Quality Control for UMTS-AMR Speech Channels

Marc Werner, Peter Vary

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

{werner | vary}@ind.rwth-aachen.de

Abstract

In UMTS speech transmissions, the Adaptive Multi-Rate (AMR) speech codec is employed to allow for a dynamic assignment of data rates to individual users. The control of AMR modes is based on quality measurements of the transmission channel and aims at the maximization of speech quality by selecting the mode which is best suited to the current interference situation. In contrast to GSM, a reduction of AMR modes also leads to an increased cell capacity in UMTS. In this paper, a decentralized method for AMR mode switching is presented which is based on individual softbit measurements at the Viterbi channel decoder and optimizes the speech quality for a wide range of channel conditions. This method was developed by optimizing the AMR mode switching thresholds with respect to PESQ speech quality scores. The second part of the paper describes a new instrumental non-intrusive method for speech quality measurement by the evaluation of UMTS transmission parameters for application to UMTS-AMR speech transmissions. High correlations with the reference PESQ scores were observed.

1. Introduction

Although data services are gaining more significance in 3G mobile communication systems like UMTS, speech telephony remains the most frequently used service. The subjective speech quality experienced by the end user can therefore be regarded as one of the most significant quality aspects of mobile communication. In the radio network, a well-designed Radio Resource Management guarantees a satisfactory transmission quality for varying link qualities and load conditions. It is of great importance that these mechanisms are designed in a way that maximizes the quality of speech while not wasting resources like power, spreading codes and data rates. Additionally, methods for an automated quality monitoring help to identify shortcomings of the radio network.

The AMR codec [1] is specified as the speech codec for GSM and UMTS. It allows a frame-wise adaptation of each user's net data rate R_x in a range between 4.75 kbit/s (mode 0) and 12.2 kbit/s (mode 7). While the higher AMR modes offer a better inherent speech quality, their robustness against interference on the radio transmission channel is weaker. In GSM systems, different levels of error-correction performance are established by different channel coding schemes of the AMR modes.

For UMTS speech channels, a constant $r = 1/3$ convolutional code is employed and the energy of net bits $\hat{x}(k)$ (fed into the speech decoder) increases with a decreasing data rate R_x. The gains in robustness are then realized by a higher spreading factor $\mathrm{SF} = 2^n$, $n \in \{2, 3, \ldots, 9\}$, or, equivalently, a greater number of bit repetitions during the rate matching procedure of UMTS (rate matching: puncturing or repetition of channel coded bits to meet the data rate R_c/SF of the UMTS physical channel; see Figure 1). At the receiver, SF chips contribute to the energy of one channel bit. Additionally, the rate dematching averages over repeated bits, thereby reducing the bit error rate. The AMR mode should always be chosen as high as possible (better inherent speech quality) but as low as necessary (sufficient robustness and/or higher UMTS cell capacity). The method of AMR mode switching is not specified in UMTS; it is, however, clear that a switching decision at the transmitter will be based on a transmission quality measurement at the receiver.

In Section 2, a switching method which we previously derived for GSM [2] is applied to the UMTS-AMR case where switching thresholds are optimized with respect to the resulting speech quality measured by the objective speech quality measure PESQ [3]. Instrumental speech quality measures like PESQ predict the results of subjective listening tests. Most of these algorithms perform a transformation of the original and degraded speech signals to an auditory representation by considering psychoacoustic effects like non-linear perception of frequency and loudness, or masking. An analysis of these representations results in a quality figure which exhibits a high correlation with results of subjective tests, expressed on the five-point MOS scale (5=excellent, 1=bad).

Objective speech quality measures which need the original speech sample as input are called 'intrusive' measures. PESQ is an intrusive measure proposed by the ITU for the measurement of telephone speech quality. 'Non-intrusive' speech quality measures are based only on information available at the receiver and are therefore suited for quality monitoring during normal operation of a mobile communication system. Approaches to the measurement of speech quality on the basis of transmission parameters have been presented for the GSM Full Rate (FR), Enhanced FR (EFR) and AMR codecs [4] [5] [2].

A new non-intrusive instrumental speech quality measure for UMTS-AMR transmissions employing the aforementioned mode switching method is presented in Section 3. It can be easily adapted to other UMTS-AMR systems.

