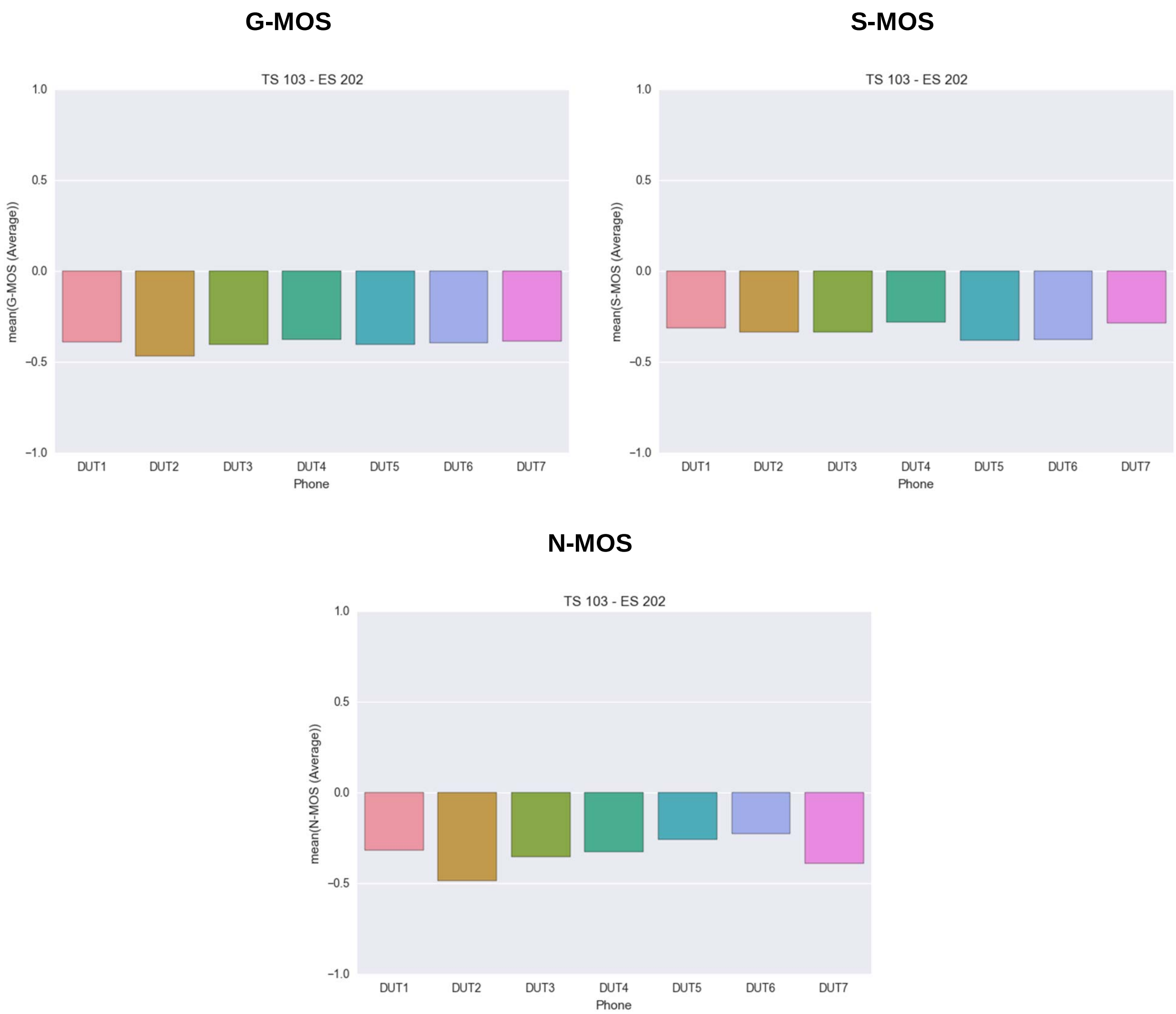


Figure 43: Differences (NB) between TS 103 224 and ES 202 396-1 methodology

4.2.3.4.3.3 Differences between TS 103 224 and ES 202 396-1 with noises from TS 103 224

When using the binaurally recorded background noises from TS 103 224 for testing, the difference between the two systems gets much smaller (see Figure 44). Again, the offset is not constant, but depends on the type of terminal and the performance for all terminals is slightly better when using the ES 202 396-1 simulation method compared to the sound field simulation technique described in TS 103 224.

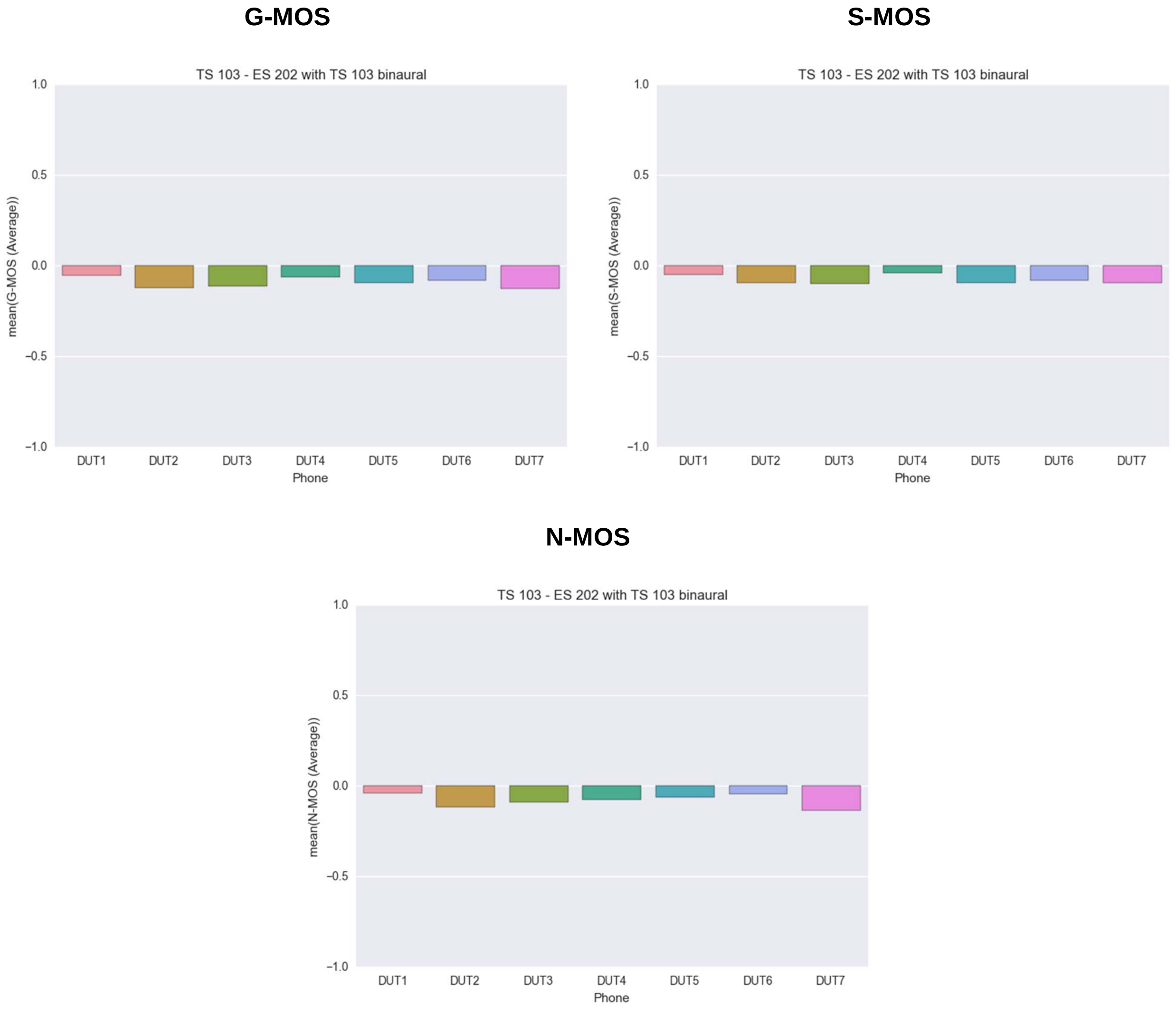


Figure 44: Differences (NB) between TS 103 224 and ES 202 396-1 methodology (binaural noises from TS 103 224 database)

4.2.3.5 Analyses of the noise spectra reproduced at the reference microphone

4.2.3.5.1 Introduction

The background noise spectra measured at the reference microphone (input signal for the analysis according to TS 103 106 [6]) of all devices under test are plotted into one diagram for all background noises which were used in this test. Each diagram contains background noise spectra from

- 6 DUTs from Lab 1 (DUT7 was not measured)
  - Included twice, since two different rooms were available
- 7 DUTs from Lab 2
- 7 DUTs from Lab 3
- 7 DUTs from Lab 4
  - Included twice, since two different setups were evaluated

4.2.3.5.2 Simulation acc. to ES 202 396-1 with recordings from TS 103 224

For all investigated background noises, quite large differences can be noticed in the spectra reproduced by the ES 202 396-1 simulation method. The differences depend to some extent on the noise type, as shown in Figure 45 to Figure 50. More uniform background noises, such as e.g., car, show less spectral variance than e.g., train station (which performs worst).

The differences in magnitude range from 5 dB up to 15 dB. The larger differences are mostly located in the low frequency domain.

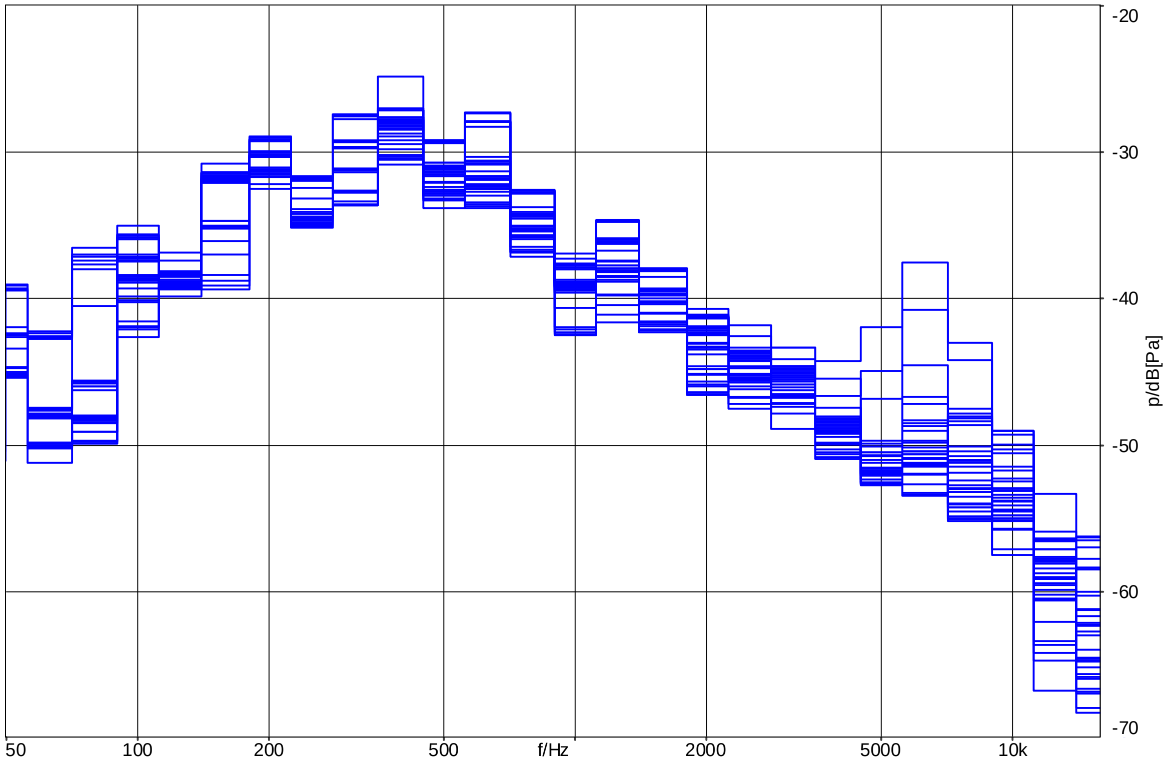


Figure 45: 1/3rd octave spectra at reference microphone and Cafeteria noise from TS 103 224 (ES 202 396-1 methodology)

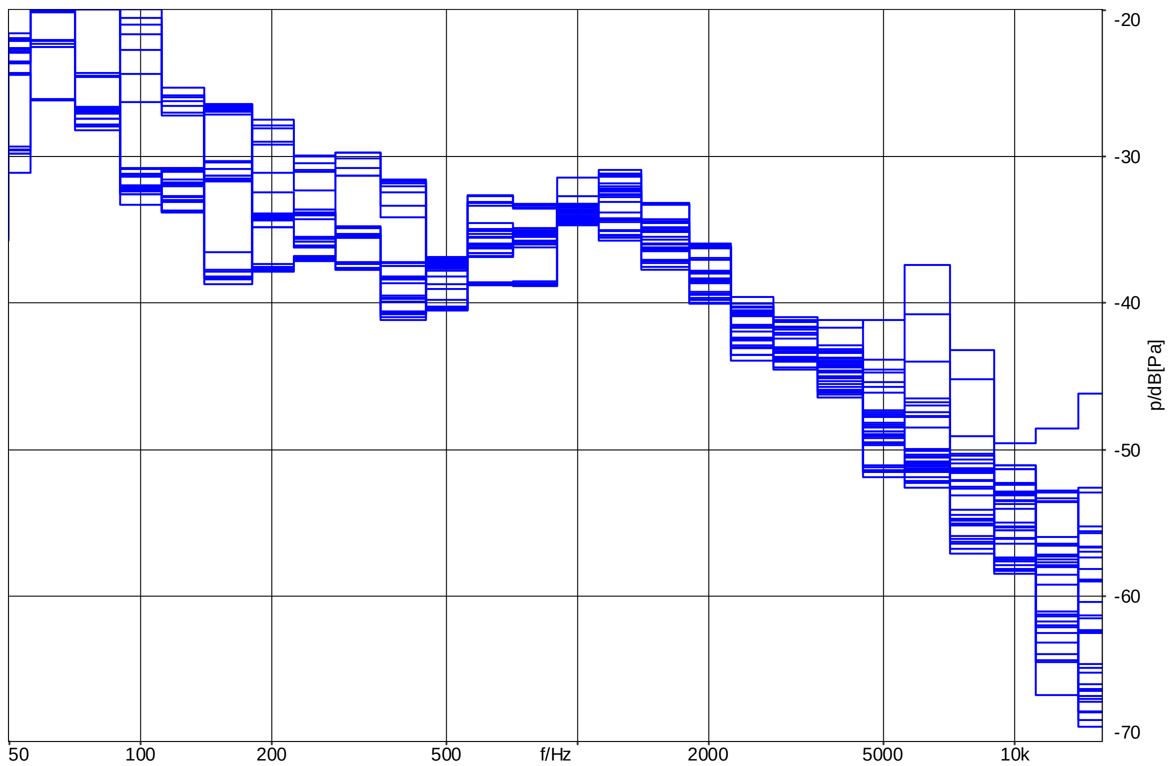


Figure 46: 1/3rd octave spectra at reference microphone and Crossroad noise from TS 103 224 (ES 202 396-1 methodology)

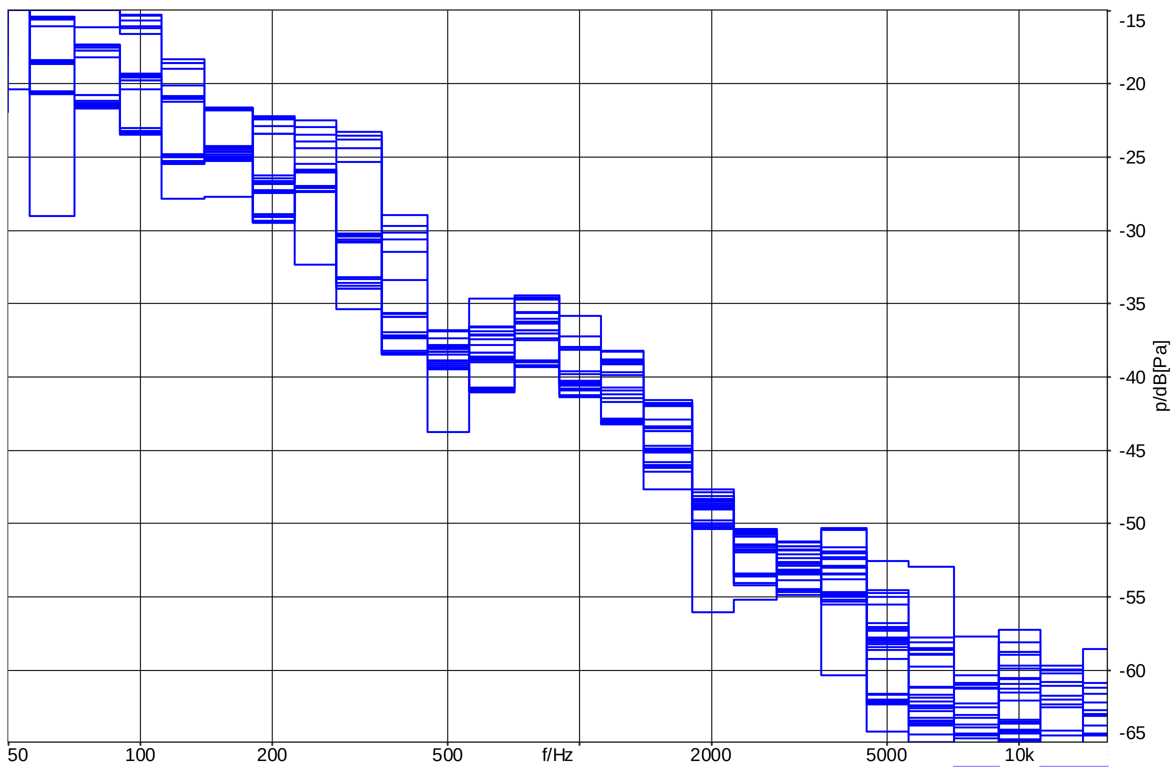


Figure 47: 1/3rd octave spectra at reference microphone and Inside Car noise from TS 103 224 (ES 202 396-1 methodology)

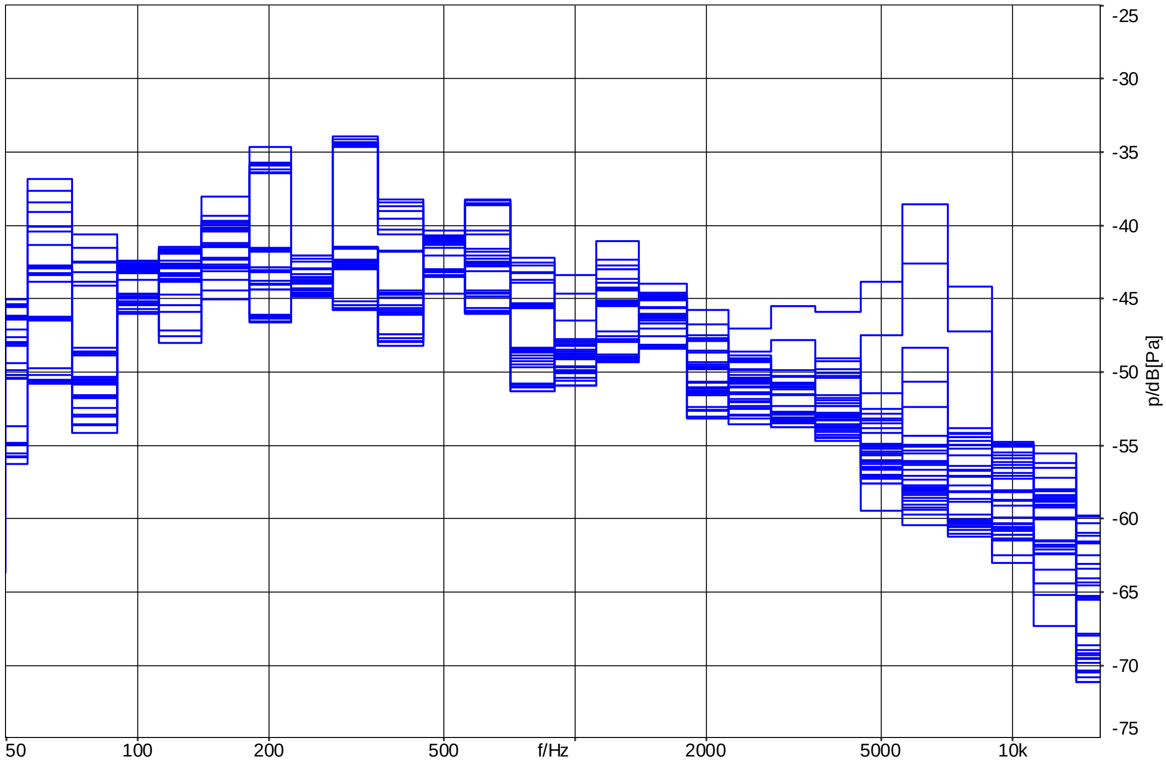


Figure 48: 1/3rd octave spectra at reference microphone and Office noise from TS 103 224 (ES 202 396-1 methodology)

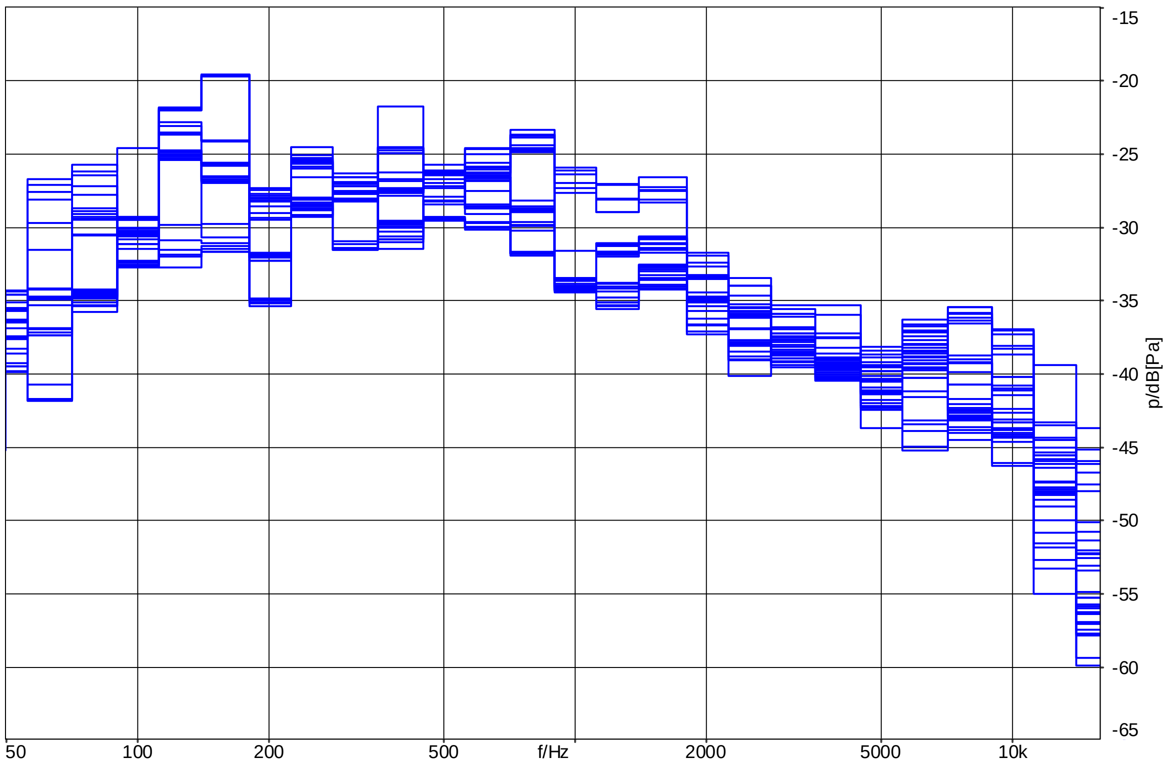


Figure 49: 1/3rd octave spectra at reference microphone and Pub noise from TS 103 224 (ES 202 396-1 methodology)

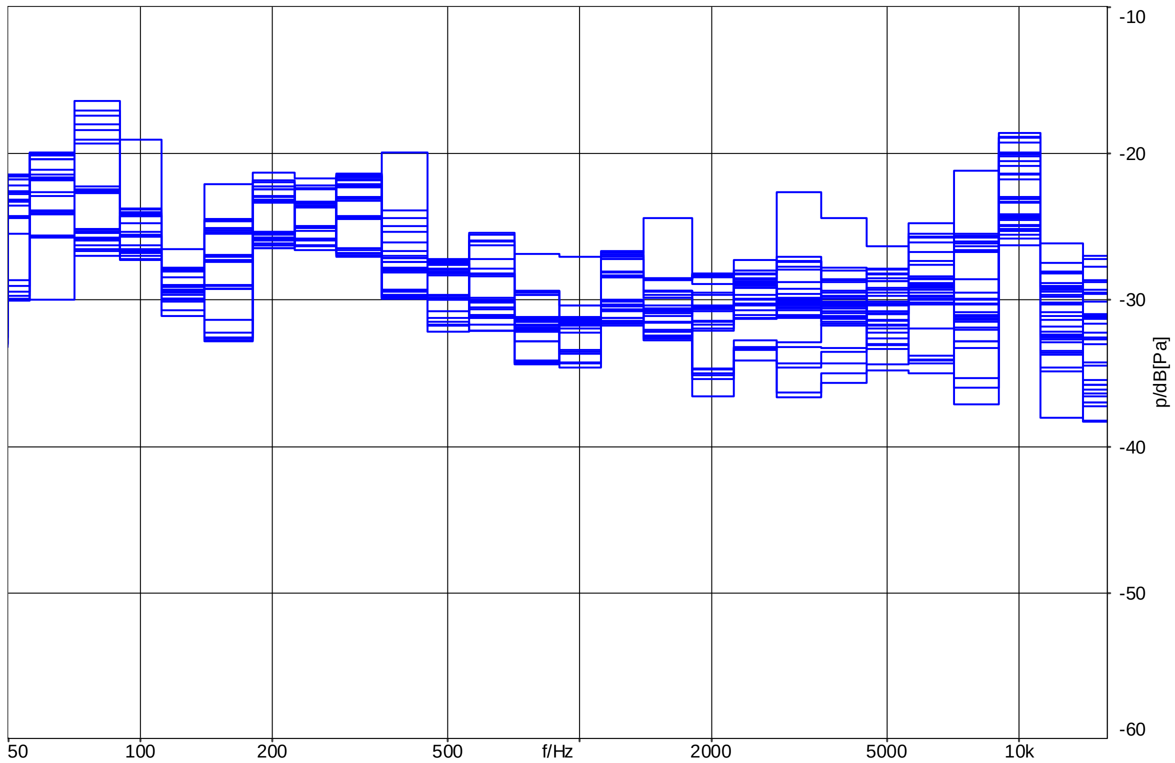


Figure 50: 1/3rd octave spectra at reference microphone and Train Station noise from TS 103 224 (ES 202 396-1 methodology)

4.2.3.5.3 Simulation and Recordings acc. to TS 103 224

The differences in the spectra reproduced by the TS 103 224 simulation method are significantly lower compared to the ones of ES 202 396-1, as shown in Figure 51 to Figure 56. The differences seem also to be independent of the noise type.

The differences range from 1 dB up to 10 dB. As expected, the sound field reproduction is highly accurate and consistent across labs in the low frequency domain up to about 2 kHz (where most energy of the noise is found). Here the spectral differences are within a range of 1 - 2 dB. Up to about 8 kHz, the differences are still less than 3 dB and from 10 - 20 kHz, the differences mostly remain below 5 dB.

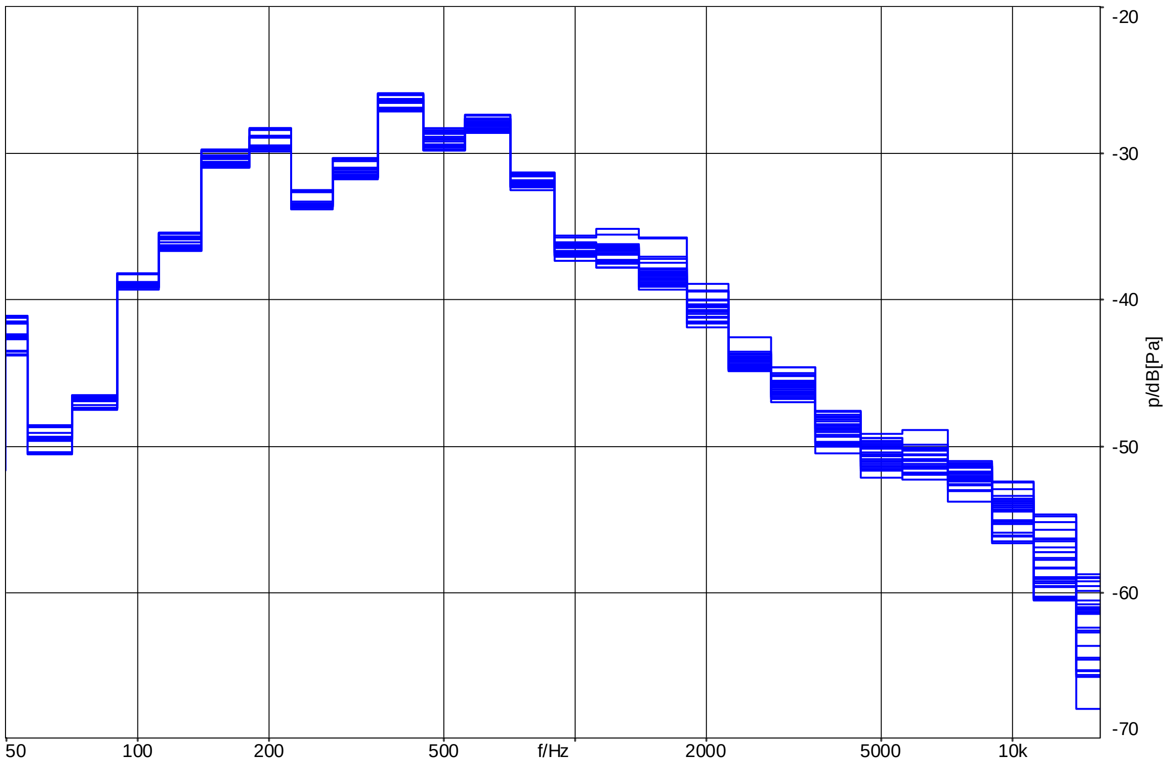


Figure 51: 1/3rd octave spectra at reference microphone and Cafeteria noise from TS 103 224 (TS 103 224 methodology)

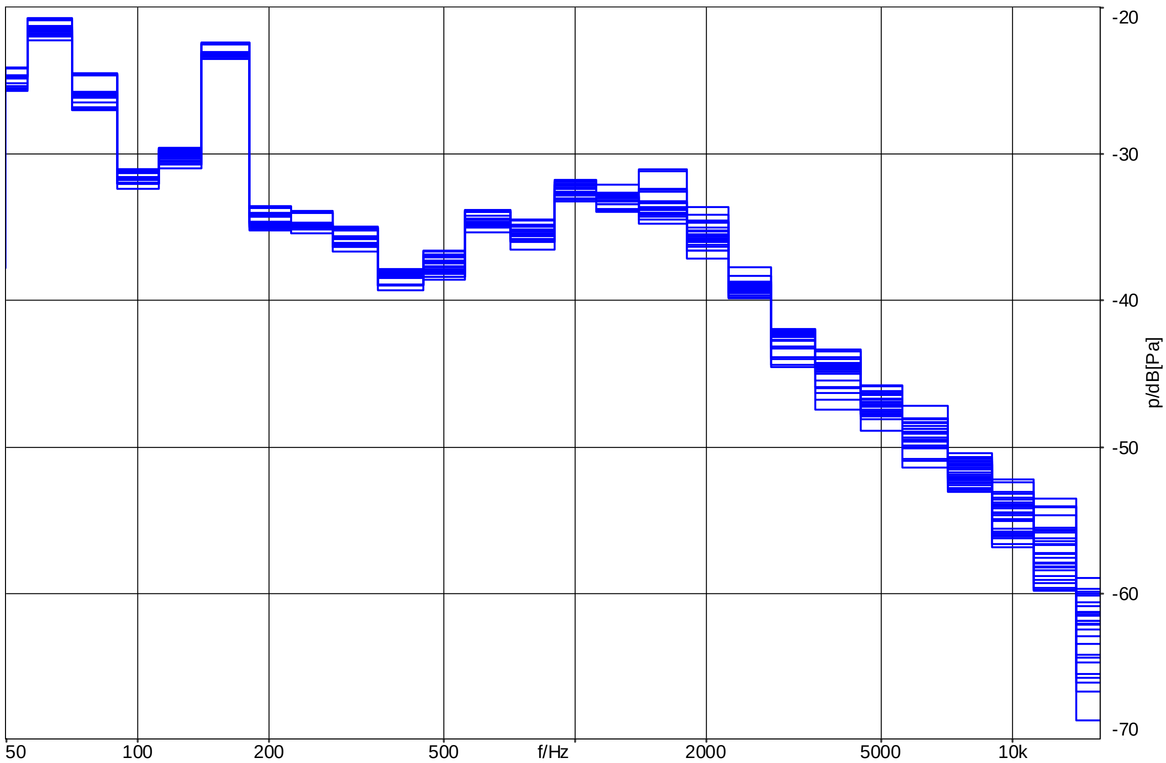


Figure 52: 1/3rd octave spectra at reference microphone and Crossroad noise from TS 103 224 (TS 103 224 methodology)

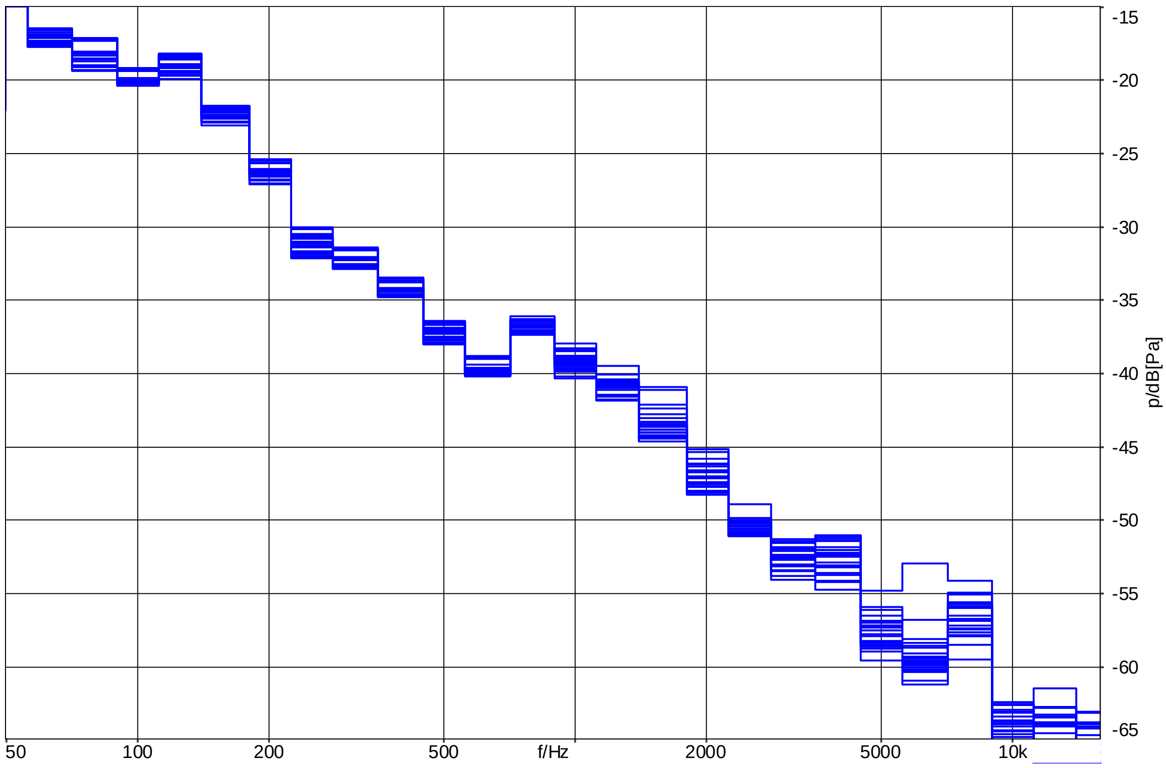


Figure 53: 1/3rd octave spectra at reference microphone and Inside Car noise from TS 103 224 (TS 103 224 methodology)

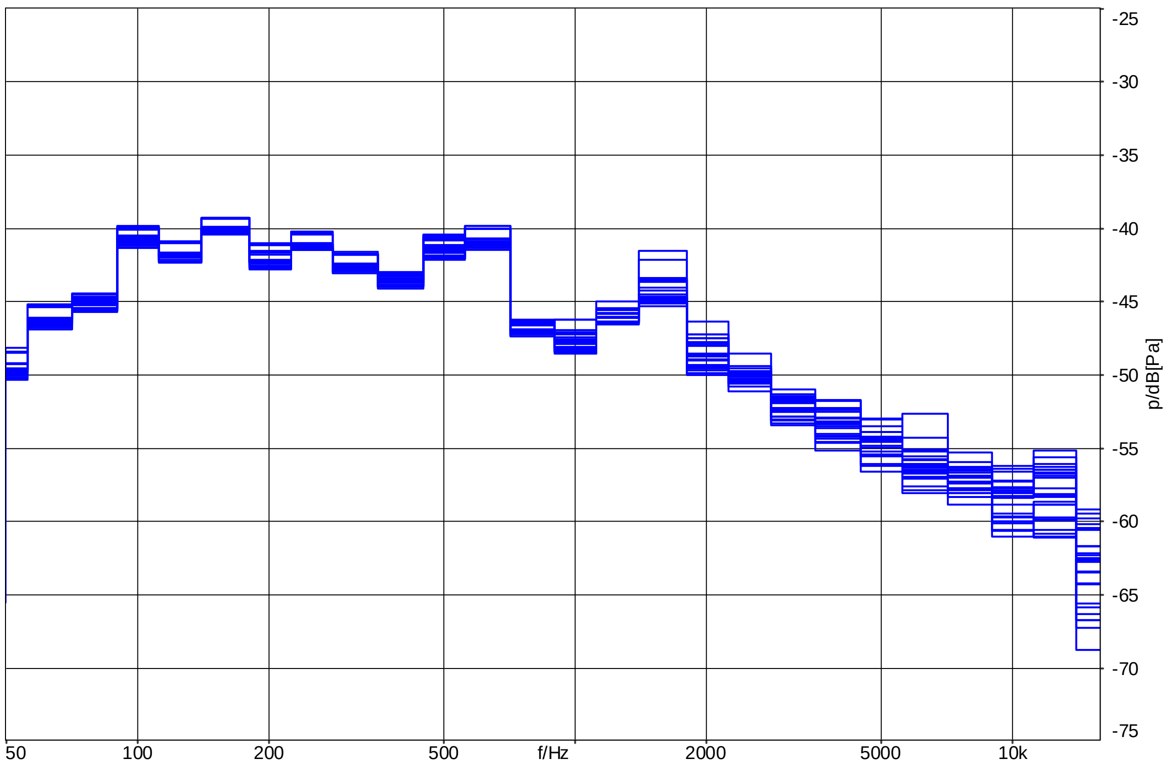


Figure 54: 1/3rd octave spectra at reference microphone and Office noise from TS 103 224 (TS 103 224 methodology)

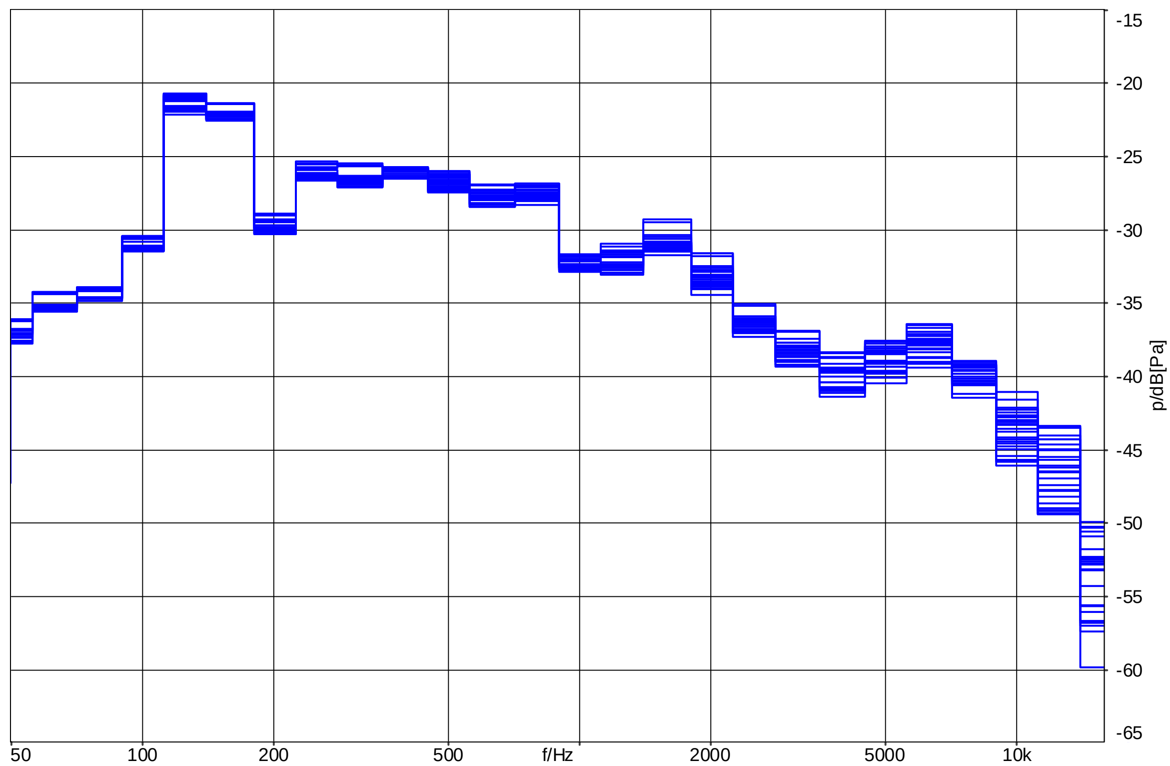


Figure 55: 1/3rd octave spectra at reference microphone and Pub noise from TS 103 224 (TS 103 224 methodology)

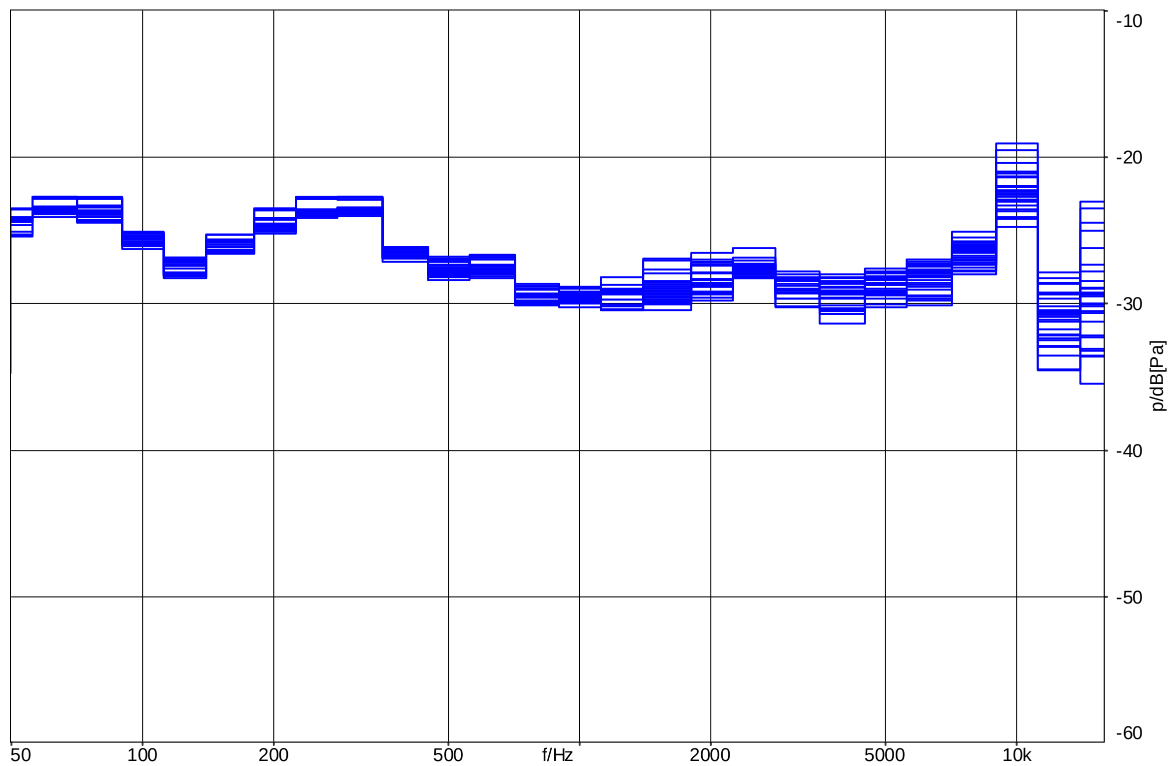


Figure 56: 1/3rd octave spectra at reference microphone and Train Station noise from TS 103 224 (TS 103 224 methodology)

4.2.4 Results for Desktop Hands-free

4.2.4.1 Introduction

In the following clauses, results of several measurements and analyses are presented for terminals in desktop hands-free mode. The test setup complies with clause 5.1.3.2 of [3]. Similar as for the handheld hands-free setup, possible differences in results across labs may be influenced and explained by:

- Calibration differences
- Setup differences
- Room differences
- Time variant behaviour of the device under test

It seems that these parameters may have impact on the results in a similar or even bigger range than the experiments which include the background noise simulation.

4.2.4.2 Comparison of inter-lab accuracy for the different background noise simulation methods

4.2.4.2.1 Wideband

4.2.4.2.1.1 No background noise

In this clause, results under silent conditions in WB mode are presented, as shown in Figure 57 to Figure 59. Basically, the variance to be expected in different labs with no background noise simulation present can be observed, as shown in Figure 57 to Figure 59.

It seems that these parameters described in clause 4.2.4.1 may have impact on the results in a similar or even bigger range than the experiments which include the background noise simulation.

