QUALITY CONTROL FOR AMR SPEECH CHANNELS IN GSM NETWORKS

Marc Werner, Thomas Junge, and Peter Vary

Institute of Communication Systems and Data Processing (ind), Aachen University, Germany

ABSTRACT

In a competitive environment of cellular radio networks, the speech quality experienced by the end user is of vital importance for the network operator. Therefore, the operator needs reliable measures for monitoring and controlling the speech quality.

For GSM mobile radio networks employing the Adaptive Multi-Rate (AMR) speech codec, a new instrumental and non-intrusive speech quality metric is proposed. 'Non-intrusive' means that the metric is based only on received transmission parameters, as e.g. the frame erasure rate, and neither needs the original nor the transmitted speech samples. However, for the validation of a new metric, standardized intrusive speech quality measures like PESQ, link-level simulations and field measurements are used.

Furthermore, for an improved control of speech quality, a novel AMR mode switching procedure is introduced which outperforms the recommended method of the GSM standard.

1. INTRODUCTION

Due to its subjective nature, speech quality is hard to quantify. Most accurate results are obtained by listening tests carried out with respect to a specified environment and testing procedure. Listening test results are summarized in a quality figure called Mean Opinion Score (MOS), an average value of assessments on a five-point scale (5 = excellent, 1 = bad). This value characterizes the quality of the device under test (e.g., a speech codec). Listening tests are expensive and time-consuming. With some sacrifice to accuracy, instrumental (objective) speech quality measures have been developed which evaluate the original and distorted digitized speech samples using a psycho-acoustically motivated algorithm to predict the outcome of subjective listening tests. An example of such objective measures is PESQ [1] [2] which has been recommended for a variety of testing situations. Most elaborate objective speech quality measures are unsuitable for quality monitoring in mobile radio networks as they need speech samples from both transmitter and receiver (they are therefore called 'intrusive' measures).

Speech quality measurement approaches for monitoring purposes should solely rely on some transmission parameters available at the receiver to enable network functions like hand-over or transmit power control. In the GSM system, these parameters include the channel bit error rate, the received power level and the distance from mobile to base station. These figures are constantly measured by the mobile station and reported back to the BTS (Base Transceiver Station) every 480 ms.

The relation of these parameters to speech quality is different for each codec used. In the case of the Adaptive Multi-Rate (AMR) codec, the net data rate for compressed speech is adapted to the current channel quality, allowing a more reliable transmission for bad radio channels and resulting in an improved quality for those cases. The AMR mode number is therefore another parameter to be taken into account.

Approaches to the measurement of speech quality on the basis of transmission parameters have been proposed for the Full Rate (FR), Enhanced Full Rate (EFR) and AMR codecs [3] [4] [5]. The metrics presented in section 2 of this paper are based on GSM measurements and bit-exact transmission simulations. The PESQ scores of simulated speech samples serve as reference speech quality values.

The AMR codec is used for quality control in digital mobile networks. In GSM, it allows a dynamic split of the gross data rate per user, 22.8 kbit/s, between speech and channel coding. A trade-off between fundamental codec speech quality and required error correction capabilities is made for changing channel conditions. AMR speech channels will also be used in the upcoming UMTS networks, complementing the flexibility on the radio link.

The AMR mode is changed at the transmitter after a mode request message from the receiver (i.e., if the transmission has been too bad or better than expected during the past time interval). An exact measurement of the current reception quality is therefore essential for the AMR operation. In the GSM case, it has been recommended [7] to evaluate a filtered version of real-valued channel samples at the receiver for this purpose. However, the correlation with the resulting speech quality is not very high in all cases. Initial studies [10] suggested that it is worthwhile to use the "soft" reliability values from the Viterbi channel decoder output as a measure for the current channel quality. It was shown that this can improve the speech quality significantly for the case of a UMTS-AMR speech channel with fixed channel coding, where the employment of different AMR modes results in different spreading factors and bit energies.

In section 3, this method is extended to the GSM-AMR case with optimized switching points for the different AMR modes offering a variable channel coding rate. It is compared to the recommended procedure [7].

