Softbit Speech Decoding: A New Approach to Error Concealment

Tim Fingscheidt and Peter Vary

Abstract*—In digital speech communication over noisy channels there is the need for reducing the subjective effects of residual bit errors which have not been eliminated by channel decoding. This task is usually called *error concealment.

We describe a new and generalizing approach to error concealment as part of a modified robust speech decoder. It can be applied to any speech codec standard and preserves bit exactness in case of an error free channel. The proposed method requires *bit reliability information* provided by the demodulator or by the equalizer or specifically by the channel decoder and can exploit additionally *a priori knowledge about codec parameters*. We apply our algorithms to PCM, ADPCM, and GSM full-rate speech coding using AWGN, fading, and GSM channel models, respectively. It turns out that the speech quality is significantly enhanced, showing the desired inherent muting mechanism or graceful degradation behavior in case of extreme adverse transmission conditions.

***Index Terms*—Error concealment, frame repetition, muting, speech decoding.**

I. Introduction

IN digital mobile communication systems the issue of speech quality is important. The best achievable speech quality is first of all determined by the speech coding algorithm. In mobile radio systems channel coding is applied to preserve the quality level over a wide range of channel characteristics.

Nevertheless, even with channel coding residual bit errors occur that may lead to severe degradation of speech quality. These annoying effects can be reduced or even be eliminated by error concealment. In cellular system standards such as GSM [1]–[3] error concealment algorithms are proposed as nonmandatory recommendations. This allows manufacturers to implement proprietary solutions to improve existing systems.

To reduce the impact of known errors (such as frame losses) on the reconstructed waveform, it is possible to apply waveform substitution techniques directly on the waveform [4]–[9], [65]. However, if a more sophisticated speech coder is used, it is usually advantageous to use the implied parametrization of the speech signal and conceal the errors at the level of the *bits* or codec *parameters*. Accordingly, the GSM standards on substitution and muting of lost frames [1]–[3] propose simple mechanisms such as parameter repetition or parameter extrapolation. They are mainly driven by a binary *bad frame indicator* (BFI) that marks the current received frame as good or bad. This BFI can be seen as a very coarse reliability information that may initiate the substitution of a complete frame even if only a few bits have been disturbed, or, on the other hand, that may declare a frame as reliable although some bits are incorrect. There are some proposals to enhance the reliability information [10]–[13]. Alternatively, BFI may not be generated explicitly. Minde *et al.* [14] compute weighting factors to perform parameter extrapolation by weighted summation over previous frames.

There are several approaches to joint source/channel coding employing *channel optimized vector quantization* (COVQ) and soft decision decoding based on the *minimum mean squared* error criterion (MS). The aim is the quantizer/encoder design for a specific channel condition [15]–[19]. However, this approach is in general not compatible with existing standards.

In this paper, a new concept of error concealment is proposed that can be applied to any speech coding algorithm. It consists of a modification of the speech *decoder* such that (real-valued) *softbits* instead of (binary) hardbits are used. A softbit can be interpreted as joint knowledge of a hardbit and its reliability, i.e., the estimated bit error probability. In contrast to BFI-driven error concealment techniques, the softbit speech decoding concept is able to process *bit-individual nonbinary* reliability information, which can be provided e.g., by a soft-output channel decoder [20], [21].

Due to practical reasons in most cases residual redundancy can be observed within the speech codec parameters. As already mentioned by Shannon this source coding sub-optimality should be exploited at the receiver "... to combat noise" [22]. We exploit the residual redundancy in terms of parameter *a priori knowledge* that is subject to algorithms based on the Bayesian framework computing *a posteriori probabilities* for individual codec parameters. The *a priori* knowledge as we use it refers to a Markov model of the speech codec parameter. It should be noted that in the literature, hidden Markov models (HMMs) are also proposed to describe the statistical behavior of the source (e.g., [23], [24]).

Finally, the speech codec parameters are estimated according to appropriate error criteria that reflect the perceived quality. Gerlach [25], [26] and Feldes [27], [28] proposed MS estimation, however still based on simplifying assumptions about the *a posteriori* probabilities. MS estimation of codec parameters gives a significant quality advantage in comparison to the alternative approaches of joint source and channel decoding exploiting source *a priori* knowledge (see e.g., [29]–[33]) as these techniques minimize bit, symbol, or path error probability.

Manuscript received February 23, 1999; revised April 18, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Peter Kroon.

T. Fingscheidt is with Siemens AG, ICM CD Mobile Phones, Munich, Germany.

P. Vary is with the Institute of Communications Systems and Data Processing, Aachen University of Technology, Aachen D-52056, Germany.

Publisher Item Identifier S 1063-6676(01)01325-6.

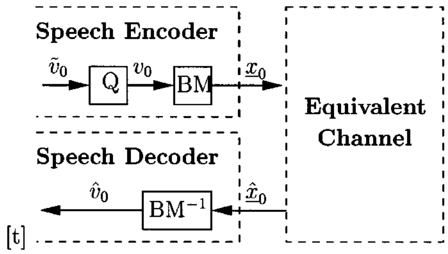

Fig. 1. Conventional hardbit speech decoding.

In the case of an error free channel situation the new softbit decoding approach preserves bit exactness as required for type approval.

II. Concept of Softbit Speech Decoding

In general, a speech encoder produces *codec parameters*, that are quantized and mapped to bit combinations. These parameters are e.g., spectral coefficients, pitch periods, gain factors, codebook entries, etc., or even quantized PCM or ADPCM samples [34], [35].

Let us consider the transmission of a specific codec parameter $\tilde{v}_n \in \mathbb{R}$ over a noisy channel as described by Fig. 1. The time index n denotes the relative discrete time instant; this can be e.g., a frame or a subframe instant. The time index $n = 0$ shall specify the present time index, accordingly $n = -1, -2, \cdots$ are time indices of previous frames or subframes. For simplicity, we restrict ourselves here to scalar quantization. In Section III-E the concept is extended to vector quantization.

The codec parameter $\tilde{v}_0$ at the present time instant $n = 0$ is quantized according to $\mathbf{Q}[\tilde{v}_0] = v_0$ with $v_0 \in$ QT (QT: quantization table). It can be represented by a quantization table index i with $i \in \{0, 1, \cdots, 2^M - 1\}$. Via bit mapping (BM) a bit combination

$$\underline{x}_0 = (x_0(0), x_0(1), \cdots x_0(m), \cdots x_0(M-1)) \quad (1)$$

consisting of M bits is assigned to each quantized parameter v_0 (or quantization table index i). The bits are assumed to be bipolar, i.e., $x)(m) \in \{-1, +1\}$. For simplicity we write $v_0 = v_0^{(i)}$ as well as $\underline{x}_0 = \underline{x}_0^{(i)}$. The transmission of this bit combination is described by the so-called *equivalent channel* which might consist of any combination of the noisy analog channel with channel (de)coding, (de)modulation, and equalization. Due to the channel noise the received bit combination $\underline{\hat{x}}_0$ is possibly not identical to the transmitted one. In a conventional decoding scheme as depicted in Fig. 1 the received bit combination $\underline{\hat{x}}_0$ is applied to table decoding (*inverse bit mapping* scheme (BM^{-1})). Thereafter, the decoded parameter $\hat{v}_0$ is used within the specific speech decoder algorithm to reconstruct speech samples. We call this conventional solution *hardbit* (speech) decoding (HD).

The proposed new approach to error concealment by softbit speech decoding (SD) is depicted in Fig. 2. It requires additionally reliability information (e.g., [31]) in terms of estimated bit error probabilities

$$\underline{p_e}_0 = (p_{e0}(0), p_{e0}(1), \cdots, p_{e0}(m), \cdots p_{e0}(M-1)) \quad (2)$$

of the received bit combination $\underline{\hat{x}}_0$. There are several possibilities to obtain such estimates. Some of them are discussed in Section III-A. The joint information of a received bit and its estimated instantaneous error probability is what we call a *softbit*

$$\{\hat{x}_0(m), p_{e0}(m)\}.$$

In the literature there are several explicit definitions of the softbit using the same information [17], [31], [36], [37].

The first algorithmic step of softbit speech decoding consists in the computation of 2^M *parameter transition probabilities*, i.e., $\mathrm{P}(\underline{\hat{x}}_0|\underline{x}_0^{(i)}$, $i \in \{0, 1, \cdots, 2^M - 1\}$. They denote the probability of the *known* received bit combination $\underline{\hat{x}}_0$ under the assumption of any *unknown* bit combination $\underline{x}_0^{(i)}$ that might have been transmitted. As shown in Section III-B we explain how to compute these probabilities using $\underline{\hat{x}}_0$ and $\underline{p_e}_0$.

