SPEECH QUALITY IMPROVEMENT IN UMTS BY AMR MODE SWITCHING

M. Werner, P. Vary

Institute of Communication Systems and Data Processing (ind)
Aachen University of Technology
{werner, vary}@ind.rwth-aachen.de

ABSTRACT

In the UMTS cellular radio system the so-called *Adaptive Multi-Rate* codec (AMR) [3] will be used for speech transmission. The bit rate of the AMR codec can be controlled by the *Radio Resource Management* (RRM). The highest speech quality can be achieved at a net bit rate of 12.2 kbit/s whereas the highest traffic capacity corresponds to the lowest bit rate of 4.75 kbit/s.

Simulations of the uplink of the UMTS physical layer have shown that the number of simultaneous speech traffic channels is limited to 73...149 per cell depending on the speech codec mode. These capacity figures are valid under some idealizing assumptions such as perfect power control and multipath compensation.

An algorithm for AMR mode switching is presented which which can significantly improve the speech quality in UMTS systems. The algorithm is controlled by evaluating the soft output likelihood values of the channel decoder.

1. INTRODUCTION

The recent UMTS standard (*Universal Mobile Telecommunications System*), of a third generation digital communication network, has defined a flexible radio access scheme based on Wideband CDMA (*Code Division Multiple Access*). Different data rates and channel coding schemes allow a large variety of different mobile services.

A crucial issue for network operators is the optimization of the service quality, especially the radio coverage, the data rates and the service dependent criteria such as speech quality or bit error rate.

These criteria depend on the Radio Resource Management, e.g. assignment of transmit powers, spreading factors and channel coding schemes to the active users in a radio cell according to individual mobile services. The influence of the corresponding parameter settings on service quality and network capacity are investigated.

2. SIMULATION OF BIT TRANSMISSION

To investigate the effects of transmission parameter settings on the speech quality, a baseband simulation model of the UMTS-FDD radio link [1] was set up in the COSSAP simulation environment [2]. The main cause of signal degradation in DS-CDMA (*Direct Sequence CDMA*) transmissions over mobile radio channels is additive uncorrelated interference. In a first approach, contermeasures such as Multi-User Detection are not considered. Furthermore, perfect power control and multipath compensation is assumed. Under this assumption, the radio channel can be modelled by Additive White Gaussian Noise (AWGN).

The basic components of the transmission model are shown in figure 1.

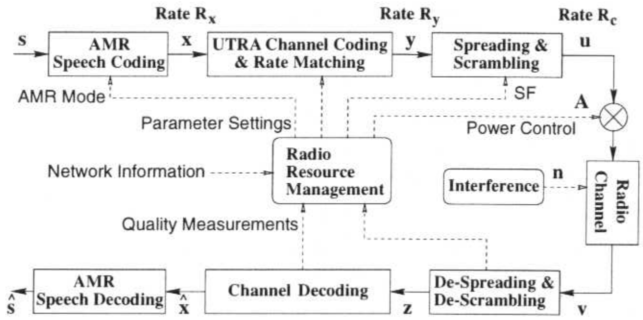

Figure 1: UMTS transmission model

The model contains the AMR speech codec with eight codec modes (0...7) with different net data rates R_x ranging from 4.75 kbit/s to 12.2 kbit/s (see table 1) [3]. The 12.2 kbit/s mode is identical to the enhanced full-rate speech codec employed in the GSM system.

AMR mode	**bit rate R_x** (kbit/s)	**speech quality** (estimated MOS)
7	12.2	3.6
6	10.2	3.5
5	7.95	3.2
4	7.4	3.2
3	6.7	3.1
2	5.9	3.0
1	5.15	2.7
0	4.75	2.6

Table 1: AMR codec modes, net rates R_x and inherent speech quality

The inherent speech quality for error-free transmission increases with net data rate, but switching to lower codec modes ensures a more reliable transmission over disturbed channels. To give some objective figures for each AMR codec mode, an estimated MOS (*Mean Opinion Score*) was calculated from the objective speech quality measure PSQM (*Perceptual Speech Quality Measure*) [4]. MOS values range from 5 (excellent) to 1 (poor). A difference in the MOS rating of 0.5 corresponds to a clearly perceivable difference of the subjective quality. But it should be noted that the estimated MOS values can be used for a relative comparison and their absolute values differ from those obtained by formal listening tests.

