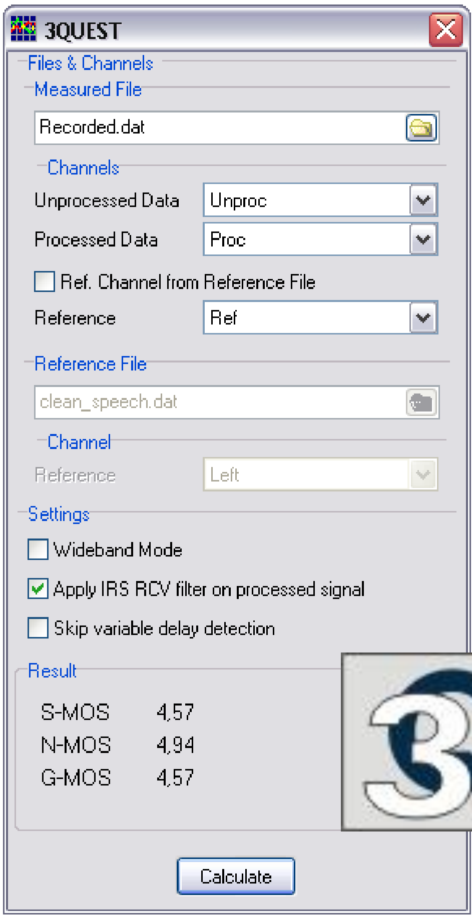
3QUEST: 3-fold Quality Evaluation of Speech in Telecommunications



HEAD acoustics
Application Note

HEAD acoustics GmbH
Ebertstraße 30a
D-52134 Herzogenrath
Tel: +49 (0) 2407-577-0
Fax: +49 (0) 2407-577-99
E-mail: telecom@head-acoustics.de
WEB: www.head-acoustics.de

3QUEST: 3-fold Quality Evaluation of Speech in Telecommunications

Contents

1. Introduction

Modern wideband communication systems like mobile phones or hands-free terminals are increasingly used in the presence of background noise. To improve the signal-to-noise ratio, the speech recorded at the terminal is often passed through noise reduction algorithms with non-linear and time-variant processing. However, such algorithms may also audibly degrade the speech quality of the transmitted signal, particularly when the background noise is time-variant or non-stationary. To judge the influence of speech processing algorithms, subjective testing according to ITU-T Recommendation P.835 is required to subjectively determine the mean opinion scores (MOS) of speech, noise and overall quality of a sample.

Based on the Relative Approach algorithm, we introduce the model described in the ETSI standard EG 202 396-3 for objectively measuring the quality of wide- and narrowband speech in noisy environments, which provides a high correlation with the subjective MOS.

This objective measurement was developed after the consecutive analysis of expert listeners, where parameters of degradation decomposition were extracted by modeling the behavior of test persons in listening test.

2. Listening Tests and Database

The quality of processed and transmitted speech in the presence of background noise is of great importance in today's communication systems. Consequently, it is highly desirable to optimize the speech quality of such systems based on objective testing methods. However, any objective model has to provide high correlation to the quality perceived subjectively. Within the ETSI STF 294 project (sponsored by eEurope [1], [2], [3]), a database including a big variety of wideband speech samples was created and subjectively evaluated based on ITU-T Recommendation P.835 [4].

These data formed the basis of a new model for predicting speech, noise and overall quality. The output of this algorithm provides three MOS (Mean Opinion Score) values for speech, noise transmission and overall quality. The test setup and the determination of these scores are different from the typical test procedure according to ITU-T Recommendation P.800 [5]. ITU-T Rec. P.835 focuses on the problem of speech quality in the presence of background noise in a more diagnostic way than P.800. The subjective scores derived from tests according to ITU-T Rec. P.835 clearly show the influence of modern noise cancelling techniques on the quality of the transmitted noise as well as on the quality of the transmitted speech. Both may impair the perceived overall speech quality. Therefore a more detailed questionnaire as well as more specific scales are used in the tests as shown in table 1.

Determination of subjective speech MOS (S-MOS)	**Determination of subjective noise MOS (N-NOS)**	**Determination of subjective global MOS (G-MOS)**
Attending ONLY to the SPEECH SIGNAL, select the category which best describes the sample you just heard. The SPEECH SIGNAL in this sample was 5 – NOT DISTORTED 4 – SLIGHTLY DISTORTED 3 – SOMEWHAT DISTORTED 2 – FAIRLY DISTORTED 1 – VERY DISTORTED	Attending ONLY to the BACKGROUND, select the category which best describes the sample you just heard. The BACKGROUND in this sample was 5 – NOT NOTICEABLE 4 – SLIGHTLY NOTICEABLE 3 – NOTICEABLE BUT NOT INTRUSIVE 2 – SOMEWHAT INTRUSIVE 1 – VERY INTRUSIVE	Select the category which best describes the sample you just heard for purposes of everyday speech communication. The OVERALL SPEECH SAMPLE was 5 – EXCELLENT 4 – GOOD 3 – FAIR 2 – POOR 1 – BAD

Table 1: Instructions and scales acc. to ITU-T P.835

Furthermore, we introduce an extension of this model which includes narrowband scenarios. For this purpose, a new database of narrowband speech samples was created and subjectively rated according to ITU-T Recommendation P.835.

For both narrowband and wideband scenarios the following ratings were derived from subjective tests:

- Noise-MOS (N-MOS);
- Speech-MOS (S-MOS);
- Global-MOS (G-MOS), the overall quality including speech and background noise.

These ratings are predicted by the objective model 3QUEST.

Different input signals are accessed during the recording process and subsequently can be used for the calculation of N-MOS, S-MOS and G-MOS (see figure 1). Beside the signal presented in the listening test (processed signal p(k), recorded in sending direction), two additional signals are used as a priori knowledge for the calculation:

- The clean speech signal c(k), which is played back via a HATS / by the speaker at the beginning of the sample generation process.
- The unprocessed signal u(k), which is recorded close to the microphone position of the handset device / hands-free telephone. Also the input signal of the terminal's microphone can be used if available. It represents the most "natural" signal which can be transmitted.

