ETSI TS 103 106 V1.1.1 (2012-08)

Technical Specification

Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods

Reference
DTS/STQ-201

Keywords
noise, quality, speech, testing, transmission

ETSI

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

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

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

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

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

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

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

Foreword

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

The present document is to be used in conjunction with the ETSI standard series EG 202 396 [i.2] to [i.4]:

Part 1: "Background noise simulation technique and background noise database";

Part 2: "Background noise transmission - Network simulation - Subjective test database and results";

Part 3: "Background noise transmission - Objective test methods".

The present document is based on the objective test method described in EG 202 396-3 [i.4] and contains modifications of the model required in order to provide a good prediction of the uplink speech quality in the presence of background noise of modern mobile terminals.

1 Scope

The present document describes testing methodologies which can be used to objectively evaluate the performance of narrowband and wideband mobile terminals for speech communication in the presence of background noise.

Background noise is a problem in mostly all situations and conditions and needs to be taken into account in both, terminals and networks. The present document provides information about the testing methods applicable to objectively evaluate the speech quality of mobile terminals with AMR and AMR-WB codecs in the presence of background noise. The present document includes:

- The method which is applicable to objectively determine the different parameters influencing the speech quality in the presence of background noise taking into account:
  - the speech quality;
  - the background noise transmission quality;
  - the overall quality.
- The description of the adaptation of the test method described in EG 202 396-1 [i.2].
- The model results in comparison with the underlying subjective tests used for the retraining of the objective model.
- The model validation results.

The present document is to be used in conjunction with:

- EG 202 396-1 [i.2] which describes a recording and reproduction setup for realistic simulation of background noise scenarios in lab-type environments for the performance evaluation of terminals and communication systems.
- EG 202 396-2 [i.3] which describes the simulation of network impairments and how to simulate realistic transmission network scenarios and which contains the methodology and results of the subjective scoring for the data forming the basis of the present document.
- EG 202 396-3 [i.4] which describes the basic objective model underlying to the Model described in the present document.
- American English speech sentences as enclosed in the present document.

2 References

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

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

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

2.1 Normative references

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

Not applicable.

2.2 Informative references

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

[i.1] 3GPP S4-120542: "Common subjective testing framework for training of P.835 test predictors".

[i.2] ETSI EG 202 396-1: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[i.3] ETSI EG 202 396-2: "Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise; Part 2: Background Noise Transmission - Network Simulation - Subjective Test Database and Results".

[i.4] ETSI EG 202 396-3: "Speech and multimedia Transmission Quality (STQ); Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods".

[i.5] ETSI TS 126 073: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; ANSI C code for the Adaptive Multi Rate (AMR) speech codec (3GPP TS 26.073)".

[i.6] ITU-T Recommendation P.835: "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm".

[i.7] ITU-T Recommendation G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[i.8] ITU-T Recommendation P.56: "Objective measurement of active speech level".

[i.9] ITU-T Recommendation P.1401: "Methods, metrics and procedures for statistical evaluation, qualifying and comparison of objective quality prediction models".

[i.10] ITU-T Recommendation G.160 Appendix II, Amendment 2: "Voice enhancement devices: Revised Appendix II - Objective measures for the characterization of the basic functioning of noise reduction algorithms".

[i.11] ITU-T Recommendation G.191: "Software tools for speech and audio coding standardization".

[i.12] Hastie, T.; Tibshirani, R.; Friedman, J.: "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", New York: Springer-Verlag, 2001.

[i.13] ITU-T Recommendation P.501: "Test Signals for Use in Telephonometry".

3 Abbreviations

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

AMR Adaptive MultiRate
AMR-WB Adaptive Multi-Rate Wideband Speech Codec
BAK Background Noise Component
dB SPL Sound Pressure Level re 20 µPa in dB
G-MOS Global MOS

NOTE: MOS related to the overall sample.

HHHF Hand-Held Hands-Free
IRS Intermediate Reference System
ITU International Telecommunication Union
ITU-T Telecommunication Standardization Sector of ITU
MOS Mean Opinion Score
MRP Mouth Reference Point

MSIN	Mobile Station Input Filter
NB	NarrowBand
N-MOS	Noise MOS

NOTE: MOS related to the noise transmission only.

NS	Noise Suppression
OVRL	Overall (speech + noise) Component
RCV	ReCeiVe
RMSE	Root Mean Square Error
RMSE*	epsilon insensitive Root Mean Square Error
SIG	SIGnal component
S-MOS	Speech MOS

NOTE: MOS related to the speech signal only.

SND	Sending Direction
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
WB	WideBand

4 Introduction

The present document describes the modifications of the EG 202 396-3 [i.4] model which were necessary to adapt to the training databases provided by the 3GPP contributors listed in Annex A. The core model itself retains mainly unmodified except the points given in the clauses below. Modifications affect the narrow- and wideband mode in different ways.

The adapted objective method described in the present document is intended to be used for all types of modern mobile terminals using different bitrates of AMR [i.5] and AMR-WB [i.7] coding.

5 Underlying speech databases and preparations

The base for each mode of the objective model (wideband/narrowband) as described in EG 202 396-3 [i.4] are listening test conducted according to ITU-T Recommendation P.835 [i.6]. From the beginning of the development, these listening test databases were designed to be a training set for predicting ITU-T Recommendation P.835 [i.6] scores. They included a huge amount of conditions (> 170) and a wide range of speech and noise quality. Besides real terminals also terminal simulations and transmission impairments were included. However, the data and processing included were based on technologies actual at the time when the standard and its updates were created.

The underlying databases for the retraining as described in the present document were created using real state-of-the-art mobile devices and thus the quality ranges yielded may not be normally distributed over all MOS scales. The context between the databases can also differ (e.g. pure handset recordings vs. mixed handset/hands-free databases). Furthermore new reference conditions extensively discussed in different standards groups and described in [i.1] were included in the tests.

Table 1: Set of reference conditions

File	SIG.	SNR	Noise Type
i01	Source (filtered)	No Noise	-
i02	Source (filtered)	0 dB	Fullsize_Car1_130Kmh_binaural
i03	Source (filtered)	12 dB	Fullsize_Car1_130Kmh_binaural
i04	Source (filtered)	24 dB	Fullsize_Car1_130Kmh_binaural
i05	Source (filtered)	36 dB	Fullsize_Car1_130Kmh_binaural
i06	NS Level 1	No Noise	-
i07	NS Level 2	No Noise	-
i08	NS Level 3	No Noise	-
i09	NS Level 4	No Noise	-
i10	NS Level 3	24 dB	Fullsize_Car1_130Kmh_binaural
i11	NS Level 2	12 dB	Fullsize_Car1_130Kmh_binaural
i12	NS Level 1	[0 dB]	Fullsize_Car1_130Kmh_binaural

Each training database was provided together with 12 reference conditions, mainly created according to the annex of [i.1], table 1 shows one possible arrangement. Although it was observed that not all reference sets included exactly the same speech material, used background noise, SNR ranges and speech distortion configuration, this data indicates which range of speech and noise degradations can be expected in the databases.

For transforming the different databases (to achieve at least approximately on a common base for the retraining of the model), thus the 12 x 3 values of the reference conditions (averaged over all samples) were used to linearly transform the subjective MOS data. In a first step, the reference conditions of all databases included in the retraining process were weighted together to an average reference condition set. The weight per database depends on the number of samples it provides for the training.

For each database, a mapping between the reference conditions and the average reference condition set is calculated. To catch also inter-relations between speech, noise and global ratings, a matrix transformation instead a per-scale regression was chosen. To compensate biases, a constant column was added to the reference set. Then a transformation Tj is calculated for each database j with reference set Rj which minimizes the distance to the average reference set A:

$$\underbrace{\begin{pmatrix} 1 & S_{i01} & N_{i01} & G_{i12} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & S_{i12} & N_{i12} & G_{i12} \end{pmatrix}}_{R_j (Ref.\ set\ j)} \times T_j = \underbrace{\begin{pmatrix} \overline{S_{i01}} & \overline{N_{i01}} & \overline{G_{i12}} \\ \vdots & \vdots & \vdots \\ \overline{S_{i12}} & \overline{N_{i12}} & \overline{G_{i12}} \end{pmatrix}}_{A\ (Avg.\ ref.\ set)} \quad (1)$$

The transformation matrix Tj (size 4 x 3) can easily be determined to:

$$T_j = \left(R_j^{\,T} \times R_j\right)^{-1} \times R_j^{\,T} \times A \quad (2)$$

If the three scales (S-MOS/N-MOS/G-MOS) are independent from each other for any database, the matrix transformation T_j equals a linear per-scale transformation. Before the retraining of the model, the transformation is applied to the whole test data on a per-sample base:

$$\underbrace{\begin{pmatrix} 1 & S_1 & N_1 & G_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & S_N & N_N & G_N \end{pmatrix}}_{S_j\ (scores\ of\ samples\ of\ database\ j)} \times T_j = \underbrace{\begin{pmatrix} \tilde{S}_1 & \tilde{N}_1 & \tilde{G}_1 \\ \vdots & \vdots & \vdots \\ \tilde{S}_N & \tilde{N}_N & \tilde{G}_N \end{pmatrix}}_{\tilde{S}_j\ (transformed\ scores\ of\ samples\ of\ database\ j)} \quad (3)$$

6 Modifications to the model described in EG 202 396-3

6.1 Prefiltering in Narrowband Mode (NB)

In the narrowband mode described in EG 202 396-3 [i.4], the listening test audio files included a far-end handset simulation, realized with an IRS RCV filter. In the requirements described in [i.4], neither for narrow- nor for wideband such a listening filter was described or used in the databases.

The narrowband mode internally filters the unprocessed and clean reference with IRS SND and IRS RCV to simulate a transmission over high-quality listening devices and network. The principle of IRS seems to be outdated, modern state-of-the-art mobiles do not have this frequency characteristic. Even more when using these newly created NB databases, where the used devices have almost flat frequency responses in sending direction.

Thus the filtering with IRS SND and RCV of the two reference signals was replaced by filtering with the MSIN [i.11] filter, which is mainly a band pass. Also no listening filter was applied to the processed signals.

6.2 Detection of the speech parts

The detection of signal parts belonging to either speech or noise was updated. Now the clean speech signal is segmented into frames and classified according to ITU-T Recommendation G.160 [i.10]. The signal parts classified as silence are assumed as background noise sections, all other frames are assumed as speech.

6.3 Speech level adjustment in wideband

The current EG 202 396-3 [i.4] implementation assumes 79 dB SPL / -15 dB Pa active speech level due to the underlying listening test based on the underlying subjective databases in the wideband model of EG 202 396-3 [i.4].

For the objective model as described in the present document the level adjustment of the recordings of the training databases was applied in such a way, that the active speech level over the full sequence test should be about 73 dB SPL / -21 dB Pa (for the listening test) as described in [i.4].

6.4 Replacement of parameter regression for S-MOS

The model described in EG 202 396-3 [i.4] calculates several parameters out of the psycho-acoustically motivated inner representation for the estimation of S- and N-MOS. The parameters are shown in tables 2 and 3. A detailed description of the calculation for the parameters can be found in [i.4].

Table 2: Extracted parameters for N-MOS

P_1	$N_{BGN,P}$	P_4	$\mu(RA_{BGN,U})$
P_2	$\mu(RA_{BGN,P})$	P_5	$\sigma^2(RA_{BGN,U})$
P_3	$\sigma^2(RA_{BGN,P})$	P_6	$\sigma^2(\Delta RA_{BGN,P-U})$

Table 3: Extracted Parameters for S-MOS

P_1	ΔSNR	P_4	$\mu(\Delta RA_{Sp,\,P\text{-}C})$
P_2	$\mu(RA_{Sp,\,P})$	P_5	$\sigma^2(\Delta RA_{Sp,\,P\text{-}C})$
P_3	$\mu(\Delta RA_{Sp,\,P\text{-}U})$	P_6	$\sigma^2(\Delta RA_{Sp,\,P\text{-}U})$

The calculation of the objective S-MOS in clause 6.5.2 of [i.4] is performed with a linear quadratic regression of the parameters mentioned above. In addition, the regression coefficients are switched with regard to the N-MOS calculated before which models the expectation to speech [i.4] quality of the listener.

The applied modification is the replacement of the linear quadratic regression with a feed forward neural network. In consequence, the switching of the regression coefficients depending on the N-MOS is removed. Only one network is trained with input (6 parameters of table 3) and output (S-MOS) data by a simple back-propagation algorithm [i.12].

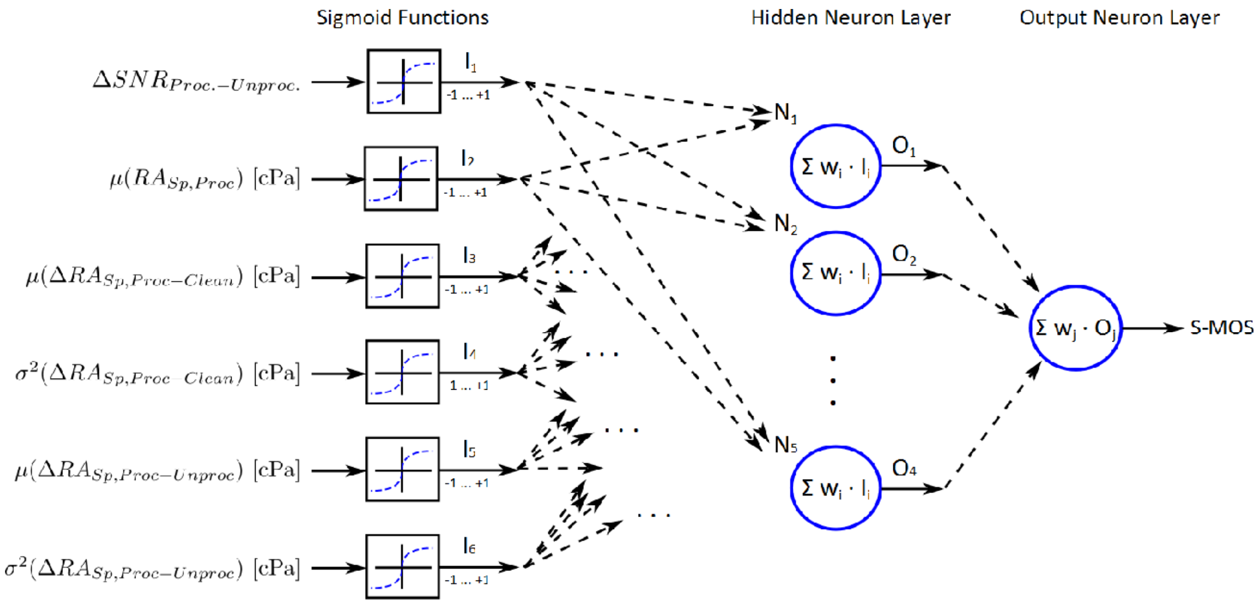


Figure 1: Structure of neural network for S-MOS

The setup of the neural network is shown in figure 1. It consists of 5 units in one hidden layer; each unit N_j includes a connection from each transformed input parameter I_i. The output O_j of each unit is calculated as the weighted sum of each input I_i using the weights w_{ij}. The outputs O_j are then weighted by w_j and summed up to the output S-MOS. Both, w_{ij} and w_j are the result of the training of the network.

The parameters according to table 3 are composed to a vector **P** including a bias as the first element:

$$\mathbf{P} = (1 \quad P_1 \quad P_2 \quad P_3 \quad P_4 \quad P_5 \quad P_6) \tag{4}$$

The output calculation of the neural network shown in figure 1 can be described as concatenated matrix operations:

$$SMOS_{objective,raw} = f_{sigmoid}\left(f_{sigmoid}\left(\frac{\mathbf{P}-\mathbf{M_{in}}}{\mathbf{S_{in}}}\right)\times\mathbf{H}\right)\times\mathbf{O} \tag{5}$$

First the parameter vector **P** is normalized to mean 0,0 and standard deviation 1,0. This is done by subtracting the average of all training data for each parameter from each item of the input parameter vector. The averages for each parameter P_i can be described as a vector, which is different for narrow- and wideband mode:

$$\mathbf{M}_{\mathbf{in},\mathbf{WB}} = (0{,}0 \quad 12{,}7309 \quad 4{,}2076 \quad -1{,}2456 \quad 0{,}8834 \quad 12{,}2522 \quad 7{,}0541)$$

$$\mathbf{M}_{\mathbf{in},\mathbf{NB}} = (0{,}0 \quad 13{,}7519 \quad 2{,}0884 \quad -0{,}3124 \quad 0{,}2511 \quad 6{,}7091 \quad 5{,}2951)$$

(6)

NOTE 1: The first element is set to zero to be compatible with the bias element in **P**.

A similar approach can be made for the standard deviation for each parameter P_i, also separated for wide- and narrowband:

$$\mathbf{S}_{\mathbf{in},\mathbf{WB}} = (1{,}0 \quad 11{,}8503 \quad 1{,}2824 \quad 1{,}1981 \quad 0{,}9572 \quad 6{,}7848 \quad 4{,}8380)$$

$$\mathbf{S}_{\mathbf{in},\mathbf{NB}} = (1{,}0 \quad 11{,}4341 \quad 0{,}4047 \quad 0{,}3877 \quad 0{,}3309 \quad 3{,}1189 \quad 2{,}5976)$$

(7)

NOTE 2: The first element is set to one to be compatible with the bias element in **P**.

After normalizing the input data, the sigmoid function $f_{sigmoid}(x)$ is applied to the each normalized parameter P_i. This ensures that each input of each neuron of the hidden layer is soft-limited to the range ±1,0 and guarantees that parameters out of the training range cannot produce an overflow which results in eventually unreasonable scores. For the current model, the hyperbolic tangent was chosen to a sigmoid function:

$$f_{sigmoid}(x) = \tanh(x)$$

(8)

Thus the input of the hidden neuron layers can also be given as a transformed parameter vector $\widetilde{\mathbf{P}}$:

$$\widetilde{\mathbf{P}} = f_{sigmoid}\left(\frac{\mathbf{P} - \mathbf{M}_{\mathbf{in}}}{\mathbf{S}_{\mathbf{in}}}\right) = (1 \quad \widetilde{P_1} \quad \widetilde{P_2} \quad \widetilde{P_3} \quad \widetilde{P_4} \quad \widetilde{P_5} \quad \widetilde{P_6})$$

(9)

NOTE 3: The sigmoid function is not applied to the bias component.

