Bad Parameter Indication for Error Concealment in Wireless Multimedia Communication

Tobias Breddermann, Stanislaus Iwelski, and Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
Email: {breddermann|iwelski|vary}@ind.rwth-aachen.de

***Abstract*—Wireless communication of multimedia signals is often based on the transmission of source codec parameters which are protected against channel noise by channel coding. Commonly, a *cyclic redundancy check* (CRC) is introduced in order to check for erroneous frames after error correction. If the CRC fails, a *bad frame indication* (BFI) flag is set which controls a codec-specific frame error concealment. This results in the substitution of complete multimedia frames by their concealed versions even if only a few parameters may be corrupted. The proposed system is based on iterative source-channel decoding. A low complexity *bad parameter indication* (BPI) scheme is introduced which identifies faulty decoded parameters and enables more effective parameter-individual concealment strategies. This concept may be applied to any multimedia codec (speech, audio, video, images).**

I. Introduction

It has been proven that iterative source-channel decoding (ISCD) [1], [2], [3] can perform close to the Shannon limit with reasonable computational complexity and bandwidth consumption. The improvement of decoding quality is achieved by utilizing residual redundancy of the source encoded signal by means of a turbo-like exchange of extrinsic information between a *soft decision source decoder* (SDSD) [4] and a *soft input soft output* (SISO) channel decoder. A successful application to the *adaptive multi-rate wideband* (AMR-WB) speech codec has been presented in [5].

Many systems apply *cyclic redundancy checks* (CRCs) to check for erroneous frames after error correction. Based on the CRC, a *bad frame indication* (BFI) flag is set which controls the concealment routine of the multimedia codec. However, this can result in a concealment of packets containing only a few faulty decoded codec parameters with having, possibly, little impact on the synthesized signal quality.

Applying ISCD provides further opportunities. It has been shown in [4] that soft bit speech decoding in conjunction with a *maximum a posteriori* (MAP) [6] parameter estimation inherently performs conventional error concealment. It has been further shown that *mean square* (MS) estimation refines this error concealment by graceful degradation of less reliable source codec parameters. However, the MAP approach can result in faulty decoded parameters and the MS approach implies major modifications within the speech decoder.

In this paper, we propose an improved MAP approach. For discovering faulty decoded parameters, the available parameter-related *a posteriori* reliabilities calculated by the SDSD are exploited. A *bad parameter indication* (BPI) flag is set which enables either standardized frame error concealment schemes as, e.g., the state machine of the AMR speech codec [7] or even novel and more sophisticated parameter-individual concealment approaches. This is achieved without spending additional data rate for error detection (no CRCs).

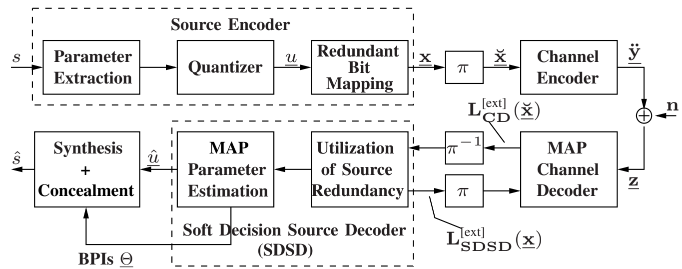

Fig. 1. Baseband model of the ISCD transceiver.

II. ISCD System Model

In Fig. 1 the baseband model of the ISCD system is depicted. A source encoder generates frames $\underline{u} = (u_1, \ldots, u_\kappa, \ldots, u_K)$ of K quantized codec-specific parameters which are, in general, non-uniformly distributed and correlated in time. The $u_\kappa \in \mathbb{U}_\kappa = \{u_\kappa^{(1)}, \ldots, u_\kappa^{(i)}, \ldots, u_\kappa^{(Q_\kappa)}\} \subset \mathbb{N}_0$ represent quantizer code book indices with $Q_\kappa = |\mathbb{U}_\kappa|$ denoting the number of code book entries. A unique bit pattern $\mathbf{x}_\kappa = (x_\kappa(1), \ldots, x_\kappa(k), \ldots, x_\kappa(M_\kappa)) \in \mathbb{X}_\kappa = \{\mathbf{x}_\kappa^{(1)}, \ldots, \mathbf{x}_\kappa^{(i)}, \ldots, \mathbf{x}_\kappa^{(Q_\kappa)}\}$ of $M_\kappa \geq \log_2 Q_\kappa$ bits is assigned to each u_κ by redundant bit mapping (RBM) [8]. Note that the bit mapping can differ from parameter to parameter. The interleaver π rearranges the frame $\underline{\mathbf{x}}$ consisting of all bit patterns $\mathbf{x}_\kappa$ in a deterministic, pseudo-random like manner to the interleaved frame $\underline{\breve{\mathbf{x}}}$. For channel encoding we use a recursive convolutional code of memory J and of rate $r^{\text{CC}} = 1$ in order to become capacity achieving [9]. Prior to transmission, the encoded bits are mapped to bipolar values $\underline{\ddot{\mathbf{y}}}$ with symbol energy $E_s = 1$.

