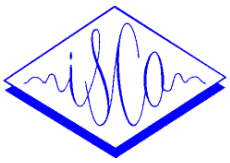
ISCA Archive
http://www.isca-speech.org/archive



First ISCA ITRW on
Auditory Quality of Systems
Akademie Mont-Cenis, Germany
April 23-25, 2003

PARAMETER-BASED SPEECH QUALITY MEASURES FOR GSM

Marc Werner[1], *Karsten Kamps*[1], *Ulrich Tuisel*[2], *John G. Beerends*[3] *and Peter Vary*[1]

[1]Institute of Communication Systems and Data Processing (ind), Aachen University, Germany
[2]E-Plus Mobilfunk GmbH, Düsseldorf, Germany
[3]TNO Telecom, The Netherlands

werner@ind.rwth-aachen.de

Abstract— **This contribution introduces non-intrusive instrumental speech quality measures for the GSM system based only on transmission parameters. The idea behind these measures is that parameters which quantify the transmission quality in terms of bit error rate, received power level, etc., may also be suitable to predict the resulting speech quality. Neither the original nor the received speech signal is needed for this kind of prediction. The proposed speech quality measures have been validated by extensive link-level simulations which are based on measurements of transmission parameters collected in a GSM-1800 network. Speech samples were produced by bit-exact transmission simulations using the measured link parameters for channel modelling. The reference speech quality assessments of these samples were carried out with the PESQ algorithm [4]. The correlation of the presented parameter measures with the intrusive PESQ measure is remarkable.**

1. INTRODUCTION

In spite of the upcoming technologies for data transmission, voice telephony is still by far the most important form of mobile communication. The improvement of speech quality is therefore an essential task within the competition of cellular network operators. While the optimization of technical radio network parameters, e.g., cell sizes or received power levels, is widely known and continuously carried out by the network operators, it is difficult to measure the resulting speech quality in an objective and automated way.

The most reliable method for evaluating the perceived speech quality of a transmission system or speech codec is the subjective assessment of speech material by a large number of persons in a listening test. The testing environment and procedures are specified by the ITU [1] [2]; a common rating method is the Absolute Category Rating (ACR) in which speech samples are marked on a five-point scale from 1 (bad) to 5 (excellent) by a number of listeners and a Mean Opinion Score (MOS) is calculated. The MOS scale is generally accepted for the appropriate description of speech quality.

Because subjective listening tests are expensive and time-consuming, instrumental speech quality measures have been developed. They allow the measurement of speech quality by analyzing the speech samples or some other related transmission system parameters.

Intrusive speech quality measurements like PSQM (*Perceptual Speech Quality Measure*) [3] [5] or the more elaborate PESQ (*Perceptual Evaluation of Speech Quality*) [4] [6] need the original and distorted speech samples. They have been developed on the basis of human auditory perception and deliver excellent correlations with subjective listening tests. The major drawback of intrusive measurements is that only pre-defined test calls can be evaluated because both the original and degraded speech samples are needed. Additional network load is generated by these measurements. Furthermore, it is sometimes difficult to see which part of the transmission chain contributes in which way to the resulting end-to-end quality.

The radio link is to be regarded as the most critical part of the GSM transmission chain with respect to speech quality. The analysis of the radio link by automated non-intrusive quality measurements is therefore a suitable and convenient option in Quality-of-Service (QoS) optimization.

Approaches to non-intrusive quality monitoring based on GSM measurement values include a statistical analysis of single parameters, e.g., a call is estimated to be of satisfactory quality if a certain threshold Bit Error Rate (BER) is not exceeded in any transmission segment. While such statistics are very convenient because the required parameters (RxQual etc.) are available at the Operation and Maintenance Centre (OMC) for the uplink direction and are also part of the GSM measurement reports from the mobile station [15], they can be unreliable and lack an accurate distinction between speech quality levels. An optimized combination of transmission parameters can serve as a good speech quality estimation. One example of such a method was presented by Karlsson et al. for a GSM system employing the Full-Rate (FR) speech codec [8] [12], and later extended by Wänstedt et al. to the Adaptive Multi-Rate (AMR) codec [9] [14]. It was shown that the correlation of the non-intrusive parameter-based measure SQI (*Speech Quality Index*) with subjective speech quality can exceed that of the psychoacoustically motivated end-to-end instrumental quality measure PSQM. However, the PSQM was not designed for the evaluation of signal distortions occurring in mobile radio communications.