The output of the hidden layer is calculated with a matrix multiplication of $\widetilde{\mathbf{P}}$ and **H**. **H** describes all weights from each input parameter to each neuron in the hidden layer. These weights are the results of the training with the back-propagation algorithm. In consequence, **H** is different for each bandwidth mode:

$$\mathbf{H}_{\mathbf{WB}} = \begin{pmatrix} -0{,}4336 & -0{,}9873 & 0{,}0091 & -0{,}0845 & 0{,}0203 \\ 0{,}1141 & -0{,}0004 & -0{,}7133 & -0{,}2798 & -1{,}8189 \\ 1{,}0265 & 0{,}5001 & 0{,}5120 & 0{,}0537 & 0{,}1265 \\ -0{,}8627 & -1{,}7518 & -0{,}0374 & -0{,}2908 & 0{,}3064 \\ 2{,}1381 & 0{,}4190 & 1{,}0715 & -1{,}6716 & 0{,}4973 \\ -1{,}3933 & 0{,}5972 & 0{,}0852 & 0{,}1977 & 0{,}2222 \\ -0{,}3793 & -1{,}7785 & -0{,}5306 & -1{,}7538 & -2{,}9630 \end{pmatrix}$$

(10)

$$\mathbf{H}_{\mathbf{NB}} = \begin{pmatrix} -0{,}3608 & -0{,}3805 & 0{,}5359 & -1{,}1131 & -0{,}1322 \\ 0{,}7348 & -4{,}4639 & -1{,}2552 & 0{,}3338 & 0{,}5452 \\ 0{,}9117 & 2{,}7177 & 0{,}8876 & 0{,}1712 & -2{,}1279 \\ -0{,}2383 & 1{,}7228 & -0{,}0354 & -1{,}0284 & 1{,}0483 \\ 1{,}4511 & 2{,}1467 & 1{,}0010 & 0{,}7356 & 0{,}1154 \\ -0{,}5573 & -0{,}6137 & -0{,}2648 & 1{,}6202 & 0{,}5966 \\ -3{,}2194 & -7{,}9575 & -0{,}7736 & -0{,}8676 & 0{,}1663 \end{pmatrix}$$

(11)

The outputs of the hidden layer are then again soft-limited with the same sigmoid function to assure a valid range (±1,0) for the output neuron layer. The five transformed output values of the hidden layer are then given to the output layer. Here the output of the neural network is calculated with another matrix multiplication with the matrix **O**, which weights the outputs of the hidden layers to an output score $SMOS_{objective,\ raw}$. This output layer matrix **O** is also given for wide- and narrowband mode independently:

$$\mathbf{O_{WB}} = (0{,}1777 \quad -0{,}2835 \quad -0{,}3147 \quad 0{,}1837 \quad -0{,}3237)$$

$$\mathbf{O_{NB}} = (0{,}3832 \quad -0{,}5250 \quad -0{,}1878 \quad -0{,}2674 \quad -0{,}1548)$$

(12)

Another part of the back-propagation algorithm is also to normalize the output data to mean 0,0 and standard deviation 1,0. To revise this step and transform the output of the neural network back to the MOS scale, the objective S-MOS is calculated from the raw score:

$$SMOS_{objective} = \max\left(1{,}0, \min\left(\mathbf{S_{out}} \cdot (SMOS_{objective,raw} + \mathbf{M_{out}})\ ,5{,}0\right)\right)$$

(13)

The objective S-MOS is calculated with M_{out} = (3,0), S_{out} = (2,0) and a hard limiter [1,0; 5,0].

6.5 Retraining of parameter regression for N-MOS and G-MOS

The objective N-MOS is the result of a linear, quadratic regression algorithm applied to the six parameters of table 2 according to equation (14):

$$NMOS = c_0 + \sum_{j=1}^{2}\sum_{i=1}^{6} c_{ji} \cdot P_i^j \quad (1)$$

(14)

The overall or global quality G-MOS is calculated by using the previously calculated N-MOS and S-MOS as input parameters for a linear quadratic regression according to equation (15):

$$GMOS = c_0 + \sum_{j=1}^{2} c_{Sj} \cdot SMOS^j + \sum_{j=1}^{2} c_{Nj} \cdot NMOS^j \quad (1)$$

(15)

The calculation steps for N-MOS and G-MOS are not modified, only the coefficients for the linear regressions according to equations (14) and (15) are adapted to the new training material. The new coefficients are given in tables 4 to 7:

Table 4: N-MOS coefficients for narrowband; Parameters P_i according to table 2

	Bias	P_1	P_2	P_3	P_4	P_5	P_6
Order j = 1	2,2231	-0,0395	-0,0359	0,2825	0,0023	-0,3959	-2,6965
Order j = 2	-	-	0,0021	-0,0239	-0,0003	0,0542	0,8684

Table 5: N-MOS coefficients for wideband; Parameters P_i according to table 2

	Bias	P_1	P_2	P_3	P_4	P_5	P_6
Order j = 1	1,4279	-0,0484	0,0994	0,2189	-0,0732	-0,3346	-1,3108
Order j = 2	-	-	-0,0018	-0,0079	0,0011	0,0891	0,2566

Table 6: G-MOS coefficients for narrowband

	Bias	S-MOS	N-MOS
Order j = 1	-0,4879	0,2647	0,8274
Order j = 2	-	0,0696	-0,0737

Table 7: G-MOS coefficients for wideband

	Bias	S-MOS	N-MOS
Order j = 1	-0,2141	0,2735	0,4542
Order j = 2	-	0,0708	-0,0065

7 Comparison of objective and subjective results after the training process

The comparison between the results of the subjective tests and the objective prediction of the conditions used in the training process are given in this clause. The metrics used in the statistical evaluation process are derived from ITU-T Recommendation P.1401 [i.9]. Besides the RMSE or RMSE* values, the difference metrics and scatterplots are given in this clause.

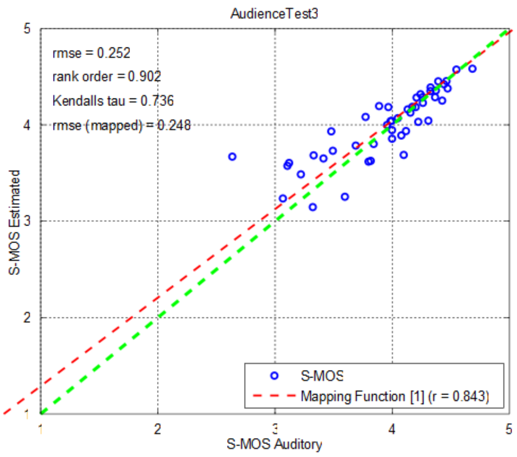
A summary of the databases and the conditions used for retraining is given in Annex A.

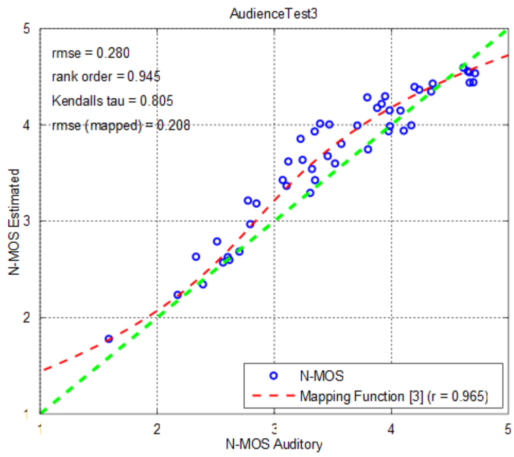
7.1 Results in wideband mode

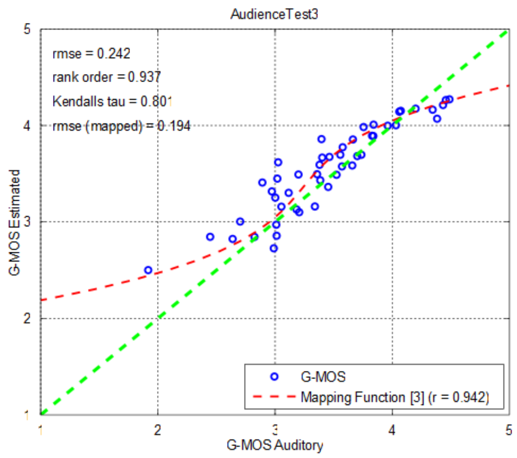
For the wideband retraining procedure two databases were not included within the training for several reasons. Removal of these databases significantly increases the performance. Further analysis is required why these databases seem to be "incompatible" with the remaining training set.

In overall, 7 databases with 387 conditions and 5 544 samples were used.

7.1.1 Results for database "Audience - Test 3"



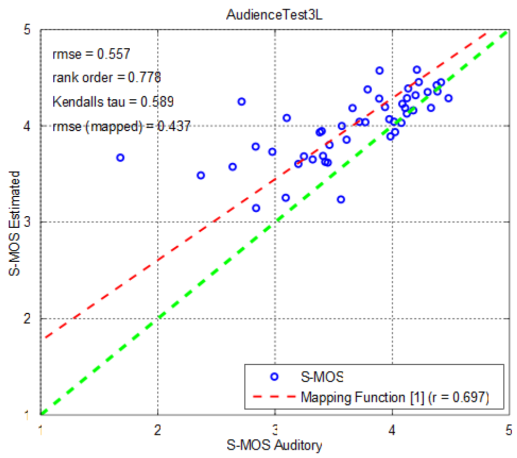


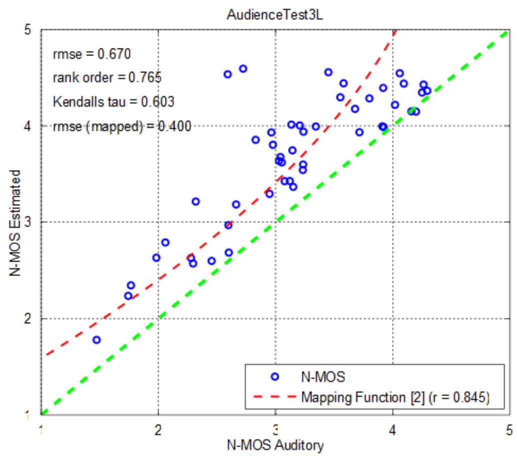


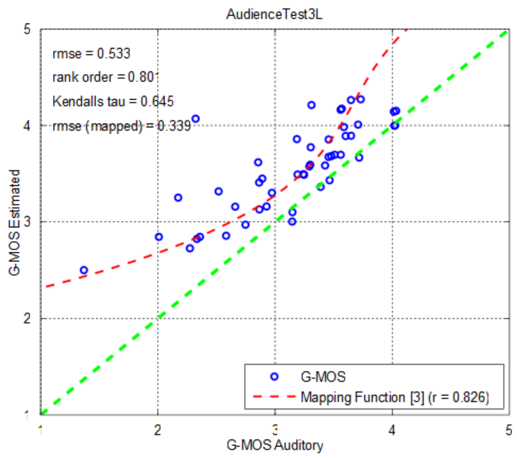
RMSE:	no Mapping	0,25	0,28	0,24
	1st Ord. Mapping	0,25	0,23	0,22
	3rd Ord. Mapping	**0,24**	**0,21**	**0,19**

RMSE*:	no Mapping	0,17	0,18	0,15
	1st Ord. Mapping	0,16	0,13	0,12
	3rd Ord. Mapping	**0,15**	**0,11**	**0,10**

7.1.2 Results for database "Audience - Test 3L" (excluded during retraining)



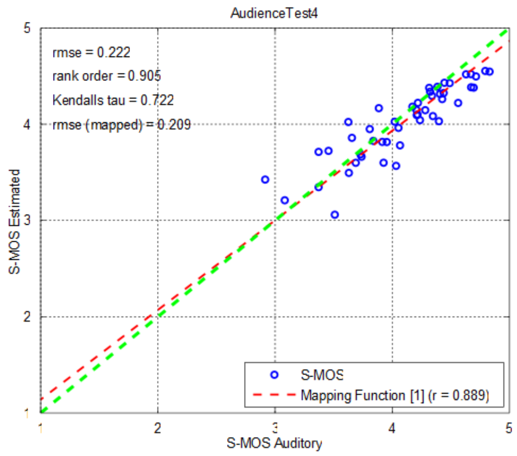


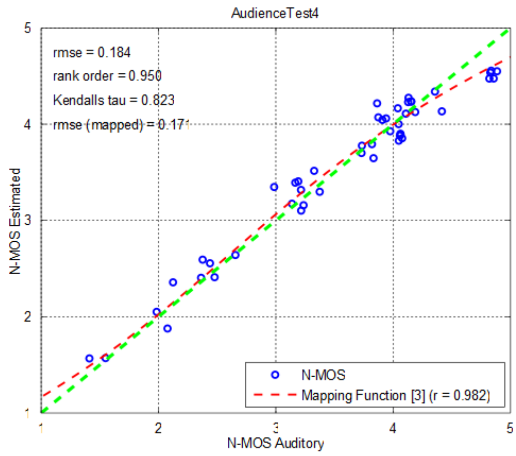


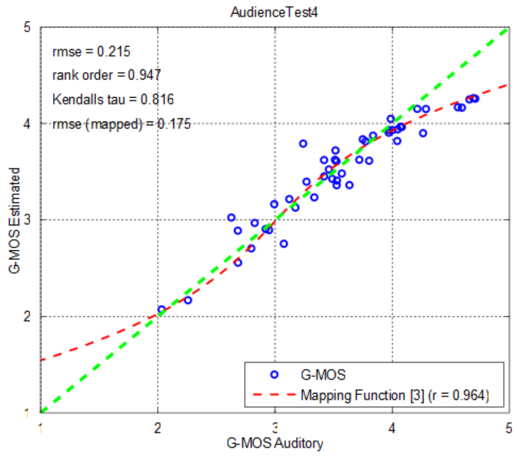
		SIG	BAK	OVRL
RMSE:	no Mapping	0,56	0,67	0,53
	1st Ord. Mapping	0,44	0,40	0,36
	3rd Ord. Mapping	**0,42**	**0,39**	**0,34**

RMSE*:	no Mapping	0,45	0,56	0,43
	1st Ord. Mapping	0,34	0,30	0,26
	3rd Ord. Mapping	**0,31**	**0,28**	**0,25**

7.1.3 Results for database "Audience - Test 4"



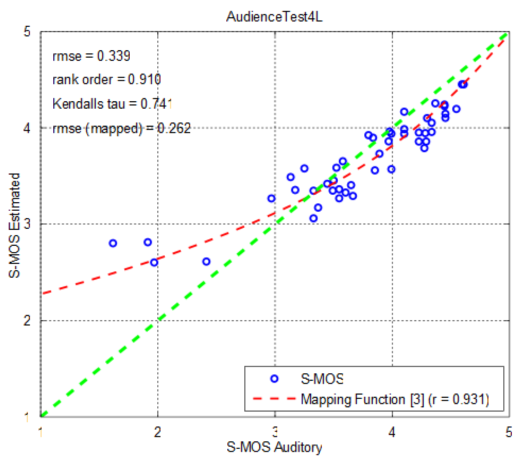


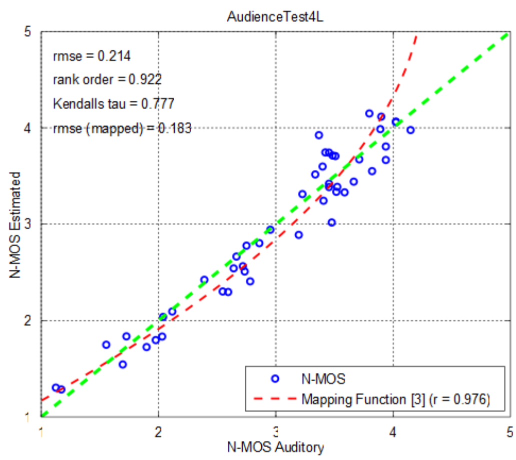


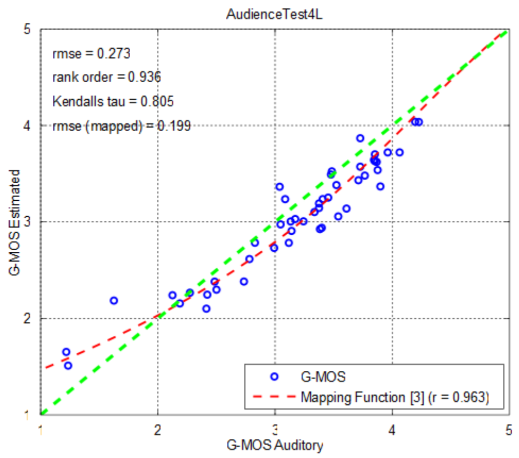
		SIG	BAK	OVRL
RMSE:	no Mapping	0,22	0,18	0,21
	1st Ord. Mapping	0,21	0,18	0,20
	3rd Ord. Mapping	**0,21**	**0,17**	**0,18**

RMSE*:	no Mapping	0,14	0,11	0,14
	1st Ord. Mapping	0,12	0,10	0,12
	3rd Ord. Mapping	**0,12**	**0,08**	**0,10**

7.1.4 Results for database "Audience - Test 4L"



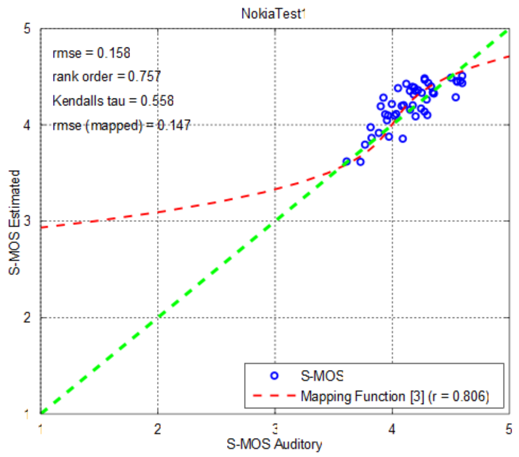


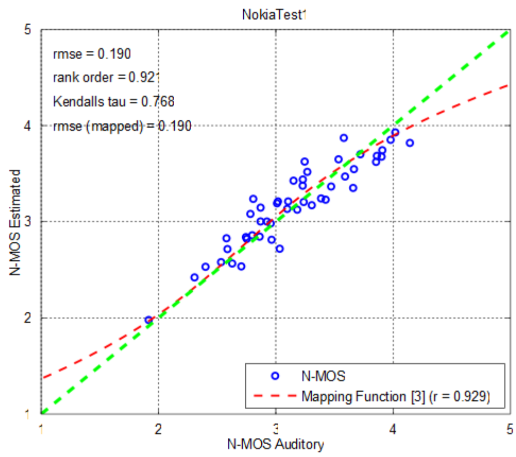


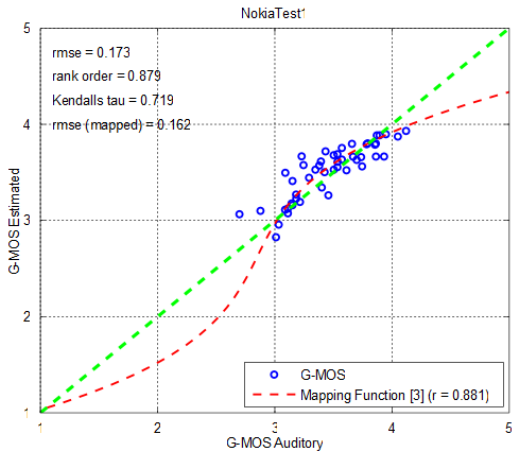
		SIG	BAK	OVRL
RMSE:	no Mapping	0,34	0,21	0,27
	1st Ord. Mapping	0,28	0,21	0,22
	3rd Ord. Mapping	**0,26**	**0,18**	**0,20**

RMSE*:	no Mapping	0,23	0,11	0,17
	1st Ord. Mapping	0,17	0,11	0,14
	3rd Ord. Mapping	**0,15**	**0,08**	**0,12**