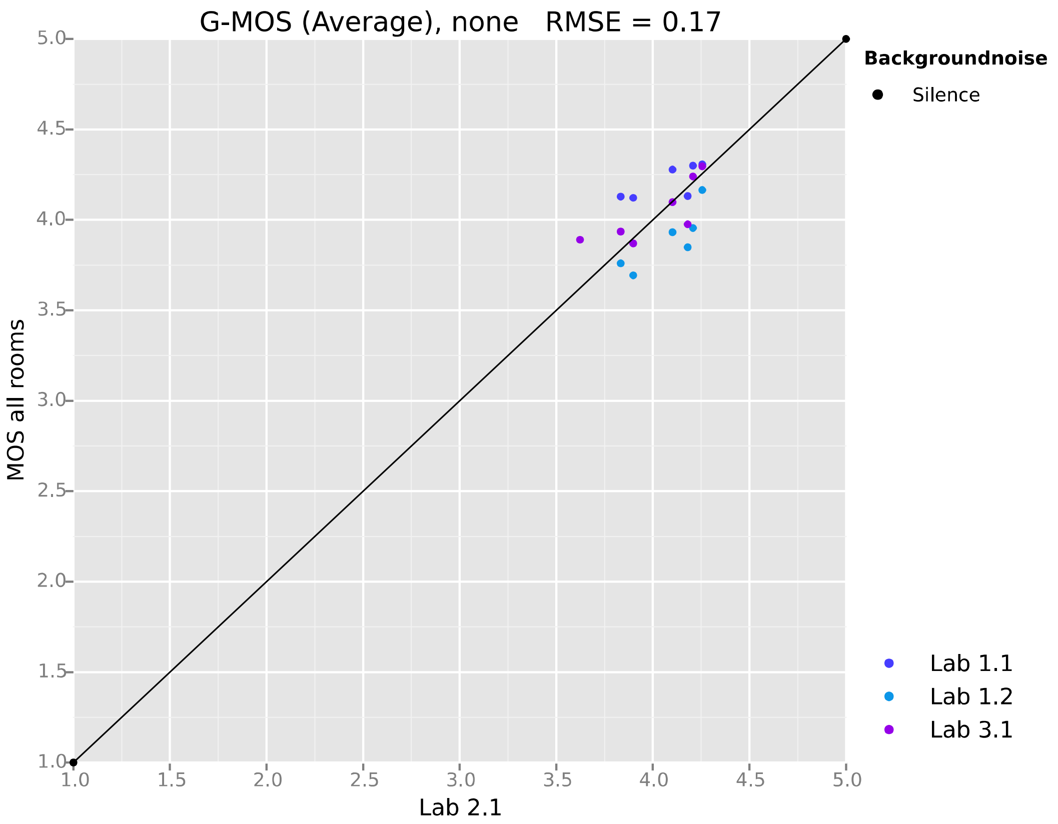


Figure 57: Correlation between G-MOS (WB) results from Lab 2.1 and other labs (Silence)

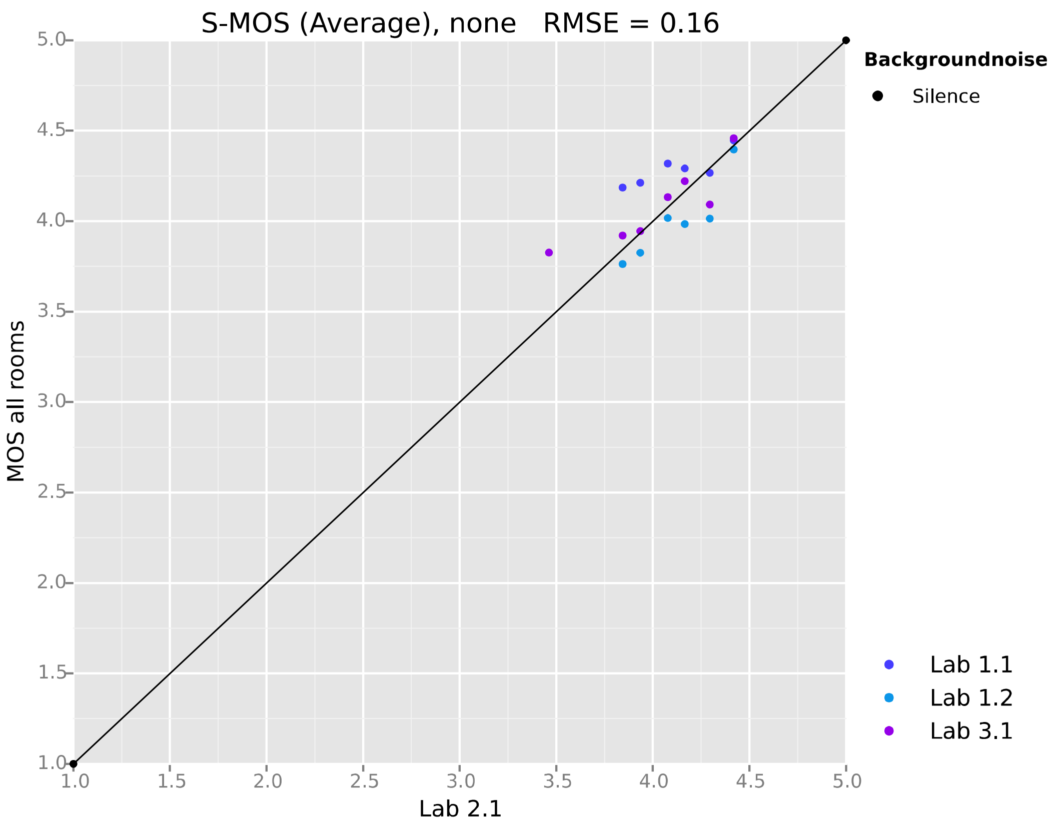


Figure 58: Correlation between S-MOS (WB) results from Lab 2.1 and other labs (Silence)

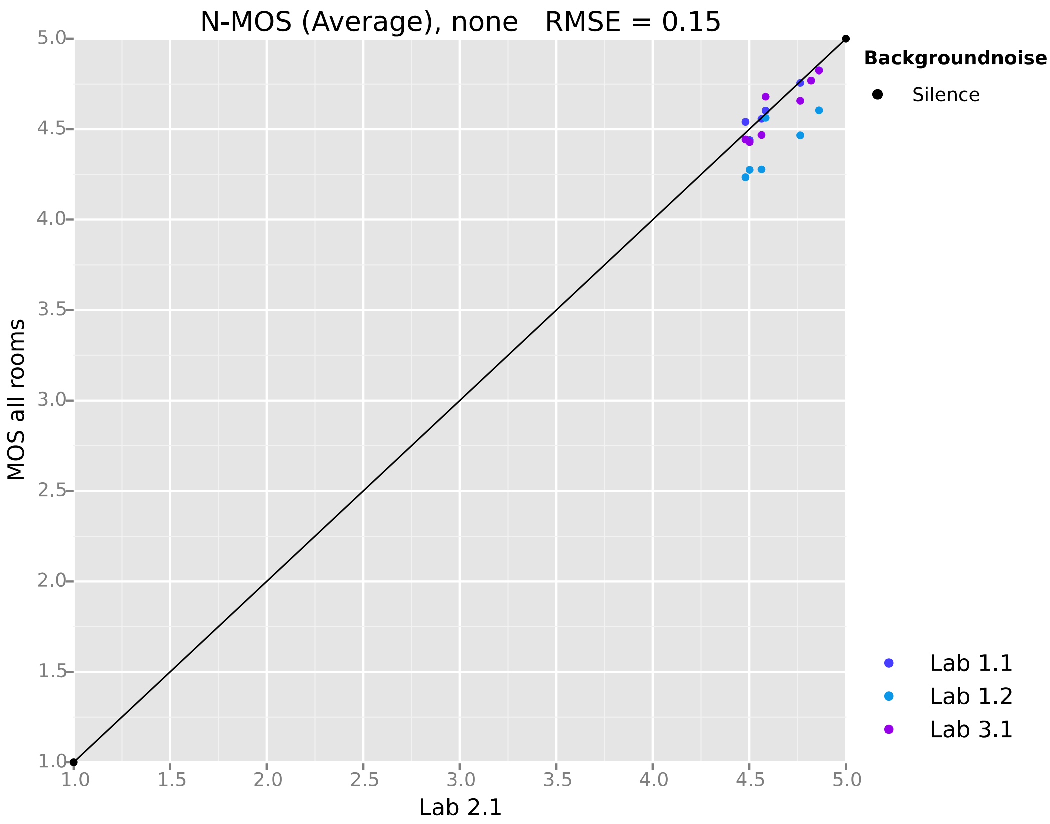


Figure 59: Correlation between N-MOS (WB) results from Lab 2.1 and other labs (Silence)

4.2.4.2.1.2 Simulation acc. to ES 202 396-1 with noises from TS 103 224 (binaural)

The results shown in Figure 66 to Figure 68 are based on using the ES 202 396-1 noise field simulation, but using the binaurally recorded background noises from TS 103 224. The following observations can be made:

- No systematic differences in results of other labs compared to Lab 2.1 can be observed.
- The RMSE values for S-, N- and G-MOS are 0.12.

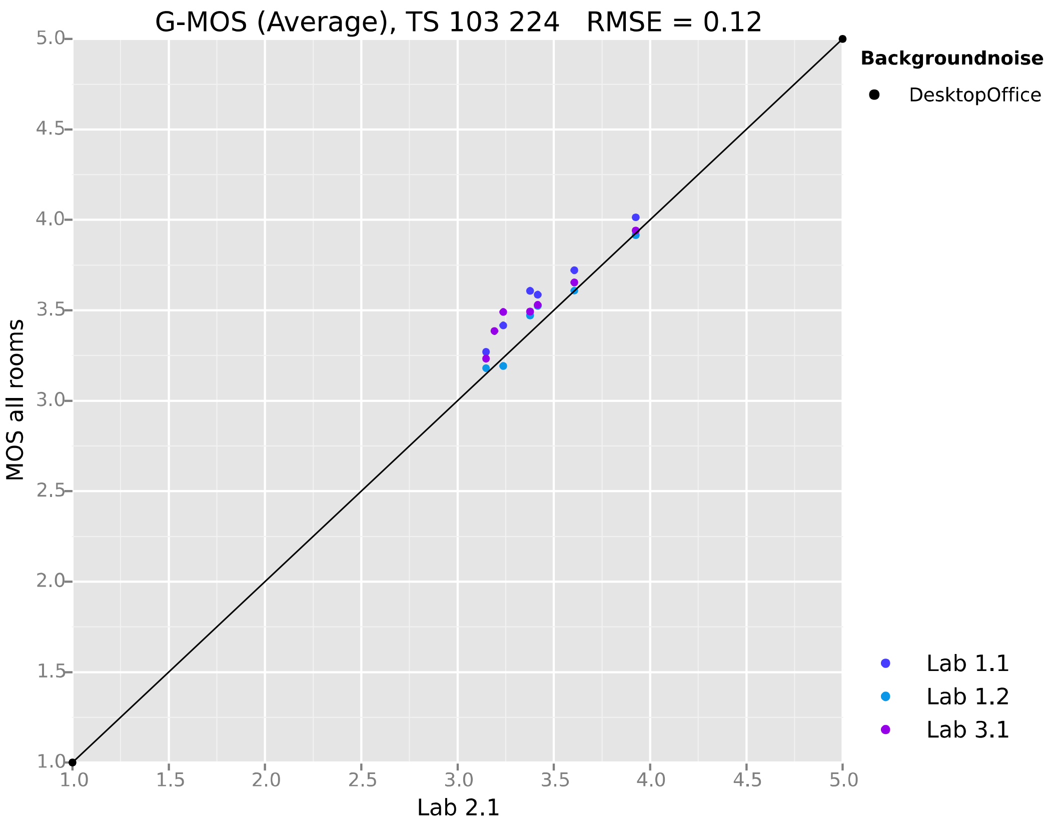


Figure 60: Correlation between G-MOS (WB) results from Lab 2.1 and other labs

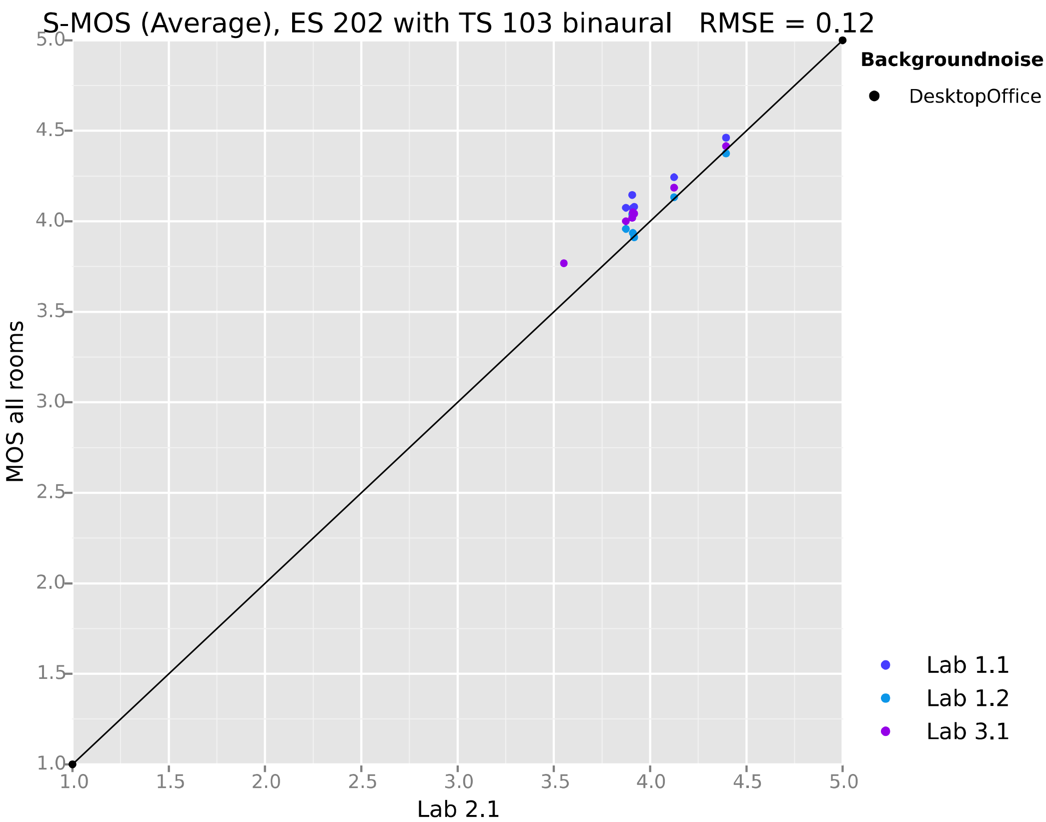


Figure 61: Correlation between S-MOS (WB) results from Lab 2.1 and other labs

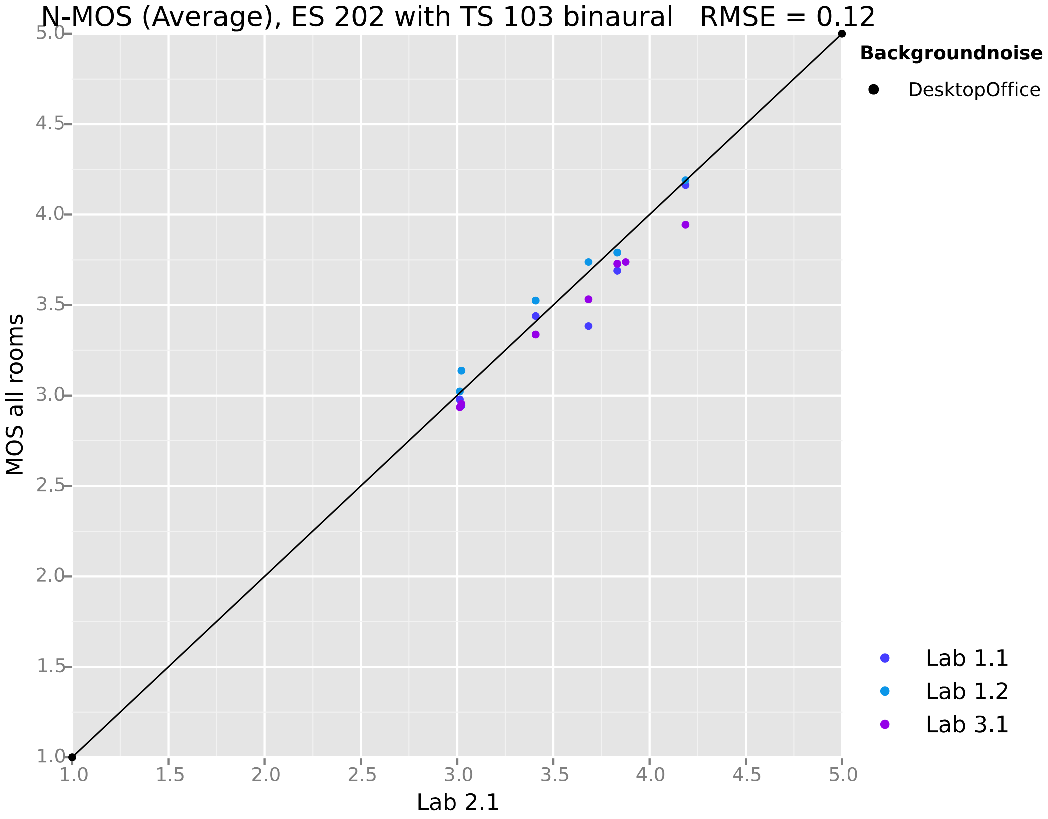


Figure 62: Correlation between N-MOS (WB) results from Lab 2.1 and other labs

4.2.4.2.1.3 Background Noises & Simulation acc. to TS 103 224

The results shown in Figure 63 to Figure 65 are based on using the TS 103 224 noise field simulation, using the 8-channel noise recording from the associated noise database. For this setup, the following observations can be made:

- No systematic differences in results of other labs compared to Lab 2.1 can be observed.
- The RMSE values for S-, N- and G-MOS are 0.07 – 0.17
- The lowest variation can be seen for N-MOS, the results are quite consistent

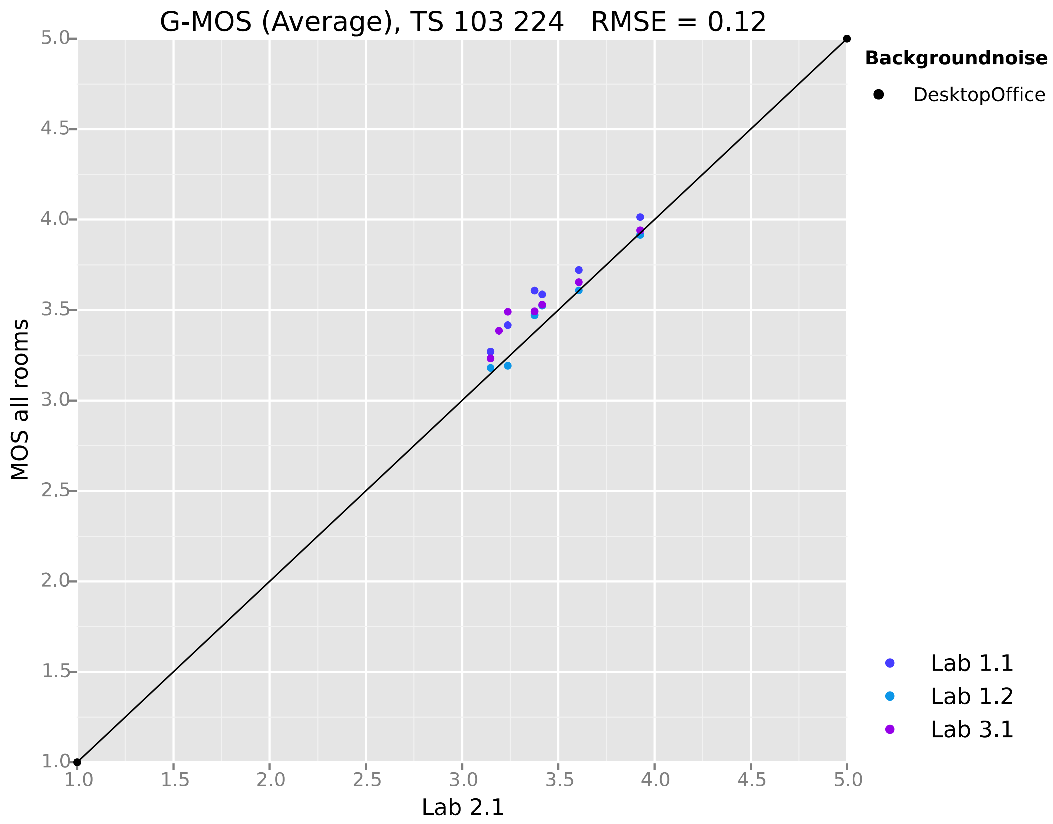


Figure 63: Correlation between G-MOS (WB) between Lab 2.1 and other labs

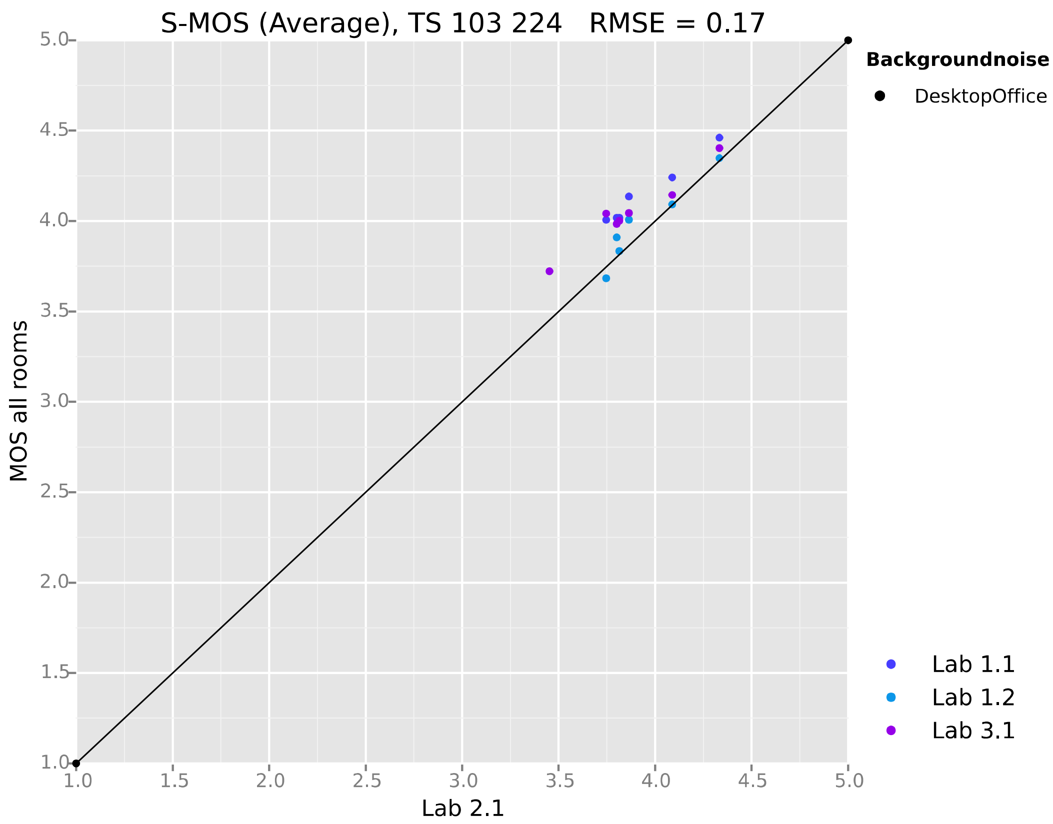


Figure 64: Correlation between S-MOS (WB) between Lab 2.1 and other labs

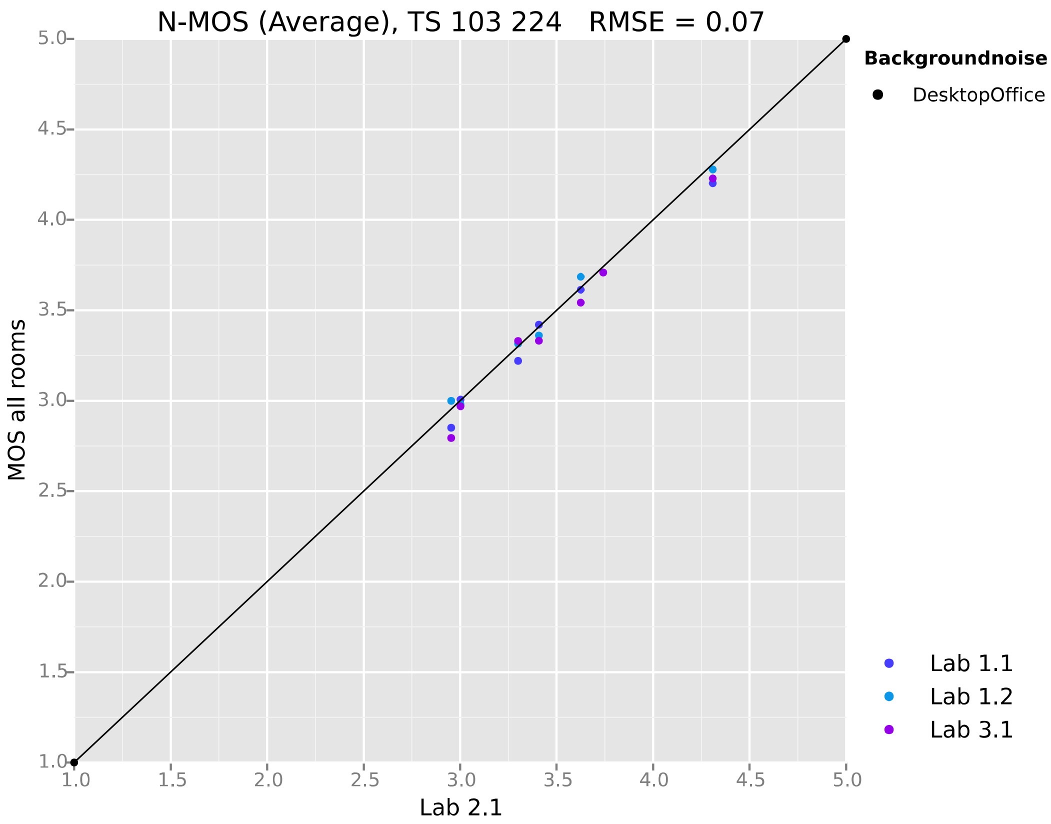


Figure 65: Correlation between N-MOS (WB) between Lab 2.1 and other labs

4.2.4.2.2 Narrowband

4.2.4.2.2.1 No background noise

In this clause, results under silent conditions in NB mode are presented, as shown in Figure 66 to Figure 68. Basically, the variance to be expected in different labs with no background noise simulation present can be observed, as shown in Figure 66 to Figure 68.

It seems that these parameters described in clause 4.2.4.1 may have impact on the results in a similar or even bigger range than the experiments which include the background noise simulation. The RMSE of the results is quite high, mainly influences by S-MOS.

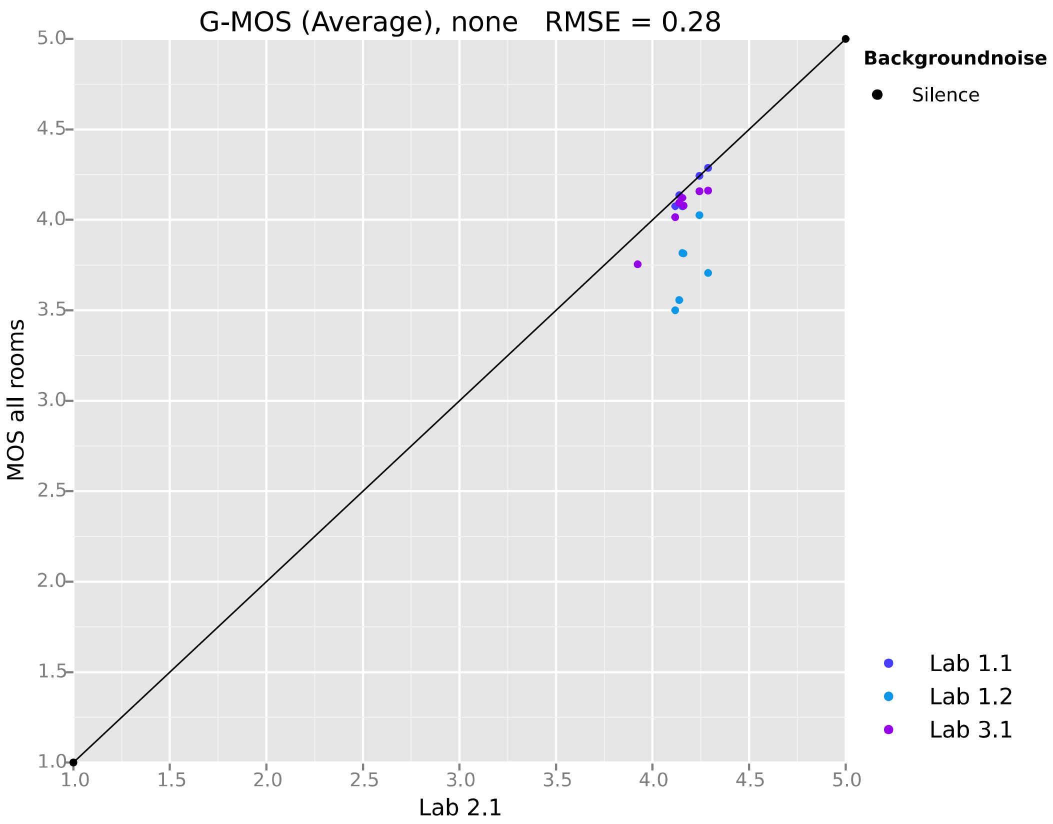


Figure 66: Correlation between G-MOS (NB) results from Lab 2.1 and other labs (Silence)

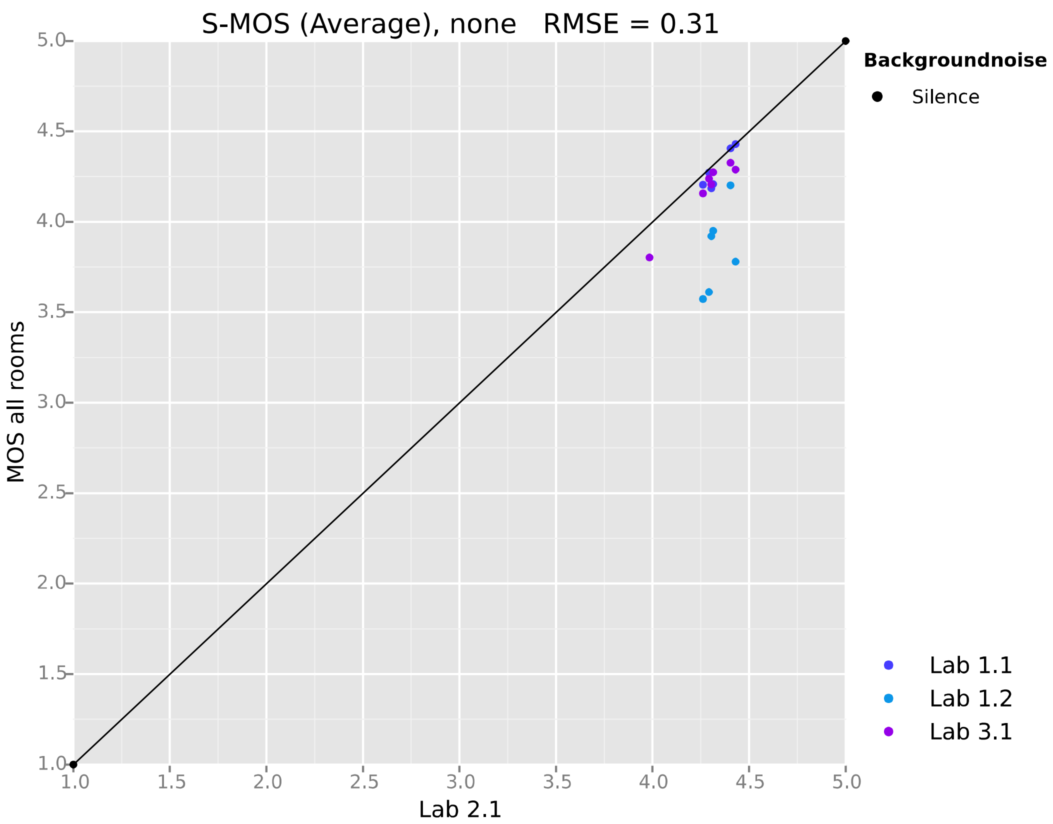


Figure 67: Correlation between G-MOS (NB) results from Lab 2.1 and other labs (Silence)

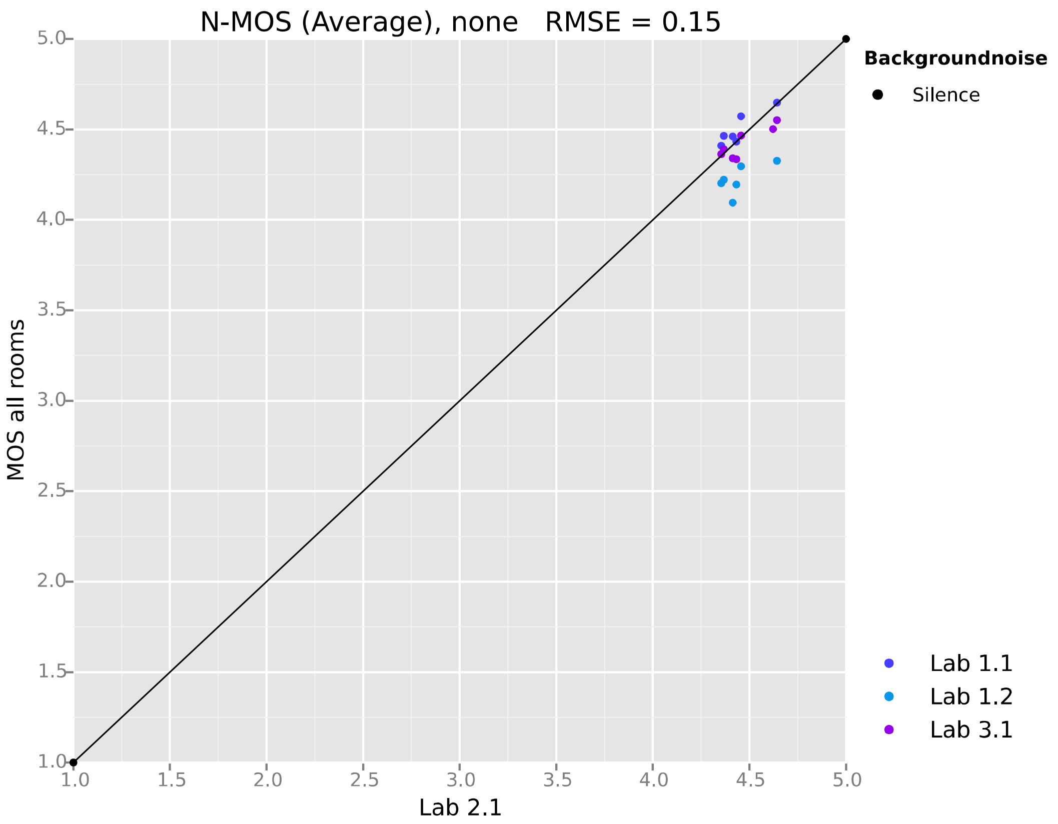


Figure 68: Correlation between G-MOS (NB) results from Lab 2.1 and other labs (Silence)

4.2.4.2.2.2 Simulation acc. to ES 202 396-1 with noises from TS 103 224 (binaural)

The results shown in Figure 69 to Figure 71 are based on using the ES 202 396-1 noise field simulation, but using the binaurally recorded background noises from TS 103 224. The following observations can be made:

- No systematic differences in results of other labs compared to Lab 2.1 can be observed.
- The RMSE values for S-, N- and G-MOS are ranging from 0.17 to 0.22.
- The results are less consistent between labs than in wideband.

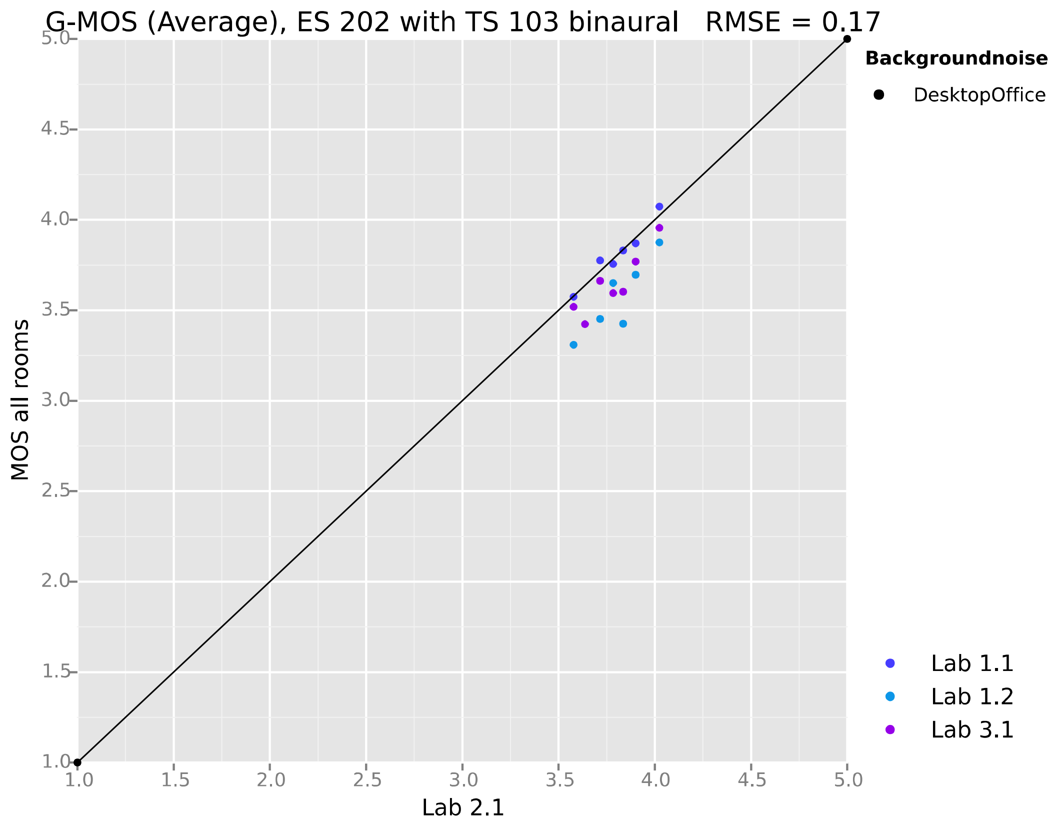


Figure 69: Correlation between G-MOS (NB) results from Lab 2.1 and other labs

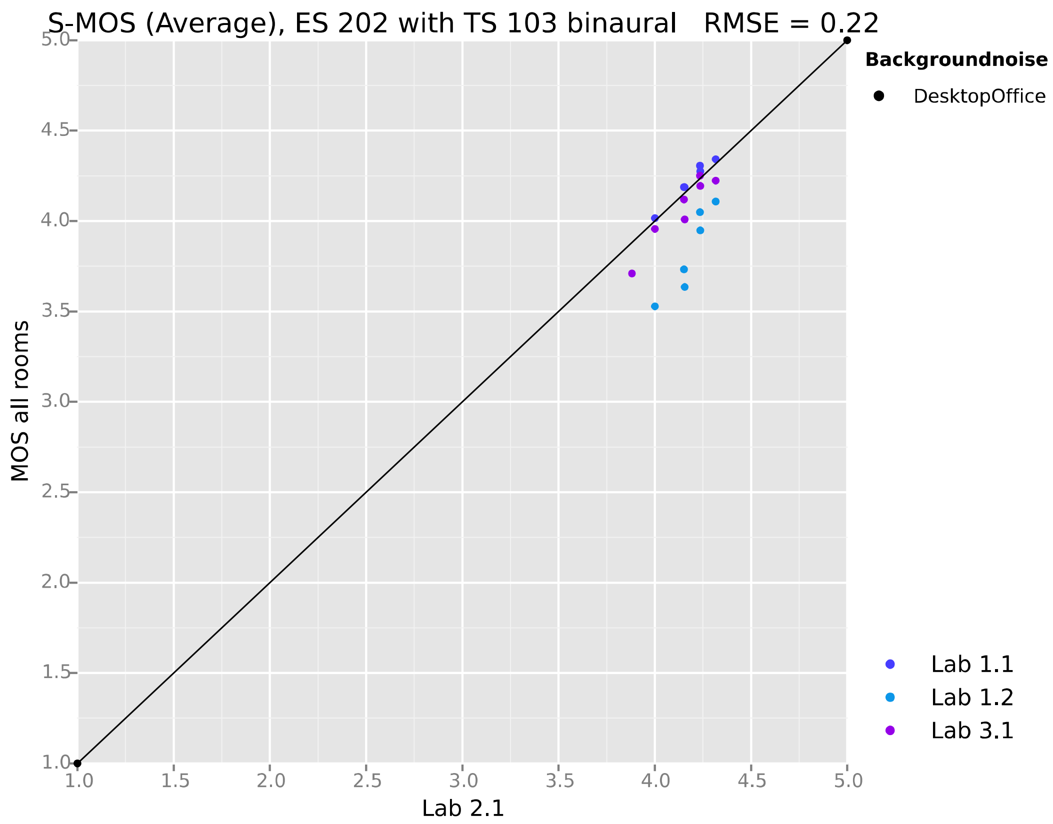


Figure 70: Correlation between S-MOS (NB) results from Lab 2.1 and other labs

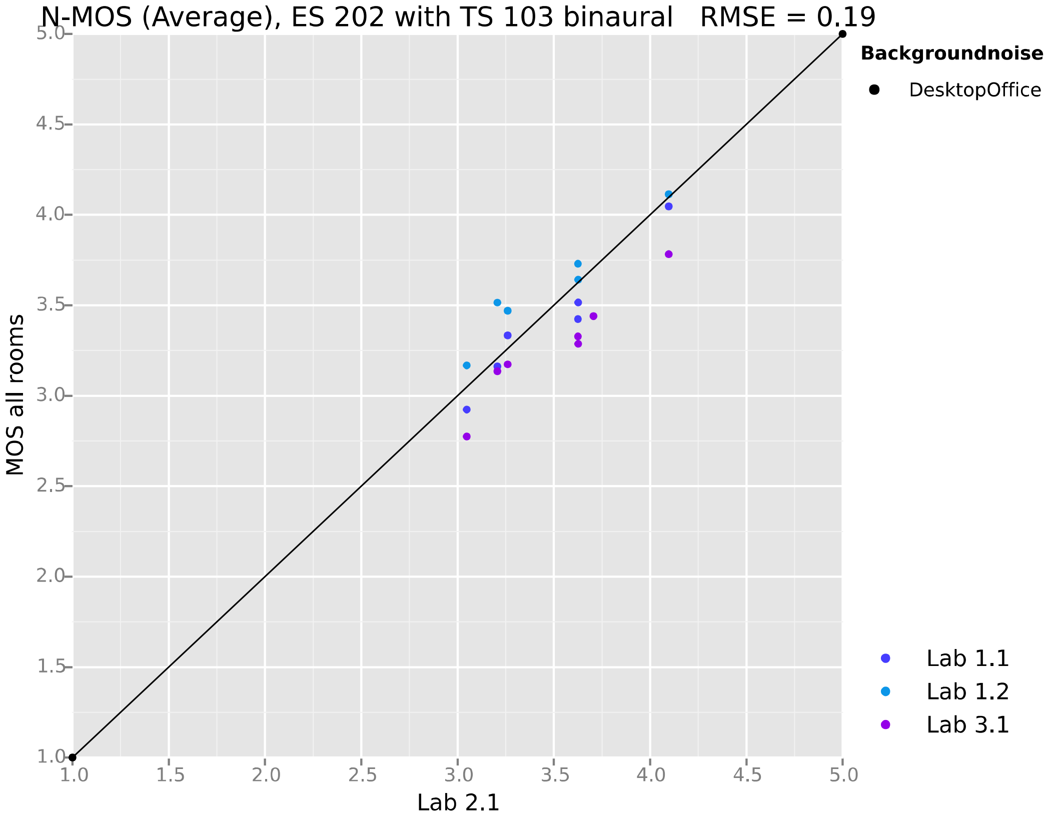


Figure 71: Correlation between N-MOS (NB) results from Lab 2.1 and other labs

4.2.4.2.2.3 Background Noises & Simulation acc. to TS 103 224

The results shown in Figure 72 to Figure 74 are based on using the TS 103 224 noise field simulation, using the 8-channel noise recording from the associated noise database. For this setup, the following observations can be made:

- No systematic differences in results of other labs compared to Lab 2.1 can be observed.
- The RMSE values for S-, N- and G-MOS are in the range of 0.13 – 0.2.
- The lowest variation can be seen for N-MOS, the results are quite consistent.

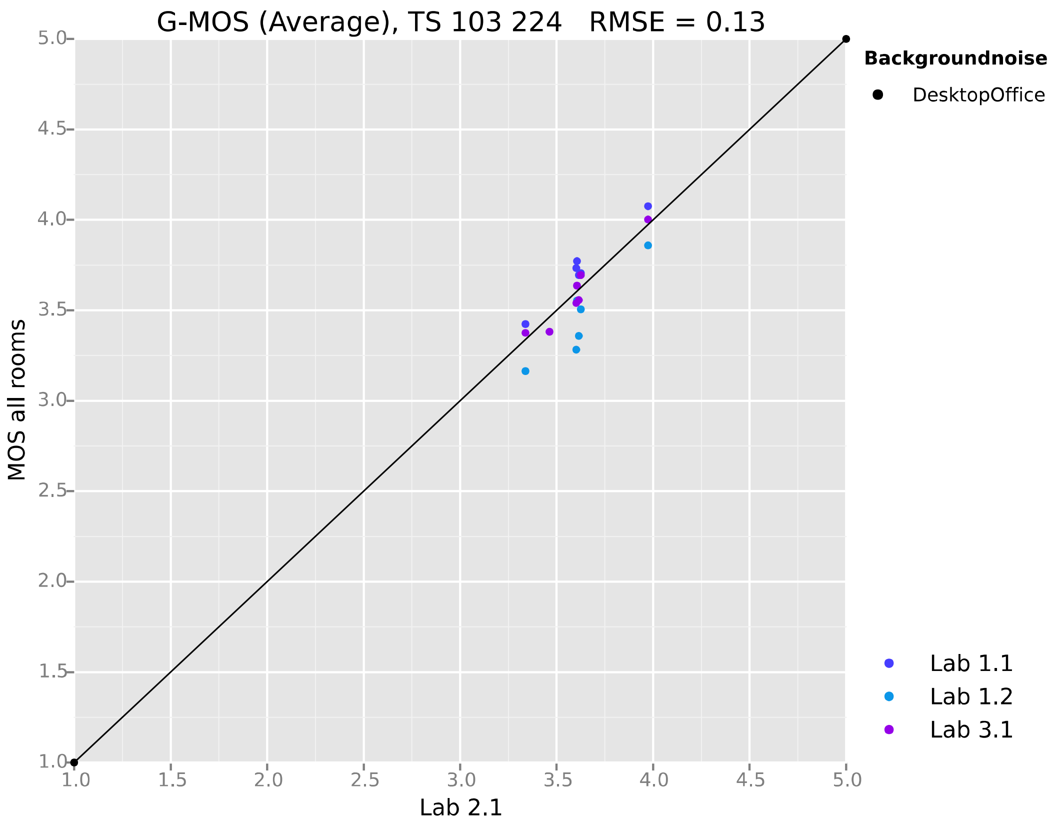


Figure 72: Correlation between G-MOS (NB) between Lab 2.1 and other labs

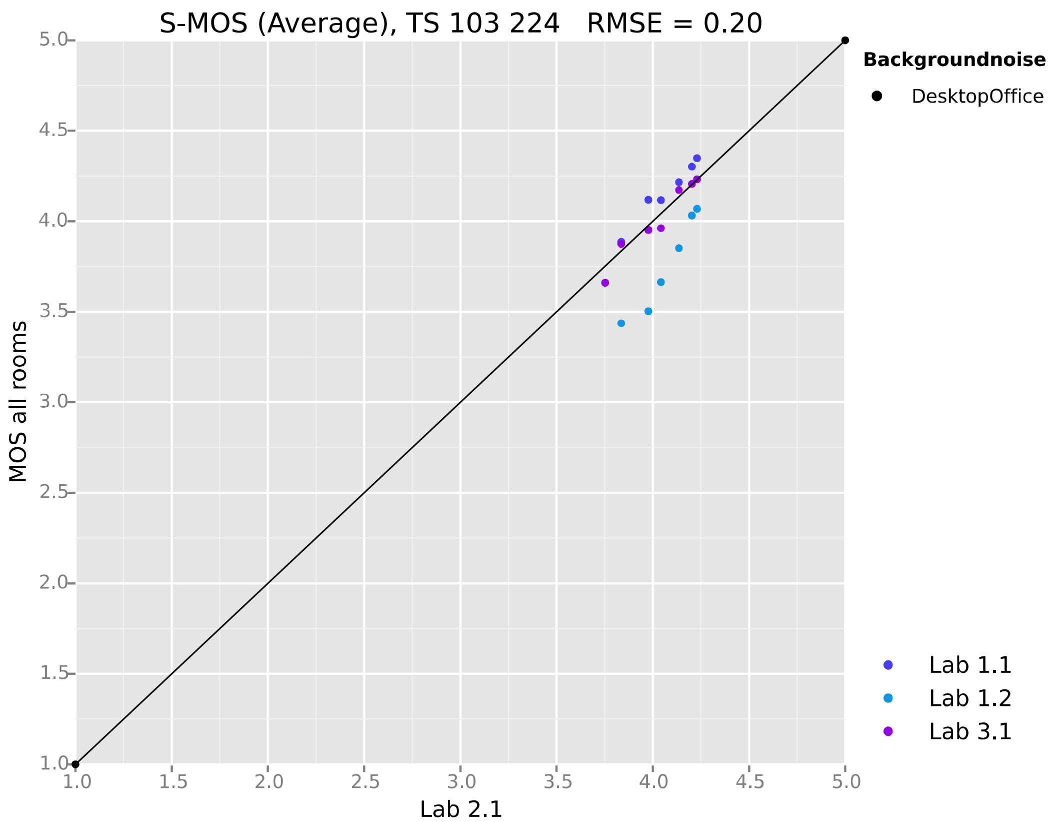


Figure 73: Correlation between S-MOS (NB) between Lab 2.1 and other labs

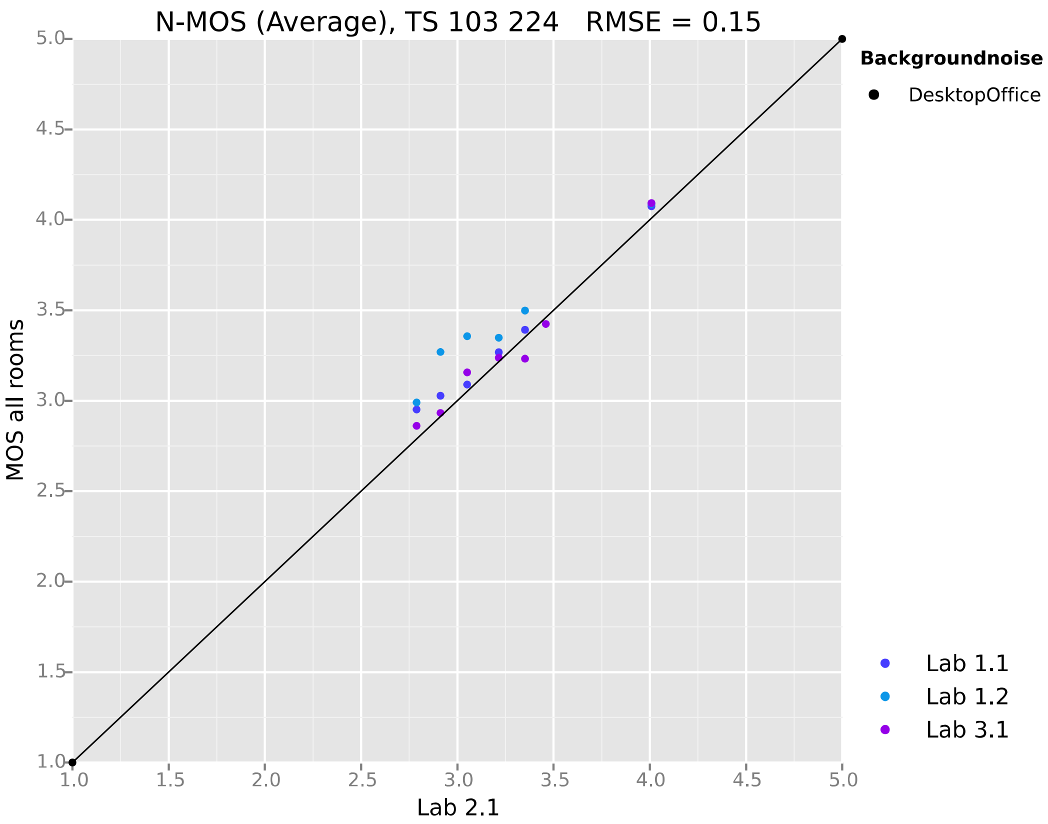


Figure 74: Correlation between N-MOS (NB) between Lab 2.1 and other labs

4.2.4.3 Comparison between ES 202 396-1 and TS 103 224 background noise simulation technique

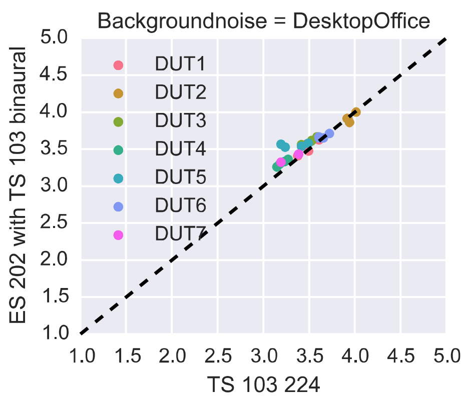
4.2.4.3.1 Introduction

In this clause, S-, N-, and G-MOS as measured in the different rooms for different terminals is analysed. The comparison is focused on the differences between background noise simulation systems. The comparison is carried out between each method and the different sets of background noises. The results measured using the ETSI ES 202 396-1 background noise simulation system are used as the reference (all noises, all labs, and all terminals). The results obtained using the ETSI TS 103 224 simulation system are plotted on the abscissa (x-axis) for comparison.