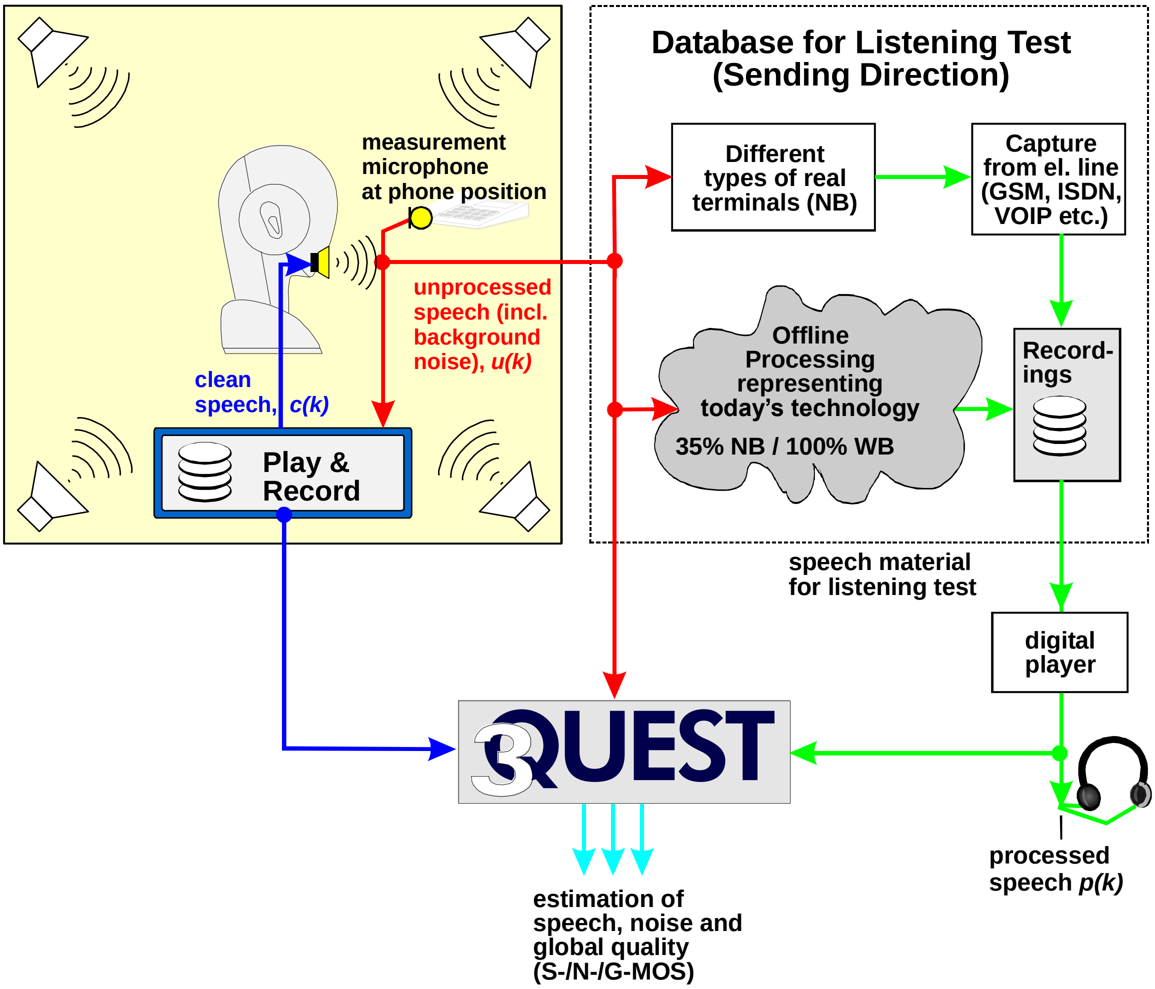


Figure 1: Recordings for listening tests acc. to ITU-T P.835 / Training of 3QUEST Algorithm

2.1. Description of Databases

The output scores of objective models for the prediction of speech quality are always related to subjective listening-only tests. Each pair of processed speech and its corresponding subjectively determined MOS values is called condition. All conditions taken from a single listening test are named as a database. The reference signals c(k) and u(k) for each condition are not band-limited and are included in the databases.

Due to the lack of freely available databases containing narrow- or wideband speech and evaluated according to ITU-T Recommendation P.835, new databases had to be created.

2.1.1. ETSI Wideband Database

The database in the ETSI STF 294 project was created in French. Overall, a male and a female speaker were used, one condition included one speaker each. For the creation of the model 179 conditions were used for training and 81 unknown conditions were used for validation. The background noises according to the table shown below were included:

Background noises	Handset	Hands-free
Car	23	22
Crossroads	18	18
Road	25	0
Office	27	23
Pub/Café	23	0
Overall	116	63

Table 2: Distribution of background noises in the ETSI STF294 database

The processing of the degraded speech files consisted of different VADs, noise suppression algorithms, network/packet loss scenarios and handset/hands-free modes. The band limitation of the processing was applied between 135Hz and 7 kHz [2]. After this processing step, the speech files were calibrated to an active speech level (ASL, ITU-T P.56) of -15 dB Pa (79 dB SPL) for the listening test performed diotically.

2.1.2. HEAD acoustics Narrowband Database

The new narrowband database provided by HEAD acoustics includes a wide variety of different impairments found in today’s communication systems including mobile and stationary handset/hands-free terminals. The following background noises were used for the recordings:

Background noises	Handset	Hands-free
Car	40	25
Crossroads	36	8
Road	43	9
Office	39	13
Pub/Café	42	10
Overall	200	66

Table 3: Distribution of background noises in the HEAD acoustics database

In the database, two sets of English speakers were used. A set consists of two male and two female speakers, with two sentences each. In each condition, one of these two sets was applied. The degradations of the speech samples were produced by noise reduction algorithms and different types of speech coding algorithms. Due to the given narrow speech bandwidth, the conditions were calibrated to an ASL (speech sequences only) of -21 dB Pa (73 dB SPL, ITU-T compliant level) and were also played back diotically during the listening test.

3. Technical Description of the 3QUEST Algorithm

In this chapter, the 3QUEST algorithm is introduced. 3QUEST is based on the “Relative Approach” algorithm, which is a patented development by HEAD acoustics. Because of the importance of this analysis, a brief description of the “Relative Approach” is given first.

3.1. The Relative Approach

The Relative Approach [6] is an analysis method developed to model a major characteristic of human hearing. This characteristic is the much stronger subjective response to distinct patterns (tones and/or relatively rapid time-varying structure) than to slowly changing levels and loudnesses. The Relative Approach analysis is based on the assumption that the human ear creates a running reference sound (an “anchor signal”) for its automatic recognition process against which it classifies tonal or temporal pattern information moment-by-moment. It evaluates the difference between the instantaneous and the estimated patterns in both time and frequency. Temporal structures and spectral patterns are important factors in deciding whether a sound is judged as annoying or disturbing [7], [8], [9].

Similar to human hearing and in contrast to other analysis methods the Relative Approach algorithm does not require any reference signal for the calculation. Comparable to the human experience and expectation, the algorithm generates an “internal reference” which can best be described as a forward estimation. The Relative Approach algorithm objectifies pattern(s) in accordance with human perception by resolving or extracting them while largely rejecting pseudostationary energy. Figure 2 shows a block diagram of the Relative Approach.

The time-dependent spectral pre-processing can either be done by a filter bank analysis or a spectral analysis based on a Hearing Model [12]. Both of them result in a spectral representation versus time.

The Relative Approach takes the absolute signal level into account. Therefore, the input data must be calibrated to a realistic listening level. Two variants of the Relative Approach can be applied to the pre-processed signal. One applies a regression versus time for each frequency band, afterwards for each time slot a smoothing versus frequency is performed. The next step is a non-linear transformation according to the Hearing Model of Sottek [12]. This output is compared to the source signal. This variant focuses on the detection of tonal components.

The second variant first smoothes versus frequency within a time slot and then applies the regression versus time. This output signal is again transformed non-linearly to the Hearing Model and compared to the output of the Hearing Model processed with smoothing versus frequency only.

Finally non-relevant components (for human hearing) are again set to zero. Thus more transient structures are detected. In general, the factors λ_1 and λ_2 describe the weighting of the Relative Approach for tonal and transient signals. For the new model $\lambda_1 = 0$ and $\lambda_2 = 0$ was chosen. Thus, the model is tuned to detect time-variant transient structures.

The result of the Relative Approach analysis is a 3D spectrograph displaying the deviation from the “close to human expectation” between the estimated and the current signal. Due to the nonlinear relationship between sound pressure and perceived loudness, the term “compressed pressure” in compressed Pascal (*cPa*) is used to scale the results.