7.1.5 Results for database "Nokia - Test 1"



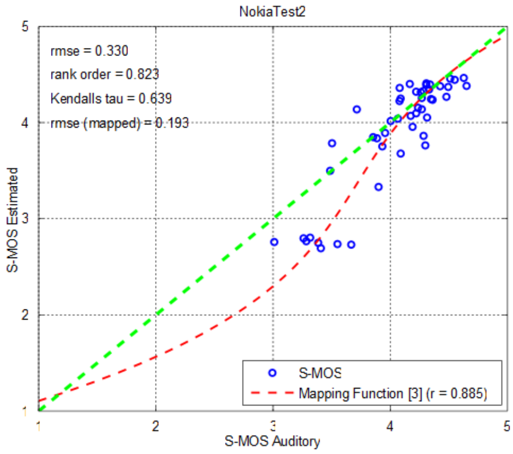


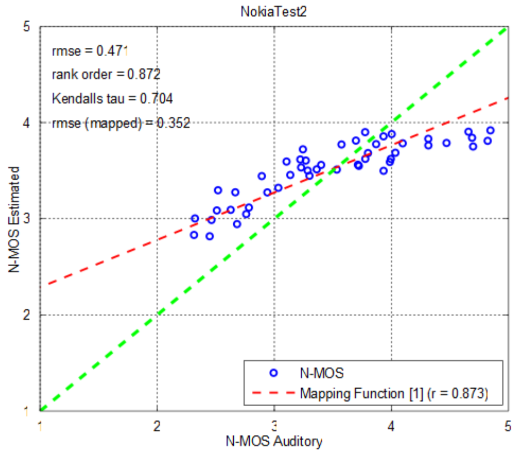


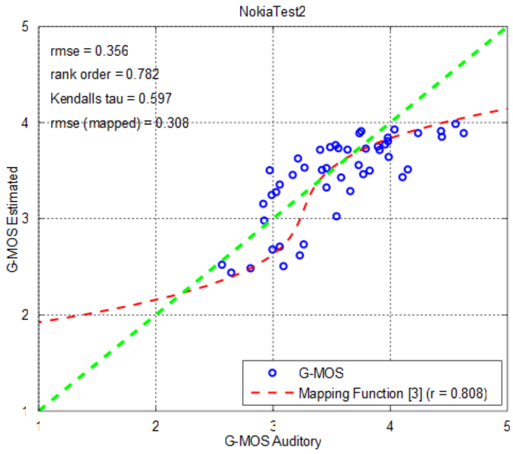
		SIG	BAK	OVRL
RMSE:	no Mapping	0,16	0,19	0,17
	1st Ord. Mapping	0,16	0,19	0,18
	3rd Ord. Mapping	**0,16**	**0,20**	**0,18**

RMSE*:	no Mapping	0,06	0,08	0,09
	1st Ord. Mapping	0,07	0,08	0,09
	3rd Ord. Mapping	**0,07**	**0,09**	**0,09**

7.1.6 Results for database "Nokia - Test 2" (excluded during retraining)



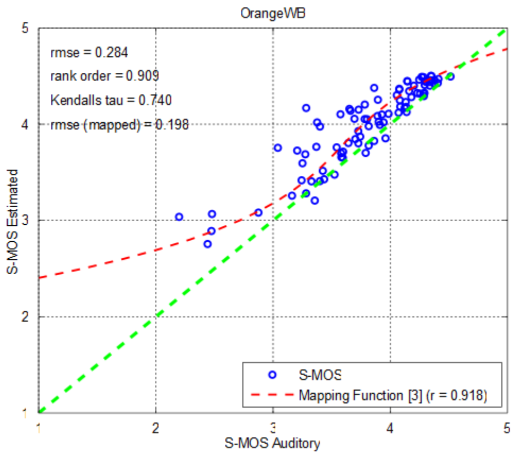


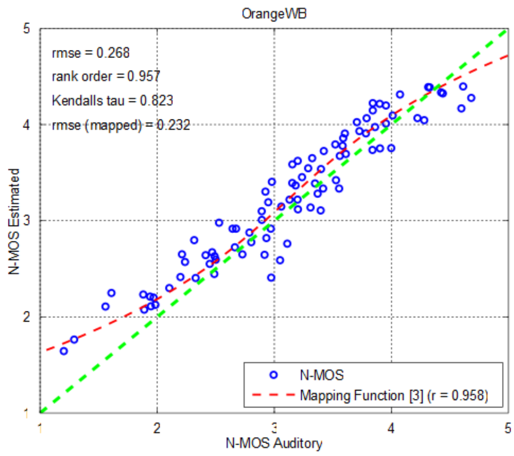


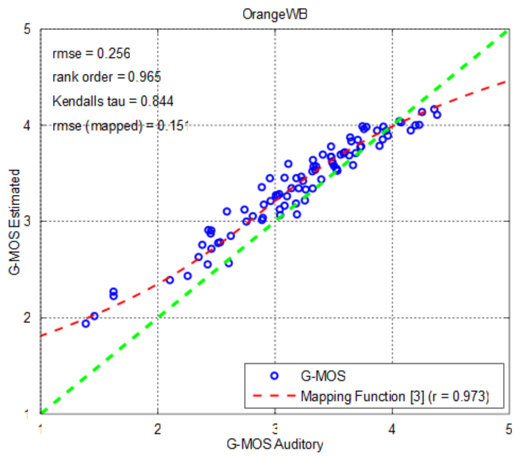
		SIG	BAK	OVRL
RMSE:	no Mapping	0,33	0,47	0,36
	1st Ord. Mapping	0,33	0,48	0,36
	3rd Ord. Mapping	**0,34**	**0,49**	**0,37**

RMSE*:	no Mapping	0,23	0,37	0,26
	1st Ord. Mapping	0,23	0,38	0,26
	3rd Ord. Mapping	**0,24**	**0,38**	**0,26**

7.1.7 Results for database "Orange"



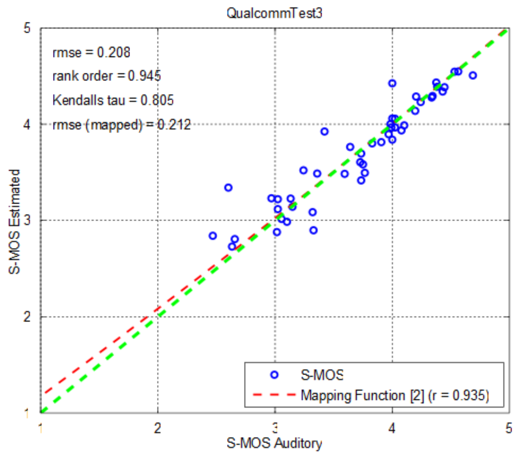


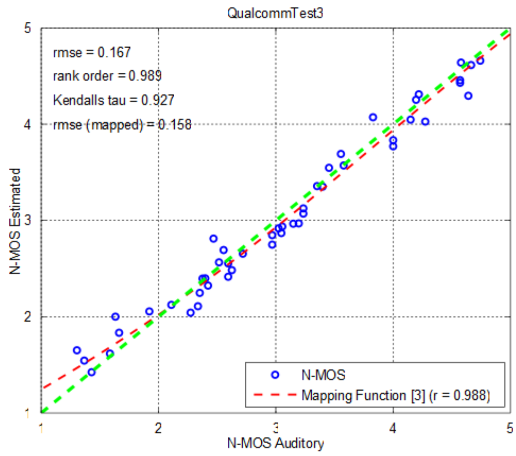


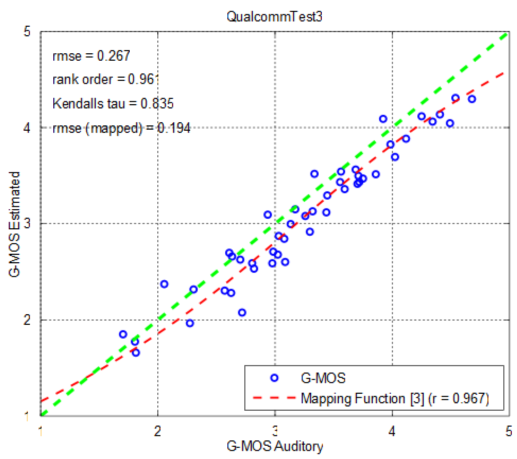
		SIG	BAK	OVRL
RMSE:	no Mapping	0,28	0,27	0,26
	1^{st} Ord. Mapping	0,20	0,24	0,16
	3^{rd} Ord. Mapping	**0,20**	**0,23**	**0,15**

RMSE*:	no Mapping	0,22	0,21	0,20
	1^{st} Ord. Mapping	0,13	0,19	0,10
	3^{rd} Ord. Mapping	**0,13**	**0,18**	**0,10**

7.1.8 Results for database "Qualcomm - Test 3"



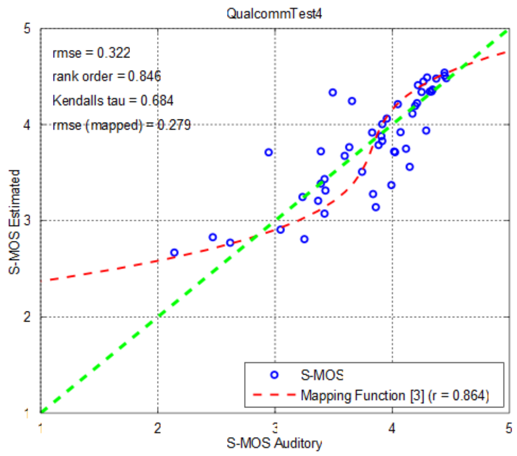


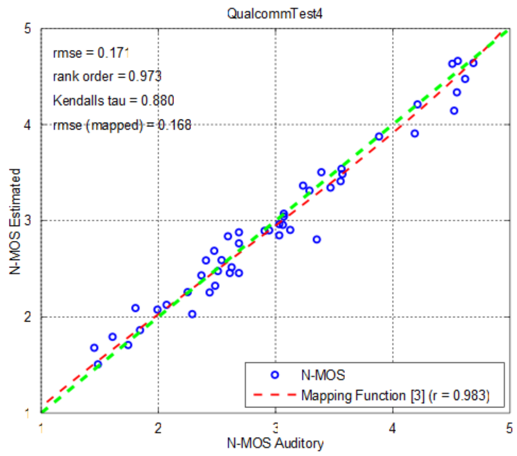


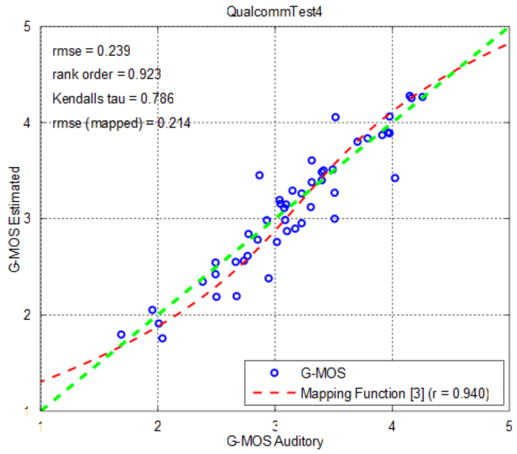
		SIG	BAK	OVRL
RMSE:	no Mapping	0,21	0,17	0,27
	1^{st} Ord. Mapping	0,21	0,17	0,28
	3^{rd} Ord. Mapping	**0,21**	**0,16**	**0,19**

RMSE*:	no Mapping	0,11	0,08	0,16
	1^{st} Ord. Mapping	0,11	0,08	0,18
	3^{rd} Ord. Mapping	**0,11**	**0,07**	**0,10**

7.1.9 Results for database "Qualcomm - Test 4"







		SIG	BAK	OVRL
RMSE:	no Mapping	0,32	0,17	0,24
	1st Ord. Mapping	0,31	0,28	0,24
	3rd Ord. Mapping	**0,28**	**0,17**	**0,21**

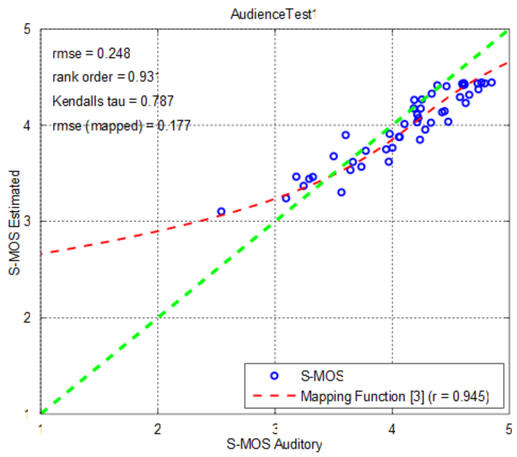
RMSE*:	no Mapping	0,22	0,09	0,16
	1st Ord. Mapping	0,21	0,18	0,14
	3rd Ord. Mapping	**0,17**	**0,08**	**0,12**

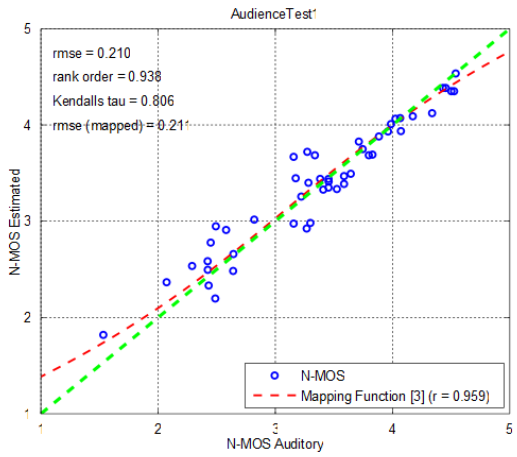
7.2 Results in narrowband mode

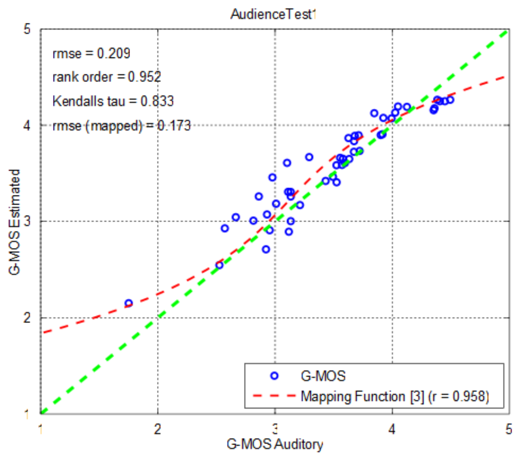
For the narrowband retraining procedure, no database was excluded.

In overall, 6 databases with 288 conditions and 3 840 samples were used.

7.2.1 Results for database "Audience - Test 1"



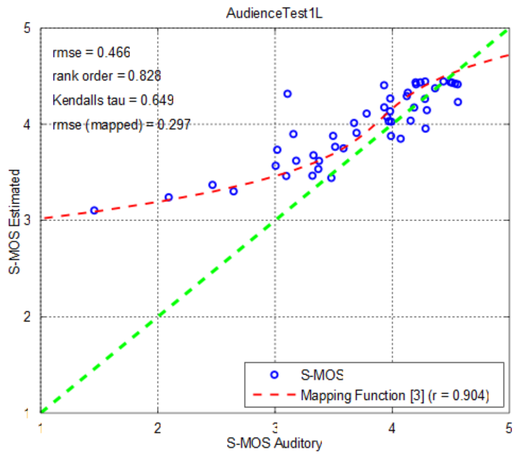


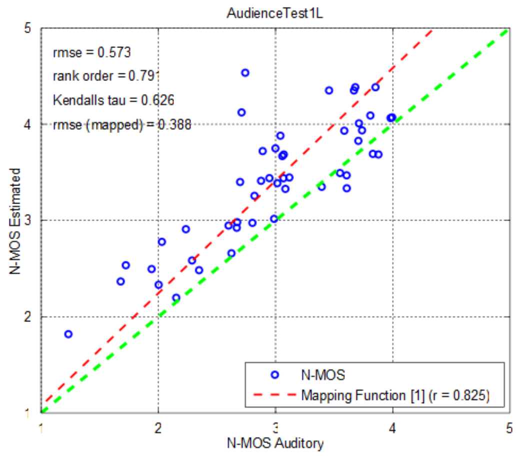


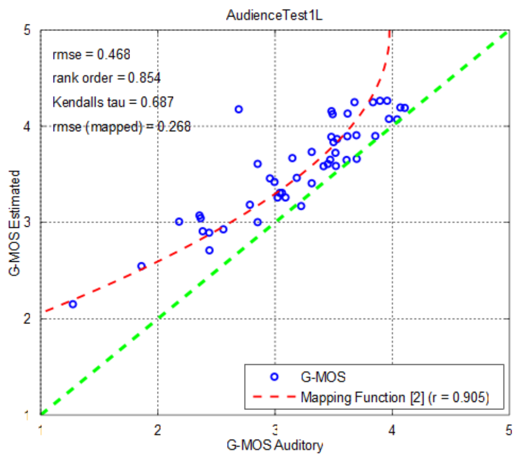
		SIG	BAK	OVRL
RMSE:	no Mapping	0,25	0,21	0,21
	1st Ord. Mapping	0,18	0,21	0,19
	3rd Ord. Mapping	**0,18**	**0,21**	**0,17**

RMSE*:	no Mapping	0,15	0,12	0,12
	1st Ord. Mapping	0,08	0,11	0,10
	3rd Ord. Mapping	**0,08**	**0,11**	**0,07**

7.2.2 Results for database "Audience - Test 1L"



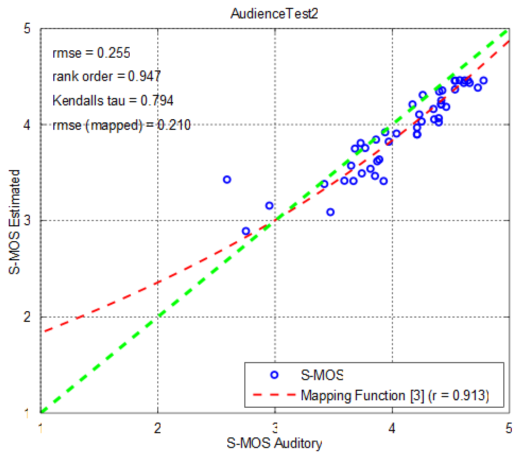


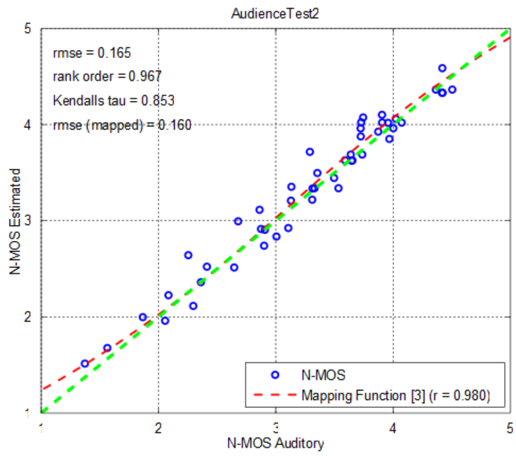


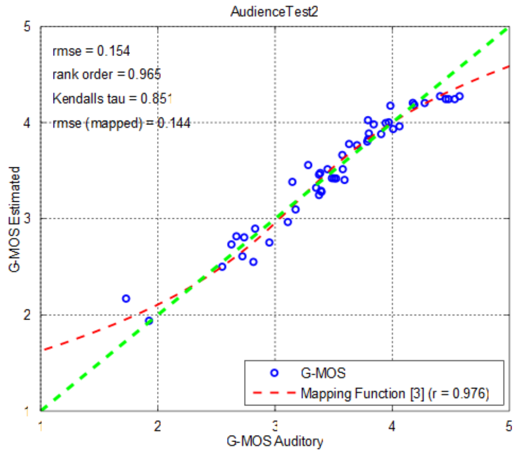
		SIG	BAK	OVRL
RMSE:	no Mapping	0,47	0,57	0,47
	1st Ord. Mapping	0,34	0,39	0,29
	3rd Ord. Mapping	**0,30**	**0,35**	**0,27**

RMSE*:	no Mapping	0,38	0,47	0,37
	1st Ord. Mapping	0,25	0,29	0,20
	3rd Ord. Mapping	**0,21**	**0,24**	**0,19**