The received symbols $\underline{\mathbf{z}}$ are transformed to log-likelihood ratios (LLRs) before being evaluated in a turbo process by exchanging *extrinsic* reliabilities between the MAP channel decoder ($\mathbf{L}_{\mathbf{CD}}^{[\text{ext}]}(\underline{\breve{\mathbf{x}}})$) and the SDSD ($\mathbf{L}_{\mathbf{SDSD}}^{[\text{ext}]}(\underline{\mathbf{x}})$). The SDSD exploits the residual redundancy of

978-1-4244-3574-6/10/$25.00 ©2010 IEEE

the source codec parameters in order to refine the *extrinsic* information which is fed back to the channel decoder. After a fixed number of iterations each parameter is estimated by the related *a posteriori* probabilities $P(u_\kappa^{(i)}|\underline{\mathbf{z}})$ of each quantizer code book index $u_\kappa^{(i)}$ according to

$$\hat{u}_\kappa = u_\kappa^{(\nu)} \quad \text{with} \quad \nu = \arg\max_i P(u_\kappa^{(i)}|\underline{\mathbf{z}}), \tag{1}$$

$$\begin{aligned} P(u_\kappa^{(i)}|\underline{\mathbf{z}}) &\approx C \cdot \exp\left(\sum_{k=1}^{M_\kappa} \frac{x_\kappa(k)}{2} L(x_\kappa(k)|\underline{\mathbf{z}})\right) \\ &= C \cdot \exp\left(R(u_\kappa^{(i)}|\underline{\mathbf{z}})\right), \end{aligned} \tag{2}$$

$$L(x_\kappa(k)|\underline{\mathbf{z}}) = L_{\text{SDSD}}^{[\text{ext}]}(x_\kappa(k)) + L_{\text{CD}}^{[\text{ext}]}(x_\kappa(k)), \tag{3}$$

where $R(u_\kappa^{(i)}|\underline{\mathbf{z}})$ is the logarithmic *a posteriori* reliability of the quantizer code book index $u_\kappa^{(i)}$. The constant factor C ensures a total probability of 1. The determination rules for the *extrinsic* information of both decoders are given in [1], [2]. After MAP parameter estimation, faulty parameters $\hat{u}_\kappa$ are indicated by a set of parameter-individual BPI flags $\underline{\Theta}$ which are determined exploiting the logarithmic *a posteriori* reliabilities $R(\hat{u}_\kappa = u_\kappa^{(\nu)}|\underline{\mathbf{z}})$ as described in Sec. III. Based on the BPIs, parameter-individual concealment strategies can be applied before signal syntheses providing a less degraded perceptual speech quality after speech decoding.

III. Bad Parameter Indication

The ISCD system introduced in Sec. II performs parameter estimation according to (1). However, in particular for bad channel qualities E_s/N_0 this estimation may fail. Then it might be beneficial to discard or replace some or even all corrupted parameters within a frame depending on their influence on the signal quality.