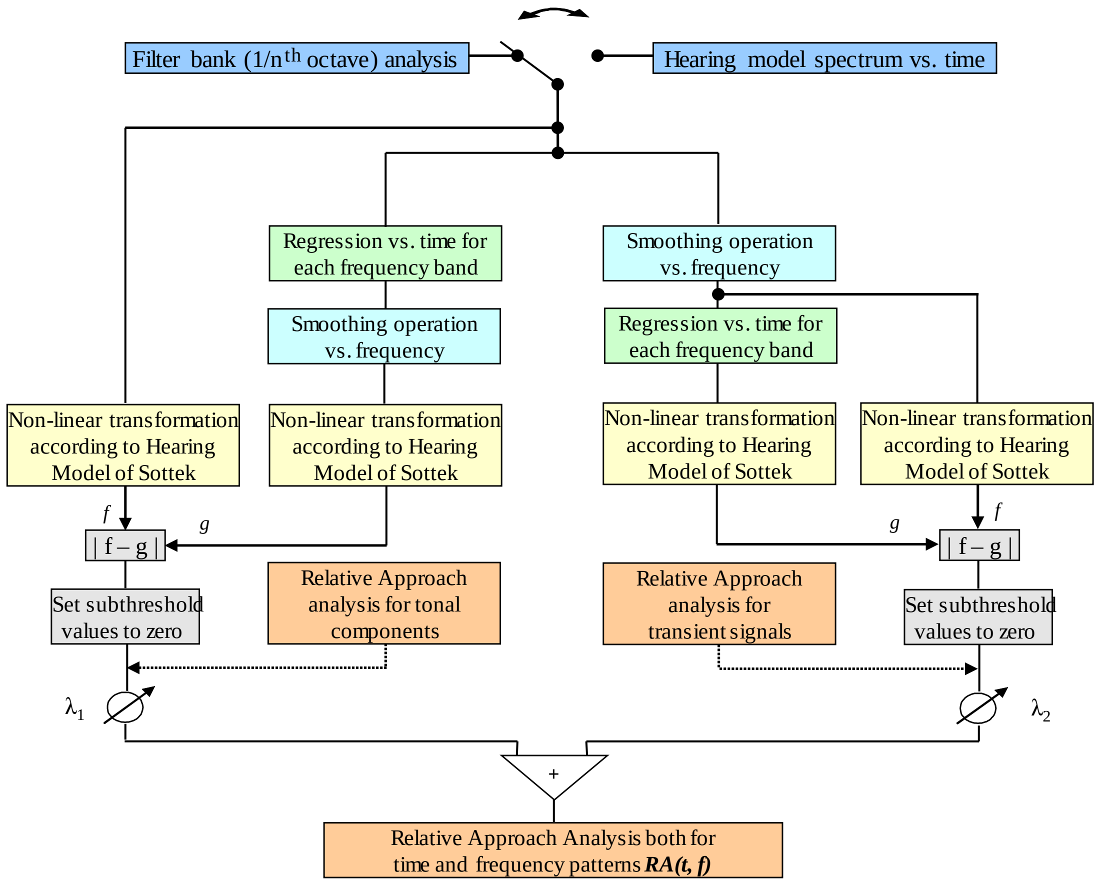


Figure 2: Block diagram of Relative Approach

A first attempt using the Relative Approach for analyzing time variant background noises was submitted as a contribution in ITU-T 2001 [10]. For time variant signals this “estimation error” can best be interpreted as the “attention” which is attracted by the patterns of the particular signal on human perception. For a consistent notation, the 3D Relative Approach representation is specified as $RA_P(t, f)$ for the processed signal, $RA_U(t, f)$ for the unprocessed signal and $RA_C(t, f)$ for the clean speech.

3.1.1. Delta Relative Approach

In addition, the human a priori knowledge about “good” sound quality for time-variant background noise and speech signals needs to be considered. Therefore the 3D Relative Approach spectrograph is calculated for a processed as well as for an unprocessed signal.

Both spectrographs can be subtracted from each other in order to determine the changes caused by the transmission. This differential analysis (Relative Approach between transmitted processed signal and undisturbed unprocessed signal) provides information on how “close to human expectation” the processed signal still is in comparison with the unprocessed signal. The calculation for this example is carried out using eq. (1):

$$\Delta RA_{P-U}(t_i, f_j) = RA_p(t_i, f_j) - RA_u(t_i, f_j) \quad \forall\, t_i, f_j \tag{1}$$

3.2. CALCULATION OF N-/S-/G-MOS

A brief description of the algorithm is given below. To determine noise, speech and overall quality, several parameters must be extracted from the signals and the Relative Approach spectrograph. A more detailed description can be found in [3].

Before the three input signals for (p(k), u(k) and c(k)) are used for calculations, some preparations have to be made which also may depend on the operation mode (wide- or narrowband). A flow chart of the complete calculation algorithm is shown in figure 3. The different preparation and calculation blocks are described in the following chapters.

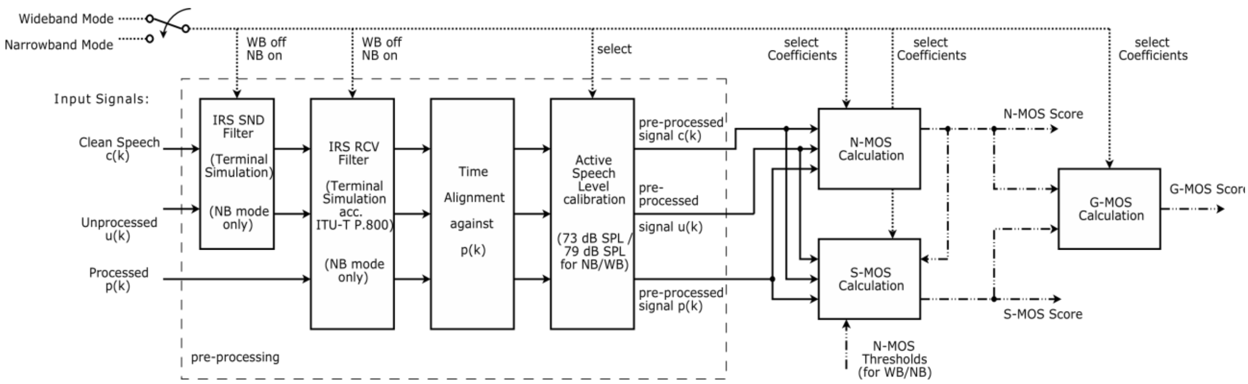


Figure 3: Flow Chart of complete 3QUEST algorithm

3.2.1. Preprocessing steps

3.2.1.1. Filtering

For the narrowband mode, the clean speech and the unprocessed signal are filtered with an intermediate reference system (IRS ITU-T P.830) in sending and receiving direction. With this preprocessing, all following analyses refer to a perfect transmission over a typical narrowband telephony network. The processed signal here is only filtered with an IRS RCV, because it was captured from the network access and therefore already includes the influence of the sending terminal.

In wideband mode, both signals are unfiltered, because here the intermediate reference system for sending and receiving was assumed as a flat transfer function between 50 Hz and 8 kHz.

3.2.1.2. Time Alignment

For wideband as well as for the narrowband modes, a time alignment must be applied. With an envelop analysis of the cross-correlation, the clean speech c(k) and the unprocessed signal u(k) are aligned against the processed signal p(k) to compensate delays.

3.2.1.3. Division into Speech Parts

For both wideband and narrowband scenarios, the clean speech signal c(k) is used to detect the speech parts. With a threshold decision in a smoothed level-versus-time representation, a nearly perfect voice activity detection (VAD) can be realized very easily. Since the signals are time-aligned, u(k) and p(k) can also be separated into parts containing either background noise or speech. All scalars, signals and spectrographs referring to parts of the signal with background noise are indexed with *BGN*. When referring to signal parts of speech, the variables are indexed with *Sp*.

3.2.1.4. Active Speech Level adjustment

After the filtering and time alignment steps, all signals (also the clean and unprocessed signal) are calibrated to a special active speech level (ASL). The signals are scaled to the listening level as it was presented in the listening tests.