7.2.3 Results for database "Audience - Test 2"



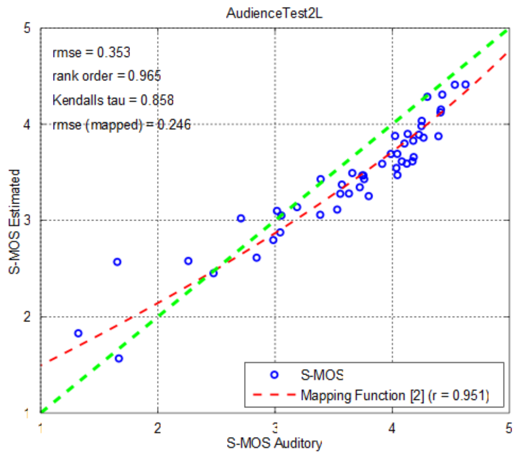


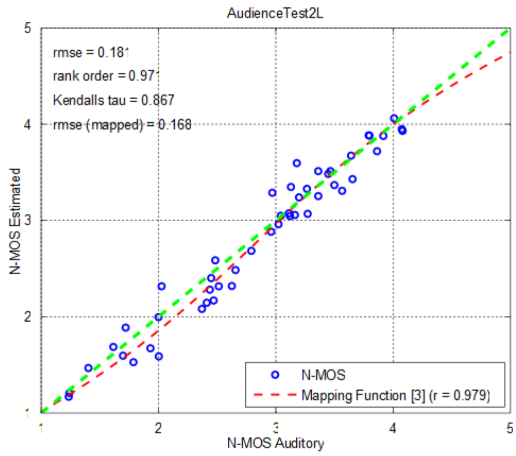


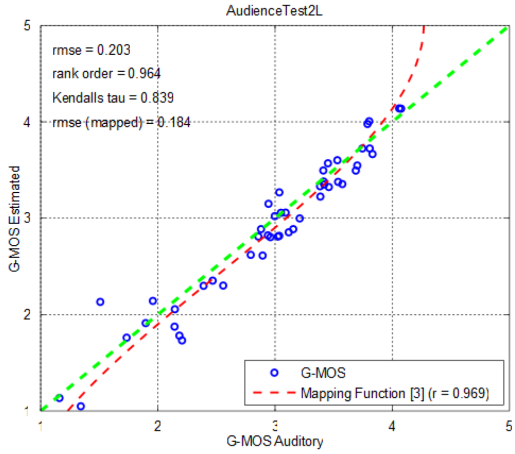
		SIG	BAK	OVRL
RMSE:	no Mapping	0,47	0,57	0,47
	1st Ord. Mapping	0,34	0,39	0,29
	3rd Ord. Mapping	**0,30**	**0,35**	**0,27**

RMSE*:	no Mapping	0,38	0,47	0,37
	1st Ord. Mapping	0,25	0,29	0,20
	3rd Ord. Mapping	**0,21**	**0,24**	**0,19**

7.2.4 Results for database "Audience - Test 2L"



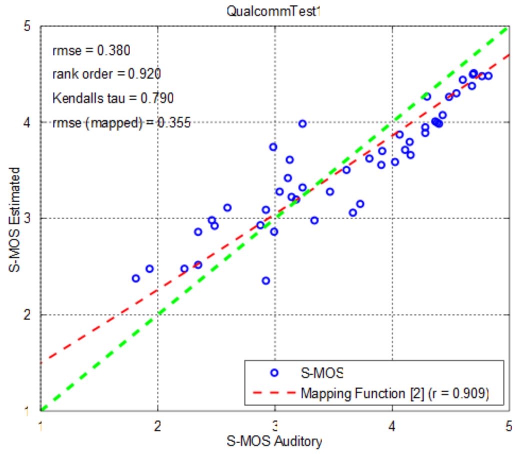


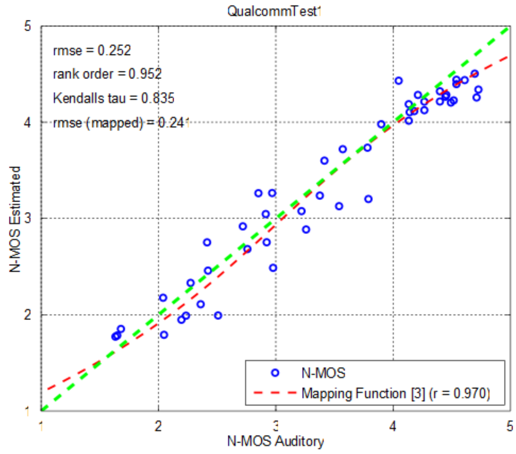


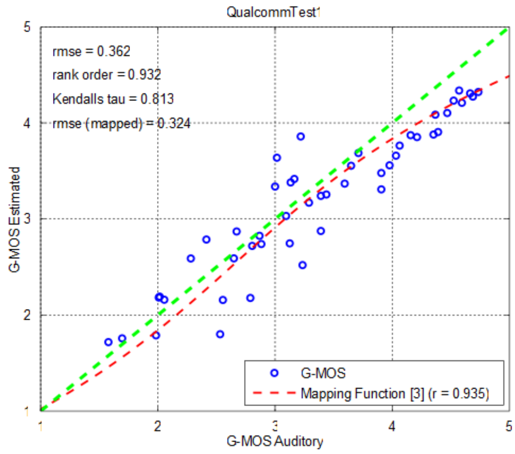
		SIG	BAK	OVRL
RMSE:	no Mapping	0,35	0,18	0,20
	1st Ord. Mapping	0,25	0,17	0,18
	3rd Ord. Mapping	**0,21**	**0,17**	**0,18**

RMSE*:	no Mapping	0,23	0,08	0,11
	1st Ord. Mapping	0,15	0,07	0,11
	3rd Ord. Mapping	**0,11**	**0,08**	**0,11**

7.2.5 Results for database "Qualcomm- Test 1"



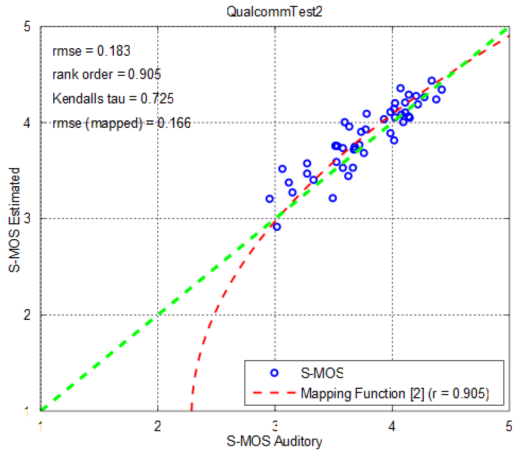


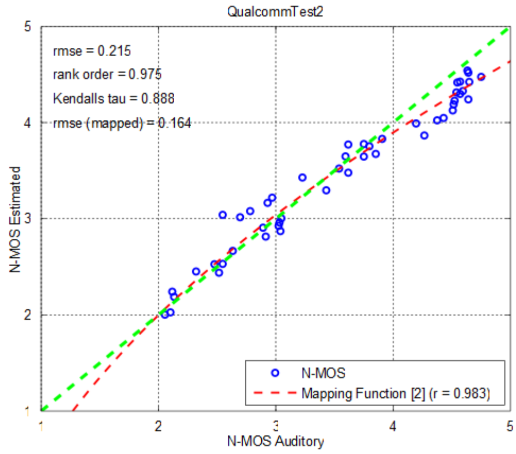


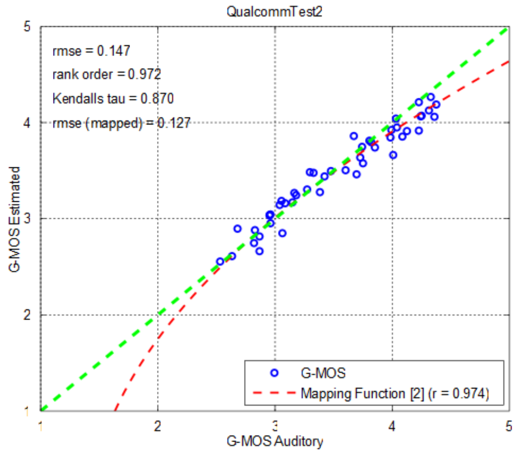
		SIG	BAK	OVRL
RMSE:	no Mapping	0,38	0,26	0,37
	1st Ord. Mapping	0,35	0,33	0,41
	3rd Ord. Mapping	**0,36**	**0,24**	**0,33**

RMSE*:	no Mapping	0,26	0,16	0,27
	1st Ord. Mapping	0,24	0,23	0,31
	3rd Ord. Mapping	**0,24**	**0,13**	**0,22**

7.2.6 Results for database "Qualcomm- Test 2"







		SIG	BAK	OVRL
RMSE:	no Mapping	0,32	0,31	0,37
	1st Ord. Mapping	0,18	0,53	0,22
	3rd Ord. Mapping	**0,18**	**0,17**	**0,16**

RMSE*:	no Mapping	0,21	0,19	0,26
	1st Ord. Mapping	0,08	0,41	0,12
	3rd Ord. Mapping	**0,08**	**0,06**	**0,08**

8 Validation results

For the validation of the model different databases were provided. The databases included different types of conditions and different terminals and simulations. The details of the validation databases are described separately for each set of databases provided by the validation labs.

8.1 Audience validation data

8.1.1 Description of tests

Four tests were conducted, two narrowband (5 & 6) and two wideband (7 & 8). In each test, the noise types listed in [i.1] were used, but the noise levels were increased by 6 dB as in five of the training databases. Six different devices, new to this sequence of validation tests, were used, again a mix of commercial and simulated handsets. All devices were tested in both handset and handheld speakerphone use cases, counterbalanced between the pair of tests at a given bandwidth.

Devices

In each experiment, six devices were evaluated, the maximum number allowed in the EATS-3 [i.1] test plan. In each experiment at one bandwidth, half of the devices were tested in handset mode and half tested in handheld speakerphone mode, in order to provide a consistent and wide range of listening conditions, so that all six devices were tested in both handset and handheld speakerphone modes across the two tests at each bandwidth. The devices included a mix of real and simulated devices with both 1- and 2-microphone noise suppression systems.

The reference conditions and noise types are as defined in table 1 of [i.1].

Reference Conditions				
File	**SIGNAL**	**SNR**	**Noise Type**	
i01	Source (filtered)	No Noise	-	
i02	Source (filtered)	0 dB	Fullsize_Car1_130Kmh_binaural	
i03	Source (filtered)	12 dB	Fullsize_Car1_130Kmh_binaural	
i04	Source (filtered)	24 dB	Fullsize_Car1_130Kmh_binaural	
i05	Source (filtered)	36 dB	Fullsize_Car1_130Kmh_binaural	
i06	NS Level 1	No Noise	-	
i07	NS Level 2	No Noise	-	
i08	NS Level 3	No Noise	-	
i09	NS Level 4	No Noise	-	
i10	NS Level 3	24 dB	Fullsize_Car1_130Kmh_binaural	
i11	NS Level 2	12 dB	Fullsize_Car1_130Kmh_binaural	
i12	NS Level 1	[0 dB]	Fullsize_Car1_130Kmh_binaural	
Test Conditions				
File	**Speech level @ MRP Handset/handsfree**	**Noise level @ HATS ear simulators with ID correction**	**Noise Type**	**Description of Noise from EG 202 396-1 [i.2]**
i13	-1,7/+1,3 dBPa	L: 75,0 dB(A) / R: 73,0 dB(A)	Pub_Noise_binaural_V2	Recording in a pub
i14	-1,7/+1,3 dBPa	L: 74,9 dB(A) / R: 73,9 dB(A)	Outside_Traffic_Road_binaural	Recording at pavement
i15	-1,7/+1,3 dBPa	L: 69,1 dB(A) / R: 69,6 dB(A)	Outside_Traffic_Crossroads_binaural	Recording at pavement
i16	-1,7/+1,3 dBPa	L: 68.2 dB(A) / R:69,8dB(A)	Train_Station_binaural	Recording at departure platform
i17	-1,7/+1,3 dBPa	L: 69,1 dB(A) / R: 68,1 dB(A)	Fullsize_Car1_130Kmh_binaural	Recording in passenger cabin
i18	-1,7/+1,3 dBPa	L: 68,4 dB(A) / R: 67,3 dB(A)	Cafeteria_Noise_binaural	Recording at sales counter
i19	-1,7/+1,3 dBPa	L: 63,4 dB(A) / R: 61,9 dB(A)	Mensa_binaural	Recording in a cafeteria
i20	-1,7/+1,3 dBPa	L: 56,6 dB(A) / R: 57,8 dB(A)	Work_Noise_Office_Callcenter_binaural	Recording in a business office

However, as noted above for these tests, the noise levels were increased by 6 dB as was done in five of the training databases.

8.1.2 Description of validation results

For each test, three scatter plots are shown, plotting the results of the predictions versus the subjective data. In each plot, three sets of data are shown, one for no mapping, one for a first-order remapping, and one for a third-order remapping. Tables of correlation, rmse, and rmse* [i.9] follow each set of scatter plots. The 1st and 3rd order remappings were derived for each experiment from the 48 test conditions, according to the procedure defined in [i.9]. The intention behind showing scatter plots for the three mapping cases is to demonstrate visually that there is only a small impact of the remapping procedure for these data.

8.1.2.1 Experiment 5: Narrowband

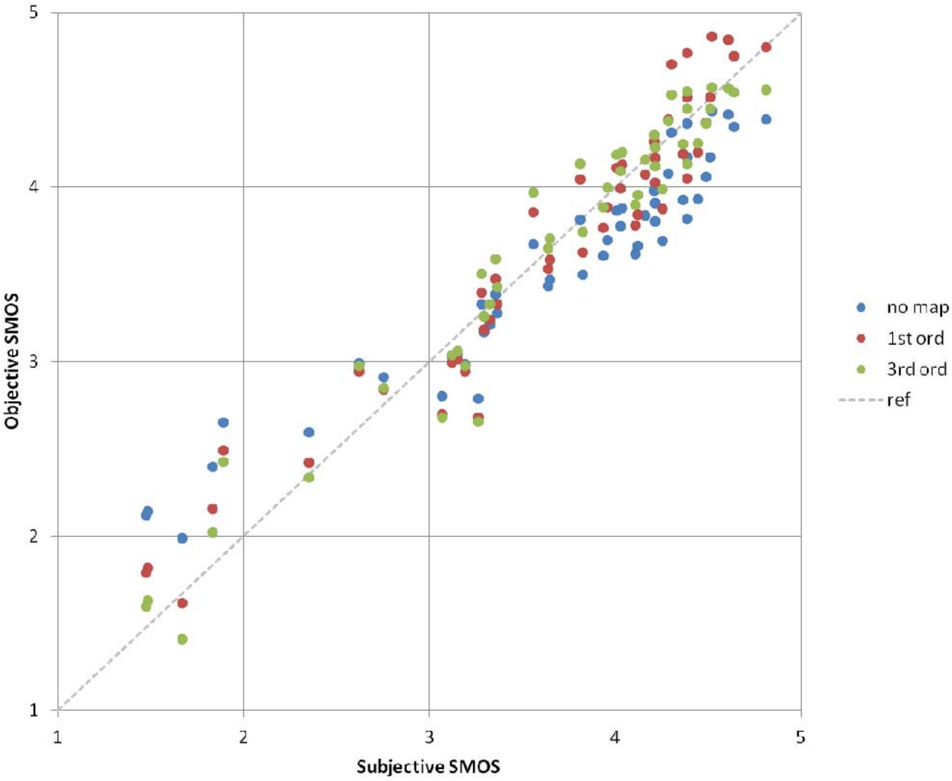


Figure 2: Experiment 5 S-MOS scatter plot

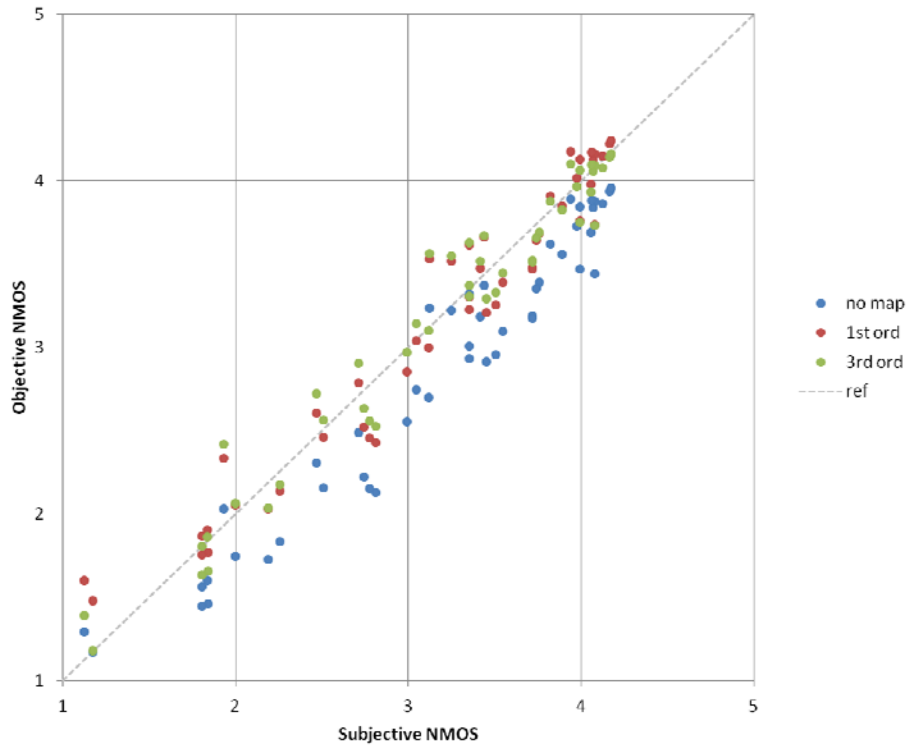


Figure 3: Experiment 5 N-MOS scatter plot

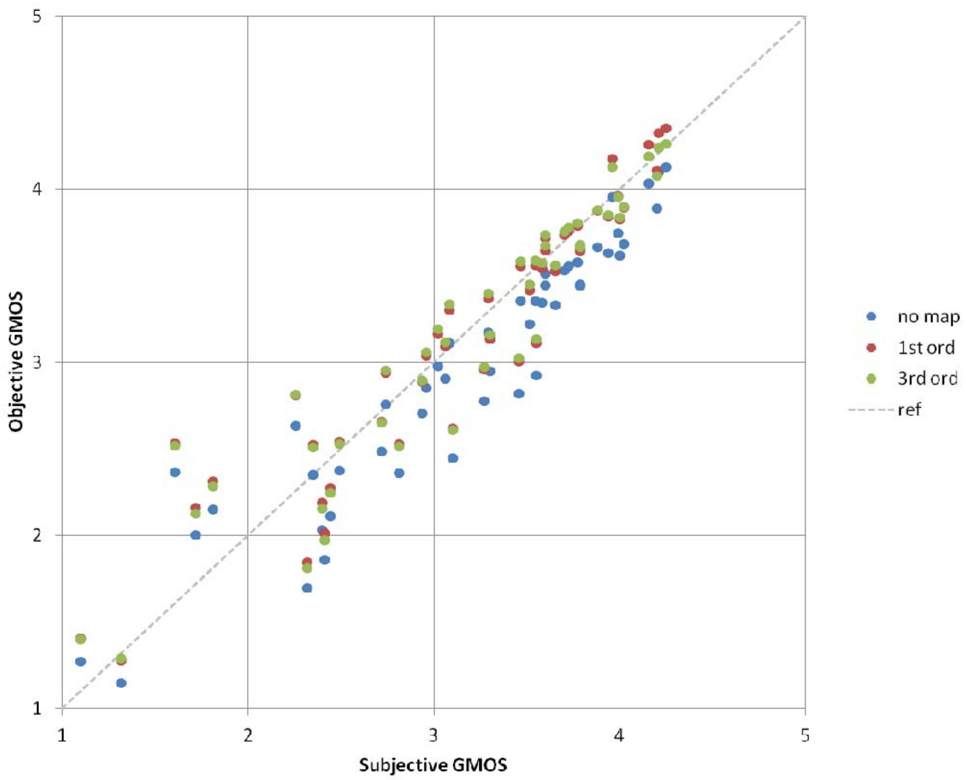


Figure 4: Experiment 5 G-MOS scatter plot

Table 8: Correlation, RMSE, and RMSE* for experiment 5

Condition	S-MOS	N-MOS	G-MOS
Correlation	0,96	0,97	0,94
RMSE, no mapping	0,35	0,36	0,33
RMSE, 1st order mapping	0,25	0,20	0,27
RMSE, 3rd order mapping	**0,22**	**0,18**	**0,28**
RMSE*, no mapping	0,24	0,25	0,23
RMSE*, 1st order mapping	0,14	0,12	0,20
RMSE*, 3rd order mapping	**0,12**	**0,10**	**0,20**