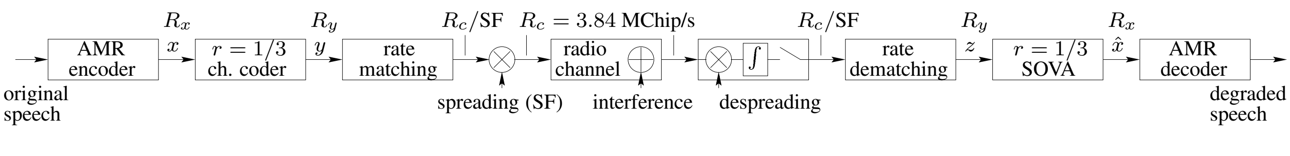

Figure 1: UMTS-AMR transmission chain

2. Optimized Speech Quality by AMR Mode Switching

The adaptation of AMR modes to changing transmission conditions should maximize the resulting speech quality by choosing a compromise between inherent speech quality and robustness against interference. It is assumed that in UMTS, a preselection of four AMR modes takes place like in GSM. As the characteristics of neighbouring AMR modes differ only slightly, the quality loss induced by the reduction of the number of possible modes from eight to four is small. On the other hand, a selection of a mode subset allows for a quicker link adaptation as the step size of net data rates becomes larger, and also reduces the number of switching instants. Our UMTS link-level simulations suggest that the selection of modes 0, 2, 5 and 7 represents a good choice, covering a wide range of transmission conditions.

The AMR switching thresholds are based on an estimation of each mode's speech quality characteristics and can be determined by link-level simulation. The method of estimating the current transmission quality is based on an evaluation of received likelihood values after Viterbi convolutional decoding.

At the convolutional channel decoder, the Euclidean distances of the real-valued received sequence $z(k)$ to possibly transmitted bit sequences $y(k)$ are determined by Viterbi decoding, and the most probable sequences $\hat{y}(k)$ and $\hat{x}(k)$ are identified (see Figure 1). In the case of the *Soft Output Viterbi Algorithm* (SOVA) [6], reliability values are calculated from the decoded path metrics for each output bit. These so-called L-values (likelihood values) may be defined as a function of conditional error probabilities $p(k)$,

$$L(k) = \ln\left(\frac{1-p(k)}{p(k)}\right) \geq 0, \tag{1}$$

where $p(k) \leq 0.5$ is the approximated a-posteriori error probability of the bit with time index k, calculated by the SOVA.

The exact calculation of L-values in the SOVA requires some prerequisites at the decoding stage [6]: The calculation of the Euclidean distance leads to exact error probabilities only for the case of an AWGN channel. Additionally, for each input bit to the SOVA decoder, an exact estimation of the current channel SNR is needed (channel state information, CSI) for a correct mapping of the path metrics to L-values. These prerequisites are often not met in real mobile radio systems. The aberration of the calculated L-values $\tilde{L}(k)$ without AWGN/CSI-condition is difficult to predict due to the recursive and nonlinear derivation rule. However, for the purpose of a frame-wise estimation of the current transmission quality, a monotonic function describing the frame-wise net bit error rate (BER) with adequate accuracy is sufficient.

An interpretation [7] of simulation results concluded that the mean over one speech frame (20 ms) of N absolute SOVA soft output bits $\tilde{L}(k)$ exhibits a very high correlation with the net BER for the case of an AWGN channel without CSI,

$$S(i) = \frac{1}{N}\sum_{k=iN}^{(i+1)N} \tilde{L}(k), \quad i = 0, 1, 2, \ldots . \tag{2}$$

This result was verified in UMTS link-level simulations with explicitly modelled intra-interference (see Figure 2). OCNS (Orthogonal Code Noise Sequence) interference signals were used, suspending the AWGN condition. The correlation coefficient ρ of the linearized relation between $S(i)$ and net BER, which was identified for all eight AMR modes and