In what follows, we will demonstrate that the logarithmic *a posteriori* reliability $R(u_\kappa^{(\nu)}|\underline{\mathbf{z}})$ can serve as a robust classifier for the identification of faulty decoded parameters. For illustration, we consider an ISCD-transmission of 40000 AMR speech frames (800 seconds of speech encoded at 12.2 kbit/s) over an AWGN channel. The AMR codec parameters are unequally protected by parameter-individual RBMs. Let us, e.g., consider parameter u_1 which corresponds to the first line spectral frequency submatrix (LSF1). The line spectral frequencies are determined by split matrix quantization of the LP filter coefficients of two consecutive LP analysis filters [10]. These parameters are quite sensitive since marginal distortions can cause filter instabilities or severe acoustic effects. Therefore, it is extremely important to identify faulty decoded parameters. The distribution of the logarithmic *a posteriori* reliability $R(u_1^{(\nu)}|\underline{\mathbf{z}})$ after MAP decoding for an AWGN channel with $E_s/N_0 = \{-1\,\text{dB}, 0\,\text{dB}, 1\,\text{dB}\}$ is shown in Fig. 2. Correctly decoded parameters $\hat{u}_1 = u_1^{(\nu)} = u_1$ are marked by gray dots and faulty decoded parameters $\hat{u}_1 = u_1^{(\nu)} \neq u_1$ by black crosses. The reliabilities of the major fraction of correctly decoded parameters rapidly increase with increasing E_s/N_0, while the reliabilities of faulty decoded parameters still stay close to zero. This behavior is generally shown by each codec parameter which enables a quite robust identification of faulty decoded parameters.

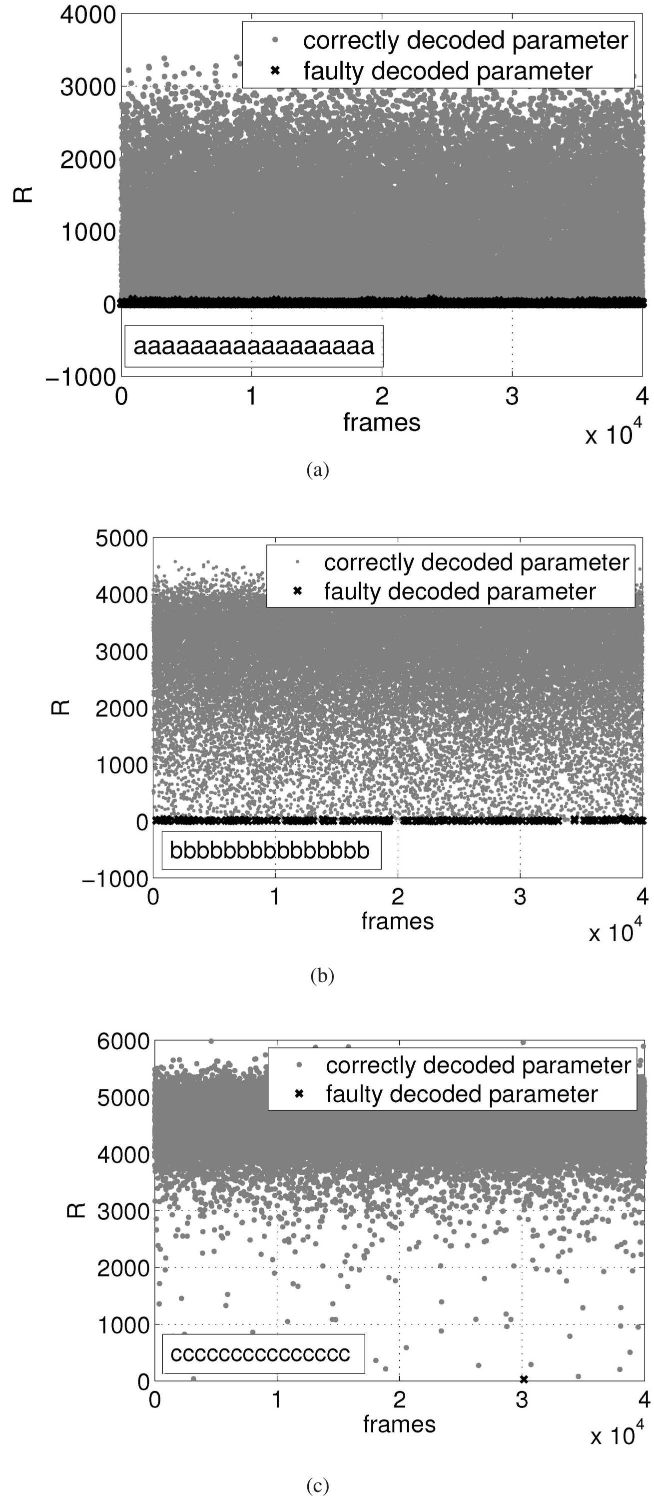

Fig. 2. $R(u_1^{(\nu)}|\underline{\mathbf{z}})$ after MAP decoding for a transmission of 40000 12.2 kbit/s AMR speech frames over an AWGN channel with $E_s/N_0 = \{-1\,\text{dB}, 0\,\text{dB}, 1\,\text{dB}\}$. Parameter u_1 corresponds to the LSF1.

