16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

A NEW OBJECTIVE MODEL FOR WIDE- AND NARROWBAND SPEECH QUALITY PREDICTION IN COMMUNICATIONS INCLUDING BACKGROUND NOISE

H.W. Gierlich, F. Kettler, S. Poschen, J. Reimes

HEAD acoustics GmbH
Ebertstrasse 30a, Herzogenrath, Germany
www.head-acoustics.de

ABSTRACT

Modern wideband communication systems like mobile phones, or hands-free terminals are more and more 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 the speech, noise and the overall quality of a sample. Based on the Relative Approach algorithm, we introduce a new model for objectively measuring the quality of wide-band speech in noisy environments which provides a high correlation with the subjective MOS.

1. INTRODUCTION

The quality of processed and transmitted speech in 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]), a database including a big variety of wideband speech samples was created and subjectively evaluated based on ITU-T Recommendation P.835 [3]. 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 as defined in [3].

We furthermore introduce an extension of this model, which includes narrowband scenarios as well. Therefore, a new database of narrowband speech samples was created and subjectively rated according to ITU-T Recommendation P.835.

2. NOMENCLATURE

The objective model determines the following scores:

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

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. 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.

3. 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.

3.1 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 following background noises were included: Inside car, Crossroad, Road, Office, Pub.

The processing of the degraded speech files consisted of different VADs, noise suppression algorithms, network/packet loss scenarios and handset/hands-free modes and were bandlimited between 135Hz and 7kHz. 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.

3.2 HEAD acoustics Narrowband Database

Due to the lack of freely available databases containing narrowband speech and evaluated according to ITU-T Recommendation P.835, a new database including 263 conditions was created. This database 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: Inside car (3 different types), Office, Road, Crossroad, Cafeteria.

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.

4. THE RELATIVE APPROACH ALGORITHM

4.1 Description of the algorithm

The Relative Approach [4] 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 [5], [6], [7].

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 pseudo-stationary energy.

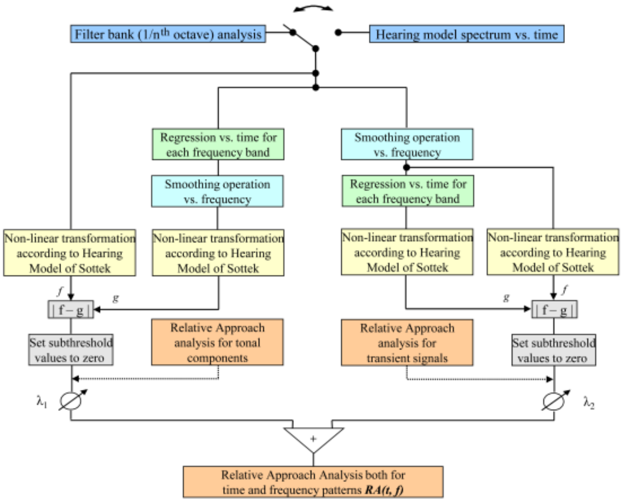
Filter bank (1/n^th octave) analysis
Hearing model spectrum vs. time
Regression vs. time for each frequency band
Smoothing operation vs. frequency
Smoothing operation vs. frequency
Regression vs. time for each frequency band
Non-linear transformation according to Hearing Model of Sottek
Non-linear transformation according to Hearing Model of Sottek
Non-linear transformation according to Hearing Model of Sottek
Non-linear transformation according to Hearing Model of Sottek
f
g
|f−g|
g
f
|f−g|
Set subthreshold values to zero
Relative Approach analysis for tonal components
Relative Approach analysis for transient signals
Set subthreshold values to zero
λ1
λ2
+
Relative Approach Analysis both for time and frequency patterns RA(t, f)

Figure 1: Block diagram of Relative Approach

Figure 1 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 [9]. 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 [9]. This output is compared to the source signal. This variant focuses on the detection of tonal components.

The second variant first smooths 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 $\lambda_2 = 1$ 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 the 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.