The wideband listening test is applied with an active speech level of 79 dB_{SPL}, in narrowband mode 73 dB_{SPL} are used.

It is assured that only the speech parts are affected by this speech level calibration.

3.2.2. Objective N-MOS and prediction results

The objective N-MOS algorithm is based on subjective listening tests and conclusions drawn from a consecutive expert listening analysis. This analysis led to the conclusion that the subjective N-MOS is affected by parameters such as background noise level, modulation / ”naturalness“ of the background noise (e.g. musical tones) and interruptions / lost packets (minor influence).

The level of the background noise N_{BGN} (in dB) is a significant parameter which influences noise quality. It is given in (2):

$$N'_{BGN} = \frac{1}{K} \sum_k p_{BGN}^2(k)$$

$$N_{BGN} = 10 \cdot \log\left(\frac{N'_{BGN}}{1Pa}\right) \tag{2}$$

where k are the sample bins during the background noise sections of the processed signal p(k).

Next, the 3D Relative Approach spectrograph is calculated for the complete unprocessed signal u(k) and processed signal p(k), resulting in $RA_u(t, f)$ and $RA_p(t, f)$. In these spectrographs, sections containing background noises are extracted using the above-mentioned above (with clean speech as perfect VAD). The marked time bins lead to the spectrograph parts $RA_{BGN,p}(t, f)$ and $RA_{BGN,u}(t, f)$. The mean μ (RA_{BGN}) and variance σ² (RA_{BGN}) of a Relative Approach spectrograph, which describe audible effects like annoying sounds and/or musical tones resulting from noise suppression algorithms and processing in general, are calculated with (3) and (4).

$$\mu = \frac{1}{A} \cdot \sum_{t_i = t_{\min}}^{t_{\max}} \sum_{\Delta f_m = \Delta f_{\min}}^{\Delta f_{\max}} RA_{BGN}(t_i, f_m) \cdot dA_j(\Delta f_m) \tag{3}$$

and

$$\sigma^2 = \left(\frac{1}{A} \cdot \sum_{t_i = t_{\min}}^{t_{\max}} \sum_{\Delta f_m = \Delta f_{\min}}^{\Delta f_{\max}} RA_{BGN}^2(t_i, f_m) \cdot dA_j(\Delta f_m)\right) - \mu^2 \tag{4}$$

with:

$$A_{ges} = \frac{1}{(t_{\max} - t_{\min})(f_{\max} - f_{\min})},$$

$$dA_j(\Delta f_m) = \Delta t \cdot \Delta f_m,$$

Δt = 6,66 ms
$\Delta f_m \neq$ constant (1/12th octave frequency band resolution)
f_{min} = 50 Hz, lower frequency of band Δf_{min},
f_{max} = 8 kHz, upper frequency of band Δf_{max},
f_m - centre frequency of band Δf_m
t_{min} and t_{max} given by the background noise section extracted before.

These parameters are calculated for the background noise sections of the 3D Relative Approach spectrographs of the processed ($RA_P(t, f)$), unprocessed ($RA_U(t, f)$) and the difference of processed and unprocessed signal ($\Delta RA_{P-U}(t, f)$). Finally, the objective N-MOS is the result of a linear, quadratic regression algorithm applied to all six parameters (see table 4) according to eq. (5)

$$NMOS = c_0 + \sum_{j=1}^{2} \sum_{i=1}^{6} c_{ji} \cdot P_i^j \qquad (5)$$

with c_0, c_{ji} : weights for each parameter P_i and regression order j

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 4: Parameters for N-MOS regression

The coefficients for the weighting of all parameters are extracted from a linear quadratic regression with the subjective scores of each listening test. In figure 4, the algorithm for the determination of the N-MOS is summarized in a flow diagram.

The N-MOS prediction results for the training data for narrow- and wideband mode are shown in figure 5.

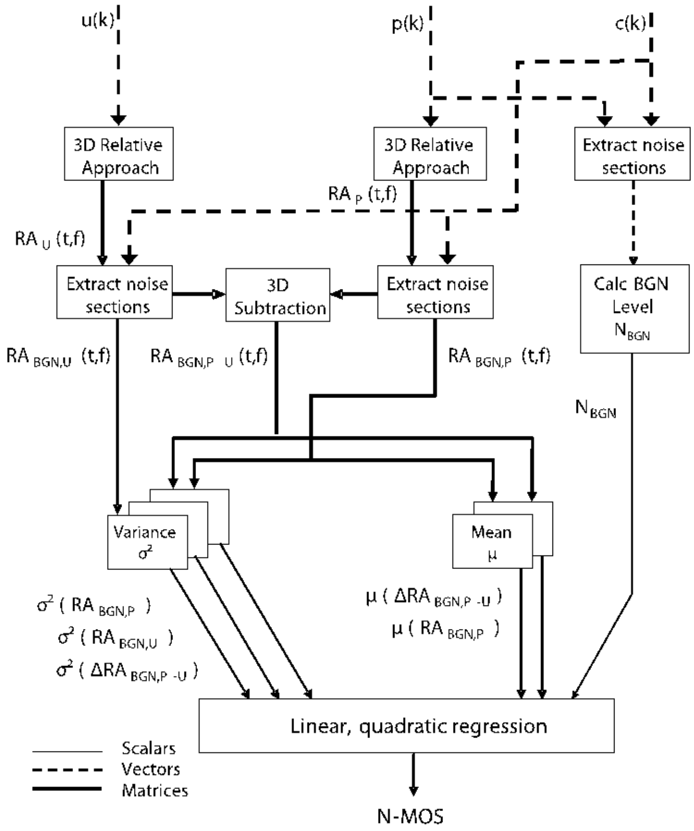


Figure 4: Block diagram of Objective N-MOS Calculation

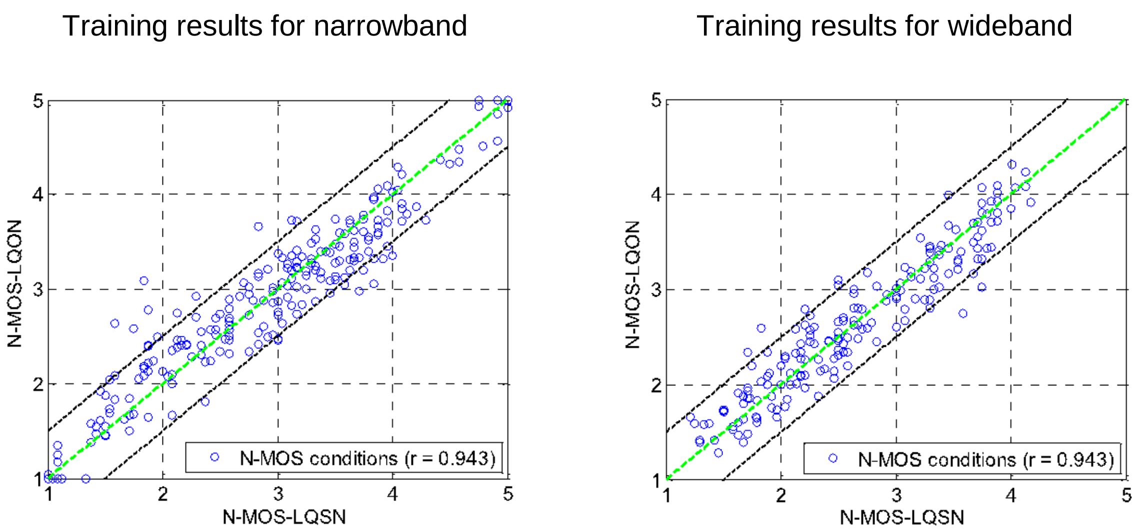


Figure 5: Training Prediction Results for N-MOS

3.2.3. Objective S-MOS and prediction results

The objective S-MOS also aims to reproduce the listening impression of the test persons in the listening test in order to provide a high correlation with the given database and also a high robustness for other databases. Various parameters were found to be relevant for the subjective S-MOS: Level and quality of processed background noise, SNR and improvement (or impairment) of SNR (between unprocessed and processed signal), interrupted or modulated sounding speech and the natural sound impression of the speech.