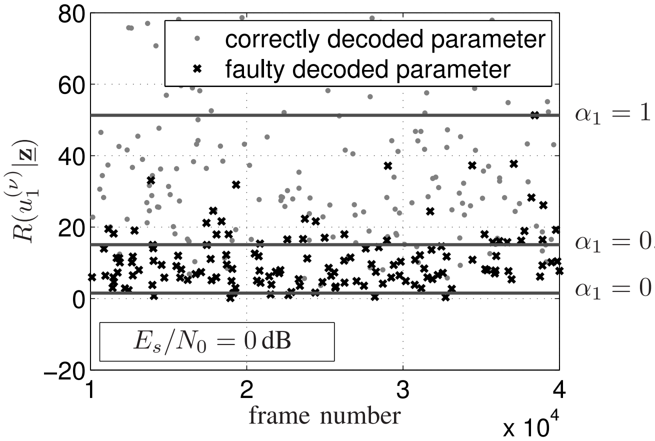

Fig. 3. Identification of faulty decoded parameters by the threshold $\Gamma_{\text{opt},1}^{[j=0\,\text{dB}]}(\alpha_1)$ for the AWGN channel at 0 dB. The control factor α_1 can be varied between 0 and 1 and is set to 0, 0.5 and 1 for illustration.

Considering Fig. 3 which is a detail of Fig. 2(b), it is obvious that there usually does not exist an optimal decision threshold which classifies correctly decoded and faulty decoded parameters without any decision errors. However, it is possible to determine optimal decision thresholds $\Gamma_{\text{opt},\kappa}^{[j]}(\alpha_\kappa)$ individually for each parameter with index κ and each channel quality j. They are optimal in the sense that they minimize specific error probability functions as mathematically expressed in what follows. The meaning of the weighting factor α_1 in Fig. 3 and its influence on the threshold $\Gamma_{\text{opt},1}^{[j=0\,\text{dB}]}(\alpha_1)$ will also be clarified in what follows.

Exploiting the parameter-individually optimized decision thresholds, a *bad parameter indication* (BPI) flag Θ_κ is defined for each codec parameter according to

$$\Theta_\kappa = \begin{cases} 1 \quad , & R(u_\kappa^{(\nu)}|\underline{\mathbf{z}}) \leq \Gamma_{\text{opt},\kappa}^{[j]}(\alpha_\kappa) \\ 0 \quad , & R(u_\kappa^{(\nu)}|\underline{\mathbf{z}}) > \Gamma_{\text{opt},\kappa}^{[j]}(\alpha_\kappa) \end{cases} , \tag{4}$$

which can control a parameter-individual concealment routine. This routine has to be applied before signal synthesis.

The parameter-specific optimal decision threshold $\Gamma_{\text{opt},\kappa}^{[j]}(\alpha_\kappa)$ for a system with receiver *channel state information* (CSI) can be computed in advance for each E_s/N_0 individually by solving the minimization problem for a sufficiently large set of training signals:

$$\Gamma_{\text{opt},\kappa}^{[j]}(\alpha_\kappa) = \arg\min_{\Gamma}\Big\{P_{\text{err},\kappa}^{[j]}(\alpha_\kappa,\Gamma)\Big\}, \forall j \in \mathbb{J}, \Gamma \in \mathbb{G}, \tag{5}$$

$$\begin{aligned} P_{\text{err},\kappa}^{[j]}(\alpha_\kappa,\Gamma) = {} & \alpha_\kappa \cdot P^{[j]}(\Theta_\kappa = 0, \hat{u}_\kappa \neq u_\kappa|\Gamma) \\ & + (1-\alpha_\kappa)\cdot P^{[j]}(\Theta_\kappa = 1, \hat{u}_\kappa = u_\kappa|\Gamma), \end{aligned} \tag{6}$$