In the second step the parameter transition probabilities $\mathrm{P}(\underline{\hat{x}}_0|\underline{x}_0^{(i)})$ have to be combined with *a priori knowledge* about the codec parameter which can be measured once in advance by processing a representative speech database by the speech encoder. The *a priori* knowledge is stored in the decoder in a ROM table. Different kinds of *a priori* knowledge are discussed in Section III-C. As a result of the second step we obtain 2^M *a posteriori probabilities* $\mathrm{P}(\underline{x}_0^{(i)}|\underline{\hat{x}}_0, \cdots)$ with $i \in \{0, 1, \cdots, 2^M - 1\}$ being the probabilities of each of the possibly transmitted bit combinations $\underline{x}_0^{(i)}$ given the received one $\underline{\hat{x}}_0$ and given possibly additional receiver information marked by "$\cdots$." Several ways of computing *a posteriori* probabilities are shown in Section III-D. Sections IV-A and IV-B give solutions to the *a posteriori* probabilities if additional information is obtainable by a block code, or by diversity reception.

The third and last block of the softbit speech decoding process according to Fig. 2 is the *parameter estimator*. The *a posteriori* probabilities are used to find optimum parameter values $v_0^{(\mathrm{est})}$ taking an appropriate error criterion into consideration. Two widely used estimators are discussed in Section III-E.

Sections IV and Section V give applications to prove the capabilities of the proposed softbit speech decoding technique.

III. Softbit Speech Decoding Algorithm

A. Bit Error Probabilities

In the following it is described by two examples how to provide reliability information in terms of error probability estimates $\underline{p_e}_0$ of the received bits $\underline{\hat{x}}_0$ (see also (2)).

1) Fading Channel with Coherent BPSK: We assume a fading channel with additive white Gaussian noise (constant power spectral density $N_0/2$) and coherent binary phase shift keying (BPSK) demodulation. This binary symmetric channel can be described by Fig. 3(a) with the time varying instantaneous bit error probability p_e or by Fig. 3(b) with the time varying fading factor a and the effective noise samples $\tilde{z}_0(m)$ at the receiver output.

The bit error probability of the hard decided bit $\hat{x}_0(m)$ can be formulated as [31], [38]

$$p_{e0}(m) = \frac{1}{1 + \exp|L_c \cdot \tilde{x}_0(m)|} \quad \text{with} \quad L_c = 4a \cdot \frac{E_b}{N_0} \quad (3)$$

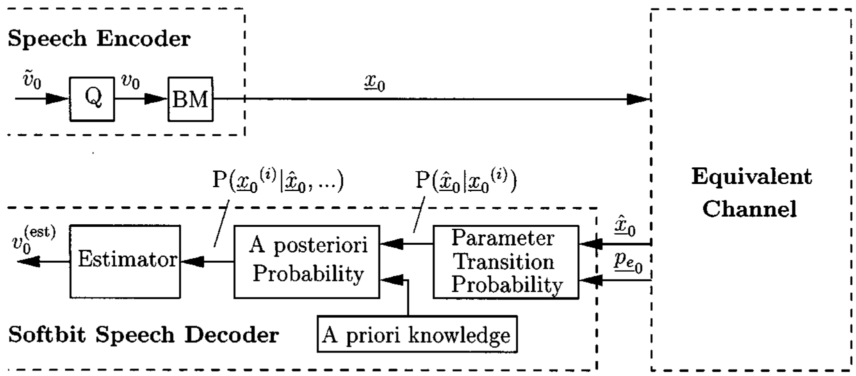

Fig. 2. New softbit speech decoding technique.

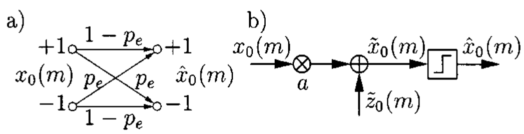

Fig. 3. (a) BSC channel and (b) fading channel with fading factor a and additive Gaussian noise $\tilde{z}_0(m)$ at the receiver output.

and $\tilde{x}_0(m)$ being the *real-valued* received value. The *channel state information* (CSI) L_c comprises the fading factor a as well as the normalized noise variance $N_0/2E_b$ of the additive Gaussian noise. Both values have to be estimated at the receiver. From (3) it is obvious that an individual instantaneous bit error probability can be assigned to each received value $\tilde{x}_0(m)$, even if the CSI is constant.

2) Transmission Scheme with Channel Coding: If the equivalent channel includes channel coding the required reliability information can be provided by a soft-output channel decoder. The best choice is the channel decoder of Bahl *et al.* [21] because it is able to yield the so-called log-likelihood values (L-values)

$$L_0(m) = \ln \frac{\mathrm{P}(x_0(m) = +1|\underline{\tilde{Y}})}{\mathrm{P}(x_0(m) = -1|\underline{\tilde{Y}})} \tag{4}$$

with $\underline{\tilde{Y}}$ being the input sequence to the channel decoder. The hardbit is simply

$$\hat{x}_0(m) = \mathrm{sign}[L_0(m)] \tag{5}$$

whereas the corresponding bit error probability can be derived from (4) as

$$p_{e0}(m) = \frac{1}{1 + \exp|L_0(m)|}. \tag{6}$$

Note that the less complex soft-output Viterbi algorithm (SOVA) [20] yields approximations to the L-values in (4)–(6).

B. Parameter Transition Probabilities

Knowing the (instantaneous) bit error probability according to (3) or (6) we get the conditional bit probability for the transition of a transmitted bit $x_0^{(i)}(m)$ to the known received bit $\hat{x}_0(m)$ as

$$\mathrm{P}(\hat{x}_0(m)|x_0^{(i)}(m)) = \begin{cases} 1 - p_{e0}(m) & \text{if } \hat{x}_0(m) = x_0^{(i)}(m) \\ p_{e0}(m) & \text{if } \hat{x}_0(m) \neq x_0^{(i)}(m). \end{cases} \tag{7}$$

If we consider the equivalent channel to be memoryless, the *parameter* transition probability reads

$$\mathrm{P}(\underline{\hat{x}}_0|\underline{x}_0^{(i)}) = \prod_{m=0}^{M-1} \mathrm{P}(\hat{x}_0(m)|x_0^{(i)}(m)). \tag{8}$$

This term includes the channel characteristics and provides the probability of a transition from any possibly transmitted bit combination $\underline{x}_0^{(i)}$, $i \in \{0, 1, \cdots, 2^M - 1\}$, to the known received bit combination $\underline{\hat{x}}_0$.

In real-world applications the assumption of a memoryless equivalent channel can be a coarse approximation, even if an interleaving scheme is employed. However, the achievable error concealment based on (8) is still very effective (see Section V).

C. A Priori Knowledge

In specifying the required *a priori* knowledge there are some degrees of freedom. In the general case we model the quantized parameter as a Markov process of Nth order according to

$$\begin{aligned} &\mathrm{P}(\underline{x}_n|\underline{x}_{n-1}, \cdots, \underline{x}_{n-N}, \underline{x}_{n-N-1}, \cdots) \\ &\quad = \mathrm{P}(\underline{x}_n|\underline{x}_{n-1}, \cdots, \underline{x}_{n-N}). \end{aligned}$$

To find out an appropriate Markov order it is convenient to measure terms such as $\mathrm{P}(\underline{x}_n)$, $\mathrm{P}(\underline{x}_n|\underline{x}_{n-1})$, and $\mathrm{P}(\underline{x}_n, \underline{x}_{n-1})$ or even higher order conditional and joint probabilities. This can be achieved by applying a large speech database to the speech encoder and by counting how often the different quantizer output symbols, or different pairs of output symbols, occur. We call $\mathrm{P}(\underline{x}_n)$ *0th order a priori knowledge* (AK0) because it gives a statistical description of a 0th order Markov process, i.e., a memoryless process. Accordingly, we call $\mathrm{P}(\underline{x}_n|\underline{x}_{n-1})$ or $\mathrm{P}(\underline{x}_n, \underline{x}_{n-1})$ *first-order a priori knowledge* (AK1) because it refers to a first-order Markov process. The decision which model should be taken is a matter of the

- observed redundancy;
- allowed complexity of the softbit speech decoder;
- tradeoff between performance and complexity.

If we model a parameter as 0th order Markov process, 2^M probabilities $\mathrm{P}(\underline{x}_n^{(i)})$ with $i \in \{0, 1, \cdots, 2^M - 1\}$ have to be stored in the decoder. With the entropy defined as

$$H(\underline{x}_n) = -\sum_{i=0}^{2^M-1} \mathrm{P}(\underline{x}_n^{(i)}) \log_2 \mathrm{P}(\underline{x}_n^{(i)}) \tag{9}$$

the redundancy of $\Delta R = M - H(\underline{x}_n)$ bits can be used for error concealment.