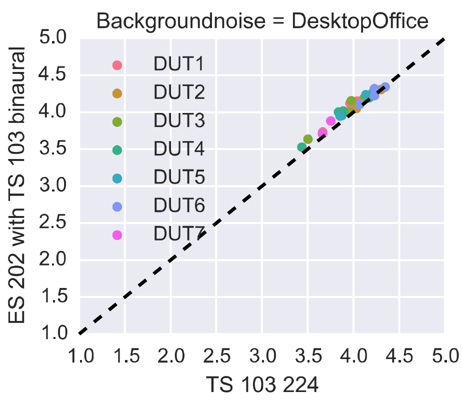
In contrast to the related analysis for handheld hands-free devices (see clause 4.2.3.3), only one noise type according to TS 103 224 is used (Office/Callcenter). The results of the different devices are colour-coded as indicated in the following figures.

4.2.4.3.2 Wideband

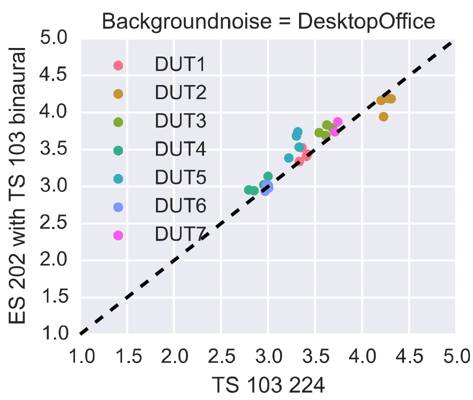
The results are shown in Figure 75. In general, a fairly good agreement between the background noise simulations can be noticed for S-MOS. Some deviations of the averaged results for N-MOS leads to subsequent deviations in G-MOS as well.



G-MOS



S-MOS

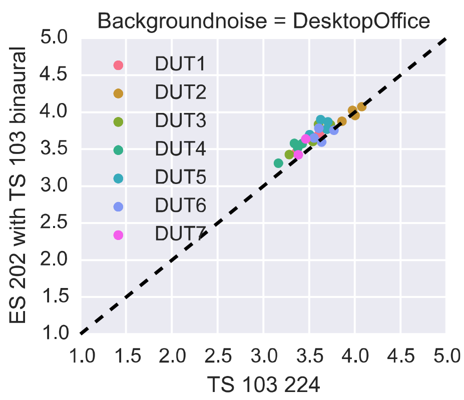


N-MOS

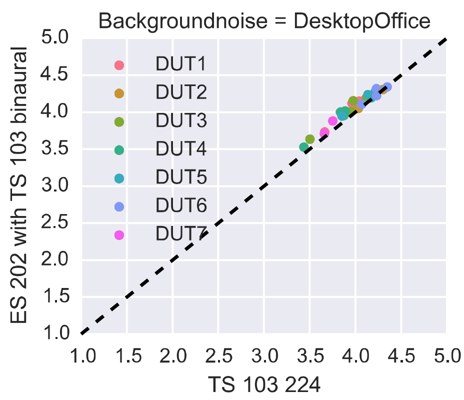
Figure 75: Correlation of G-MOS (WB) results between TS 103 224 and ES 202 396-1 methodology (noise from TS 103 224)

4.2.4.3.3 Narrowband

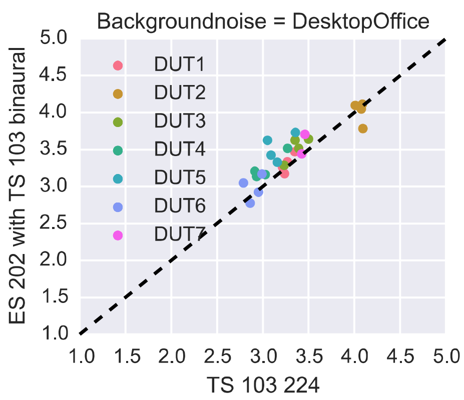
Similar as in wideband mode, a fairly good agreement between the background noise simulations can be observed for S-MOS in Figure 76. Some deviation of the averaged results in N-MOS again leads to subsequent deviations in G-MOS as well.



G-MOS



S-MOS



N-MOS

Figure 76: Correlation of G-MOS (NB) results between TS 103 224 and ES 202 396-1 methodology (noise from TS 103 224)

4.2.4.4 Analysis results according to TS 26.131 & TS 26.132

4.2.4.4.1 Introduction

In this clause, the speech quality analysis as specified in [3] and [2] is carried out. Averaging the results across all noise conditions provides one single S-MOS, N-MOS and G-MOS for each device.

4.2.4.4.2 Wideband

4.2.4.4.2.1 Comparison of absolute Results

The results in Figure 77 show the same tendency for all phones when measured using the different background noise simulation setups. The rank order in performance remains unchanged.

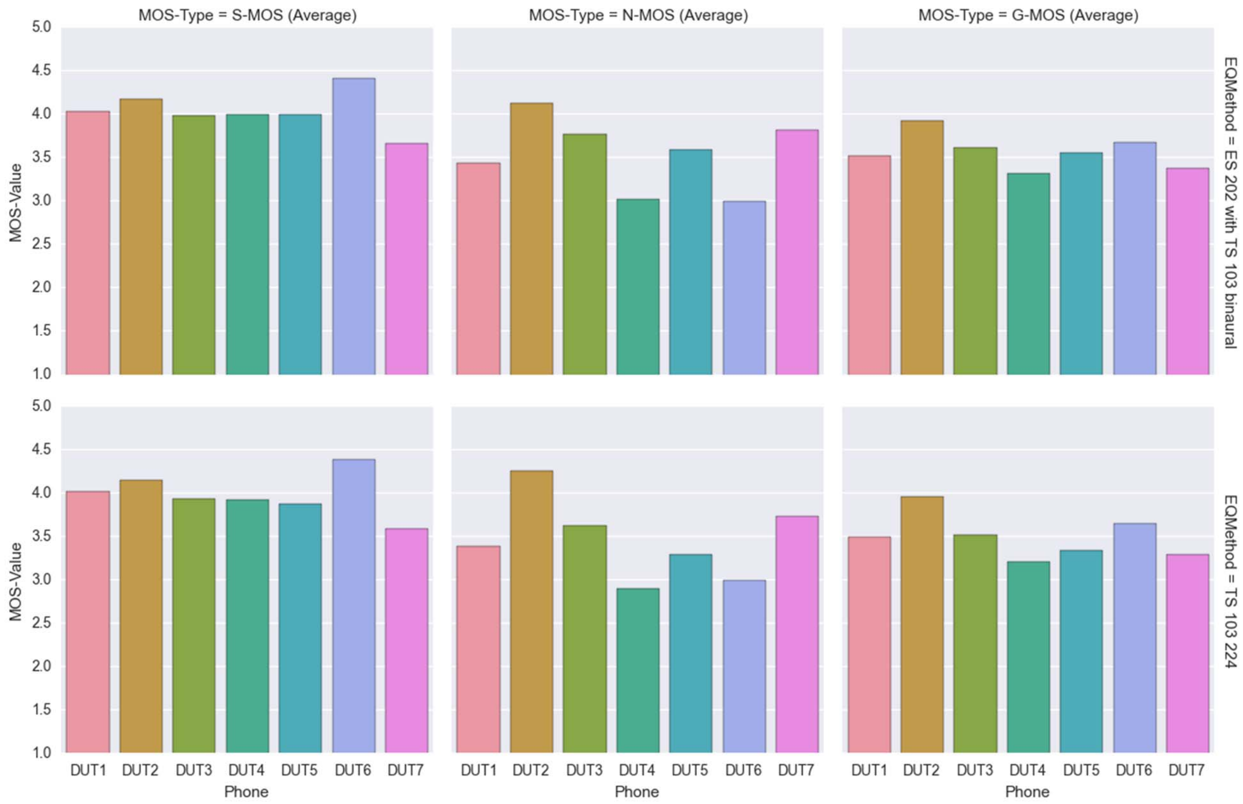


Figure 77: S-/N-/G-MOS (WB) averaged across background noises and all labs

4.2.4.4.2.2 Differences between TS 103 224 and ES 202 396-1 with noises from TS 103 224

When using the binaurally recorded background noises from TS 103 224 for the tests the differences when comparing ES 202 396-1 with TS 103 224 (see Figure 78) are slightly higher than the deviations observed in HHHF mode (see Figure 41). The offset is not constant, but depends on the type of terminal. Except for DUT2, the performance for all terminals is slightly better when using the ES 202 396-1 simulation method compared to the sound field simulation technique described in TS 103 224.

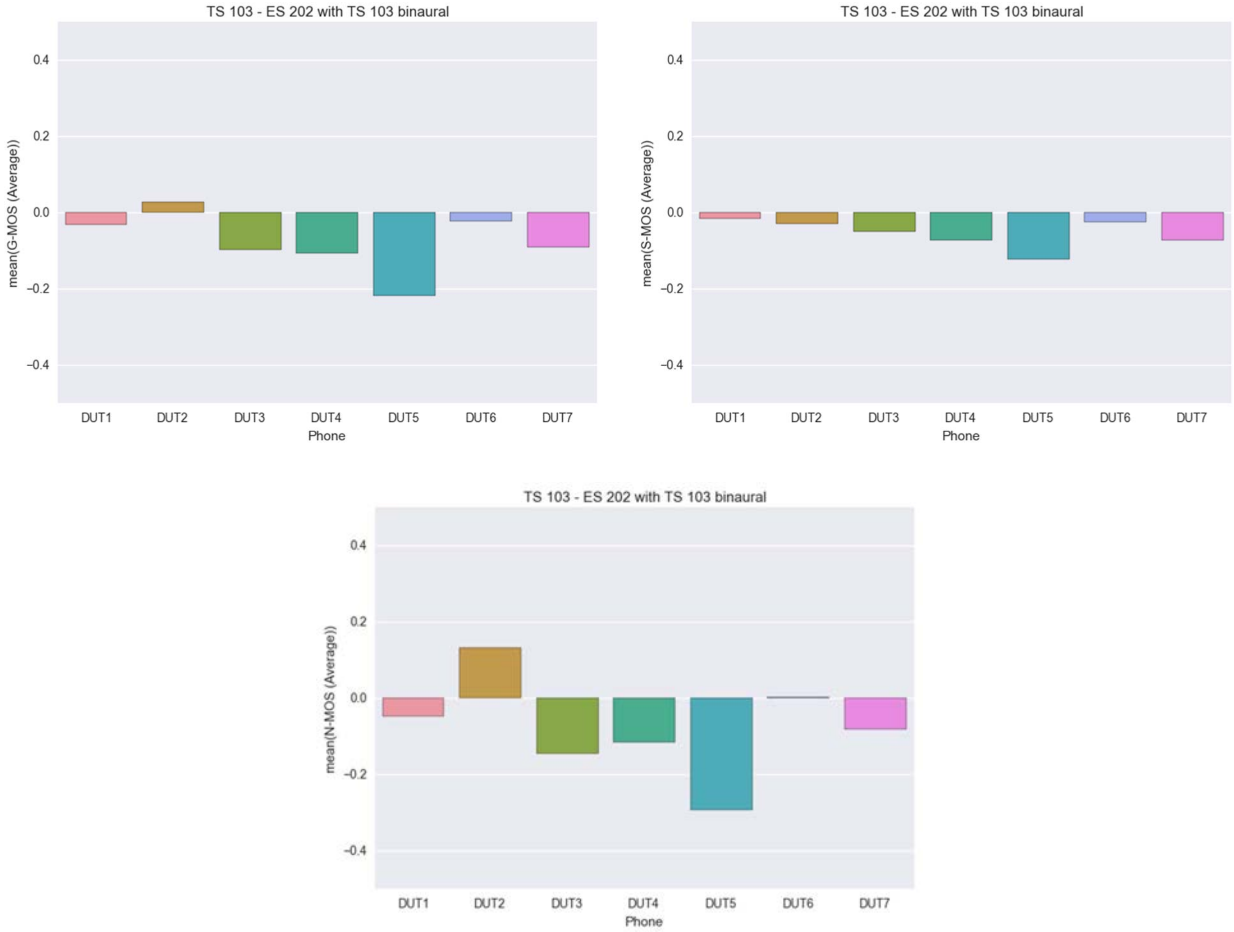


Figure 78: Differences (WB) between TS 103 224 and ES 202 396-1 methodology (binaural noises from TS 103 224 database)

4.2.4.4.3 Narrowband

4.2.4.4.3.1 Comparison of absolute Results

The results in Figure 79 show the same tendency for all phones when measured using the different background noise simulation setups. The rank order in performance remains unchanged.

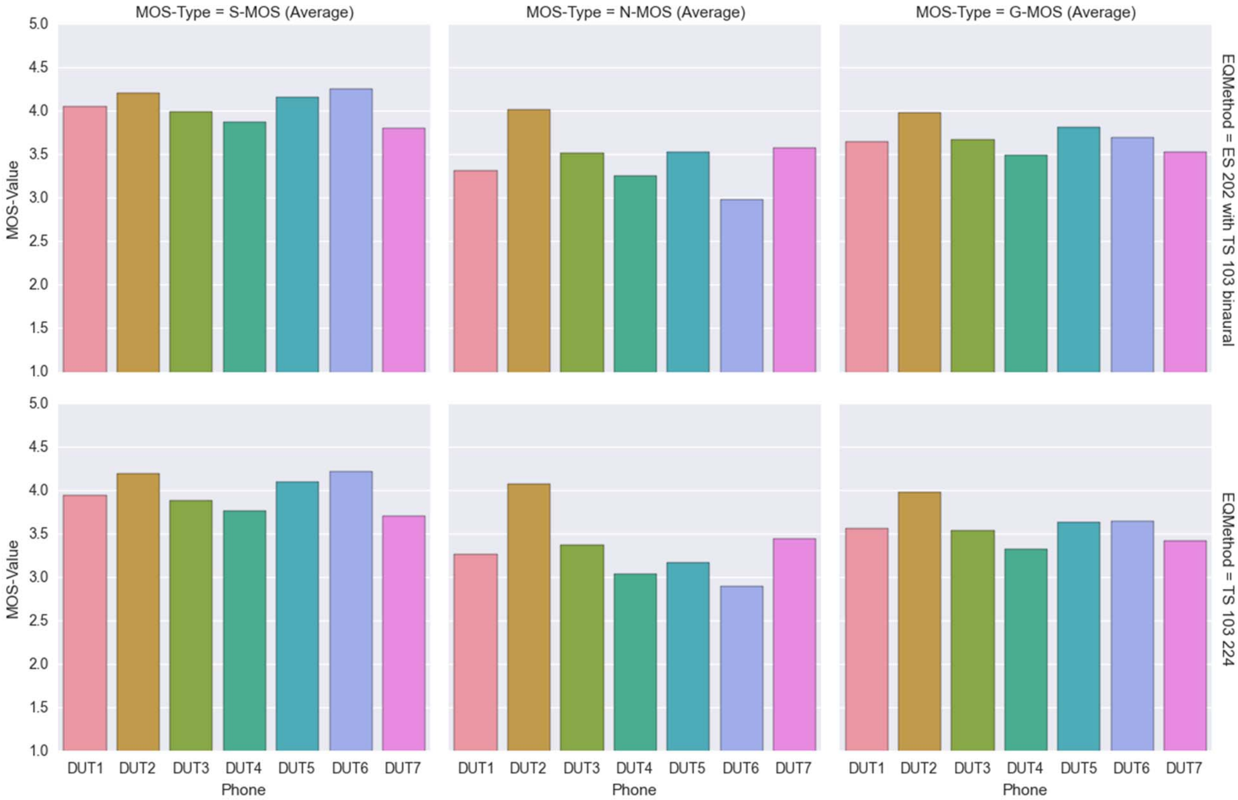


Figure 79: S-/N-/G-MOS (NB) averaged across background noises and all labs

4.2.4.4.3.2 Differences between TS 103 224 and ES 202 396-1 with noises from TS 103 224

When using the binaurally recorded background noises from TS 103 224 for the tests, the differences when comparing ES 202 396-1 with TS 103 224 (see Figure 80) are slightly higher than the deviations observed in HHHF mode (see Figure 44). The offset is not constant, but depends on the type of terminal. Except for DUT2, the performance for all terminals is slightly better when using the ES 202 396-1 simulation method compared to the sound field simulation technique described in TS 103 224.

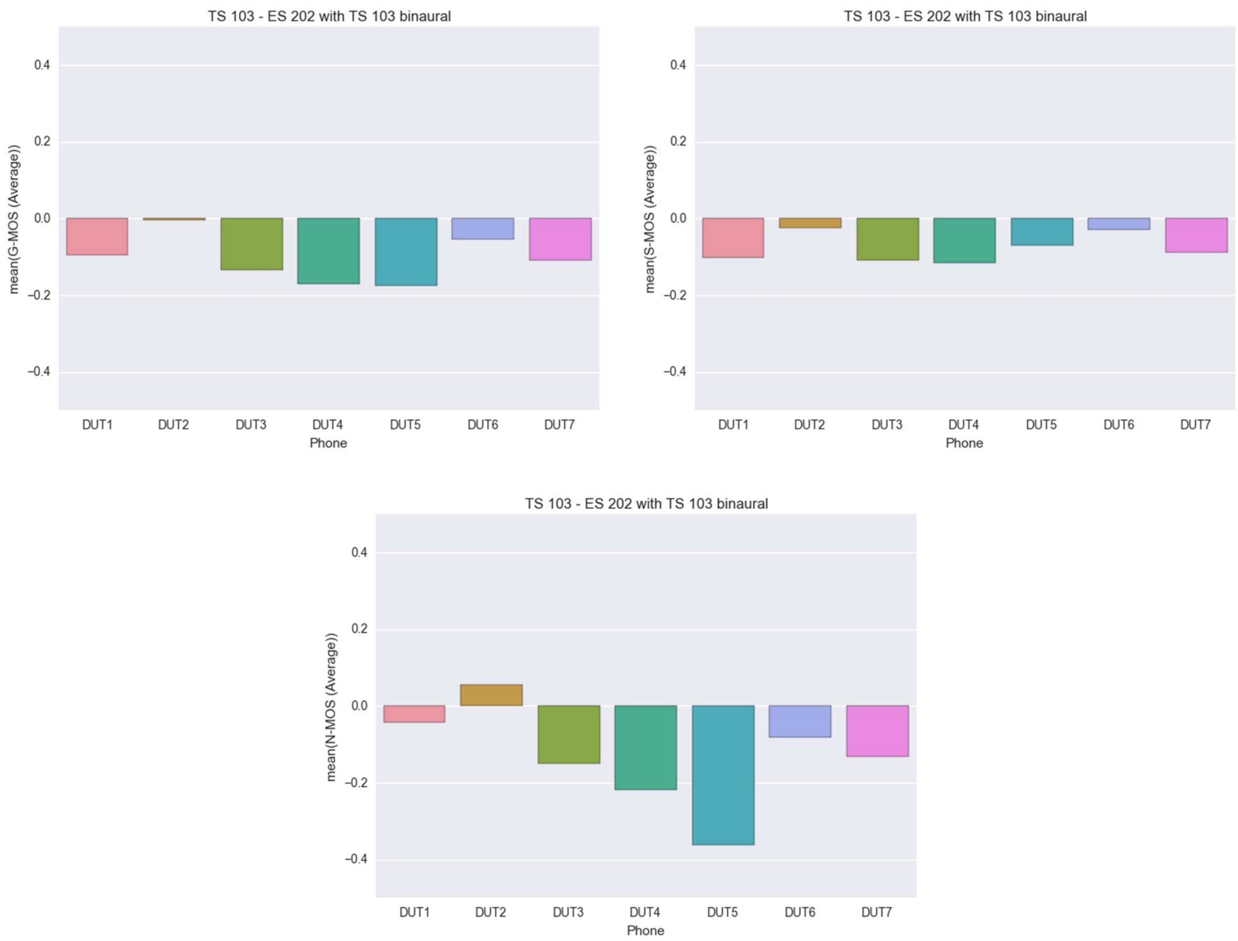


Figure 80: Differences (NB) between TS 103 224 and ES 202 396-1 methodology (binaural noises from TS 103 224 database)

4.2.4.5 Analyses of the noise spectra reproduced at the reference microphone

4.2.4.5.1 Introduction

The background noise spectra measured at the reference microphone (input signal for the analysis according to TS 103 106 [6]) of all devices under test are plotted into one diagram. In contrast to the corresponding analysis for HHHF (see clause 4.2.3.5), only one noise type according to TS 103 224 (Office/Callcenter) was evaluated.

4.2.4.5.2 Simulation acc. to ES 202 396-1 with recordings from TS 103 224, reference recording at place of DUT-microphone

As already seen for the HHHF experiments (see clause 4.2.3.5.2), quite large differences can be observed in Figure 81 for the spectra reproduced by the ES 202 396-1 simulation method in the different labs.

The differences range from 5 dB up to 15 dB. The largest differences are located in the low frequency range.

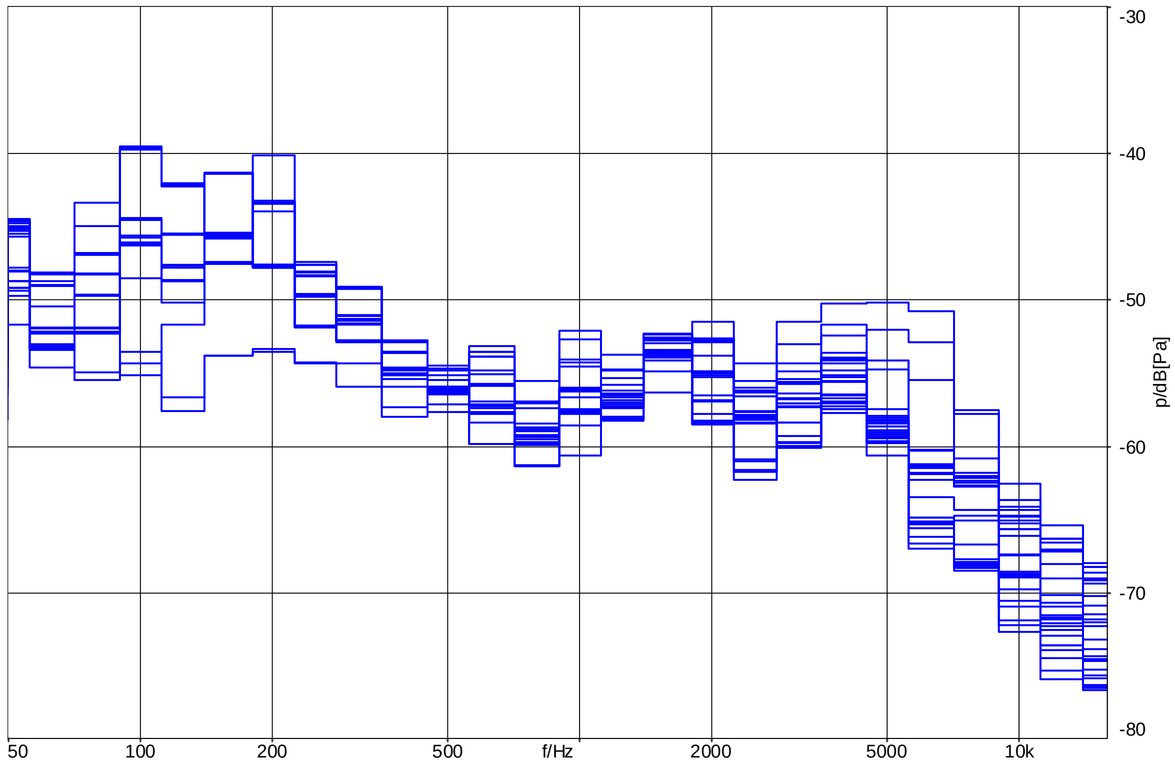


Figure 81: 1/3rd octave spectra at reference microphone and Office noise from TS 103 224 (ES 202 396-1 methodology)

4.2.4.5.3 Simulation & Recordings acc. to TS 103 224, reference recording at place of DUT-microphone

The differences in the spectra reproduced by the TS 103 224 simulation method are significantly lower compared to the ES 202 396-1 simulation method, as shown in Figure 82.

The differences range from 1 dB up to 8 dB. As expected, the sound field reproduction is highly accurate and consistent across labs in the low frequency domain up to about 2 kHz (where most energy of the noise is found). Here the spectral differences are within a range of 1 - 2 dB. Up to about 8 kHz, the differences are still less than 5 dB and from 10 - 20 kHz, the differences mostly remain below 5 dB.

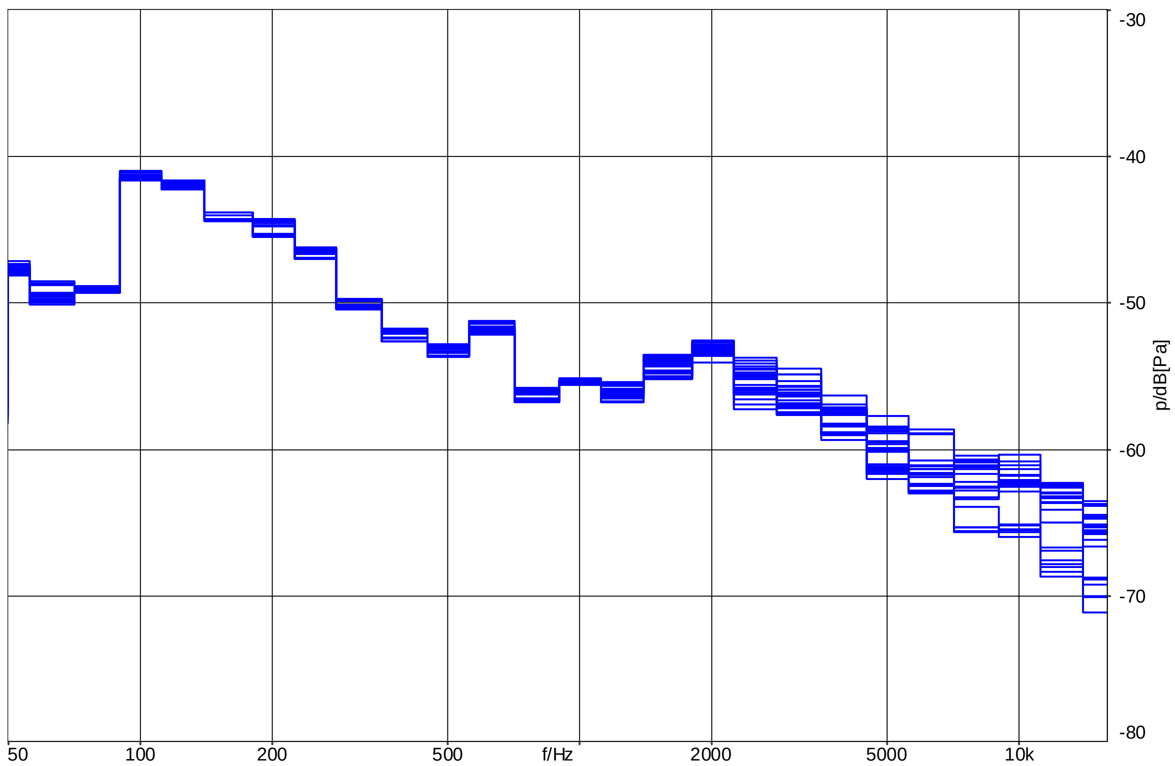


Figure 82: 1/3rd octave spectra at reference microphone and Office noise from TS 103 224 (TS 103 224 methodology)

5 Ambient Noise Testing in Handset Mode

5.1 Introduction

The investigations described in the following clauses refer to ambient noise testing of terminals in handset mode.

5.2 Analysis of modified ambient noise playback systems

5.2.1 Motivation

The noise field simulation system according to ETSI ES 202 396-1 [4] is currently used for handset UE testing in TS 26.132 [3]. Recent investigations (see clause 4.2.3.5.2) indicated drawbacks regarding the reproducibility across different labs, presumably due to manual steps in the equalization process. On the other hand, the system according to ETSI TS 103 224 [5] provides an improved reproduction accuracy by using automated equalization, as e.g. described in clause 4.2.3.5.3. Thus, it seems reasonable to investigate this noise field simulation also for handset testing.

However, one of the issues to resolve is the usage of background noise scenarios for testing. The simulation according TS 103 224 requires suitable eight-channel signals, recorded at the defined positions of reproduction. ES 202 396-1 utilizes two-channel/binaural recordings, which are not available in the aforementioned eight-channel format. For handset UE testing according to TS 26.132, two possible approaches for the usage of ETSI TS 103 224 may be considered:

1) Usage of the noise field simulation and eight-channel background noise scenarios from TS 103 224.

2) Another solution is to retain the binaural scenarios from ES 202 396-1, while using the noise field simulation TS 103 224.

The latter solution may be realized with the recently introduced extension of TS 103 224 (clause 7). This study provides some initial results of this flexible noise field simulation, which utilizes various microphone-/loudspeaker combinations.

To investigate the consistency of the reproduced noise field between measurement rooms, identical measurements were carried out in two measurement rooms with several configurations. Comparisons of the resulting sound fields as well as deviation metrics are presented in the following.

5.2.2 Test Setup

5.2.2.1 Configurations

The recently introduced clause 7 of TS 103 224 provides a flexible and more generic method for noise field simulations. An almost arbitrary number of microphones and loudspeakers can be combined here. For the current evaluation, the two HATS microphones (type 3.3 ears, ID equalization) are used as equalization points. For the playback, the loudspeaker setups 4.1 and 4.0 (four loudspeakers with/without subwoofer) according to ES 202 396-1 are considered. In addition, the default 8.0-setup as described in TS 103 224 is evaluated. Table 6 summarizes the configurations used for this investigation.

Table 6: Noise field simulation configurations

ID	Noise simulation spec.	Loudspeaker setup
A	ES 202 396-1	4.1
B	TS 103 224	4.0
C	TS 103 224	4.1
D	TS 103 224	8.0

Figure 83 illustrates the principle of the different two-point equalizations according to [5] more in detail. In general, all investigated measurement rooms are equipped with eight loudspeakers and a sub-woofer. Depending on the selected configuration, either four (labelled as 1-4 in Figure 83) or eight (labelled as 1-8 in Figure 83) channels are used for equalization and playback. For configuration C, the subwoofer is used as a fifth channel.

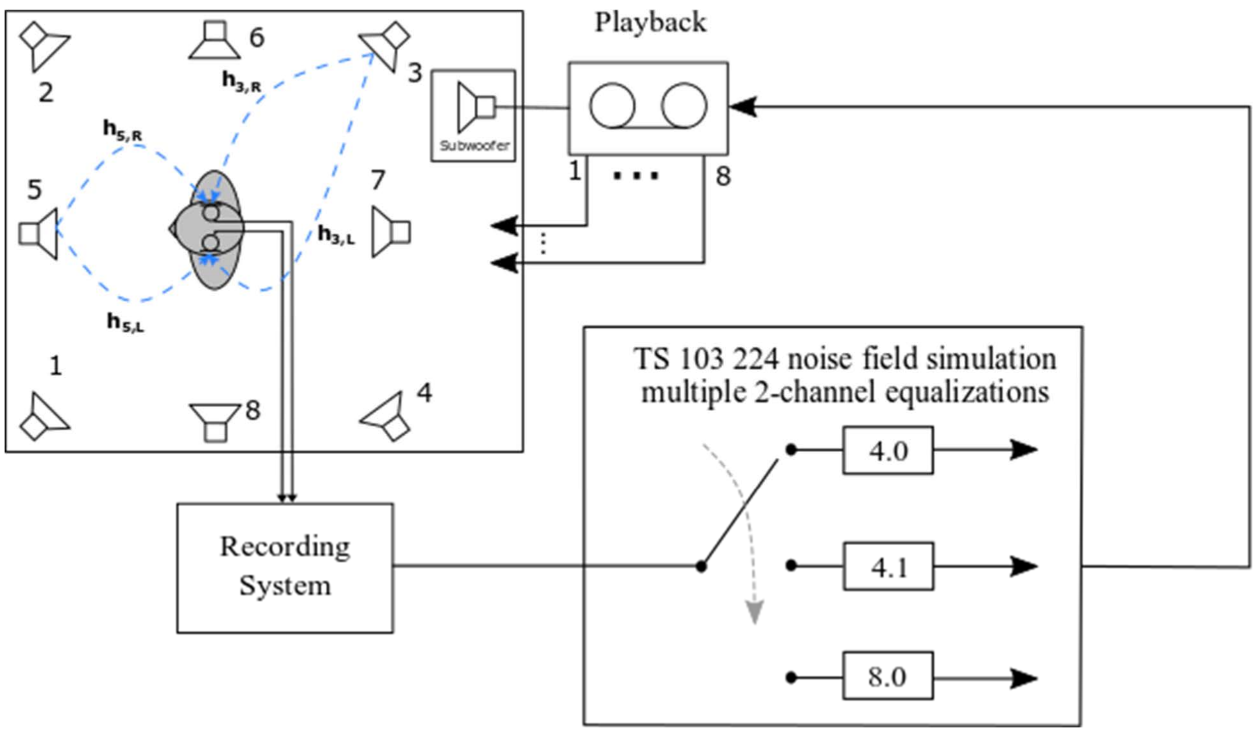


Figure 83: Possible flexible equalization configurations

For configuration A, the noise field equalization according to ETSI ES 202 396-1 (4.1 setup) is applied according to clause 6.3 of [4]. In a first step, each loudspeaker is individually compensated regarding level and frequency response to the corresponding ear in a range of 120 Hz and 20 kHz. Then the two left-handed (1 and 4 in Figure 83) and two right-handed (2 and 3 in Figure 83) pre-compensated loudspeakers are jointly equalized to the left and right ear. The compensation for level and frequency response is then repeated for the subwoofer in a frequency range between 30 Hz and 120 Hz. The four loudspeaker channels are then delayed by manually determined values. In the two rooms, different delay values were used, but are close to the recommendations as described in [4]. Finally, the overall frequency responses (left output to left ear, right output to right ear) are manually adjusted with IIR filters to comply with the specified tolerance of ±3 dB in the range of 50 Hz and 10 kHz.

For the configuration B, C and D, sweep-based impulse responses are measured from each used loudspeaker to the two ear microphones. Figure 83 depicts two examples: the two impulse responses $h_{3,L}$ and $h_{3,R}$ are measured between loudspeaker #3 (which is used for configurations B, C and D) and left/right ears. Another example in Figure 83 illustrates the impulse responses $h_{5,L}$ and $h_{5,R}$ between loudspeaker #5 and left/right ears, which are only used for configuration D.

After pre-processing of the impulse responses (clause 6.2.3 of [5]), the inversion filters for each output channel are calculated via matrix inversion (see clause 6.2.4 of [5]) between 50 Hz and 20 kHz, including the modifications for the flexible setups as described in clause 7.4 of [5]. For configuration C, the inversion filters of the sub-woofer are determined in the range of 40 Hz and 120 Hz.

In a final filter adjustment step, all loudspeakers for the configurations B, C and D are active and a suitable real noise recording is used as the so-called "reference signal", as described in clause 7.3 of [5]. For the current investigation, the first 10.0s of Pub Noise according to ETSI ES 202 396-1 were found to be appropriate for this purpose.

NOTE: Configuration D seems to be heavily overdetermined: In total, 16 impulse responses are collected for an equalization of only two microphones. However, in this and all other configurations, the accuracy and tolerances as specified in clause 7.5 in [5] were met for both rooms.

5.2.2.2 Evaluation of sound field

In order to investigate the reproduction accuracy across measurement rooms, all introduced equalization methods are analysed at microphone positions, which were not part of the equalization procedure. For this purpose, the eight positions of the fixed array for handset-type and headset terminals as shown in Figure 84 are recorded. These represent relevant positions, where the microphones of the test devices are usually located.

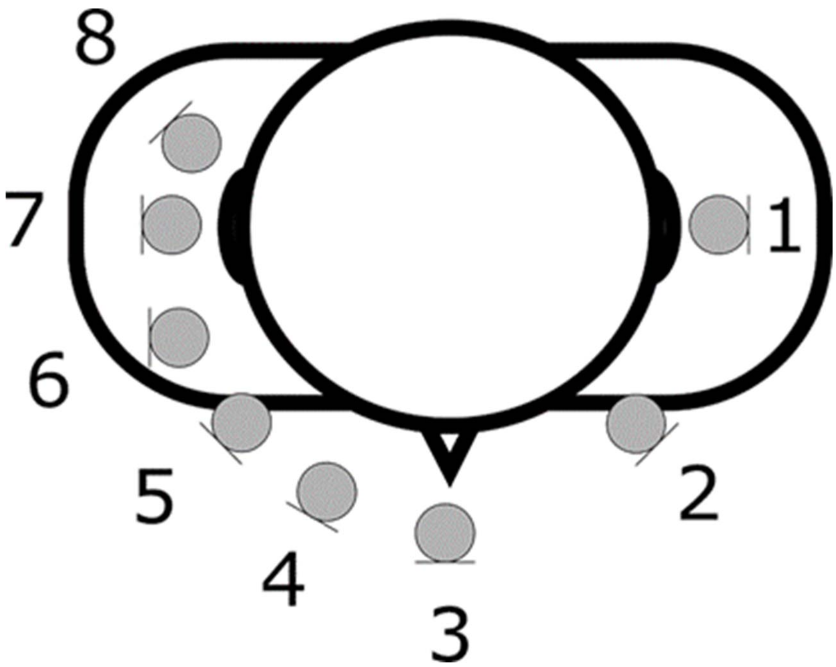


Figure 84: Microphone positions according to [5]

5.2.3 Noise Types

The four noise types as shown in Table 7 are used for the evaluation of the reproduction accuracy.

NOTE: The test procedures as described in clauses 7/8/9/10.12.1 of [3] (speech quality in the presence of ambient noise) are specified with eight noises. The four noise types of Table 7 were selected as an arbitrary, but also representative subset.

Table 7: Investigated noise types acc. to ETSI ES 202 396-1

Filename	Duration	Level
Pub_Noise_binaural_V2	30 s	L: 75.0 dB(A) R: 73.0 dB(A)
Outside_Traffic_Road_binaural	30 s	L: 74.9 dB(A) R: 73.9 dB(A)
Train_Station_binaural	30 s	L: 68.2 dB(A) R: 69.8 dB(A)
Fullsize_Car1_130Kmh_binaural	30 s	L: 69.1 dB(A) R: 68.1 dB(A)

5.2.4 Measurement rooms

The dimensions of the measurement rooms evaluated in this study are described in Table 8.

NOTE: For the present investigation, two rooms (room 1 and 4) were considered. In order to keep the naming / numbering convention consistent across related contributions and studies, the same names as in previous work are used.

Table 8: Measurement rooms

Name	Length [m]	Width [m]	Height [m]	Comment
Room 1	2.40	3.40	2.05	Semi-anechoic
Room 4	1.80	2.40	2.05	Semi-anechoic, rather small chamber

5.2.5 Results

For each measurement room and each background noise, the transfer function in 1/3rd octave bands between the binaural source file and the eight recorded positions was evaluated. Each spectrum of the eight positions is then referenced to the corresponding spectrum of the binaural reference signal. Channel numbers 1 to 3 of the recordings are referenced to the left ear signal, while 4 to 8 are referenced to the right ear signal. Note that this procedure does not change the relation between the different configurations, as the reference signal is identical in all cases.

For the sake of clarity, two simplifications are made for illustration and discussion of the results:

- Only the two microphone positions 5 and 8 according to Figure 84 are considered in the following. These two positions represent typical locations of primary/bottom (pos. 5) and secondary/top (pos. 8) microphones of a device under test.

- Only the background noise Road is presented in this clause. For information, Annex B provides all result curves from all combinations of noise field configurations and background noises.

Figure 85 and Figure 86 show the referenced spectra for each configuration (A, B, C and D according to Table 6) in the two investigated rooms 1 and 4. Each plot contains separate curves for the positions / array channels 5 and 8. Mainly the consistency between rooms is of interest. It is not expected that curves are close to 0 dB for all frequencies, since the curves represent transfer functions between different points in space. Referencing the spectra are mainly intended to perform an energy normalization for better comparison of the different channels.

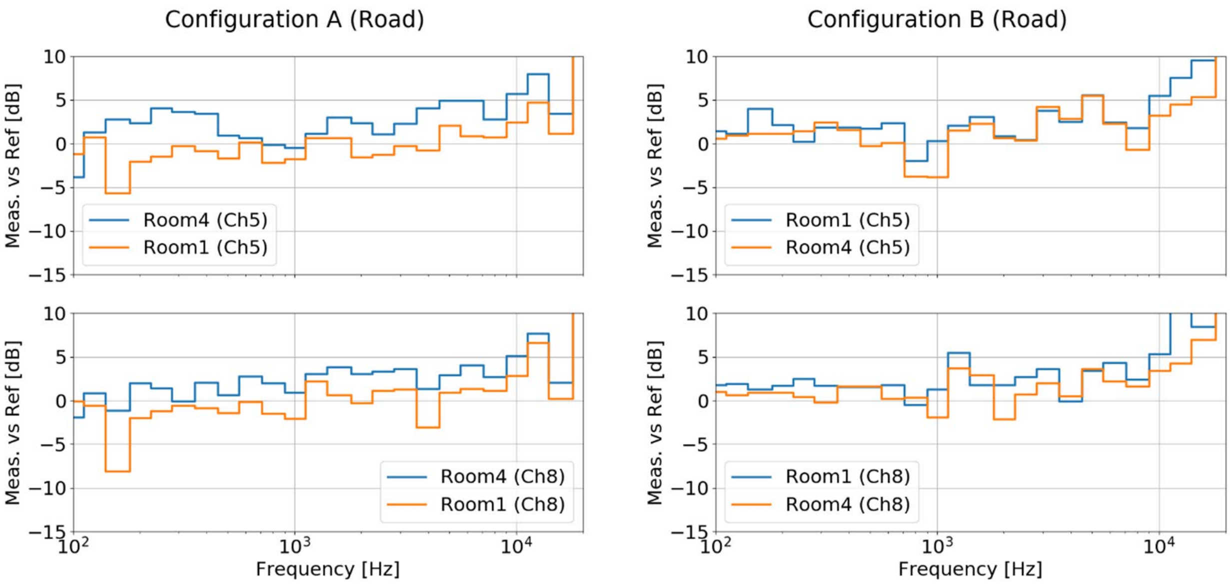


Figure 85: Referenced spectra of microphones 5 and 8 for configuration A and B, *Road* noise

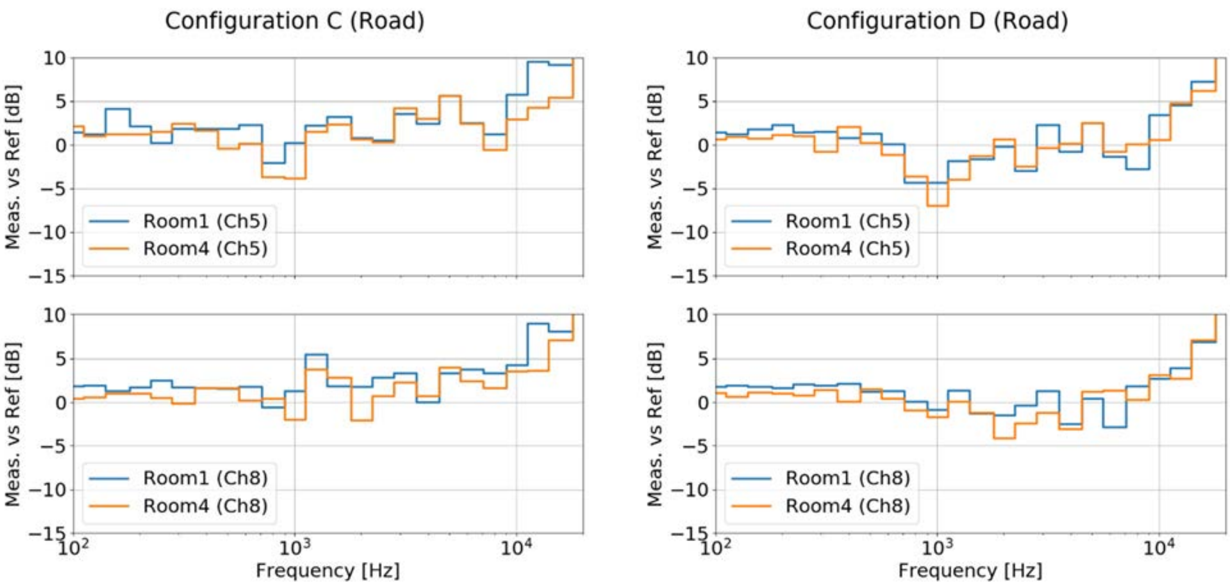


Figure 86: Referenced spectra of microphones 5 and 8 for configuration C and D, *Road* noise

In order to quantify the observed differences between the rooms, the maximum and average deviation across frequencies is calculated for each configuration, background noise and microphone position. Table 9 shows the deviation metrics, which are averaged across all eight microphone positions. Additionally, Table 10 provides the deviation metrics averaged across all noise types.

Table 9: Deviations between two rooms, per setup and per background noise

BGN	Setup	Max. Deviation [dB]	Avg. Deviation [dB]
Car 130kmh	A	11.2	3.2
	B	4.6	1.6
	C	4.4	1.5
	D	3.8	1.2
Pub	A	7.8	2.8
	B	5.5	1.6
	C	5.1	1.7
	D	3.5	1.2
Road	A	7.5	2.9
	B	5.7	1.6
	C	5.0	1.5
	D	3.8	1.2
Train Station	A	8.0	3.3
	B	6.5	1.6
	C	5.4	1.5
	D	4.2	1.2

Table 10: Deviations between two rooms, per setup

Setup	Max. Deviation [dB]	Avg. Deviation [dB]
A	8.6	3.1
B	5.5	1.6
C	5.0	1.5
D	3.8	1.2

5.2.6 Discussion

For configuration A (simulation according to ES 202 396-1, 4.1 loudspeaker setup), a similarity between the sound fields is clearly present, but also larger deviations occur across the entire frequency range. Up to 7.5 dB difference between the two rooms can be overserved, the average deviation is 2.9 dB.

For configuration B (simulation according to TS 103 224, 4.0 loudspeaker setup), the two noise fields at the considered positions are more aligned than for configuration A, but still show some moderate deviations up to 5.7 dB. Compared to configuration A, the average deviation is reduced by almost the half (1.6 dB).

For configuration C (simulation according to TS 103 224, 4.1 loudspeaker setup), almost no differences to configuration B can be identified. The maximum deviation decreases slightly from 5.7 dB to 5.0 dB, while the average is almost the same (1.5 dB). The additionally used sub-woofer seems to have not much impact on the noise field simulation, since it mainly affects the low-frequencies.

For configuration D (simulation according to TS 103 224, 8.0 loudspeaker setup), the smallest deviation in maximum (3.8 dB) and average (1.2 dB) can be observed, i.e. the sound fields are better aligned across rooms than in all other configurations. The increased number of loudspeakers seem to improve the reproduction accuracy and provide the most consistent results across different measurement rooms.

Similar results can be observed for residual noise types, as shown in Table 9. For configuration A, the car noise which includes a substantial amount of low-frequency energy, shows the largest maximum and average deviations. These cannot be identified for the flexible configurations B, C and D. Here the deviation metrics are similar to the other noise types.

The averaged deviation metrics according to Table 10 show a similar trend as the per-noise results. The three flexible setups (configurations B, C and D) indicate an improved reproduction accuracy compared to configuration A. Here both deviation metrics are approximately halved (A: 8.6 dB Max. / 3.1 Avg.) compared to the most advanced configuration D (3.8 dB Max. / 1.2 dB Avg.).

In general, both deviation metrics decrease with an increasing number of loudspeakers. While the maximum deviation could be decreased from 5.5 dB for configuration B to 3.8 dB for configuration D, the improvement in average is rather small (from 1.6 dB for B to 1.2 dB for D).

5.2.7 Conclusions

This study presented analysis results of several variants of the flexible background noise simulation according to clause 7 of TS 103 224, which are compared to the default setup of ES 202 396-1. The four setups introduced allow the backward compatible usage of the binaural sound sources of ES 202 396-1. The analysis was carried out by recording the noise signals with microphones close to typical terminal microphone locations. These locations were not included in the equalization process, which was exclusively performed at the DRP of the HATS. The reproduction accuracy across different measurement rooms / labs at typical terminal microphone locations was investigated.

For the test conditions used in this study, the default setup according to ES 202 396-1 leads to largest deviations between the two measurement rooms. The reproduction methods using the TS 103 224 equalization method provide more consistent sound fields around the HATS across measurement rooms and deviations between the different test rooms can be reduced. The reproduction accuracy across the investigated rooms can be improved further by increasing the number of loudspeakers, as shown by comparing configuration B and D. However, this observation may need further verification by additional measurements / comparisons.

Still, it should be noted that the noise fields at the locations compared in this study are unknown and possibly incorrect since the recordings were made purely at the DRP of the artificial head. In the presented study, only the similarity across different rooms was considered, not the absolute reproduction accuracy. However, based on the obtained results, the flexible equalization setup may represent a promising alternative to the one according to ES 202 396-1, which is currently used in TS 26.132 for testing UEs in handset mode.

5.3 Noise Field Simulations in different Labs

5.3.1 Introduction

The study described in the following clauses investigate at the reproducibility of the background noise configurations among various lab by assessing the difference in the noise spectra between measurement at the HATS ear microphones and the reference noise.

5.3.2 Noise Field Simulations

5.3.2.1 System according to ES 202 396-1 (binaural noise sources)

The noise field simulation system according to ETSI ES 202 396-1 [4] is currently used for handset UE testing in TS 26.132 [3]. It utilizes a 4.1 loudspeaker setup (four satellites and one sub-woofer) for reproduction and is equalized in order to play back binaurally recorded signals. The equalization procedure is conducted in several automated steps, but also manual adjustments for cross-talk cancellation are necessary.

However, recent investigations (see clause 4.2.3.5.2) indicated drawbacks regarding the reproducibility across different labs, presumably due to manual steps in the equalization process.

The setup is denoted as ES202 in the following clauses.

5.3.2.2 System according to TS 103 224 (8-channel sources)

The noise field simulation according to ETSI TS 103 224 [5] provides an improved reproduction accuracy by:

- An eight-channel microphone array (instead of two ear microphones) as shown in Figure 87, developed in particular for testing of handset and handheld hands-free devices,
- Eight loudspeakers (instead of four loudspeakers and a subwoofer),
- An automated procedure for room equalization and cross-talk cancellation (instead of manual adjustment).