The parameter-based mapping functions presented in this paper are based on a large database of transmission parameters from a GSM network, described in Section 2, and on bit-exact speech transmission simulations referring to the recorded parameter values. Apart from the actual output speech samples, some additional transmission parameters were obtained from the simulations. The output speech samples were evaluated with the PESQ algorithm. However, subjective listening test evaluations were not performed within this study. The PESQ scores serve as reference speech quality values on the Mean Opinion Score (MOS) scale used in subjective listening tests [1].

The correlations of single GSM parameters with the objective speech quality are analyzed in Section 3. The method of averaging the parameter progression per speech sample is of vital importance for a good correlation. The mean value per speech sample must be calculated for each parameter using some sort of averaging over the sample period. Since linear averaging does not correspond well to auditive perception, in which extreme values are over-emphasized, the so-called L_P-norm was employed.

In Section 4, the individual parameters are combined using optimized mapping functions which maximize the linear correlation with objective speech quality.

2. GSM LINK PARAMETERS

The analysis of transmission parameters was based on downlink measurement data collected within a GSM-1800 network in Germany. The data covers more than 400 measurement sessions, each containing multiple speech calls. Apart from link quality measurements, the recorded GSM parameters include information about the current radio cell, geographic position and time. The parameters that were identified to be particularly relevant for the resulting speech quality include:

RxQual: The channel bit error rate is averaged over an interval of 480 ms and mapped to the logarithmic RxQual parameter with eight bit error rate ranges according to Table 1. RxQual serves as an estimate of current channel quality during an active call. In the GSM system, values below four are desirable, because at a gross BER of less than 1.6%, nearly all bit errors within the most important class-I-bits can be corrected by the channel decoder. Due to a high base station density in the regarded area, a large fraction of RxQual measurement data exhibits small values.

RxLev: The received power level at the mobile station is measured in dBm (relative to 1 mW) and mapped linearly to an RxLev index ranging from 0 to 63 in 1 dBm steps (see Table 2). The minimum value specified in the GSM standard [10] ranges from -104 to -100 dBm (RxLev$>$6...10). Measurements are reported every 480 ms. The received power level describes the radio channel in terms of path loss and slow fading. It is not a measure of signal-to-interference ratio (SIR), but really an expression of the sum of the desired signal plus interference. A high correlation with the resulting speech quality is therefore only expected for the case that the interference is low and relatively constant, e.g., in a GSM system with a large cluster size.

RxQual level	BER from	to	mean
RxQual = 0		$<$0.2%	0.14%
RxQual = 1	0.2%	0.4%	0.28%
RxQual = 2	0.4%	0.8%	0.57%
RxQual = 3	0.8%	1.6%	1.13%
RxQual = 4	1.6%	3.2%	2.26%
RxQual = 5	3.2%	6.4%	4.53%
RxQual = 6	6.4%	12.8%	9.05%
RxQual = 7	$>$12.8%		18.10%

Table 1: RxQual levels and corresponding BER ranges [11]

RxLev level	received power level from	to
RxLev = 0		$<$ -110 dBm
RxLev = 1	-110 dBm	-109 dBm
RxLev = 2	-109 dBm	-108 dBm
...	...	...
RxLev = 61	-50 dBm	-49 dBm
RxLev = 62	-49 dBm	-48 dBm
RxLev = 63	$>$ -48 dBm	

Table 2: RxLev levels and corresponding power ranges [11]

TA: The Timing Advance (TA) parameter describes the distance between mobile station and base station. Its resolution is roughly $\pm$ 550 m, which corresponds to the distance of radio waves propagation during the duration of one bit (3.69 μs). Transmission quality is expected to be better for low distances between mobile and base station, and a high correlation with the RxLev parameter is obvious. For urban measurement environments, the TA resolution was observed to be insufficient: About 95% of TA values were equal to 0 or 1. For this reason, the TA parameter was excluded from the further analysis.