2. PARAMETER-BASED SPEECH QUALITY MEASURES

The development of speech quality metrics relying only on GSM-EFR transmission parameters has been presented previously [5]. It was shown that it is possible to predict the subjective speech quality very accurately. The metric optimization procedure for the EFR codec is extended here to the case of the AMR codec.

Functions of GSM transmission parameters which maximize the linear correlation coefficient ρ with the reference PESQ scores are identified. The maximization of this correlation is based on averaging functions for individual parameters over each speech sample and on the linearization of the mapping functions between parameter norms and quality values.

The parameters found to exhibit the highest correlation with the resulting speech quality are:

1. RxQual: The channel bit error rate (BER) is averaged over an interval of 480 ms and mapped to the logarithmic RxQual parameter with eight BER ranges [6]: RxQual $= 0$ corresponds to a BER from 0% to 0.2%, while RxQual $= 7$ indicates a BER $> 12.8\%$.

2. AMR mode: If the AMR codec is employed in GSM networks, the chosen AMR mode is recorded. The lower the mode, the higher the error correction capabilities, but the lower the inherent speech quality. The GSM-AMR implementation allows a selection of four (out of eight possible) different speech- and channel coding schemes. A combination of AMR modes $7, 6, 5$, and 0 was identified by simulation to deliver an optimum speech quality for a broad range of channel conditions.

3. FER: The Frame Erasure Rate for speech frames is generated from Bad Frame Indicators (BFI) of the CRC for class-I-bits.

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

To obtain reference PESQ scores, bit-exact link-level simulations of the GSM system were performed using the CoCentric System Studio [8] software. The transmission simulation includes the AMR speech and channel coders, a dynamic radio channel (the channel BER from GSM field measurements served as an input to the channel simulation), demodulator/equalizer, Viterbi channel and speech decoders. For the AMR mode switching, an optimized algorithm presented in section 3 was employed. The resulting speech samples were recorded and analyzed with the PESQ measure. Some additional parameters like FER and AMR mode were also derived from the simulations. Although the FER is not part of the standard GSM measurement report, FER values are usually computed within the Operation and Maintenance Center (OMC).

As a first step, the progression of the parameters $\zeta_i(k)$ (like RxQual, FER, etc.) was identified for each speech sample i (k serves as a discrete time index). An average value of each parameter was obtained by calculating the so-called L_P-norms per speech sample,

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

for various values of P. 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. Large 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 f of degree $m = 4$:

$$M_i \approx f(L_P(\zeta_i(k))) \qquad (2)$$

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

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

As an example, the correlation of the AMR mode and the speech quality score is examined. Although a lower AMR mode performs better in terms of error correction, the inherent speech quality becomes lower. This fact can be observed in Figure 2, where the $L_{0.2}$ norms of the AMR modes per speech sample are plotted without linearization against the respective PESQ values. The linearization polynomial f_{AMRmode} is indicated as well. After linearization, i.e., taking the function values $f_{\text{AMRmode}}(L_{0.2}(\zeta_i(k)))$, the correlation with the PESQ scores amounts to $\rho = 0.9259$.

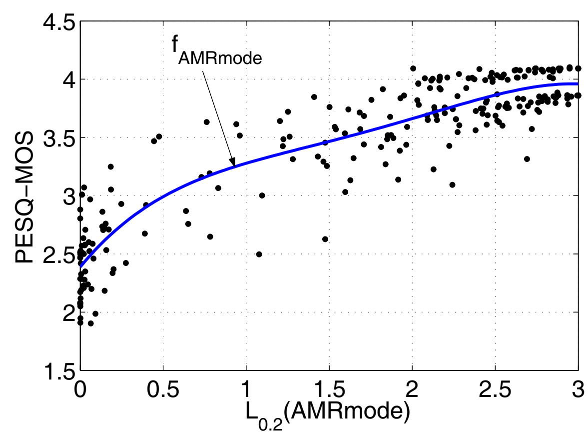

Figure 1: Correlation of AMR and PESQ (mode norms $\in [0\ldots3]$ calculated from codec mode indices $\{0, 1, 2, 3\}$)

Table 1 gives an overview of all obtained parameter correlations, using optimum L_P-norms, after linearization by individual polynomials, for the parameters RxQual, FER, MnMxLFER and AMRmode.