The measurement of transmission quality and the assignment of codec modes are studied in section 3.

The UTRA (*UMTS Terrestrial Radio Access*) channel coder is specified as an $r = 1/2$ or $r = 1/3$ convolutional coder with constraint length $k = 9$ (depending on the kind of service, turbo coding and no coding at all are possible). In this contribution, convolutional encoding and frame-by-frame soft-input soft-output Viterbi decoding was implemented. For BPSK/QPSK modulation used in UMTS, coding gains in the range of 4.5...6.5 dB of E_b/N_0 can be realized for the $r = 1/3$ code.

For physical CDMA-transmission, channel coded bits y of rate R_y are multiplied, or "spread" with a chip sequence c of higher rate R_c than the bit sequence y. The rate ratio R_c/R_y is called *Spreading Factor* (SF). The chip rate in UMTS is constant at $R_c = 3.84$ Mchip/s [5], and Spreading Factors are powers of two, SF $\in \{2^2 \dots 2^8\}$. Consequently, only certain discrete gross bit rates R_y are allowed to obtain the fixed chip rate R_c by spreading with the possible SFs:

$$R_c = \mathrm{SF} \cdot R_y \tag{1}$$

Table 2 shows the possible gross bit rates R_y for the allowed spreading factors.

SF	256	128	64	32	16	8	4
$\boldsymbol{R_y}$ (kbit/s)	15	30	60	120	240	480	960

Table 2: Spreading Factors and gross rates R_y for $R_c = 3.84$ Mchip/s

In order to obtain one of these discrete gross rates for all kinds of possible net rates R_x, a rate matching algorithm punctures or repeats bits after channel coding. In the case of puncturing, some channel coded bits are not transmitted. As the rate matching scheme is known at the receiver, the positions of punctured bits can be filled up again before channel decoding in the receiver. For the insertion of these bits, an amplitude of maximum uncertainty (± 0.0 for bipolar transmission) is used. The gross bit rate R_y after rate matching takes one of the values given in table 2.

The Spreading Factor SF is proportional to transmitter E_b/N_0 ratio,

$$E_b = \frac{R_y}{R_x} \cdot E_s = \frac{R_y}{R_x} \cdot \mathrm{SF} \cdot E_c \tag{2}$$

where E_b is the energy per information bit x, E_s is the energy per channel coded bit y, and E_c is the chip energy.

SF is also inversely proportional to the gross bit rate R_y (see equation 1).

If no puncturing is needed, the repetition of channel coded bits y by rate matching for a smaller SF yields the same gain in E_s as a higher SF because the energy of the repeated bits can also be used for estimation in the receiver. In both cases, the transmitter signal u is composed of the same number of chips for each bit y.

Under most conditions a better performance was observed when employing a more powerful channel coding scheme, resulting in a lower E_s, a lower SF and higher channel bit rate R_y, but also a better error correction. For example, the $r = 1/3$ code gains approximately 1 dB in E_b/N_0 compared to the $r = 1/2$ code with higher spreading or stronger repetition.

The assignment of different spreading and scrambling sequences makes the transmitted signals uncorrelated, or even orthogonal. OVSF (*Orthogonal Variable Spreading Factor*) codes, derived from Walsh codes, are used for spreading while PN sequences are employed for scrambling.

In the downlink, orthogonal OVSF codes are assigned for signal separation (orthogonality in OVSF codes holds only for synchronous transmission). In the uplink, different unsynchronized users' signals are separated by uncorrelated scrambling codes.

Interference in the downlink direction is caused by multipath propagation and signals from other sectors. The signal separation properties lead to better performance in the downlink. The UMTS radio transmission quality is usually limited by the uplink.