A first attempt using the Relative Approach for analyzing time variant background noises was submitted as a contribution in ITU-T 2001 [8]. 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, below the 3D Relative Approach representation is specified as $\mathrm{RA}_P(t_i, f_j)$ for the processed, $\mathrm{RA}_U(t_i, f_j)$ for the unprocessed signal and $\mathrm{RA}_C(t_i, f_j)$ for the clean speech.

4.2 Δ 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 what has changed due to the transmission. This differential analysis (Δ Relative Approach, between the transmitted processed signal and the undisturbed unprocessed signal) provides the information how “close to the human expectation” the processed signal still is compared to the unprocessed signal. The calculation for this example is carried out using (1)

$$\Delta\mathrm{RA}_{P-U}\left(t_i, f_j\right) = \mathrm{RA}_P\left(t_i, f_j\right) - \mathrm{RA}_U\left(t_i, f_j\right) \quad \forall \;\; t_i, f_j \qquad (1)$$

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

Below a brief description of the algorithm is given. 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 [1] and [2].

5.1 Preprocessing steps

5.1.1 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.

5.1.2 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.

5.1.3 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 also $u(k)$ and $p(k)$ can 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*.

5.2 Objective N-MOS

The objective N-MOS algorithm is based on the results of subjective listening tests and conclusions drawn from a consecutive expert listening analysis. This analysis pointed out that parameters like absolute background noise level, modulation / "naturalness" of the background noise (e.g. musical tones) and interruptions / lost packets (minor influence) lead to the subjective N-MOS. The absolute level of the background noise N_{BGN} (in dB) is a significant parameter which influences noise quality. It is given in (3):

$$N'_{\mathrm{BGN},P} = \frac{1}{K} \cdot \sum_k p^2_{\mathrm{BGN}}(k) \qquad (2)$$

$$N_{\mathrm{BGN},P} = 10 \cdot \log\left(\frac{N'_{\mathrm{BGN}}}{1\ \mathrm{Pa}}\right) \qquad (3)$$

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 $\mathrm{RA}_U(t,f)$ and $\mathrm{RA}_P(t,f)$. In these spectrographs, sections containing background noises are extracted with the method described in 5.1.3. The marked time bins lead to the spectograph parts $\mathrm{RA}_{\mathrm{BGN},P}(t,f)$ and $\mathrm{RA}_{\mathrm{BGN},U}(t,f)$.

The mean $\mu\left(\mathrm{RA}_{\mathrm{BGN}}\right)$ and variance $\sigma^2\left(\mathrm{RA}_{\mathrm{BGN}}\right)$ of a Relative Approach spectograph describe audible effects like annoying sounds and/or musical tones resulting from noise suppression algorithms and processing in general.

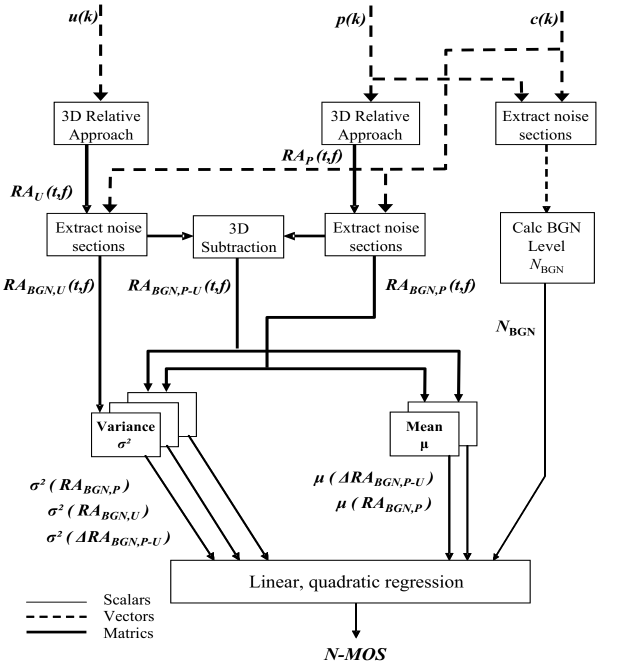

Figure 2: Block diagram of Objective N-MOS Calculation

$$\mu\left(\mathrm{RA}_{\mathrm{BGN}}\right) = \frac{1}{A} \cdot \sum_{t_i} \sum_{\Delta f_j = \Delta f_{\min}}^{\Delta f_{\max}} \mathrm{RA}_{\mathrm{BGN}}\left(t_i, f_j\right) \cdot dA_j \qquad (4)$$