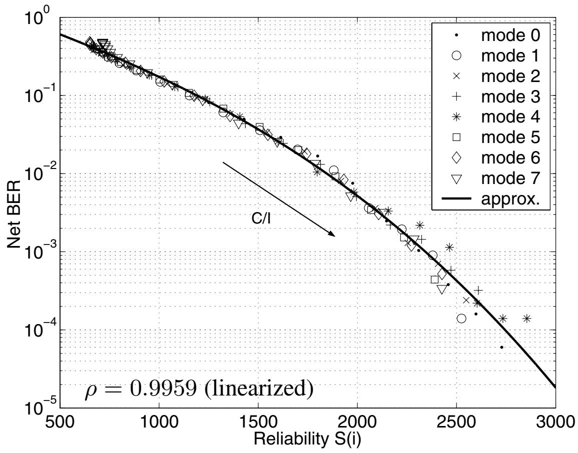

Figure 2: Correlation of mean likelihood values and net BER

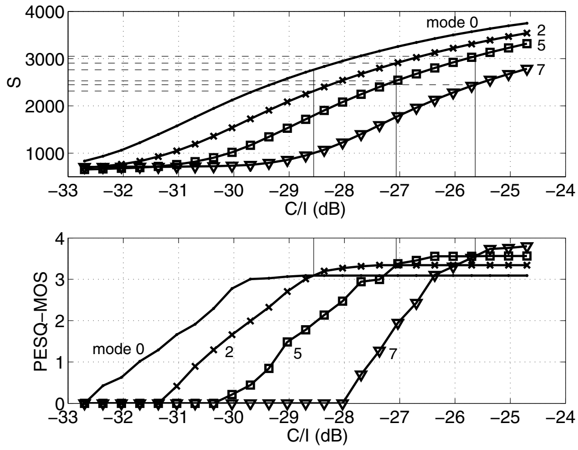

Figure 3: Optimum switching thresholds in different domains

well approximated by a second-degree polynomial, is sufficient for an exact estimation of the frame-wise transmission quality ($\rho = 0.9959$). The absolute values of S depend on the particular implementation of the channel decoder.

To exploit the correlation between mean likelihood values and net BER, optimum switching thresholds were identified that maximize the speech quality in the S domain. The procedure of identifying these thresholds is depicted in Figure 3. The optimum thresholds are derived from intersections of the PESQ curves in the lower diagram of Figure 3. At these intersections, the AMR mode should be changed to remain on the upper envelope of PESQ curves. The C/I values corresponding to these intersections are transferred to the upper diagram, where a mapping to the S domain takes place using the curve for the current AMR mode. The procedure is illustrated in Figure 3 with thin lines. The C/I values in Figure 3 represent raw carrier-to-interference ratios without considering orthogonality and spreading gains.

The three lower switching thresholds for the different AMR modes are closely spaced together. This was to be expected: The effective channel codes for the individual AMR modes are slightly different due to the different rate matching. However, the mapping of S values to net BER is independent of the AMR mode. The net BER for which a first measurable degradation of speech quality occurs is equal for all AMR modes at about 10^{-4}. This yields the same optimum S-limits for all AMR modes. The three upper switching thresholds are placed at an offset S value from the lower thresholds, resulting from the shift in the initial modes.

In link-level simulations covering multiple male and female speech samples, mean upper and lower switching thresholds of $S_\text{u} = 3200$ and $S_\text{l} = 2500$ were identified, including an appropriate hysteresis. If these thresholds are exceeded for, e.g., a certain fraction α of an observation interval T_0, a mode change is initiated. The choice of T_0 represents a trade-off between a quick adaptation of modes and a prevention of excessive switching. The settings $T_0 = 0.5$ s and $\alpha = 0.5$ were used in the simulations described here.

3. Parameter-based Speech Quality Measures for UMTS-AMR

UMTS transmission or measurement parameters available for a speech quality analysis at the receiver and their reporting frequency depend on the manufacturer's hard- and software implementation and on the network configuration [8]. Some parameters specified for measurement reports of the mobile station and for measurements at the base station are given in Table 1.

Quantity	Downlink Parameter	Uplink Parameter
power	received total power, rec. CPICH power, transmitted power	received total power, trans. carrier power, trans. code power
C/I	CPICH-E_c/N_0	SIR, SIR-Error
BER/FER	transport channel BLER	transport ch. BER, physical ch. BER

Table 1: Some UMTS measurement parameters [8] (CPICH: UMTS Common Pilot Channel)

Taking into account previous results regarding parameter-based speech quality measures for GSM [2], a preselection of suitable UMTS parameters for a prediction of the resulting speech quality can be undertaken. It was shown that parameters describing the total received power exhibit a relatively low correlation with the resulting speech quality.

The Outer Loop Power Control measures the uplink received signal-to-interference power ratio (SIR) and compares the measurements to a target value. On the downlink, the block error rate (BLER) is measured instead. The uplink BLER can also be derived from BER estimations [9].