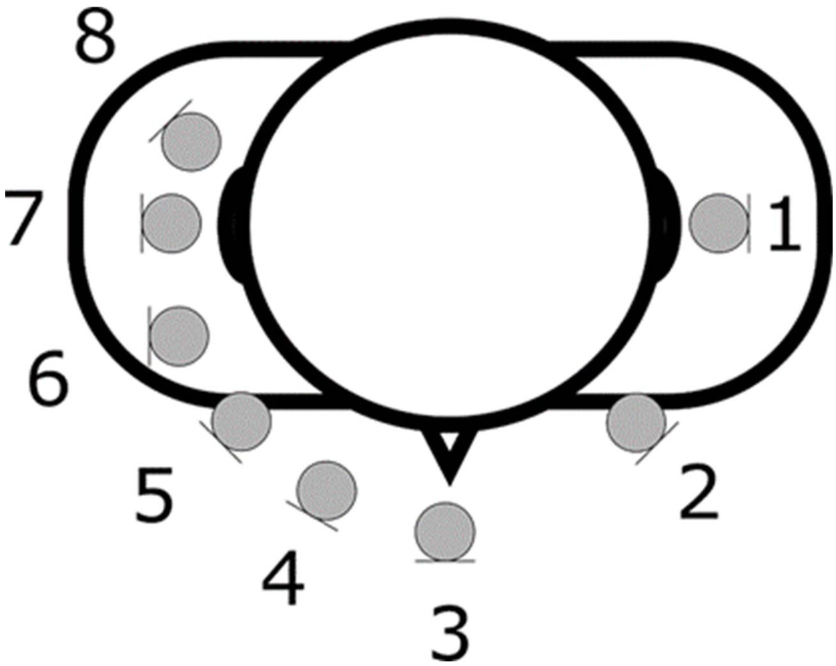


Figure 87: Microphone array for handset testing according to ETSI TS 103 224

Due to the different simulation system, the types of noise sources are different as well – instead of binaural noise recordings, eight-channel signals recorded with the same microphone array have to be used. For this purpose, ETSI TS 103 224 provides a sound file database including similar scenarios as in ETSI ES 202 396-1.

The setup is denoted as TS103-HS in the following clauses.

5.3.2.3 System according to TS 103 224 (binaural noise sources)

Beside the default handset and hands-free setups, the recently updated version of ETSI TS 103 224 [5] provides an extension for so-called "flexible setups". Here an arbitrary number of N microphones and M loudspeakers may be used for the equalization procedure, i.e. the automated equalization procedure is generalized to other configurations (with the restriction of $N \geq 2$ and $M \geq N$). As microphone inputs, the two ears of the HATS were used (N=2). This allows the usage of the same noise types as for ES202.

Figure 83 illustrates the principle of the different two-point equalizations more in detail. In general, all investigated measurement rooms are equipped with eight loudspeakers and a sub-woofer in order to realize the setups required for ES202 and TS103-HS. Depending on the selected configuration, either four (labelled as 1-4 in Figure 83) or eight (labelled as 1-8 in Figure 83) channels are used for equalization and playback. The subwoofer may also be used as a fifth output channel.

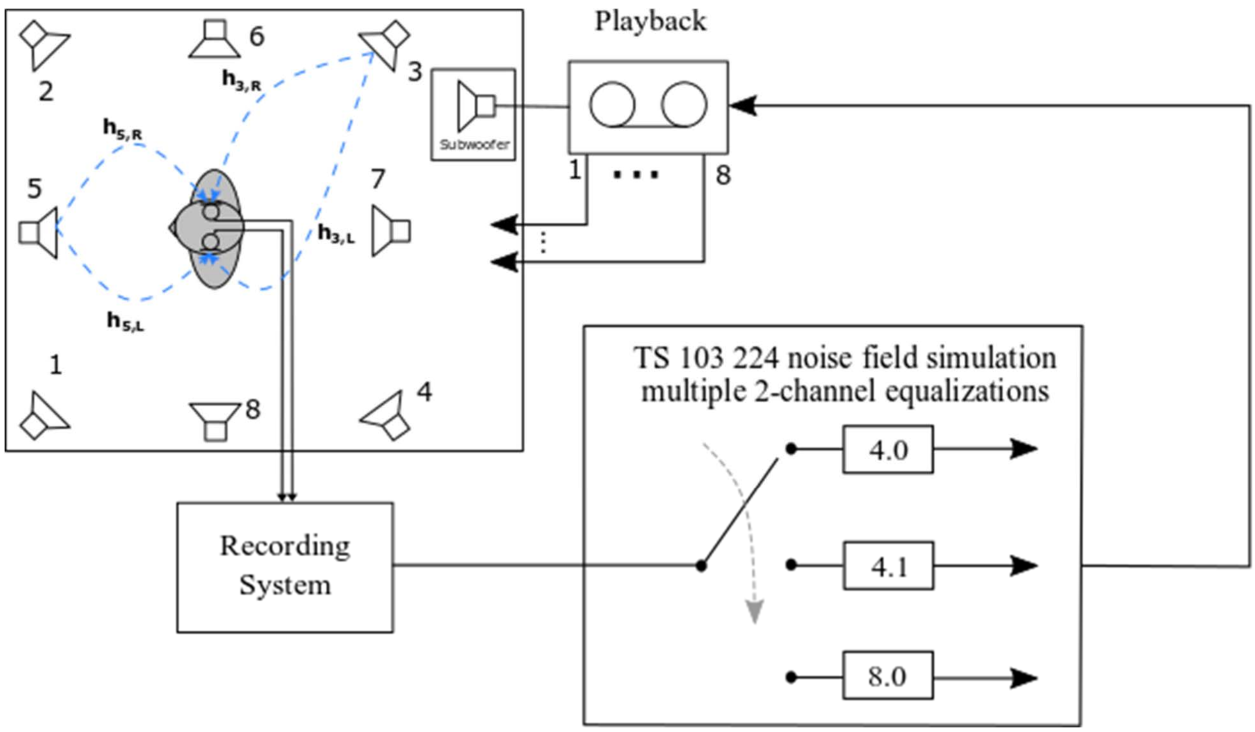


Figure 88: Possible flexible equalization configurations

For the present study, the following setups with M ≥ 4 loudspeakers were investigated:

- 4.1 (four loudspeaker, one sub-woofer – as specified in ETSI ES 202 396-1)
- 4.0 (without sub-woofer)
- 8.0 (same loudspeaker setup as specified in default setup of ETSI 103 224)

The flexible methodology requires the usage of a so-called "reference sound", which should be a typical example of the microphone configuration. It seems trivial to simply use an arbitrary binaural recording for this purpose; however, several pitfalls were discovered during the equalization across several measurement rooms:

- Since the equalization procedure optimizes transfer functions and filter coefficients based on this sound, it is desirable to select a binaural noise which is not included in the later evaluation,
- The reference sound should provide sufficient "binaural content" in order to ensure adequate cross-talk cancellation between left and right ear,
- The reference sound should provide sufficient energy across all frequency bands in order to accurately determine a transfer function between source and recorded channel.

Based on these observations, the binaural noise *TrainStation_bin* was selected, which is provided in TS 103 224 as the binaural version of the eight-channel recording (see also Table 13).

The setups are denoted as TS103-4.0, TS103-4.1 and TS103-8.0 in the following clauses.

5.3.3 Noise Types

According to the descriptions of clauses 7.12.1 (NB), 8.12.1 (WB) and 9.12.1 (SWB/FB) of [2], the binaural noise types as shown in Table 7 are used for the evaluation of noise field simulations ES202, TS103-4.0, TS103-4.1 and TS103-8.0. For the eight-channel system TS103-HS, these identical noise scenarios are not available in the background noise database of ETSI TS 103 224.

However, the noise database here provides similar / related scenarios, which may be used instead. Table 12 shows the eight noise types which are proposed and used for the present study. The database also provides binaural versions of the eight-channel recordings as shown in Table 13.

NOTE 1: Since the noise recordings are in 8-channel-format, only level information for channel 1 (close to left ear) and 7 (close to right ear) are provided. More information on the specific recordings can be found in clause 8 of ETSI TS 103 224.

NOTE 2: In the noise database of TS 103 224, there is no equivalent noise condition for "mensa". The noise "sales counter" is used instead.

Table 11: Noise types acc. to ETSI ES 202 396-1 (binaural)

Filename	Duration	Level (L & R)
Pub_Noise_binaural_V2	30 s	L: 75.0 dB(A) R: 73.0 dB(A)
Outside_Traffic_Road_binaural	30 s	L: 74.9 dB(A) R: 73.9 dB(A)
Outside_Traffic_Crossroads_binaural	20 s	L: 69.1 dB(A) R: 69.6 dB(A)
Train_Station_binaural	30 s	L: 68.2 dB(A) R: 69.8 dB(A)
Fullsize_Car1_130Kmh_binaural	30 s	L: 69.1 dB(A) R: 68.1 dB(A)
Cafeteria_Noise_binaural	30 s	L: 68.4 dB(A) R: 67.3 dB(A)
Mensa_binaural	22 s	L: 63.4 dB(A) R: 61.9 dB(A)
Work_Noise_Office_Callcenter_binaural	30 s	L: 56.6 dB(A) R: 57.8 dB(A)

Table 12: Noise types acc. to ETSI TS 103 224 (8-channel versions)

Filename	Duration	Level Channel 1 & 7
Pub_handset	30 s	1: 77.2 dB(A) 7: 76.0 dB(A)
Roadnoise_handset	30 s	1: 72.8 dB(A) 7: 73.0 dB(A)
Crossroadnoise_handset	30 s	1: 70.6 dB(A) 7: 71.2 dB(A)
TrainStation_handset	30 s	1: 78.9 dB(A) 7: 78.8 dB(A)
FullSizeCar_130_handset	30 s	1: 68.5 dB(A) 7: 70.8 dB(A)
SalesCounter_handset	30 s	1: 66.6 dB(A) 7: 66.6 dB(A)
Cafeteria_handset	30 s	1: 70.0 dB(A) 7: 70.6 dB(A)
Callcenter2_handset	30 s	1: 60.2 dB(A) 7: 60.2 dB(A)

Table 13: Noise types acc. to ETSI TS 103 224 (binaural versions of Table 12)

Filename	Duration	Level (L & R)
Pub_bin	30 s	L: 77.7 dB(A) R: 76.2 dB(A)
Roadnoise_bin	30 s	L: 74.0 dB(A) R: 74.1 dB(A)
Crossroadnoise_bin	30 s	L: 70.8 dB(A) R: 71.6 dB(A)
TrainStation_bin	30 s	L: 78.1 dB(A) R: 78.8 dB(A)
FullSizeCar_130_bin	30 s	L: 68.7 dB(A) R: 70.7 dB(A)
SalesCounter_bin	30 s	L: 66.7 dB(A) R: 66.6 dB(A)
Cafeteria_bin	30 s	L: 69.8 dB(A) R: 70.3 dB(A)
Callcenter2_bin	30 s	L: 60.2 dB(A) R: 60.0 dB(A)

5.3.4 Measurement rooms

5.3.4.1 Geometry

The parameters of the measurement rooms evaluated in this study are described in Table 8. All rooms are equipped with sub-woofer setup for the playback according to ES 202 396-1 [4] and with eight satellite loudspeakers in overall for the playback according to TS 103 224 [5].

Table 14: Measurement rooms

Company	Name	Length [m]	Width [m]	Height [m]	Comment
HEAD acoustics GmbH	Room#1	2.40	3.40	2.05	Same chamber type and design as Room 2
	Room#2	2.40	3.40	2.05	Same chamber type and design as Room 1
	Room#3	2.90	3.10	2.05	Different manufacturer than Room 1,2,4
	Room#4	1.80	2.40	2.05	Smallest chamber
Intel	Room#5	5.57	4.67	2.93	Trapezoidal room
	Room#6	3.60	3.00	2.05	

Figure 89 to Figure 91 depict schematic representations of the six rooms, including positions of the eight loudspeakers, subwoofer and location of HATS.

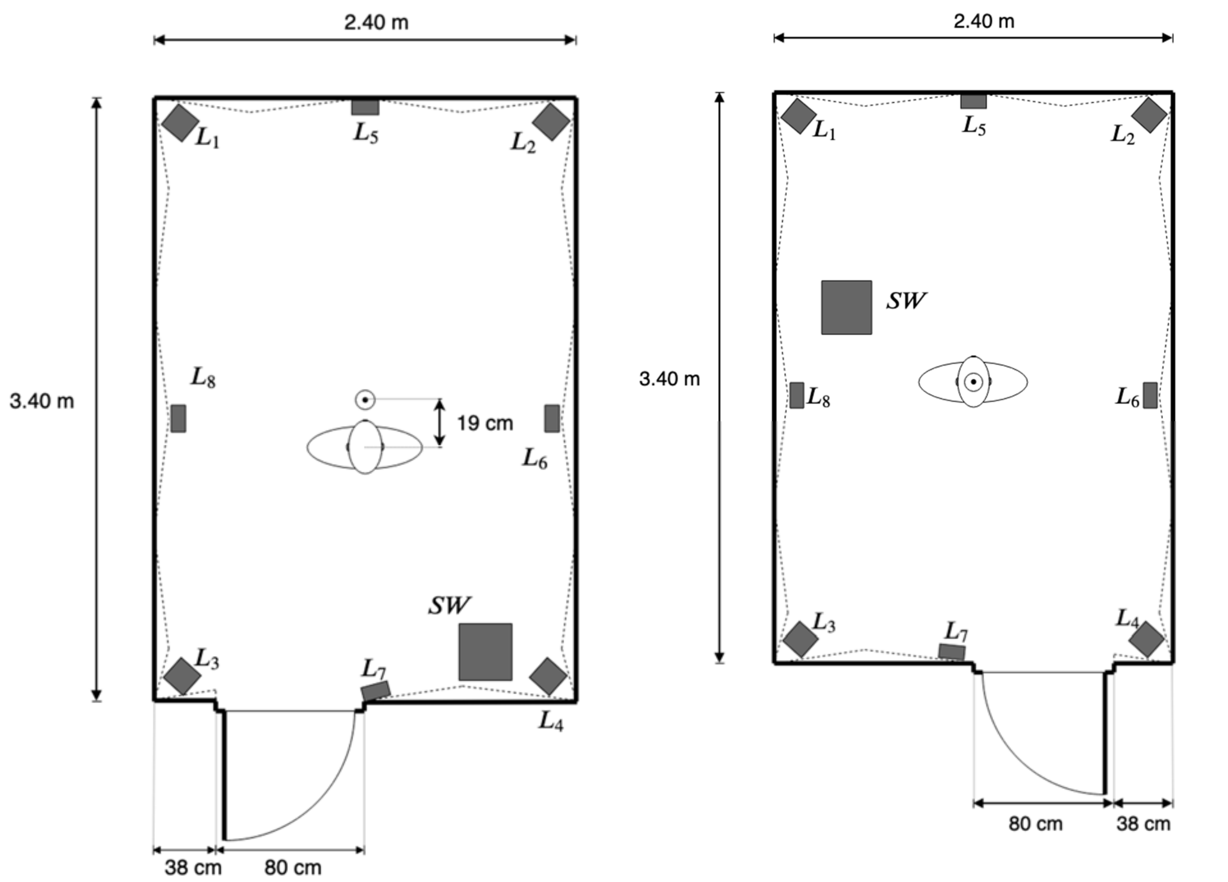


Figure 89: Schematic representation of Room#1 and Room#2

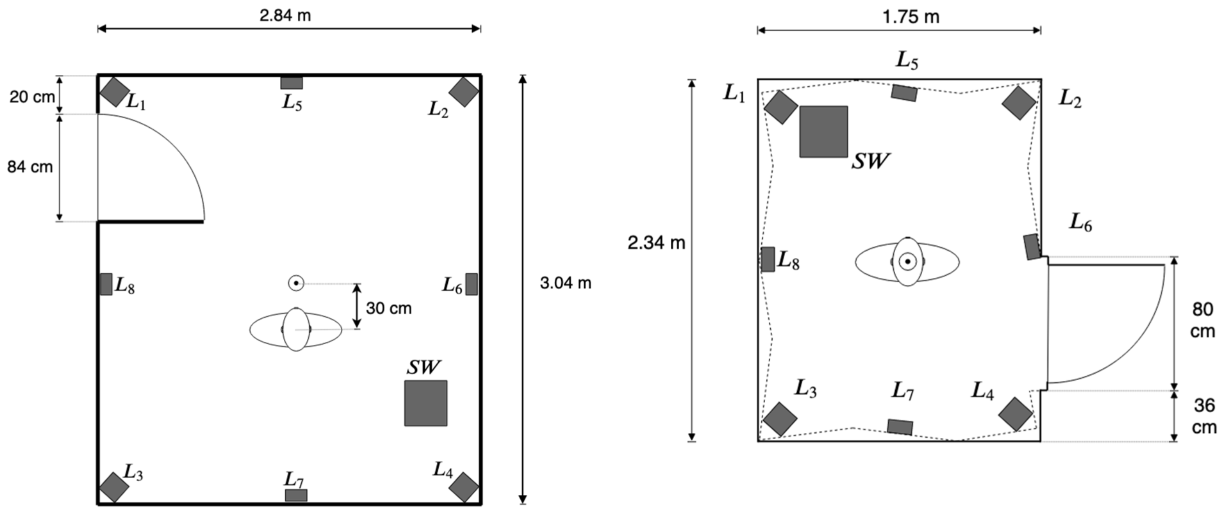
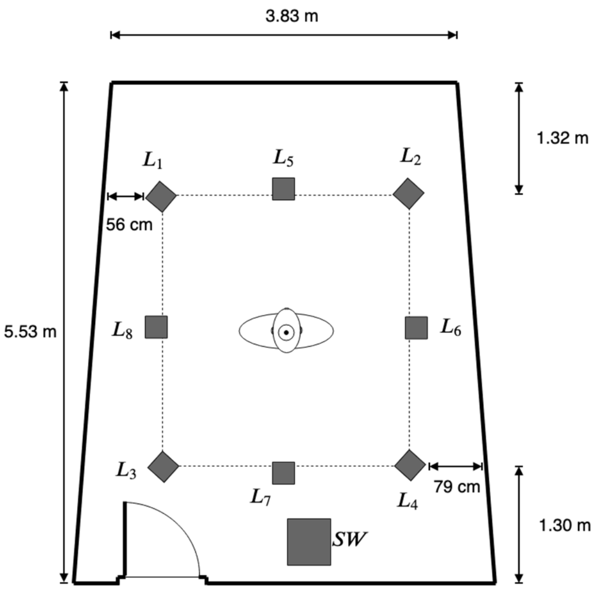


Figure 90: Schematic representation of Room#3 and Room#4



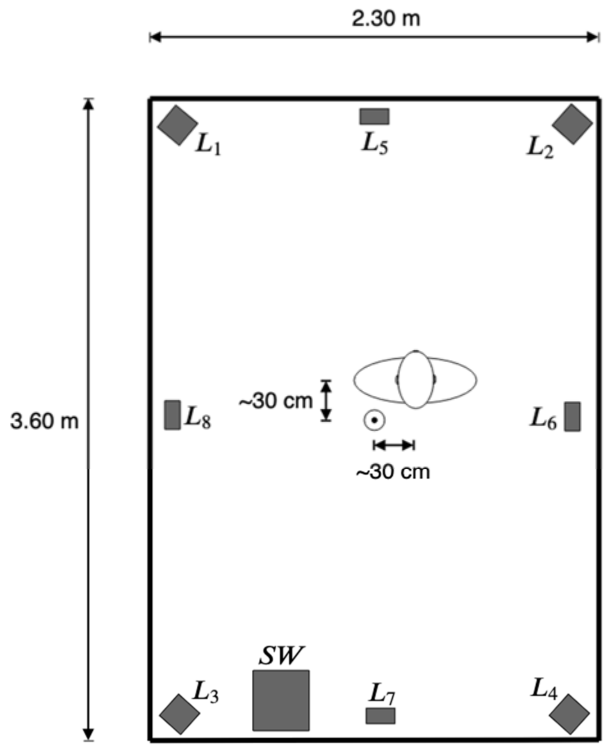


Figure 91: Schematic representation of Room#5 and Room#6

5.3.4.2 Acoustic Room Parameters

In both noise field simulation specifications, certain objectives regarding reverberation time (RT60) and clarity (C80) of the measurement room are provided. These acoustic room parameters are typically derived from impulse response, which can be measured in different ways. Manufacturers of measurement chambers may provide such parameters in the data sheet, but it is often unclear how it was measured. For self-build chambers, these numbers have to be measured anyway.

In order to avoid ambiguous statements about these acoustic parameters, they were measured in the same way across all labs. During the equalization procedure according to TS 103 224, sweep signals from each loudspeaker to each equalization microphone are captured, which were reused for the determination of the acoustic room parameters. By using the eight-channel microphone array with eight loudspeakers, 64 impulse responses are calculated per room. For each of these impulse responses, the parameters RT60, C80 and C50 (for information) are calculated. The average values for all rooms are shown in Table 15.

Table 15: Acoustic parameters of labs (average)

Name	RT60 [ms]	C80 [dB]	C50 [dB]
Room#1	115.5	31.6	24.3
Room#2	117.8	32.1	24.3
Room#3	54.9	35.8	30.3
Room#4	88.2	34.4	30.0
Room#5	107.2	29.2	23.6
Room#6	119.5	26.1	22.2

5.3.4.3 Diffusivity Delays

The equalization procedure according to ES 202 396-1 utilizes four additional delays for each loudspeaker output channel. This step increases the diffusivity of the playback and may reduce cross-talk between the ears microphones. Table 16 provides the delay values used for each room.

Table 16: Diffusivity delays (in ms)

Name	Front/Rear Channel	Left Channel	Right Channel
Room#1	Front	0	11
	Rear	17	29
Room#2	Front	17	29
	Rear	0	11
Room#3	Front	0	11
	Rear	17	29
Room#4	Front	0	11
	Rear	17	29
Room#5	Front	0	11
	Rear	19	30
Room#6	Front	0	1
	Rear	1	2

5.3.4.4 Equalization Results

The equalization procedures of the five noise field simulations as described in clause 2 were successfully applied for all rooms. The binaural as well as the eight-channel transfer functions passed the tolerance of +/- 3 dB across the frequency range 50 Hz to 10 kHz, which is specified in ES 202 396-1 as well as in TS 103 224.

Figure 92 to Figure 96 show the validation results of the equalization procedure for all investigated simulation systems and measurement rooms. Note that Figure 92 to Figure 95 show the left and right ear equalization curves, while Figure 96 shows the eight microphone equalization curves of the spatial array.

It can be seen that binaural equalization using TS 103 224-4.0, TS103-4.1 and TS103-8.0 system yield smoother equalization compared to ES202.

NOTE: Due to time constraints, Room#6 could not perform equalization and measurements with TS103-4.0.

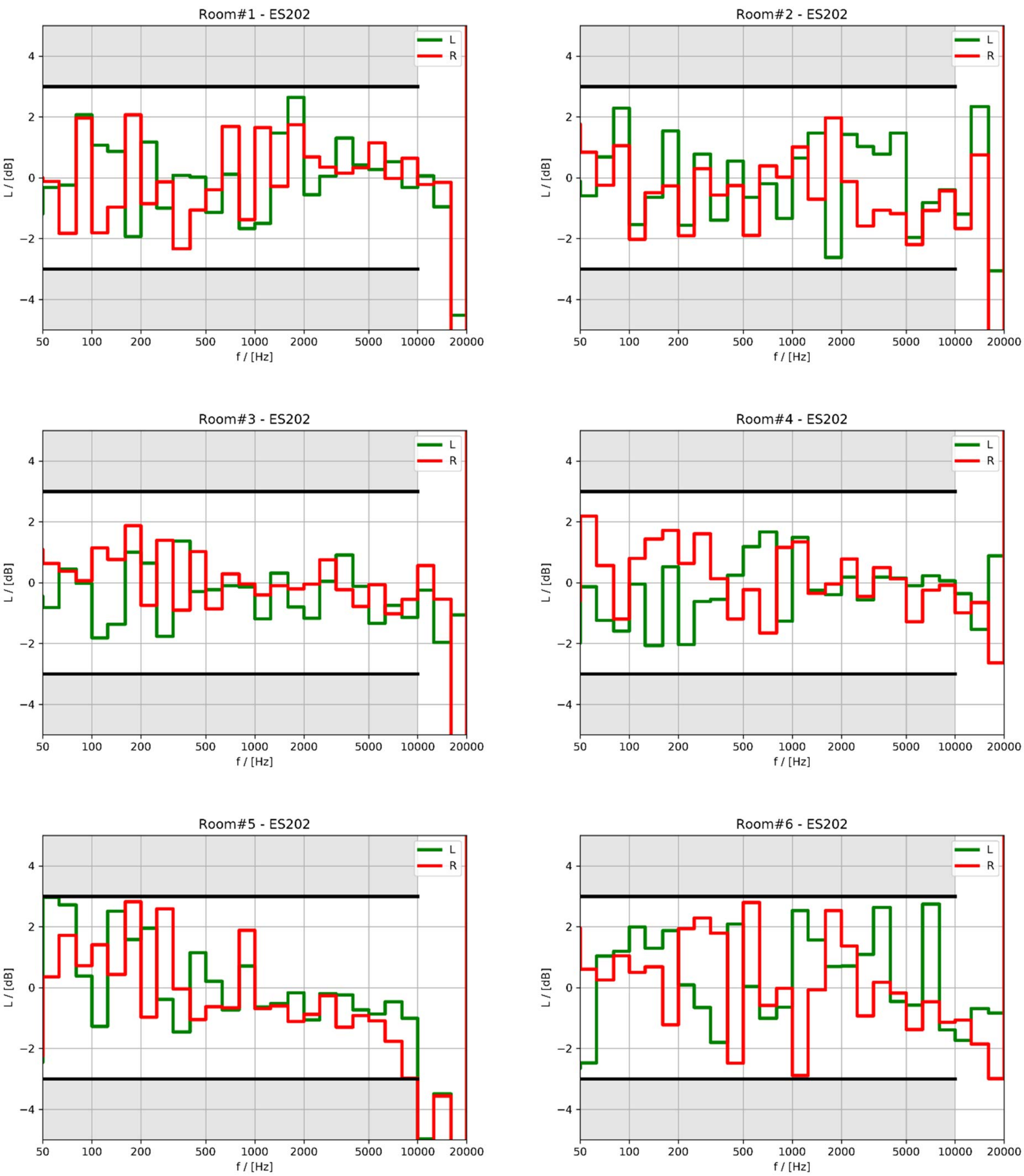


Figure 92: Validation results of equalization (ES202)

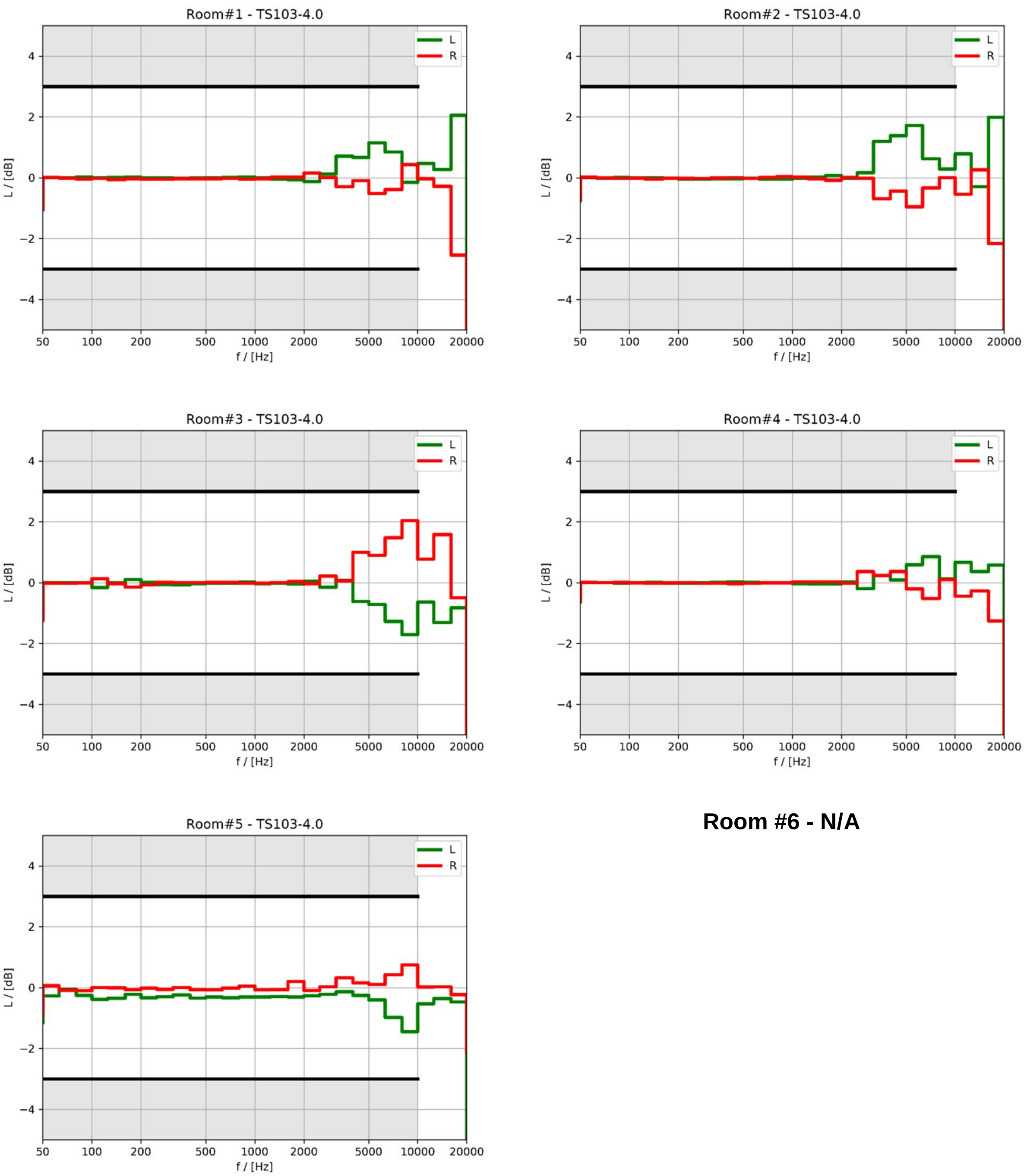


Figure 93: Validation results of equalization (TS103-4.0)

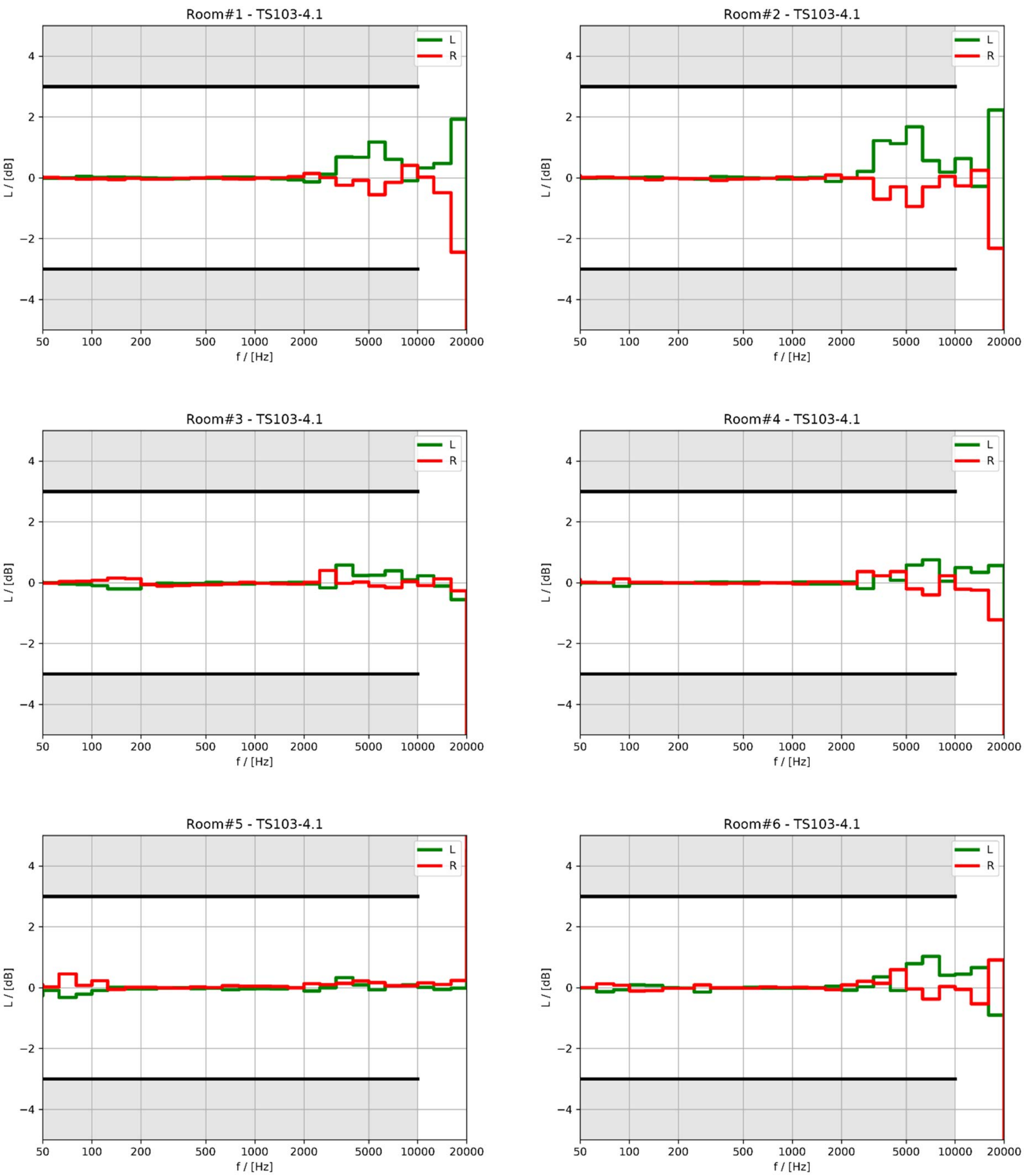


Figure 94: Validation results of equalization (TS103-4.1)

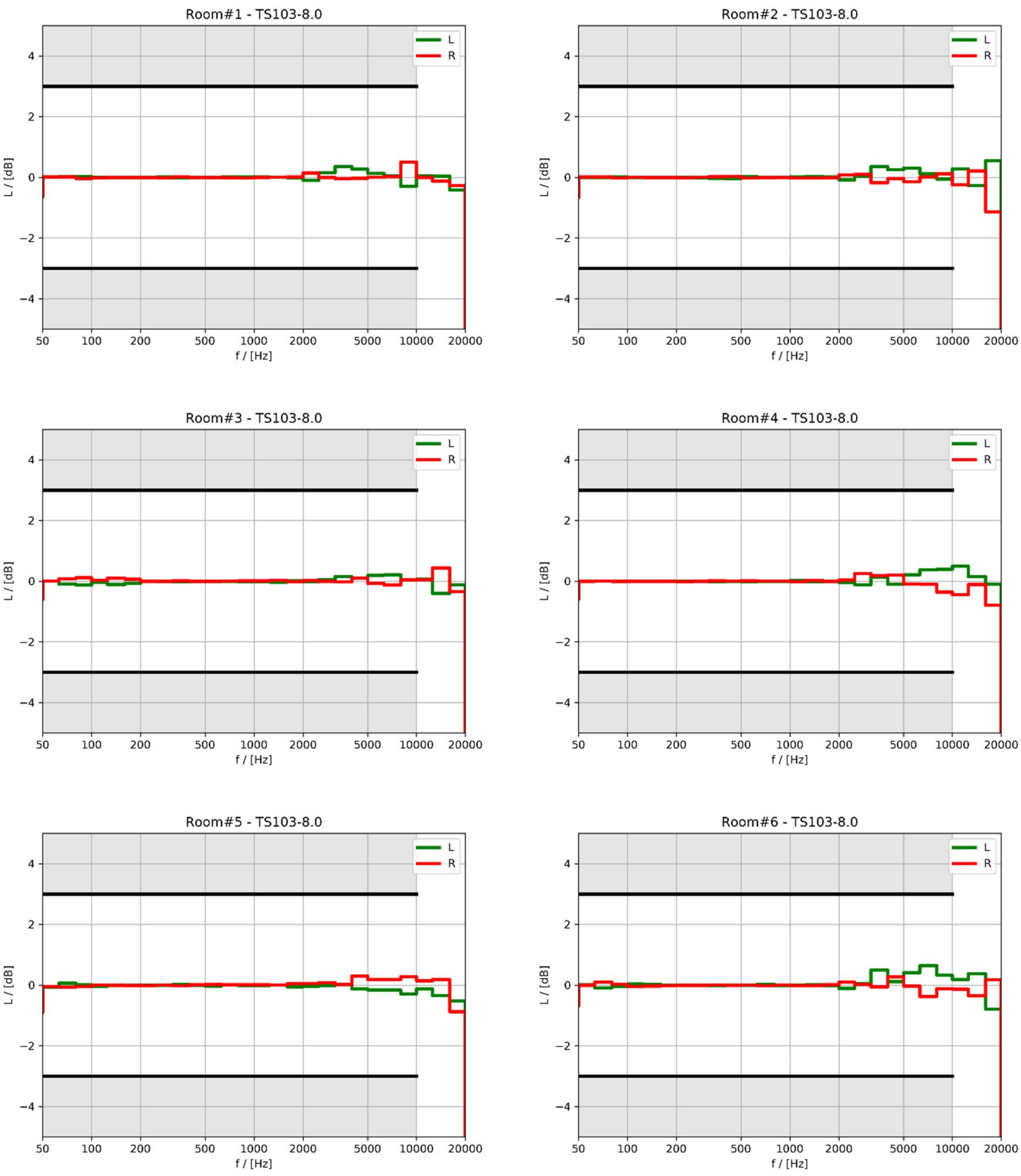


Figure 95: Validation results of equalization (TS103-8.0)

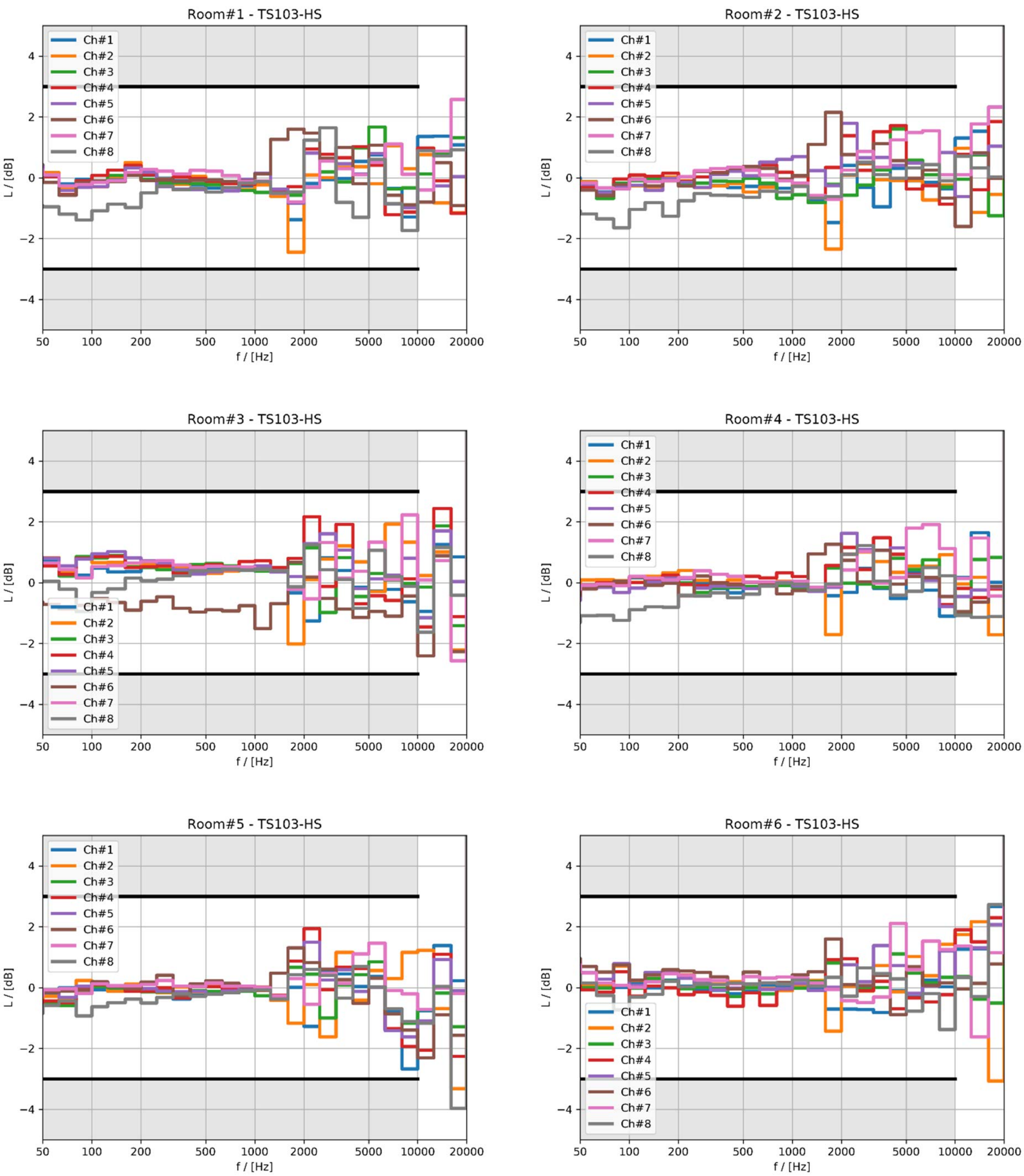


Figure 96: Validation results of equalization (TS103-HS)

5.3.5 Test Methodology

5.3.5.1 Spectral Analysis

In order to investigate and quantify the reproduction accuracy of the noise field simulations, for each system, lab and noise type the background noise is recorded with the HATS ears. For each single measurement, the 1/3rd octave spectrum for left and right ear microphone is calculated from the recordings and is then referenced to the corresponding spectrum of the original noise source. If the noise field reproduction was perfect, the difference between the measured spectrum and the original reference spectrum should be zero.

For TS103-HS, the binaural versions of the eight-channel recordings are used as the reference spectrum. Table 8 summarizes the various configurations tested in the current investigation.

Table 17: Configurations of sources and references

Noise field simulation	Equalization Points	Noise source from	Recording point	Reference noise
ES202	Ear Mics (2)	ES 202 396-1 (binaural, Table 7)	Ear Mics	ES 202 396-1 (binaural, Table 7)
TS103-4.0	Ear Mics (2)	ES 202 396-1 (binaural, Table 7)	Ear Mics	ES 202 396-1 (binaural, Table 7)
TS103-4.1	Ear Mics (2)	ES 202 396-1 (binaural, Table 7)	Ear Mics	ES 202 396-1 (binaural, Table 7)
TS103-8.0	Ear Mics (2)	ES 202 396-1 (binaural, Table 7)	Ear Mics	ES 202 396-1 (binaural, Table 7)
TS103-HS	Mics of array (8)	TS 103 224 (8 Channels, Table 12)	Ear Mics	TS 103 224 (binaural, Table 13)

The following definitions apply for the determined spectra and differences:

$O_X(f_i)$ Original 1/3rd octave spectrum for left (X=L) or right (X=R) ear at frequency index i.

$M_X(f_i)$ Measured 1/3rd octave spectrum for left (X=L) or right (X=R) ear at frequency index i.

$H_X(f_i) = {M_X(f_i)}/{O_X(f_i)}$ 1/3rd octave spectral distance for left (X=L) or right (X=R) ear at frequency index i.

5.3.5.2 Error Metrics

In order to quantify the accuracy of each system several error metrics are calculated on the difference between the measured and reference spectrum. Based on the definitions above, the metrics are specified in equations (1,2) for left (X=L) or right (X=R) ear (all on dB scale):

Standard deviation:

$$Std_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(20 \cdot \log_{10}\left(H_X(f_i)\right) - Avg_X\right)^2} \quad (1)$$

Absolute maximum deviation:

$$\mathrm{AbsMax}_X = \max_{i\in[1,N]}\left|20 \cdot \log_{10}\left(H_X(f_i)\right)\right| \quad (2)$$

With:

$Avg_X = \frac{1}{N}\sum_{i=1}^{N} 20 \cdot \log_{10}\left(H_X(f_i)\right)$ Average of spectral difference in dB

f_1 Minimum frequency used for metric calculation (50 Hz)

f_N Maximum frequency used for metric calculation (10 kHz)

N Number of frequencies included in range

Figure 97 depicts two example results of two measured different test cases. The graph displays two spectral differences for the left and right ear. The graph on the left provides a "good" example (regarding metrics), where the transfer function for the noise "Train Station" is almost flat, and results for left and right ear are consistent, whereas the right one shows a "bad" example (regarding metrics), where the transfer function of the noise "Full-size car 130 km/h" shows quite substantial deviations. In addition, there are noticeable differences between left and right ear. Table 18 provides the corresponding metrics for left and right ears.

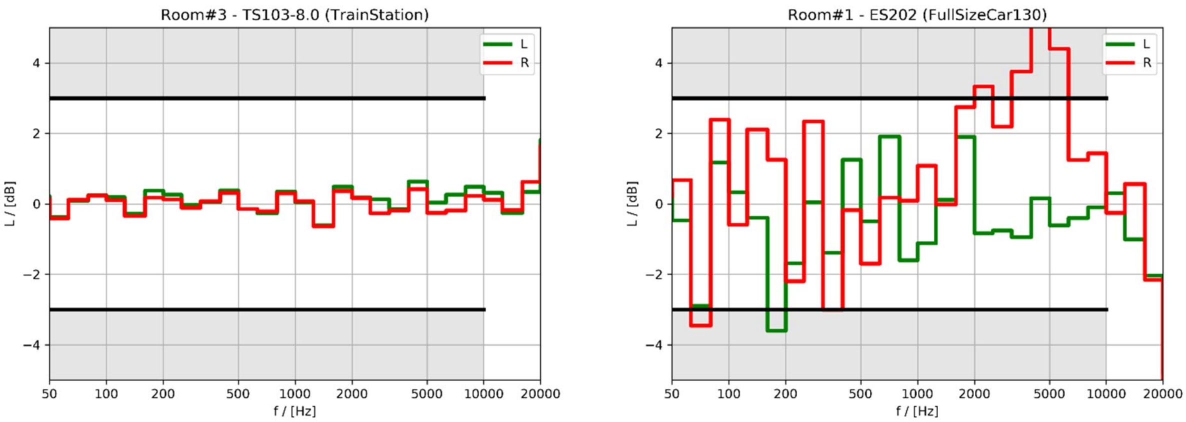


Figure 97: Example for "good" (left) and "bad" (right) binaural transfer function

Table 18: Deviation metrics for two examples

	AbsMax [dB]		Std [dB]	
Example	**L**	**R**	**L**	**R**
"bad"	3.60	6.62	1.28	2.30
"good"	0.63	0.63	0.29	0.27

5.3.6 Measurement Results

5.3.6.1 Overall Results

Table 19 shows the overall results for all the rooms, noise and both ears combined for the five different background noise (BGN) configurations. For both metrics AbsMax and Std, the maximum and average across all included cases. The lower the number the better.

Table 19: Overall results for the five BGN configuration

System	Max (AbsMax) [dB]	Avg (AbsMax) [dB]	Max (Std) [dB]	Avg (Std) [dB]
ES202	12.03	4.38	3.72	1.57
TS103-4.0	6.31	2.25	1.43	0.66
TS103-4.1	6.45	2.04	1.77	0.59
TS103-8.0	4.05	1.70	1.48	0.46
TS103-HS	9.67	3.74	2.21	1.06

NOTE 1: For Room#6, there is no data for TS103-4.0 system available.

NOTE 2: For Room#5, the noise *Train_Station_binaural* is missing in TS103-4.1.

NOTE 3: Due to the different equalization points and different noise types used in TS103-HS, this configuration may not be included in all analyses shown in the following clauses.

The ES202 configuration provides the worst metrics, while TS103-8.0 has the best ones (by a factor of about 3 compared to ES202). TS103-4.0 and TS103-4.1 have quite similar results, but still not as accurate as TS103-8.0.

The results of the TS103-HS seem, at first glance, not as good as the binaural configurations of the TS103. But since the ear microphones were not used as equalization points in TS103-HS, higher deviations than for binaural equalizations are expected. Nevertheless, results are still better than for ES202.

Table 20: Details for AbsMax metric for the five BGN configuration

System	Max (AbsMax) [dB]	Room	Noise	Frequency [Hz]	Ear
ES202	12.03	#3	Pub	50	R
TS103-4.0	6.31	#2	FullSizeCar130	5000	R
TS103-4.1	6.45	#2	FullSizeCar130	5000	R
TS103-8.0	4.05	#4	FullSizeCar130	5000	R
TS103-HS	9.67	#6	FullSizeCar130	10000	R

Table 20 provides additional information about the maximum AbsMax (first column of Table 19). This aggregated metric can be tracked to a certain measurement room, noise type, frequency band and ear channel. For ES202, the pub noise causes the maximum value at 50 Hz and right ear. For TS103-4.0/-4.1/8.0, the right ear is affected as well, but here the car noise shows the largest deviations, commonly at 5 kHz. The same observation can be made for TS103-HS, but at 10 kHz.

An additional representation of the results is depicted in Figure 98. Here the numbers of spectral differences are counted, which pass the equalization requirement of +/- 3 dB according to ETSI ES 202 396-1 [4] and ETSI TS 103 224 [5]. With eight noise types and two channels (left and right ear), 16 pass/fail checks are evaluated for each per noise field simulation.

The noise field simulation ES202 provides some contrary results: room #3 and #4 provide 10-11 passes, but also one room with only one pass (#6) and even two rooms (#2 and #5) with no pass at all.

With the binaural equalizations according to TS 103 224, the number of passes increases clearly. At least 10 out of 16 checks pass for all rooms. For Room#3, the configurations TS103-4.1 and TS103-8.0 pass all possible 16 checks.

Even though the noise field simulation TS103-HS was not equalized at the ear microphones, at least three (rooms #2 and #4) and for some cases between six and nine (rooms #1, #3, #5 and #6) passes can be noticed.