parameter ζ	$\hat{P}_\zeta$ EFR	$\rho_\zeta(\hat{f}, M_i)$ EFR	$\hat{P}_\zeta$ AMR	$\rho_\zeta(\hat{f}, M_i)$ AMR
RxQual	6	0.9419	14	0.9317
FER	0.5	0.9632	1	0.9556
MnMxLFER	1	0.9383	1	0.9282
AMRmode	–	–	0.2	0.9259

Table 1: $\hat{P}$-values of L_P-norms, and resulting correlation ρ (after linearization by $\hat{f}$)

The large optimum P-values for RxQual indicate that outliers are perceived more strongly than it is suggested by the numerical value of this parameter. Note that for the FER parameter, $L_{0.5}(\text{FER}) = \sqrt{\text{FER}}$ and $L_1(\text{FER}) = \text{FER}$, because the constituent elements are taken from the set $\{0, 1\}$ or $\{\text{BFI, no BFI}\}$ only.

The above parameters were combined to obtain objective non-intrusive parameter-based speech quality metrics. The MSECT

(Minimum Mean Square Error Coordinate Transformation) [9] procedure was employed to optimize the mapping function with respect to a minimum mean squared error between the parameter vectors and estimated MOS scores from the PESQ algorithm.

The resulting speech quality measures are of the form

$$\begin{aligned} \text{SQM}_{\text{EFR}} &= T_1 \cdot f_1(L_6(\text{RxQual})) + T_2 \cdot f_2(\sqrt{\text{FER}}) \\ &+ T_3 \cdot f_3(L_1(\text{MnMxLFER})) + T_4 \end{aligned} \quad (3)$$

$$\begin{aligned} \text{SQM}_{\text{AMR}} &= T_5 \cdot f_5(L_6(\text{RxQual})) + T_6 \cdot f_6(\text{FER}) \\ &+ T_7 \cdot f_7(L_1(\text{MnMxLFER})) \\ &+ T_8 \cdot f_8(L_{0.2}(\text{AMRmode})) + T_9 \end{aligned} \quad (4)$$

with optimized values for T_ζ, where the value ranges of f_ζ are comparable. The weighting factors T_ζ in Eqns. 3 and 4 indicate a prominent importance of the parameter FER.

A fraction of the available speech samples and PESQ scores was chosen as training data for the MSECT algorithm. The correlation coefficient of SQM_{EFR} and the PESQ values was $\rho_{\text{EFR}} = 0.9527$ based only on datasets excluding the training data. For the extension to the AMR codec, the resulting correlation coefficient of SQM_{AMR} is $\rho_{\text{AMR}} = 0.9516$.

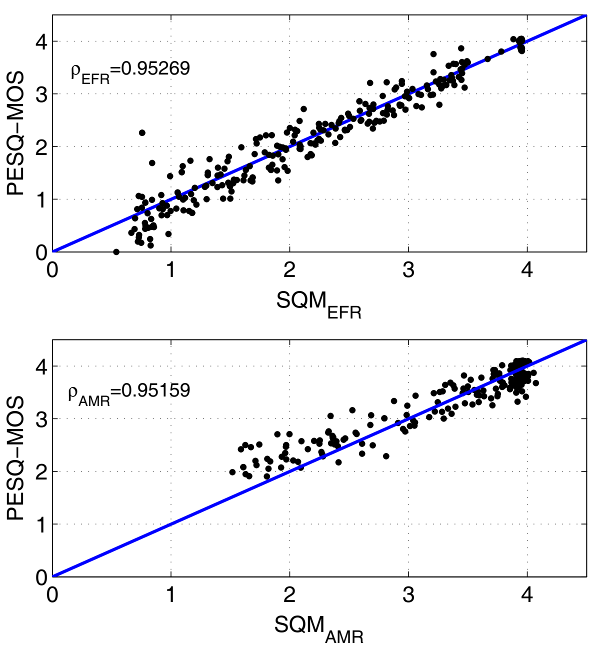

Figure 2: Correlation of SQM_{EFR}/SQM_{AMR} and PESQ

These correlations are depicted as scatter-plots in Figure 2. The high degree of correlation can be clearly observed for both the EFR (upper figure) and the AMR speech quality metric (lower figure).