Similar to the N-MOS calculation, the S-MOS algorithm is also designed to reproduce the above-mentioned parameters. The difference between the SNR of the unprocessed and the processed signal (DSNR) is one of the extracted parameters. It is determined by considering the energies in the speech (Sp) and background noise (BGN) parts according to eq. (2) and (7).

$$SNR = 10 \cdot \log\left(\frac{(S+N)'_{SP} - N'_{BGN}}{N'_{BGN}}\right) \tag{7}$$

The determination of SNRU is done likewise. The ΔSNR is then given in (12):

$$\Delta SNR = SNR_p - SNR_u \tag{8}$$

In order to cover the influence of signal processing on the sound of the transmitted signal, the modulation and “naturalness” (potentially impaired e.g. by noise reduction algorithms) the Relative Approach and the Δ Relative Approach are used.

Equivalent to eq. (3) and (4), mean and variance of the Relative Approach spectrographs of $RA_{Sp,\ P}$, $\Delta RA_{BGN,\ P\text{-}C}$, $\Delta RA_{BGN,\ P\text{-}U}$ within the speech parts can be determined. Again, this results in six parameters P_i for a linear, quadratic regression (compare table 5):

A seventh indirect input parameter for the regression is the N-MOS. Test persons tend to expect high quality speech if the background noise sounds pleasant at the beginning of a 3D sample.

Vice versa: if the background noise sounds unpleasant, the speech sound is also expected to be impaired. For the determination of the S-MOS, this continuous weighting of N-MOS is “quantized” into three ranges:

- High N-MOS → high speech quality expected ($N\text{-}MOS > N\text{-}MOS_{high}$).
- Average N-MOS → several influences need to be considered ($N\text{-}MOS_{low} \leq N\text{-}MOS \leq N\text{-}MOS_{high}$)
- Low N-MOS → low speech quality expected ($N\text{-}MOS < N\text{-}MOS_{low}$).

Depending on the N-MOS of a condition, the parameters Pi are more or less important. To map this dependency of the N-MOS in the calculation of the S-MOS, for each interval a different set of weighting coefficients for the regression is chosen. The determination of the S-MOS is given in (9).

$$SMOS = c_{R,0} + \sum_{j=1}^{2}\sum_{i=1}^{6} c_{R,i,j} \cdot P_n^j \tag{9}$$

with $c_{R,0}, c_{R,i,j}$: weighting coefficients for parameters, extracted with linear quadratic regression, extracted from subjective data and R = 1,2,3 : N-MOS interval index (low, mid, high).

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

Table 5: Parameters for S-MOS regression

The best fitting values for N-MOS_{low} and N-MOS_{high} can also be extracted from the results of the listening tests. To achieve a uniform regression when mapping the parameters to the subjective ratings, the amount of conditions in each N-MOS interval should be equal. A flow diagram of the complete algorithm is shown in figure 6. The S-MOS prediction results for the training data for narrow- and wideband mode are shown in figure 7.

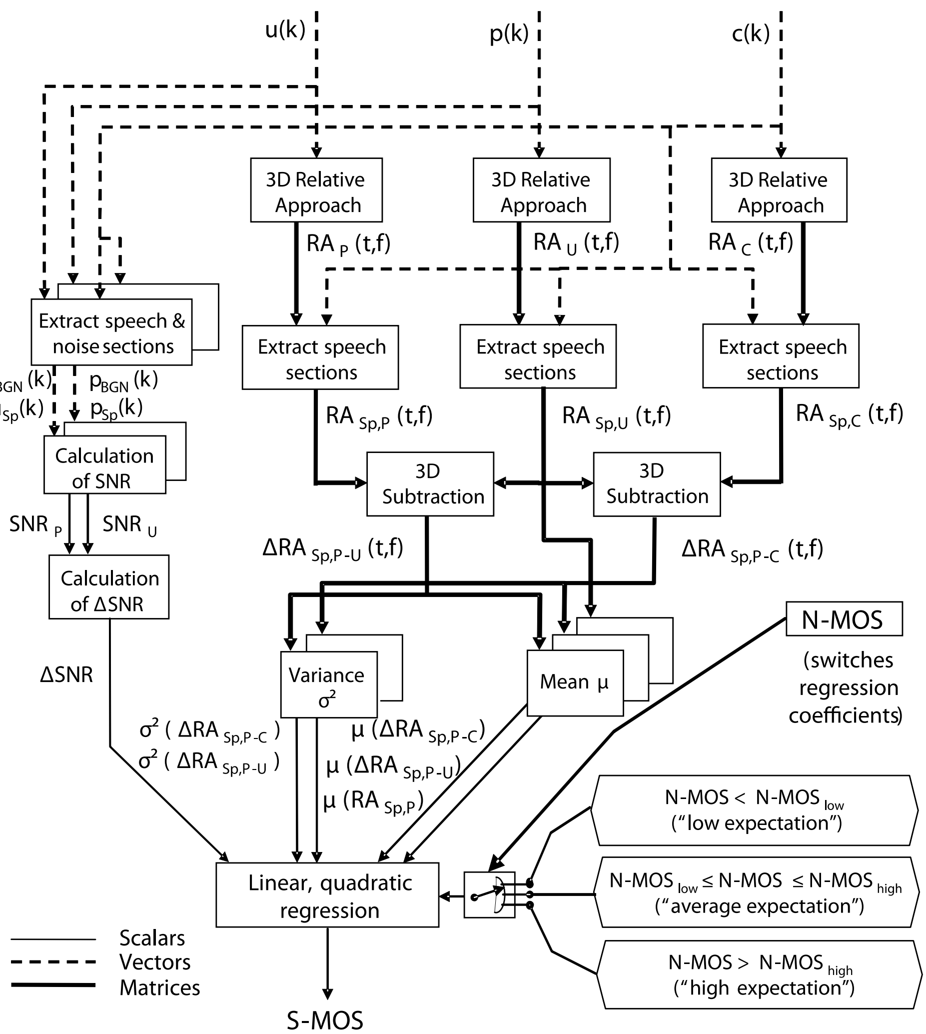
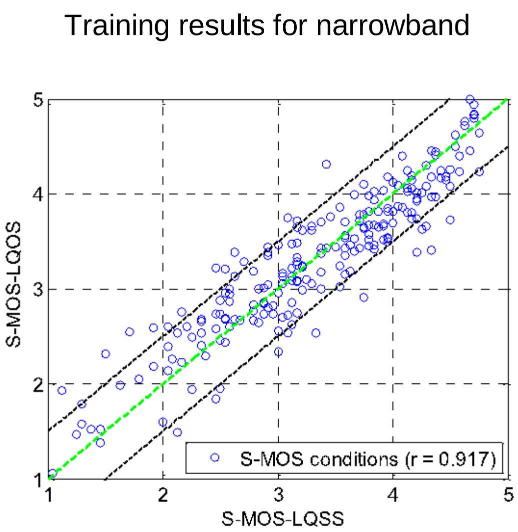