If a parameter is modeled as first order Markov process, 2^{2M} probabilities $\mathrm{P}(\underline{x}_n^{(i)}|\underline{x}_{n-1}^{(j)})$ with $i, j \in \{0, \cdots, 2^M - 1\}$ have to be stored in the decoder. Then a redundancy of $\Delta R = M - H(\underline{x}_n|\underline{x}_{n-1})$ bits can be used for error concealment, with the conditional entropy

$$H(\underline{x}_n|\underline{x}_{n-1}) = -\sum_{i=0}^{2^M-1}\sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_n^{(i)}, \underline{x}_{n-1}^{(j)}) \cdot \log_2 \mathrm{P}(\underline{x}_n^{(i)}|\underline{x}_{n-1}^{(j)}). \tag{10}$$

This can be extended to even higher Markov orders, while the storage requirements are $2^{M(N+1)}$ words. In Section V examples of the appropriate Markov order are given.

D. *A Posteriori Probabilities*

For the estimation of a speech codec parameter at the receiver, *a posteriori* probabilities providing information about any possibly transmitted bit combination $\underline{x}_0^{(i)}$ are required.

1) Approach with no A Priori Knowledge: If there is no *a priori* knowledge available about the regarded speech codec parameter, it has to be assumed that the quantizer output symbols are uncorrelated and equally likely. In this case only the channel dependent information can be exploited in terms of softbits. The required probabilities are

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0) = C \cdot \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(i)}) \tag{11}$$

with the normalization constant

$$C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(l)})}.$$

Here and in the following the constant C is always used to normalize the left hand side *a posteriori* probabilities such that $\Sigma_{i=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \cdots) = 1$. It should be noted that in practical coding schemes the assumption of equally likely quantizer outputs is usually not met. The widely used Lloyd-Max quantizers [39] yield e.g., identical quantization error variance contributions of each quantization interval i but not identical probabilities $\mathrm{P}(\underline{x}_0^{(i)})$.

2) Approach with 0th Order A Priori Knowledge (AK0): If there is 0th order *a priori* knowledge available the Bayes rule yields *a posteriori* probabilities

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0) = C \cdot \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(i)}) \cdot \mathrm{P}(\underline{x}_0^{(i)}) \quad \text{with} \quad C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(l)}) \cdot \mathrm{P}(\underline{x}_0^{(l)})}. \tag{12}$$

If all quantized parameter values are equally likely, i.e., if $M = H(\underline{x}_n)$ or equivalently $\mathrm{P}(\underline{x}_0^{(i)}) = 2^{-M}$ holds then (11) and (12) are identical.

3) Approach with First-Order A Priori Knowledge (AK1): The *a posteriori* term in (12) assumes successive bit combinations to be independent. If there is residual correlation of first order, i.e., if $H(\underline{x}_n) > H(\underline{x}_n|\underline{x}_{n-1})$, the *a posteriori* probabilities can be extended to regard this correlation as well.

The maximum information that is available at the decoder consists of the complete sequence of the already received bit combinations $\hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{x}}_{-2}, \cdots = \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}$, where $\hat{\underline{X}}_{-1}$ includes the complete history of the received bit combinations until the previous time instant $n = -1$. Given the first order *a priori* knowledge $\mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)})$ the *a posteriori* probabilities $\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ exploiting this complete history can be computed by the recursion as shown in (13) at the bottom of the page. If there are no residual correlations in the quantized parameter, then $\mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)}) = \mathrm{P}(\underline{x}_0^{(i)})$ and (13) simplifies to (12). It should be emphasized that (13) denotes an extrapolation which does not require any algorithmic delay (see also [21], [29]).

If a delay of a single frame or subframe is allowed, interpolation instead of extrapolation can be done by using

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_{+1}, \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot \mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) \cdot \sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_{+1}|\underline{x}_{+1}^{(l)}) \cdot \mathrm{P}(\underline{x}_{+1}^{(l)}|\underline{x}_0^{(i)}) \tag{14}$$

with $\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ taken from recursion (13). Here a future bit combination $\hat{\underline{x}}_{+1}$ has to be received before the *a posteriori* probabilities for the present parameter can be computed. Due to the independence from the time index n the *a priori* knowledge $\mathrm{P}(\underline{x}_{+1}|\underline{x}_0)$ equals the already known term $\mathrm{P}(\underline{x}_0|\underline{x}_{-1})$.

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot \mathrm{P}(\hat{\underline{x}}_0|\hat{\underline{x}}_0^{(i)} \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}^{(j)}|\hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2} \quad \text{with}$$
$$C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(l)} \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(l)}|\underline{x}_{-1}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}^{(j)}|\hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})} \tag{13}$$

If a delay of two frames or subframes is allowed we get

$$\begin{aligned}\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_{+2},\hat{\underline{x}}_{+1},\hat{\underline{x}}_0,\hat{\underline{X}}_{-1}) &= C\cdot\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\hat{\underline{X}}_{-1})\\ &\cdot\sum_{g=0}^{2^M-1}\mathrm{P}(\hat{\underline{x}}_{+2}|\underline{x}_{+2}^{(g)})\cdot\sum_{h=0}^{2^M-1}\mathrm{P}(\underline{x}_{+2}^{(g)}|\underline{x}_{+1}^{(h)})\\ &\mathrm{P}(\hat{\underline{x}}_{+1}|\underline{x}_{+1}^{(h)})\cdot\mathrm{P}(\underline{x}_{+1}^{(h)}|\underline{x}_0^{(i)}).\end{aligned}\tag{15}$$

Two future bit combinations $\hat{\underline{x}}_{+1}$ and $\hat{\underline{x}}_{+2}$ must be known. Theoretically, *a posteriori* probabilities exploiting an arbitrary number of future received bit combinations can be specified [40], but there is a complexity problem. In Section IV it will be shown for the first order Markov case that with a delay of only one parameter instant most of the quality gain which is possible according to the more general case of (15) can be achieved.

Even for higher order *a priori* knowledge extrapolative and interpolative *a posteriori* probabilities can be computed [40]. Because of the immense computational complexity and storage capacity this is not really of practical interest for $N \geq 2$.

E. Parameter Estimation

If any of the previously specified *a posteriori* terms $\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\cdots)$ is computed, the parameter value itself can be estimated. The arbitrary estimation error criterion should reflect the impact of parameter errors on the subjective speech quality.

For a wide area of speech codec parameters the *minimum mean square* (MS) error criterion is appropriate. These parameters may be PCM speech samples, spectral coefficients, gain factors, etc. In contrast to that the estimation of a pitch period should be performed according to the *maximum a posteriori* (MAP) estimator. In the following we discuss these two well known estimators in the context of speech codec parameter estimation.

1) MAP Estimation: The MAP estimator follows the criterion

$$v_0^{(\mathrm{MAP})}=v_0^{(\nu)}\quad\text{with}\quad \nu=\arg\max_i\ \mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\cdots).\tag{16}$$

Independent of the type of the speech codec parameter, a MAP estimation always minimizes the probability of an erroneous decoded parameter [41]. The optimum decoded parameter in a MAP sense $v_0^{(\mathrm{MAP})}$ equals one of the codebook/quantization table entries. This is of great advantage for some applications.

2) MS Estimation: The optimum decoded parameter in a mean square sense equals

$$v_0^{(\mathrm{MS})}=\sum_{i=0}^{2^M-1} v_0^{(i)}\cdot\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\cdots).\tag{17}$$

According to the orthogonality principle of the linear mean square estimation (see, e.g., [41]) the variance of the estimation error $e^{(\mathrm{MS})}=v_0^{(\mathrm{MS})}-v_0$ is $\sigma^2_{e^{(\mathrm{MS})}}=\sigma_v^2-\sigma^2_{v^{(\mathrm{MS})}}$ with σ_v^2 being the variance of the undisturbed parameter v_n and $\sigma^2_{v^{(\mathrm{MS})}}$ denoting the variance of the estimated parameter $v_n^{(\mathrm{MS})}$. Because of $\sigma^2_{e^{(\mathrm{MS})}}\geq 0$ we can state that the variance of the estimated parameter is smaller than or equals the variance of the error free parameter.

For the worst-case channel with $p_e=0.5$ the *a posteriori* probabilities simplify to $\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\cdots)=\mathrm{P}(\underline{x}_0^{(i)})$. If in this case the unquantized parameter $\tilde{v}_n$ as well as the quantization table entries $v^{(i)}$ are distributed symmetrically around zero the MS estimated parameter according to (17) is attenuated to zero (by weighted averaging). These symmetries are often found for gain factors in CELP coders or they can be created by involving the sign of the excitation signal into the estimation of the gain factor. Thus the MS estimation of gain factors results in an inherent muting mechanism providing a *graceful degradation* of the speech quality. This is one of the main advantages of softbit speech decoding.