where $P^{[j]}(\Theta_\kappa = 0, \hat{u}_\kappa \neq u_\kappa|\Gamma)$ denotes the false positive probability at one distinct E_s/N_0 value $j \in \mathbb{J}$ meaning that the κ-th parameter is incorrectly assumed to be error-free decoded given a decision threshold Γ. $\mathbb{J}$ and $\mathbb{G}$ are the sets of all considered channel qualities and thresholds, respectively. $P^{[j]}(\Theta_\kappa = 1, \hat{u}_\kappa = u_\kappa|\Gamma)$ is the false negative probability. The parameter-specific weighting factor α_κ controls the influence of both probabilities on the minimization result. This is essential in order to adjust the decision for parameters of different importance. A factor of $\alpha_\kappa = 0$ minimizes the probability that correctly decoded parameters are classified as being erroneously (false negative). This might be a suitable choice for parameters with little impact on the instantaneous signal quality. In order to guarantee for a minimum number of faulty decoded parameters used for signal synthesis (false positive), a factor of $\alpha_\kappa = 1$ should be selected. However, the number of correctly decoded parameters which are discarded will increase in this case. The minimum total error probability $P^{[j]}(\Theta_\kappa = 0, \hat{u}_\kappa \neq u_\kappa|\Gamma) + P^{[j]}(\Theta_\kappa = 1, \hat{u}_\kappa = u_\kappa|\Gamma)$ is obtained by a control factor $\alpha_\kappa = 0.5$. It is worth noting, that an optimal system performance in terms of perceptual multimedia quality implies an optimally tuned control factor for each codec parameter individually. All three special cases are shown for the AMR example in Fig. 3. The optimal choice lies somewhere in between and depends on the impact of faulty decoded LSF1s on the speech quality.

Note, that the bad frame indication method proposed above induces an extremely low additional complexity. This is an important property for real-time applications. Almost all computational complexity is concentrated in the determination of suitable decision thresholds done in advance by the full search algorithm given in (6).

IV. Evaluation

In order to show the effectiveness of the BPI in conjunction with a parameter-individual error concealment, we have compared three ISCD systems performing AMR speech transmission (12.2 kbit/s) and error concealment. The first one applies only the MAP estimation (MAP system), the second one performs additionally the BFI controlled standardized AMR frame error concealment [7] (BFI system) and the third one a parameter-individual concealment based on the BPI flag defined in (4) for $\alpha_\kappa = 0.1, \forall\kappa$ (BPI system). This value for α_κ results from simulations and secures a very robust identification of faulty decoded parameters. The BPI system conceals all parameters which are indicated by the BPI using the subroutines of the AMR concealment separately as stand-alone parameter concealment routines. The AMR frame error concealment substitutes erroneous frames based on a state machine with seven states which comprises the previous frame into the concealment procedure. The underlying parameter-specific concealment can be summarized as follows:

1) The LSFs of the previous frame are shifted towards their mean before being applied for substitution,
2) the pitch gain and the code book gain are limited, reduced or muted,
3) the pitch delay values are replaced by the past value from the 4th subframe of the previous frame and
4) the received fixed codebook indices from the erroneous frame are used as being received (no concealment).

Note, that the modality of the concealment depends on the state of the state machine and does not necessarily comprise

TABLE I
REDUNDANT BIT MAPPINGS APPLIED TO THE AMR-NB CODEC AT 12.2 KBIT/S.

AMR-NB Mode	Parameter	total bits per frame uncoded / after RBM
12.2 kbit/s	5 LSFs	38 / 195
	Pitch delay	30 / 120
	Pitch gain	16 / 128
	Fixed code book index	140 / 200
	Code book gain	20 / 128
	Total	**244 / 771**

all steps. Exploiting the BPI flag, it is possible to establish state machines for each parameter type individually. In this case, if a faulty code book gain is detected, only this gain has to be adapted leaving all other parameters unaffected. It is worth noting, that one faulty decoded LSF enforces the substitution of all LSFs in order to avoid filter instabilities. Since no concealment is performed for faulty decoded fixed code book indices, a BPI flag is only defined for the rest of the parameters.