The UMTS BLER is equivalent to the GSM FER (Frame Erasure Rate) parameter if one AMR speech frame is transmitted within one TTI (Transmission Time Interval), which is the specified procedure for the UMTS reference channels of $R_x =$ 12.2 kbit/s [10]. Excellent quality correlations had been observed for the GSM FER and the derived parameters LFER, MnMxLFER and MxLFER [2]. The LFER measures the sequence lengths of consecutively erased speech frames, and the MxLFER is the maximum of all LFER lengths within a speech sample. The MnMxLFER records the four LFER maxima from the first, second, third and fourth quarter of the speech sample. These FER derivations are motivated by the tendency of human perception to over-emphasize strong local disturbances. The AMR mode itself is taken into account as a transmission parameter as well. A system with dynamic adaptation of AMR modes to the current channel quality as described in Section 2 is assumed here.

Parameter ζ	$\hat{P}$	$\rho(\hat{f},\, M_i)$
FER	4	0.9752
LFER	10	0.9239
MxLFER	–	0.9331
MnMxLFER	1	0.9525
AMRmode	2	0.8761

Table 2: Optimum P-values and correlation coefficients

In link-level simulations of the UMTS radio transmission, the correlation of the above parameters with PESQ speech quality scores was analysed. The simulations were performed using a transmission model for the UMTS 12.2 kbit/s downlink reference channel [10] within the Synopsys System Studio software package [11], extended by AMR coding and decoding functionality. The simulations employ a spreading factor of 128 and a TTI of 20 ms. Over 800 simulations were carried out under different channel conditions. A first-order Markoff process was implemented to introduce additional slow fading effects. Regarding multipath, it was assumed that all components can be resolved by the rake receiver so that nearly the total signal power of all paths is available. Apart from the received speech itself, the Frame Erasure flags $\in \{0, 1\}$ were recorded during the simulation at constant time intervals of 20 ms.

The parameter progression $\zeta_i(k)$ of the parameter $\zeta \in$ {FER, LFER, MxLFER, MnMxLFER, AMRmode} is evaluated for each speech sample i of duration T_i. To compare these parameter progressions with the single PESQ speech quality score M_i, L_P norms of $\zeta_i(k)$ are calculated:

$$L_P(\zeta_i(k)) = \left[\frac{1}{N} \sum_{k=1}^{N} (\zeta_i(k))^P \right]^{1/P}, \quad P \in \mathbb{R}^+. \quad (3)$$

The reason for calculating different L_P norms for each parameter progression is the different psychoacoustic perception of parameter outliers. Large values of P amplify parameter variations whereas small P-values have a smoothing effect.

Monotonic linearization polynomials f were identified for each parameter ζ and different P-values to approximate the L_P norms of all speech samples i to the reference PESQ values M_i in an MMSE sense:

$$M_i \approx f(L_P(\zeta_i(k))) \quad \forall i. \quad (4)$$

The correlation coefficient $\rho(f(L_P(\zeta_i(k))),\, M_i)$ is calculated after a transformation of the L_P norms to the PESQ domain by the polynomial f. The optimum P-value $\hat{P}$ which maximizes the correlation is identified along with the corresponding linearization polynomial $\hat{f}$ for each parameter ζ. The values for $\hat{P}$ and ρ are given in Table 2.

By evaluating only the FER, a good speech quality estimation rule can be derived:

$$\text{UFSQM} = f_0(L_4(\text{FER})), \quad (5)$$

where f_0 represents the optimum linearization polynomial $\hat{f}$ for the FER parameter, and $L_4(\text{FER}) = \text{FER}^{1/4}$ because the constituent elements are taken from $\{0, 1\}$. The correlation of the UFSQM measure with PESQ scores ($\rho = 0.9603$) is indicated in the upper diagram of Figure 4.

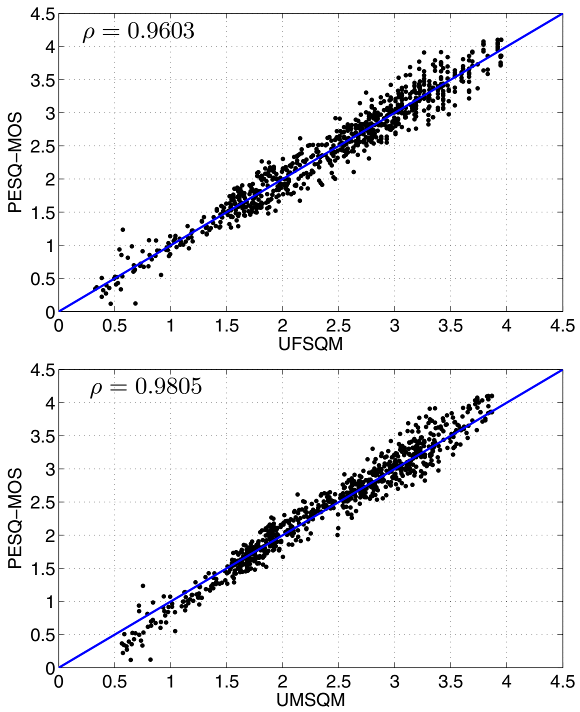

Figure 4: Correlation of two new UMTS speech quality measures with PESQ scores: UFSQM, eq. (5); UMSQM, eq. (7)