8.1.2.2 Experiment 6: Narrowband

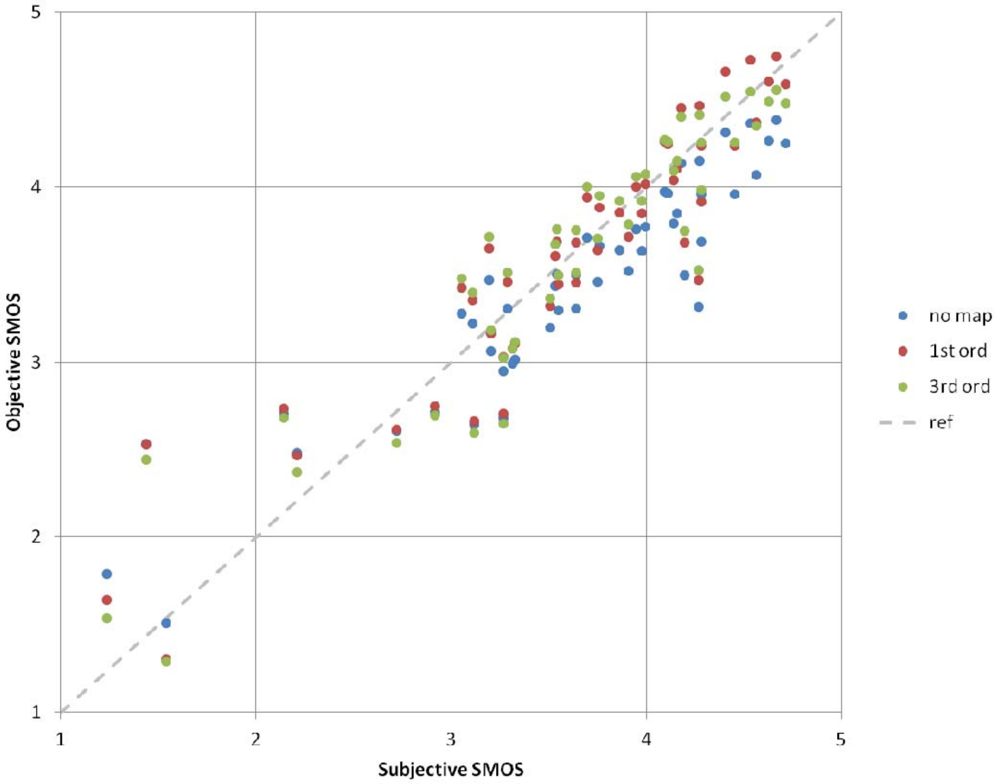


Figure 5: Experiment 6 G-MOS scatter plot

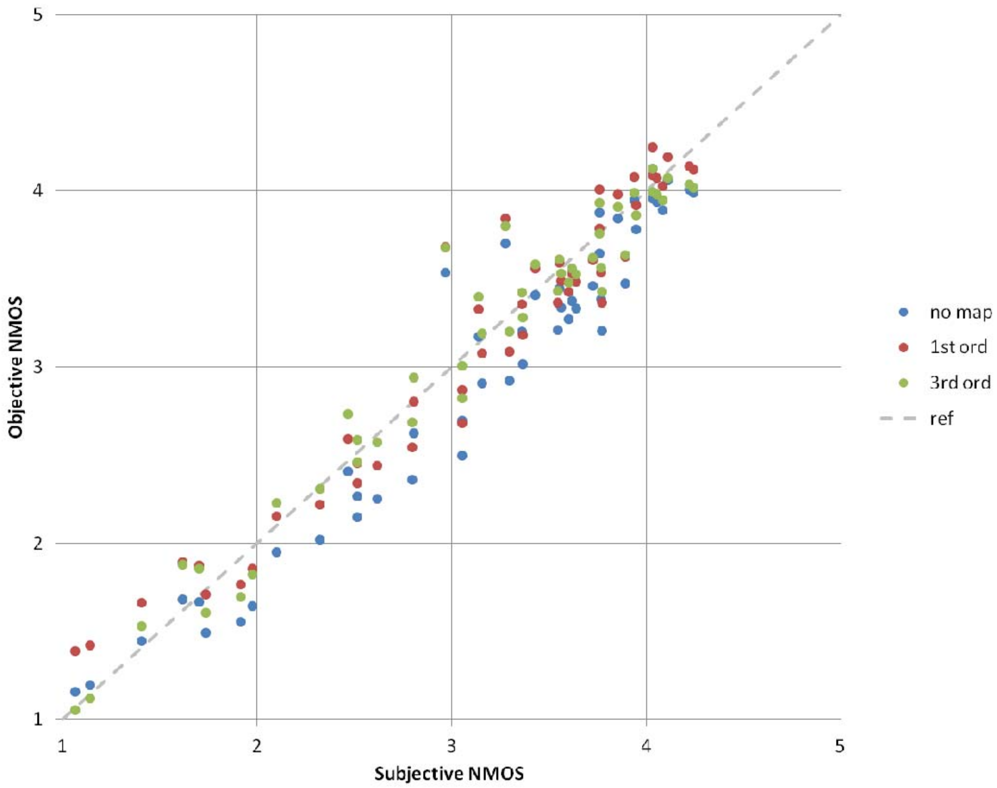


Figure 6: Experiment 6 N-MOS scatter plot

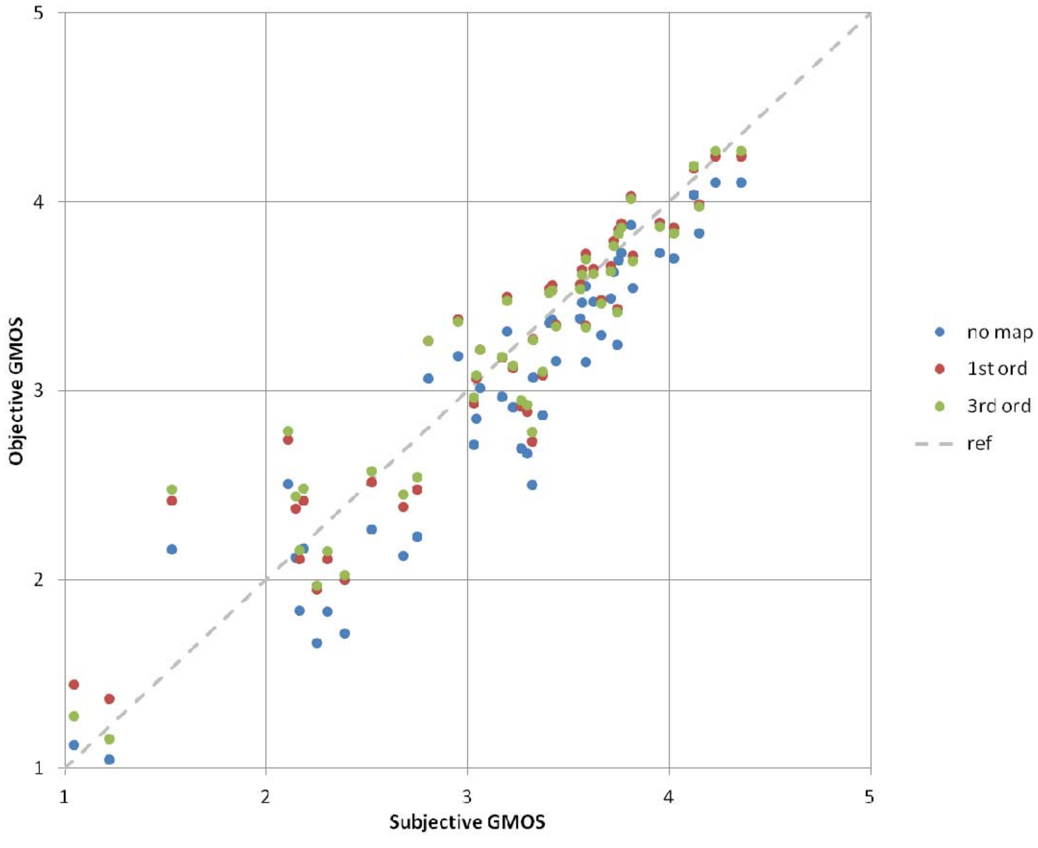


Figure 7: Experiment 6 G-MOS scatter plot

Table 9: Correlation, RMSE, and RMSE* for Experiment 6

Condition	S-MOS	N-MOS	G-MOS
Correlation	0,93	0,97	0,93
RMSE, no mapping	0,38	0,28	0,35
RMSE, 1st order mapping	0,32	0,22	0,28
RMSE, 3rd order mapping	**0,32**	**0,20**	**0,28**
RMSE*, no mapping	0,28	0,18	0,25
RMSE*, 1st order mapping	0,22	0,14	0,19
RMSE*, 3rd order mapping	**0,22**	**0,12**	**0,20**

8.1.2.3 Experiment 7: Wideband

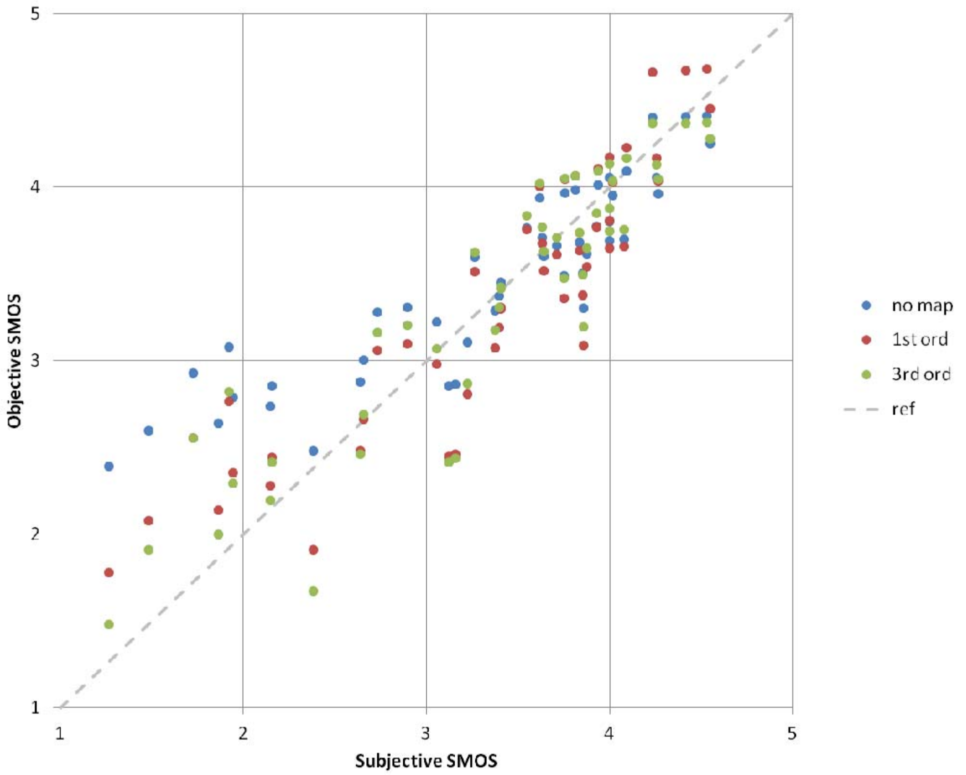


Figure 8: Experiment 7 S-MOS scatter plot

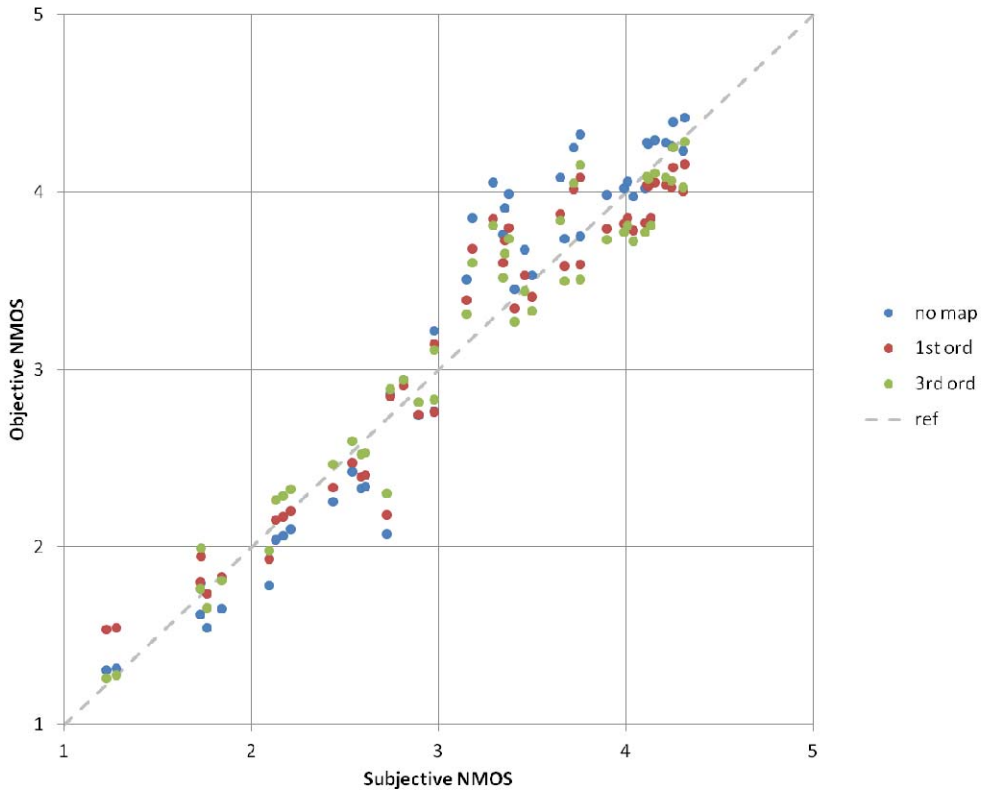


Figure 9: Experiment 7 N-MOS scatter plot

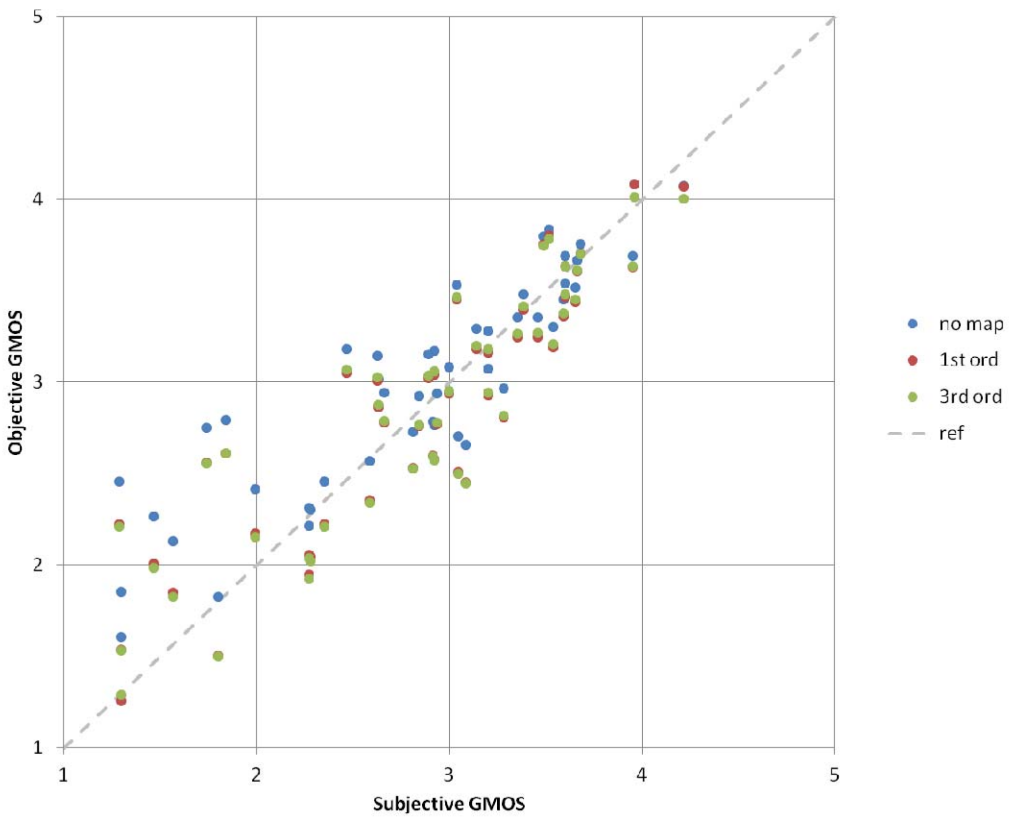


Figure 10: Experiment 7 G-MOS scatter plot

Table 10: Correlation, RMSE, and RMSE* for Experiment 7

Condition	S-MOS	N-MOS	G-MOS
Correlation	0,90	0,96	0,89
RMSE, no mapping	0,46	0,29	0,39
RMSE, 1st order mapping	0,37	0,24	0,35
RMSE, 3rd order mapping	**0,36**	**0,22**	**0,36**
RMSE*, no mapping	0,36	0,20	0,32
RMSE*, 1st order mapping	0,26	0,13	0,26
RMSE*, 3rd order mapping	**0,25**	**0,12**	**0,27**