Figure 6: Block diagram of Objective S-MOS Calculation



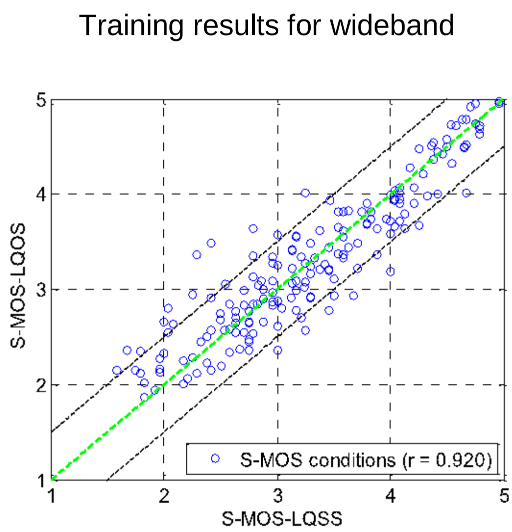


Figure 7: Training Prediction Results for S-MOS

3.2.4. Objective G-MOS

The overall or global quality G-MOS can best be calculated by using the previously calculated N-MOS and S-MOS as input parameters for a linear quadratic regression. Subjects combine speech and noise quality to a “global” overall quality. The N-MOS and S-MOS algorithms consider all perceptual influences, thus they are the only input parameters for the G-MOS algorithm. The objectively determined G-MOS then results in (10).

$$GMOS = c_0 + \sum_{j=1}^{2} c_{Sj} \cdot SMOS^{\,j} + \sum_{j=1}^{2} c_{Nj} \cdot NMOS^{\,j} \qquad (10)$$

with c_0, $c_{S,j}$, $c_{N,j}$: weights of parameters, extracted with linear regression from subjective data.

The G-MOS prediction results for the training data for narrow- and wideband mode are shown in figure 8.

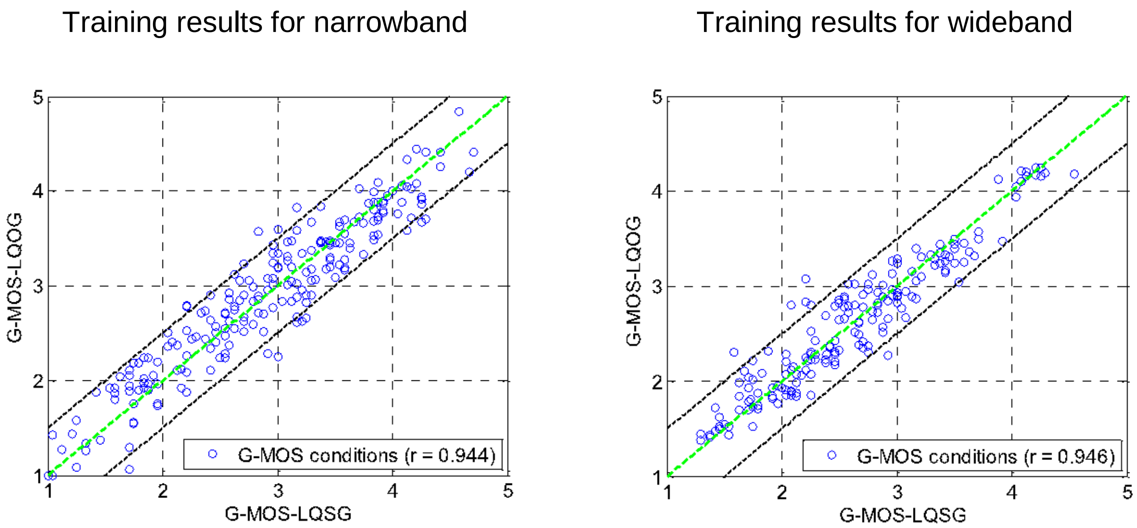


Figure 8: Training Prediction Results for G-MOS

3.3. Validation of the model

To verify the reliability and robustness of the 3QUEST algorithm, a validation was carried out for both operation modes (narrow- and wideband). For this analysis, a certain amount of conditions were retained and were unknown to the algorithm within the training phase.

In the ETSI STF294 project (wideband mode), these validation data were retained (and later on checked) by an external project partner, so that it was guaranteed these data were kept unknown in the training phase.

As mentioned before, the training database included 179 conditions. In addition, 81 validation conditions were available to test the model. For these unknown data, the measurement also yielded high correlation with the subjective MOS from the listening test. The comparison metrics are shown in table 6, the corresponding scatter plots in figure 9.

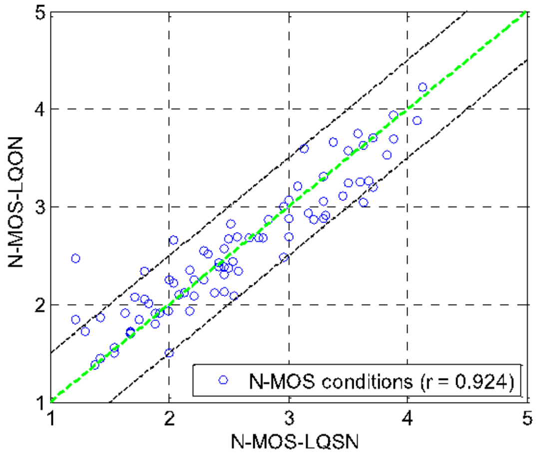


Figure 9a: N-MOS Validation results for wideband mode

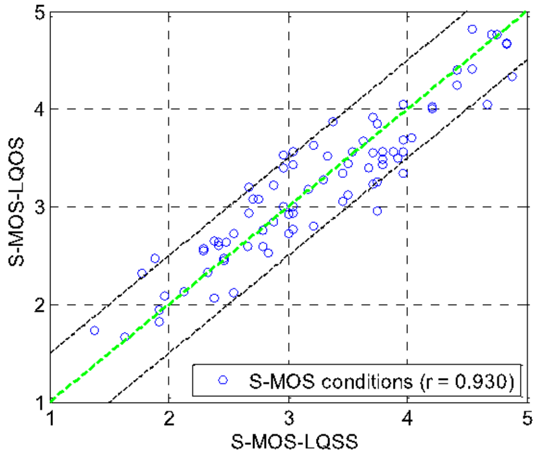


Figure 9b: S-MOS Validation results for wideband mode

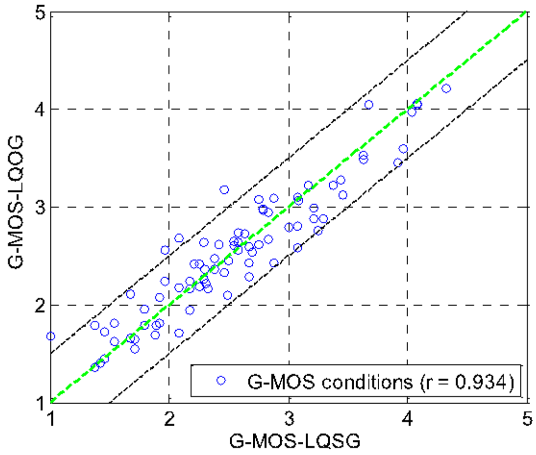


Figure 9c: G-MOS Validation results for wideband mode

Comparison metrics		
	Correlation	**RMSE**
S-MOS	93.0%	0.33
N-MOS	92.4%	0.32
G-MOS	93.4%	0.28

Table 6: Metrics for wideband validation

The narrowband extension of the model was developed by HEAD acoustics, including the newly created listening test database. For this purpose, the validation process was slightly modified: The database originally included 263 conditions, but only 213 were used to train the model. The remaining 50 conditions were randomly chosen and were tested to the already trained model.

