Softbit Speech Decoding With Re-Quantization

S. Heinen, M. Adrat, P. Vary, *W. Xu
Institute of Communication Systems and Data Processing
Aachen University of Technology
e-mail: {heinen,adrat,vary}@ind.rwth-aachen.de
WWW: http://www.ind.rwth-aachen.de
*Dept. of Mobile Phone Development, Siemens AG
Hofmannstr. 51, 81359 Munich
wen.xu@mch.siemens.de

Abstract — **In digital mobile communication usually speech decoding algorithms are closely combined with error concealment techniques to reduce the subjective effects of residual bit errors. In general these error concealment techniques require reliability information which is provided by the channel decoder. The amount of additional information may vary from 1 bit per received frame, e.g. in GSM a BFI (bad frame indicator) for a speech frame of 260 bits (Full Rate (FR) codec) or for a frame of 244 bits (Enhanced Full Rate (EFR) codec), up to several bits per received bit for *Softbit Speech Decoding*. In applications such as the GSM infrastructure where the channel decoder and the speech decoder are placed at spatially distant locations, the additional data rate required to transmit the reliability information might not be available.**

Therefore, we break up the close connection between error concealment and speech decoding and move the error concealment unit to the base transceiver station where the channel decoder is located. As the error concealment algorithm produces estimated speech parameters with high amplitude resolution a re-quantization of these estimated speech parameters is required which implies a loss in estimation quality. In this paper it is shown that this loss in quality is negligible in most practical cases. In addition, methods for a reduction of the re-quantization loss are discussed.

I. Introduction

Speech coding is indispensable to achieve a required bandwidth efficiency in applications where bandwidth is a limited resource, as e.g. in digital mobile telecommunication. In contrast to fixed network transmission, the mobile radio channel suffers from a variety of adverse effects such as multi-path propagation and Doppler spread that make it very hard to guarantee data transmission at low error rates. This can only be achieved by a combination of interleaving, equalization and channel coding.

Especially at high compression rates (e.g. 0.5-1.5 bits per speech sample), which are achievable by modern CELP codecs, the compressed speech data becomes extremely vulnerable. Therefore, in current mobile systems powerful *Forward Error Correction* (FEC) schemes, usually based on convolutional codes, are applied for protection. In order to adapt such channel codes to the source codec, *Unequal Error Protection* (UEP) [1] is used, i.e. very significant bits of a coded speech segment are highly protected, whereas less significant bits have weaker error protection.

However, under severely disturbed transmission conditions the described measures of error protection will partially fail. In this case, residual bit errors remain in the bit stream delivered to the source decoder, which tend to cause very annoying artifacts in the synthesized speech unless additional error concealment techniques are applied. A rather simple, but effective example of error concealment is the *Bad Frame Indicator* (BFI) mechanism of GSM. A *Cyclic Redundancy Check* (CRC) over the 50 most sensitive bits of a coded speech frame is used in addition to reliability information derived from the equalizer or channel decoder, to classify the received frames as either “good” or “bad”. This coarse reliability information is forwarded to the speech decoder, which reacts on a bad frame either by repeating the last speech frame classified as good, or by muting if too many subsequent bad frames are received.

This ad hoc approach can be improved by forwarding reliability information not only for a complete frame but for speech parameters or even single bits and by exploiting this information for parameter estimation. This *Softbit Speech Decoding* approach [2, 3, 4] is furthermore able to exploit residual redundancy in the source coded data which is due to the instationarity of the speech signal.

Considering the practical implementation of *Softbit Speech Decoding*, the extensive use of reliability information at the speech decoder will require a multiple of the transmission bandwidth on the link between the channel decoder and the speech decoder compared to a BFI solution. While this is easily solved in a mobile phone where both components are probably located on the same chip or in the same DSP program, a problem will occur in the network, since channel coding is implemented in the *Base Transceiver Station* (BTS) and speech coding in the *Transcoder/Rate Adapter Unit* (TRAU) located at the *Base Station Controller* (BSC) or at the *Mobile Switching Center* (MSC), or in between. So channel decoding and speech decoding take place at a distance of typically several kilometers, which have to be linked e.g. by leased lines.