To calculate the correlation of the above GSM link parameters with the objective speech quality, speech samples were produced that reflect the transmission conditions characterized by the measurements. These samples were generated using a bit-exact GSM speech transmission simulation. Channel degradations are best described by the SIR or the resulting channel BER. The RxLev Parameter does not serve as a good SIR indicator in all cases. Therefore, the TUx channel model recommended for simulations [10] was replaced by an equivalent binary bit error channel which adapts the error rate every 480 ms corresponding to the measured RxQual values. The bit errors are distributed evenly

over each 480 ms interval. This is not the case in real transmissions, where burst errors occur due to fast and slow fading. However, the de-interleaver at the receiver side spreads error bursts over the transmission frames so that bit errors are nearly independent after the de-interleaving. Simulations of the TUx channel with de-interleaver confirmed that this simplification is permissible with respect to the measured speech quality.

Simulations were performed using the CoCentric System Studio software [16]. The GSM transmission model includes the EFR (*Enhanced Full Rate*) speech codec [13], channel coding, frame building, an equivalent binary channel and the decoding elements at the receiver side. At the channel decoding stage, a BFI (*Bad Frame Indication*) signal is generated for any speech frame in which the class-I-bits could not be correctly decoded. In this case, the speech decoder performs error concealment by repeating the last correct frame or by muting if too many subsequent frames have been replaced. Several thousands of male and female speech samples were generated from the measurement data. Each sample has a duration of approximately 9 s.

Obviously, the BFI rate, or Frame Erasure Rate (FER), and its distribution within the speech sample, are of great relevance for the speech quality. Therefore the FER and some new derivations were included as GSM parameters, although they had not been part of the original measurements:

FER: Frame Erasure Rate for speech frames,

LFER: Length of Erased Frames, mean sequence length of consecutively erased speech frames in the speech sample,

MxLFER: Maximum Length of Erased Frames, maximum sequence length of erased speech frames,

MnMxLFER: Mean of Maximum Length of Erased Frames, a combination of local maximum sequence lengths of erased speech frames for four intervals of equal length. The maximization over short periods was regarded to be similar to the human perception of severe signal distortions.

Although the FER is not part of the standard GSM downlink measurement report, FER values for the uplink are usually stored within the OMC and an OMC function often exists which estimates the downlink FER. This feature depends on the OMC manufacturer.

3. CORRELATION OF PARAMETERS AND SPEECH QUALITY

To express the degree of correlation between two data vectors $\boldsymbol{u}$ and $\hat{\boldsymbol{u}}$, the correlation coefficient ρ is calculated:

$$\rho = \left| \frac{\sum_{i=1}^{n} u_i \cdot \hat{u}_i}{\sqrt{\sum_{i=1}^{n} u_i^2 \cdot \sum_{i=1}^{n} \hat{u}_i^2}} \right| \in [0,1] \tag{1}$$

For the sake of simplicity, we define the correlation coefficient to be an absolute value and drop the sign of ρ. In Eq. 1, u_i are n zero-mean reference vector elements normalized by their standard deviation, and $\hat{u}_i$ the corresponding estimation values. It should be ensured that both vectors cover their complete range of values, and that the two vectors exhibit a linear dependency. In this study, we maximize the correlation of GSM parameters (or functions thereof) with the reference PESQ speech quality scores. A value of $\rho = 1$ represents perfect correlation, and for $\rho = 0$ the two vectors are uncorrelated. Instrumental quality measures should have a correlation coefficient of at least 0.9 with respect to the results of subjective quality tests.

The procedure to estimate the correlation coefficient ρ of the GSM parameters and the objective speech quality is based on the averaging functions for individual parameters per speech sample and on the linearization of the mapping functions between parameters and speech quality.

In the original data, parameter measurements were recorded at irregular time intervals, ranging from 1/8 s to 1 s. As a first step, the progression of the parameters $\zeta_i(k)$ described in Section 2 was identified for each speech signal i. The variable k serves as a discrete measurement (time) index.