The basic ISCD system is equally configured in all cases. Unequal error protection is realized by parameter-individual RBMs which apply rate-flexible linear block codes for each parameter $\kappa = 1 \ldots K$. In this paper, we have used the rate-constrained design specification given in [11] for the design of these codes. The advantage is that the parity bits can easily be generated using simple binary operations and the generator matrices do not need to be explicitly stored. The parameter-individual code rates of the RBMs for a 20 ms speech frame are heuristically optimized and summarized in Table I. A convolutional IIR rate-1 channel encoder with octal generator polynomial $\mathbf{G} = \left(\frac{10}{17}\right)_8$ and constraint length $K = 4$ is used as inner channel encoder. That leads to a total number of $771 + 3$ tail bits per frame and thus to a gross bit rate of 38.7 kbit/s which corresponds to the typical bit rate of the UMTS standard for 12.2 kbit/s speech before rate matching to code rate 1/3.

In a first evaluation, the indication error rates induced by the BPI system are compared to the parameter error rates (PER) of the MAP and the BPI system. A sequence of 400000 AMR speech frames is transmitted over an AWGN channel with power spectral density $\sigma_n^2 = N_0/2$. All systems perform 10 ISCD iterations. The BFI system ($P(\underline{\hat{u}} \neq \underline{u})$) generates the highest error rate due to the rejection of complete erroneous frames irrespective of their number of faulty decoded parameters. The PER of the MAP system ($P(\hat{u} \neq u)$) is in the order of the total indication error rate of the BPI system (P_{err}). The big advantage of the novel BPI system is the extremely low false positive probability P_{fp} which results in a robust rejection of faulty decoded parameters ($P_{\text{fp}} \ll P(\hat{u} \neq u)$) avoiding highly audible artifacts. The quite high false negative probability $P_{\text{fn}} = P_{\text{err}} - P_{\text{fp}} \approx P_{\text{err}}$ inducing unnecessary parameter concealments is not that critical (less audible) due to the applied error concealment. For bad channel qualities this method converges to the BFI method in terms of speech quality, due to the high number of distorted parameters per frame. Note, that this behavior is achieved without additional data rate for error detection (CRC). For good channel qualities, only the discovered fraction of distorted parameters and the false negative part of the parameters have to be concealed. Consequently, the rate of concealed parameters approximately results in a total of two times the PER of the MAP system. Nevertheless, this number is still much lower than the number of concealed parameters in the BFI case. Furthermore, it could be confirmed by informal listening tests that the perceptual speech quality provided by the parameter-individual concealment outperforms the speech quality delivered by the MAP system, although the total error rate of the BPI system is twice as high after concealment. This is due to the possibly deep impact of distorted parameters on the synthezised speech quality which is avoided by the BPI system as we have demonstrated by an informal listening test.

This listening test was based on the transmission of 20 different speech files (8 s each) over an AWGN channel now performing only 2 ISCD iterations at the receiver. The reduced number of iterations entails a flatter turbo cliff in the E_s/N_0 range between 1.5 dB and 3 dB which simplifies the highlighting of differences between all systems. The channel quality E_s/N_0 was varied between 2.3 dB and 2.8 dB (upper waterfall region). The speech quality provided by all methods was scored between 1 (poor) and 5 (excellent) by each of the 30 listeners.

Figure 5 shows the results of the informal listening test. It indicates that most of the listeners prefer the BPI concealment (75 %), while the BFI concealment is only prefered in approximately 8 % of the test cases. For the evaluated scenarios (upper waterfall, i.e., low number of corrupted parameters), even the MAP system has reached 17 %. Note that this will surely not be the case for bad channel qualities. In this case, the MAP system might cause higly audible artifacts due to the increasing number of distorted parameters. Then it might be more expedient in terms of perceptual signal quality to substitute erroneous frames by their concealed versions instead of using all distorted parameters for signal syntheses. Nevertheless, best performance can be expected by the BPI system for all channel qualities.

V. CONCLUSION

It has been shown that the parameter-related logarithmic *a posteriori* reliabilities enables a robust identification of erroneous codec parameters and, consequently, the application of parameter-individual concealment routines. This is achieved with low complexity and no additional data rate for CRCs. The proposed method provides a highly enhanced multimedia quality for wireless communication especially for low channel qualities, i.e., within the waterfall region.

REFERENCES

[1] M. Adrat, P. Vary, and J. Spittka, “Iterative Source-Channel Decoder using Extrinsic Information from Softbit-Source Decoding,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, Salt Lake City, Utah (USA), 2001, pp. 2653–2656.