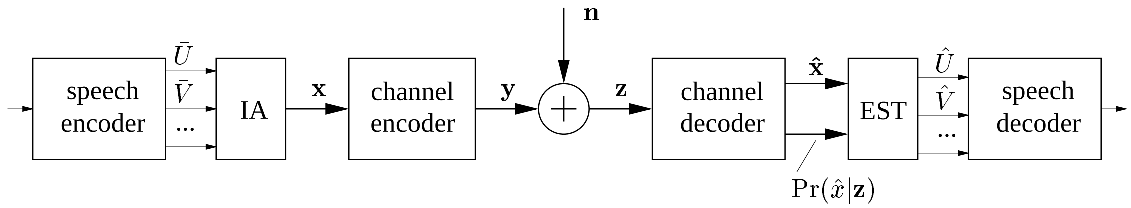

Fig. 1: Transmission system
(IA: index assignment, EST: estimation)

As specified in the current GSM standard, the data transport between BTS and TRAU is constrained to be maximally 16 kbit/s (by using the so-called 16 kbit/s A_{bis} sub-multiplexing). Since almost all of the 16 kbit/s data rate has been occupied by the signaling bits and the speech (hard) bits (e.g. 13 kbit/s in the case of the Full Rate Codec), the transportation of reliability information (soft bits) from the channel decoder to the source decoder becomes impossible.

To overcome this problem we propose a new system configuration which associates error concealment by *Softbit Speech Decoding* directly with the channel decoder. This requires that the parameters estimated by the error concealment are adequately re-quantized before transmitted to the speech decoder. When carefully designed, this approach can yield almost no degradation in terms of estimation gain compared to the solution without re-quantization.

In section II we give a quick review of *Softbit Speech Decoding*. Section III describes the implementation aspects in detail. As we will see, the new system configuration requires a re-quantization of the estimated speech parameters. Section IV will focus the impacts on the estimation quality caused by this re-quantization and will provide design rules for quantizers adequate for re-quantization. Further enhancement of the re-quantization quality is considered in section V. Finally, in section VI we present simulation results that verify our considerations.

II. SOFTBIT SPEECH DECODING

Recently parameter estimation was proposed to realize *Softbit Speech Decoding* [2, 3]. This approach becomes especially attractive when used together with efficient soft output decoders for convolutional codes [5, 6, 7].

Figure 1 depicts a model of the considered transmission system. In the *index assignment* (IA) block to each of the quantized source coding parameters $\bar{U}, \bar{V}, ..$ (either scalar or vectors), a value-specific bit pattern is assigned, e.g. parameter $\bar{U}$ is encoded by a word of M bits $u_1, u_2, .., u_M$. The bits $u_1, .., u_M, v_1, ..$ are compiled to a block $\mathbf{x}$ and are convolutionally encoded. The coded bits $\mathbf{y}$ are transmitted over the channel, which is described by the additive noise $\mathbf{n}$. The channel decoder processes the received values $\mathbf{z}$ and yields two outputs: Hard decision bits $\hat{\mathbf{x}}$ and reliability information in terms of *a posteriori* probabilities $\Pr(\hat{x}|\mathbf{z})$ for single bits. This information enables the estimation unit to compute parameter estimates $\hat{U}, \hat{V}, \ldots$

Parameter estimation requires the computation of **parameter** *a posteriori* probabilities, which can be achieved by combining the corresponding **bit** *a posteriori* probabilities produced by the channel decoder. Under the assumption of statistical independence of the M distinct error processes[1] that influence the parameter bits u_i we get the *a posteriori* probability of a specific quantizer reproduction level $\bar{U}_j$

$$\Pr(\bar{U}_j|\mathbf{z}) \approx C \cdot p_{\bar{U}}(\bar{U}_j) \cdot \prod_{i=1}^{M} \frac{\Pr(u_i|\mathbf{z})}{p_{u_i}(u_i)} \quad , \qquad (1)$$

where $p_{\bar{U}}(\cdot)$ is the *a priori* distribution of the quantized speech parameter and $p_{u_i}(\cdot)$ the *a priori* distribution of the i-th bit of its binary representation. The constant C is determined by the condition $\sum_j \Pr(\bar{U}_j|\mathbf{z}) = 1$. The computation of the *a posteriori* probabilities can also take into consideration the time correlation of parameter values of successive speech segments. In this case the process $\bar{U}$ must be modeled by a Markov chain rather than by a pdf [2]. A further enhancement of *Softbit Speech Decoding* can be achieved by application of *Source Optimized Channel Codes* (SOCCs) [8, 9]