To study the correlation of transmission parameters with the reference PESQ values M_i, an average value of each parameter was obtained by calculating the L_P-norms per speech sample

$$L_P(\zeta_i(k)) = \left[\frac{1}{N} \sum_{k=1}^{N} \left(\zeta_i(k)\right)^P \right]^{1/P} \tag{2}$$

for exponents $P \in \{1/10, 1/9, \ldots 1/2, 1, 2, \ldots 9, 10\}$. In the above expression, the L_1-norm corresponds to the arithmetic mean and the L_2-norm is equivalent to the quadratic mean of $\zeta_i(k)$. The reason for using various L_P-norms is that for each parameter, variations and outliers may be perceived in a different way with respect to the resulting speech quality. High values for P emphasize parameter variations.

For each parameter and each value of P, the mapping function f which fits the L_P-norms to the objective PESQ quality values M_i, over all speech samples i, was approximated with respect to a minimum mean squared error using a polynomial of degree $m \in \{2 \ldots 6\}$:

$$M_i \approx f(L_P(\zeta_i(k))) \tag{3}$$

The resulting correlation coefficient $\rho(f(L_P(\zeta_i(k))),\ M_i)$ was calculated for all values of P and m, and optimum values $\hat{P}$ and $\hat{m}$ together with the corresponding linearization polynomial $\hat{f}$ were identified which maximize the correlation.

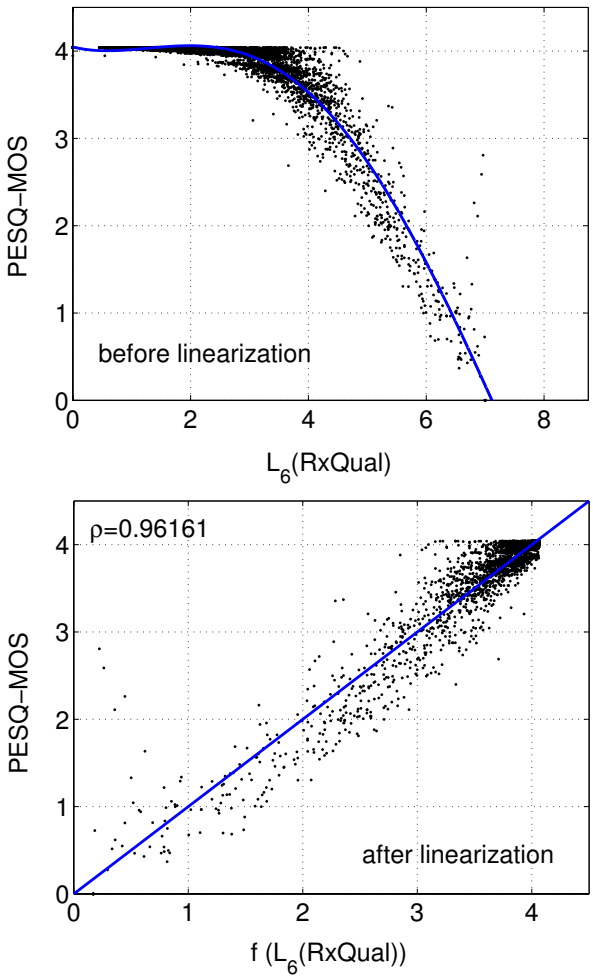


Figure 1: RxQual–PESQ correlation after polynomial linearization: Transformation of RxQual (L_6-norms)

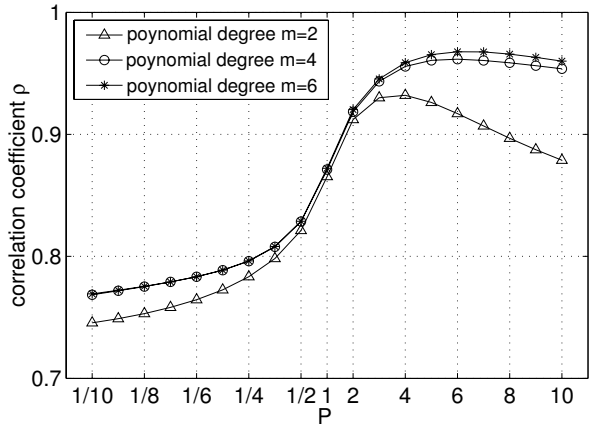


Figure 2: RxQual–PESQ correlations for different L_P-norms and polynomial degrees

An example of the polynomial fitting and selection of optimal values for P and m is depicted in Figures 1 and 2.