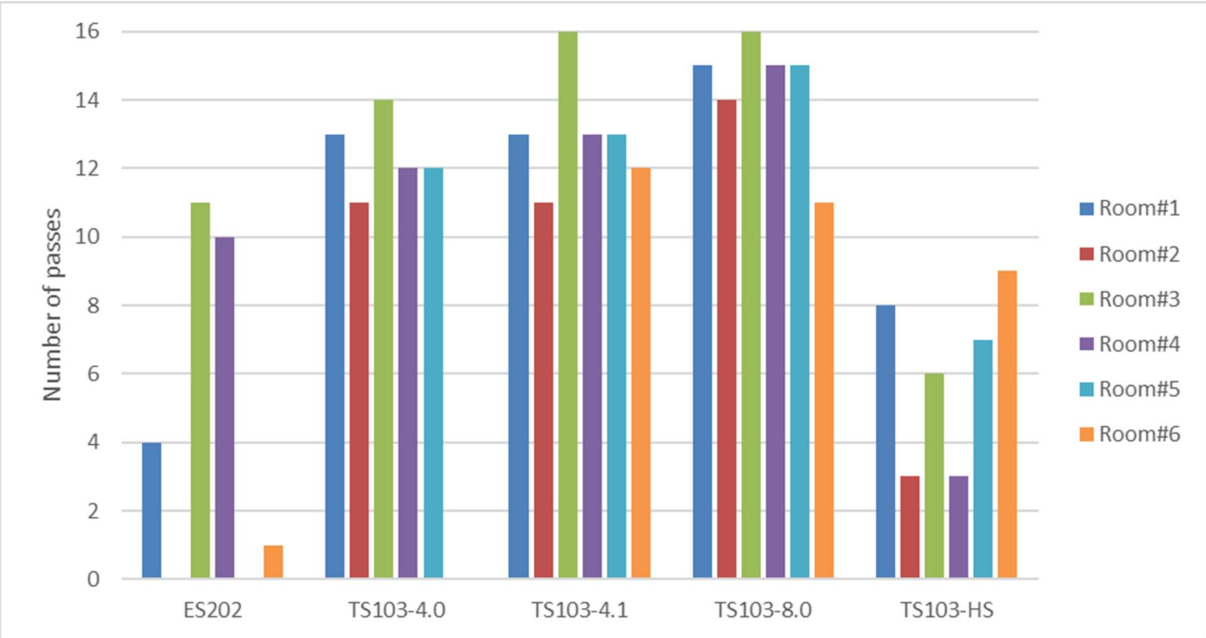


Figure 98: Number of analyses passing the equalization requirement

5.3.6.2 Results per Room

Figure 99 and Figure 100 show maximum AbsMax and average Std, respectively, for all noise types and ears as a function of the BGN system and room.

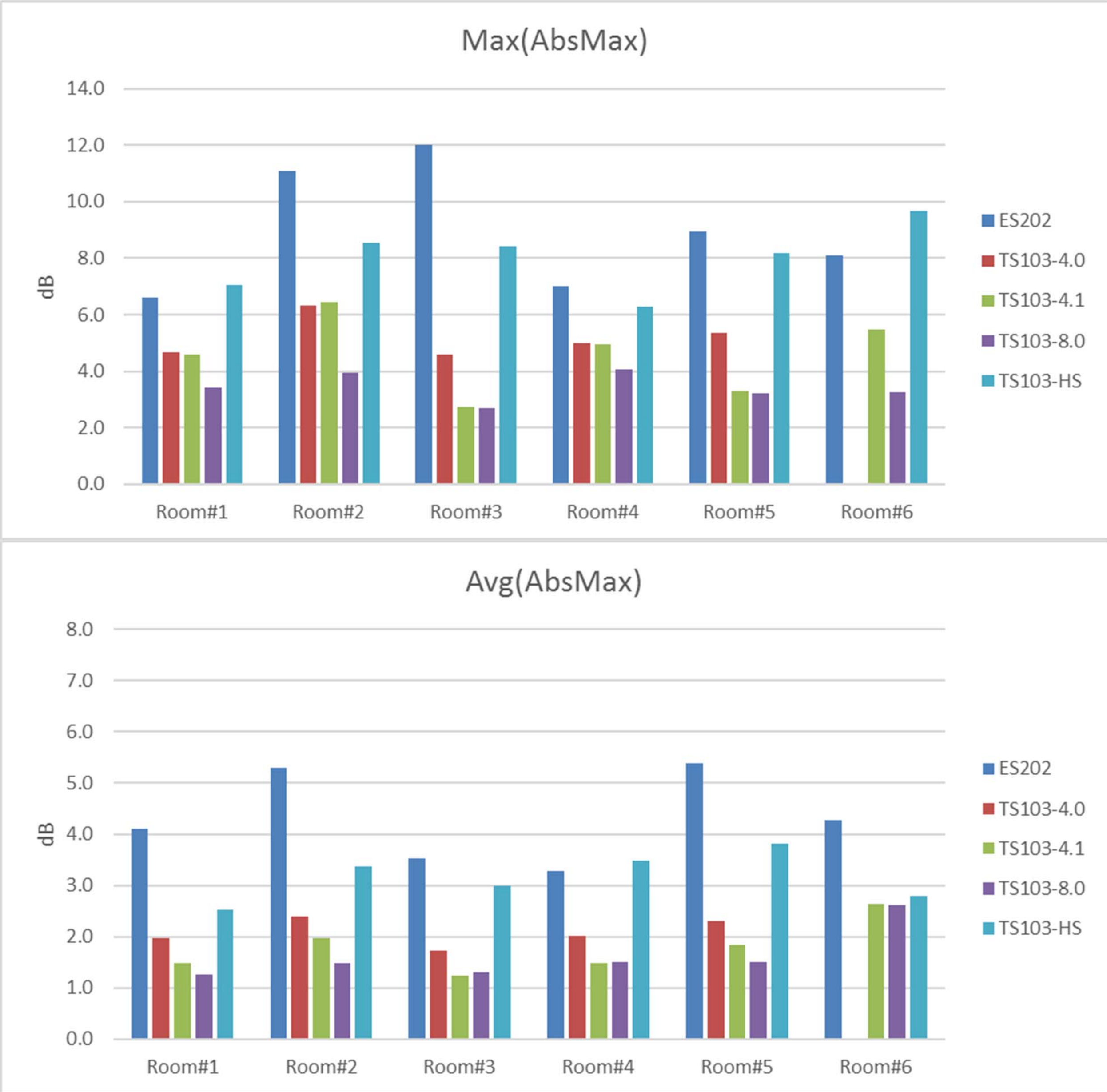


Figure 99: Maximum (top) and average (bottom) AbsMax deviation for all noises and ear depending on the room and BGN system

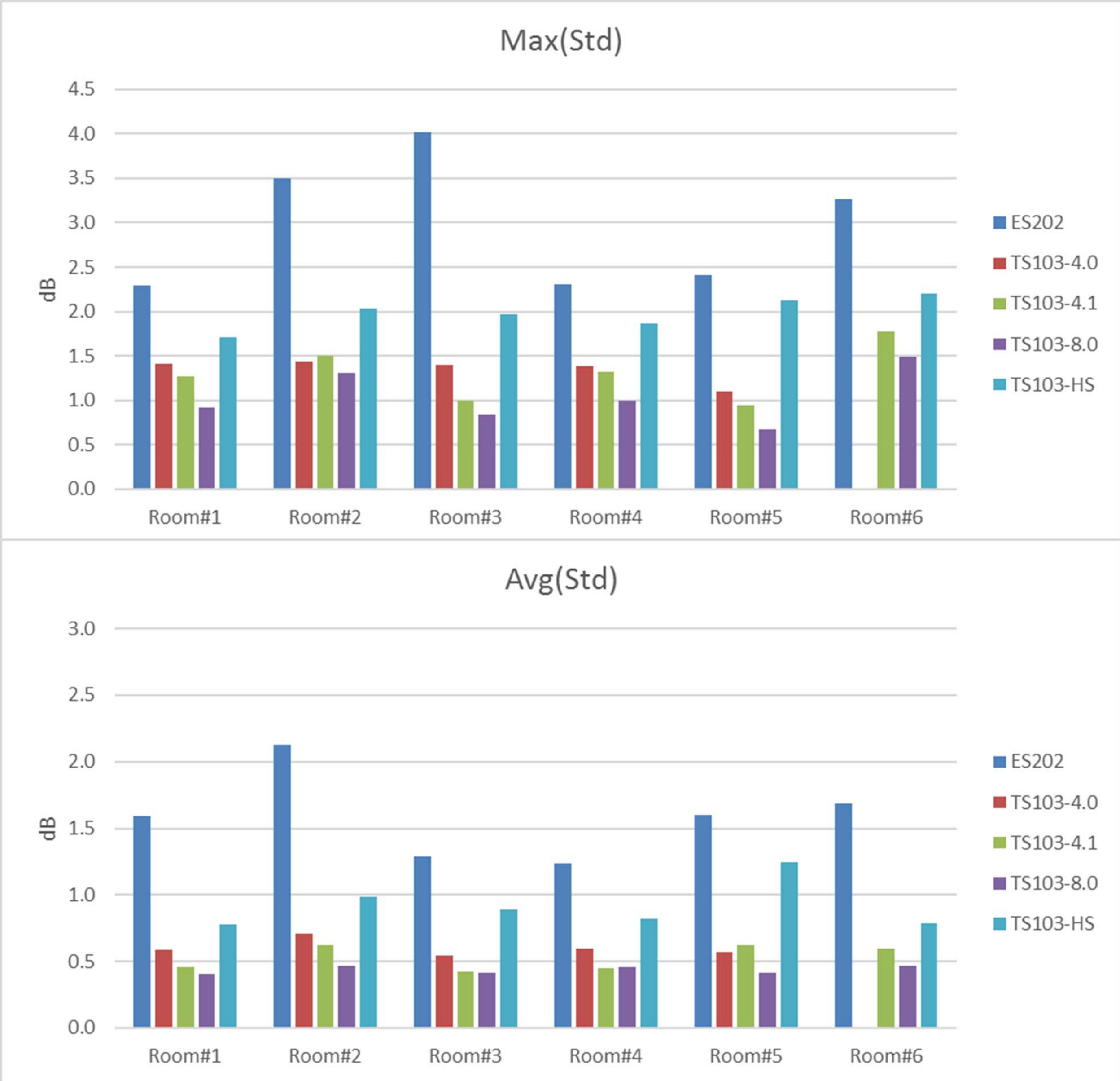


Figure 100: Maximum (top) and average (bottom) Std for all noises and ear depending on the room and BGN system

It can be seen that the behaviour between the rooms are consistent, with the ES202 giving the worst results in terms of AbsMax and Std. Both decrease consistently from TS103-4.0 to TS103-8.0, whereas the latter one provides the best results.

Due to the different equalization points, increased error metrics TS103-HS are expected. However, across most rooms, the metrics are better than for ES202.

5.3.6.3 Results per Background Noise

Figure 101 and Figure 102 show maximum AbsMax and average Std, respectively, for all rooms and ear in function of the BGN systems and noise types.

NOTE: Since TS103-HS is used with different noises than the binaural systems, it is excluded in this analysis.

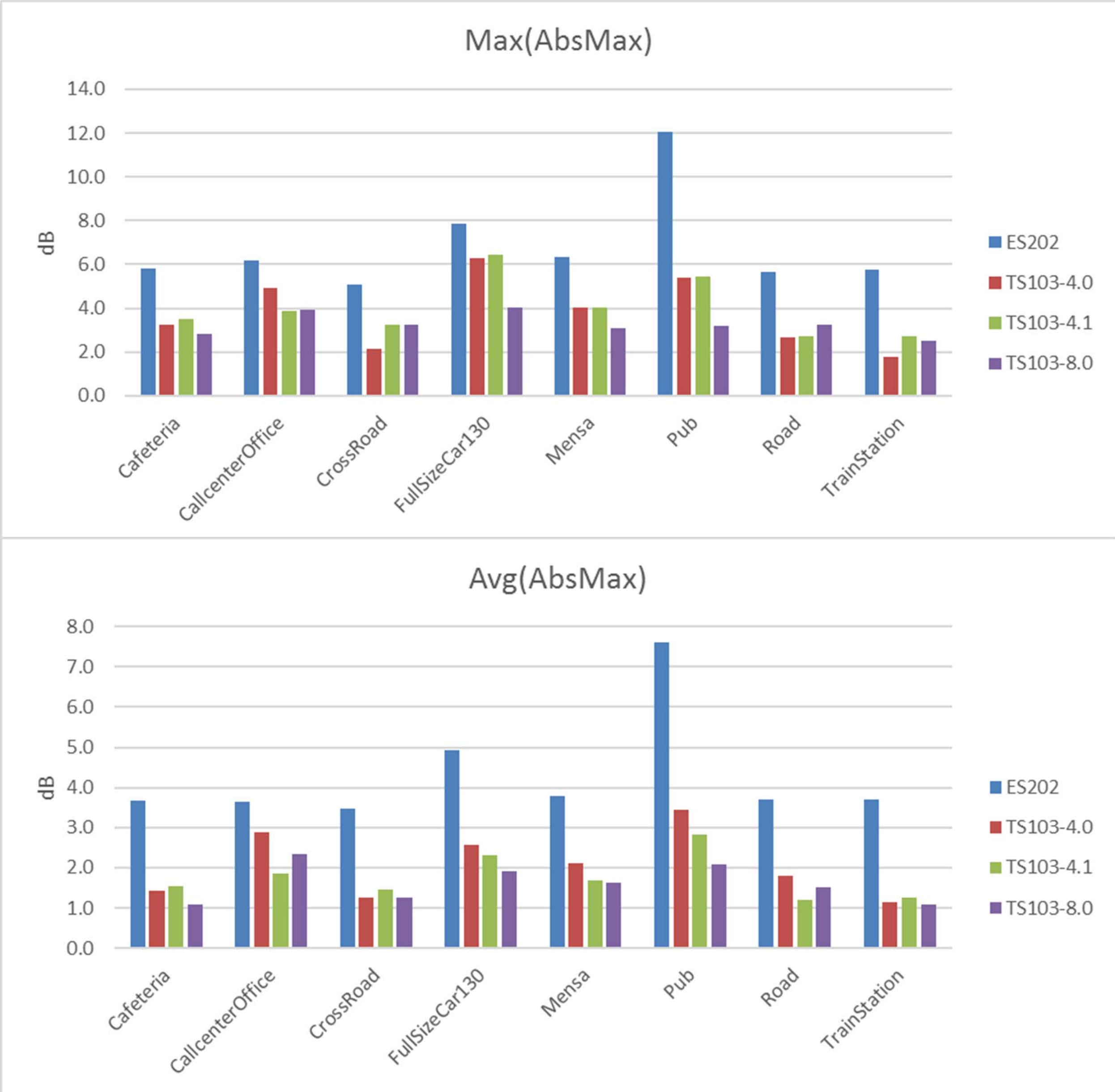


Figure 101: Maximum (top) and average (bottom) AbsMax for all rooms and ear depending on the noise type and BGN system

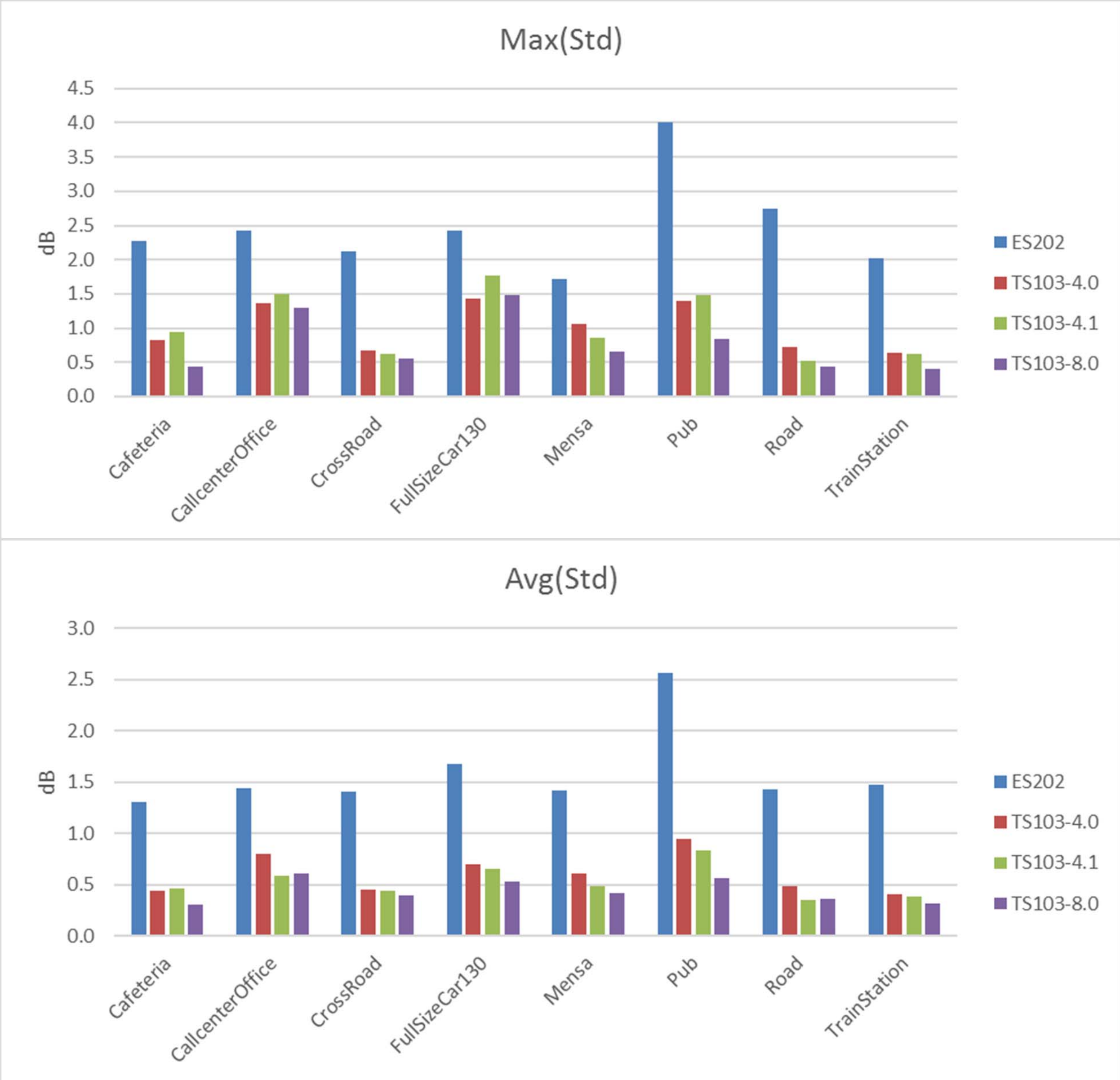


Figure 102: Maximum (top) and average (bottom) Std for all rooms and ear depending on the noise type and BGN system

Similar trends as before can be observed: ES202 provides the worst results in all four metrics across all noise types. In particular, the error metrics here vary a lot across noises; the Pub noise always obtains much worse results here.

On the other hand, the method TS103-8.0 again obtains the best results, TS103-4.0 and TS103-4.1 are slightly worse. For these configurations a variance across noises is visible as well, but less pronounced.

5.3.6.4 Impact of Bandwidth

In this clause the effect of the bandwidth (BW) is examined, which is used to calculate the error metrics. The limit f_N = 10 kHz was chosen according to the validation requirement described in [5]. However, most of the energy in the considered noise signals are located in lower frequencies. In order to investigate the dependency of the error metrics on the frequency, all of them are recalculated with f_N = 4 kHz.

Figure 103 and Figure 104 show AbsMax and average Std, respectively, for all rooms and ears as a function of the BGN system and BW limit.

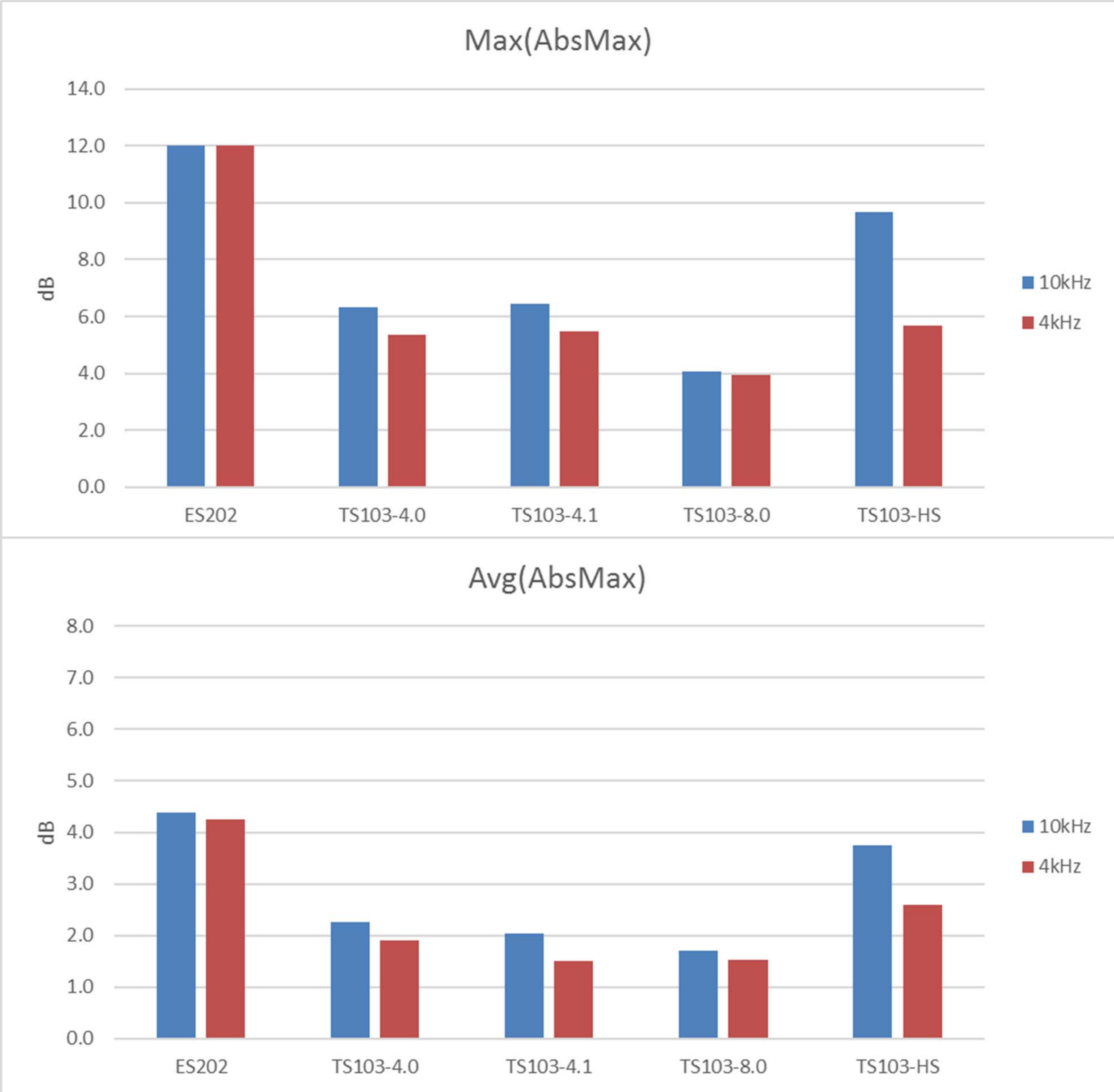


Figure 103: Effect of bandwidth on Maximum (top) and average (bottom) AbsMax for all rooms, noise and ear depending on the BGN system

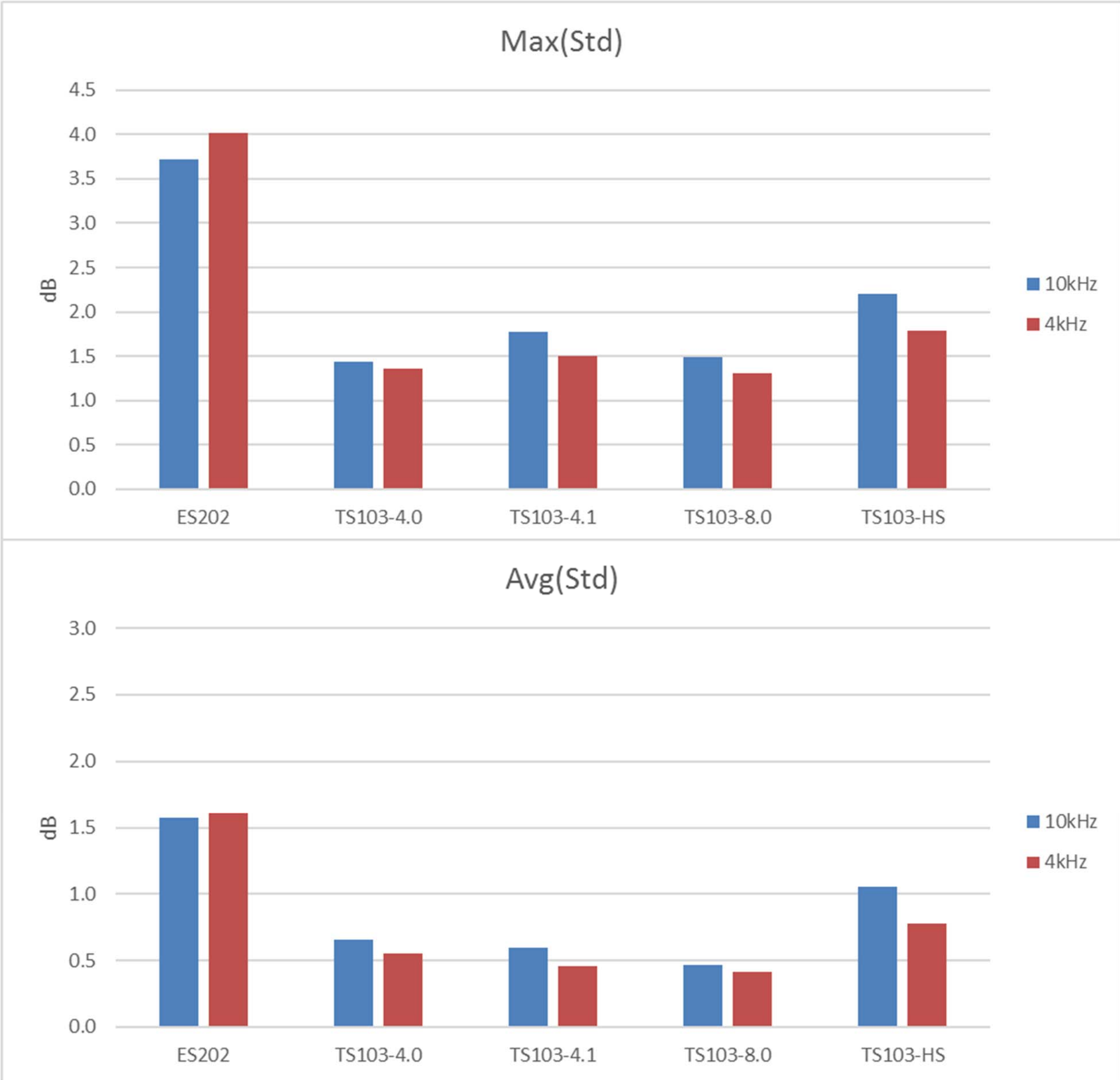


Figure 104: Effect of bandwidth on Maximum (top) and average (bottom) Std for all rooms, noise and ear depending on the BGN system

It can be noticed that the largest differences between the limits 10 kHz and 4 kHz occur for TS103-HS. Results are much closer to the binaural equalizations TS103-4.0/4.1/8.0. This indicates that most spectral differences of TS103-HS occur in the frequency range 4-10 kHz, respectively that the system reproduces the noise field quite adequate in the low frequency domain. This could be explained by the difference in the calibration point (ears vs. microphone array), which will be more noticeable in high frequency.

For the systems TS103-4.0/4.1/8.0, error metrics also decrease from 10 kHz to 4 kHz. Since these configurations are based on the same equalization methodology as TS103-HS, also these results seem plausible.

On the other hand, error metrics for ES202 do not improve between the two different limits. Only for Avg(AbsMax), there is a very slight improvement. For Std metrics, results even get slightly worse. This indicates that most of the spectral differences occur in the low frequency domain between 50 Hz and 4 kHz.

5.3.7 Discussion

The present study investigated six labs/measurement rooms regarding the reproduction accuracy of noise field simulations. Four binaural (ES202, TS103-4.0, TS103-4.1 and TS103-8.0) and one microphone-array-based (TS103-HS) equalization methods were used. Eight noises each were evaluated and analysed based on spectral difference metrics between recording and original sound.

In overall, the configurations TS103-4.0, TS103-4.1 and TS103-8.0 provided consistently better results than ES202 with respect to the error metrics defined in 5.3.5.2. A subsequent analysis across labs and noise types confirmed this trend.

In several cases, ES202 was even outperformed by TS103-HS with respect to the error metrics defined in 5.3.5.2, in particular for the lower frequency domain below 4 kHz - even though this method was not equalized at the ear microphones. Since the asymmetric microphone array used for TS103-HS provides more equalization points on the right side of HATS (used for testing handset devices), error metrics as defined in 5.3.5.2 may also be even better when leaving out the left ear.

However, the results of this investigation may only give a note/hint about the variance of different rooms and noise field simulations on acoustic performance evaluation of handset terminals according to TS 26.131/132. Only a subsequent measurement series with multiple devices may confirm the impact on handset terminals.

5.4 Round Robin Test on Noise Field Simulations with UEs

5.4.1 Introduction

For device testing in handset mode, the reproduction accuracy of the system according to ETSI ES 202 396-1 [1] was evaluated across labs with a round-robin test. Several variants of the noise field simulation systems according to ETSI TS 103 224 [3] were taken into account as well: the eight-channel microphone array as well as binaural equalization are considered here, too.

Up to five noise field simulations were tested per bandwidth, device and room. More than 5100 measurements were collected, each with a duration of at least 80 seconds. This corresponds to more than 113h of recording time.

The following clauses present detailed results of this round robin test, the corresponding test plan is provided in Annex C.

5.4.2 Noise Field Simulations

An overview about the investigated noise field simulations is provided in Table 21. The short names in the first column are used in the following clauses to differentiate between the systems. More detailed descriptions are given in clause 5.3.2. Eight noise types according to Tables 11, 12 and 13 as per clause clause 5.3.2 were used for the different noise field simulation to generate the recordings.

Table 21: Overview of noise field simulations

Noise field simulation (short name)	Equalization Points	Noise source from
ES202	Ear Mics (2)	ES 202 396-1 (binaural)
TS103-4.0	Ear Mics (2)	ES 202 396-1 (binaural)
TS103-4.1	Ear Mics (2)	ES 202 396-1 (binaural)
TS103-8.0	Ear Mics (2)	ES 202 396-1 (binaural)
TS103-HS	Mics of array (8)	TS 103 224 (8 Channels)

5.4.3 Devices

The devices according to Table 22 were available for testing. In addition to the test plan, one more device was tested in some of the labs.

Table 22: Devices for round robin test

Device ID	Manufacturer	Year	VoLTE possible	Test in NB	Test in WB	Test in SWB
DUT 02	A	2018	Yes (only for EVS- SWB)	Yes	Yes	Yes
DUT 03	B	2018	Yes	Yes	Yes	Yes
DUT 08	C	2015	No	Yes	Yes	No
DUT 09	D	2012	No	Yes	Yes	No
DUT 10	E	2014	No	Yes	Yes	No
DUT 05	F	2018	Yes	Yes	Yes	Yes
DUT 04	F	2017	Yes	Yes	Yes	Yes
DUT 07	G	2018	Yes	Yes	Yes	Yes
DUT 06	H	2018	Yes	Yes	Yes	Yes
DUT 39	I	2017	Yes (not for EVS- SWB)	Yes	Yes	No

Four out of these ten devices could not be evaluated in SWB mode due to several reasons (at least in conjunction with test equipment):

- DUT 09 and 10 do not provide LTE functionality (no VoLTE possible).
- DUT 08 offers VoLTE calls in the settings, but none of the codecs worked.
- DUT 39 supports VoLTE calls, but EVS-SWB codec seems not available at all.

5.4.4 Rooms/Labs

The participating labs/rooms were identical to the ones described in clause 5.3.4. A detailed description regarding geometry, acoustic room parameters, and results of the noise field equalization procedures can be found here.

Due to internal restructuring and time constraints, room #6 could only evaluate two devices (DUT 02 & 03) and room #5 could not run the tests for DUT 39.

5.4.5 Additional measurements

Four devices of the round robin test were kindly provided by CTIA CPWG, sub-working group “Audio”. In return, they asked to provide some basic measurements according to TS 26.132 [3] in NB, WB and SWB mode. These additional measurements were addressed described in the test plan (see Annex C) and were collected in advance to the ambient noise testing.

Sending Direction:

- DUT Delay (compensated for test equipment delay)
- Frequency Response
- Loudness Rating
- Predicted speech quality acc. to ITU-T P.863 V2.4

Receiving Direction:

- DUT Delay (compensated for test equipment delay)
- Frequency Response
- Loudness Rating (at nominal level)
- Predicted speech quality acc. to ITU-T P.863 V2.4 (at nominal level)

5.4.6 Results for noise field simulation ES 202 396-1

Table 23 to Table 25 show the results of S-MOS and N-MOS of the noise field simulation according to ES 202 396-1 [4] (labelled as ES202 in the following) for NB, WB and SWB mode for all devices and all measurements rooms. Each number is calculated as the average across the eight noise types as described in Table 1 of the test plan (which equals the noise types of TS 26.132 [3]). In addition, the average across rooms is provided in an additional column.

Table 23: RR-Test results for ES 202 396-1 noise field simulation (NB)

	S-MOS							N-MOS						
Room	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**
DUT														
DUT 02	4.19	4.18	4.08	4.21	4.27	3.47	4.07	3.57	3.73	3.30	3.48	3.71	4.09	3.65
DUT 03	4.09	4.12	3.94	3.95	4.21	4.04	4.06	4.32	4.48	4.31	4.30	4.42	4.54	4.40
DUT 04	3.75	3.88	3.80	3.74	4.04		3.84	4.48	4.45	4.36	4.40	4.52		4.44
DUT 05	3.99	3.89	3.97	3.95	4.16		3.99	4.10	4.11	3.93	3.91	4.08		4.03
DUT 06	3.66	3.75	3.90	3.93	4.06		3.86	4.25	4.39	4.32	4.43	4.50		4.38
DUT 07	3.96	4.00	3.95	4.02	4.12		4.01	4.48	4.53	4.41	4.51	4.52		4.49
DUT 08	4.06	4.04	4.04	4.06	4.22		4.08	3.47	3.69	3.43	3.47	3.83		3.58
DUT 09	4.11	4.19	4.03	4.07	4.20		4.12	3.07	3.51	3.11	3.12	3.59		3.28
DUT 10	3.97	4.01	3.88	3.94	3.99		3.96	3.71	3.95	3.69	3.79	3.99		3.83
DUT 39	4.08	4.22	4.16	4.15			4.15	4.08	4.17	3.98	4.03			4.07

Table 24: RR-Test results for ES 202 396-1 noise field simulation (WB)

	S-MOS							N-MOS						
Room	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**
DUT														
DUT 02	3.95	4.01	3.90	3.98	4.06	4.08	4.00	3.94	4.06	3.70	3.78	4.05	4.37	3.98
DUT 03	3.87	3.96	3.90	3.91	4.04	3.96	3.94	4.44	4.41	4.34	4.32	4.43	4.41	4.39
DUT 04	3.84	3.97	3.87	3.88	4.02		3.92	4.18	4.26	4.04	4.04	4.33		4.17
DUT 05	3.79	3.95	3.83	3.91	4.01		3.90	4.14	4.29	3.89	4.06	4.13		4.10
DUT 06	3.94	3.96	3.80	3.84	3.94		3.90	4.32	4.49	4.35	4.39	4.51		4.41
DUT 07	4.00	4.02	3.95	4.04	4.09		4.02	4.16	4.24	4.06	4.11	4.26		4.17
DUT 08	3.77	3.91	3.77	3.84	4.04		3.87	3.81	4.00	3.76	3.70	4.10		3.88
DUT 09	3.95	4.06	3.89	3.94	4.08		3.98	3.26	3.57	3.31	3.36	3.63		3.43
DUT 10	3.85	3.94	3.78	4.00			3.89	4.24	4.41	4.21	3.75			4.15
DUT 39	3.85	3.97	3.93	3.94			3.92	4.26	4.32	4.18	4.24			4.25

Table 25: RR-Test results for ES 202 396-1 noise field simulation (SWB)

	S-MOS							N-MOS						
Room	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**	**R#1**	**R#2**	**R#3**	**R#4**	**R#5**	**R#6**	**Avg.**
DUT														
DUT 02	3.72	3.84	3.72	3.69	3.91	3.81	3.78	4.18	4.32	4.05	4.27	4.37	4.28	4.24
DUT 03	3.76	3.85	3.76	3.74	4.03	3.88	3.84	4.23	4.28	4.11	4.17	4.34	4.33	4.24
DUT 04	3.92	4.01	3.82	3.76	3.99		3.90	3.99	4.05	3.69	3.87	4.10		3.94
DUT 05	3.95	3.95	3.89	3.87	4.06		3.94	3.74	3.91	3.49	3.78	3.87		3.76
DUT 06	3.52		3.53	3.48	3.69		3.56	4.48		4.47	4.55	4.57		4.52
DUT 07	3.86	3.97	3.75	3.66	4.05		3.86	4.08	4.16	3.94	4.03	4.21		4.08

5.4.7 Results for noise field simulation TS 103 224 (binaural)

As described in the test plan, also the flexible equalization methods according to ETSI TS 103 224 [5] with different loudspeaker setups were investigated in the round robin test (labelled as TS103-4.0, TS103-4.1 and TS103-8.0 in the following). Since the noise sources used for these equalizations are identical to ES202, similar results are expected.

5.4.7.1 Narrowband Mode

5.4.7.1.1 Results for all noise types, rooms and devices

Figure 105 to Figure 107 provide the results of the three binaural equalization methods compared to the data of ES202 (cf. Table 23) for NB mode. Each combination of device, room and background noise represents one dot in these scatter plots.

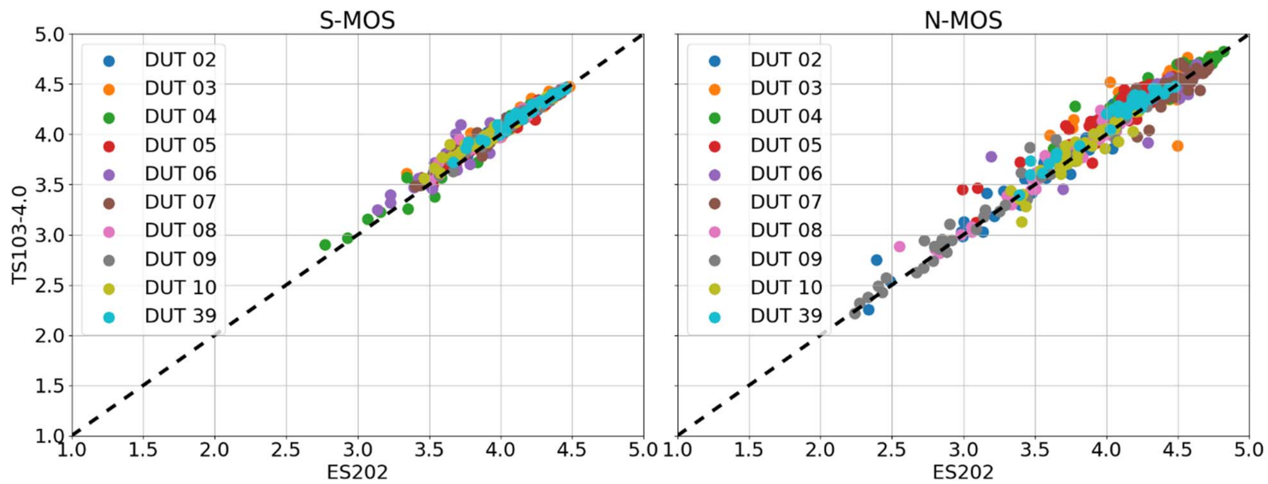


Figure 105: Results S-/N-MOS for TS103-4.0 vs. ES202 (NB)

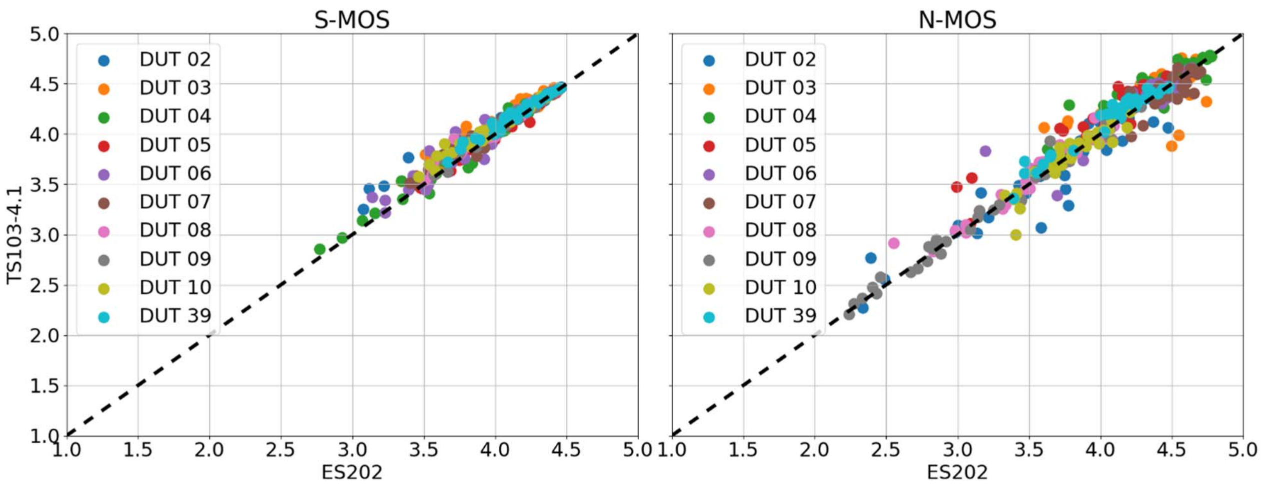


Figure 106: Results S-/N-MOS for TS103-4.1 vs. ES202 (NB)

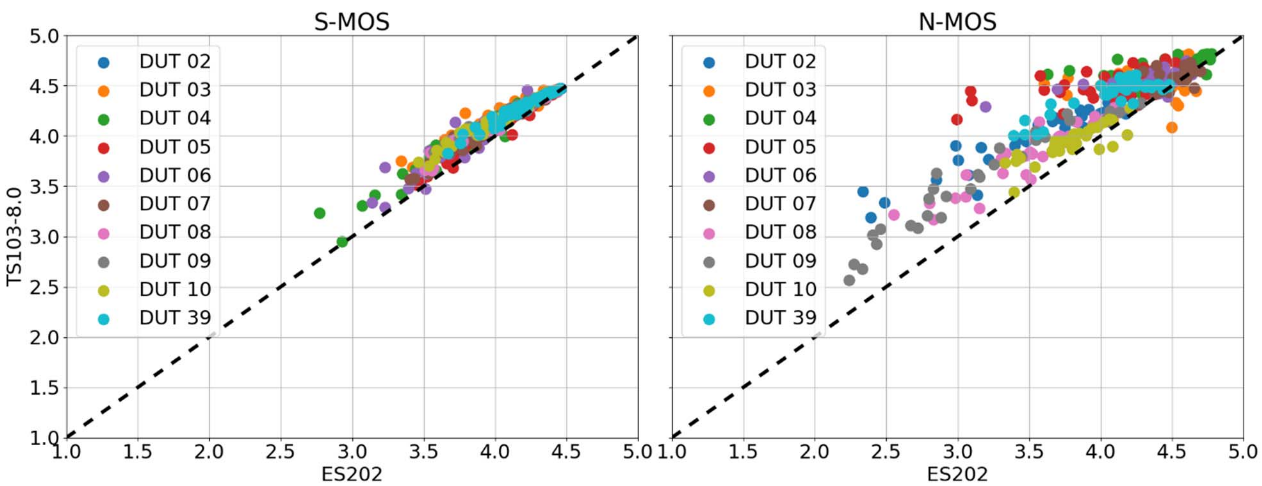


Figure 107: Results S-/N-MOS for TS103-8.0 vs. ES202 (NB)

5.4.7.1.2 Results for all rooms and devices (average across noises)

Figure 108 to Figure 110 provide the results averaged across the eight noise types for NB mode. Each combination of device and room represents one dot in these scatter plots.

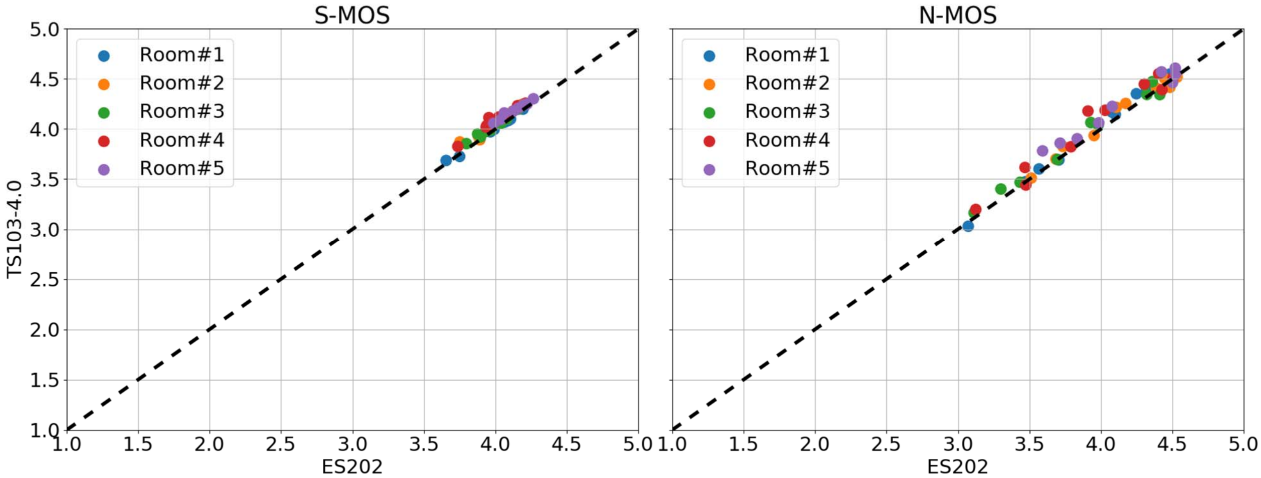


Figure 108: Average results S-/N-MOS for TS103-4.0 vs. ES202 (NB)

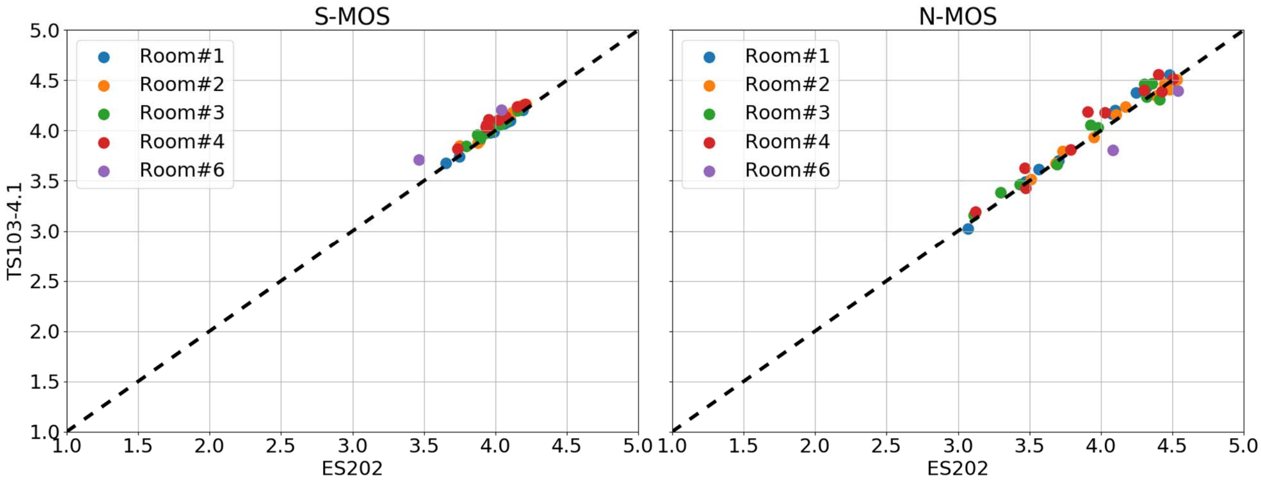


Figure 109: Average results S-/N-MOS for TS103-4.1 vs. ES202 (NB)

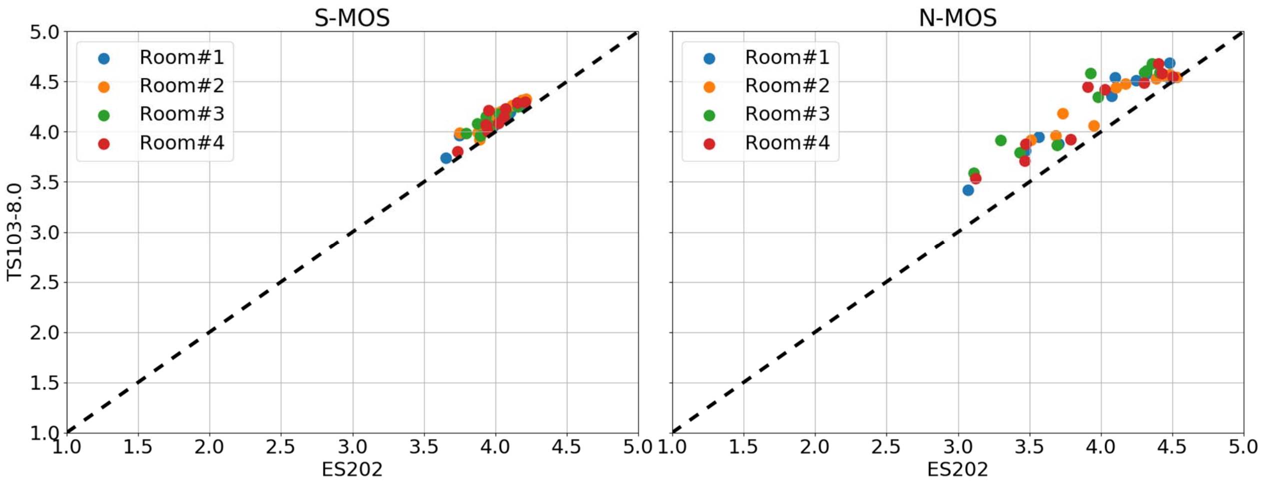


Figure 110: Average results S-/N-MOS for TS103-8.0 vs. ES202 (NB)

5.4.7.2 Wideband Mode

5.4.7.2.1 Results for all noise types, rooms and devices

Figure 111 to Figure 113 provide the results of the three binaural equalization methods compared to the data of ES202 (cf. Table 24) for WB mode. Each combination of device, room and background noise represents one dot in these scatter plots.