On the other hand, if the channel is free of errors ($p_e=0$) and $\underline{x}_0^{(\kappa)}$ has been transmitted, then the parameter transition probabilities are zero except $\mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(\kappa)})=1$. This directly yields $\mathrm{P}(\underline{x}_0^{(\kappa)}|\hat{\underline{x}}_0,\cdots)=1$ while all other *a posteriori* probabilities become zero. As a consequence, the MAP estimator as well as the MS estimator yield the correct parameter value $v^{(\mathrm{MAP})}=v^{(\mathrm{MS})}=v$. This is equivalent to bit exactness in clear channel situations.

Finally it should be mentioned that all of the above discussed algorithms can be used in the case of vector quantization. The only difference to scalar quantization consists in the estimation step. Let us consider a P-tuple of codec parameters $\tilde{\underline{v}}_0\in\mathbb{R}^P$ which is coded by M bits. The quantized parameter vector is then $\mathbf{Q}[\tilde{\underline{v}}_0]=\underline{v}_0$ with $\underline{v}_0\in$ CB (CB: codebook). MAP estimation simply yields a parameter vector $\underline{v}_0^{(\mathrm{MAP})}$ instead of a scalar whereas MS estimation can be formulated as

$$\underline{v}_0^{(\mathrm{MS})}=\sum_{i=0}^{2^M-1}\underline{v}_0^{(i)}\cdot\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0,\cdots).\tag{18}$$

IV. Application to Model Parameters

In this section, we want to demonstrate the capabilities of softbit speech decoding by applying it to an artificial Gaussian parameter. We focus on a single parameter rather than on speech reconstruction, thus we use here the term *softbit decoding* instead of softbit *speech* decoding. The parameter $\tilde{v}_n$ which is taken from a first order autoregressive process with correlation factor $\rho_{\tilde{v}\tilde{v}}=0.9$ is quantized by a scalar Lloyd-Max quantizer (LMQ) [39] using M bits. As we are not primarily interested in the optimization of the bit mapping scheme we choose the natural binary code (NBC) [42]. We employ BPSK modulation over an AWGN channel and coherent demodulation, i.e., Fig. 3(b) with $a=1$. The CSI in terms of L_c in (3) is assumed to be ideally known and thus the results presented in Figs. 4–6 as well as Fig. 7 can be interpreted as upper bounds for a practical implementation with estimation of the CSI.

As measure of quality we choose the global signal-to-noise ratio on parameter level

$$\mathrm{SNR}=10\log_{10}\frac{E\{\tilde{v}^2\}}{E\{(v^{(\mathrm{est})}-\tilde{v})^2\}}\quad[\mathrm{dB}]\tag{19}$$

which is henceforth called *parameter SNR*. In the following $v^{(\mathrm{est})}$ may denote either $\hat{v}$ if the SNR of hardbit decoding is

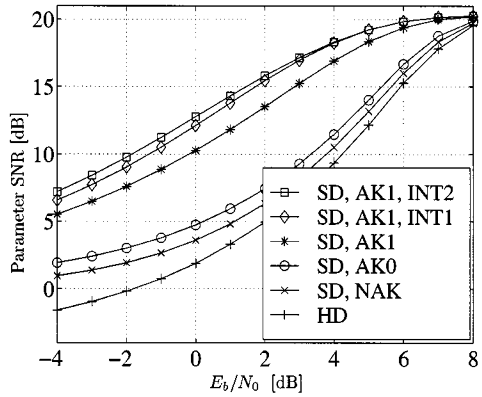

Fig. 4. SNR performance of softbit decoding without channel coding (LMQ, $M = 4$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation). See Table I for explanation of the legend.

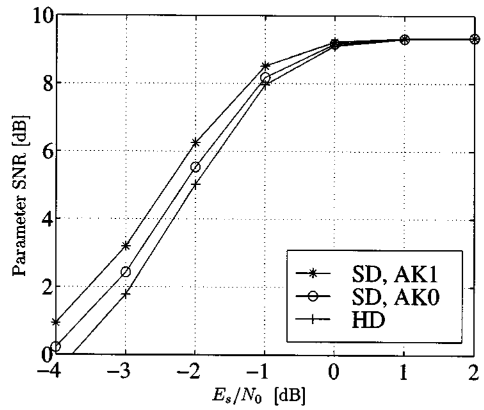

Fig. 5. SNR performance of softbit decoding in connection with channel coding using the soft-output Viterbi algorithm (SOVA) [20] for channel decoding (LMQ, $M = 2$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation), $E_s = rE_b = \frac{1}{2}E_b$. See Table I for explanation of the legend.

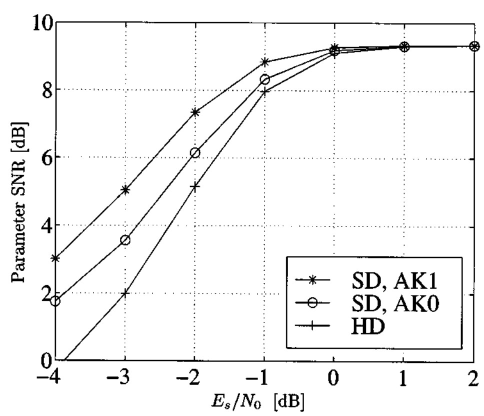

Fig. 6. SNR performance of softbit decoding in connection with channel coding using a sequential decoder similar to the algorithm of Bahl *et al.* [21] for channel decoding (LMQ, $M = 2$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation), $E_s = rE_b = \frac{1}{2}E_b$ because of $r = 1/2$. See Table I for explanation of the legend.

computed or $v^{(\mathrm{MS})}$ in the case of softbit decoding using a MS estimator.

For the typical predictive speech coding schemes we are interested in we found by informal listening tests that the parameter SNR—especially the SNR of the most sensitive parameters—seems to be a reasonable measure of *parameter* quality even for sporadic but extreme parameter errors.

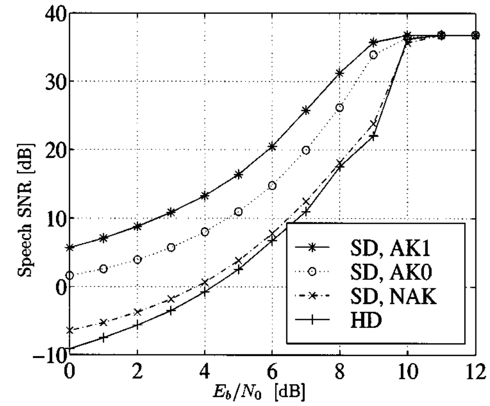

Fig. 7. SNR performance of A-law PCM softbit speech decoding: See Table I for explanation of the legend.

A. Softbit Decoding without Channel Coding

In a first simulation we do not use any channel coding. Thus the input bit $x_0(m)$ and the output bit $\hat{x}_0(m)$ of the channel in Fig. 3(b) belong to the bit combinations of the quantized codec parameters v_0 and $\hat{v}_0$, respectively. The softbit decoding (SD) techniques use softbits (or bit error probabilities) as given in Section III-A1. In the example of Fig. 4 we choose $M = 4$. Under clear channel conditions the SNR is 20.22 dB.

While the SNR of hardbit decoding (HD) decreases rapidly with decreasing channel quality (E_b/N_0 ratio) softbit decoding (SD) allows gains depending on the amount of parameter *a priori* knowledge which is used. By exploiting only the soft information at the channel output (i.e., no *a priori* knowledge, NAK) the SNR performance is slightly improved in comparison to hardbit decoding (HD). A further improvement can be gained by using the 0th order *a priori* knowledge (SD,AK0) which is corresponding to the best case if the correlation is not taken into consideration or if the parameter is uncorrelated (AK0, i.e., $\rho_{\tilde{v}\tilde{v}} = 0$). However, if we exploit the high correlation of the model parameter according to $\rho_{\tilde{v}\tilde{v}} = 0.9$ then the first-order *a priori* knowledge allows significant additional gains (SD,AK1). At the expense of additional delay the parameter SNR can be improved even more if interpolation schemes are used (SD,AK1,INT1 or SD,AK1,INT2).

The conclusion from that experiment is that in case of a highly correlated parameter, softbit decoding by MS allows an improvement of the parameter SNR of about 6..10 dB or a corresponding E_b/N_0 gain of about 3.6 dB.