The effect of linearization is shown in Figure 1. The relation between the resulting speech quality (PESQ-MOS) and the L_6-norm of RxQual is depicted for a subset of speech samples as a scatter-plot. The distribution of points in the upper diagram indicates a nonlinear dependency and therefore a low linear correlation of the parameter norms with the PESQ scores. After the mapping polynomial f has transformed the L_6-norms on the x-axis in the lower diagram, the correlation coefficient ρ increases significantly.

Figure 2 shows the dependency of the resulting correlation $\rho(f(L_P(\zeta_i(k))), M_i)$ on the P-value of the L_P-norm and on the polynomial degree m, for the parameter RxQual. It can be observed that the highest correlation is obtained for $\hat{P} = 6$ and $\hat{m} = 6$, but a lower-degree polynomial with $m = 4$ reduces the correlation only very slightly. To simplify the obtained measures, a polynomial degree of $m = 4$ was chosen, resulting in a correlation loss well below 1%.

It should be noted that, the deterministic linearization function f does not change the general dependency between speech quality and parameter value itself but improves the linear correlation measure. On the other hand, the optimization of P offers a real correlation gain.

Table 3 gives an overview of the obtained parameter correlations, using optimum L_P-norms, after linearization by individual polynomials of degree $m = 4$. It can be observed that all transmission parameters except RxLev exhibit a high correlation with the objective speech quality, especially RxQual, FER and MnMxLFER. RxLev is a measure of the attenuation property of the radio channel, which is a cause for signal degradation. All other parameters represent the signal impairment effects at the receiver and are therefore better suited to characterize the received signal quality.

parameter ζ	$\hat{P}_\zeta$	$\rho_\zeta(f, M_i)$ after lin.
RxQual	6	0.9419
RxLev	0.25	0.6781
FER	0.5	0.9632
LFER	6	0.8864
MnMxLFER	1	0.9383
MxLFER	n.a.	0.9088

Table 3: $\hat{P}$-values of L_P-norms, and resulting correlation ρ (after linearization by f)

A large optimum P-value of $\hat{P} = 6$ for RxQual indicates that outliers are perceived more strongly than it is suggested by the numerical value of this parameter. Note that for the FER, the $L_{0.5}$-norm corresponds to the square root, because

the constituent elements are taken from the set $\{0, 1\}$ or {BFI, no BFI} only. For MxLFER, the L_P-norm is not applicable because only a single value per speech sample is available.

4. PARAMETER-BASED SPEECH QUALITY MEASURES

The GSM transmission parameters RxQual, FER and MnMxLFER exhibit a very high correlation with the resulting speech quality in terms of PESQ scores. These parameters were combined to obtain an objective non-intrusive parameter-based speech quality measure.

To find a suitable combination rule, the MSECT (*Minimum Mean Square Error Coordinate Transformation*) [17] procedure was employed.

Multidimensional data in pre-defined categories within a source space of dimension D is mapped onto target positions in a target space of dimension $N < D$. The mapping function is optimized with respect to a minimum mean squared error between the mapping points and the specified target positions of training datasets.

The optimal mapping function is of the form

$$\boldsymbol{c} = \boldsymbol{T} \cdot \boldsymbol{v} + \boldsymbol{o} \tag{4}$$

where source vectors $\boldsymbol{v}$ are mapped in a linear way to target vectors $\boldsymbol{c}$, i.e., an optimal mapping matrix $\boldsymbol{T}$ and offset vector $\boldsymbol{o}$ are identified by the algorithm. This procedure is based on training datasets for which the target positions are already known.