One possible decision rule can be derived from *Maximum A Posteriori* (MAP) criterion, which minimizes the symbol probability of error:

$$\hat{U}_{MAP} = \arg\max_j \Pr(\bar{U}_j|\mathbf{z}) \quad . \qquad (2)$$

From estimation theory it is well known that the optimal estimator in the *Minimum Mean Square Error* (MMSE) sense is given by

$$\hat{U}_{\text{MMSE}} = \mathrm{E}\left\{\bar{U}|\mathbf{z}\right\} = \sum_j \bar{U}_j \cdot \Pr(\bar{U}_j|\mathbf{z}) \quad . \qquad (3)$$

It has to be noted, that due to equation (3) $\hat{U}_{\text{MMSE}}$ can take values other than quantization table or codebook entries. As the speech decoder usually is able to process these real-valued estimates directly, no further quantization is required. In the next section however, we will consider a practical implementation that makes such a re-quantization necessary.

For the following derivations, two limits with respect to the channel quality are of importance. If the transmission channel is error-free, the *a posteriori* probabilities

[1] This is the case if the bits of one parameter have a mutual distance of at least $5L$ bit positions in the block $\mathbf{x}$, where L is the constraint length of the convolutional code. This can be achieved, if the compilation of x includes an adequate interleaving.

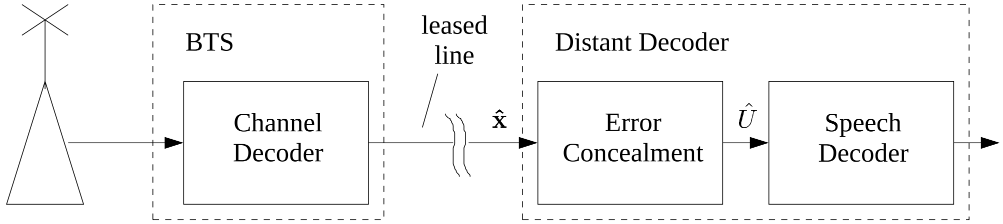

Fig. 2: Conventional receiver site of transmission system

take the values

$$\Pr(\bar{U}_j|\mathbf{z}) = \begin{cases} 1 & \bar{U}_j \quad \text{sent} \\ 0 & \bar{U}_k,\ k \neq j \quad \text{sent} \end{cases} \quad . \tag{4}$$

Hence, $\hat{U}_{\text{MMSE}}$ is equal to the originally quantized parameter value $\bar{U}$. In turn, for very bad conditions we get

$$\Pr(\bar{U}|\mathbf{z}) = p_{\bar{U}}(\bar{U}) \quad , \tag{5}$$

i.e. the *a posteriori* probabilities are equal to their corresponding *a priori* probabilities.

III. Implementation Aspects

We will now consider the problems implied by a practical implementation of *Softbit Speech Decoding* in a cellular network. A typical receiver structure in the uplink is depicted in figure 2.

Channel decoding is already performed in the BTS in order to reduce the transmission bandwidth necessary to forward the received bit stream to the distant speech decoder, which is usually placed at some network switching or inter-working unit or in between.

To implement error concealment by *Softbit Speech Decoding* in the way depicted in Figure 2 it would be necessary to transmit reliability information in addition to the channel decoded (hard) bit stream. This consumes a large amount of bandwidth which in some cases is not available. E.g. in the GSM networks, the maximum data rate between BTS and TRAU is usually constrained to be 16 kbit/s and in the case of the Half Rate transmission often even 8 kbit/s, by use of the 16 and 8 kbit/s A_{bis} sub-multiplexing. On the other hand, when each of the channel decoded (speech parameter) soft bits is represented/quantized by 3 bits, then for a 13 kbit/s speech codec (as in the case of the ETSI FR codec) a data of 13×3 = 39 kbit/s had to be transmitted. As a result, error concealment by *Softbit Speech Decoding* in the conventional way of associating it to the speech decoder is not implementable in the current GSM infrastructure.

Therefore, we propose an alternative solution if the channel decoder and the speech decoder are at spatially distant locations, namely, the estimator is associated with the channel decoder. Figure 3 depicts such a modified system configuration.