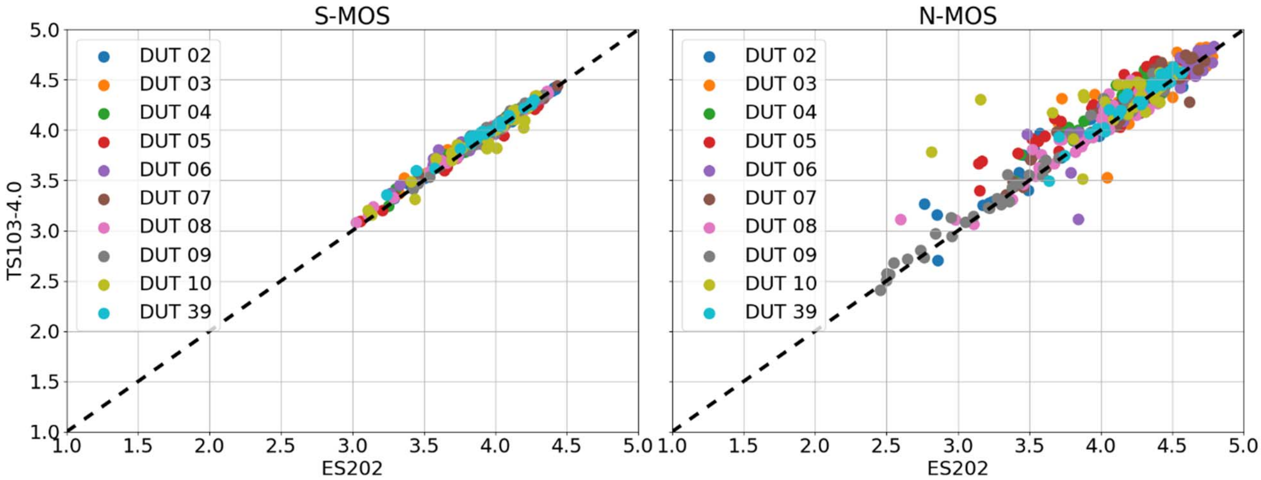


Figure 111: Results S-/N-MOS for TS103-4.0 vs. ES202 (WB)

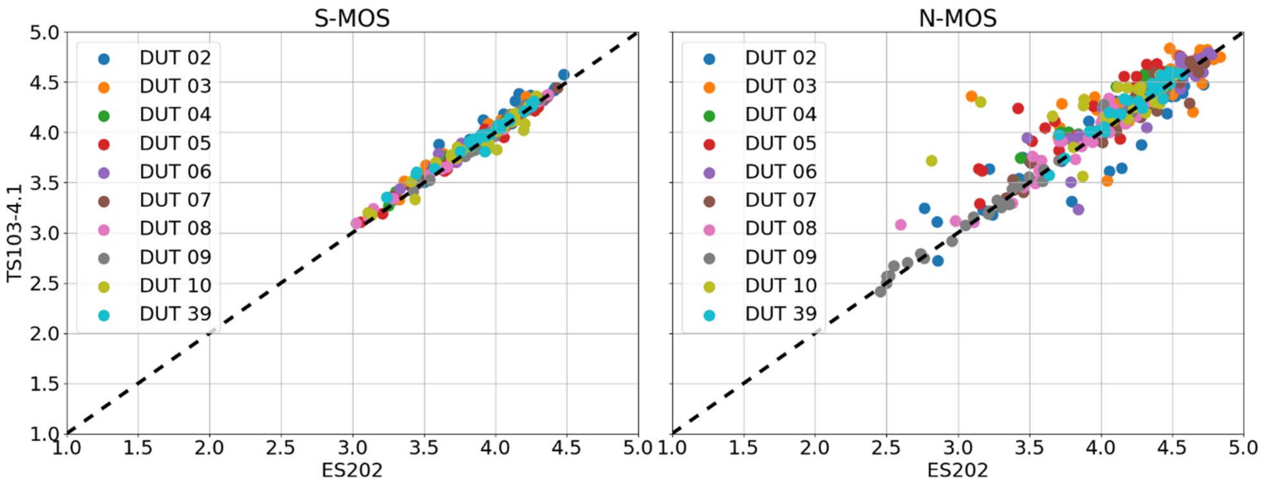


Figure 112: Results S-/N-MOS for TS103-4.1 vs. ES202 (WB)

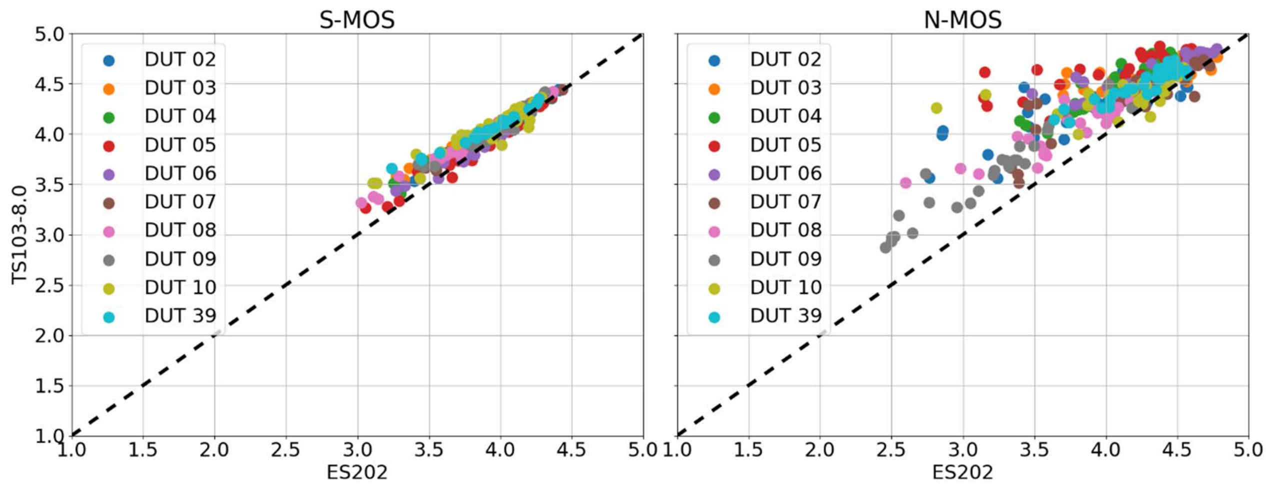


Figure 113: Results S-/N-MOS for TS103-8.0 vs. ES202 (WB)

5.4.7.2.2 Results for all rooms and devices (average across noises)

Figure 114 to Figure 116 provide the results averaged across the eight noise types for WB mode. Each combination of device and room represents one dot in these scatter plots.

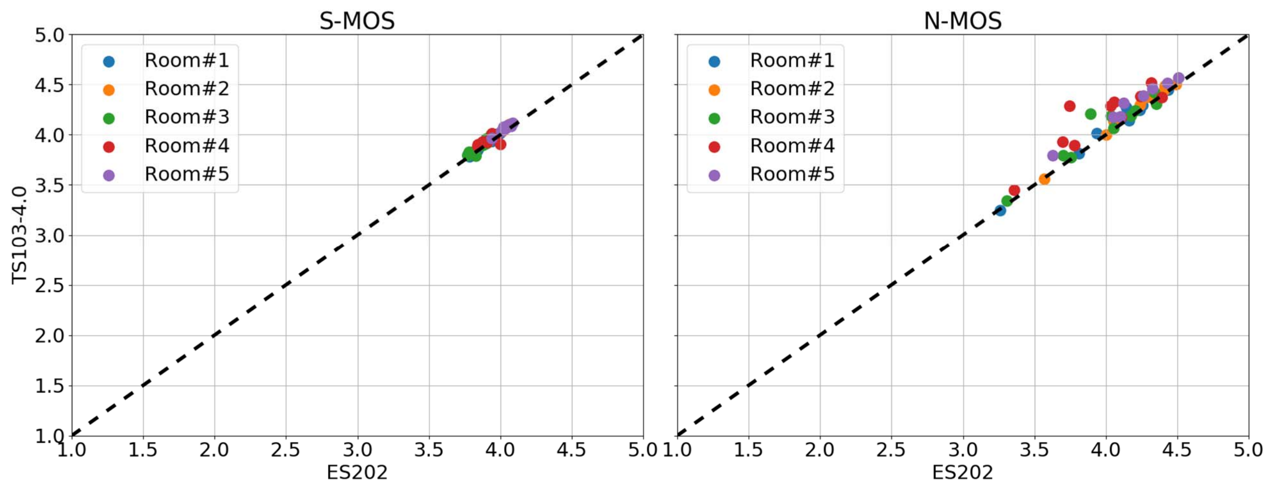


Figure 114: Average results S-/N-MOS for TS103-4.0 vs. ES202 (WB)

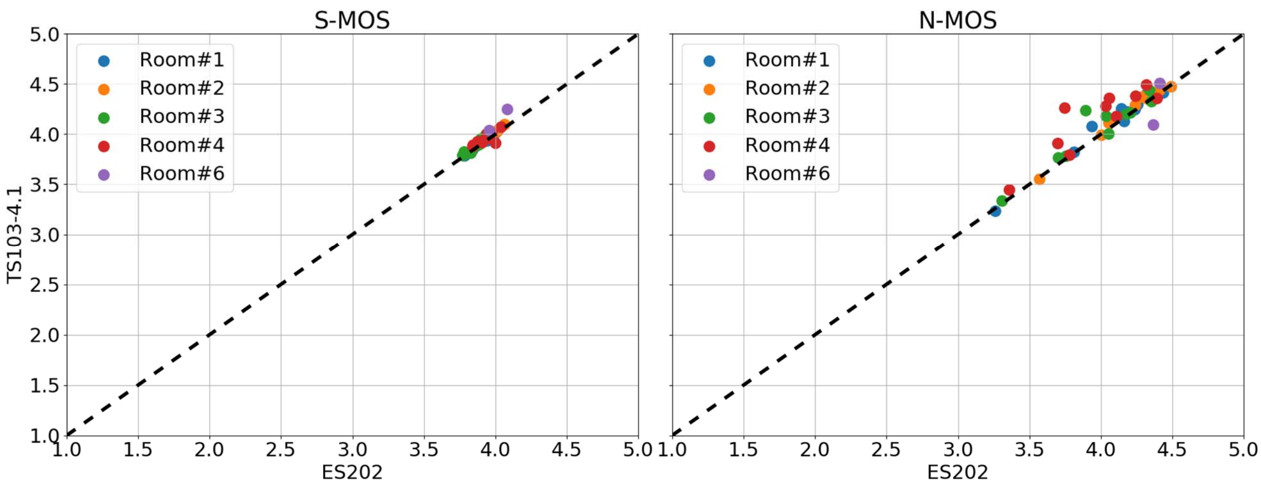


Figure 115: Average results S-/N-MOS for TS103-4.1 vs. ES202 (WB)

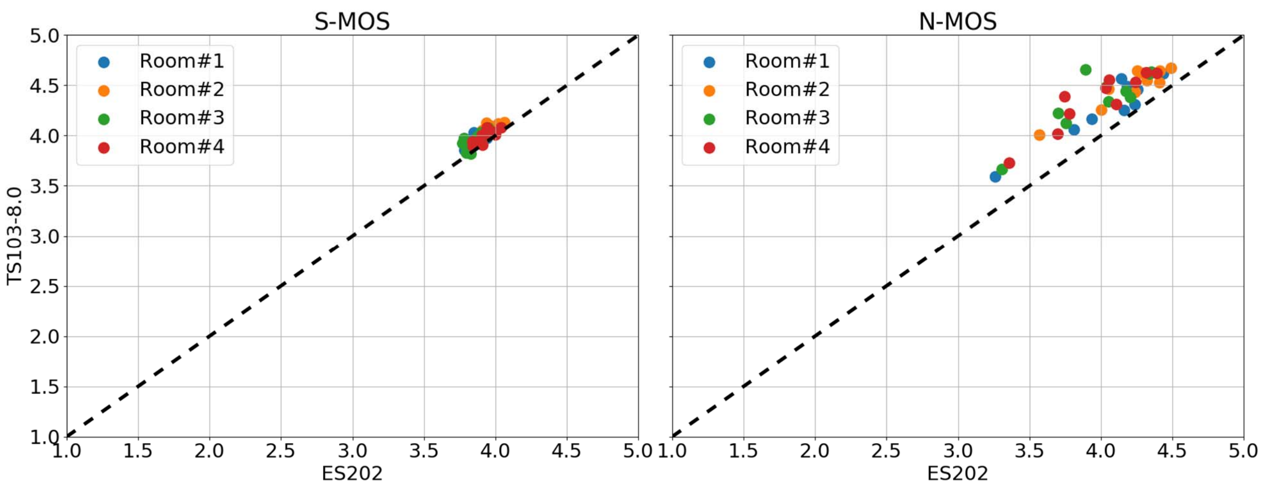


Figure 116: Average results S-/N-MOS for TS103-8.0 vs. ES202 (WB)

5.4.7.3 Super-wideband Mode

5.4.7.3.1 Results for all noise types, rooms and devices

Figure 117 to Figure 119 provide the results of the three binaural equalization methods compared to the data of ES202 (cf. Table 25) for SWB mode. Each combination of device, room and background noise represents one dot in these scatter plots.

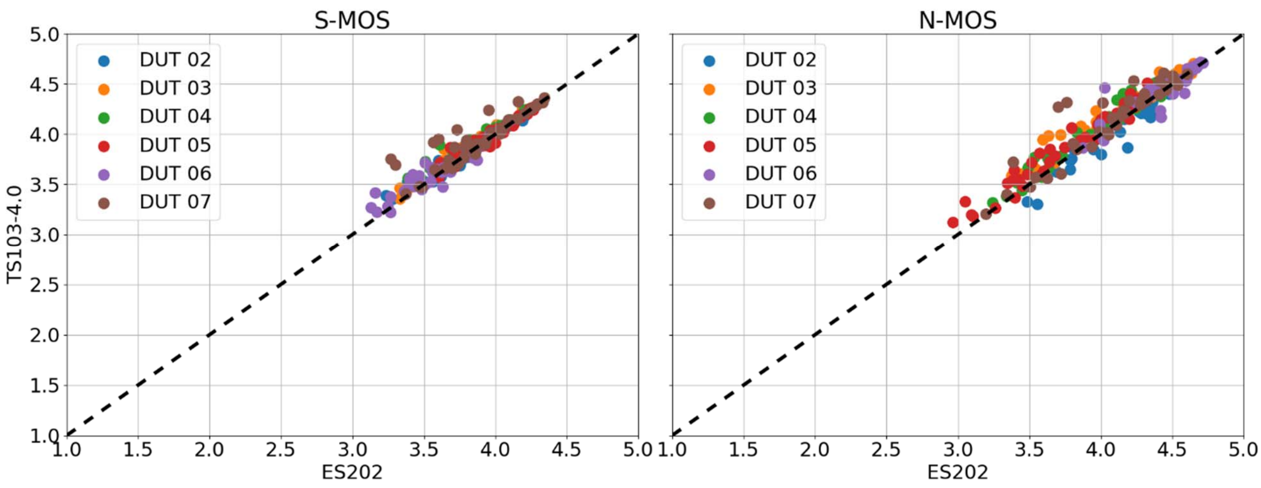


Figure 117: Results S-/N-MOS for TS103-4.0 vs. ES202 (SWB)

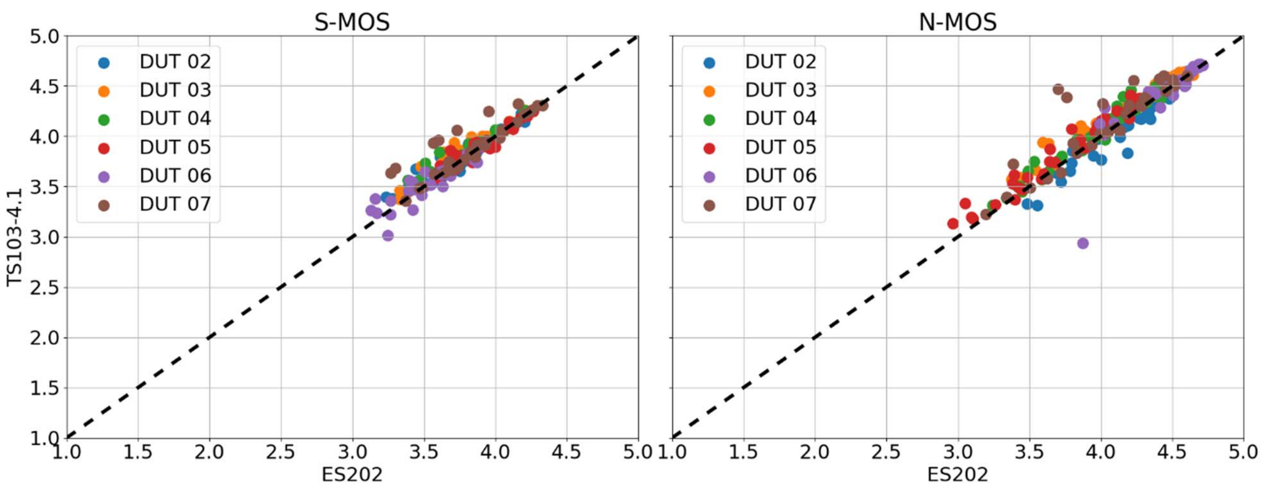


Figure 118: Results S-/N-MOS for TS103-4.1 vs. ES202 (SWB)

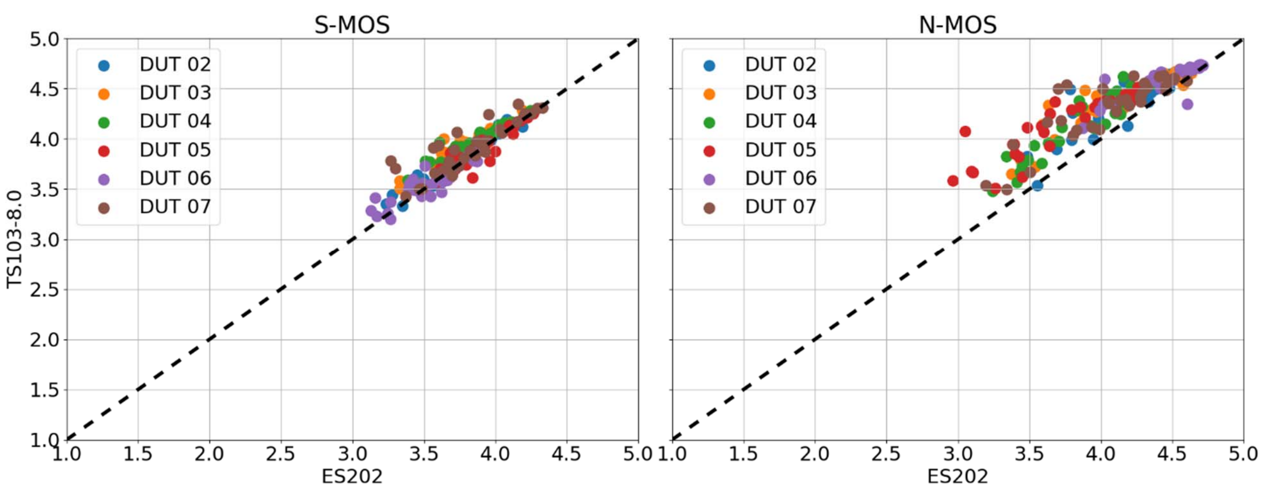


Figure 119: Results S-/N-MOS for TS103-8.0 vs. ES202 (SWB)

5.4.7.3.2 Results for all rooms and devices (average across noises)

Figure 120 to Figure 122 provide the results averaged across the eight noise types for SWB mode. Each combination of device and room represents one dot in these scatter plots.

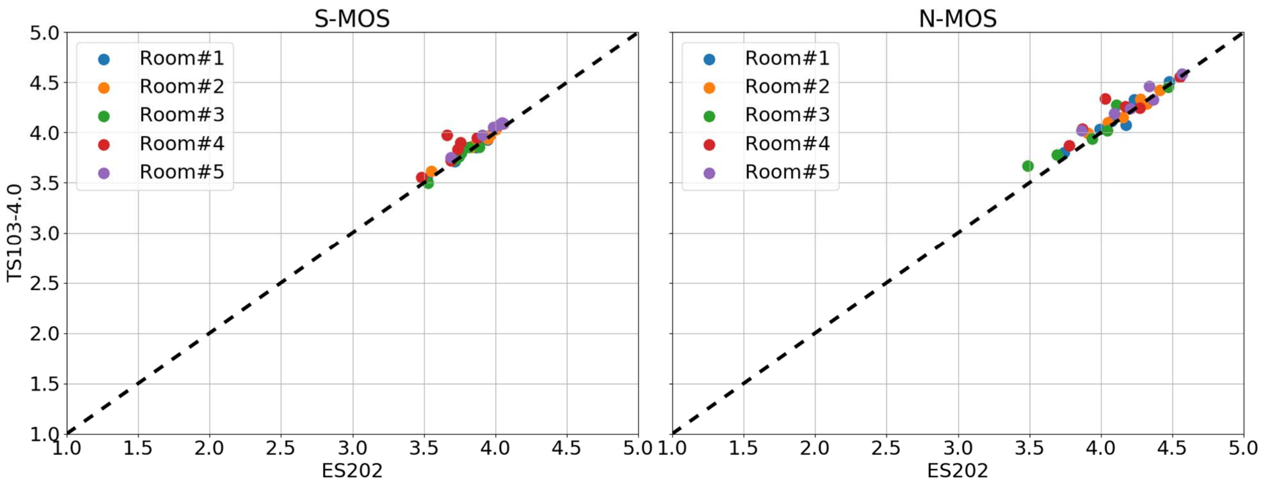


Figure 120: Average results S-/N-MOS for TS103-4.0 vs. ES202 (SWB)

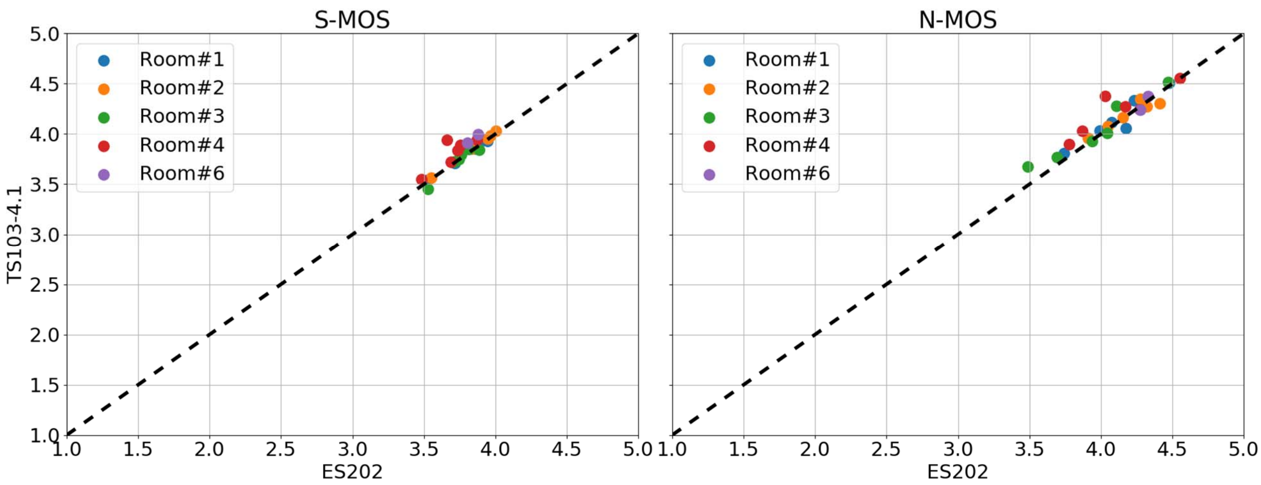


Figure 121: Average results S-/N-MOS for TS103-4.1 vs. ES202 (SWB)

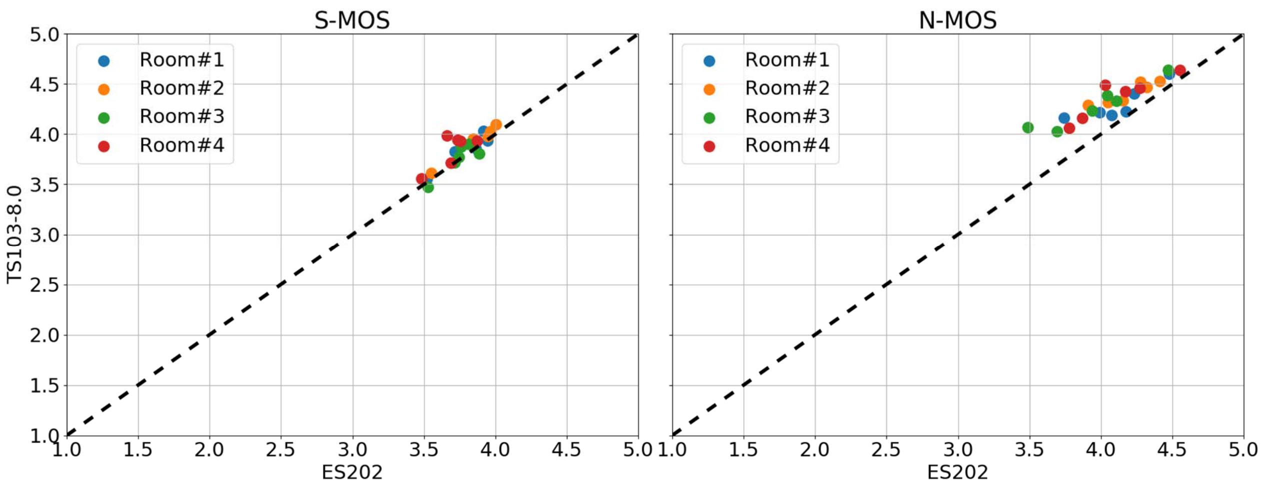


Figure 122: Average results S-/N-MOS for TS103-8.0 vs. ES202 (SWB)

5.4.8 Results for noise field simulation TS 103 224 (8-channel)

As described in the test plan, also the equalization method based on eight-channel microphone array and eight loudspeakers according to ETSI TS 103 224 [5] were investigated in the round robin test (labelled as TS103-HS in the following). Since the noise sources used for this system are different from ES202, results averaged across noises might have a bias or offset.

Figure 123 to Figure 125 provide the results averaged across the eight noise types for NB, WB and SWB mode as listed in Table 12. Similar as in the previous clauses, the abscissas axes show the corresponding result for ES202 for each device. Since different underlying noise types are compared here, offsets or shifts might be included in the data. To illustrate such possible shifts, a linear regression line and corresponding coefficients are depicted in the plots as well.

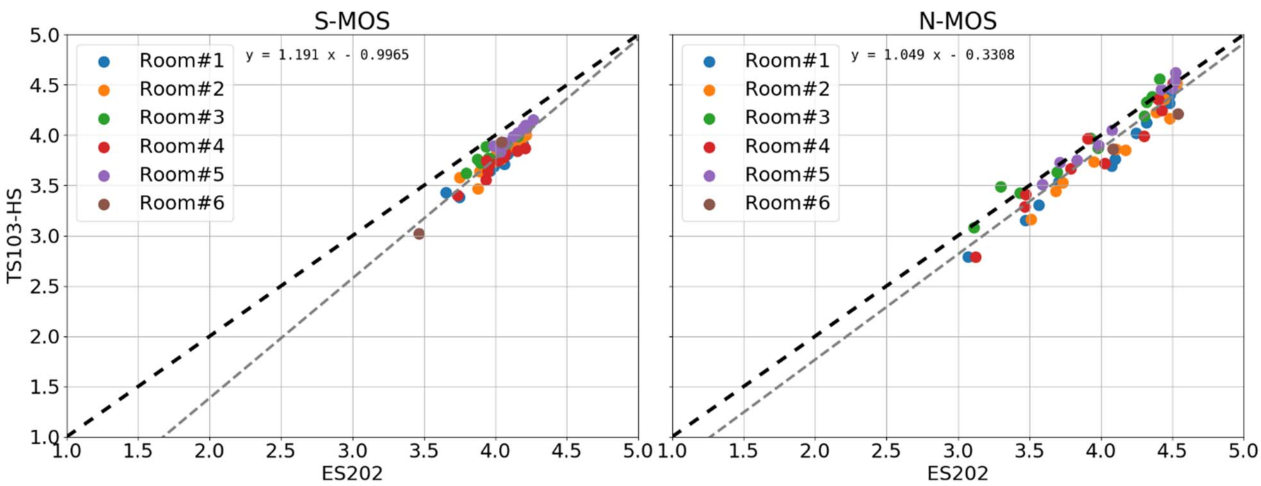


Figure 123: Average results S-/N-MOS for TS103-HS vs. ES202 (NB)

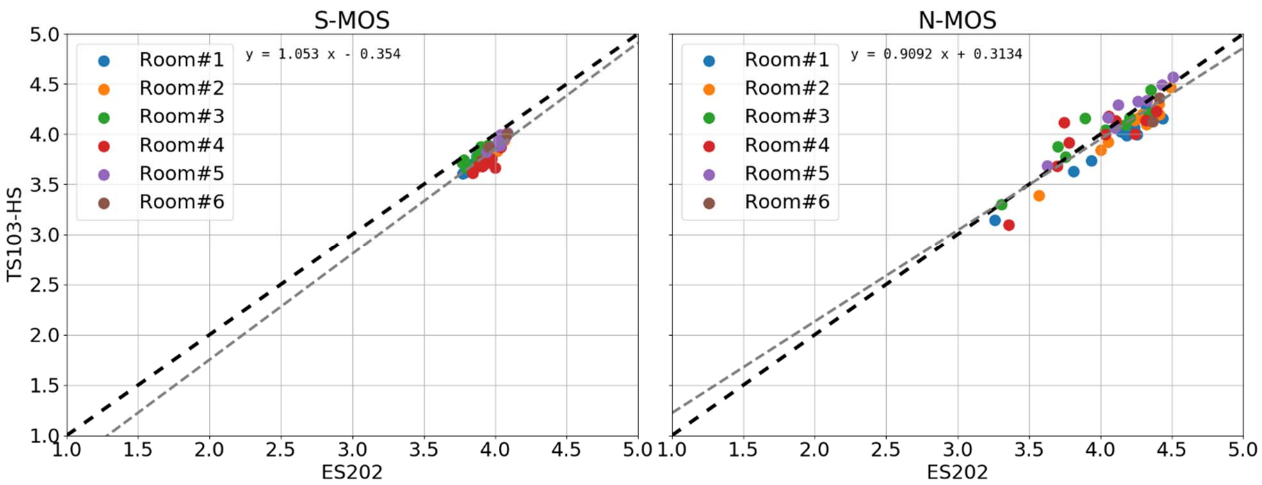


Figure 124: Average results S-/N-MOS for TS103-HS vs. ES202 (WB)

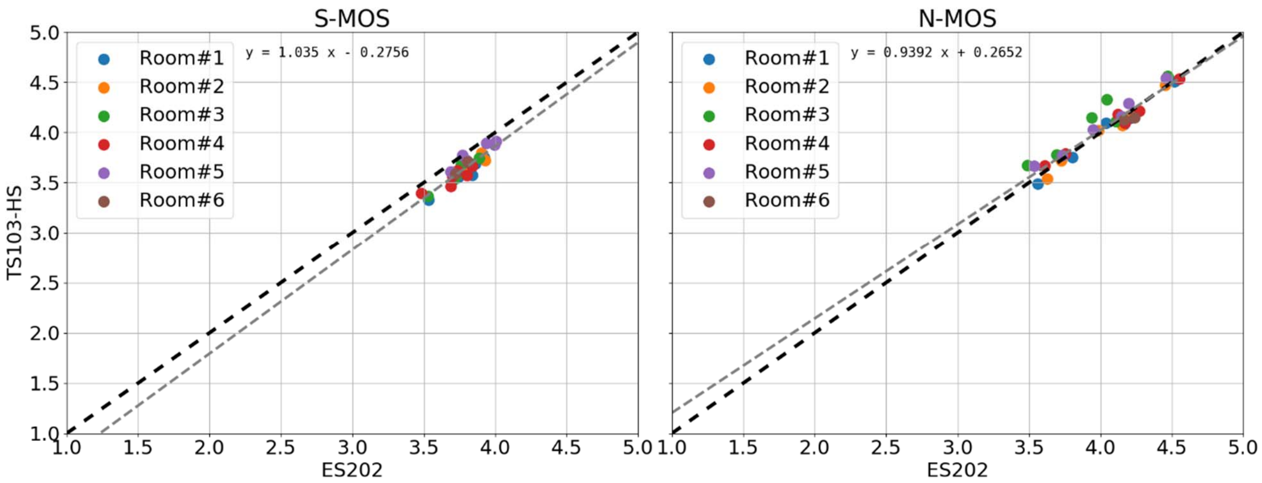


Figure 125: Average results S-/N-MOS for TS103-HS vs. ES202 (SWB)

5.4.9 Outlier Analysis

5.4.9.1 Overview

One major goal of the conducted round robin test is the analysis of consistency across labs. Inconsistencies in results may be explained by several reasons, variance of positioning or inaccurately configured equipment. Examples:

- Measurement equipment components:
  - Calibration and equalization of mouth output may be outdated/incorrect.
  - (Re-)Positioning of additional measurement microphone (at device input microphone) may vary.
- Noise field simulation
  - Equalization procedure includes manual adjustment (for ES202).
  - Calibration of equalization microphones (e.g., ears, microphone array) may be outdated/incorrect.
  - Sound field correction of microphones may be incorrectly configured/applied.
  - Even slight physical modifications of measurement chamber may influence the equalization.

- Devices under test:
    - Noise reduction signal processing may behave highly non-linear and time-invariant.
    - Call may drop close to the end of a (longer) measurement series (observed quite often, especially for VoLTE and EVS-SWB). The prediction model determines S-MOS and N-MOS (close) to 1.0, but due to averaging, this may not be directly noticeable in the overall results.

Since the recordings obtained during the round robin test also include side-information (like e.g., silence recordings or acoustic/unprocessed signals close to the device input are available for all devices), more detailed analyse are possible in order to investigate if outliers/unexpected results can be explained by the one or other reason. Several examples of observed inconsistencies in this test are described in the following clauses.

5.4.9.2 Inconsistency in mouth playback

For Room#5, it was noticed that results for S-/N-MOS were consistently higher than for all other rooms. It was assumed that the level of the mouth playback was not correctly adjusted. By analysing the active speech level according to Recommendation ITU-T P.56 [7], this could only partially be confirmed. Figure 126 shows the ASL across devices and measurement rooms. In most cases, the violet bar (room #5) provides the highest speech level. However, for DUT08, this trend cannot be confirmed and the speech level is identical for all rooms, but could be due to specific gain control.

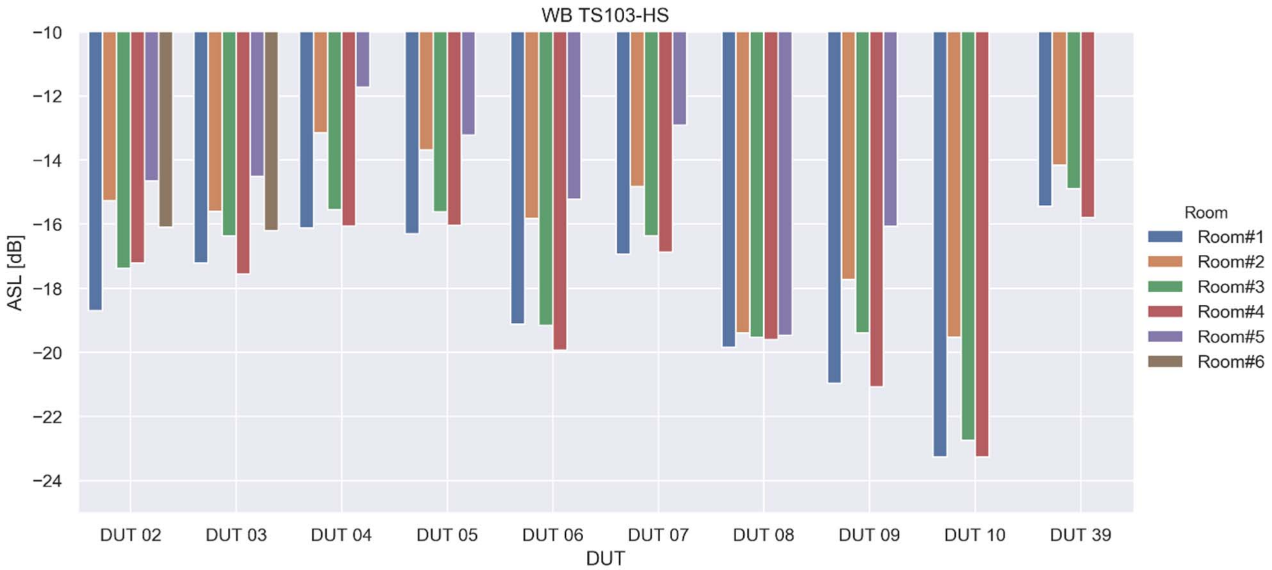


Figure 126: ASL of uplink signal in silence

In order to investigate this issue more in detail, an analysis of the unprocessed signal was conducted. According to the measurement description of ETSI TS 103 106 [6] and TS 26.132 [3], this signal is captured acoustically with a reference microphone close to the input microphone of the DUT. With an equalized and calibrated mouth playback, any observed variance in these recordings (e.g. due to positioning of the reference microphone) are independent of the signal processing of the DUT. Figure 127 shows a 1/3rd octave band analysis of the unprocessed signal in silence for DUT 08. In contrast to the previous analysis, here a difference of 5-6 dB compared to the four other rooms over the whole frequency range can be observed. This also implies that DUT 08 obviously has a quite strong automatic gain control.

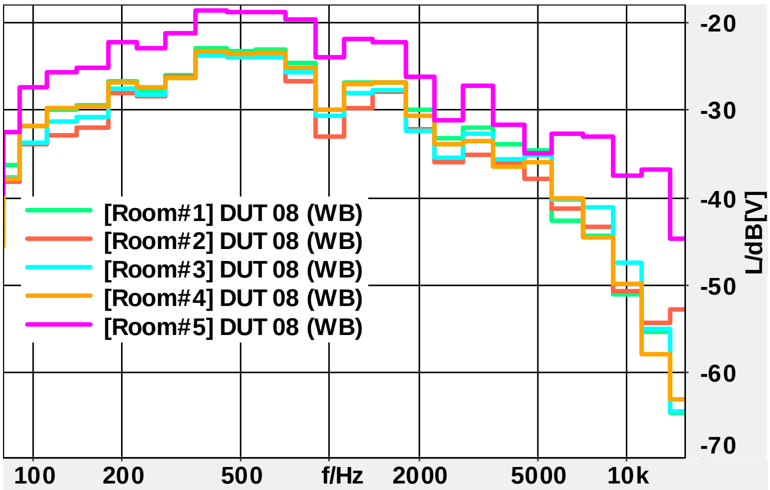


Figure 127: Average 1/3rd octave spectrum of silence recording for DUT 08

To quantify this observation in terms of active speech level, the unprocessed signals in silence of DUT 08 of all rooms and bandwidth modes (NB, WB and SWB) were analysed. Table 26 provides the levels according to Recommendation ITU-T P.56 [7]. Here the numbers confirm the spectrum analysis, here the difference of 5-6 dB can be observed as well.

Table 26: ASL of unprocessed signals in silence for DUT 08

Active Speech Level [dB]			
Room	**Min.**	**Avg.**	**Max.**
Room#1	-12.17	-12.09	-12.00
Room#2	-13.41	-13.38	-13.35
Room#3	-13.11	-13.10	-13.10
Room#4	-12.50	-12.40	-12.29
Room#5	-7.61	-7.60	-7.59

Finally, the ASL analysis was conducted for all rooms, devices and bandwidths, results are shown in Table 27. The highest speech level obtained for Room#5 (-4.43 dB Pa) is even higher than the source file level at MRP (-4.7 dB Pa), which seem not reasonable at all.

Table 27: ASL of unprocessed signals in silence for all devices

Active Speech Level [dB] - Average vs. DUT/Bandwidth				
Room	**Min.**	**Avg.**	**Max.**	**Count**
Room#1	-13.45	-12.11	-10.17	26
Room#2	-13.41	-11.24	-7.39	26
Room#3	-13.63	-11.96	-10.68	28
Room#4	-15.29	-13.06	-9.96	26
Room#5	-8.84	-7.42	-4.43	23
Room#6	-11.99	-11.24	-9.55	6

Again, these numbers support the initial assumption that there is an issue with the mouth playback calibration and/or equalization. Unfortunately, the set-up in room#5 is no longer available for debugging; for this reason, Room#5 is removed for some further analyses.

5.4.9.3 Inconsistency of Device

State-of-the-art signal processing in mobile devices may be quite complex, especially for the task of noise reduction. Since it cannot be considered as linear and time-invariant, testing of ambient noise performance is conducted for different noises and across numerous speech samples to compensate for such uncertainty. However, in some exceptional cases, even such advanced tests may fail.

As an example, an outlier of Figure 111 is highlighted in Figure 128 (DUT 10 in WB mode). For ES202, N-MOS was determined as 2.82, for TS103-4.0 to 3.78 (Pub noise). A difference of approximately 1.0 MOS for the same room, noise type, and bandwidth mode seems to be quite high. In addition, this deviation could not be observed in NB mode.

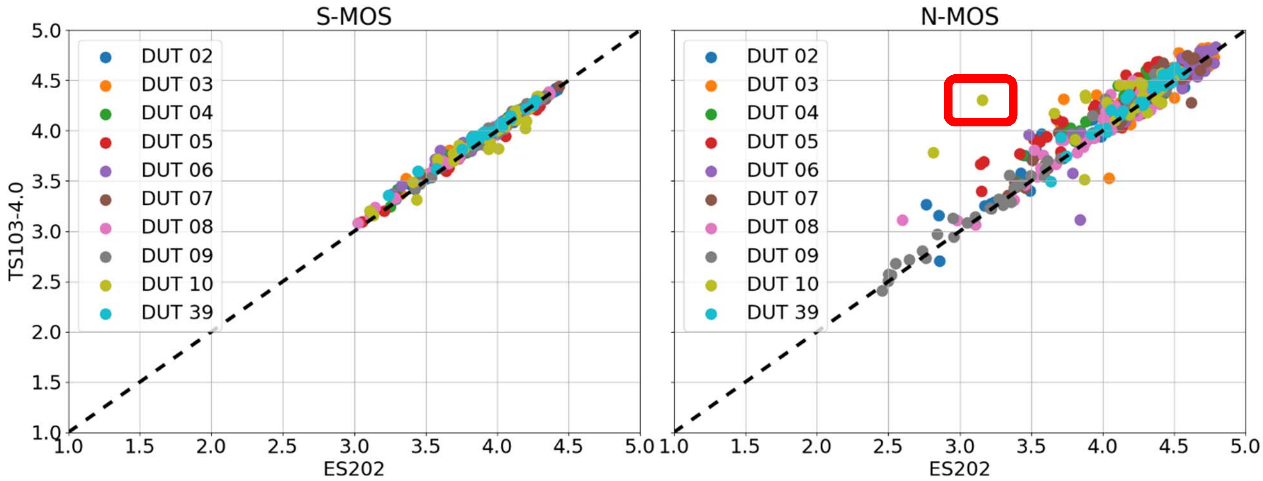


Figure 128: Results S-/N-MOS for TS103-4.0 vs. ES202 (WB)

This outlier was investigated further in detail. First, the level vs time analysis (time constant 35 ms) was calculated for the uplink signal. Figure 129 shows an excerpt of the recording for DUT 10, room#4 and Pub noise type in WB mode. It can be observed that the speech level is almost identical in both cases, but the residual noise in the speech pauses differ quite a lot. Obviously, the lower performance in N-MOS results from a much higher noise level for ES202.

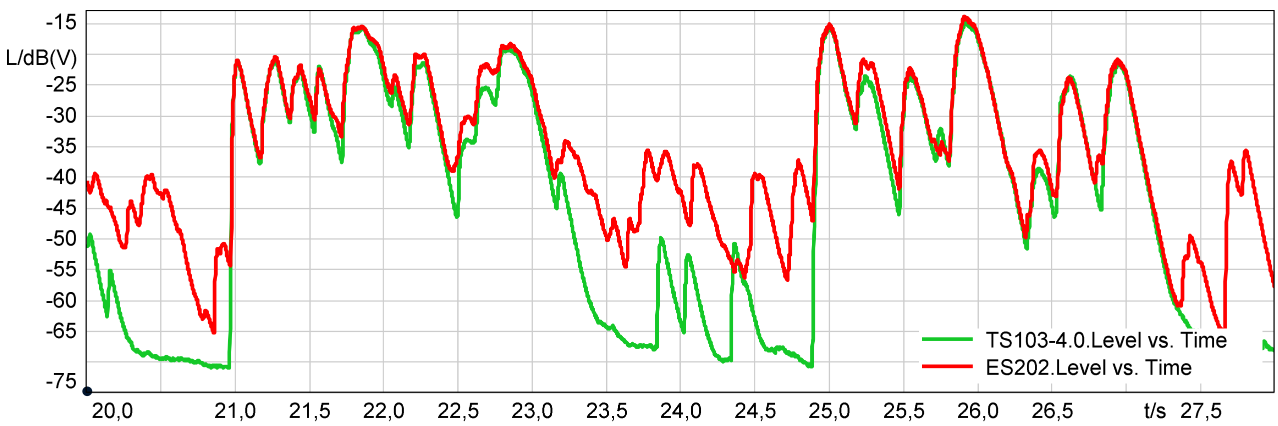


Figure 129: Level vs time analysis (example) ES202 vs TS103-4.0 (room#4, Pub Noise)

As a second analysis, the unprocessed signal of this recording was analysed. The measurement sequence includes some additional trailing pause (80.0-83.0 s), where no speech is active. For this time range, the 1/3rd octave band spectrum was calculated for the noise field simulations ES202 and TS103-4.0 for the two measurements in NB and WB mode, as shown in Figure 130.

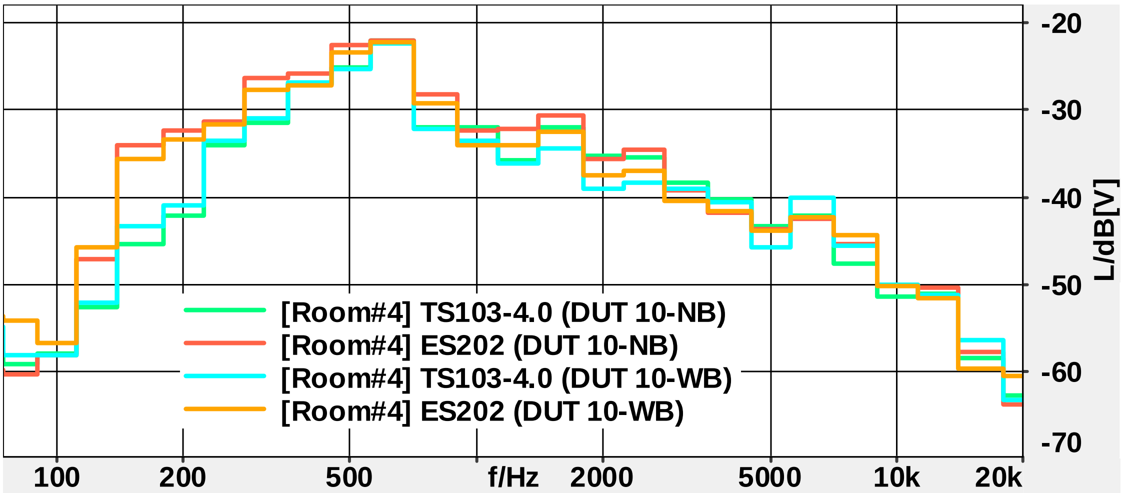


Figure 130: 1/3rd octave band spectrum of noise-only time range of room#4 (Pub Noise)

As expected, the sound field reproduction is consistent between NB and WB measurements. However, level differences between ES202 and TS103-4.0 are clearly visible, especially for low frequencies (about 10 dB at 150 Hz). The fullband level is determined to 76 dB SPL for ES202 and 77 dB SPL for TS103-4.0. The signal processing seems to react quite sensitive to this very singular deviation in the spectrum, while other devices perform very similar for ES202 and TS103-4.0 (and TS103-4.1) setups with these noise fields.

5.4.10 Variance across rooms

5.4.10.1 Overview

In order to analyse the variance across rooms, the average results across all rooms are use as reference for comparisons in the following (denoted as "Room#Avg"). In the following scatter plots, results of all bandwidth modes (NB, WB and SWB) are included. For each comparison, the singular as well as the averaged results across background noises are reported.

NOTE 1: Due to the increased mouth playback level, the results of room#5 are removed for the following analyses.

NOTE 2: Due to the low number of comparisons, the results of room#6 are removed for the following analyses.

5.4.10.2 ES 202 396-1

Figure 131 to Figure 134 provide the comparison results for ES202 of rooms #1 to #4 versus the average across all rooms. The upper sub-figures shows the results for each single noise type (one dot per DUT and noise type) and the lower sub-figures show the averaged results across eight noise types (one dot per DUT).