8.1.2.4 Experiment 8: Wideband

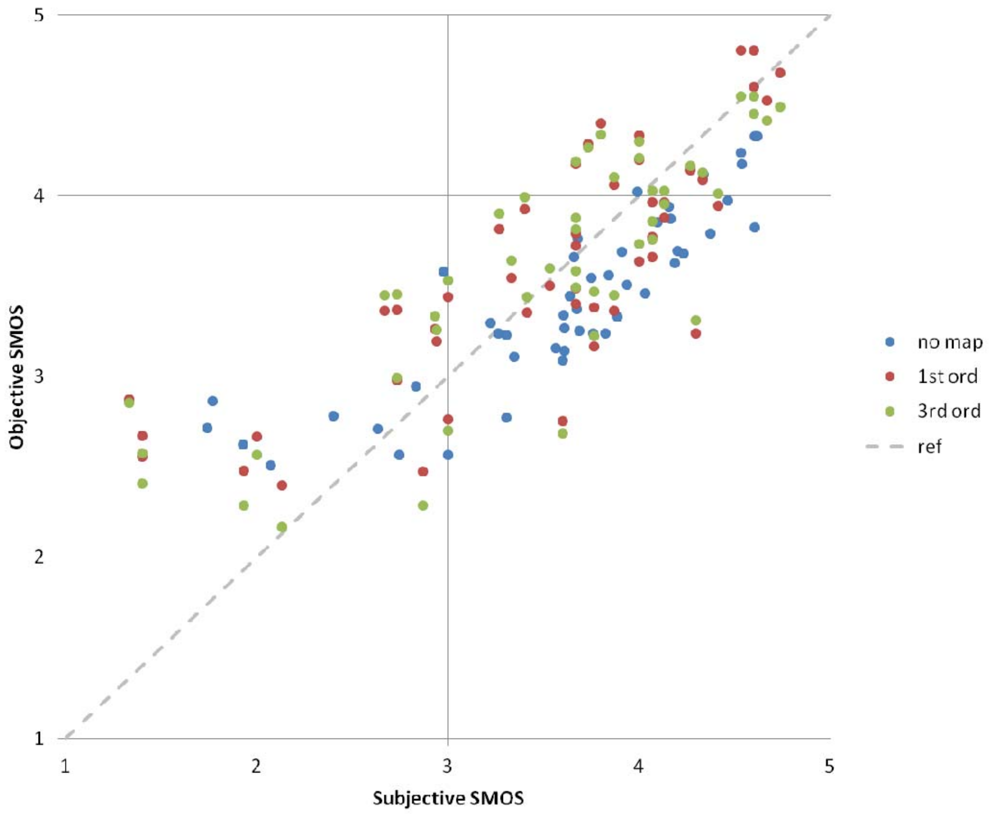


Figure 11: Experiment 8 S-MOS scatter plot

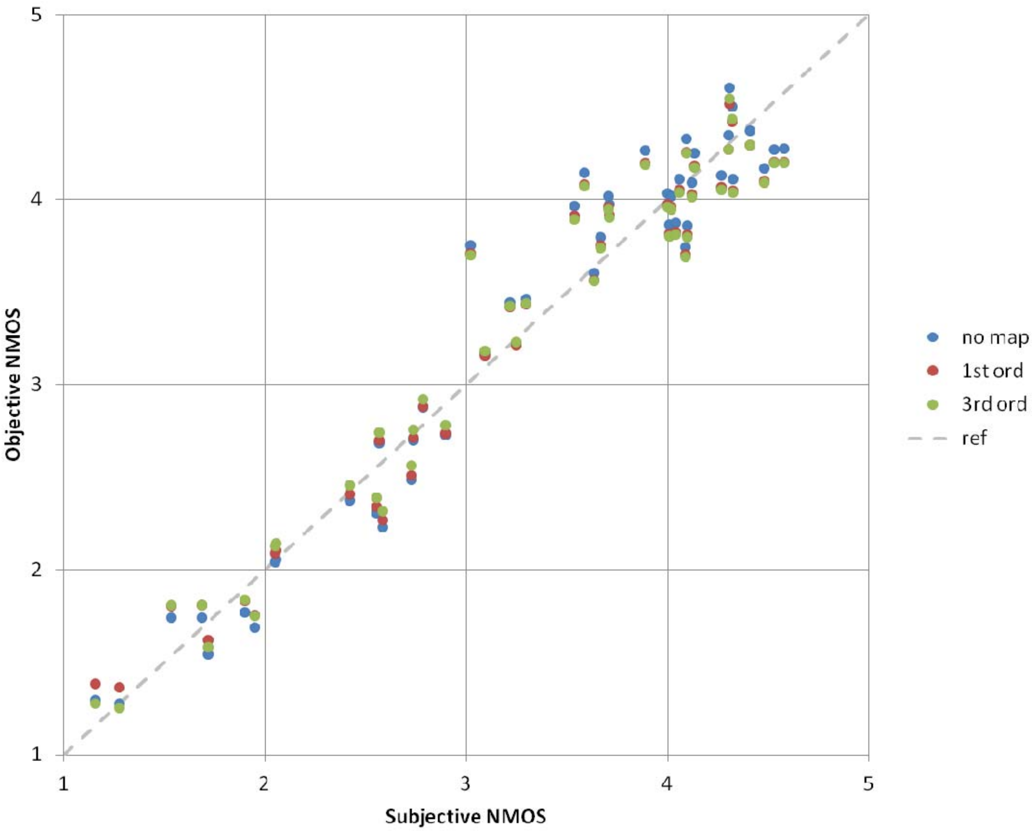


Figure 12: Experiment 8 N-MOS scatter plot

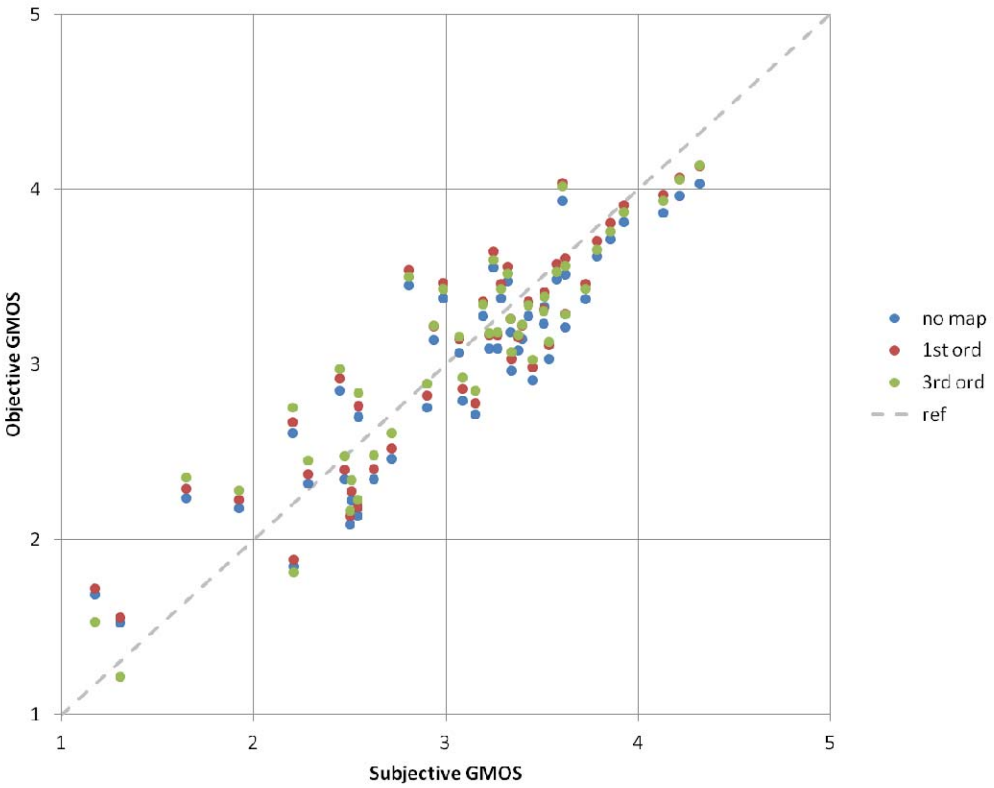


Figure 13: Experiment 8 G-MOS scatter plot

Table 11: Correlation, RMSE, and RMSE* for Experiment 8

Condition	S-MOS	N-MOS	G-MOS
Correlation	**0,87**	**0,97**	**0,90**
RMSE, no mapping	0,45	0,24	0,31
RMSE, 1st order mapping	0,38	0,23	0,31
RMSE, 3rd order mapping	**0,37**	**0,24**	**0,30**
RMSE*, no mapping	0,32	0,14	0,20
RMSE*, 1st order mapping	0,26	0,14	0,20
RMSE*, 3rd order mapping	**0,26**	**0,14**	**0,20**

8.2 Orange validation data

8.2.1 Description of tests

The Orange validation database includes six wideband mobile devices, and three noises from EG 202-396-1 [i.4] at nominal level are used (see table 12). As for speech samples, four talkers are used: two males and two females, with two sentences for each talker. The resulting tests conditions are summarized in table 2. Except for f3, all talkers come from ITU-T Recommendation P.501 [i.13].

Table 12: Noise names and descriptions for Orange validation database

Noise type	Description	EG 202 396-1 [i.2] filename
Crossroad	Recording at pavement	Outside_Traffic_Crossroads_binaural
Mensa	Recording in a cafeteria	Mensa_binaural
Pub	Recording in a Pub	Pub_Noise_binaural_V2

Table 13: Definition of tests conditions parameters for Orange WB validation test

Test conditions	Number	Designation
Noises	3	N1, N2, N3
SNR	1	Nominal level
Devices	6	D1, …, D6
Talkers	4	m1, m2, f2, f3
Sentences per talker	2	s1, s2

All test conditions were processed with the 4 talkers and 2 sentences. Level adjustment was performed as described in EATS-3.

Reference conditions which incorporate a spectral subtraction based distortion were included in the test and are listed in table 14. These reference conditions are exactly the same as the one provided in EATS-3, table 2 of [i.1].

Table 14: Reference set conditions for wideband testing

Reference Conditions			
File	**SIG.**	**SNR**	**Noise Type**
i01	Source (filtered)	No Noise	-
i02	Source (filtered)	10 dB	Outside_Traffic_Crossroads_binaural
i03	Source (filtered)	20 dB	Outside_Traffic_Crossroads_binaural
i04	Source (filtered)	30 dB	Outside_Traffic_Crossroads_binaural
i05	Source (filtered)	40 dB	Outside_Traffic_Crossroads_binaural
i06	NS Level 1, 2nd set of parameters	No Noise	-
i07	NS Level 2, 2nd set of parameters	No Noise	-
i08	NS Level 3, 2nd set of parameters	No Noise	-
i09	NS Level 4, 2nd set of parameters	No Noise	-
i10	NS Level 3, 2nd set of parameters	30 dB	Outside_Traffic_Crossroads_binaural
i11	NS Level 2, 2nd set of parameters	20 dB	Outside_Traffic_Crossroads_binaural
i12	NS Level 1, 2nd set of parameters	10 dB	Outside_Traffic_Crossroads_binaural

8.2.2 Description of validation results

Scatter plots on a per condition basis are provided in figures 15 to 17: they show the distribution over the quality range for the three dimensions (Speech, Noise, Overall quality).

The RMSE and RMSE* performance parameters specified in [i.9] were computed. Results before mapping and after monotonic 3rd order mapping are presented in tables 15 and 16 respectively. The Pearson correlation is also reported in table 17. These results are meeting the performance requirements specified for RMSE and RMSE* on the 3rd order remapping, as given in [i.9].

Table 15: Statistical analysis results before mapping

	S-MOS	N-MOS	G-MOS
RMSE	0,68	0,29	0,62
RMSE*	0,58	0,23	0,53

Table 16: Statistical analysis results after monotonic 3rd order mapping

	S-MOS	N-MOS	G-MOS
RMSE	0,38	0,23	0,29
RMSE*	0,30	0,16	0,21

Table 17: Pearson correlation (after monotonic 3rd order mapping)

	S-MOS	N-MOS	G-MOS
before mapping	0,90	0,97	0,90
after monotonic 3rd order mapping	0,91	0,98	0,93

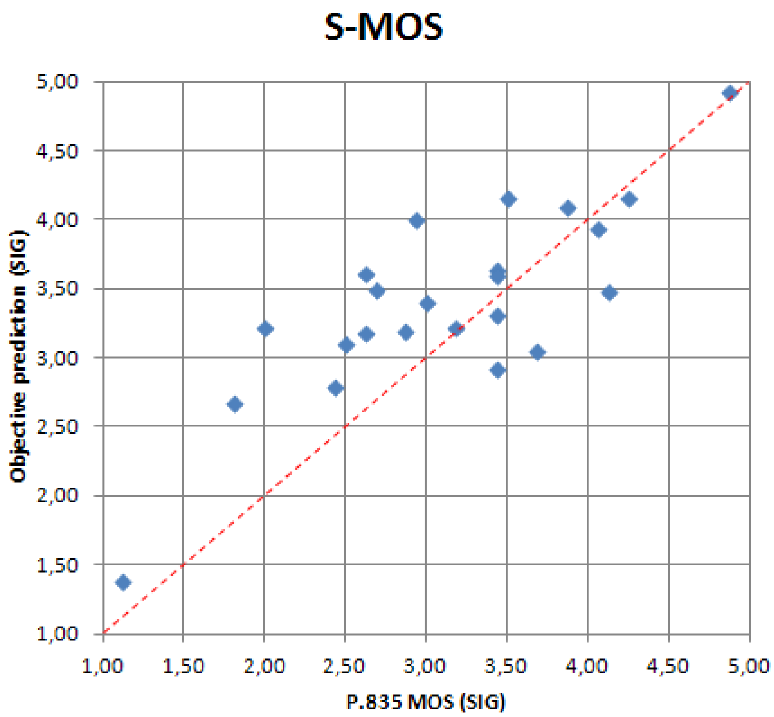


Figure 14: S-MOS scatter plot

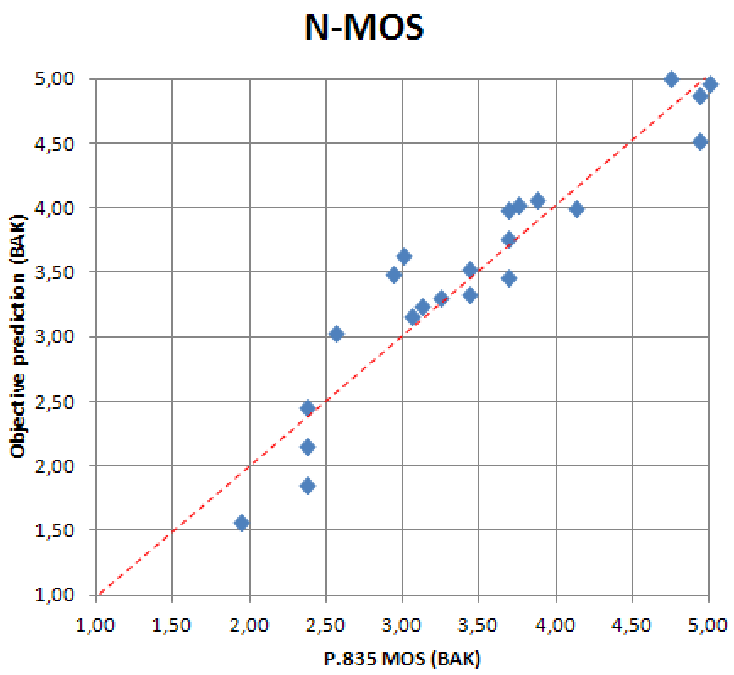


Figure 15: N-MOS scatter plot

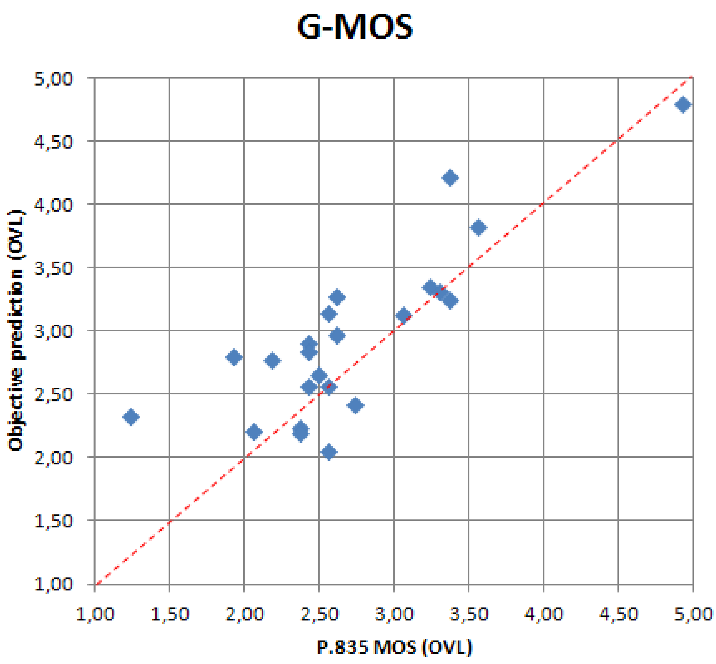


Figure 16: G-MOS scatter plot

8.3 Qualcomm validation data

8.3.1 Description of tests

Two narrowband experiments following the EATS-3 subjective test plan [i.1] were conducted. The test set-up, background noise reproduction calibration and levels, noise types and convergence sequencing are according to the EATS-3 subjective test plan [i.1], except where noted. The reference conditions are according to [i.1], table 1.

In the first validation experiment (Exp 6), 2 devices were tested with 7 noise types and a clean condition (no noise added). The devices were tested in the following modes:

- Handset with AMR 12,2 kbps
- Handset with AMR 5,9 kbps
- Handheld Hands-free with AMR 5,9 kbps

resulting in a total of 48 test conditions. The inclusion of AMR 5,9 kbps was used in order to increase the range of degradations for the validation tests. Commercial devices in a call with a CMU200 network simulator were used.

In the second validation experiment (Exp 7), 1 device was tested with 7 noise types and a clean condition (no noise added). The device was tested in the following modes:

- Handset with AMR 12,2 kbps
- Handset with AMR 5,9 kbps
- Handheld Hands-free with AMR 5,9 kbps
- Handset with AMR 12,2 kbps (Noise levels increased by 6 dB)
- Handset with AMR 5,9 kbps (Noise levels increased by 6 dB)
- Handheld Hands-free with AMR 5,9 kbps (Noise levels increased by 6 dB)

resulting in a total of 48 test conditions. A commercial device in a call with the CMU200 network simulator was used.

The same reference set (exact same signals) was used in the narrowband experiments reported in previous contributions in order to keep consistency and facilitate any necessary mapping or normalization of the data.

Tables 18 and 19 detail the conditions for both experiments.

Table 18: Summary of experimental conditions for EXP 6 (NB)

Experiment	6
Number of devices	2 (HS AMR 12.2; HS AMR 5.9; HHHF AMR 5.9)
Number of noise conditions per device	8 noise conditions
Number of reference conditions	12
Number of test conditions	48
Number of talkers	4
Number of samples per talker	4
Number of votes per condition	128
Method of presentation	Diotic
Presentation level (for -26 dBov)	73dBSPL
Headphones	HD280 PRO
Reference set	According to table 1 and batch processing script in section 8.3 of [i.1].
Noise conditions	Pub_Noise_binaural_V2
	Outside_Traffic_Road_binaural
	Outside_Traffic_Crossroads_binaural
	Clean (no noise)
	Fullsize_Car1_130Kmh_binaural
	Cafeteria_Noise_binaural
	Mensa_binaural
	Work_Noise_Office_Callcenter_binaural

Table 19: Summary of experimental conditions for EXP 7 (NB)

Experiment	7
Number of devices	1 (HS AMR12.2; HS AMR5.9, HHHF AMR12.2, HHHF AMR5.9)
Number of noise conditions per device	16 noise conditions
Number of reference conditions	12
Number of test conditions	48
Number of talkers	4
Number of samples per talker	4
Number of votes per condition	128
Method of presentation	Diotic
Presentation level (for -26 dBov)	73dBSPL
Headphones	HD280 PRO
Reference set	According to table 1 and batch processing script in section 8.3 of [i.1].
Noise conditions	Pub_Noise_binaural_V2 (nominal and +6 dB)
	Outside_Traffic_Road_binaural (nominal and +6 dB)
	Outside_Traffic_Crossroads_binaural (nominal and +6 dB)
	Clean (no noise, two different recordings)
	Fullsize_Car1_130Kmh_binaural (nominal and +6 dB)
	Cafeteria_Noise_binaural (nominal and +6 dB)
	Mensa_binaural (nominal and +6 dB)
	Work_Noise_Office_Callcenter_binaural (nominal and +6 dB)

The results for Experiment 6 and 7 are summarized in figures 17 and 18. The results for S-MOS (SIG), N-MOS (BAK) and G-MOS (OVRL) of 60 conditions (being 48 test and 12 reference conditions) are reported for each experiment. Results are sorted by OVRL.