B. Softbit Decoding with Channel Coding

Now softbit decoding in connection with channel coding shall be investigated. For this purpose we realize the equivalent channel of Fig. 2 as a combination of a convolutional channel encoder, an AWGN channel, and a channel decoder with soft decisions. We choose $M = 2$ and a convolutional encoder with a coding rate of $r = 1/2$ and constraint length $L = 5$. In Fig. 5 a soft-output Viterbi algorithm (SOVA) [20] is employed, in Fig. 6 a sequential realization of the channel decoder according to Bahl *et al.* [21] is used. In the clear channel condition the parameter SNR provided by a 2 bit Lloyd-Max quantization is 9.3 dB.

Figs. 5 and 6 show that softbit decoding is able to reduce the gradient of quality loss beyond the typical threshold of channel decoding. The algorithm of Bahl *et al.* performs much better than the soft-output of the SOVA algorithm which approximates the optimum log-likelihood values of (4). According to Fig. 6 the softbit decoding scheme based on the first order model (SD,AK1) allows E_s/N_0 or E_b/N_0 gains of 1 dB and more.

Nevertheless, it is interesting to compare also Figs. 4 and 6 as they both are showing simulation results with the same gross bit rate of 4 bit/parameter. If we assume the same transmitter power in both cases we have to compare E_b/N_0 in Fig. 4 with E_s/N_0 in Fig. 6. It turns out that softbit decoding without channel coding is able to provide a comparable or even much better SNR at *all* channel conditions than the system comprising a channel coding scheme. This is valid under the assumption of a certain residual parameter correlation that is used by softbit decoding but not used within channel decoding.

It can be concluded that softbit decoding seems to be a promising alternative to channel coding if the parameter correlation is high. In this case the available bit rate might be used completely for (redundant) source encoding such that in good channel conditions the quality becomes significantly higher as with channel coding and in bad channel conditions the quality will be as good or even better as with channel coding [43]–[46].

V. Application to Speech Codecs

A. PCM

The PCM standard G.711 [34] provides a single parameter—the speech sample itself—quantized by $M = 8$ bit. Sundberg *et al.* [47], [48] investigated already a soft decision demodulation for PCM coded speech, where the probability density function of speech samples is exploited as *a priori* knowledge. When a likely transmission error is identified, the corresponding PCM word is rejected by the receiver and replaced by a predictor estimate.

We simulated an A-law PCM transmission over an AWGN channel as in Section IV-A. The CSI is again assumed to be ideally known. We measured entropy values of $H(\underline{x}_n) = 7.83$ bit and $H(\underline{x}_n|\underline{x}_{n-1}) = 6.5$ bit if pause segments were removed from the speech data base. Including speech pauses both values are much lower depending on the percentage of speaker activity. This indicates that the usage of at least 0th order *a priori* knowledge is recommendable. The difference between entropy and conditional entropy reflects the amount of redundancy due to correlation that can be used by first-order *a priori* knowledge.

Fig. 7 shows four different cases in terms of speech SNR (here: equal to parameter SNR) as a function of the E_b/N_0 ratio. In any case, the quality of the MS estimated speech degrades asymptotically to 0 dB with decreasing E_b/N_0, i.e., the inherent muting mechanism of MS estimation. As in Fig. 4 the shape of the curves strongly depends on the order of exploited *a priori* knowledge. In comparison to hardbit decoding (HD) softbit decoding by MS estimation without *a priori* knowledge (SD,NAK) leads to a small SNR gain of about $1 \cdots 2$ dB, while the exploitation of *a priori* knowledge allows gains of up to 10 dB (0th order), and up to 15 dB (first-order), respectively. This corresponds to a significant enhancement of the perceived speech quality although the model of the speech as Markov process of first-order is actually too simple. Further improvements can be gained by increasing the model order [49].

B. ADPCM

To enhance digital enhanced cordless telecommunication (DECT [50]), personal handy phone system (PHS [51]), or personal access communications system (PACS [52]) systems by error concealment we designed a softbit decoder for the G.726 ADPCM standard [35]. There is a single $M = 4$ bit parameter representing a residual signal sample in the logarithmic domain. We measured entropy values of $H(\underline{x}_n) = 3.62$ bit and $H(\underline{x}_n|\underline{x}_{n-1}) = 3.57$ bit. Thus 0th order *a priori* knowledge can exploit $(4 - 3.62)$ bit $= 0.38$ bit of redundancy while first order *a priori* knowledge may exploit $(4 - 3.57)$ bit $= 0.43$ bit of redundancy. This indicates that the SD,AK1 scheme will only be slightly better than the SD,AK0 scheme.

The softbit decoder performs the parameter estimation of the residual signal sample in the linear domain instead of the logarithmic domain. This allows to employ a MS estimator because the squared residual signal error shows strong correlation to the speech quality. Furthermore, the reconstruction of the recursively computed ADPCM scale factor y [35] is very sensitive to bit errors. Thus, we provide an additional MS estimation of the term 2^y.

To simulate fading situations comparable to DECT channel characteristics we assume an FSK modulation over a nonfrequency selective fading channel with 2-path selection diversity and perfect frame synchronization. The reliability information is available only once per frame of 320 bits, i.e., a single estimate of a bit error probability per frame is available to the softbit speech decoder.

Fig. 8 shows the simulation results indicating the superiority of softbit speech decoding (SD,AK0) over hardbit speech decoding (HD). As reference a simple frame repetition mechanism with muting (FR) was simulated. In comparison to that the speech quality degrades much smoother by softbit speech decoding.

C. GSM Full-Rate Codec

We applied the concept of softbit speech decoding to the GSM full-rate speech decoder [53]–[55]. There are several codec parameters of which entropy values are listed in Table II. It is obvious that for most of the parameters a 1st order *a priori* knowledge will be helpful. Exceptions are the RPE grid and the RPE pulses. Nevertheless, for simplicity we assume each codec parameter being modeled as a first-order Markov process.

As an example, the LAR no. 1 as one of the subjectively most important parameters provides a redundancy of 1.54 bit. This is more than 25%. Even speech codecs with lower bit rates than GSM full-rate provide a very high amount of residual redundancy within the spectral parameters: Alajaji *et al.* [32] found, e.g., about 29% of redundancy for the LSP's of the FS 1016 CELP [56] due to nonuniform distribution (0th order *a priori* knowledge concerning time) and due to intraframe correlation (first-order *a priori* knowledge concerning correlation to LSPs of the same frame).

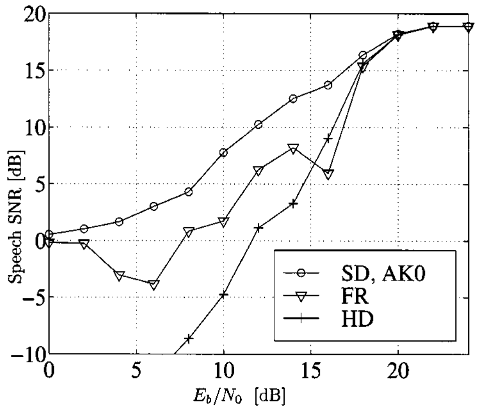

Fig. 8. SNR performance of ADPCM softbit speech decoding, only a single reliability information to any DECT frame of 10 ms (320 bits) is available. See Table I for explanation of the legend.

TABLE I
ABBREVIATIONS

SD	:	Softbit source decoding using ...
AK1	:	1st order a priori knowledge in eq. (13)
AK0	:	0th order a priori knowledge in eq. (12)
NAK	:	no a priori knowledge, eq. (11)
INT2	:	interpolation by exploiting two future bit combinations, eq. (15)
INT1	:	interpolation by exploiting one future bit combination, eq. (14)
BC	:	block coded parameter, eqs. (20) to (22)
RPT	:	repeated parameter transmission, eq. (24)
FR	:	Frame repetition mechanism
HD	:	Hardbit source decoding by table lookup

It turns out that the noninteger GSM codec parameters can be well estimated using a MS estimator. In contrast to that, the estimation of a pitch period (LTP lag) or the RPE grid position should be performed by a MAP estimator.

In Fig. 9, the results of a complete GSM simulation using the COSSAP GSM library [57] with speech and channel (de)coding, (de)interleaving, (de)modulation, a channel model, and equalization are depicted. The channel model represents a typical case of an urban area (TU) with six characteristic propagation paths [58] and a user speed of 50 km/h (TU50). Soft-output channel decoding is carried out by the algorithm of Bahl *et al.* [21]. The reference conventional GSM decoder performs error concealment by a frame repetition (FR) algorithm as proposed in [1]. The bad frame indicator (BFI) is simply set by the evaluation of the *cyclic redundancy check*(CRC).

The SNR surely is not the optimum measure for speech quality. However, informal listening tests show the superiority of the softbit speech decoder in comparison to the conventional decoding scheme in all situations of vehicle speeds and C/I ratios. The new error concealment by softbit speech decoding provides quite a good subjective speech quality down to C/I $= 6$ dB, whereas the conventional frame repetition produces severe distortions already at C/I $= 7$ dB. Even long error bursts caused by a low vehicle speed can be decoded sufficiently by the new technique. In the softbit decoding simulation, the hard annoying clicks in the case of CRC failures and the synthetic sounds of the frame repetition disappeared completely and turned into a slightly noisy or modulated speech. This gives a significant enhancement of the speech quality.