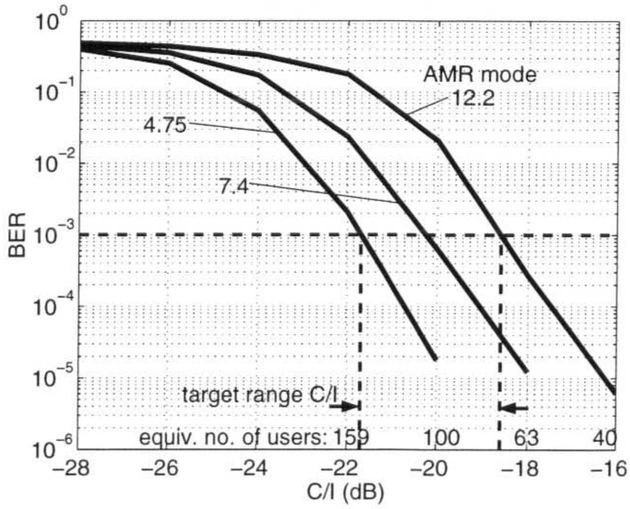

Figure 2: BER vs. channel C/I; bit rates of AMR codec: 4.75, 7.4 and 12.2 kbit/s

Figure 2 shows the residual bit error rate (remaining bit errors in the sequence $\hat{x}$ after channel decoding) for a UMTS uplink speech transmission employing an $r = 1/3$ convolutional coder and soft-input/soft-output Viterbi decoder. The ratio of carrier to interference power C/I is on the horizontal axis.

The bit error rate (BER) decreases as the net data rate R_x of the AMR speech codec decreases (due to a higher possible SF and larger bit energy), as long as the SF is chosen appropriately (i.e. no puncturing is needed). Gross bit rates exceeding 3.84 Mchips/s/SF require puncturing. If too many bits have been punctured, a sufficient quality may not be achieved even for good channels. For the results in figure 2, the highest possible SF (without puncturing) was chosen for each AMR mode: SF=128 for the 4.75 and 7.4 kbit/s modes and SF=64 for $R_x = 12.2$ kbit/s.

For the AMR codec a good speech quality is achieved for residual BERs lower than 10^{-3}, which is equivalent to a channel quality in terms of the C/I ratio of $-21.7 \ldots -18.6$ dB, depending on the AMR codec mode. If each user signal is received with the same power at the base station (ideal power control), C/I can be directly mapped to a number of active users N, included in figure 2, with

$$C/I\,(\mathrm{dB}) = 10 \log 1/(N-1) \qquad (3)$$

The maximum number of simultaneous uplink speech channels for the selected codec modes and the corresponding total data rates per cell are shown in table 3.

AMR mode	R_x	N	$N \cdot R_x$
0	4.75 kbit/s	149	708 kbit/s
4	7.4 kbit/s	106	784 kbit/s
7	12.2 kbit/s	73	891 kbit/s

Table 3: Number of uplink channels and total rates

3. SPEECH QUALITY IMPROVEMENT BY AMR MODE SWITCHING

Under the influence of increasing interference, the quality of speech transmissions employing the higher AMR modes (e.g. 12.2 kbits/s) will fall below an acceptable level because of rising BER. Therefore, the AMR codec mode should be changed towards a lower net rate R_x and better error correcting capabilities if the interference, i.e. the traffic, is increasing. If, on the other hand, interference decreases, higher codec modes can be chosen to obtain a better speech quality.

For the assignment of suitable AMR codec modes, an algorithm for estimation of BER at the receiver was developed. The target BER range for acceptable speech quality was set to $3 \cdot 10^{-4} \ldots 3 \cdot 10^{-3}$ for all codec modes. Although there are control channels for which transmitted bits are known at the receiver, it is not feasible to measure the BER directly using control channels because SFs and transmission powers of data and control channels may differ. Additionally, for sufficient statistics, the time consumed by the bit error counting procedure would only allow a very slow adaptation of AMR modes.