A speech quality measure based on a combination of multiple parameters can be generated by optimizing a mapping function from the multi-dimensional vector of their linearized L_P-norms to the one-dimensional quality score domain. For this purpose, the MSECT algorithm (Minimum Mean Square Error Coordinate Transformation) [12], a multi-dimensional MMSE mapping procedure with categorized reference points, was employed as in earlier studies for GSM [2].

For the combination measure, the two parameters FER and MnMxLFER which exhibit the highest PESQ correlation were selected in addition to the AMR mode:

$$\mathbf{v} = \begin{pmatrix} f_1(L_4(\text{FER})) \\ f_2(L_1(\text{MnMxLFER})) \\ f_3(L_2(\text{AMRmode})) \end{pmatrix}, \tag{6}$$

$$\text{UMSQM} = \mathbf{T} \cdot \mathbf{v} + o, \tag{7}$$

with a mapping vector $\mathbf{T}$ and offset value o generated by the MSECT algorithm. The correlation of the UMSQM measure with PESQ values was enhanced to $\rho = 0.9805$. It is depicted in the lower diagram of Figure 4.

4. Remarks

The presented correlation coefficients were calculated on the basis of speech samples that had *not* been used during the generation of the speech quality measures. The coefficients of linearization polynomials as well as mapping vectors and offset values depend on the specific configuration of a UMTS link (spreading factor, rate matching, power control, etc.). For different network configurations, their values can be calculated with the described procedures.

The correlation of the presented non-intrusive speech quality measures with PESQ scores was maximized. However, PESQ scores themselves are only a (good) approximation of the MOS scores from subjective listening tests. Therefore, the correlation figures of the presented measures with listening test MOS scores might be slightly lower than the correlation coefficients presented.

5. Conclusions

An AMR mode switching algorithm for the UMTS-AMR system was proposed which is based on a quality analysis of UMTS speech transmissions. The switching thresholds in the domain of likelihood values were selected with respect to an optimum speech quality. In GSM systems, the employed switching procedure delivers a performance gain over the standardized method [2]. Its reliable operation in UMTS systems was verified by link-level simulations. In loaded cells, a reduction of the AMR data rate also increases the user capacity of the system.

Systematic mapping functions of UMTS-AMR transmission parameters were presented which allow a non-intrusive estimation of the objective speech quality in UMTS. High correlations with the PESQ reference speech quality measure were achieved by linearization and L_P-norm averaging methods. Quality metrics and the switching algorithm were based on extensive link-level simulations.

6. References

[1] 3G TS 26.071, “AMR Speech Codec - General Description”, ver. 3.0.1, Release 99.
[2] M. Werner, T. Junge, P. Vary: “Quality Control for AMR Speech Channels in GSM Networks”, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004.
[3] 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”, 2001.
[4] A. Karlsson et al.: “Radio Link Parameter based Speech Quality Index - SQI”, Proc. IEEE Workshop on Speech Coding, 1999.
[5] S. Wänstedt 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, 2002.
[6] J. Hagenauer, P. Hoeher.: “A Viterbi Algorithm with Soft-Decision Outputs and its Applications”, Proc. IEEE Globecom, 1989.
[7] M. Werner, P. Vary: “Speech Quality Improvement in UMTS by AMR Mode Switching”, Proc. 4th Intl. ITG-Conference on Source and Channel Coding, 2002.
[8] 3G TS 25.215, “Physical layer - Measurements (FDD)”, ver. 3.4.0, Release 99.
[9] K. Heck et al.: “Diversity Effects on the Soft Handover Gain in UMTS networks”, Proc. IEEE Vehicular Technology Conference (VTC) Fall, 2002.
[10] 3G TS 34.121, “Terminal Conformance Specification; Radio Transmission and Reception (FDD)”, ver. 3.9.0, R99.
[11] Synopsys, Inc.: “CoCentric System Studio User Guide”, ver. 2004.09.
[12] A. Zahorian, A.J. Jagharghi: “Minimum Mean Square Error Transformations of Categorical Data to Target Positions”, IEEE Trans. on Signal Processing, Vol. 40, No. 1, Jan. 1992.