$$P\left(\mathrm{RA}_{\mathrm{BGN}}\right) = \frac{1}{A} \cdot \sum_{t_i} \sum_{\Delta f_j = \Delta f_{\min}}^{\Delta f_{\max}} \mathrm{RA}^2_{\mathrm{BGN}}\left(t_i, f_j\right) \cdot dA_j \qquad (5)$$

$$\sigma^2\left(\mathrm{RA}_{\mathrm{BGN}}\right) = P\left(\mathrm{RA}_{\mathrm{BGN}}\right) - \mu\left(\mathrm{RA}_{\mathrm{BGN}}\right)^2 \qquad (6)$$

with

$$A = \frac{1}{\left(\sum_{t_i} \Delta t\right) \cdot \left(f_{\max} - f_{\min}\right)} \qquad (7)$$

$$dA_j = \Delta t \cdot \Delta f_j \qquad (8)$$

t_i : Time bins in background noise sections
Δt : Duration of a single time bin ($6.67\,ms$)
Δf_j : Bandwidth of frequency bin f_j ($^1/_{12}$th octave bands)

For wideband mode:
$f_{\min} = 50$ Hz, lower frequency of band $\Delta f_{\min}$
$f_{\max} = 8000$ Hz, upper frequency of band $\Delta f_{\max}$
For narrowband mode:
$f_{\min} = 200$ Hz, lower frequency of band $\Delta f_{\min}$
$f_{\max} = 3600$ Hz, upper frequency of band $\Delta f_{\max}$

These parameters are calculated for the background noise sections of the 3D Relative Approach spectographs of the processed ($\mathrm{RA}_{\mathrm{BGN},P}$), unprocessed ($\mathrm{RA}_{\mathrm{BGN},U}$) and the difference of processed and unprocessed signal ($\Delta\mathrm{RA}_{\mathrm{BGN},P-U}$). Finally the objective N-MOS is the result of a linear, quadratic regression algorithm applied to all six parameters (see table 1) according to (9).

$$\text{N-MOS} = c_0 + \sum_{j=1}^{2} \sum_{i=1}^{6} c_{j,i} \cdot P_i^j \qquad (9)$$

P_1	$N_{\mathrm{BGN},P}$	P_4	$\mu\left(\mathrm{RA}_{\mathrm{BGN},U}\right)$
P_2	$\mu\left(\mathrm{RA}_{\mathrm{BGN},P}\right)$	P_5	$\sigma^2\left(\mathrm{RA}_{\mathrm{BGN},U}\right)$
P_3	$\sigma^2\left(\mathrm{RA}_{\mathrm{BGN},P}\right)$	P_6	$\sigma^2\left(\Delta\mathrm{RA}_{\mathrm{BGN},P-U}\right)$

Table 1: Parameters for N-MOS calculation

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

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 2, the algorithm for the determination of the N-MOS is summarized in a flow diagram.

5.3 Objective S-MOS

The objective S-MOS is also aiming to reproduce the listening impression of the test persons in the listening test, to provide a high correlation to 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 also the S-MOS algorithm is designed to reproduce the above parameters. The difference between the SNR of the unprocessed and the processed signal (ΔSNR) is one of the extracted parameters. It is determined by considering the energies in the speech (Sp) and background noise (BGN) parts according to (3) and (12).

$$(S+N)'_{P,\mathrm{Sp}} \quad = \quad \frac{1}{K}\cdot\sum_k p^2_{\mathrm{Sp}}(k) \qquad (10)$$

$$\mathrm{SNR}_P \quad = \quad 10\cdot\log\left(\frac{(S+N)'_{P,\mathrm{Sp}} - N'_{BGN,P}}{N'_{BGN,P}}\right) \qquad (11)$$