The soft output likelihood values of the Viterbi decoder were identified as a measure that shows good correlation with residual BER and is available at the receiver. Furthermore, the mapping of likelihood values to BER proved to be independent of codec mode and SF. Path metrics for the encoded bits were calculated using the Euclidean distances between received and most likely transmitted sequences. From these metrics, likelihood figures corresponding to the log-likelihood values [6] for the decoded bits are generated.

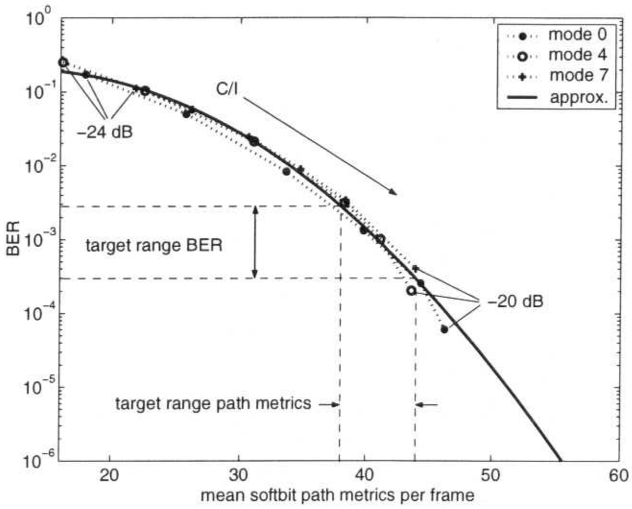

Figure 3: Mean absolute likelihood values vs. BER

Figure 3 shows the correlation of frame-wise calculated likelihood figure means (horizontal axis) and residual BER. One frame consists of the information bits representing 20 ms of speech. For each of the AMR codec modes 0, 4 and 7 the measurement points are connected by a dotted line. An approximation of these curves was obtained by a least-square-error fit of the simulation points to an exponential second-order polynomial (solid line):

$$\log_{10}(\mathrm{BER}) \approx -0.0029\, m^2 + 0.0728\, m - 1.1494 \qquad (4)$$

with m representing the likelihood figure means.

The target BER range and mean likelihood range are represented by the dashed lines.

The operation of the AMR switching algorithm is shown in figure 4 for the example of a C/I ratio decreasing with time. A decreasing C/I could be caused by increasing interference I at a constant received power C (perfect power control).

The channel C/I example is shown in figure 4a. The degradation imposed on the signal resulted in frame-wise likelihood means depicted in figure 4b. The limits for optimal operation, mapped from the desired BER range of figure 3, are drawn as dashed lines. Each time a frame-wise