VI. Extended Capabilities of Softbit Speech Decoding

A. Decoding of a Block Code

The concept of softbit speech decoding can be extended to exploit other types of information than the received bit combination $\hat{\underline{x}}_0$ and the appropriate bit error probabilities $\underline{p_e}_0$. We consider the application of block codes to individual codec parameters. Usually, this can be described as mapping of the parameter bit combination $\underline{x}_0^{(i)}$ (M bits) to the valid codewords $y_0^{(i)}$ ($K > M$ bits) with $i \in \{0, 1, \cdots, 2^M - 1\}$.

Based on the concept of softbit speech decoding [49] algorithmic proposals and applications for softbit decoding of block codes have been published by Görtz [59], Heinen *et al.* [60], and more generally in [61]. The *a posteriori* probabilities exploiting 1st order *a priori* knowledge for softbit decoding of block coded parameters are given by the recursion [61]

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{y}}_0, \hat{\underline{Y}}_{-1}) = C \cdot \mathrm{P}(\hat{\underline{y}}_0|y_0^{(i)}) \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}^{(j)}|\hat{\underline{y}}_{-1}, \hat{\underline{Y}}_{-2}). \quad (20)$$

These *a posteriori* probabilities are subject to the parameter estimation as described in Section III-E.

If a *systematic* block code is used with the K bit codewords $y_0^{(i)} = (\underline{x}_0^{(i)}, \underline{z}_0^{(i)})$ including $K - M$ explicit parity bits $\underline{z}_0^{(i)}, i \in \{0, 1, \cdots, 2^M - 1\}$, recursion (20) becomes [59]

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{z}}_0, \hat{\underline{X}}_{-1}, \hat{\underline{Z}}_{-1}) = C \cdot \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(i)}) \cdot \mathrm{P}_{sum}^{(i)} \cdot \mathrm{P}(\hat{\underline{z}}_0|\underline{z}_0^{(i)}) \quad (21)$$

with

$$\mathrm{P}_{sum}^{(i)} = \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}^{(j)}|\hat{x}_{-1}, \hat{\underline{z}}_{-1}\hat{\underline{X}}_{-2}, \hat{\underline{Z}}_{-2}).$$

Equation (21) can be generalized to [61]

$$\mathrm{P}(\underline{x}_0^{(i)}|\hat{\underline{x}}_0, \hat{\underline{z}}_0, \hat{\underline{X}}_{-1}, \hat{\underline{Z}}_{-1}) = C \cdot \mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(i)}) \cdot \mathrm{P}_{sum}^{(i)} \cdot \sum_{l=0}^{2^{K-M}-1} \mathrm{P}(\hat{\underline{z}}_0|\underline{z}_0^{(l)}) \cdot \mathrm{P}(\underline{z}_0^{(l)}|\underline{x}_0^{(i)}) \quad (22)$$

with a summation over the $K - M$ explicit parity bits $\underline{z}_0 z^{(l)}, l \in \{0, 1, \cdots, 2^{K-M} - 1\}$. The *code a priori knowledge* $\mathrm{P}(\underline{z}_0^{(l)}|\underline{x}_0^{(i)})$ describes this block code in a statistical manner. Comparing recursions (13) with (21), (22) the block code's influence can be interpreted as an enhancement of the parameter transition probabilities $\mathrm{P}(\hat{\underline{x}}_0|\underline{x}_0^{(i)})$.

TABLE II
NUMBER OF BITS M, ENTROPY $H(\underline{x}_n)$, CONDITIONAL ENTROPY $H(\underline{x}_n|\underline{x}_{n-1})$ OF THE PARAMETERS OF THE GSM FULL-RATE CODEC [53], [54]

$\left[\frac{\text{bit}}{\text{parameter}}\right]$	LAR No.								LTP		RPE		
	1	2	3	4	5	6	7	8	Lag	Gain	Grid	Max.	Pulse
M	6	6	5	5	4	4	3	3	7	2	2	6	3
$H(\underline{x}_n)$	5.43	4.88	4.75	4.53	3.73	3.76	2.84	2.88	6.31	1.88	1.96	5.39	2.86
$H(\underline{x}_n\|\underline{x}_{n-1})$	4.46	4.29	4.18	4.09	3.37	3.39	2.49	2.46	5.75	1.74	1.96	4.29	2.86

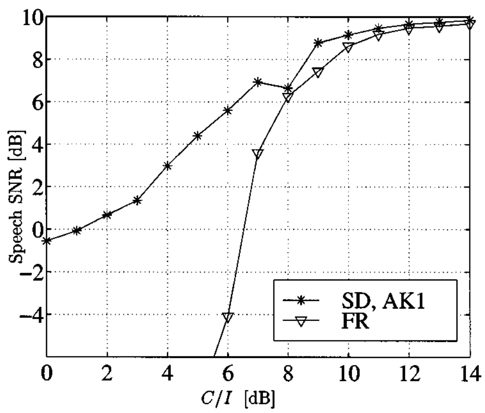

Fig. 9. SNR performance of GSM full-rate softbit speech decoding, TU50 channel. See Table I for explanation of the legend.

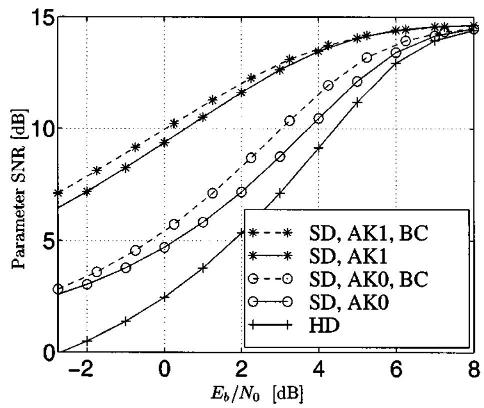

Fig. 10. SNR performance of softbit decoding with and without block code BC (LMQ, $M = 3$, $K = 4$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation). See Table I for explanation of the legend.

A first example of a block code is the simple even parity code of rate $r = M/K$ with $K - M = 1$. Here, recursion (21) can be used.

However, the "parity bit" can alternatively be chosen dependent on the unquantized parameter vector $\underline{\tilde{v}}_0$ rather than on the bit combination $\underline{x}_0$ of the quantized parameter vector $\underline{v}_0$. An example code for $K - M = 1$ is, e.g.,

$$z_0 = \begin{cases} +1, & \text{if } \|\underline{\tilde{v}}_0 - \underline{\overline{v}}_0\|^2 < d^2 \\ -1, & \text{if } \|\underline{\tilde{v}}_0 - \underline{\overline{v}}\|^2 \geq d^2 \end{cases} \tag{23}$$

with $\|\underline{\tilde{v}}_0 - \underline{\overline{v}}_0\|^2$ being the Euclidean distance of $\underline{\tilde{v}}_0$ to its mean $\underline{\overline{v}}_0$. The threshold d of this *source optimized* block code should be adjusted optimally via simulation over a noisy channel model. For a unit variance Gaussian parameter that is scalar quantized by an $M = 3$ bit Lloyd-Max quantizer and transmitted over an AWGN channel, we found an optimum threshold value of $d = 1.1$. This yields the measured code *a priori* knowledge

$$\mathrm{P}(z_0^{(l)}|\underline{x}_0^{(i)}) = \begin{bmatrix} 1 & 0.8953 & 0 & 0 & 0 & 0 & 0.8953 & 1 \\ 0 & 0.1047 & 1 & 1 & 1 & 1 & 0.1047 & 0 \end{bmatrix}$$

with $l \in \{0, 1\}$ for decoding by (22).

Fig. 10 shows simulation results to this source-optimized block code. The parameter and the channel are chosen as in Fig. 4, except the quantization is done with $M = 3$ bit. The block code thus has a code rate of $r = M/K = 3/4$. The dashed curves obtained by block coding show coding gains of up to 0.75 dB as compared to the uncoded transmission. This is especially true if no parameter correlations are exploited by softbit decoding (SD,AK0,BC).

B. Systems with Diversity Reception

In transmission systems with diversity reception softbit speech decoding is able to combine a number of bitstreams with their appropriate bit error probabilities in the sense of *maximum ratio combining* (see, e.g., [62]–[64]). This is possible under the assumption of statistically independent transmission paths.