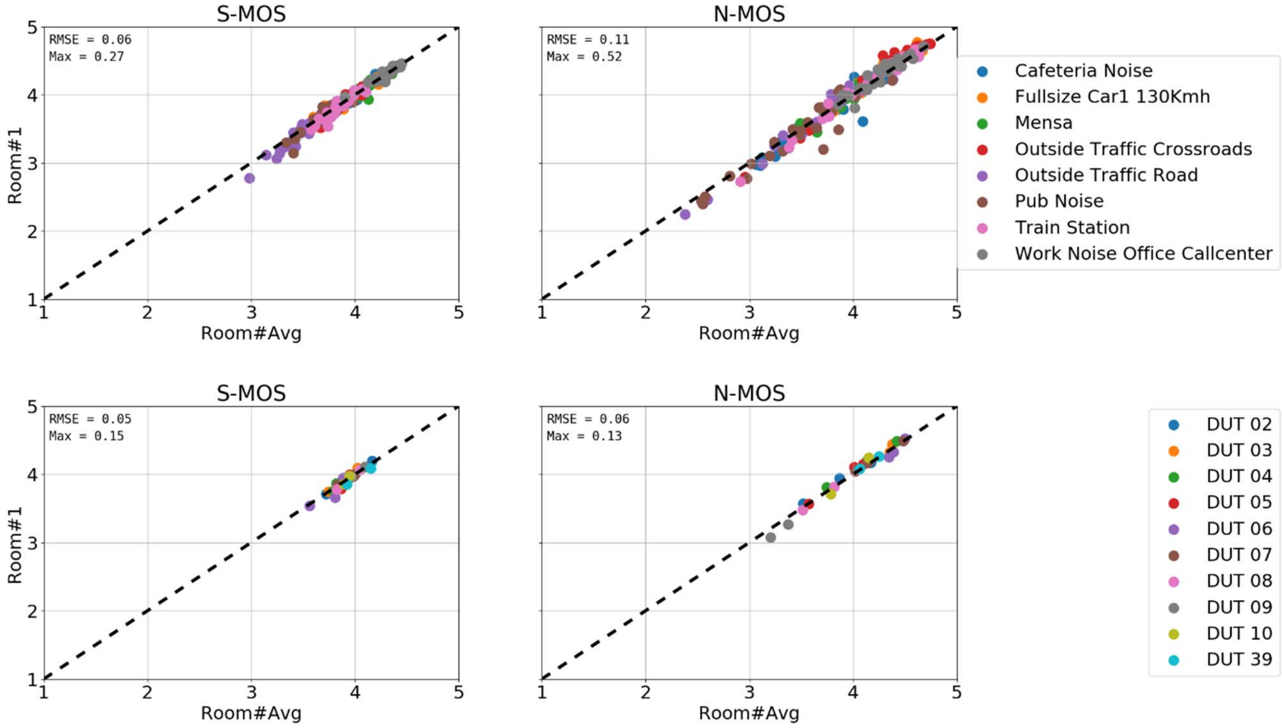


Figure 131: Comparison of Room#1 vs Room#Avg (ES202)

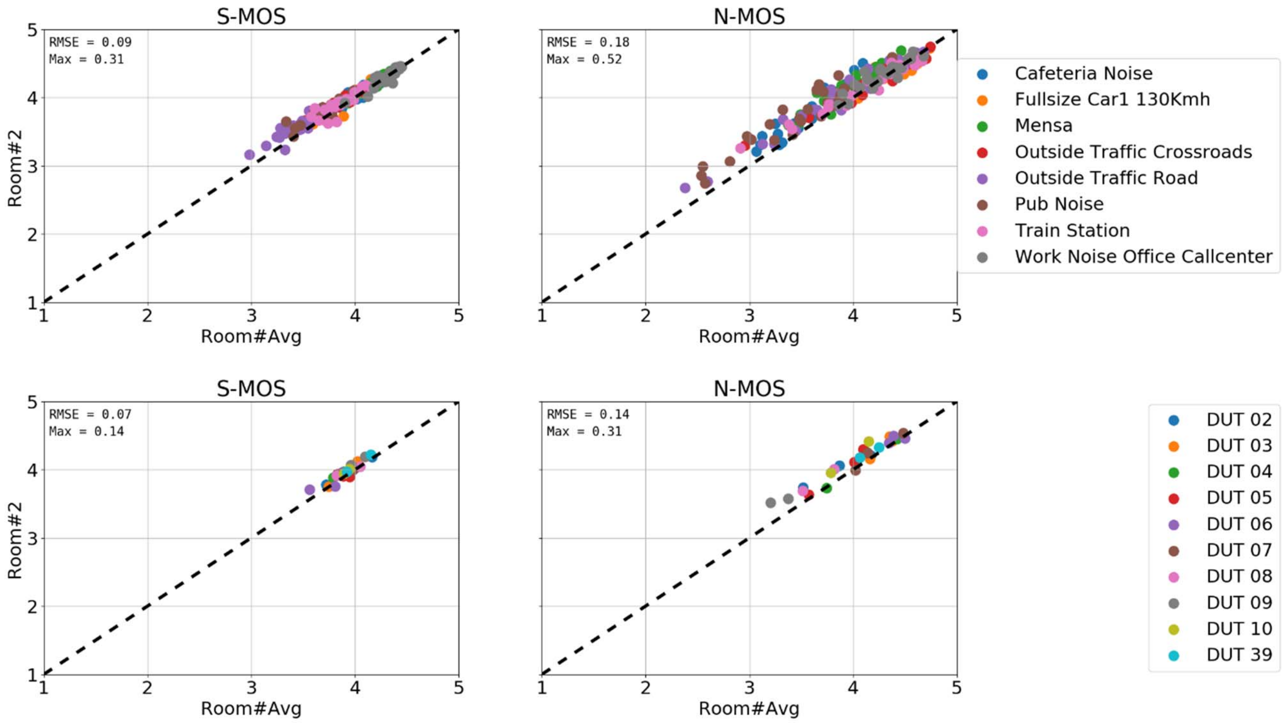


Figure 132: Comparison of Room#2 vs Room#Avg (ES202)

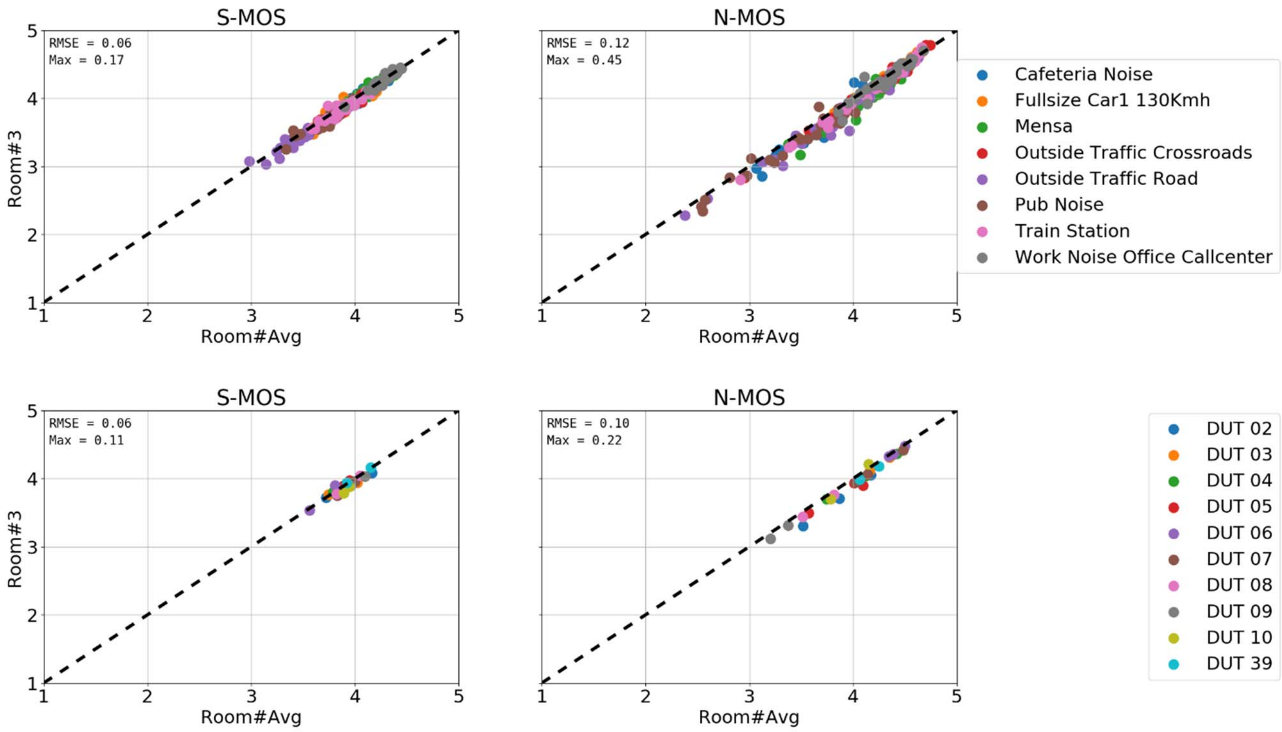


Figure 133: Comparison of Room#3 vs Room#Avg (ES202)

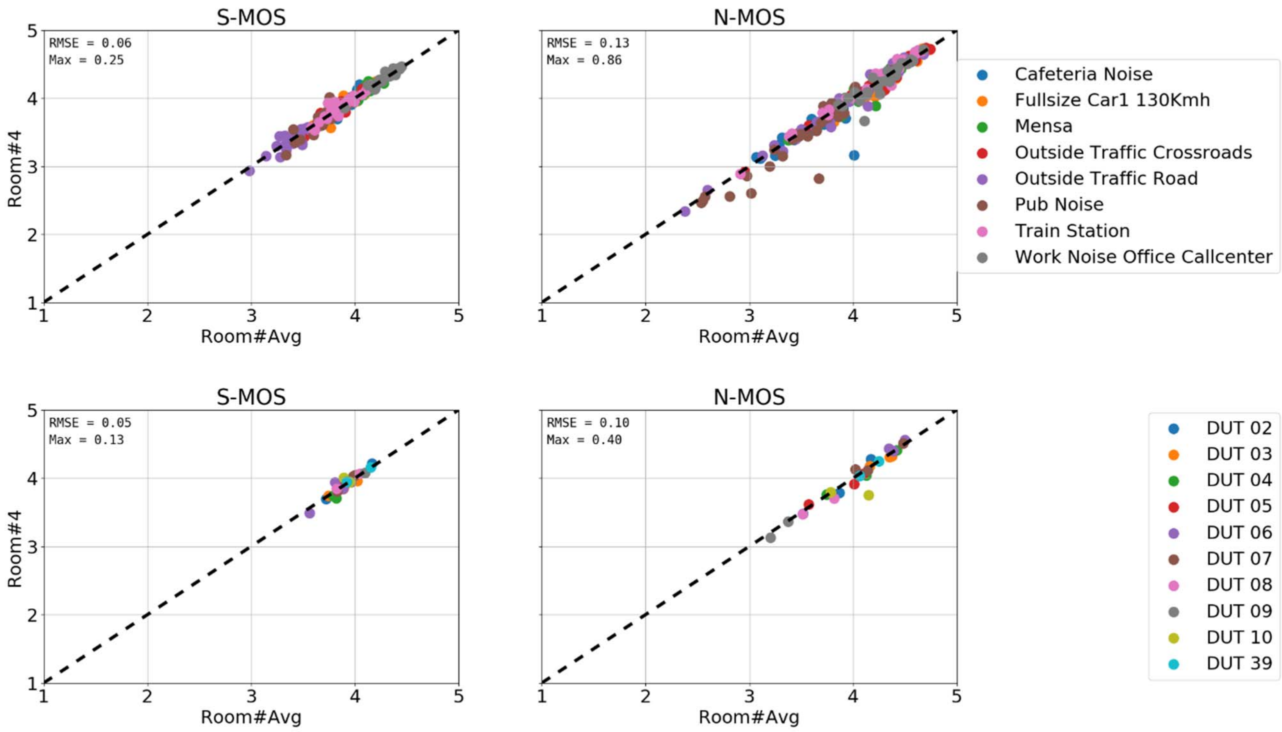


Figure 134: Comparison of Room#4 vs Room#Avg (ES202)

5.4.10.3 TS 103 224 (binaural), 4.0 setup

Figure 135 to Figure 138 provide the comparison results for TS103-4.0 of rooms #1 to #4 versus the average across all rooms. The upper sub-figures shows the results for each single noise type (one dot per DUT and noise type) and the lower sub-figures show the averaged results across eight noise types (one dot per DUT).

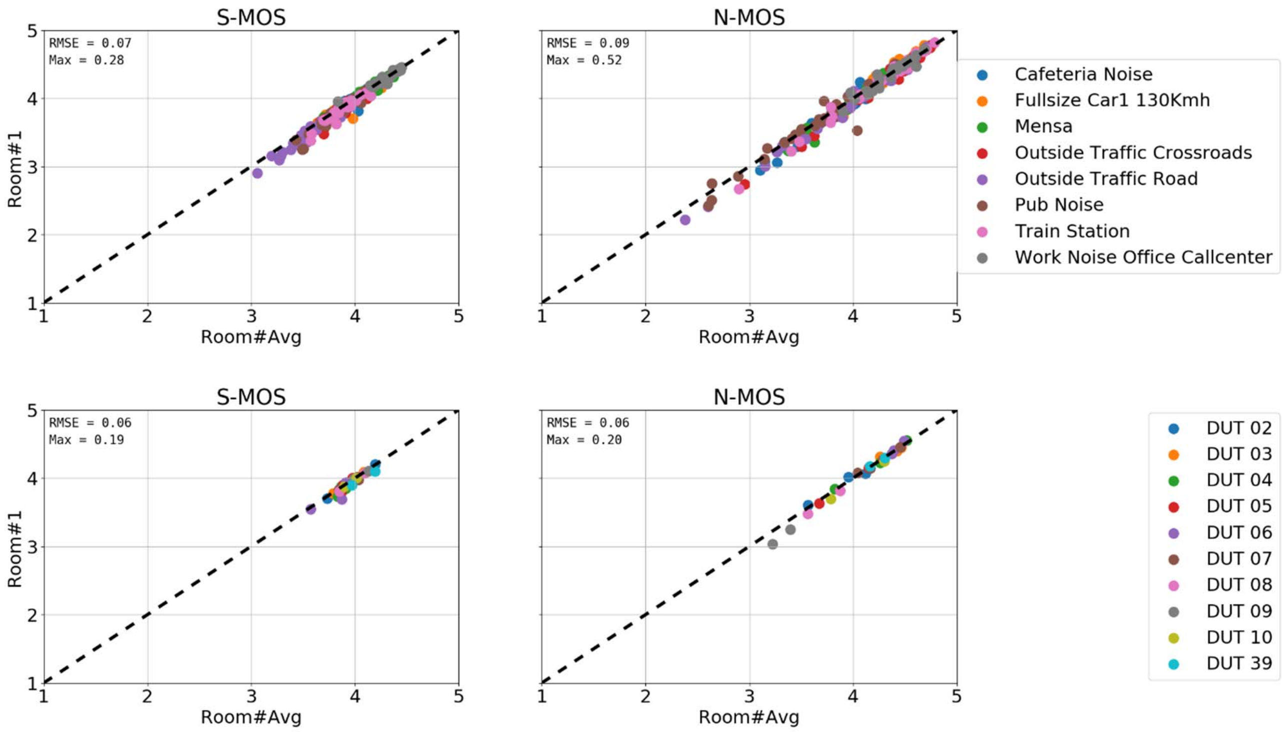


Figure 135: Comparison of Room#1 vs Room#Avg (TS103-4.0)

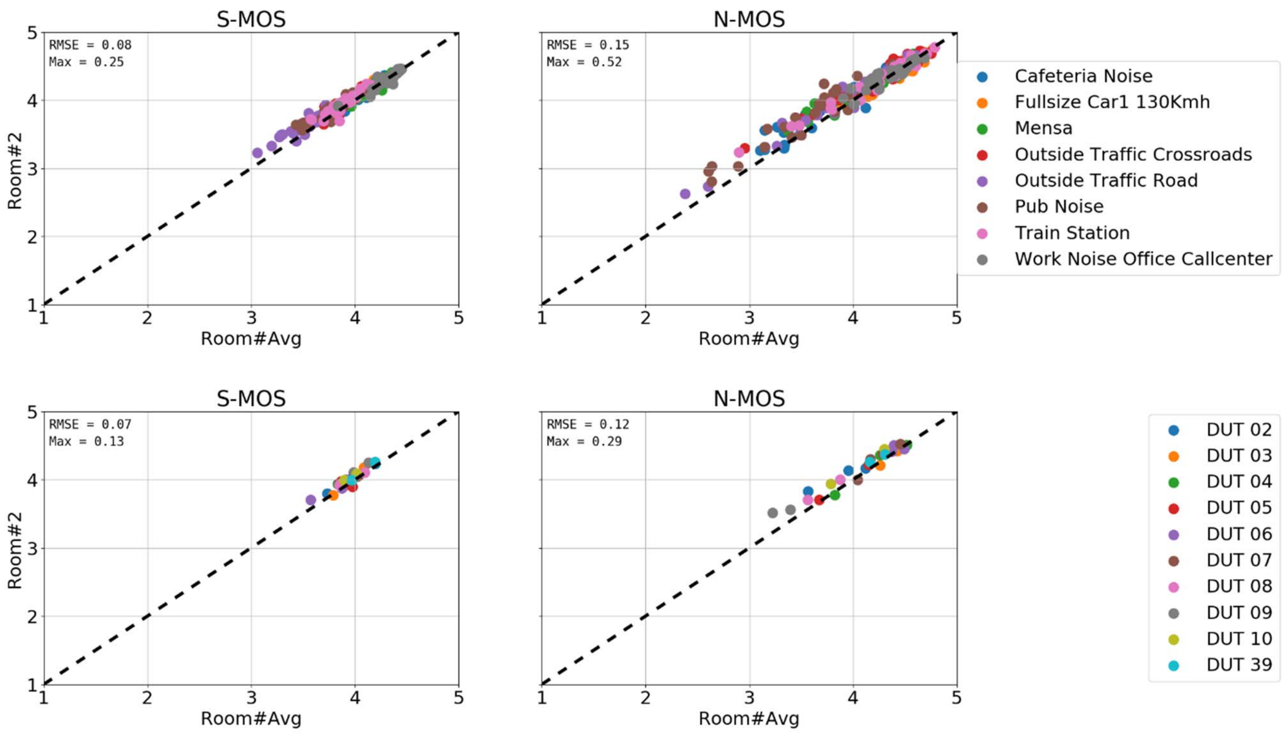


Figure 136: Comparison of Room#2 vs Room#Avg (TS103-4.0)

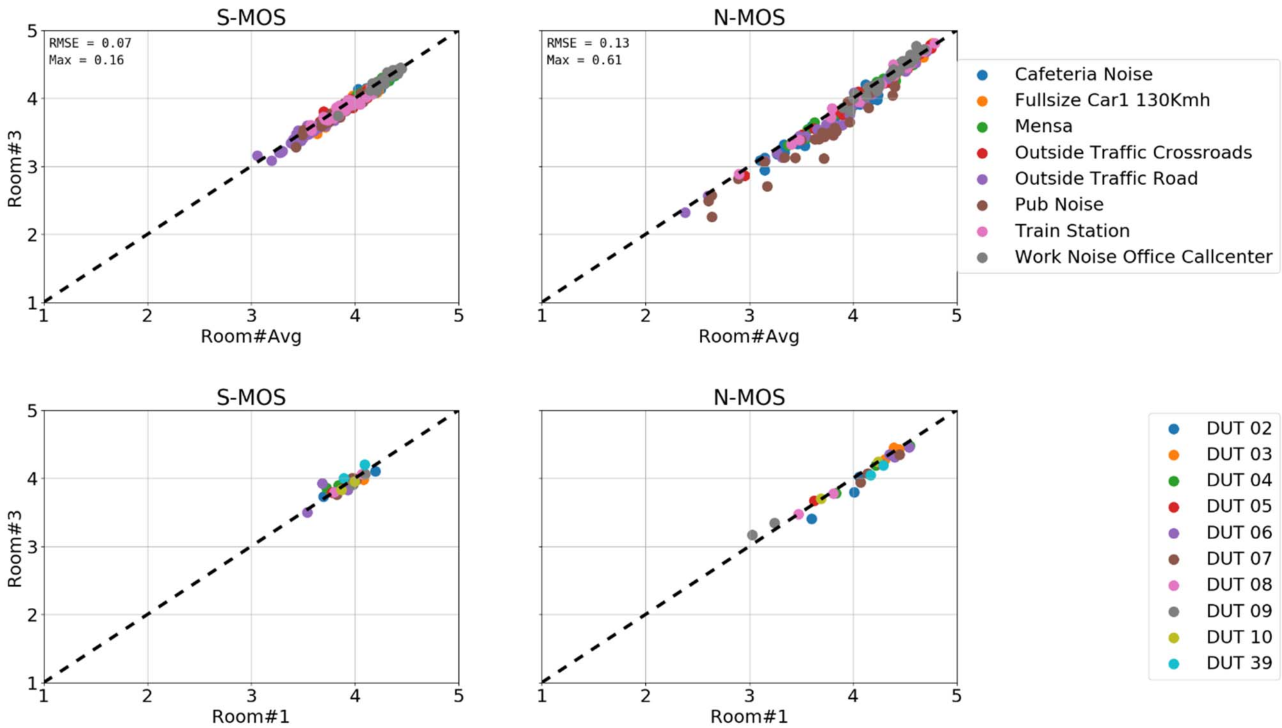


Figure 137: Comparison of Room#3 vs Room#Avg (TS103-4.0)

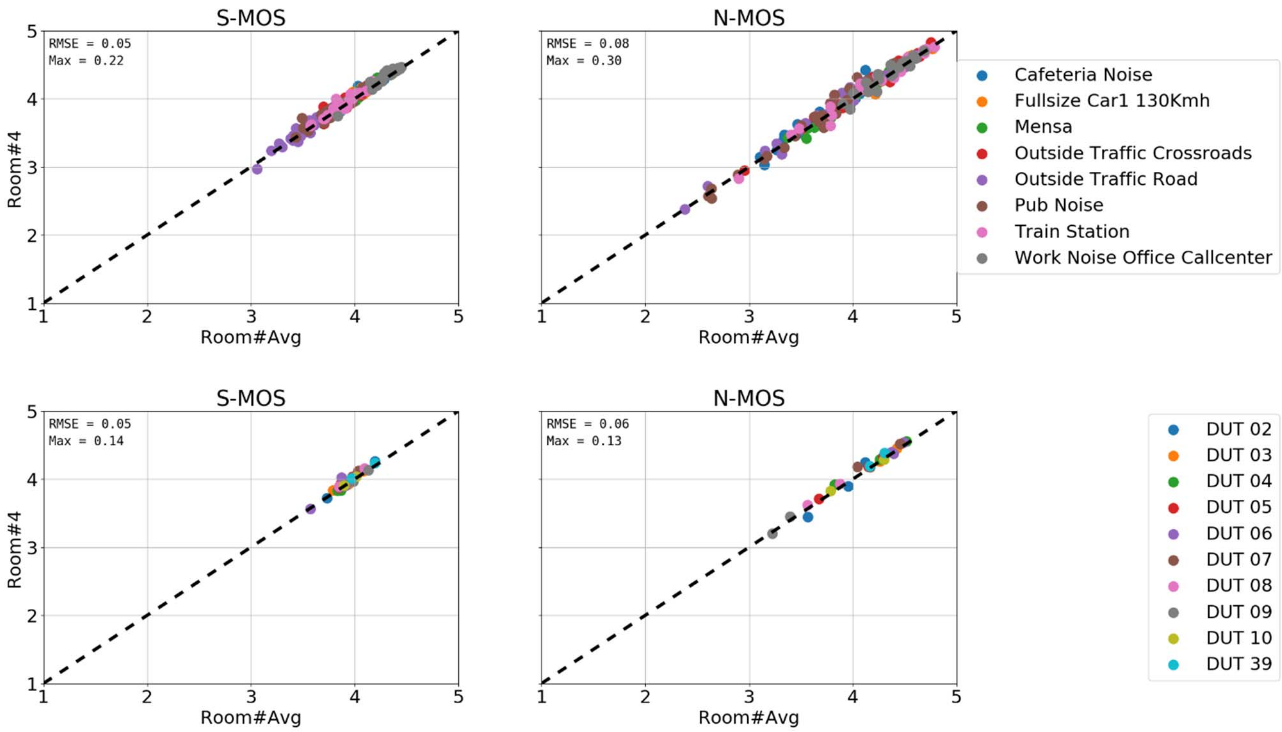


Figure 138: Comparison of Room#4 vs Room#Avg (TS103-4.0)

5.4.10.4 TS 103 224 (binaural), 4.1 setup

Figure 139 to Figure 142 provide the comparison results for TS103-4.1 of rooms #1 to #4 versus the average across all rooms. The upper sub-figures shows the results for each single noise type (one dot per DUT and noise type) and the lower sub-figures show the averaged results across eight noise types (one dot per DUT).

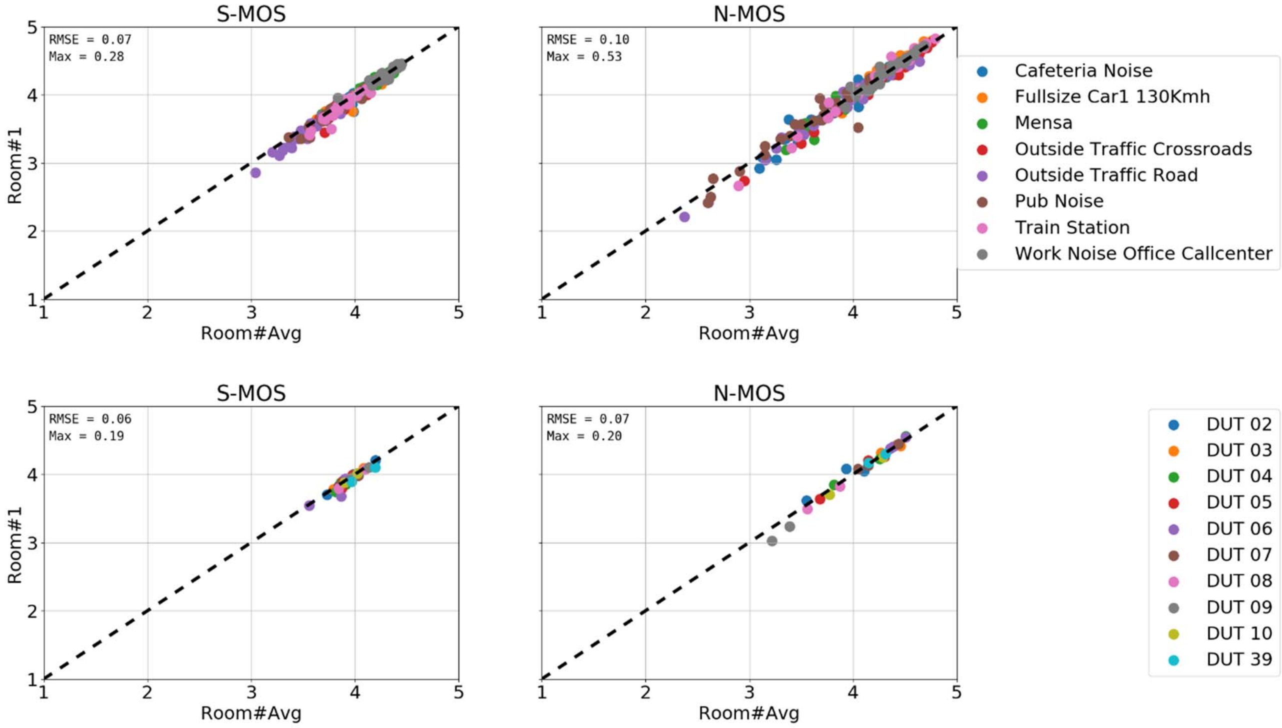


Figure 139: Comparison of Room#1 vs Room#Avg (TS103-4.1)

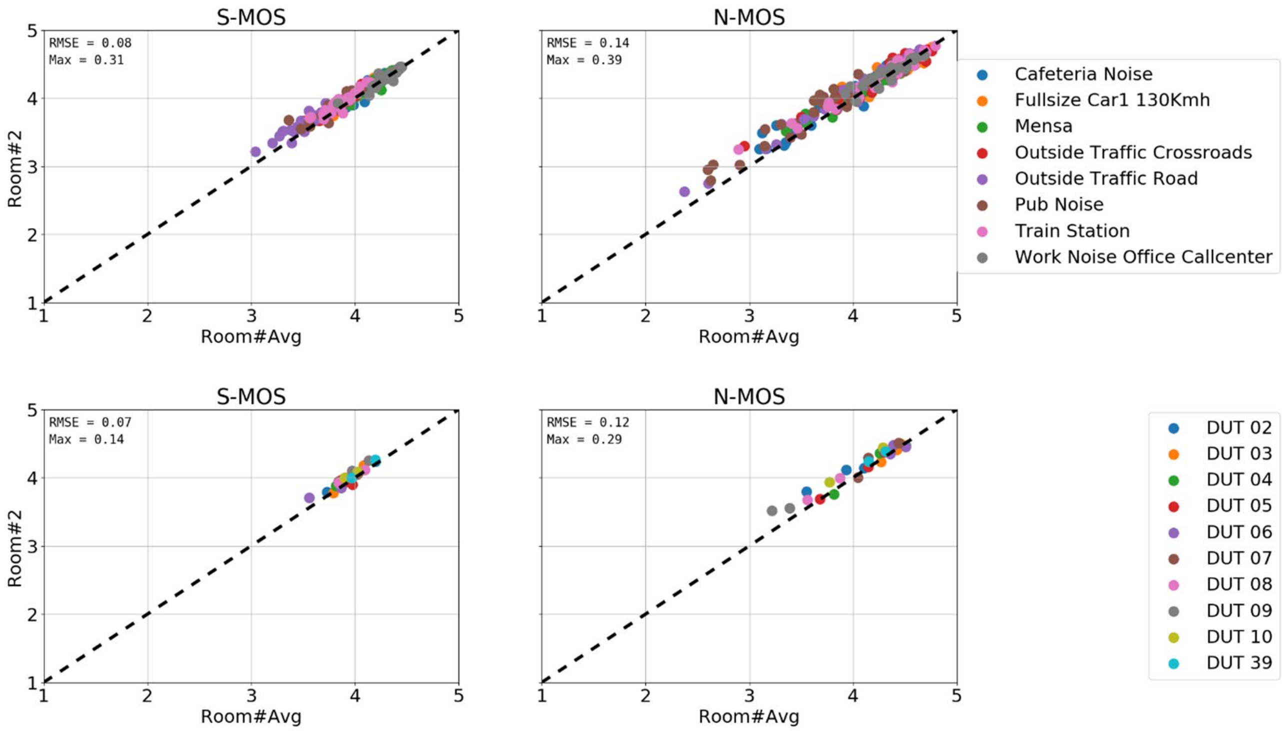


Figure 140: Comparison of Room#2 vs Room#Avg (TS103-4.1)

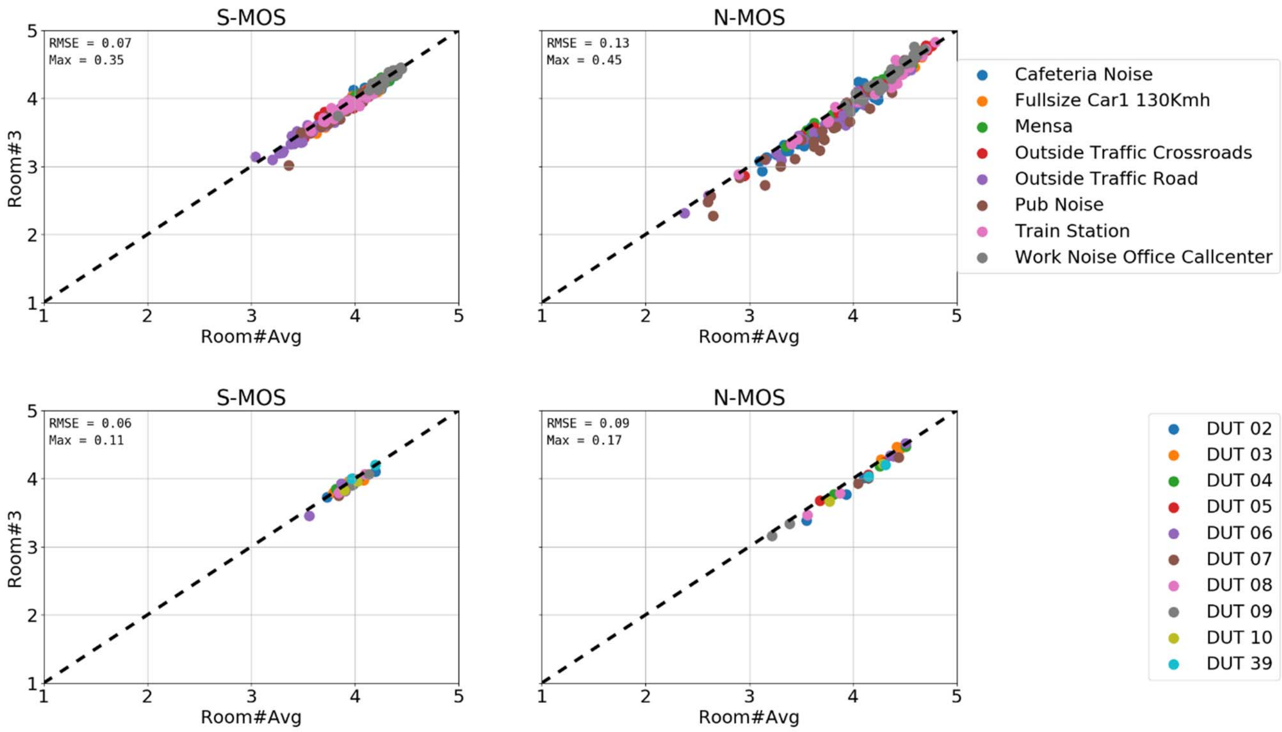


Figure 141: Comparison of Room#3 vs Room#Avg (TS103-4.1)

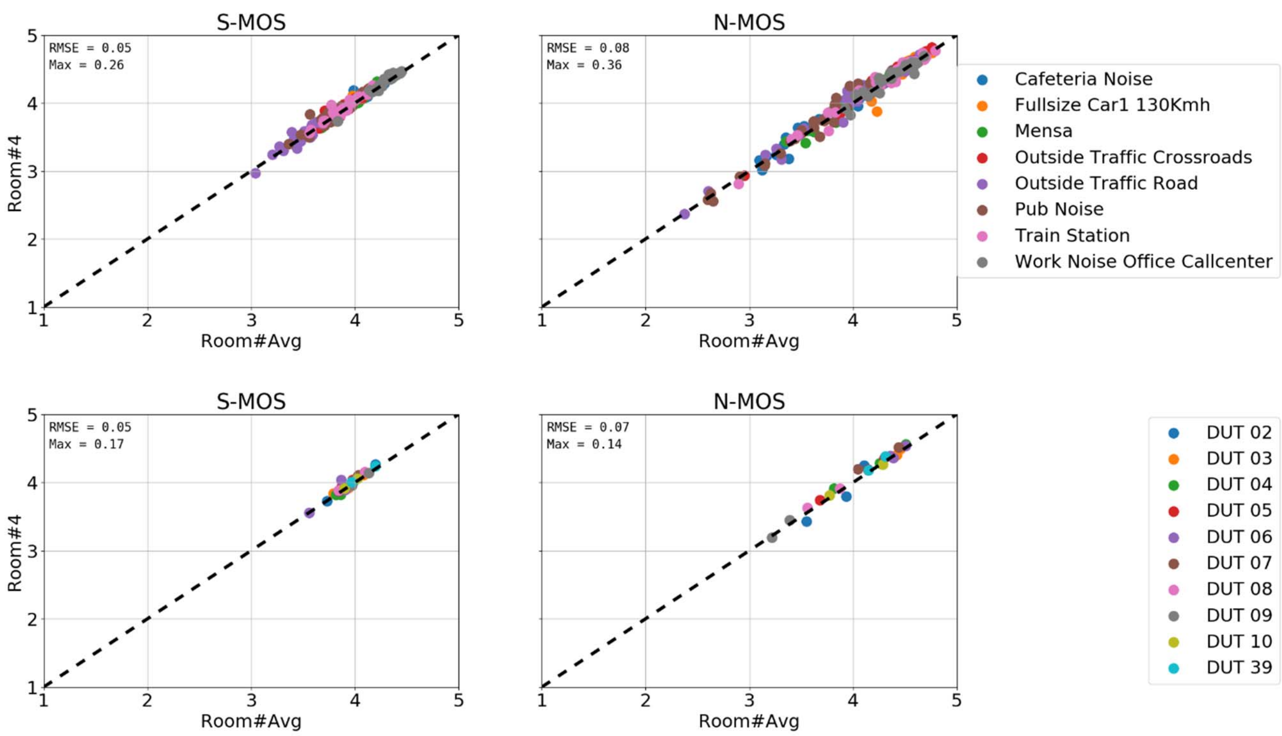


Figure 142: Comparison of Room#4 vs Room#Avg (TS103-4.1)

5.4.10.5 TS 103 224 (binaural), 8.0 setup

Figure 143 to Figure 146 provide the comparison results for TS103-8.0 of rooms #1 to #4 versus the average across all rooms. The upper sub-figures shows the results for each single noise type (one dot per DUT and noise type) and the lower sub-figures show the averaged results across eight noise types (one dot per DUT).

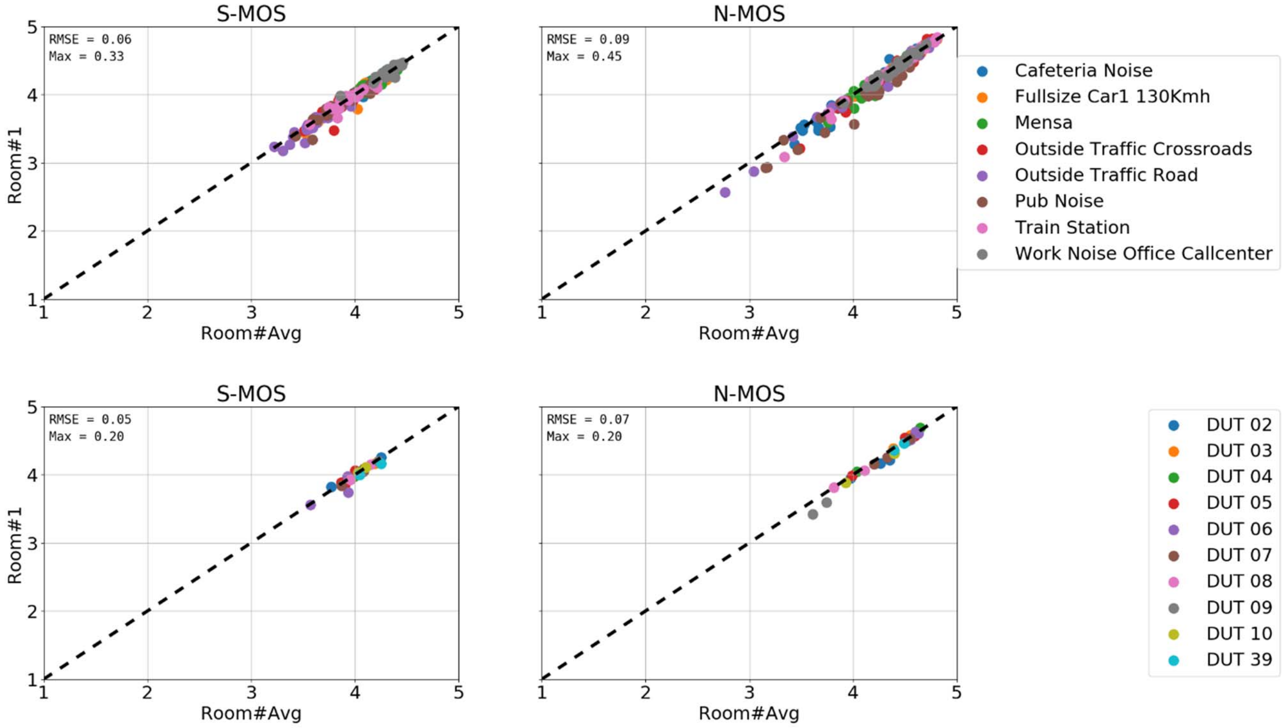


Figure 143: Comparison of Room#1 vs Room#Avg (TS103-8.0)

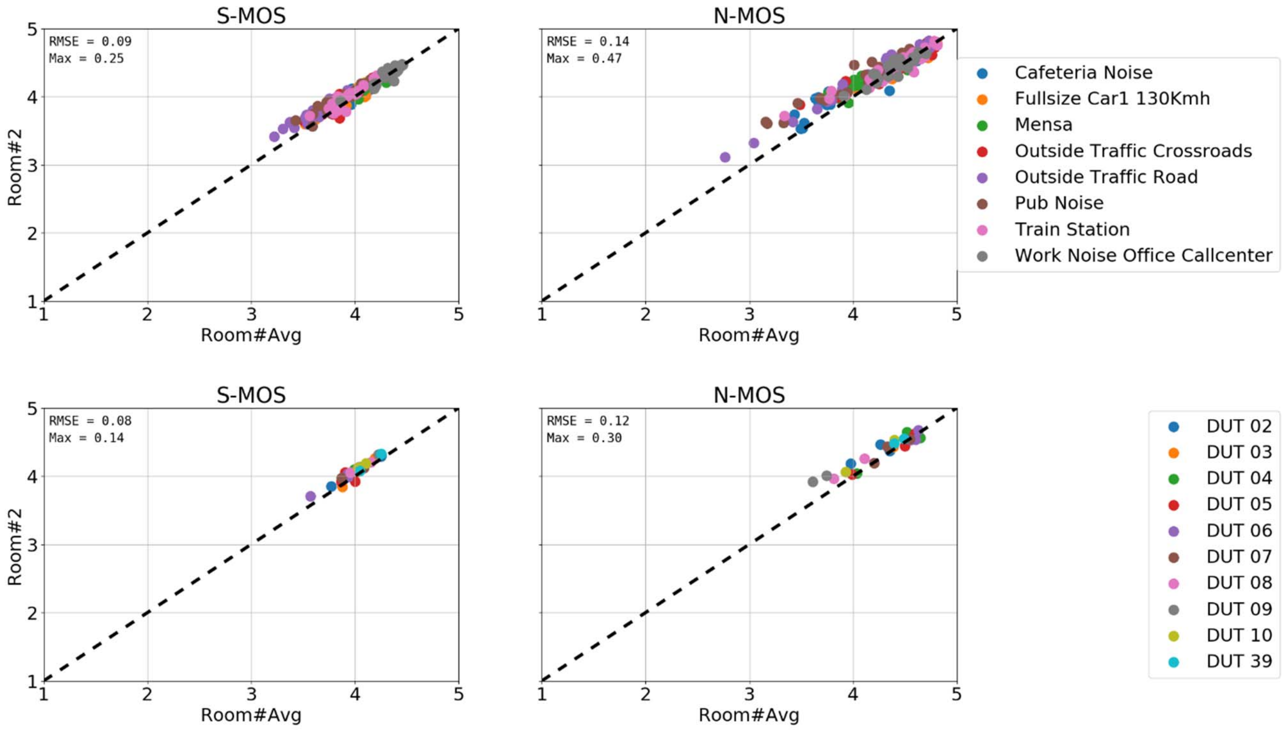


Figure 144: Comparison of Room#2 vs Room#Avg (TS103-8.0)

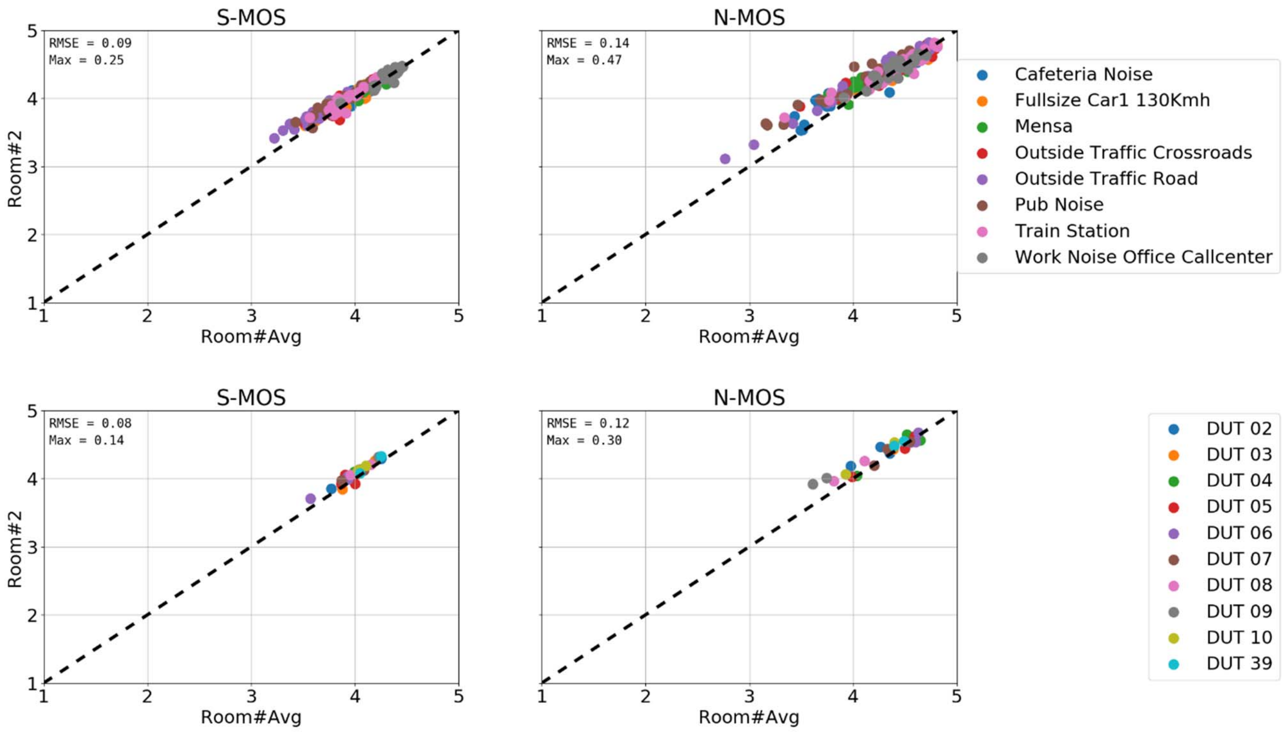


Figure 145: Comparison of Room#3 vs Room#Avg (TS103-8.0)

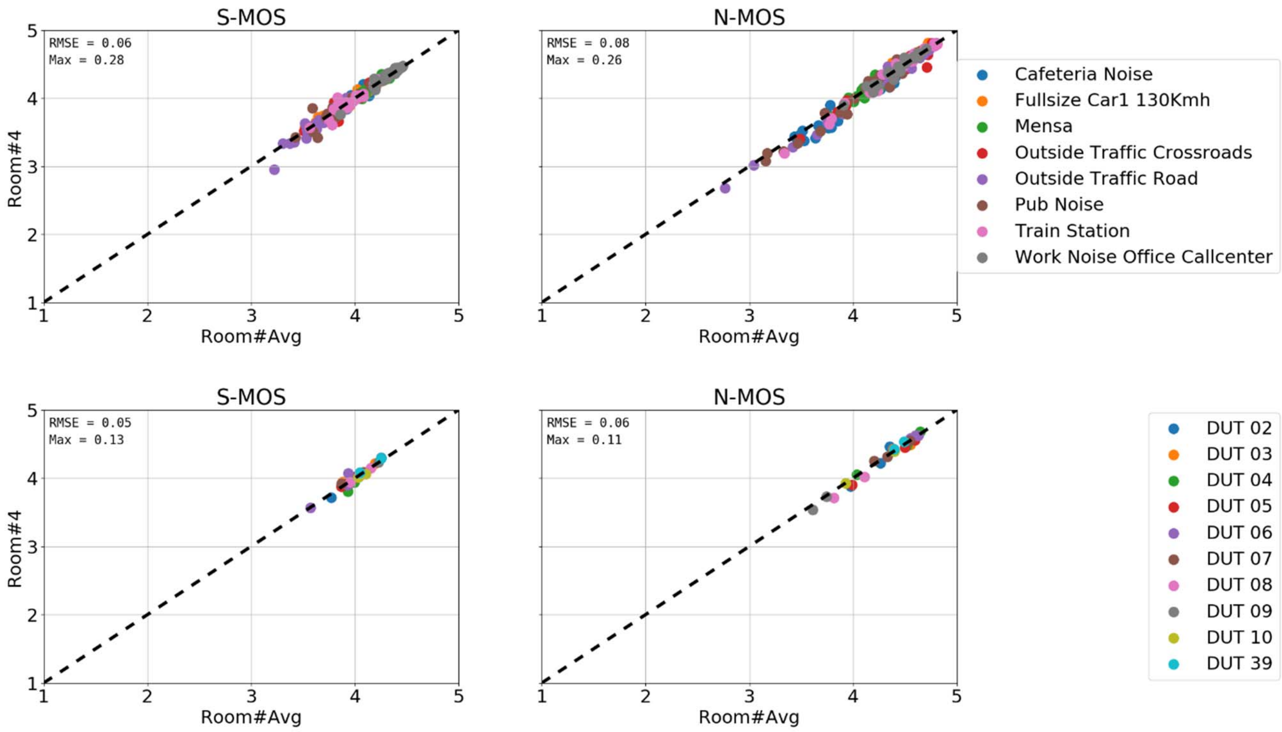


Figure 146: Comparison of Room#4 vs Room#Avg (TS103-8.0)

5.4.10.6 TS 103 224 (8-channel)

Figure 147 to Figure 150 provide the comparison results for TS103-HS of rooms #1 to #4 versus the average across all rooms. The upper sub-figures shows the results for each single noise type (one dot per DUT and noise type) and the lower sub-figures show the averaged results across eight noise types (one dot per DUT).