Because of the adaptation of the speech codec to the current channel condition, the AMR transmission yields a better speech quality than the constant-rate EFR transmission. This quality improvement is visible in Figure 2: All quality scores are above 1.5 for the AMR codec. The range of channel quality conditions was identical for both simulation groups.

It should be noted that the correlation values presented above are only valid for the employed mode switching algorithm, described in section 3.

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 different.

3. AMR MODE SWITCHING ALGORITHM

The originally proposed AMR mode switching procedure [7] estimates the current channel quality by analyzing filtered soft channel values at the demodulator/equalizer output. It was observed from transmission simulations that these channel values do not exhibit a high correlation with the resulting speech quality in all cases. Therefore a different approach was developed [10] for the case of UMTS-AMR transmissions using a constant convolutional channel code of rate $r = 1/3$ for all AMR modes (a more reliable transmission of lower AMR modes is in this case achieved by a higher spreading factor, resulting in a greater number of chips per information bit and a higher bit energy). A monotonic function of the absolute soft output of the Viterbi channel decoder was shown to be highly correlated to the residual BER, which in turn is a good indicator of the expected speech quality. A delay/hysteresis was implemented by allowing a switching of modes at the end of each analysis window of 25 speech frames (0.5 s).

The approach to use soft samples from the channel decoder output is transferred here to the GSM-AMR case with different channel codes for each mode. The decision for switching to another AMR mode is based on an interval of N bit-wise soft values at the Viterbi decoder output, where N represents the number of bits per speech frame in the selected AMR mode. The soft values are represented in the log-likelihood domain of L-values. The absolute L-values are summed over one speech frame of duration $T = 20$ ms,

$$S = \sum_{i=kN}^{(k+1)N} |L(i)| \quad , k = 0, 1, 2, \ldots \quad (5)$$

Theoretically, S takes on values from $[0, \infty]$. In the examined systems, values between 20 and 6000 were observed. It has been shown [10] that the BER can be expressed accurately by a monotonic function of S. This function, however, is dependent on the channel coding scheme employed. While for the UMTS case, the channel coding had been constant ($r = 1/3$ convolutional code), the GSM-AMR implementation takes into account the selection of four different speech- and channel coding schemes.

A quality-based method for defining the mode switching and hysteresis values in the S domain was developed. These S-values were identified from a speech quality analysis of simulated AMR speech samples.

The method is illustrated in Figure 3, which depicts the relation of channel BER, S-values and resulting PESQ scores. In the upper part of Figure 3, initial thresholds for downward switching (dashed lines) and upward switching (solid lines) are indicated. The downward thresholds were placed in the channel BER domain at the intersections of the speech quality functions for different AMR modes, resulting in an optimal mode for all BER conditions. A switching hysteresis is realized by an offset placement of the upward thresholds.

For the example of the threshold for switching from mode 7 to 6, the identification of the equivalent S value is depicted

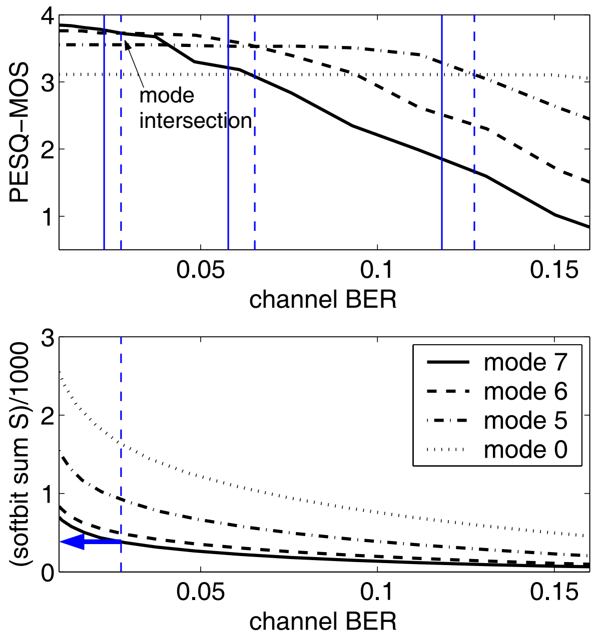

Figure 3: Example of switching threshold definition

in the lower part of Figure 3. It is selected as the S function value at which the BER threshold crosses the corresponding BER-S curve of the current AMR mode. The initial threshold values were slightly adjusted to maximize the resulting speech quality. A mode switching event occurs if more than half of the S measures within a sliding 0.5 s interval exceed the mode-specific upper or lower limit.