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

$$\Delta\mathrm{SNR} \quad = \quad \mathrm{SNR}_P - \mathrm{SNR}_U \qquad (12)$$

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 (4) and (6), mean and variance of the Relative Approach spectographs of $\mathrm{RA}_{\mathrm{Sp},P}$, $\Delta\mathrm{RA}_{\mathrm{Sp},P-C}$, $\Delta\mathrm{RA}_{\mathrm{Sp},P-U}$ within the speech parts can be determined. This results again in six parameters P_i for a linear, quadratic regression (compare table 2):

P_1	$\Delta\mathrm{SNR}$	P_4	$\mu\left(\mathrm{RA}_{\mathrm{Sp},P-C}\right)$
P_2	$\mu\left(\mathrm{RA}_{\mathrm{Sp},P}\right)$	P_5	$\sigma^2\left(\mathrm{RA}_{\mathrm{Sp},P-C}\right)$
P_3	$\mu\left(\mathrm{RA}_{\mathrm{Sp},P-U}\right)$	P_6	$\sigma^2\left(\mathrm{RA}_{\mathrm{Sp},P-U}\right)$

Table 2: Parameters for S-MOS calculation

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 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 continous weighting of N-MOS is "quantized" into three ranges:

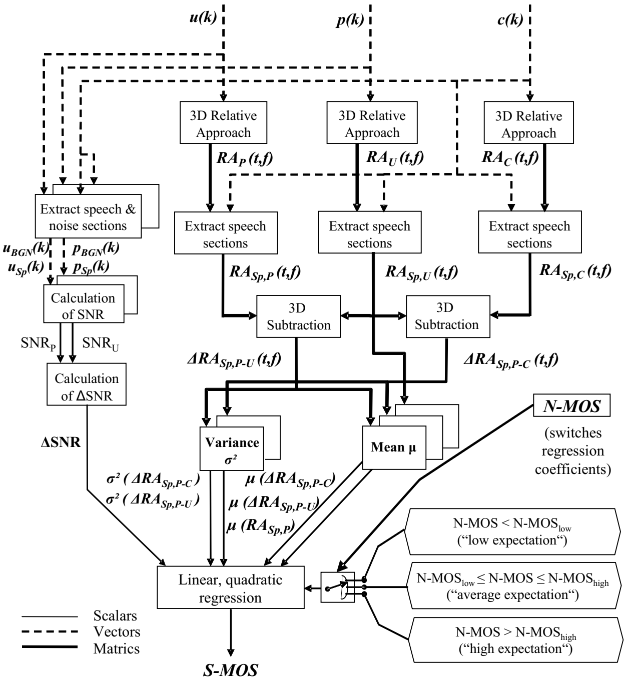
u(k)
p(k)
c(k)
3D Relative Approach
3D Relative Approach
3D Relative Approach
$RA_P\,(t,f)$
$RA_U\,(t,f)$
$RA_C\,(t,f)$
Extract speech & noise sections
Extract speech sections
Extract speech sections
Extract speech sections
$u_{BGN}(k)$
$p_{BGN}(k)$
$u_{Sp}(k)$
$p_{Sp}(k)$
$RA_{Sp,P}\,(t,f)$
$RA_{Sp,U}\,(t,f)$
$RA_{Sp,C}\,(t,f)$
Calculation of SNR
3D Subtraction
3D Subtraction
$\mathrm{SNR_P}$
$\mathrm{SNR_U}$
$\Delta RA_{Sp,P\text{-}U}\,(t,f)$
$\Delta RA_{Sp,P\text{-}C}\,(t,f)$
Calculation of ΔSNR
N-MOS
(switches regression coefficients)
ΔSNR
Variance σ^2
Mean μ
$\sigma^2\,(\,\Delta RA_{Sp,P\text{-}C}\,)$
$\sigma^2\,(\,\Delta RA_{Sp,P\text{-}U}\,)$
$\mu\,(\Delta RA_{Sp,P\text{-}C})$
$\mu\,(\Delta RA_{Sp,P\text{-}U})$
$\mu\,(RA_{Sp,P})$
N-MOS < $\text{N-MOS}_{\text{low}}$ ("low expectation")
$\text{N-MOS}_{\text{low}}$ ≤ N-MOS ≤ $\text{N-MOS}_{\text{high}}$ ("average expectation")
N-MOS > $\text{N-MOS}_{\text{high}}$ ("high expectation")
Linear, quadratic regression
Scalars
Vectors
Matrics
S-MOS