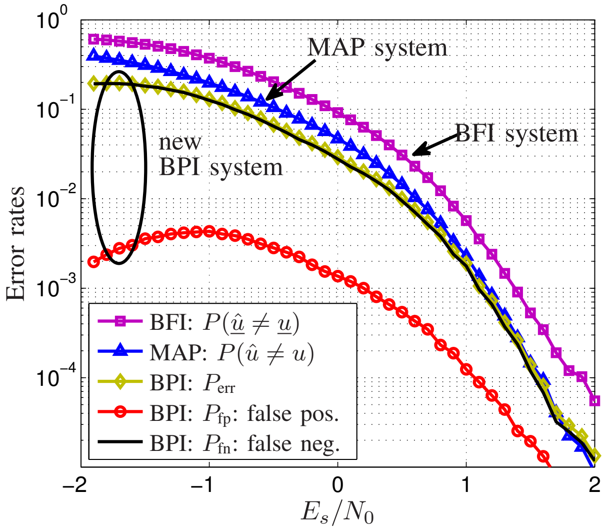

Fig. 4. Error rates of the BPI, BFI and MAP system (10 ISCD iterations). AMR speech at 12.2 kbit/s is transmitted over an AWGN channel.

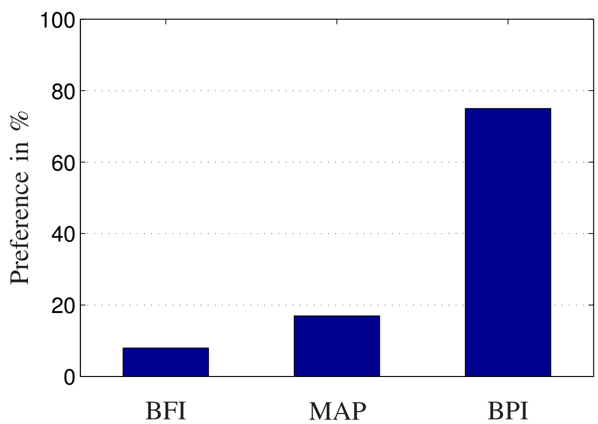

Fig. 5. Preference of the different concealment methods in %. E_s/N_0 is varied between 2.3 dB and 2.8 dB (upper waterfall region). Two ISCD iterations are carried out. The speech quality evaluation is done by 30 listeners.

[2] N. Görtz, "Iterative Source-Channel Decoding Using Soft-In/Soft-Out Decoders," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 173.

[3] M. Adrat, T. Clevorn, and L. Schmalen, "Iterative Source-Channel Decoding & Turbo DeCodulation," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. John Wiley & Sons, Ltd, 2008, ch. 13, pp. 365–398.

[4] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 240–251, March 2001.

[5] N. Othman, M. El-Hajjar, O. Alamri, and L. Hanzo, "Soft-Bit Assisted Iterative AMR-WB Source-Decoding and Turbo-Detection of Channel-Coded Differential Space-Time Spreading Using Sphere Packing Modulation," in *Proc. 65th IEEE Vehicular Technology Conference (VTC-Spring)*, 2007, pp. 2010–2014.

[6] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate (Corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, March 1974.

[7] ETSI TS 126 091 V7.0.0 (2007-06), *Digital Cellular Telecommunications System (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR Speech Codec; Error Concealment of Lost Frames (3GPP TS 26.091 Version 7.0.0 Release 7)*, June 2007.

[8] T. Clevorn, P. Vary, and M. Adrat, "Iterative Source-Channel Decoding Using Short Block Codes," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, 2006, pp. 221–224.

[9] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic Information Transfer Functions: A Model and Two Properties," in *Proc. Conference on Information Sciences and Systems (CISS)*, Princeton, New Jersey (USA), March 2002.

[10] ETSI EN 301 704 V7.2.1 (2000-04), *Digital Cellular Telecommunications System (Phase 2+); Adaptive Multi-Rate (AMR) Speech Transcoding (GSM 06.90 Version 7.2.1 Release 1998)*, April 2000.

[11] T. Breddermann, H. Lüders, P. Vary, I. Aktas, and F. Schmidt, "Iterative Source-Channel Decoding with Cross-Layer Support for Wireless VoIP," in *Proc. International ITG Conference on Source and Channel Coding*, Siegen, Germany, January 2010.