It can be seen that both experiments exercised the entire range of degradations for the SIG, BAK and OVRL scales. About 67 % of the scores for OVRL are > 3,0 in both tests. This is in contrast with previous experiments conducted by the source where 3,0 represented the median of the scores for OVRL. This effect is observed despite an attempt to increase the range of degradations by including raised noise levels and AMR 5,9 kbps speech coding.

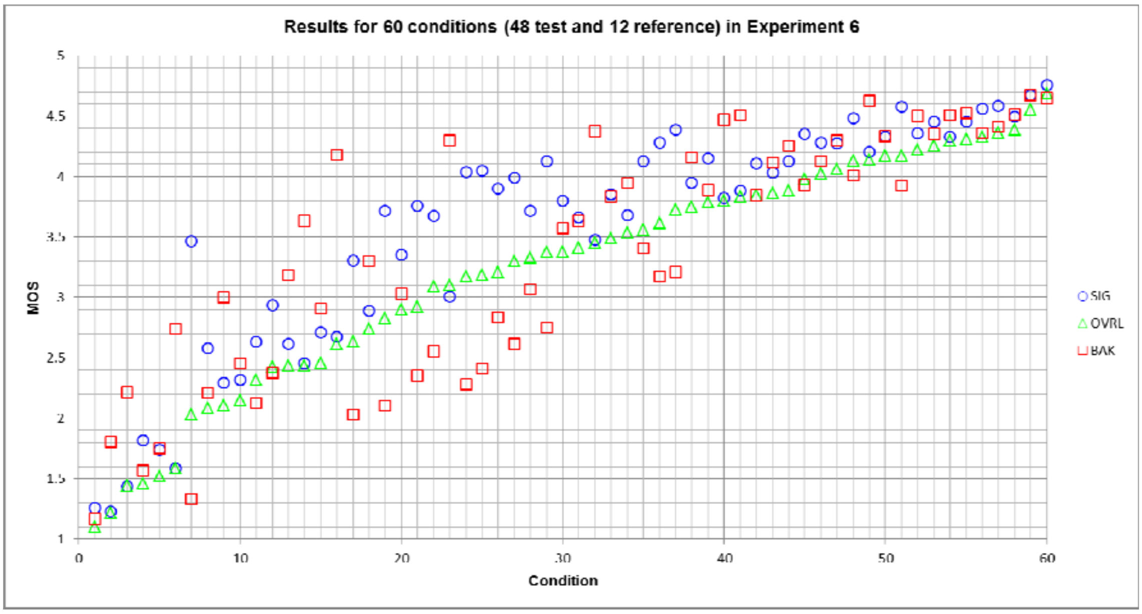


Figure 17: Results of Experiment 6

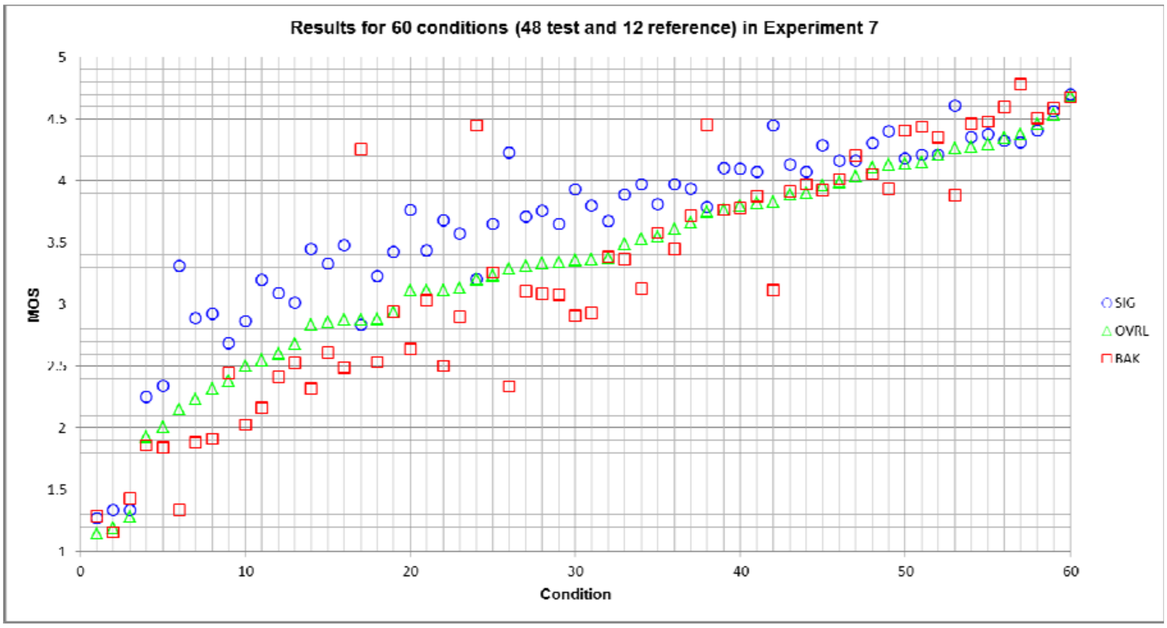


Figure 18: Results of Experiment 7

8.3.2 Description of validation results

Each individual sample used in Experiments 6 and 7 was processed by HEAD Acoustics GmbH using the re-trained P.835 objective predictor model. An average of the objective scores per condition (average of the scores of 16 samples), as well as the 95 % confidence interval was computed and plotted against the results of the subjective test. Scatter plots for N-MOS, S-MOS and G-MOS are shown in figures 19 to 24.

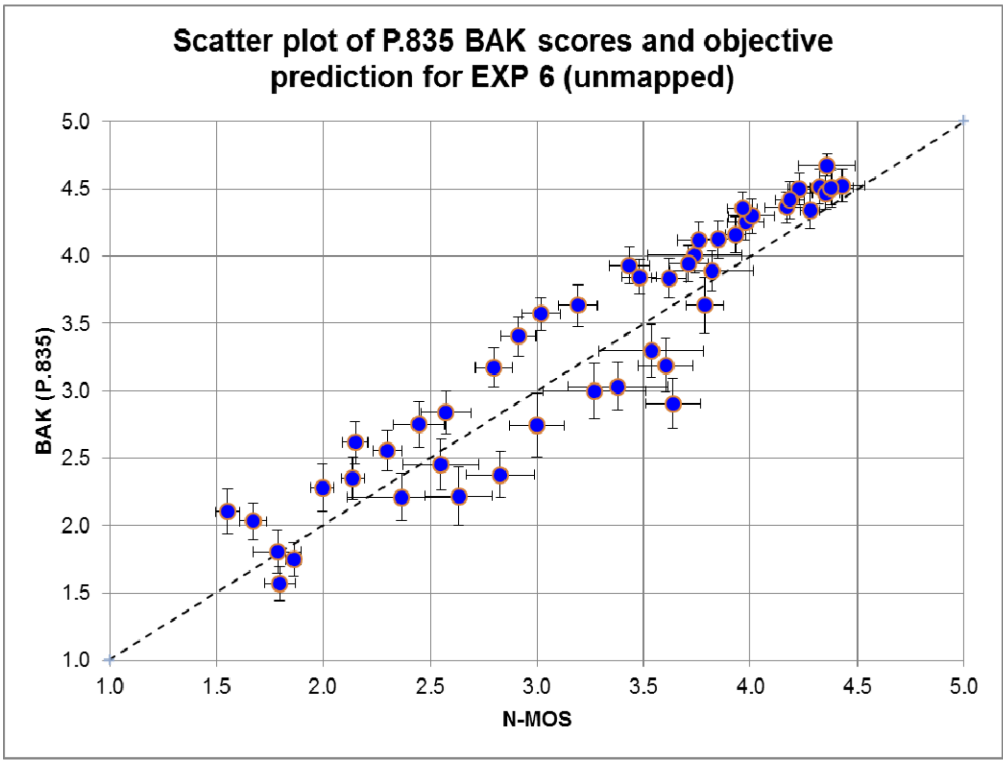


Figure 19: Experiment 6 N-MOS scatter plot

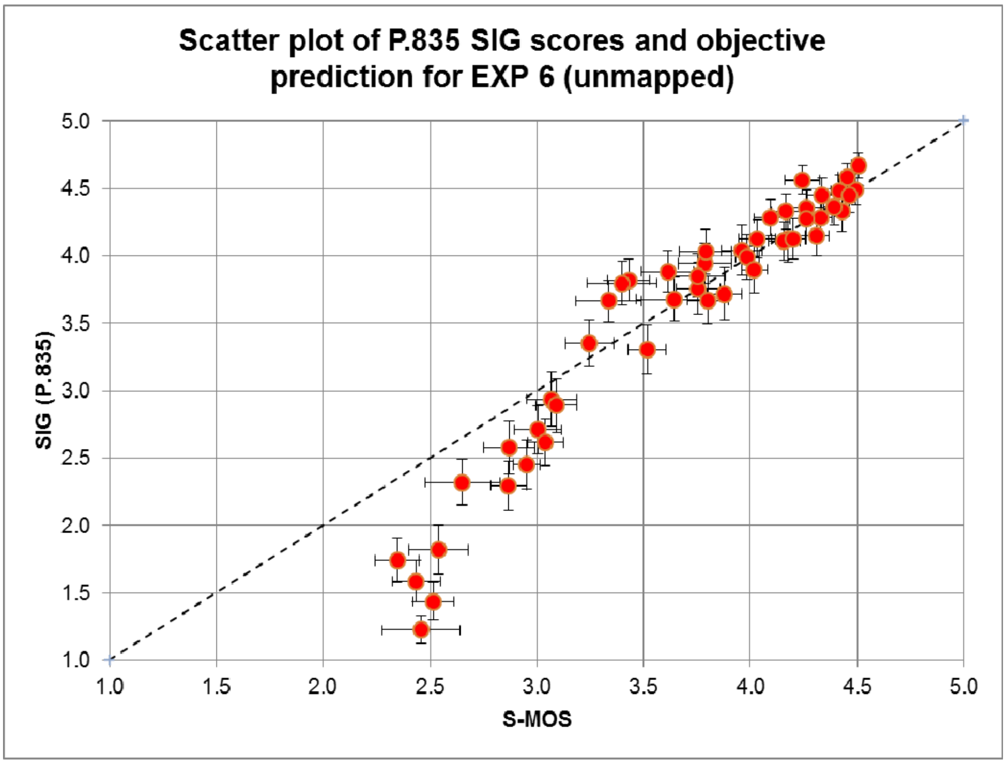


Figure 20: Experiment 6 S-MOS scatter plot

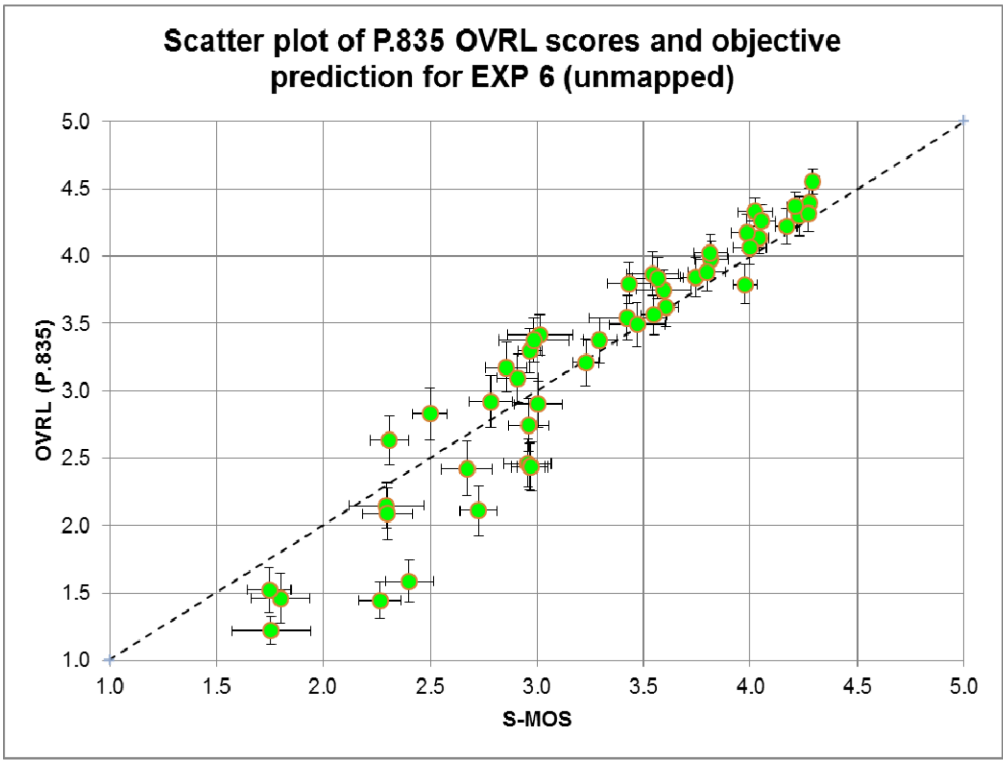


Figure 21: Experiment 6 G-MOS scatter plot

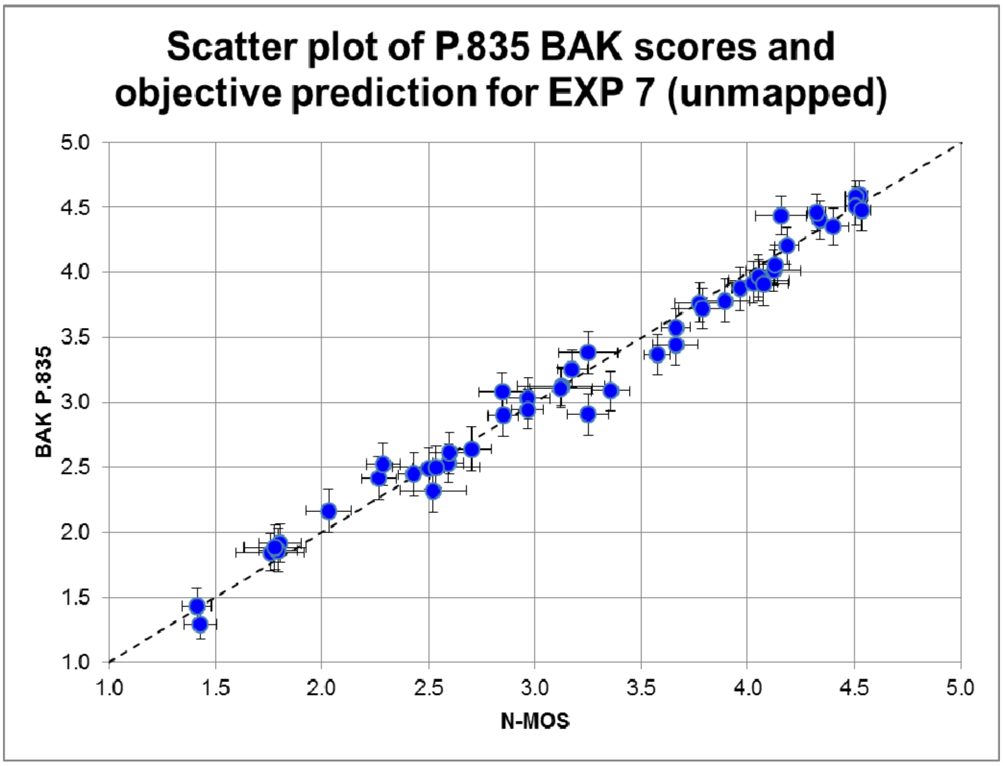


Figure 22: Experiment 7 N-MOS scatter plot

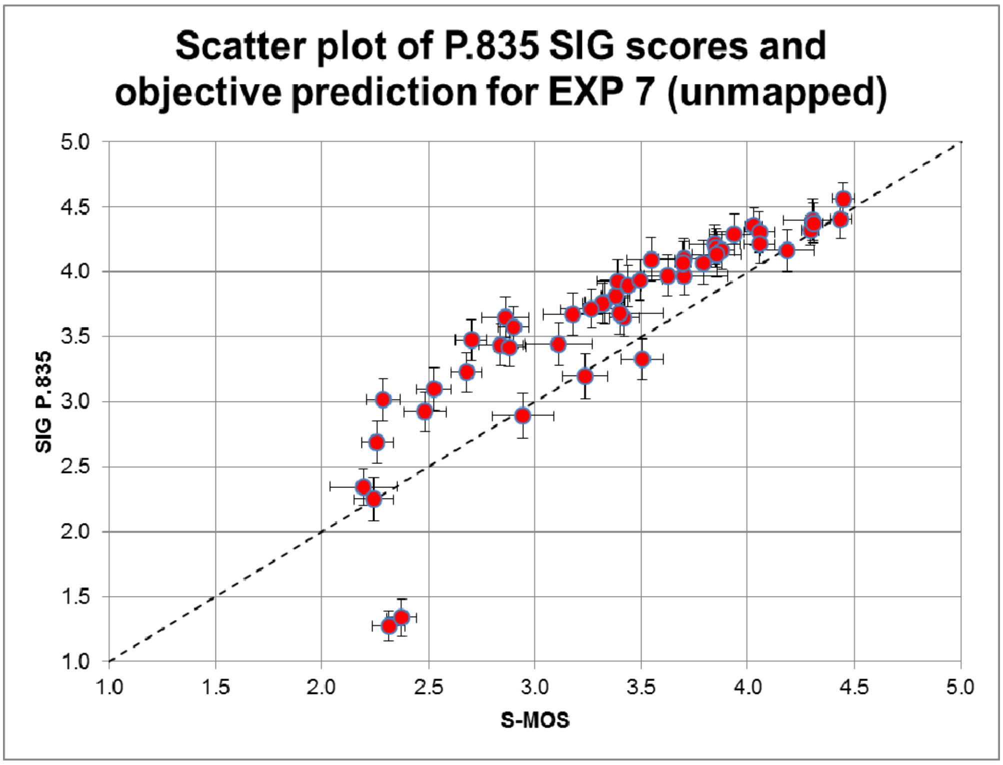


Figure 23: Experiment 7 S-MOS scatter plot

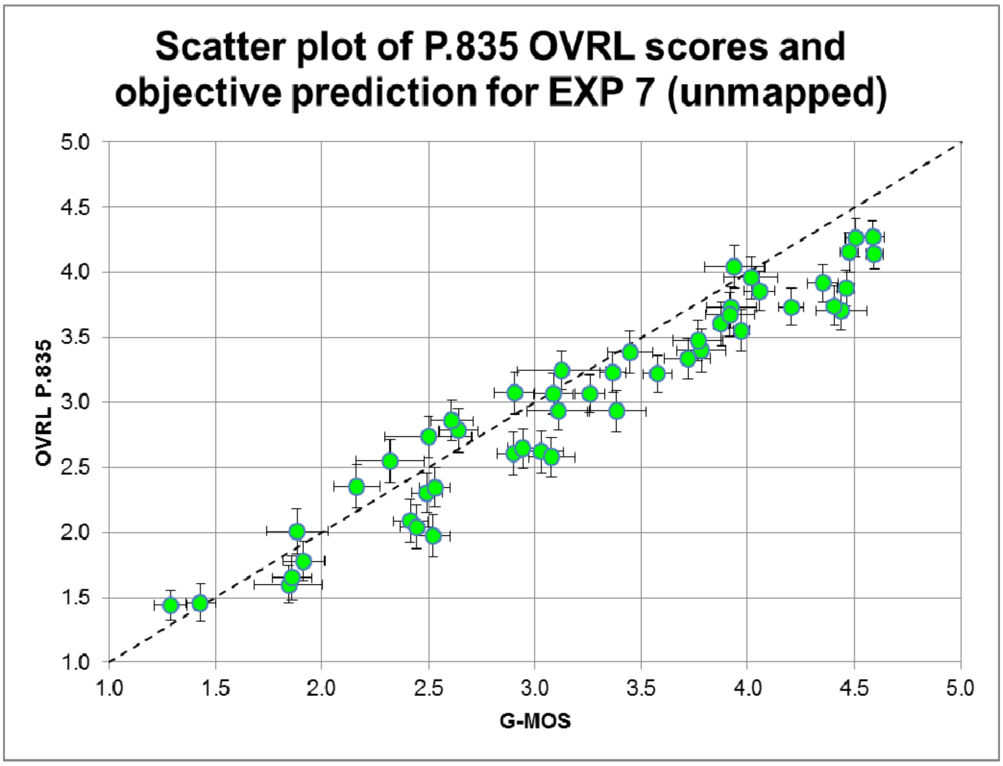


Figure 24: Experiment 7 G-MOS scatter plot

The Pearson correlation coefficient, RMSE and RMSE* performance parameters specified in [i.9] were computed for both validation databases and reported in tables 20 and 21 along with results before and after 1st and 3rd order mapping.