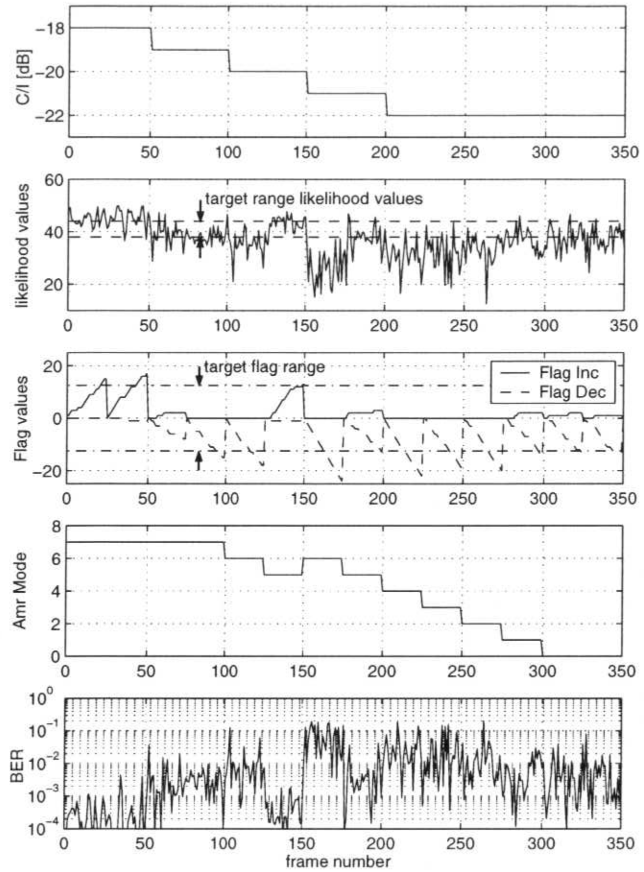

Figure 4: AMR mode control operation
a) channel C/I
b) frame-wise likelihood means
c) mode switching flags
d) chosen AMR mode
e) estimated frame BER

likelihood mean is above the upper limit, a flag for the increase of the AMR mode is set. Conversely, a decrease flag is set if likelihood values are below the lower limit. Every 25 frames (equivalent to 0.5 s for 20 ms speech frames), the AMR mode is changed if more than 50% of the 25 previous frame flags have been set. Figure 4c shows the cumulative sums of increase and decrease flags, reset every 25 frames. The AMR mode (figure 4d) changes if the cumulative sum exceeds an absolute value of 12.5. The estimated residual frame BER is shown in figure 4e.

In this example, a sufficient speech quality is maintained by selecting lower AMR modes as the channel quality degrades. An estimated MOS value of 2.69 was found for a speech sample transmitted over the UMTS model employing the AMR mode switching scheme and the above C/I example. This compares to an unacceptable 1.48 for the same C/I example and transmission at a constant AMR mode of 7. The MOS values for the received speech samples were estimated using the PSQM algorithm [4].

The described method for switching the AMR mode by integrating and dumping over 25 flag values is a simple and robust approach with low complexity. Other averaging methods, such as the calculation of a running average over the likelihood values, did not result in further speech quality enhancements.

In general, the improvements in speech quality obtained by this switching algorithm are substantial especially for adverse time-varying channel conditions. The presented AMR mode switching procedure can be realized as a Radio Resource Management function which interacts with other functions like power control and SF assignment to ensure reliable transmissions for a dynamic network load.

4. CONCLUSIONS

Simulation results have shown that the maximum number of speech channels and the transmission quality in the UMTS uplink depends on the AMR codec mode. Interference by other users was modelled as AWGN. It is a subject of current studies to take the multipath/RAKE channel into consideration.

Under adverse channel conditions, the proposed AMR mode switching procedure helps to maintain a target level of speech quality. It makes use of likelihood figures from the output of the channel decoder as a good estimate of current transmission quality.

UMTS offers a higher service flexibility and elaborate RRM mechanisms like the assignment of data rates, transmit powers and coding schemes. Significant improvements to the performance of UMTS are expected from the implementation of system enhancements like adaptive RAKE signal combination, macro diversity and Multi-User Detection.

5. ACKNOWLEDGEMENTS

The authors would like to thank Alok Gauba, Jan Fuchs, Jörn Breier, and Markus Zensen for their supporting investigations for this contribution.

REFERENCES

[1] 3rd Generation Partnership Project (3GPP) TS 25.201, *Universal Mobile Telecommunications System (UMTS); Physical layer - General Description*, Ver.3.1.0, Rel.99.

[2] Synopsys, Inc.: *COSSAP Documentation*, Ver.3.5a.

[3] 3rd Generation Partnership Project (3GPP) TS 26.071, *Universal Mobile Telecommunications System (UMTS); AMR Speech Codec; General Description*, Ver.3.0.1, Rel.99.

[4] ITU-T Rec. P.861, *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs*, 1998.

[5] 3rd Generation Partnership Project (3GPP) TS 25.213, *Universal Mobile Telecommunications System (UMTS); Spreading and modulation (FDD)*, Ver.3.2.0, Rel.99.

[6] Hagenauer, J., Höher, P.: *A Viterbi algorithm with soft-decision outputs and its applications*, Proc. Globecom 1989, Dallas, USA, 1989, pp. 1680-1686