The MSECT method is applied to the given task of mapping parameter vectors to estimated MOS scores. In this application, parameter groups resulting in different speech quality levels are regarded as the categories of the source space. Distinct MOS values serve as target positions in the one-dimensional target space. L_P-norms of the chosen parameters ζ can be optionally linearized by their polynomial f_ζ before serving as input vectors.

The resulting speech quality measure is of the form

$$\begin{aligned} \text{SQM} &= T_1 \cdot f_1(L_6(\text{RxQual})) + T_2 \cdot f_2(\sqrt{\text{FER}}) \\ &+ T_3 \cdot f_3(L_1(\text{MnMxLFER})) + B \end{aligned} \tag{5}$$

with optimized values for T_1, T_2, T_3, and B, where the value ranges of $f_{1\ldots3}$ are comparable. The weighting factors T_i in Eq. 5 indicate a prominent importance of the parameter FER. The value of T_2 is more than four times larger than that of the MnMxLFER and RxQual weights which are in the same range.

Approximately 2% of the available speech samples and PESQ scores were chosen as training data for the MSECT algorithm. When evaluating the prediction performance of the resulting mapping function, the training data should normally be excluded from the correlation calculations. Two correlation coefficients of SQM and the PESQ values were calculated: $\rho_{\text{incl}} = 0.9648$ on the basis of the complete data (including the training datasets), and $\rho_{\text{excl}} = 0.9527$ based only on the datasets excluding the training data.

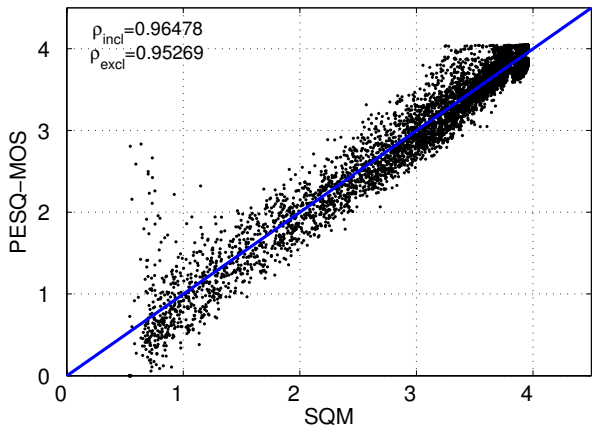


Figure 3: Correlation of SQM and PESQ

The correlation is depicted as a scatter-plot consisting of more than 58 000 points in Figure 3. A high degree of correlation can be clearly observed. However, a small subset of about 40 points ($< 0.07\%$) does not seem to match the prediction model very well. The corresponding speech samples exhibiting poor SQM scores but high PESQ values are subject to severe time clipping. This is a known issue in the PESQ version used for this study and has been corrected in a newer version [7]. The SQM measure evaluates these samples correctly.

Of the three GSM parameters chosen for the SQM, the square root of FER possesses the highest correlation with PESQ values. A simple method to estimate the speech quality is therefore to evaluate this parameter on its own. Because the correlation coefficient of $\sqrt{\text{FER}}$ (without linearization) and PESQ is already very high (see Table 3), the polynomial f can be discarded in this case. A simple measure is thus obtained by

$$\text{SQM}_\text{F} = A \cdot \sqrt{\text{FER}} + C \tag{6}$$

For optimized values of A and C, determined by exploiting 2% of the measurement data, correlations with PESQ of up to $\rho_{\text{excl}} = 0.9600$ were observed. Figure 4 illustrates this case.