Let us regard a combination of two received bit combinations $\underline{\hat{x}}_0$ and $\underline{\hat{y}}_0$ whose transmitted bit combinations are the same: $\underline{x}_0 = \underline{y}_0$. This may be a diversity reception problem, but it may also be a problem of how to combine the information of a *repeated* transmission of the same bit combination via a *single* memoryless channel. Thus we discuss the combination problem like a repeated parameter transmission.

If 0th order *a priori* knowledge is available, the *a posteriori* probabilities read (SD,AK0,RPT)

$$\mathrm{P}(\underline{x}_0^{(i)}|\underline{\hat{x}}_0, \underline{\hat{y}}_0) = \frac{1}{C} \cdot \mathrm{P}(\underline{\hat{x}}_0|\underline{x}_0^{(i)}) \cdot \mathrm{P}(\underline{\hat{y}}_0|y_0^{(i)}) \cdot \mathrm{P}(\underline{x}_0^{(i)}). \tag{24}$$

If first-order *a priori* knowledge is available the *a posteriori* probabilities are given recursively by (SD,AK1,RPT)

$$\begin{aligned} &\mathrm{P}(\underline{x}_0^{(i)}|\underline{\hat{x}}_0, \underline{\hat{X}}_{-1}, \underline{\hat{y}}_0, \underline{\hat{Y}}_{-1}) \\ &\quad = \frac{1}{C} \cdot \mathrm{P}(\underline{\hat{x}}_0|\underline{x}_0^{(i)}) \cdot \mathrm{P}(\underline{\hat{y}}_0|y_0^{(i)}) \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0^{(i)}|\underline{x}_{-1}^{(j)}) \\ &\quad\quad \cdot \mathrm{P}(\underline{x}_{-1}^{(j)}|\hat{x}_{-1}, \underline{\hat{X}}_{-2}, \underline{\hat{y}}_{-1}, \underline{\hat{Y}}_{-2}). \end{aligned} \tag{25}$$

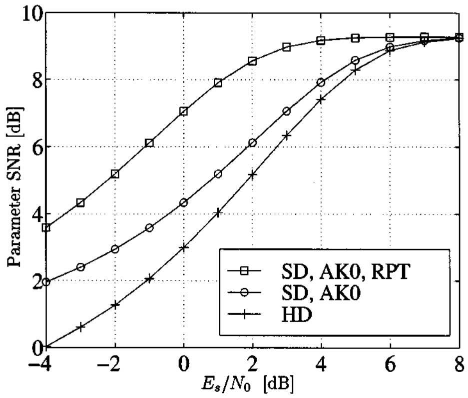

Fig. 11. SNR performance of softbit decoding using parameter repetition or 2-path diversity reception (LMQ, $M = 2, \rho_{\tilde{v}\tilde{v}} = 0$, MS estimation). See Table I for explanation of the legend.

Fig. 11 shows simulation results of a repeated parameter transmission, or, equivalently, to a two-path diversity reception and subsequent softbit decoding. The parameter and the channel are chosen as in Fig. 4, except the quantization is done with $M = 2$ bit and the parameter is uncorrelated ($\rho_{\tilde{v}\tilde{v}} = 0$). As expected, the quality gain of softbit decoding in the case of a repeated parameter transmission (SD,AK0,RPT) compared to softbit decoding of a single received bit combination (SD,AK0) is tremendous and amounts to 3 dB in terms of E_s/N_0 or up to about 3 dB in terms of SNR.

VII. Conclusions

We proposed the new concept of softbit speech decoding which comprises previous empirically motivated approaches to error concealment as special cases. This new concept allows the consequent usage of soft information at all stages of a receiver—not only within equalization and channel decoding, but also within speech decoding.

The softbit speech decoding is driven by

- bit reliability information from the channel (decoder);
- residual redundancy of the codec parameters.

For different coding schemes and different channel models with and without channel coding we discussed how to obtain and how to use these informations. Based on this information, the softbit speech decoding process consists of the four steps

- calculation of parameter transition probabilities;
- computation of *a posteriori* probabilities;
- estimation of codec parameters;
- conventional speech decoding.

We derived algorithms for computing *a posteriori* probabilities taking different degrees of *a priori* knowledge into consideration. Two common estimators were discussed showing that the mean square estimator is able to yield the desired graceful degradation of speech in case of a decreasing quality of the transmission link because of its inherent muting mechanism.

In addition, we showed how block coding of single codec parameters, repeated parameter transmission, or diversity reception can be taken into consideration for quality improvement by softbit speech decoding.

Softbit decoding can be applied to any speech coding algorithm without modifications at the transmitter side. For noisy channels and various speech codecs we demonstrated the significant improvement of the speech quality. In clear channel conditions, bit exactness is preserved.

Finally it should be mentioned that softbit source decoding can also be applied to audio and video coding schemes and that under certain conditions channel coding might be replaced by redundant source encoding in combination with softbit source decoding.

Acknowledgment

The authors would like to thank D. Herzberg, O. Scheufen, J. A. Andonegui, B. Dortschy, and S. Heinen for a lot of programming and inspiring discussions as well as the reviewers and the editor for several valuable suggestions.

References

[1] *Recommend. GSM 06.11 Substitution and Muting of Lost Frames for Full Rate Speech Traffic Channels*, ETSI TC-SMG, 1992.

[2] *Eur. Digital Cellular Telecommun. System Half Rate Speech Part 3: Substitution and Muting of Lost Frames for Half Rate Speech Traffic Channels (GSM 06.21)*, ETSI TM/TM5/TCH-HS, 1995.

[3] *Digital Cellular Telecommunications System: Substitution and Muting of Lost Frames for Enhanced Full Rate (EFR) Speech Traffic Channels (GSM 06.61)*, ETSI TC-SMG, 1996.

[4] N. S. Jayant and S. W. Christensen, “Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure,” *IEEE Trans. Commun.*, vol. COMM-29, pp. 101–109, Feb. 1981.

[5] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W.-C. Wong, “Waveform substitution techniques for recovering missing speech segments in packet voice communications,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1440–1448, Dec. 1986.

[6] R. A. Valenzuela and C. N. Animalu, “A new voice-packet reconstruction technique,” in *Proc. Int. Conf. Acoustics, Speech, Signal Processing ’89*, Glasgow, U.K., May 1989, pp. 1334–1336.

[7] N. Erdöl, C. Castelluccia, and A. Zilouchian, “Recovery of missing speech packets using the short-time energy and zero-crossing measurements,” *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 295–303, July 1993.

[8] A. Stenger, K. B. Younes, R. Reng, and B. Girod, “A new error concealment technique for audio transmission with packet loss,” in *Signal Processing VIII, Proc. EUSIPCO’96*, Trieste, Italy, Sept. 1996, pp. 1965–1968.

[9] K. Clüver, “An ATM speech codec with improved reconstruction of lost cells,” in *Signal Processing VIII, Proc. EUSIPCO’96*, Trieste, Sept. 1996, pp. 1641–1643.

[10] D. Sereno, “Frame substitution and adaptive post-filtering in speech coding,” in *Proc. EUROSPEECH’91*, Genoa, Italy, Sept. 1991, pp. 595–598.

[11] H.-Y. Su and P. Mermelstein, “Improving the speech quality of cellular mobile systems under heavy fading,” in *Proc. Int. Conf. Acoustics, Speech, Signal Processing ’92*, San Francisco, CA, 1992, pp. II 121–II 124.

[12] I. Heikkila, H. Jokinen, and J. Ranta, “A signal quality detecting circuit and method for receivers in the GSM system,” Eur. Patent 0 648 032 A1, Oct. 1993.

[13] K. Järvinen and J. Vainio, “Detection of defective speech frames in a receiver of a digital speech communication system,” World Patent WO 96/09 704, Sept. 1994.

[14] T. B. Minde, P. Mustel, H. Nilsson, and T. Lagerqvist, “Soft error correction in a TDMA radio system,” World Patent WO 95/16 315, Dec. 1993.

[15] N. Farvardin and V. Vaishampayan, “Optimal quantizer design for noisy channels: An approach to combined source-channel coding,” *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827–838, Nov. 1987.

[16] N. Farvardin, “A study of vector quantization for noisy channels,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 799–809, July 1990.

[17] M. Skoglund and P. Hedelin, “Vector quantization over a noisy channel using soft decision decoding,” in *Proc. Int. Conf. Acoustics, Speech, Signal Processing ’94*, vol. 5, Adelaide, Australia, Apr. 1994, pp. 605–608.