To avoid statistical outliers, this process was done with permuted divisions between validation and training conditions and over a large amount of iterations (>500). The mean of the correlation coefficients and RMSE are given in table 6, the scatter plots of an average example is given in figure 10.

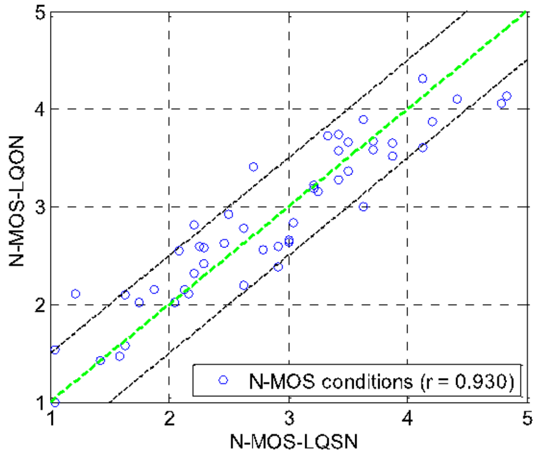


Figure 10a: N-MOS Validation results for narrowband mode

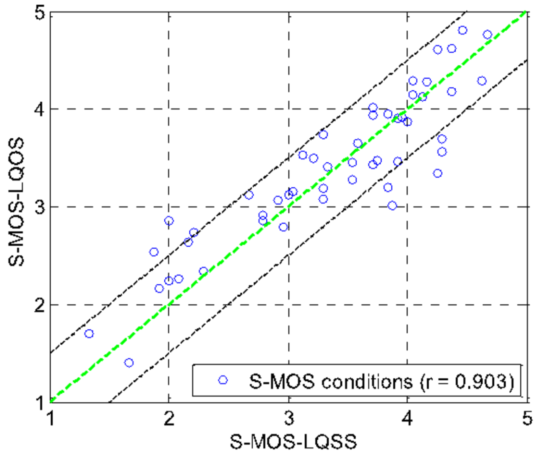


Figure 10b: N-MOS Validation results for narrowband mode

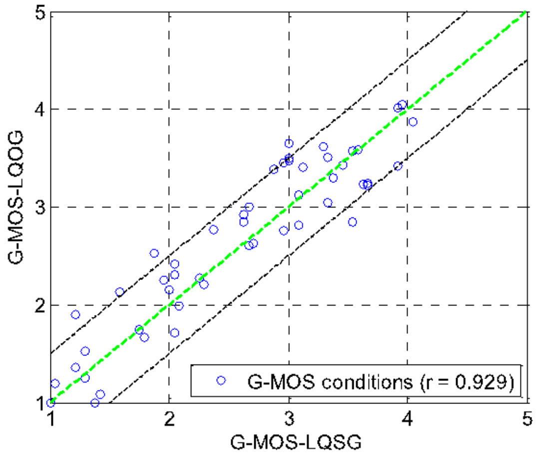


Figure 10c: N-MOS Validation results for narrowband mode

Comparison metrics		
	Correlation	**RMSE**
S-MOS	90.0%	0.37
N-MOS	93.5%	0.35
G-MOS	93.2%	0.36

Table 7: Metrics for narrowband validation

An additional validation was carried out in cooperation with France Telecom in the context of an ETSI STQ meeting [11]. The algorithm was tested by an independent lab with unknown narrowband databases. These unknown databases included several noise reduction systems applied on different noises and included English and French speakers (male and female). With these database deviations the results also showed a good performance.

4. Typical Applications for 3QUEST Measurements and ACQUA Configurations

For acquisition of the required signals for the 3QUEST algorithm the communication analysis system ACQUA is recommended in combination with the background noise simulation systems HAE-car or HAE-BGN.

The processed signal of the DUT (device under test) can be recorded electrically by ACQUA via several types of network access (e.g. radio tester). The unprocessed signal is recorded acoustically with a measurement microphone at the position of the DUT microphone.

Typical example configurations for 3QUEST measurements of car hands-free terminals and mobile phones are shown below in figures 11 and 12.