As already mentioned, this requires a re-quantization of the estimated speech parameters. On one hand, the accuracy of this re-quantization can be made arbitrarily high to avoid a loss of quality, but on the other hand, the transmission bandwidth has to be considered.

IV. Influence of Re-Quantization

The performance degradation caused by re-quantization of the estimated parameter values will now be discussed. Figure 4 shows an equivalent transmission system for the codec parameter U. The total quantization caused by the source encoder and the re-quantization are modeled as additive noise n_q and n_r, respectively. Furthermore, the equivalent channel consisting of channel coder, physical channel, equalizer, channel decoder and parameter estimation is also modeled by additive noise n_c.

It is obvious that the noise power caused at the encoder site $N_q = \text{E}\{n_q^2\}$ solely depends on the speech codec used and is independent of the channel conditions. The noise power $N_c = \text{E}\{n_c^2\}$ of the equivalent channel is mainly determined by the carrier-to-interferer-ratio (C/I) of the physical channel. If the quantizer used for re-quantization has the same reproduction levels as the transmitter site quantizer, the noise power $N_r = \text{E}\{n_r^2\}$ caused by re-quantization will depend on both, the performance of the quantizer and the quality of the physical channel, as shown below:

A. Noise-free transmission ($C/I \to \infty$)

In case of a noise-free transmission ($N_c = 0$) the output values of the estimator $\hat{U}$ are identical to the quantized values $\bar{U}$. Then, using a quantizer for re-

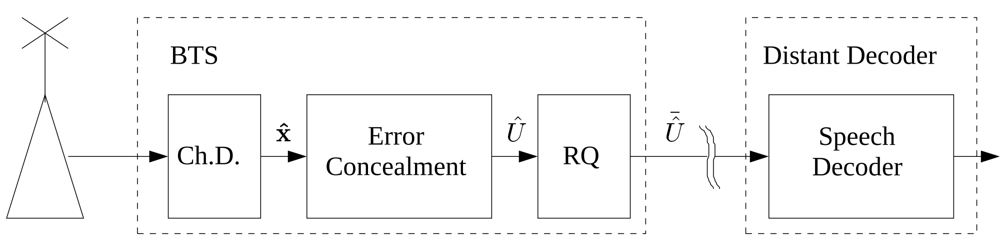

Fig. 3: Transmission system with re-quantization
(Ch.D.: channel decoder, RQ: re-quantization)

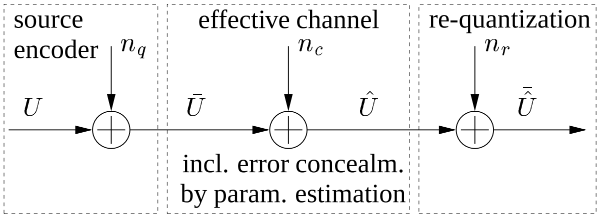

Fig. 4: Substitute model for a speech codec parameter transmission

quantization that contains all encoder site reproduction levels the estimated parameter $\hat{U}$ can be reproduced exactly by re-quantization, i.e. $\hat{\bar{U}} = \hat{U}$. Therefore, it can be stated that

$$N_r|_{C/I\to\infty} = 0 \quad . \tag{6}$$

B. Transmission over a disturbed channel

If the parameter $\bar{U}$ is transmitted over a disturbed channel, because of (5), the output values of the MS estimator $\hat{U}$ are shifted towards the expectation[2] $\mathrm{E}\{\bar{U}\}$ by increasing channel noise. Hence, for very bad channels the estimator outputs a constant value which consequently results in the constant re-quantization error

$$N_r|_{C/I\to-\infty} = [\mathrm{E}\{\bar{U}\} - \bar{U}_{opt}]^2 \quad . \tag{7}$$

In equation (7) $\bar{U}_{opt}$ represents the reproduction level used for $\mathrm{E}\{\bar{U}\}$. Furthermore, the noise power of the channel N_c is

$$N_c|_{C/I\to-\infty} = \sigma^2_{\bar{U}} \quad , \tag{8}$$

where $\sigma^2_{\bar{U}}$ denotes the variance of the quantized parameter source.