Table 20: Performance of the objective predictor on NB validation database from EXP6

	Condition	S-MOS	N-MOS	G-MOS
	Correlation	0,96	0,95	0,95
RMSE:	no Mapping	0,37	0,32	0,32
	1st Ord. Map.	0,26	0,30	0,28
	3rd Ord. Map	0,19	0,30	0,28
RMSE*:	no Mapping	0,28	0,20	0,22
	1st Ord. Map.	0,17	0,18	0,18
	3rd Ord. Map	0,09	0,17	0,18

Table 21: Performance of the objective predictor on NB validation database from EXP7

	Condition	S-MOS	N-MOS	G-MOS
	Correlation	0,87	0,99	0,97
RMSE:	no Mapping	0,45	0,13	0,36
	1st Ord. Map.	0,36	0,13	0,19
	3rd Ord. Map	0,33	0,12	0,16
RMSE*:	no Mapping	0,33	0,04	0,23
	1st Ord. Map.	0,28	0,04	0,12
	3rd Ord. Map	0,25	0,04	0,07

9 Application of the retrained model

In order to avoid ambiguities in the results the objective model should be applied in the way it was applied during the training process which also reflects the listening test:

1) The speech samples used in conjunction with the model should be the ones used in the subjective tests: 16 sentences of male and female speakers, American English.

2) The results should be calculated on a per sentence basis and averaged over all 16 samples.

3) The background noises to be used in conjunction with the model shall be taken from EG 202 396-1 [i.2].

4) The setup is according to EG 202 396-1 [i.2].

Annex A (normative): Summary of Retraining Databases

					References															
Database	Lab	Test #	BW	# Noise Types	NSLevel	Ref SNR [dB]	# Reference Conditions	# Test Conditions	# Devices	Use case	Listening Instrument	Listening Mode	Presentation Level [dBSPL]	# talkers	# samples per talker	# of listeners	# votes per sample	# votes per condition	Signals available	Contribution
1	Audience	1	NB	8 (replace Crossroads with clean speech)	Table 1	0, 12, 24, 36	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120322
2	Audience	2	NB	8 (replace Crossroads with clean speech)	Table 1	0, 12, 24, 36	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120322
3	Audience	3	WB	8 (replace Crossroads with clean speech)	Table 1	10, 20, 30, 40	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120322
4	Audience	4	WB	8 (replace Crossroads with clean speech)	Table 1	10, 20, 30, 40	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120322
5	Qualcomm/Dynastat	1	NB	6 (Pub, Road, Train, Car, Mensa, clean speech)	Table 1	0, 12, 24, 36	12	48	8	HS	HD25	diotic	73	4 (2M, 2F)	8	32	4	128	CMU_OUT, PRI_MIC_IN, MRP	S4-120375
6	Qualcomm	2	NB	8 (replace Train with clean speech)	Table 1	0, 12, 24, 36	12	48	3	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN, MRP	S4-120375
7	Qualcomm	3	NB	8 (replace Train with clean speech)	Table 1	0, 12, 24, 36	12	48	6	HS	HD280 PRO	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN, MRP	S4-120375
8	Orange SA	1	WB	5 (Car, Road, Train, Cafeteria, Office)	Table 2	10, 20, 30, 40	12	90	6	HS	HD25	monaural	79	6 (3M, 3F)	2	24	24	288	CMU_OUT, PRI_MIC_IN	SA-120348
9	Qualcomm	4	WB	8 (replace Train with clean speech)	Table 2	10, 20, 30, 40	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN, MRP	S4-120467
10	Qualcomm	5	WB	8 (replace Train with clean speech)	Table 1	10, 20, 30, 40	12	48	6	HS	HD280 PRO	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN, MRP	S4-120619
11	Audience	1A	NB	8 (noise level +6dB)	Table 1	0, 12, 24, 36	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120655
12	Audience	2A	NB	8 (noise level +6dB)	Table 1	0, 12, 24, 36	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120655
13	Audience	3A	WB	8 (noise level +6dB)	Table 1	10, 20, 30, 40	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120655
14	Audience	4A	WB	8 (noise level +6dB)	Table 1	10, 20, 30, 40	12	48	6	HS & HHHF	HD280 PRO	diotic	73	4 (2M, 2F)	2	32	16	128	CMU_OUT, PRI_MIC_IN	S4-120655
15	NOKIA Corp/Dynastat	1	WB	8	Table 1	0, 12, 24, 36	12	48	6	HS	HD25	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN	S4-120813
16	NOKIA Corp/Dynastat	2	WB	8	Table 1	0, 12, 24, 36	12	48	6	HS	HD25	diotic	73	4 (2M, 2F)	4	32	8	128	CMU_OUT, PRI_MIC_IN	S4-120813

Annex B (normative): Test vectors for model verification

The test vectors for verification of an objective model implementation are given in this annex. A model claiming to be compatible with the present document shall achieve all scores with an accuracy of ±0,1 MOS.

B.1 Audience test vectors

The validation results below are for signals used in the validation Experiments 6 (NB) and 8 (WB) as reported in clause 9. The reference conditions and noise types are as described in [i.1], but with levels of noise increased by 6 dB. The [six] devices have been tested in a mix of handset and handheld speakerphone use cases. Predictions from the model are presented at both sample and condition level. For each Experiment, 50 sample files and value sets are provided for validation of implementations of this model.

The test vectors can be downloaded here:
http://docbox.etsi.org/stq/Open/TS%20103%20106%20Wave%20files/Annex_B1_1_2_Audience%20Verification%20Data/

Table B.1: Audience experiment 6 test vectors and objective scores to be achieved by an objective model implementation

				Per sample		
Noise	**Device**	**talker**	**sample**	**SMOS**	**NMOS**	**GMOS**
cafeteria	A	m1	s1	3,21	3,05	2,87
car	A	m1	s2	3,31	3,34	3,05
crossroad	A	f1	s4	3,82	3,35	3,42
crossroad	A	f2	s3	3.08	3,29	2,88
mensa	A	f1	s5	4,09	3,19	3,57
mensa	A	m2	s4	3,65	3,39	3,31
office	A	f2	s6	4,39	3,53	3,97
pub	A	f1	s6	2,77	2,81	2,49
pub	A	m1	s6	2.92	3,04	2,68
traffic	A	f1	s7	3,09	2,82	2,69
train	A	f2	s8	3,22	3,36	3,00
cafeteria	B	f1	s1	3,56	3,19	3,16
car	B	m1	s3	3,60	3,53	3,32
crossroad	B	m1	s3	3,20	3,49	3,03
mensa	B	f1	s5	3,91	3,26	3,46
office	B	f2	s6	4,36	3,52	3.94
pub	B	m1	s6	2,80	2,45	2,32
traffic	B	m1	s8	2,15	2,31	1,93
train	B	m1	s1	2,50	3,02	2,44
cafeteria	B	f2	s2	4,06	4,09	3,85
car	B	f2	s3	3,98	4,17	3,80
mensa	B	m2	s4	4,35	3,85	4,03
office	B	m2	s6	4,27	3,78	3,94
pub	B	m2	s7	3,88	3,96	3,66
traffic	B	m2	s8	3,13	2,98	2,78
train	B	m2	s8	3,42	3,94	3,32
cafeteria	D	m2	s2	3,64	2,89	3,09
car	D	m2	s3	3,00	4,02	3,07
crossroad	D	m2	s4	3,91	3,84	3,66
mensa	D	m2	s5	3,84	4,05	3,66
office	D	m2	s6	4,38	3,72	4,02
pub	D	m2	s7	3,58	3,85	3,41
traffic	D	m2	s8	2,92	2,68	2,51
train	D	m2	s8	3,36	3,79	3,23
cafeteria	E	m2	s2	2,54	1,96	1,90
car	E	m2	s3	2,57	1,94	1,90
crossroad	E	m2	s4	2,96	1,72	1,98

				Per sample		
Noise	**Device**	**talker**	**sample**	**SMOS**	**NMOS**	**GMOS**
mensa	E	m2	s5	3,04	2,08	2,25
office	E	m2	s6	3,90	2,38	3,03
pub	E	m2	s7	1,12	1,00	1,00
traffic	E	m2	s8	1,68	1,19	1,01
train	E	m2	s8	2,89	1,37	1,68
cafeteria	F	m2	s2	3,60	2,62	2,93
car	F	m2	s3	3,94	2,61	3,19
crossroad	F	m2	s4	3,93	2,60	3,18
mensa	F	m2	s5	3,89	2,48	3,08
office	F	m2	s6	4,15	3,57	3,77
pub	F	m2	s7	2,88	1,71	1,92
traffic	F	m2	s8	1,76	1,52	1,27
train	F	m2	s8	2,43	2,04	1,89

Table B.2: Audience experiment 8 test vectors and objective scores to be achieved by an objective model implementation

				Per sample		
Noise	**Device**	**talker**	**sample**	**SMOS**	**NMOS**	**GMOS**
cafeteria	A	m2	s1	3,30	3,34	2,93
car	A	m1	s2	3,66	4,24	3,54
crossroad	A	m2	s3	3,57	3,57	3,22
mensa	A	f1	s4	3,55	3,56	3,20
mensa	A	m1	s5	3,90	3,96	3,60
office	A	f1	s5	4,09	4,18	3,83
office	A	f2	s6	4,33	4,13	3,98
pub	A	m2	s6	2,21	3,73	2,32
traffic	A	m2	s7	1,96	3,32	1,98
train	A	m1	s1	2,93	4,06	2,96
cafeteria	B	f1	s2	3,38	3,27	2,96
car	B	m2	s3	3,76	3,97	3,51
crossroad	B	m2	s3	3,52	3,58	3,19
mensa	B	f1	s5	3,66	3,65	3,31
mensa	B	m2	s4	3,65	3,83	3,38
office	B	f2	s5	3,77	3,81	3,46
office	B	m1	s5	4,36	4,00	3,95
pub	B	f2	s7	2,86	2,25	2,14
traffic	B	m1	s7	2,07	2.90	1.88
train	B	f2	s8	2,93	4,10	2,97
cafeteria	B	f2	s2	4,10	4,29	3,87
car	B	m1	s2	4,30	4,62	4,14
crossroad	B	m1	s3	4,17	4,40	3,96
mensa	B	f1	s4	4,12	4,21	3,86
office	B	f1	s6	4,39	4,54	4,17
pub	B	m1	s6	3,15	4,07	3,12
traffic	B	m2	s7	2,20	3,73	2,31
train	B	f1	s1	3,21	4,38	3,27
cafeteria	D	f2	s1	3,52	4,14	3,40
car	D	f2	s3	3,44	4,49	3,48
mensa	D	f2	s4	3,66	3,63	3,31
office	D	f1	s6	4,37	4,55	4,16
pub	D	f1	s7	2,71	3,92	2,75
traffic	D	f2	s7	2,19	3,71	2,30
train	D	m1	s8	3,65	4,37	3,58
cafeteria	E	m2	s2	3,00	2,32	2,28
car	E	f2	s3	3,24	2,20	2,40
mensa	E	f1	s5	4,15	1,91	2,90
office	E	f1	s6	3,80	2,41	2,89
pub	E	m2	s6	2,04	1,26	1,09
traffic	E	f2	s8	2,53	1,56	1,59
train	E	m2	s1	2,69	1,58	1,71
cafeteria	F	m1	s1	3,66	2,53	2,84

				Per sample		
Noise	Device	talker	sample	SMOS	NMOS	GMOS
car	F	f2	s2	3,69	2,67	2,92
mensa	F	f2	s5	3,83	2,94	3,14
office	F	m1	s6	4,30	3,19	3,58
office	F	m2	s5	4,35	3,29	3,66
pub	F	m2	s7	2,93	2,16	2,16
traffic	F	f1	s8	2,88	2,12	2,10
train	F	f1	s8	4,34	3,14	3,59

B.2 Orange test vectors

A subset of Orange validation database, comprised of the three scores [S-MOS, N-MOS, G-MOS] and the associated audio material [Clean, Noisy Input, Noise-reduced output] for each sample is provided for purposes of validation. This subset covers as much as possible the entire quality range and includes samples of conditions 2, 10, 19, 23, 26 and 30 as detailed in table B.3.

The test vectors can be downloaded here:

http://docbox.etsi.org/stq/Open/TS%20103%20106%20Wave%20files/Annex_B2_Orange%20Verification%20Data/

Table B.3: Orange test vectors and objective scores to be achieved by an objective model implementation

					Per sample			Per condition		
File name	**Noise**	**Device**	**talker**	**sample**	**S-MOS**	**N-MOS**	**G-MOS**	**S-MOS**	**N-MOS**	**G-MOS**
1bff2s01.c02	Mensa	D1	f2	s1	4,23	3,69	3,73			
1bff2s02.c02	Mensa	D1	f2	s2	3,67	4,38	3,60			
1bfm1s01.c02	Mensa	D1	m1	s1	4,23	2,93	3,42			
1bfm1s02.c02	Mensa	D1	m1	s2	4,23	4,15	3,92			
1bfm2s01.c02	Mensa	D1	m2	s1	4,36	4,43	4,11			
1bfm2s02.c02	Mensa	D1	m2	s2	4,16	3,84	3,74	4,14	3,90	3,75
1bff2s01.c10	Crossroads	D4	f2	s1	3,48	3,07	2,95			
1bff2s02.c10	Crossroads	D4	f2	s2	3,99	3,66	3,55			
1bfm1s01.c10	Crossroads	D4	m1	s1	3,60	2,84	2,94			
1bfm1s02.c10	Crossroads	D4	m1	s2	3,76	3,41	3,29			
1bfm2s01.c10	Crossroads	D4	m2	s1	4,02	2,70	3,17			
1bfm2s02.c10	Crossroads	D4	m2	s2	3,96	2,94	3,23	3,80	3,10	3,19
1bff2s01.c19	No noise	Source	f2	s1	4,78	4,62	4,48			
1bff2s02.c19	No noise	Source	f2	s2	4,80	4,48	4,44			
1bfm1s01.c19	No noise	Source	m1	s1	4,77	3,49	4,04			
1bfm1s02.c19	No noise	Source	m1	s2	4,78	4,49	4,43			
1bfm2s01.c19	No noise	Source	m2	s1	4,81	4,63	4,50			
1bfm2s02.c19	No noise	Source	m2	s2	4,79	4,36	4,39	4,79	4,34	4,38
1bff2s01.c23	No noise	NS Level 1 1LeLevel 1	f2	s1	2,75	4,61	3,03			
1bff2s02.c23	No noise	NS Level 1	f2	s2	2,72	4,69	3,04			
1bfm1s01.c23	No noise	NS Level 1	m1	s1	2,86	4,35	3,02			
1bfm1s02.c23	No noise	NS Level 1	m1	s2	2,70	4,56	2,98			
1bfm2s01.c23	No noise	NS Level 1	m2	s1	2,31	4,58	2,71			
1bfm2s02.c23	No noise	NS Level 1	m2	s2	2,73	4,67	3,04	2,68	4,58	2,97
1bff2s01.c26	Crossroads	Source	f2	s1	4,71	2,33	3,50			
1bff2s02.c26	Crossroads	Source	f2	s2	4,74	2,24	3,48			
1bfm1s01.c26	Crossroads	Source	m1	s1	4,70	2,21	3,44			
1bfm1s02.c26	Crossroads	Source	m1	s2	4,68	2,20	3,41			
1bfm2s01.c26	Crossroads	Source	m2	s1	4,69	2,30	3,47			
1bfm2s02.c26	Crossroads	Source	m2	s2	4,70	2,33	3,49	4,70	2,27	3,46
1bff2s01.c30	Crossroads	NS Level 1	f2	s1	3,56	1,74	2,41			
1bff2s02.c30	Crossroads	NS Level 1	f2	s2	3,31	1,80	2,26			
1bfm1s01.c30	Crossroads	NS Level 1	m1	s1	2,90	1,76	1,95			
1bfm1s02.c30	Crossroads	NS Level 1	m1	s2	3,34	1,78	2,27			
1bfm2s01.c30	Crossroads	NS Level 1	m2	s1	2,24	1,73	1,46			
1bfm2s02.c30	Crossroads	NS Level 1	m2	s2	3,10	1,88	2,14	3,08	1,78	2,08

Annex C (normative): Speech material to be used for objective testing

The following speech samples are used in conjunction with the model: 4 talkers (2 Males/2 Females), 8 Harvard sentences per talker, each sample is 4 sec duration.

The first 4 sentences are used during the adaptation period of the noise canceller under test, the remaining 16 samples are used for calculating the objective scores.

The speech samples can be downloaded here:

http://docbox.etsi.org/stq/Open/TS%20103%20106%20Wave%20files/Annex_C_Dynastat%20Speech%20Data/

Seq	Sample	Harvard Sentence	
1	m1s8	We tried to replace the coin but failed.	Preliminary (convergence)
2	f1s8	A rod is used to catch pink salmon.	
3	m2s8	Corn cobs can be used to kindle a fire.	
4	f2s8	The crooked maze failed to fool the mouse.	
5	m1s1	The empty flask stood on the tin tray.	
6	f1s1	It is easy to tell the depth of a well.	
7	m2s1	Acid burns holes in wool cloth.	
8	f2s1	Note closely the size of the gas tank.	
9	m1s2	He broke a new shoelace that day.	
10	f1s2	The box was thrown beside the parked truck.	
11	m2s2	Eight miles of woodland burned to waste.	
12	f2s2	Mend the coat before you go out.	
13	m1s3	The urge to write short stories is rare.	
14	f1s3	Four hours of steady work faced us.	
15	m2s3	A young child should not suffer fright.	
16	f2s3	The stray cat gave birth to kittens.	
17	m1s4	The pirates seized the crew of the lost ship.	
18	f1s4	The boy was there when the sun rose.	
19	m2s4	The fruit of a fig tree is apple shaped.	
20	f2s4	The frosty air passed through the coat.	

History

Document history		
V1.1.1	August 2012	Publication