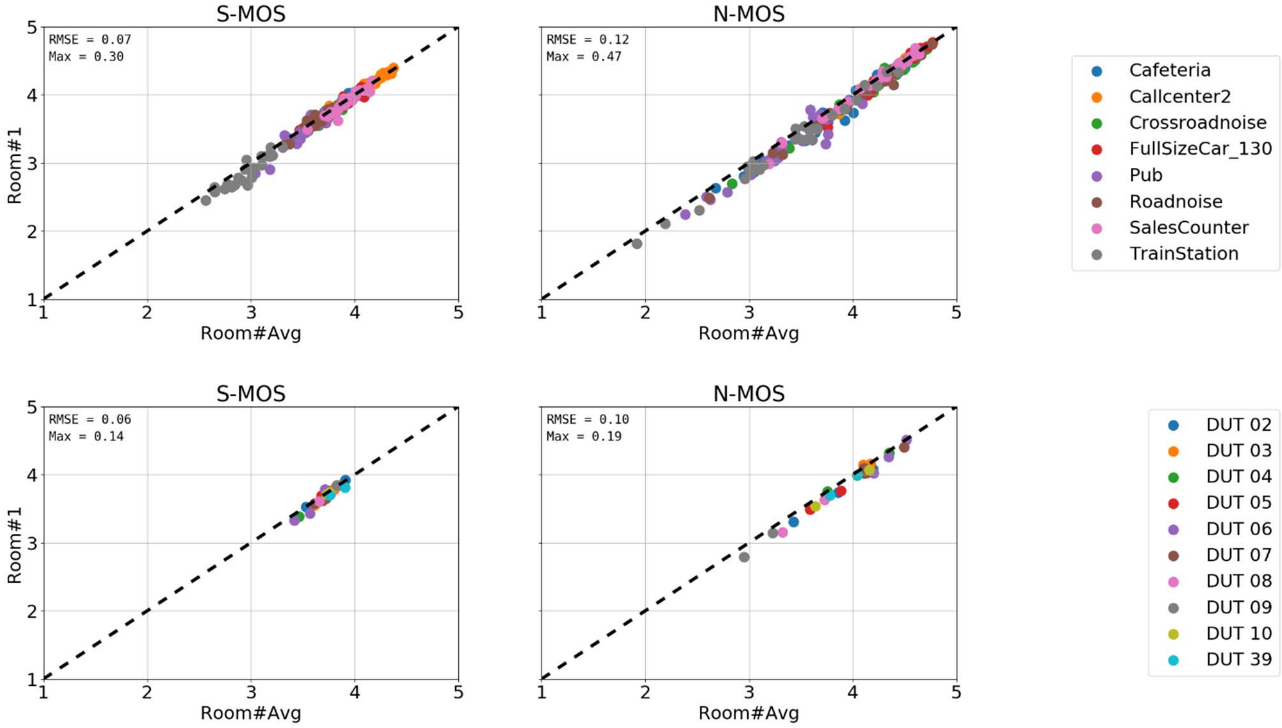


Figure 147: Comparison of Room#1 vs Room#Avg (TS103-HS)

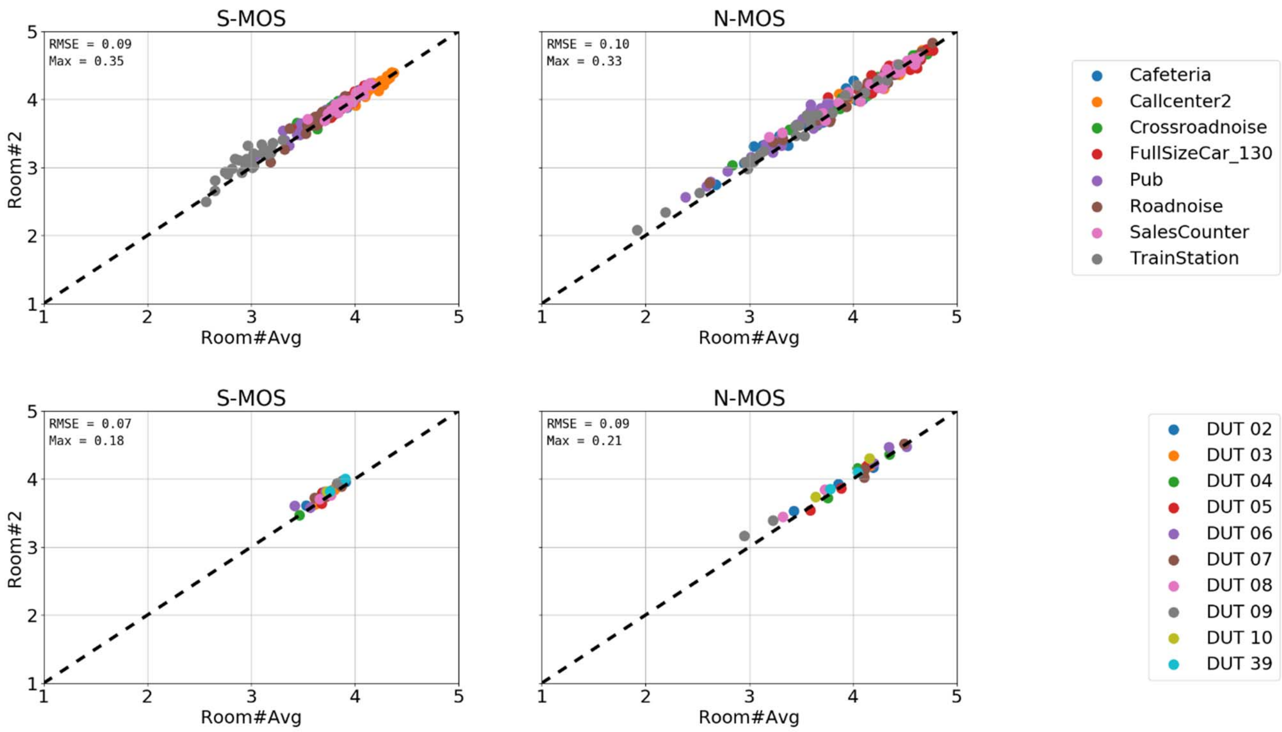


Figure 148: Comparison of Room#2 vs Room#Avg (TS103-HS)

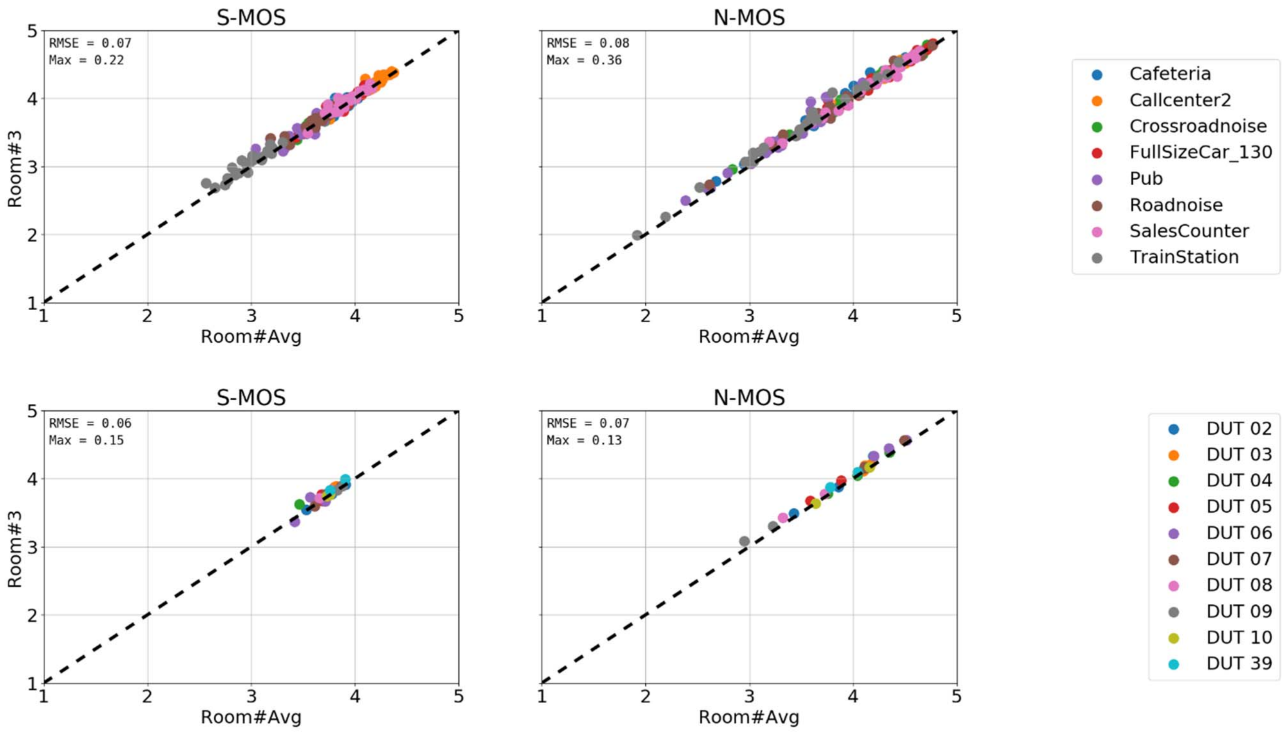


Figure 149: Comparison of Room#3 vs Room#Avg (TS103-HS)

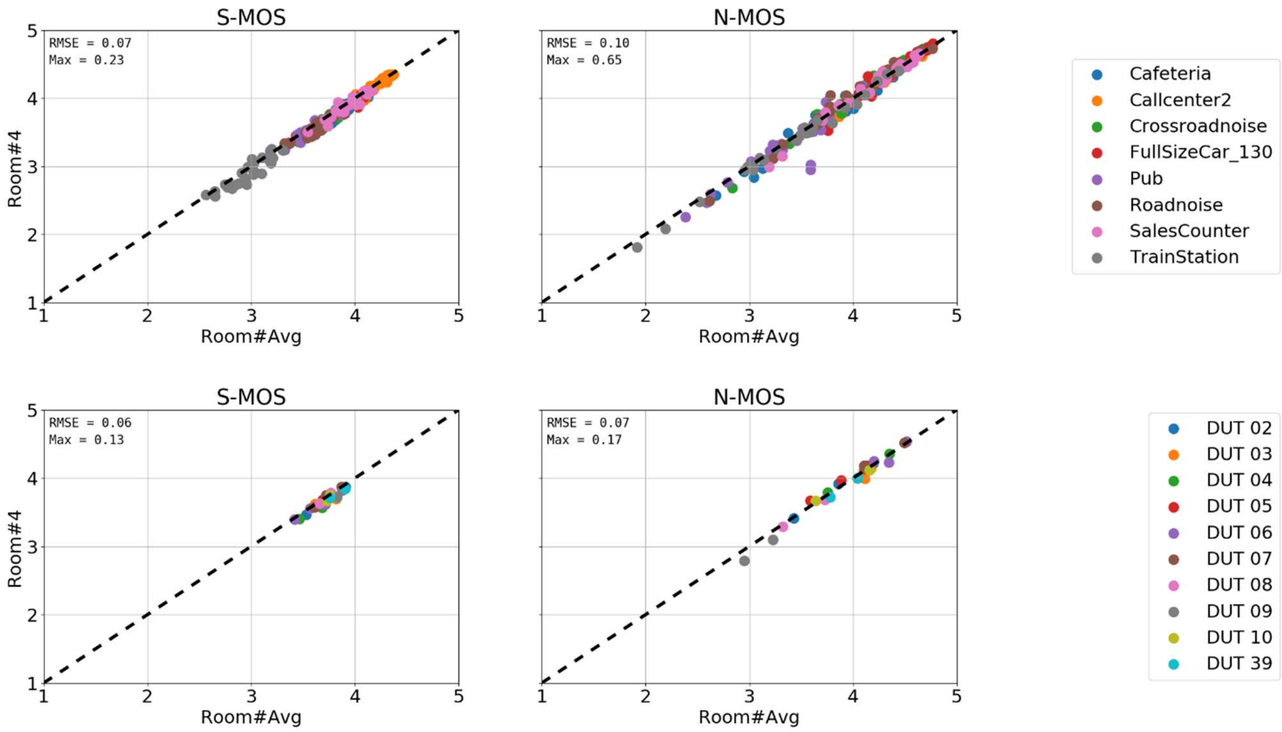


Figure 150: Comparison of Room#4 vs Room#Avg (TS103-HS)

5.4.10.7 Average Variance

In order to quantify the variance of the noise field simulation systems, two performance metrics are taken into account. First, RMSE is calculated for each room comparison shown in the previous clauses 5.4.10.2 to 5.4.10.6. The per-background noise as well as per-device results are calculated separately. In addition, also the absolute maximum deviation (AbsMax) between the rooms is used as a performance metric.

For sake of clarity, the metrics are then aggregated across the four room comparisons: RMSE is averaged and for AbsMax, the maximum is provided. This determines two metrics for S- and N-MOS and for each noise field simulation. Table 28 and Table 29 provide these aggregated result metrics.

Table 28: Average deviation metrics for per-background noise comparison

	RMSE		ABSMAX	
Attribute	**N-MOS**	**S-MOS**	**N-MOS**	**S-MOS**
System				
ES202	0.13	0.07	0.86	0.31
TS103-4.0	0.11	0.07	0.61	0.28
TS103-4.1	0.11	0.07	0.53	0.35
TS103-8.0	0.09	0.07	0.47	0.33
TS103-HS	0.10	0.07	0.65	0.35

Table 29: Average deviation metrics for per-device comparison

	RMSE		ABSMAX	
Attribute	**N-MOS**	**S-MOS**	**N-MOS**	**S-MOS**
System				
ES202	0.10	0.06	0.40	0.15
TS103-4.0	0.08	0.06	0.29	0.19
TS103-4.1	0.09	0.06	0.29	0.19
TS103-8.0	0.07	0.06	0.30	0.20
TS103-HS	0.08	0.06	0.21	0.18

5.4.11 Discussion/Summary

5.4.11.1 Comparisons of TS 103 224 (binaural) noise field simulations to ETSI ES 202 396-1

The round robin test investigated several noise field simulations with real mobile phone devices in multiple labs/rooms. The noise field simulation ES202 and the methodology as currently specified in TS 26.132 [3] serves as a kind of baseline in order to compare per-DUT performance to different simulations.

The binaural equalizations TS103-4.0 and TS103-4.1 provide almost identical results as ES202 across all bandwidths and can be regarded as equivalent. This is not a surprising result, since loudspeaker setups are comparable and equalization procedure of these systems aim to the same target as ES202.

TS103-8.0 seems to provide too optimistic (and thus, non-equivalent) results compared with ES202, TS103-4.0 and TS103-4.1 (see clause 5.4.7). Even though TS103-8.0 provided the most accurate reproduction quality at the artificial ears (see clause 5.3), it seems like that the noise field at the terminal input microphones deviate significantly from ES202, TS103-4.0 and TS103-4.1. However, for all noise field simulations based on binaural recordings, the sound field close to the input microphone of the devices is still not considered at all.

5.4.11.2 Comparison of TS 103 224 (8-channel) noise field simulation to ETSI ES 202 396-1

For TS103-HS, the binaural noise types cannot be used for the noise field simulation and in consequence similar noise scenarios available in TS 103 224 [5] were selected prior to the round robin test. Nevertheless, in general the per-DUT results obtained with TS103-HS are comparable to ES202. Some bandwidth-dependent observations can be made:

- NB: A shift (slope ~1.2) for S-MOS can be observed (see Figure 123), leading to decreased results compared to the corresponding per-DUT results of ES202. For N-MOS, a similar but less pronounced bias of approximately -0.2 MOS can be determined.
- WB: A slight bias for S-MOS (≈ -0.1 MOS) is visible, while results for N-MOS are almost free of shift/bias (see Figure 124).
- SWB: In this case, S- and N-MOS are quite close. Only S-MOS shows a very slight decrease in results compared to the ones of ES202 (see Figure 125).

In general, the selected noise types with TS103-HS provide similar performance results like ES202. Even though some linear shifts can be observed, the rank order remains comparable. A different selection of noise types may lead to even more similar results to ES202 (by e.g., selecting less or different noises).

5.4.11.3 Outlier Analysis

In a dedicated outlier analysis (clause 5.4.9), several issues causing inconsistencies could be found. While in some cases the measurement equipment was not correctly calibrated, also the non-linear and/or time-variant behaviour of the noise reduction provided by the devices may lead to differences across labs. Some example indicated that the noise field simulations TS103-4.0 and TS103-4.1 might help to reduce variance across labs.

5.4.11.4 Variance across rooms/labs

When comparing the variance (clause 5.4.10), the inter-lab consistency is in general adequate, but also shows some larger deviations. As expected, the deviation metrics improve when the per-device results are investigated, which confirms the currently specified average across background noises in TS 26.132 [3].

For the RMSE metric, all system perform in general similar, except that the ES202 method has a slightly higher variance for N-MOS. For the AbsMax metric, S-MOS performs similar for all systems, with somewhat lower performance for ES202 and TS103-8.0. For N-MOS, the largest outlier can be observed for ES202: With AbsMax = 0.86 (N-MOS) for the per-background noise results and AbsMax = 0.40 (N-MOS) for the per-DUT results, the possible differences between labs are rather high.

The best performance metrics in overall are provided by the TS103-8.0 method. However, as mentioned before, this method, depending on bandwidth, shows some slight deviations of the measured terminal performance results compared to the ES202 method (see clause 5.4.7).

The performance metrics of the TS103-HS method are close to the ones of TS103-4.0/-4.1/-8.0. With AbsMax = 0.21 (N-MOS), the lowest maximum difference is achieved for the per-DUT results. Since different (and in some cases, louder) noise types were used, also a larger overall spread in S- and N-MOS scores can be observed (see scatter plots in clause 5.4.10.6), which does not affect the reproduction accuracy across labs.

However, the data collected in the round robin test indicated that eight-channel and flexible configurations of the noise field simulation according to ETSI TS 103 224 [5] may help to reduce the variance across labs (even though in some cases less pronounced).

6 Conclusions

6.1 Conclusions on handheld and desktop hands-free mode

The following conclusions can be drawn from the Round Robin Experiment:

- In general, the inter-lab consistency of the results is good for both noise reproduction systems.
- When comparing the background noises from ES 202 396-1 [4] and TS 103 224 [5], it can be observed that the corresponding noises from both data sets lead to similar measurement results – apart from two cases where the nature of the background noises and in particular the signal levels differ: Cafeteria noise and train station noise. In those cases, the absolute measurement values changed but their rank order did not.
- When using the binaural background noises from TS 103 224 [5] for the testing, the observed differences between the reproduction methods according to ES 202 396-1 and TS 103 224 are rather small. The offset is not constant, it depends on the type of terminal. The performance scores for all terminals are slightly higher when using ES 202 396-1 in hand-held hands-free mode. For desktop hands-free, the variations observed are slightly larger, but not systematic. The inter-lab correlation obtained by TS 103 224 is higher than with ES 202 396-1. For ES 202 396-1, RMSE between 0.13 and 0.25 MOS are observed, while for TS 103 224 the RMSE decreases to a range of 0.07 to 0.2 MOS.
- The differences between the two systems observed in hand-held hands-free mode when testing according to the current version of TS 26.131 [2] and TS 26.132 [3] are quite small (when using the background noises from TS 103 224). In this case, S-, N- and G-MOS are averaged across all evaluated background noises for a per-device comparison. Differences within a range of 0.1 – 0.2 MOS were found, whereas most of the deviation seem to depend on the terminal. The performance for all terminals is slightly improved when using the simulation method according to ES 202 396-1 compared to the one according to TS 103 224 [3].
- When testing in Desktop hands-free mode, the differences according to TS 26.131 and TS 26.132 (averaging S-, N- and G-MOS across all background noises per device) slightly increase compared to hand-held hands-free mode. However, it should be noted that only two labs with three different rooms were participating in this experiment with a rather limited amount of data.
- The accuracy and consistency of the background noise reproduction at the terminal is improved especially in the low frequency domain below 2 kHz when using TS 103 224. Spectral differences measured at the reference microphone drop from 5-15 dB per 1/3rd octave for the ES 202 396-1 simulation method to 1-5 dB when using the TS 103 224 simulation method.
- It was discussed during the meetings and telephone conferences, if it might be possible to reduce this variability of the ES 202 396-1 sound field reproduction further by defining and validating the delays between the loudspeakers in a different way than it is currently described in ES 202 396-1. Whether this is possible or not and if this leads to less variability was unknown at the time of this study. Any requests regarding these two specifications for noise field simulation needs to be addressed by the responsible committee ETSI TC STQ.

6.2 Conclusions on handset mode

Clause 5 of the present document reports on investigations of different noise field simulation methods for various testing purposes. The feasibility study FS_ANTeM resulted in the following main findings, which are based on the analysis of several round robin experiments:

- Beside the noise field simulation according to ETSI ES 202 396-1 [4] (denoted as ES202), which is currently used for ambient noise testing in TS 26.132 [3], .several variants of the more recent method according to ETSI TS 103 224 [5] were investigated:
  - The eight-channel system (denoted as TS103-HS) was designed in particular for handset testing, with several recording/reproduction microphones close to typical input microphones of mobile phones and an automated procedure for calibration/equalization. However, since different noise recordings than currently specified in TS 26.132 [3] are required for this system, it may be difficult to achieve equivalence to ES202 (e.g., regarding measurement results and performance requirements).
  - In order to re-use the binaural noise signal of ES202 and to keep the advantages (like e.g. the automated equalization procedure) of the TS103-HS methodology, several so-called flexible configurations according to ETSI TS 103 224 [5] were investigated as well. Up to three loudspeaker setups were used per lab (denoted as TS103-4.0, TS103-4.1 and TS103-8.0).
- When analysing purely the sound field reproduction accuracy (at the equalization microphones, without devices involved), the method according to ES202 indicated quite large spectral deviations compared to the equalization target. These depend on the type of noise, the accuracy of the manual equalization procedure and the lab/measurement room.
- The accuracy of the pure sound field reproduction (at the equalization microphones, without devices involved) could be significantly improved by the introduced methods TS103-4.0, TS103-4.1 and TS103-8.0 compared to ES202. As expected, the accuracy improves from lower to higher numbers of loudspeakers, leading to best reproduction for the TS103-8.0 method.
- Even though the eight-channel noise field simulation TS103-HS was not equalized to the artificial ear microphones, it provided only slightly worse results for the pure sound field reproduction than TS103-4.0, TS103-4.1 and TS103-8.0. The automated equalization and calibration procedure of all methods based on ETSI TS 103 224 [5] improve the reproduction accuracy across labs.
- The large inter-lab deviations for ES202 observed in the sound field reproduction is not fully reflected in the results of the round robin test conducted with mobile phones in different bandwidth modes. A possible explanation could be that the noise reduction built into the mobile devices are not as sensitive as expected to differences in spectral magnitude and/or diffusivity of the noise field. Even though in some cases the absolute variance is rather small and might not be crucial regarding e.g., a possible fail of the test; in many other cases, the ES202 method provided higher variance across labs compared to the other simulations.
- When comparing S- and N-MOS obtained in the round robin test by the ES202 simulation (real devices, different bandwidth modes) to the methods TS103-4.0 and TS103-4.1, equivalent per-device results (averaged across noise types) are obtained. In some isolated cases, some outliers caused by the ES202 method could even be identified and were not present with TS103-4.0/-4.1.
- Even though the method TS103-8.0 provides the best sound field reproduction, results for S- and N-MOS are too optimistic compared to ES202 (and to TS103-4.0/-4.1 as well). A possible explanation for this could be that the sound field is quite accurate at the DRP, but also seems to cause a quite different sound field at the input microphones of the handset device under test.
- Analysis of S- and N-MOS with the eight-channel noise field simulation TS103-HS obtained similar, but not equivalent results. The per-DUT performance shows a shift to ES202, but the rank order is in general maintained. This may be explained by the different noise types used for the TS103-HS system. Different types or numbers of noises could lead to results closer to the ones obtained by the currently used ES202 method.
- One conclusion that can be drawn from the above is that for the handset mode, it would be possible to allow ambient noise testing using TS103-4.0 and/or TS103-4.1 methods as defined in clause 5.3.2.3 (clause 7 of ETSI TS 103 224 [5]). Identical noise files and identical performance requirements (regarding S-MOS/N-MOS, as specified in TS 26.131 [2]) as for the currently specified method according to ETSI ES 202 396-1 [4] can be used.

Annex A: Collection of reports of round robin test in HHHF mode

A.1 Introduction

The results of the round robin test described in clause 4.2 were presented more in detail across several temporary meeting documents. This annex lists these reports as well as the corresponding test plan and work item description. The documents are provided as an electronic attachment.

A.2 List of documents

SP-140740: "New work item on Acoustic Test Methods and Performance Objectives for Speakerphone Performance in Noisy Environments (ATeMPO_SPINE)".

S4-AHQ099: "Proposed test plan for a Round Robin Test for comparison of background noise simulations – Rev. 1".

S4-151040: "ATeMPO_SPINE round-robin tests conducted at Sony".

S4-151363: "Report on ATeMPO_SPINE round-robin tests conducted at Orange".

S4-151354: "ATeMPO-SPINE round-robin tests conducted at Knowles".

S4-151343: "Report on ATeMPO_SPINE round-robin tests conducted at HEAD acoustics".

S4-151365: "Results of the Round Robin Test on Different Background Noise Simulation Techniques for Hand-Held Hands-Free Terminals".

Annex B: Detailed results for modified ambient noise playback

B.1 Introduction

As described in clause 5.2.5, all result curves from all combinations of noise field configurations and background noises are provided in this Annex.

B.2 Configuration A

Figure B.1 shows the referenced spectra according to clause 5.2.5 for configuration A (ES 202 396-1 with 4.1 setup).

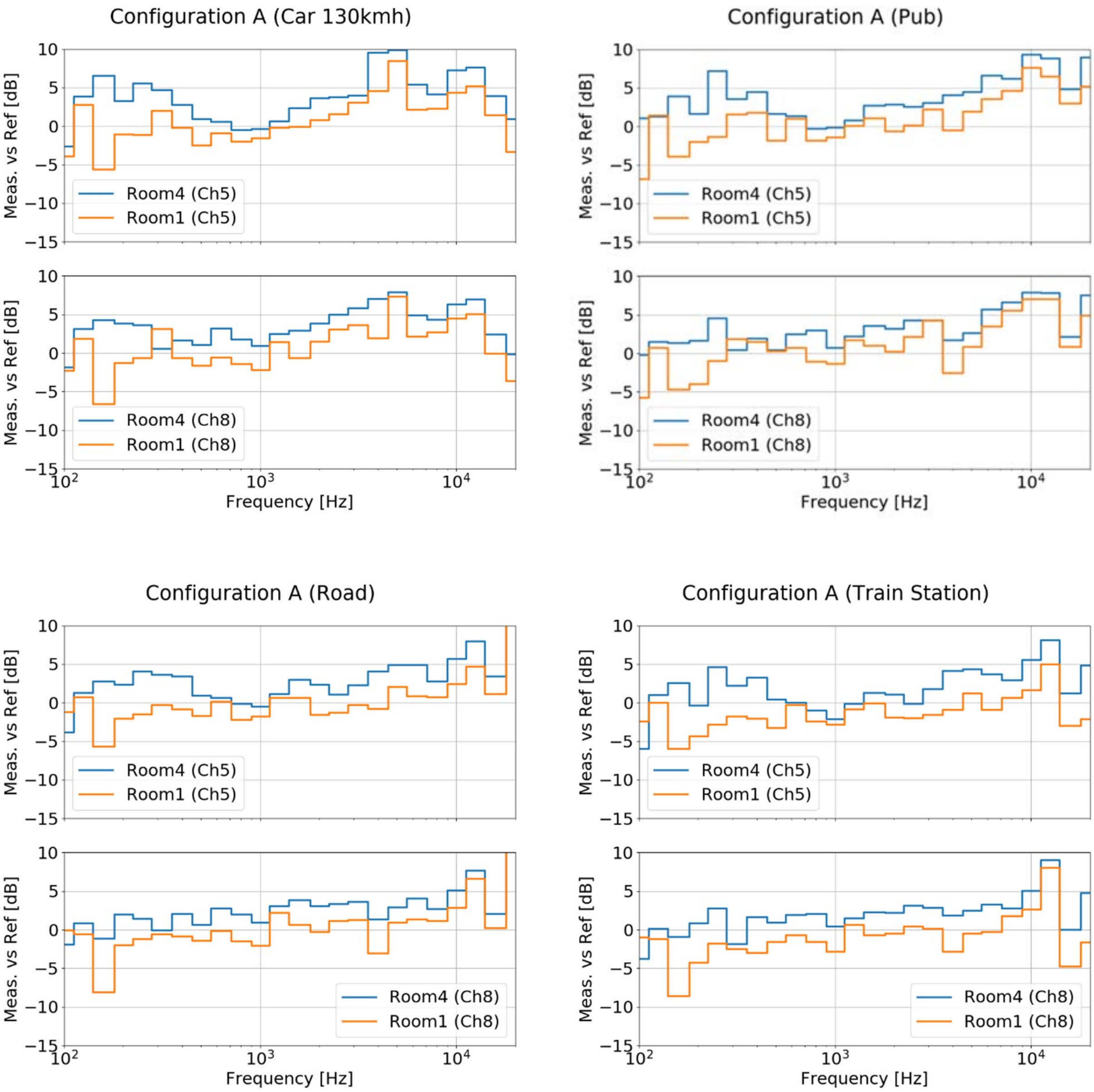


Figure B.1: Referenced spectra for configuration A, all noises

B.3 Configuration B

Figure B.2 shows the referenced spectra according to clause 5.2.5 for configuration B (TS 103 224 with 4.0 setup).

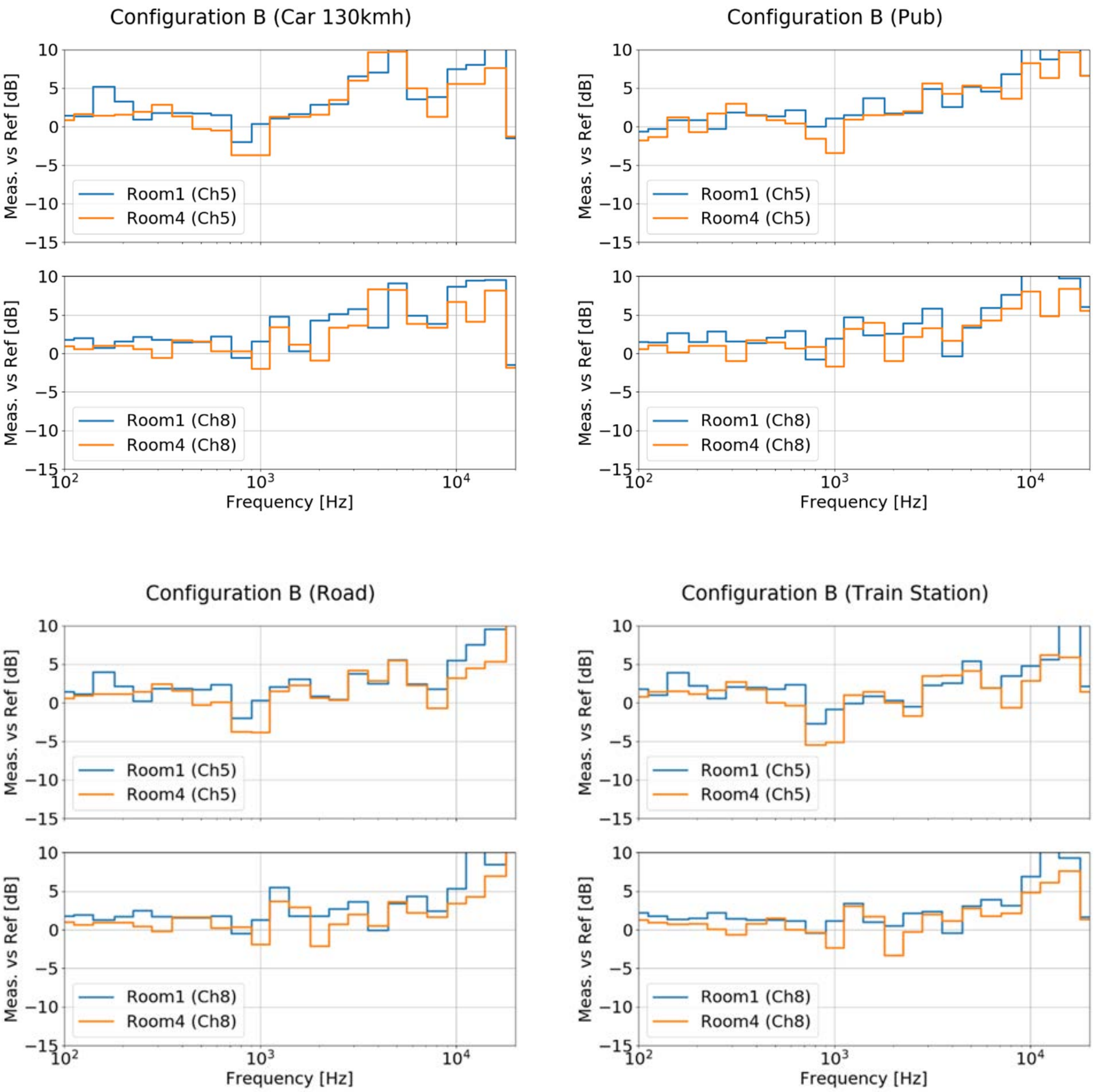


Figure B.2: Referenced spectra for configuration B, all noises

B.4 Configuration C

Figure B.3 shows the referenced spectra according to clause 5.2.5 for configuration C (TS 103 224 with 4.1 setup).

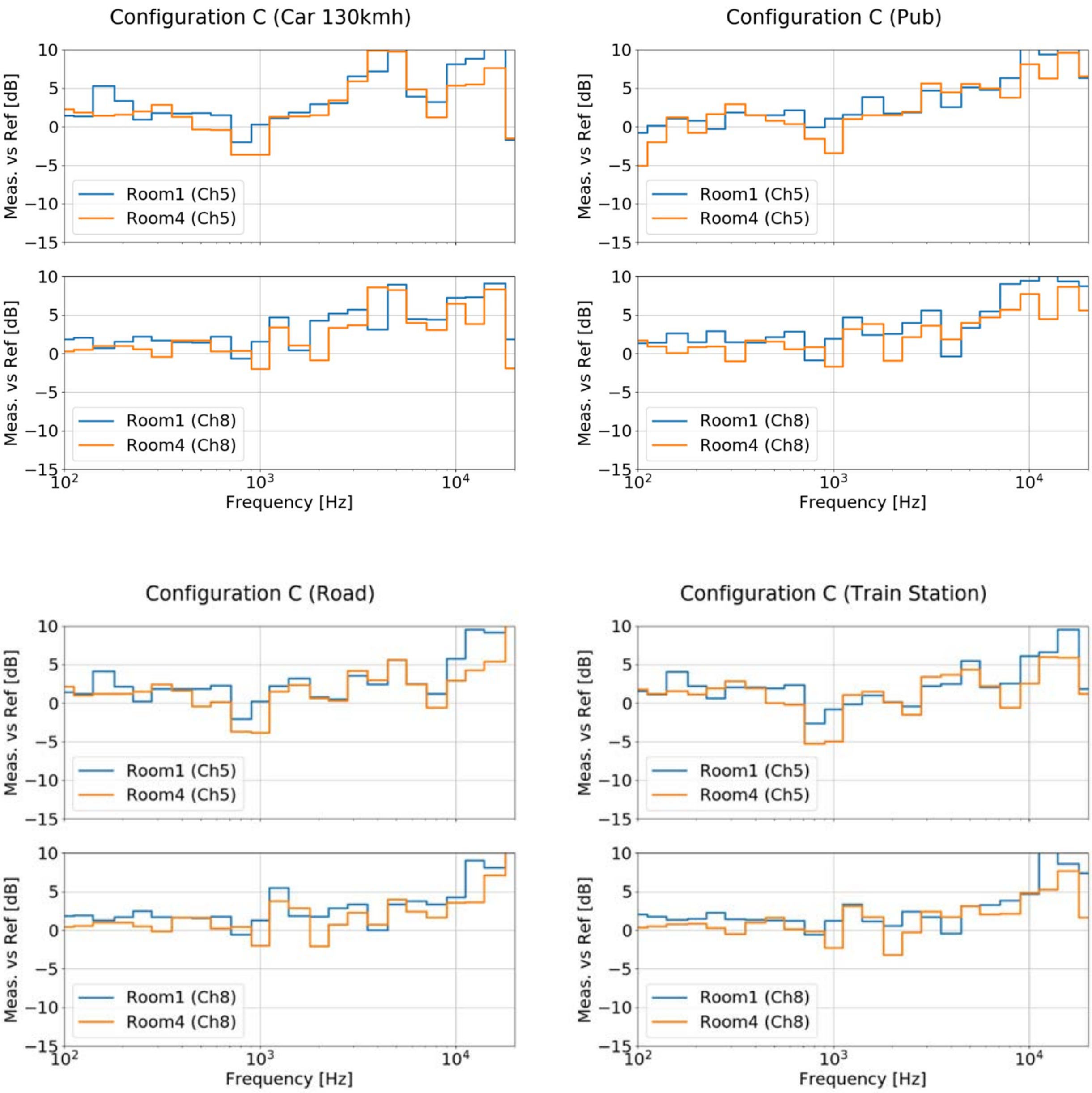


Figure B.3: Referenced spectra for configuration C, all noises

B.5 Configuration D

Figure B.4 shows the referenced spectra according to clause 5.2.5 for configuration D (TS 103 224 with 8.0 setup).

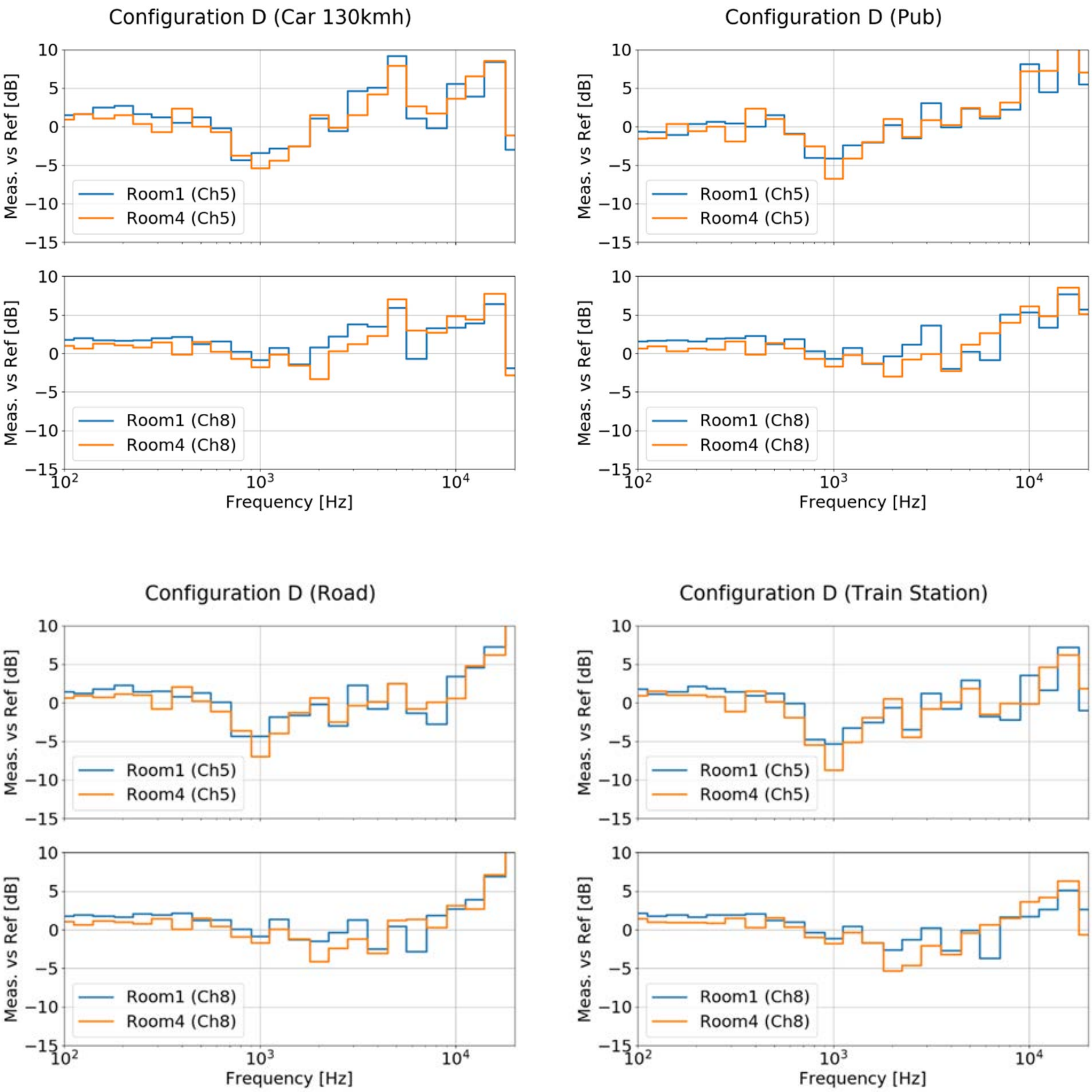


Figure B.4: Referenced spectra for configuration D, all noises

Annex C: Test plan for a Round-Robin-Test - Comparison of noise field simulations for handset mode

C.1 Introduction

Questions to answer within the round robin test:

- How good is the reproducibility of the ETSI ES 202 396-1 and ETSI TS 103 224 background noise (BGN) simulation methods across different labs in handset mode?
- Is the flexible method according to TS 103 224 able to provide more consistent results across labs due to the automated equalization procedure – while using the same noise types as for ES 202 396-1?

C.2 Noise Simulation Setup

At least three noise field simulation systems shall be investigated for each considered measurement room:

- ES 202 396-1 (as currently specified by 3GPP TS 26.132 clause 7.12.1, 8.12.1, 9.12.1)
- TS 103 224 (8-channel setup)
- TS 103 224 (flexible setup using the two HATS microphones as equalization points and 4.1 setup). In this case, the ES 202-396-1 noises are used.
- (Optional) TS 103 224 (flexible setup using the two HATS microphones as equalization points and 8.0 setup). In this case the ES 202-396-1 noises are used.
- (Optional) TS 103 224 (flexible setup using the two HATS microphones as equalization points and 4.0 setup). In this case the ES 202-396-1 noises are used.

C.3 Noise Types

Eight noises per noise field simulation:

- binaural recordings from ES 202 396-1 (as currently specified by 3GPP TS 26.132 in clauses 7.12.1, 8.12.1, 9.12.1)
- 8-channel recordings from TS 103 224, listed in Table B.1 below
- 1 silent condition (speech only)

NOTE: The silent condition is also used for the calibration procedure of the SWB-prediction model according to TS 103 281. In contrast to the noisy measurements, the playback level of speech shall be -4.7 dBPa (skipping the Lombard gain, as specified in TS 103 281).

Table B.1: Noise types according to TS 103 224 for round-robin-test

Filename	Duration	Level Channel 1 & 7
Pub_handset	30 s	1: 77.2 dB(A) 7: 76.0 dB(A)
Roadnoise_handset	30 s	1: 72.8 dB(A) 7: 73.0 dB(A)
Crossroadnoise_handset	30 s	1: 70.6 dB(A) 7: 71.2 dB(A)
TrainStation_handset	30 s	1: 78.9 dB(A) 7: 78.8 dB(A)
FullSizeCar_130_handset	30 s	1: 68.5 dB(A) 7: 70.8 dB(A)
SalesCounter_handset	30 s	1: 66.6 dB(A) 7: 66.6 dB(A)
Cafeteria_handset	30 s	1: 70.0 dB(A) 7: 70.6 dB(A)
Callcenter2_handset	30 s	1: 60.2 dB(A) 7: 60.2 dB(A)

C.4 Bandwidths & Codec setting

The following bandwidth modes and corresponding codecs shall be used:

1) NB mode: AMR-NB codec at 12.2 kbit/s
2) WB mode: AMR-WB codec at 12.65 kbit/s
3) SWB mode: EVS-SWB codec at 24.4 kbit/s

It was observed that only a limited number of commercially available phones support the EVS-SWB codec in conjunction with test SIM cards. Even though the default bitrate of 24.4 kbit/s is required for TS 26.132, several of the few available devices only operate at 13.2 kbit/s bitrate. Such devices and recordings should not be excluded from the round robin test, as long as the bitrate is reported for each device and all participating labs use the same bitrate. It has to be ensured that the device will otherwise work properly.

For sake of simplification of the measurement setup, the call for NB and WB between phones and radio testers should preferably be established via LTE or WLAN. In case the device cannot connect in this connection mode, circuit switched operation could be used instead – as long as codec and bitrate are consistently configured between these connection modes. All labs shall use the same radio configuration for a given device.

C.5 Devices

The target is to have at least four commercially available devices per bandwidth mode. There may be devices which are tested only in specific bandwidth modes, which lead to more than four devices in sum.

If possible, devices with varying noise reduction performance should be selected (e.g., aggressive and/or very low reduction, older phones). Since it is expected that EVS-SWB codec is available in recent devices of high quality, this demand will most likely not be realized in SWB mode.

The devices according to Table B.2 are currently available for testing.

Table B.2: Devices for round robin test

Device Nbr	Manufacturer	Year	AMR-NB/WB via VoLTE	Test in NB	Test in WB	Test in SWB
1	A	2018	no	yes	yes	yes
2	B	2018	yes	yes	yes	yes
3	C	2015	no	yes	yes	no
4	D	2012	no	yes	yes	no
5	E	2014	no	yes	yes	no
6	F	2018	yes	yes	yes	yes
7	F	2017	yes	yes	yes	yes
8	G	2018	yes	yes	yes	yes
9	H	2018	yes	yes	yes	yes

C.6 Rooms

For each participating lab, at least one measurement room/chamber is investigated. Information to be reported about each room/chamber:

- Room size
- Room acoustics: C80 and RT60 (measurements will be provided as part of the test suite)
- Results of equalizations:
  - TS 103 224 (8 loudspeakers/microphone array): result curve of automated equalization procedure for all 8 microphones (magnitude-only).
  - ES 202 396-1: final equalization curve for left and right ear.
  - TS 103 224 flexible (4.1/4.0/8.0 loudspeaker configuration, left and right ear): result curve of automated equalization procedure for both ears.

Six rooms are going to be tested (four from HEAD acoustics GmbH, two from Intel).

C.7 Test Setup

The following items should be taken into account during testing:

- Equipment, handset positioning, etc. according to TS 26.132 clause 5.1
- Test Sequences & speech level according to TS 26.132 clauses 7.12.1, 8.12.1, 9.12.1
- Test suite to be provided by HEAD acoustics GmbH.
- Additional and useful basic measurements according to TS 26.132 shall be conducted and results be reported:
  - Sending Direction:
    - Delay
    - Frequency Response
    - Loudness Rating

- Receiving Direction:
  - Delay
  - Frequency Response
  - Loudness Rating
  - For network access via LTE/WLAN: compensation of clock skew.
  - Predicted speech quality acc. to ITU-T P.863 V2.4 (send and receive)

C.8 Loudspeaker position

Compliant to requirements of ETSI EG 202 396-1 and ETSI TS 103 224. Loudspeaker and HATS position setup shall be reported by participating labs.

C.9 Analysis

Speech quality test method according to TS 26.131/132 shall be used:

- NB & WB Mode: ETSI TS 103 106
- SWB Mode: ETSI TS 103 281 (model A)

NOTE: Even though the unprocessed signal (captured by reference microphone) is not used for the SWB/FB prediction model, it shall be recorded in the same way as for NB and WB. Debugging and further analyses of the reproduced sound field are possible with this extra information.

C.10 Outlier

When conducting measurements as automated as possible (see also next section), several errors like e.g. wrong codec, call abort, playback of wrong or non-present noise type, etc. may be discovered only at a later stage. Strictly spoken, such recordings do neither comply with TS 26.131/132 not with this document and thus, must be repeated. If the commonly used terminals and/or lab access are not available anymore, the recordings shall be excluded for further analysis.

For otherwise valid recordings, outliers regarding S- and N-MOS are difficult to identify during testing - especially when no comparison data of another lab is available. If particular recordings or devices should be included or excluded will be discussed in the group.

C.11 Test procedure

In order to reduce variance in test results, the following test procedures are highly recommended:

- All DUTs should be tested sequentially within one measurement room
- All bandwidth modes per DUT should be tested sequentially
- All noise field simulations per DUT and bandwidth mode should be tested sequentially within one call.

Details of steps for one DUT:

1. Mount DUT A
2. Make a call in NB
   a) Run measurements: Delay, SLR, Frequency response
   b) Run Silence recording
   c) Run ES 202 396 BGN tests (8 noises)
   d) Run TS 103 224 BGN tests (8 noises)
   e) Run TS 103 224 with ES 202 396 noises (8 noises) (optional: repeat with 8.0 loudspeaker setup)
3. Make a call in WB
   a) Run Delay, SLR Frequency measurement
   b) Run Silence recording
   c) Run ES 202 396 BGN tests (8 noises)
   d) Run TS 103 224 BGN tests (8 noises)
   e) Run TS 103 224 with ES 202 396 noises (8 noises) (optional: repeat with 8.0 loudspeaker setup)
4. Make a call in SWB
   a) Run Delay, SLR Frequency measurement
   b) Run Silence recording
   c) Run ES 202 396 BGN tests (8 noises)
   d) Run TS 103 224 BGN tests (8 noises)
   e) Run TS 103 224 with ES 202 396 noises (8 noises) (optional: repeat with 8.0 loudspeaker setup)
5. Unmount DUT A
6. Repeat step 1 for each DUT (B, C, D, …)

The order of the steps listed above also helps to reduce overhead of workload by minimizing…

- Changing / (Re-)mounting DUTs to the HATS
- Switching of operational modes and codecs per DUT
- The number of calls between DUT and network simulator

Time estimation for measurements for each bandwidth mode for each DUT:

- Connecting DUT to system simulator, mounting of phone, configuration of measurement system for the current bandwidth mode / DUT (~ 1h / 60min)
- Preparation measurements: at least some basic parameters for sending directions should be measured (frequency response, SLR, delay). In addition, a speech recording in silence shall be conducted (~ 10 minutes).
- Eight noise measurements per bandwidth mode / DUT / noise field simulation. Each measurement is about 90 seconds (~15 minutes), excluding calculation time for the results.
- Calculation time of the analyses is not regarded here, since this can be done offline at a later stage.

Depending on the test equipment available in each lab, an initial setup time of up to 2 days may be needed for each measurement room.

- Check that all the devices can make a VoLTE call with CMW (This step is taking much longer than anticipated)
- Run ES calibration (1/2 day)
- Run TS calibration (1/4 day)
- Run TS calibration using ES noise (1/4 day)

C.12 Example Estimations

Based on the aforementioned estimations, the overall measurement time O is approximately given by:

O = (D * B * 60min) + (D * B) * (10 minutes + N * 15 minutes) * R

With:

D Number of devices

B Number of bandwidth modes

N Number of noise field simulation systems

R Number of evaluated measurement rooms

EXAMPLE 1: A single device (D=1) is tested in NB, WB and SWB mode (B=3) with two noise systems (N=2) in one measurement room (R=1). The overall time O is given by:

O = (D * B * 60min) + (D * B) * (10min + N * 15min) * R
O = (1*3*60) + (1*3) * (10 + 2*15) * 1 = 300min (~ 5h)

For each bandwidth (B=1), the time estimation would be:
O = (1*1*60) + (1*1) * (10 + 2*15) * 1 = 100min (~ 1.66h)

NOTE: It may be possible to significantly reduce testing time from 5h to 3-4h, since setup time per DUT should decrease (so far: 60min per DUT and bandwidth) once it is successfully connected to the radio tester in one bandwidth mode. Here a worst-case estimation is assumed.

EXAMPLE 2: Four devices (D=4) tested in all bandwidths (B=3), all noise described systems (N=4) and in two rooms (R=2) leads to:

O = (D * B * 60min) + (D * B) * (10min + N * 15min) * R
O = (4*3*60) + (4*3) * (10 + 4*15) * 2 = 2400min ~ 40h

Annex D: Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2019-09	SA#85	SP-190643				Presented to TSG SA#85 for information	1.0.0
2020-03	SA#87-e	SP-200050				Presented to TSG SA#87-e for approval	2.0.0
2020-03	SA#87-e	SP-200050				Approved by TSG SA#87-e	16.0.0
2022-04	-	-	-	-	-	Update to Rel-17 version (MCC)	**17.0.0**

History

Document history		
V17.0.0	May 2022	Publication