Figure 5 illustrates a typical behavior of the three different noise components. We performed a three-dimensional vector quantization at rate 2 bit/dimension of a white Gaussian source with variance $\sigma^2_U = 1$. The codebook indices were bit-wise transmitted over an AWGN channel and at the receiver parameter estimation and re-quantization were applied. The E_s/N_0 ratio on the AWGN channel here serves as a simplified counterpart of the C/I on real mobile channels. For the re-quantization we used the same vector quantizer as at the transmitter. It can be seen that the noise due to re-quantization tends to the finite value

$$[\mathrm{E}\{\bar{U}\} - \bar{U}_{opt}]^2 \quad . \tag{9}$$

For all considered channel conditions the distortion due to channel noise is about one magnitude greater than the re-quantization noise. N_{tot} reveals the total end-to-end distortion.

[2] If a Lloyd-Max quantizer is used, it also identical to the expectation value $\mathrm{E}\{U\}$ of the unquantized parameter U.

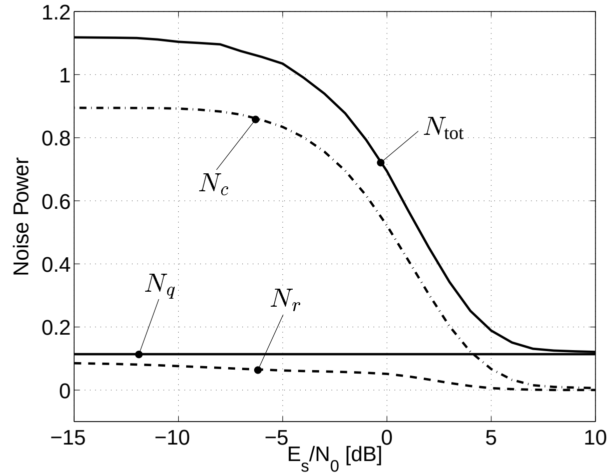

Fig. 5: N_q: quantization noise, N_c: channel noise, N_r: re-quantization noise, N_{tot}: end-to-end noise

C. Parameter Signal-To-Noise Ratio

If $\mathrm{E}\{\bar{U}\} = 0$ and if we assume statistical independence of the three error processes depicted in figure 4, the overall signal-to-noise ratio (SNR) is given by

$$\mathrm{SNR} = \frac{\sigma^2_{\bar{U}}}{N_q + N_c(C/I) + N_r(C/I)} \quad . \tag{10}$$

Using the bounds derived for a noise-free channel (6) yields

$$\mathrm{SNR}|_{C/I\to\infty} = \frac{\sigma^2_{\bar{U}}}{N_q} \quad . \tag{11}$$

For moderate distorted channels the SNR is dominated by $N_c + N_q \gg N_r$, hence

$$\mathrm{SNR} \approx \frac{\sigma^2_{\bar{U}}}{N_q + N_c} \quad . \tag{12}$$

whereas by inserting equations (7) and (8) we get for a very bad channel

$$\mathrm{SNR}|_{C/I\to-\infty} = \frac{\sigma^2_{\bar{U}}}{N_q + \sigma^2_{\bar{U}} + N_r|_{C/I\to-\infty}} \approx 1 \quad , \tag{13}$$

because of $\sigma^2_{\bar{U}} \gg N_q + N_r|_{C/I\to-\infty}$.

These considerations show that the parameter SNR for the overall transmission system is essentially independent of the noise power N_r caused by re-quantization. Only under moderate channel conditions the quantizer must fulfill the requirement

$$N_q \ll N_c \quad . \tag{14}$$

As we will show by simulations, this is fulfilled for Gaussian distributed sources by using quantizers at $R > 1$ bit/sample.

V. Enhancement of Re-Quantization Quality

If low rate quantizers at $R < 1$ bit/sample are used for encoder site quantization as well as for re-quantization, condition (14) is violated for moderately disturbed channels, and a significant loss in quality due to re-quantization would be the consequence. To overcome this problem, the accuracy of re-quantization must be enhanced by employing a quantizer with higher resolution.

The crucial point for the design of such a quantizer that all encoder site reproduction levels must be included in the re-quantizer codebook. Otherwise equation (6) would be invalid, and even under noise-free conditions a performance degradation would be the case. A training algorithm for such re-quantizer codebooks can easily be derived from the well-known LBG algorithm [10].