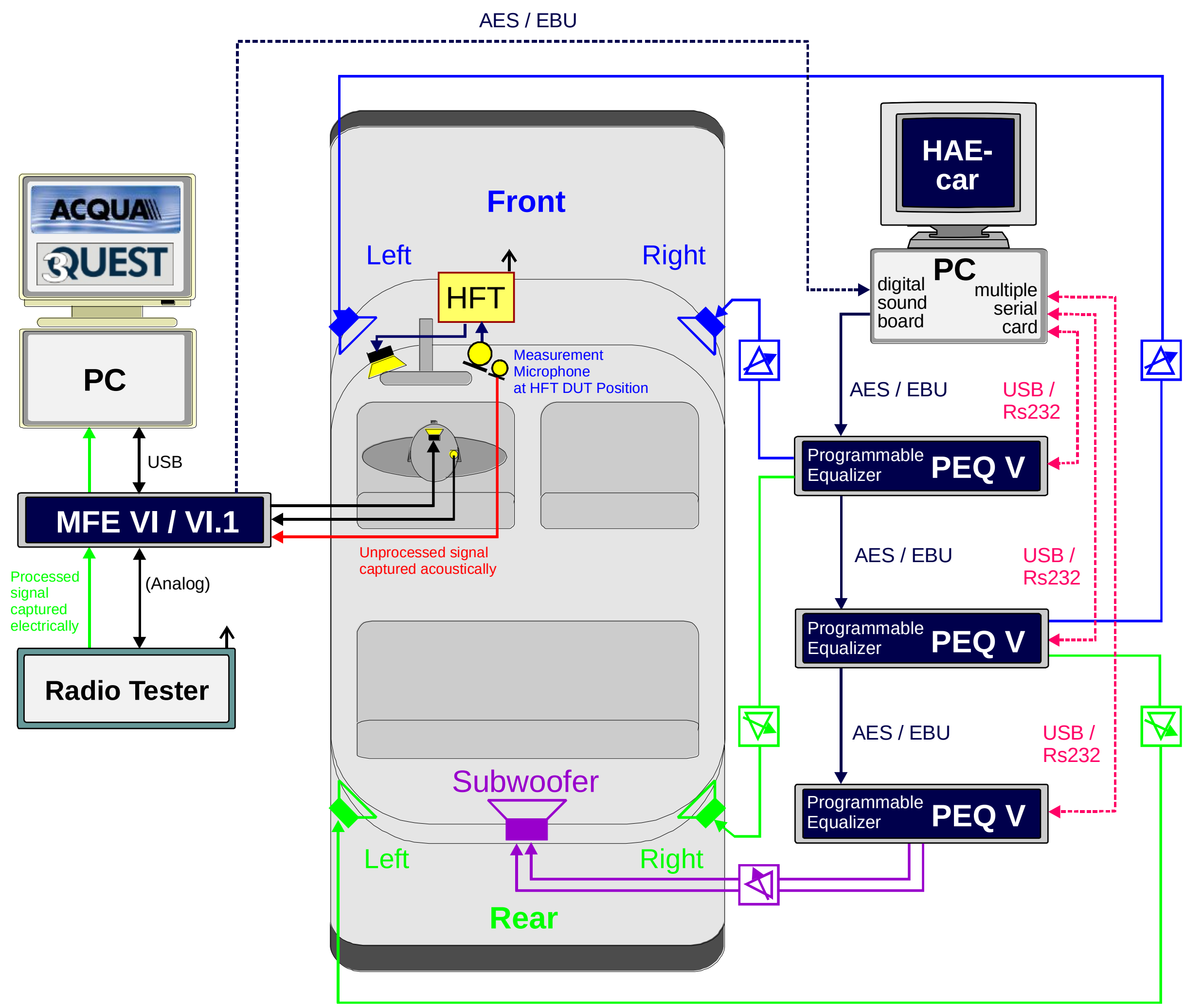


Figure 11: Typical setup for measurements of car hands-free terminals

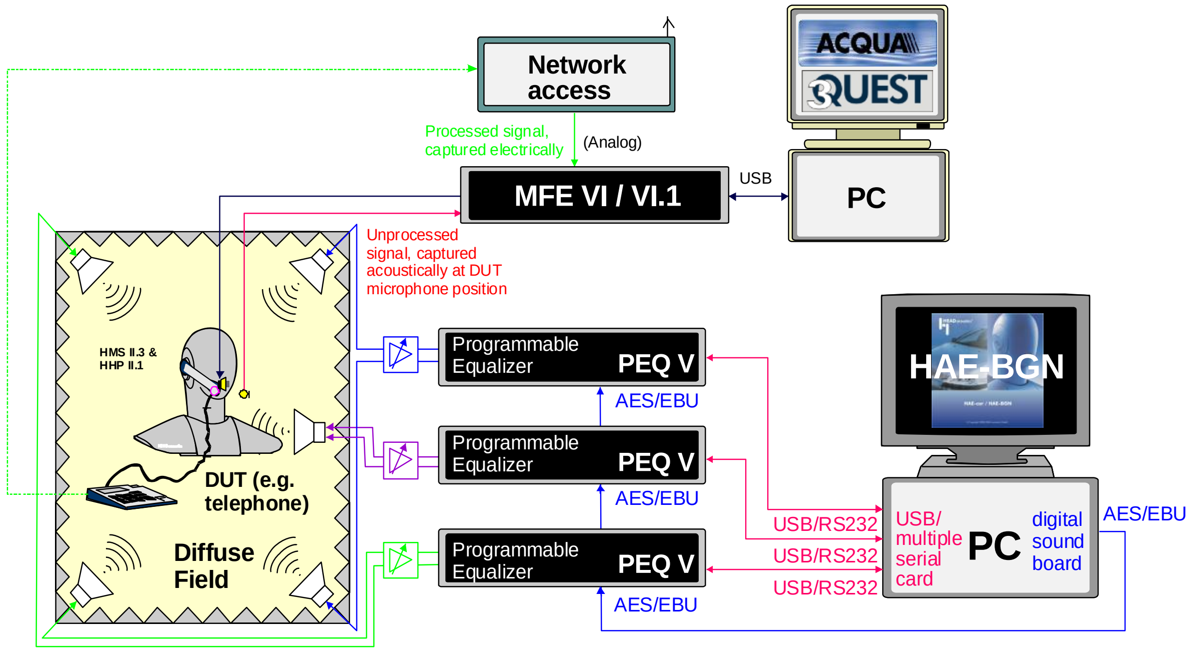


Figure 12: Typical setup for measurements of telephones acc. to ETSI EG 202 396-1 [1]

For electrical to electrical measurements, the setup shown in figure 12 can be modified by using a reference network access (e.g. high-quality ISDN phone) instead of the DUT (device under test). This modified setup is used to make recordings with separate background noises. These recordings can be considered as showing neither distortions nor signal processing effects. Subsequently, these recordings are used as source files for electrical to electrical measurements (IP gateways, IAD's etc.) as shown in figure 13. For measuring IP devices in sending direction MFE VIII is strongly recommended.

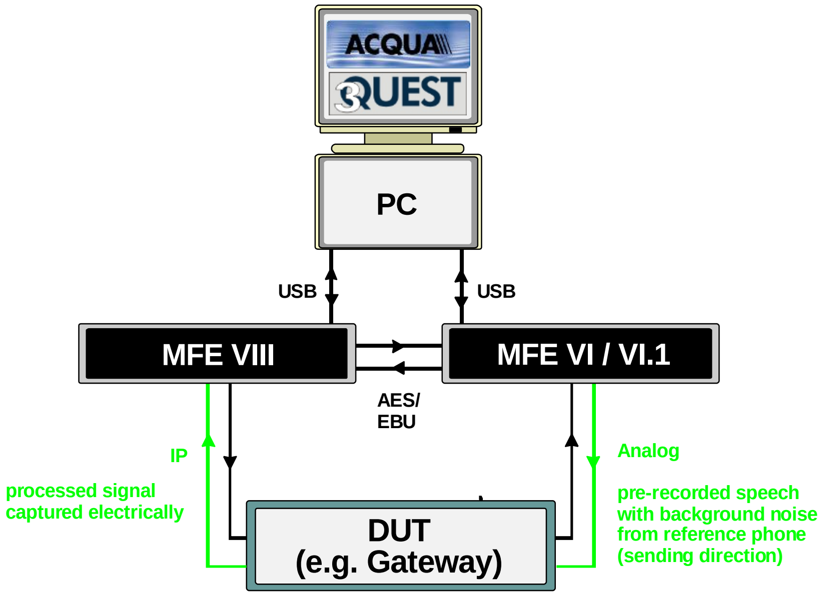


Figure 13: Additional setup for electrical to electrical measurements

5. Listening Examples of HEAD acoustics narrowband database

		Subjective MOS			Objective MOS		
	Noise	**S**	**N**	**G**	**S**	**N**	**G**
	Car 1	4.7	4.7	4.6	4.8	4.8	4.8
	Cafeteria	4.2	2.2	3.1	4.0	2.1	3.1
	Office	4.0	2.8	3.5	4.2	2.6	3.5
	Car 2	2.0	2.3	1.9	2.3	2.3	1.8
	Road	1.4	1.3	1.0	1.3	1.0	1.0

6. References

[1] ETSI EG 202 396-1, Background Noise Transmission - Network Simulation - Background noise simulation technique and background noise database

[2] ETSI EG 202 396-2, Background Noise Transmission - Network Simulation - Subjective Test Database and Results

[3] ETSI EG 202 396-3, Background Noise Transmission - Network Simulation - Objective Test Methods

[4] ITU-T Recommendation P.835, Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm

[5] ITU-T Recommendation P.800: "Methods for subjective determination of transmission quality"

[6] K. Genuit, "Objective Evaluation of Acoustic Quality Based on a Relative Approach", InterNoise 1996, Liverpool, UK

[7] F. Kettler, H.W. Gierlich, F. Rosenberger "Application of the Relative Approach to Optimize Packet Loss Concealment Implementations", DAGA, March 2003, Aachen, Germany

[8] R. Sottek, K. Genuit "Models of Signal Processing in Human Hearing", International Journal of Electronics and Communications (AEÜ), vol. 59, 2005, p. 157-165

[9] R. Sottek, W. Krebber, G. Stanley "Tools and Methods for Product Sound Design of Vehicles", SAE International - Document 2005-01-2513

[10] ITU-T Recommendation SG 12 Contribution 34 "Evaluation of the quality of background noise transmission using the 'Relative Approach' "

[11] Validation results of the modified ETSI EG 202 396-3 objective speech quality model for narrowband data", ETSI STQ Meeting #29, 15-19 September 2008, Oslo, Norway

[12] R. Sottek "Modelle zur Signalverarbeitung im menschlichen Gehör ", PHD thesis, RWTH Aachen, 1993

[13] Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

[14] J. Berger, „Instrumentelle Verfahren zur Sprachqualitätsschätzung - Modelle auditiver Tests", PHD thesis, Kiel 1998, ISBN 3-8265-4091-3

[15] J. Berger, „Results of objective speech quality assessment including receiving terminals using the advanced TOSQA20001", ITU-T Contribution 12/2000, COM 12-20-E