This AMR mode switching method ensures that the AMR modes follow the rule of maximizing the resulting speech quality. Optimized sets of switching thresholds were determined for the selected AMR modes.

An example of the new AMR mode switching operation is shown in Figure 4 for an example channel quality progression taken from GSM measurements. Compared to the old method,

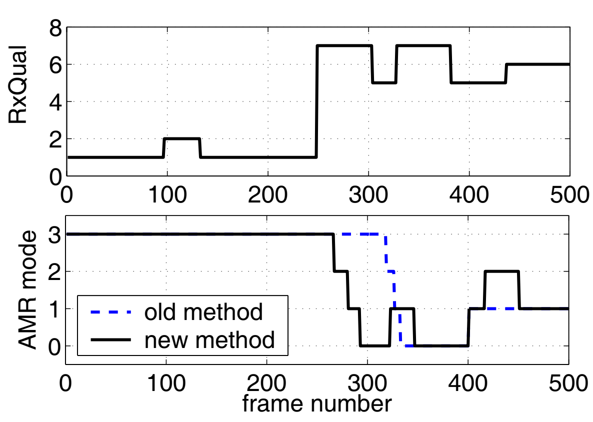

Figure 4: Example of the AMR mode switching

a quicker reaction to varying channel qualities is observed. The PESQ score is improved from 2.748 to 3.011 in this example.

The difference in speech quality due to the new AMR mode switching algorithm (compared to the old method) was calculated for an equally distributed range of channel conditions. A significant improvement due to the new switching algorithm is observed especially for bad and rapidly changing channels. For these cases, the speech quality profits from a faster and more accurate switching procedure. For the channel conditions which resulted in a poor or bad speech quality (PESQ score with old switching method < 2.5), an improvement was observed in 78% of samples. The average improvement was $E\{\Delta\text{PESQ}\} = 0.168$. For good channel conditions, the average speech quality remained roughly on the same level ($E\{\Delta\text{PESQ}\} = 0.02$), because mode switching occurs more seldom.

4. CONCLUSIONS

Empirical mapping functions of GSM-EFR and GSM-AMR transmission parameters were presented which allow a non-intrusive estimation of the objective speech quality in GSM telephony. High correlations with a reference speech quality measure were achieved by linearization and L_P-norm averaging methods. The proposed mode switching algorithm for the GSM-AMR system was shown to be superior to the standard method especially for the case of bad transmission channel characteristics. Both the quality metric and switching algorithm were based on extensive GSM measurements and link-level simulations for development and evaluation. The proposed measurement and control methods can be easily applied to existing GSM-AMR networks and allow an efficient monitoring and optimization of speech quality.

REFERENCES

[1] 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.

[2] Rix, A. W. et al. (I) / Beerends, J. G. et al. (II): PESQ, the new ITU standard for objective measurement of perceived speech quality, Parts I and II, *J. Audio Eng. Soc., vol. 50*, Oct. 2002.

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

[4] 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, Jan. 2002.

[5] Werner, M. et al.: Parameter-based Speech Quality Measures for GSM, *Proc. IEEE Intl. Symposium Personal, Indoor and Mobile Radio Communications (PIMRC)*, Beijing, Sep. 2003

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

[7] ETSI Recommendation GSM 05.09, Link Adaptation, Version 8.5.0, 1999.

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

[9] Zahorian, A., Jagharghi, A. J.: Minimum Mean Square Error Transformations of Categorical Data to Target Positions, *IEEE Trans. Signal Processing, Vol. 40, No. 1*, Jan. 1992.

[10] Werner, M., Vary, P.: Speech Quality Improvement in UMTS by AMR Mode Switching, *Proc. 4th Intl. ITG-Conference on Source and Channel Coding*, Berlin, Jan. 2002