Further enhancement can be achieved for bad channels. Due to (9) the re-quantization noise N_r in general tends to a non-zero value when going to bad channel conditions. This problem can easily be solved by adding $\mathrm{E}\{\bar{U}\}$ as a further reproduction vector to the re-quantization codebook. Repeating the simulation described in Section IV.B and applying such a modified re-quantization codebook yielded the curves shown in Figure 6. Now N_r has a local maximum and actually vanishes for $E_s/N_0 \to -\infty$.

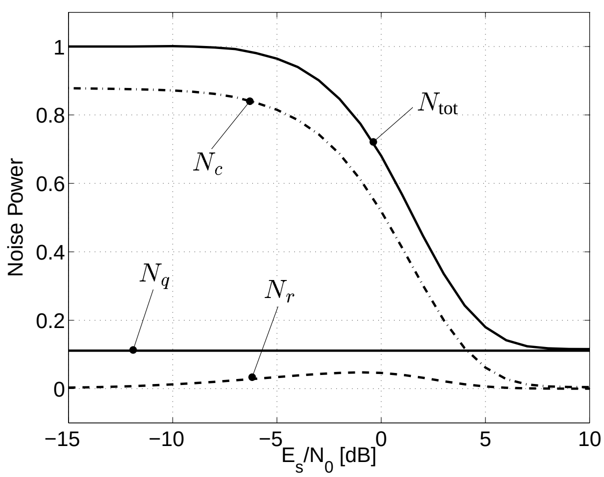

Fig. 6: Distortions using an extended re-quantization codebook

Of course, a higher re-quantization accuracy implies also an increased bandwidth demand on the link between the error concealment unit at the base station and the distant speech decoder (see Figure 3). But as we will see in the next section, 1 or 2 additional bits per speech parameter are in most cases sufficient, and as only few speech parameters are quantized at very low rates, the resulting bandwidth increase should be negligible.

One example how to merge the described enhanced re-quantization technique with other system aspects might be adaptive multi-rate speech coding (such as GSM-AMR), where several single rate codecs together build one codec family. For each family member, the speech codec contains a specific quantizers, one for each coding rate. In some cases these quantizers constitute a hierarchy such, that on one hierarchy level all reproduction centroids of the previous level are included. Then high hierarchy level quantizers could be used for enhanced re-quantization, without explicit modifications of the speech decoder.

VI. Simulations

The results of chapters IV and V shall now be verified by simulations. To model the disturbed transmission of speech parameters, we will transmit a Gaussian distributed scalar parameter U with $\sigma_U^2 = 1$ over an AWGN channel. At the decoder the parameter values are MS estimated and re-quantized.

Figure 7 shows the simulation result using a three dimensional vector quantizer (VQ) with 2 bit/sample for quantization. The same quantizer was employed for re-quantization. The loss in speech quality compared to unquantized *Softbit Speech Decoding* is at most 0.3 dB, which is tolerable in most applications.

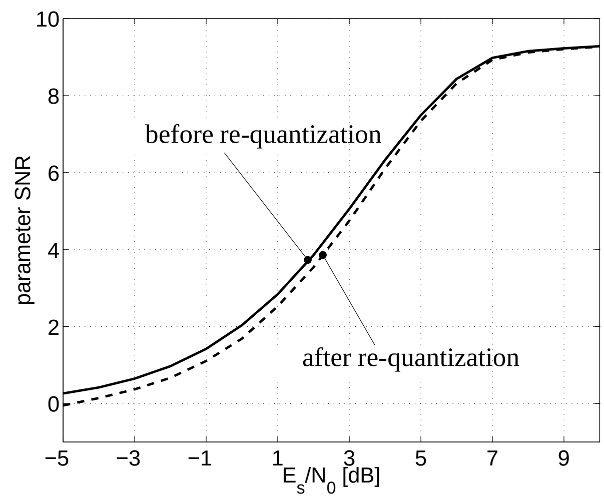

Fig. 7: Parameter-SNR for 3-dimensional VQ at 2 bit/sample before and after re-quantization

In contrast, figure 8 shows the simulation results for a three dimensional vector quantizer with only 1 bit/sample. Due to the low rate re-quantization the parameter SNR decreases about 1 dB under bad channel conditions. If a higher rate re-quantizer with e.g. 2 bit/sample is used, the loss in re-quantization quality can be considerably reduced.