Figure 3: Block diagram of Objective S-MOS Calculation

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

Depending on the N-MOS of a condition, the parameters P_i 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 coefficents for the regression is chosen. The determination of the S-MOS is given in (13).

$$\text{S-MOS} \quad = \quad c_{R,0} + \sum_{j=1}^{2}\sum_{i=1}^{6} c_{R,i,j}\cdot P_i^j \qquad (13)$$

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

The best fitting values for $\text{N-MOS}_{\text{low}}$ and $\text{N-MOS}_{\text{high}}$ 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 3.

5.4 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 (14).

$$\text{G-MOS} = c_0 + \sum_{j=1}^{2} c_{S,j} \cdot \text{S-MOS}^j + \sum_{j=1}^{2} c_{N,j} \cdot \text{N-MOS}^j \quad (14)$$

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

6. PREDICTION RESULTS

In general, a certain amount of conditions of a database is used to train the model. The prediction of these conditions, which are part of the model are called "Training Results". Unknown databases or conditions are used for validation in order to evaluate the prediction accuracy of the model. These prediction results are called "Validation Results".

6.1 Results for Wideband Data

The model was trained with a certain amount of given randomized data (179 conditions). The rest of the databases is used for own validation only. During the development of the algorithm in the STF 294 project the subjective S-, N- and G-MOS results of 81 conditions remained unknown to HEAD acoustics until the end of the algorithm development.

	Training		Validation	
	corr.	RMSE	corr.	RMSE
S-MOS	91.2%	0.37	93.0%	0.33
N-MOS	94.3%	0.27	92.4%	0.32
G-MOS	94.6%	0.25	93.5%	0.28

Table 3: Correlation and RMSE of prediction for ETSI wideband database

In table 3 the correlation coefficient and the root mean square error (RMSE) between the subjective and objective MOS data is shown.

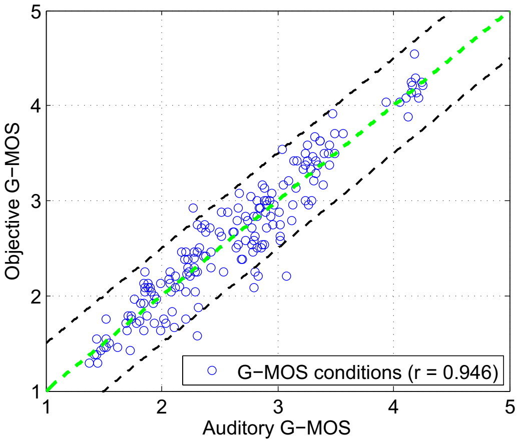

Figure 4: Correlation between subjective and objective G-MOS

As an example, the per-sample correlation of the global MOS for the training data is shown in figure 4.

6.2 Results for Narrowband Data

Overall, there are 263 conditions in the narrowband database. The training of the model was done with 213 randomly chosen conditions, the remaining 50 conditions were used to test the model against unknown, retained data (in terms of data which was not used to train the model). Again, for the training and the validation results, the correlation met the demands for an objective model (see table 4).

	Training		Validation	
	corr.	RMSE	corr.	RMSE
S-MOS	91.6%	0.37	90.0%	0.45
N-MOS	94.2%	0.33	93.5%	0.35
G-MOS	94.3%	0.31	93.2%	0.36

Table 4: Correlation and RMSE of prediction for narrowband database

7. CONCLUSION

We introduced a new model for predicting speech, background noise and overall quality. The output scores refer to the ITU-T Recommendation P.835, and are more meaningful than a single MOS value resulting from a traditional P.800 listening test. The algorithm is able to predict wideband and narrowband scenarios as well, using narrowband and wideband mapping functions to adapt the model to the different listening tests. The prediction results for both training and validation data show a high correlation between objective and subjective ratings.

REFERENCES

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

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

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

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

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

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

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

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

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