[18] M. Skoglund, "On Soft Decoding and Robust Vector Quantization," Ph.D. dissertation, Chalmers Univ. Technol., Göteborg, Sweden, Mar. 1997.
[19] V. Cuperman, F.-H. Liu, and P. Ho, "Robust vector quantization for noisy channels using soft decision and sequential decoding," *Eur. Trans. Telecommun.*, vol. 5, pp. 7–18, Sept. 1994.
[20] J. Hagenauer and P. Hoeher, "A viterbi algorithm with soft-decision outputs and its applications," in *Proc. GLOBECOM'89*, Dallas, TX, Nov. 1989, pp. 1680–1686.
[21] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.
[22] C. E. Shannon, "A mathematicl theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.
[23] P. Ligdas, N. Seshadri, and W. Turin, "Statistical methods for speech transmission using hidden markov models," in *Proc. 31st Conf. Information Sciences Systems*, Mar. 1997, pp. 546–551.
[24] W. Turin, *Digital Transmission Systems: Performance Analysis and Modeling*: McGraw-Hill, 1998.
[25] C. G. Gerlach, "A probabilistic framework for optimum speech extrapolation in digital mobile radio," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '93*, vol. 2, Minneapolis, MN, Apr. 1993, pp. 419–422.
[26] ——, "Beiträge zur Optimalität in der codierten Sprachübertragung," Ph.D. dissertation, Aachener Beiträge zu digitalenNachrichtensystemen, Aachen, Germany, 1996.
[27] S. Feldes, "Enhancing robustness of coded LPC-spectra to channel errors by use of residual redundancy," in *Proc. EUROSPEECH'93*, Berlin, Germany, Sept. 1993, pp. 1147–1150.
[28] ——, "Restredundanzbasierte Rekonstruktionsverfahren bei der digitalen Sprachüberetragung in Mobil-funksystemen," Ph.D. dissertation, Tech. Univ. Darmstadt, Germany, 1994.
[29] N. Phamdo and N. Farvardin, "Optimal detection of discrete markov sources over discrete memoryless channels—Applications to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 186–193, Jan. 1994.
[30] W. Xu, J. Hagenauer, and J. Hollmann, "Joint source-channel decoding using residual redundancy in compressed images," in *Proc. Int. Conf. Communication '96*, June 1996, pp. 142–148.
[31] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.
[32] F. I. Alajaji, N. C. Phamdo, and T. E. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325–336, Sept. 1996.
[33] S. Heinen, A. Geiler, and P. Vary, "MAP channel decoding by exploiting multilevel source a priori knowledge," in *Proc. EPMCC*, Bonn, Germany, Oct. 1997, pp. 467–473.
[34] *ITU-T Recommend. G.711, Pulse Code Modulation (PCM) of Voice Frequencies*, 1972.
[35] *ITU-T Recommend. G.726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (AD-PCM)*, 1990.
[36] J. Hagenauer, "Vom Aalogwert zum bit und zurück," *Frequenz*, vol. 51, pp. 211–227, Sept. 1997.
[37] P. Vary and T. Fingscheidt, "From soft decision channel decoding to soft decision speech decoding," in *Proc. 2nd ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung*. Aachen, Germany, Mar. 1998, pp. 155–162.
[38] J. Hagenauer, "Viterbi decoding of convolutional codes for fading- and burst-channels," in *Proc. 1980 Zurich Seminar Digital Commun.*, Zürich, Switzerland, 1980, pp. G2.1–G2.7.
[39] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.
[40] T. Fingscheidt, "Softbit-Sprachdecodierung in digitalen Mobilfunksystemen," Ph.D. thesis, Aachener Berträge zu digitalen Nachrichtensystemen, Aachen, Germany, 1998.
[41] J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory*. Tokyo, Japan: McGraw-Hill/Kogakusha, 1978.
[42] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[43] T. Fingscheidt, P. Vary, and J. A. Andonegui, "Robust speech decoding: Can error concealment be better than error correction?," in *Proc. 2nd ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung*, Aachen, Germany, Mar. 1998.
[44] S. Heinen, T. Hindelang, and P. Vary, "Channel codes for softbit source decoding: Estimation of correlated parameters," in *Proc. 3rd ITG Conf. Source and Channel Coding*, Munich, Germany, Jan. 2000, pp. 259–266.
[45] T. Hindelang, J. Hagenauer, and S. Heinen, "Source-controlled channel decoding: Estimation of correlated parameters," in *Proc. 3rd ITG Conf. Source Channel Coding*, Munich, Germany, Jan. 2000, pp. 251–258.
[46] T. Fingscheidt, T. Hindelang, R. V. Cox, and N. Seshadri, "Combined source/channel decoding: When minimizing bit error rate is suboptimal," in *Proc. 3rd ITG Conf. Source Channel Coding*, Munich, Germany, Jan. 2000, pp. 273–277.
[47] W. C. Wong, R. Steele, and C.-E. Sundberg, "Soft decision demodulation to reduce the effect of transmission errors in logarithmic PCM transmitted over Rayleigh fading channels," *AT&T Bell Labs. Tech. J.*, vol. 63, pp. 2193–2213, Dec. 1984.
[48] C.-E. Sundberg, "Soft decision demodulation for PCM encoded speech signals," *IEEE Trans. Commun.*, vol. 26, pp. 854–859, June 1978.
[49] T. Fingscheidt and P. Vary, "Robust speech decoding: A universal approach to bit error concealment," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '97*, vol. 3, Munich, Germany, Apr. 1997, pp. 1667–1670.
[50] *Digital Enhanced Cordless Telecommunications Common Interface*, ETS 300 175, 1992.
[51] *Personal Handy Phone System*, RCR STD-28, 1995.
[52] *Personal Access Communications System Unlicensed*, ANSI J-STD-014, 1995.
[53] *Recommend. GSM 06.10 GSM Full Rate Speech Transcoding*, 1992.
[54] T. Fingscheidt and O. Scheufen, "Robust GSM speech decoding using the channel decoder's soft output," in *Proc. EUROSPEECH'97*, Rhodes, Greece, Sept. 1997, pp. 1315–1318.
[55] ——, "Error concealment in the GSM system by softbit speech decoding," in *Proc. 9th Aachen Symp. Signaltheorie*, Aachen, Germany, Mar. 1997, pp. 229–232.
[56] *Details to assist in implementation of federal standard 1016 CELP*, 1992.
[57] *COSSAP—Model Libraries Volume 3*. Mountain View, CA: Synopsys, Inc., 1995.
[58] *Recommend. GSM 05.05 Radio Transmission and Reception*, ETSI TC-SMG, 1992.
[59] N. Görtz, "How to combine forward error detection and bit-reliability information in source decoding," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998.
[60] S. Heinen, M. Adrat, O. Steil, P. Vary, and W. Xue, "A 6.1 to 13.3 kbit/s variable rate CELP codec (VR-CELP) for AMR speech coding," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '99*, Phoenix, AZ, Mar. 1999.
[61] T. Fingscheidt, S. Heinen, and P. Vary, "Joint speech codec parameter and channel decoding of parameter individual block codes (PIBC)," in *IEEE Workshop Speech Coding*, Porvoo, Finland, June 1999, pp. 75–77.
[62] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE Press, 1974.
[63] J. D. Parsons and J. G. Gardiner, *Mobile Communication Systems*. Glasgow, U.K.: Blackie, 1989.
[64] W. Papen, "Makrodiversität und Signalkombination in zellularen digitalen Mobilfunksystemen," Ph.D. dissertation, Aachener Beiträge zu digitalen Nachrichtensystemen, Aachen, Germany, 1996.
[65] T. Fingscheidt, P. Vary, and J. A. Andonegui, "Robust speech decoding: Can error concealment be better than error correction?," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing '98*, Seattle, WA, May 1998, pp. 169–172.

Tim Fingscheidt received the Dipl.-Ing. degree in electrical engineering in 1993 from Aachen University of Technology, Aachen, Germany. His work at the Institute of Communication Systems and Data Processing was on speech and channel coding algorithms for wireless communications with main focus on error concealment techniques. In 1998, he received the Ph.D. degree with the dissertation "Softbit speech decoding in mobile communication systems."

He then joined the Speech Processing Software and Technology Research Department, AT&T Labs, Florham Park, NJ, as a Consultant for research on joint source/channel coding and speech enhancement in wireless systems. Since October 1999, he has been with the Signal Processing Department, Siemens AG, ICM CD Mobile Phones, Munich, Germany.

Peter Vary received the Dipl.-Ing. degree in electrical engineering in 1972 from the University of Darmstadt, Darmstadt, Germany. In 1978, he received the Ph.D. degree from the University of Erlangen-Nuremberg, Germany.

In 1980, he joined Philips Communication Industries (PKI), Nuremberg, where he became Head of the Digital Signal Processing Group. Since 1988, he has been Professor at Aachen University of Technology, Aachen, Germany, and Head of the Institute of Communication Systems and Data Processing. His main research interests are speech coding, channel coding, error concealment, adaptive filtering for acoustic echo cancellation and noise reduction, and concepts of mobile radio transmission.