VII. Conclusions

In conventional mobile communication, speech decoding error concealment techniques are spatially and algorithmically closely associated to the speech decoder. In our contribution we show, that this close association can be avoided, and that it is possible to make the error

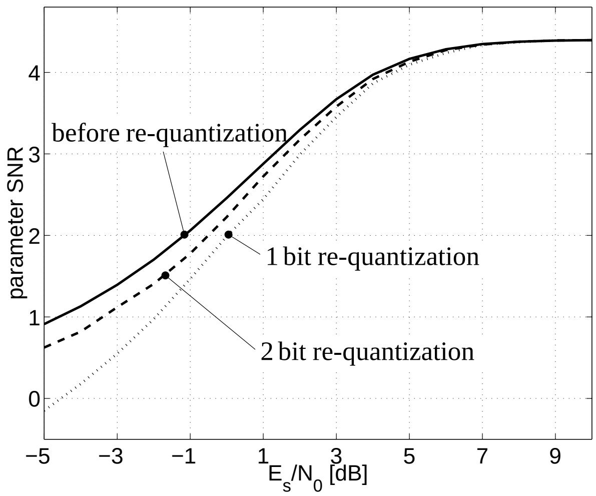

Fig. 8: Parameter-SNR for 3-dimensional VQ at 1 bit/sample before and after re-quantization at 1-2 bit/sample

concealment unit a part of the channel decoder, which plays an important role for the economic implementation of error concealment by *Softbit Speech Decoding* in a cellular network.

To do so, a re-quantization of estimated speech parameters is necessary. By analytical considerations, we derived approximations for the quality degradation due to this re-quantization compared to unquantized *Softbit Speech Decoding*. It turned out, that in most cases the re-quantization loss is negligible small. For cases where this is not valid, we proposed a method to enhance the re-quantization quality.

VIII. Acknowledgement

The authors would like to thank the team of Prof. J. Hagenauer (TU Munich) for many fruitful discussions.

IX. References

[1] Joachim Hagenauer, “Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications,” *IEEE Trans. Communications*, vol. 36, pp. 389–400, Apr. 1988.

[2] T. Fingscheidt and P. Vary, “Robust Speech Decoding: A Universal Approach to Bit Error Concealment.” In Proceedings of ICASSP’97, Vol. 3, pp. 1667-1670, Munich, April 1997.

[3] C. Gerlach, “A Probabilistic Framework For Optimum Speech Extrapolation in Digital Mobile Radio.” In Proceedings of ICASSP’93, pp. II.419-II.422, Minneapolis, April 1993.

[4] S. Heinen, M. Adrat, O. Steil, P. Vary, and W. Xu, “A 6.1 to 13.3-kb/s Variable Rate CELP Codec (VR-CELP) for AMR Speech Coding,” in *Proceedings ICASSP*, vol. 1, (Phoenix, Arizona), pp. 9–12, Mar. 1999.

[5] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate.” In IEEE Transactions on Information Theory, pp. 20:284-287, March 1974.

[6] J. Hagenauer and P. Höher, “A Viterbi Algorithm with Soft-Decision Outputs and its Applications.” In Proceedings of GlobeCom, pp. 1680-1686, November 1989.

[7] S. Heinen, A. Geiler, and P. Vary, “MAP Channel Decoding by Exploiting Multilevel Source A Priori Knowledge,” in *Proceedings European Personal Mobile Communications Conference (EPMCC)*, (Bonn), pp. 467–473, Oct. 1997.

[8] S. Heinen, S. Bleck, and P. Vary, “Robust Speech Transmission Over Noisy Channels Employing Non-Linear Block Codes,” in *Proceedings Speech Coding Workshop 1999*, (Porvoo, Finland), IEEE, June 1999.

[9] S. Heinen, T. Hindelang, and P. Vary, “Channel Codes for Softbit Source Decoding: Estimation of Correlated Parameters,” in *3rd ITG Conference Source and Channel Coding*, (Munich), ITG, Jan. 2000.

[10] Y. Linde, A. Buzo, and R. M. Gray, “An Algorithm for Vector Quantizer Design,” *IEEE Trans. Communications*, vol. 28, pp. 84–95, Jan. 1980.