It should be noted that the two instrumental measures presented above are only valid for one configuration of a GSM radio transmission network. For other networks, e.g., employing a different speech codec, noise reduction or echo cancelling algorithms, a new measure can be found by applying the described procedures and algorithms to new

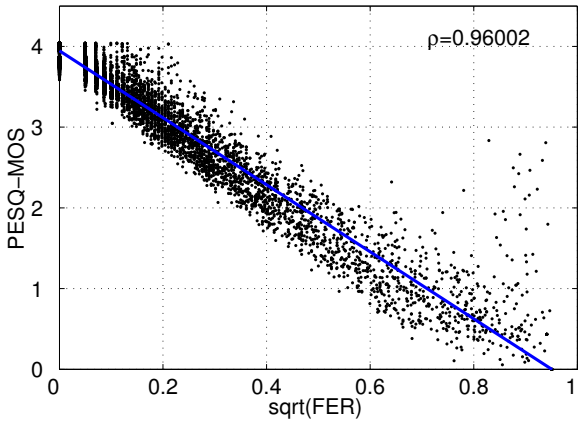


Figure 4: Correlation of $\sqrt{\text{FER}}$ and PESQ

training data. It is a subject of current studies to validate the simulated FER values using measured BFI progressions and to compare the generated speech samples to GSM network recordings.

Secondly, the given correlations are calculated with respect to the instrumental speech quality measure PESQ only. The correlation of the presented measures with listening test results might be slightly lower.

5. CONCLUSION

Two empirical mapping functions were presented which allow a non-intrusive estimation of the objective speech quality in GSM telephony, taking as input only GSM transmission parameters. The mapping functions were identified on the basis of extensive GSM measurements and link-level simulations.

The proposed methods allow accurate, fast, automated and economical quality analysis and optimization for network operators. The required measurement parameters are either available or easy to determine at the OMC level and can be combined instantly using one of the two proposed combination methods.

Although the described methods are based on short speech samples, they can be extended to longer speech transmissions. In this case, the quality of single sentences can be measured and combined in a suitable way. On the other hand, for the estimation of the conversational quality of a complete voice call, further aspects like delay, double talk etc. should be included in the analysis.

REFERENCES

[1] ITU-T Recommendation P.800, Methods for subjective determination of transmission quality, Geneva 1996.

[2] ITU-T Recommendation P.830, Subjective performance of telephone-band and wideband digital codecs, Geneva 1996.

[3] ITU-T Recommendation P.861 (withdrawn), Objective quality measurement of telephone-band (300-3400 Hz) speech codecs, Geneva 1998.

[4] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, Geneva 2001.

[5] Beerends, J.G., Stemerdink, J.A.: A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation, *Journal Audio Eng. Soc., Vol. 42, No. 3, S. 115-123*, March 1994.

[6] Rix, A. W. et al.: Perceptual Evaluation of Speech Quality (PESQ) - A New Method for Speech Quality Assessment of Telephone Networks and Codecs, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, 2001.

[7] KPN: User's Guide Perceptual Evaluation of Speech Quality Acoustic (PESQ Acoustic), April 2002.

[8] Karlsson, A. et al.: Radio Link Parameter based Speech Quality Index - SQI, *Proc. 1999 IEEE Workshop on Speech Coding*, Porvoo, Finland.

[9] Wänstedt, S. et al.: Development of an Objective Speech Quality Measurement Model for the AMR Codec, *Proc. Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN)*, Prague, January 2002.

[10] ETSI Recommendation GSM 05.05, Radio Transmission and Reception, Version 8.5.0, 1999.

[11] ETSI Recommendation GSM 05.08, Radio Sub-System Link Control, Version 8.4.0, 1999.

[12] ETSI Recommendation GSM 06.10, GSM Full Rate (EFR) speech transcoding, Version 3.2.0, 1995.

[13] ETSI Recommendation GSM 06.60, Enhanced Full Rate (EFR) speech transcoding, Version 8.0.1, 1999.

[14] ETSI Recommendation GSM 06.90, Adaptive Multi-Rate (AMR) speech transcoding, Version 7.2.1, 1998.

[15] Mouly, M., Pautet, M. B.: The GSM System for Mobile Communications, *(published by the authors)*, Palaiseau, 1992.

[16] Synopsys, Inc.: CoCentric System Studio User Guide, Version 2001.08.

[17] Zahorian, A., Jagharghi, A. J.: Minimum Mean Square Error Transformations of Categorical Data to Target Positions, *IEEE Trans. on Signal Processing, Vol. 40, No. 1, S. 13-23*, New York, January 1992.