ROBUST SPEECH TRANSMISSION OVER NOISY CHANNELS EMPLOYING NON-LINEAR BLOCK CODES

S. Heinen, S. Bleck, P. Vary

Institute of Communication Systems and Data Processing
Aachen University of Technology
heinen@ind.rwth-aachen.de
vary@ind.rwth-aachen.de

ABSTRACT

In medium to low bit rate speech codecs the speech signal is represented by a set of parameters. The most important concept is presently *Code Excited Linear Predictive* (CELP) coding. A speech segment of typically 10 to 20 ms is described in terms of prediction coefficients, gain factors and excitation vectors. Due to the high compression rates (0.5-1.5 bits per speech sample) the parameters are partly highly sensitive against channel noise.

In this paper we present a new error protection technique, that is based on a joint optimization of parameter quantization and a redundant non-linear block coding scheme. For parameter reconstruction, the principle of soft bit source decoding is applied. The proposed technique can be used in combination with conventional error protection such as convolutional coding and allows a flexible subdivision of the gross data rate for source coding and error protection.

1. 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 contrary 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, channel coding and equalization.

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) is used, i.e. significant bits of a coded speech segment are highly protected, whereas less significant bits have weaker error protection.

We present a new, parameter-oriented, non-linear precoding technique, combined with a soft bit parameter estimation unit at the receiving point [1]. By this an implicit UEP for the bits of a source parameter is achieved, such that the commonly observed MSB-LSB characteristic of the parameter bits is removed and the mean parameter distortion due to the erroneous transmission is minimized. In contrast to *Channel Optimized Vector Quantization* (COVQ) [2], where the parameter quantizer itself is adapted to the noisy channel, the presented precoding technique operates with a source optimized quantizer.

A further advantage of the presented non-linear precoding technique is its high flexibility in terms of partitioning a given gross data rate for source coding and error protection. Here it is even possible to implement non-integer source coding rates while fully exploiting the fractional part to increase the error robustness. This flexibility makes non-linear precoding attractive for AMR systems where the partitioning into the source and the channel coding rate is performed adaptively during operation.

The paper is structured as follows: First, we give a brief review of soft bit source decoding. In section 3 we present the principle of linear precoding [3, 4] which is generalized to the new non-linear precoding scheme in section 4. Simulation results will be given in section 5.

2. SOFT BIT SOURCE DECODING

Recently [5, 1] parameter estimation was proposed to realize soft bit source decoding. This approach becomes especially attractive when used together with efficient soft output decoders for convolutional codes [6, 7].
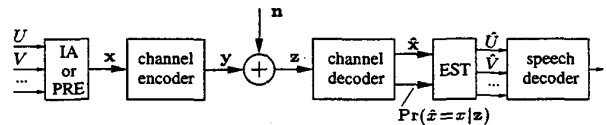

Figure 1: Transmission system
(IA: index assignment, PRE: precoding, EST: estimation)

Fig. 1 depicts a model of the transmission system. In the *index assignment* (IA) block to each of the (vector-)quantized CELP parameters $U, V, ..$, a value-specific bit pattern is assigned, e.g. parameter U is encoded by a group 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 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 $\Pr(\hat{x}=x|\mathbf{z})$ for single bits. This information enables the estimation unit to compute parameter estimates $\hat{U}, \hat{V}, ...$

Parameter estimation requires the computation of parameter *a posteriori* probabilities from the 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

---

[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.

parameter bits u_i we get

$$\Pr(\hat{U}=U|\mathbf{z}) \approx C \cdot p_U(\hat{U}) \cdot \prod_{i=1}^{M} \frac{\Pr(\hat{u}_i=u_i|\mathbf{z})}{p_{u_i}(\hat{u}_i)}, \qquad (1)$$

where $p_U(\cdot)$ is the *a priori* distribution of the speech parameter and $p_{u_i}(\cdot)$ the *a priori* distribution of its i-th bit. The constant C is determined by the condition $\sum_{\hat{U}} \Pr(\hat{U}=U|\mathbf{z}) = 1$. The computation of the *a posteriori* probabilities can also include the time correlation of parameter values of successive speech segments. In this case the process U must be modelled by a Markov chain rather than by a simple pdf [1].

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

$$\hat{U}_{MAP} = \arg \max_{\hat{U}} \Pr(\hat{U}=U|\mathbf{z}). \qquad (2)$$

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

$$\hat{U}_{MMSE} = E\left\{\hat{U}|\mathbf{z}\right\} = \sum_{\hat{U}} \hat{U} \cdot \Pr(\hat{U}=U|\mathbf{z}). \qquad (3)$$

The estimated parameter value can directly be processed by the speech decoder. In applications where channel decoder and speech decoder are at spatially distant locations, a problem may occur due to the increased data rate that has to be transmitted to transport the reliability information to the speech decoder. This can easily be solved by moving the estimator from the speech decoder site to the channel decoder site. Then the estimated parameter values $\hat{U}, \hat{V}, ..$ have to be re-quantized. To avoid performance degradation, the quantizer codebook must contain all original centroids of the corresponding parameter quantizer in the speech encoder. In most cases a one-to-one copy of this initial quantizer is sufficient for the re-quantization.

3. LINEAR PRECODING

The estimation gain of the approach presented in the last section can significantly be improved by redundant precoding. Linear precoding was recently presented in [3, 4]. In linear precoding an adequate linear code is used at the transmitter site to generate parity bits for the parameter bits u_i. A simple example for such a code is the *Single Parity Check* (SPC) code, which adds a single parity bit u_p, computed as the modulo 2 sum ($\oplus$) of all parameter bits: $u_p = \sum_{\oplus i=1}^{M} u_i$. The parity bit is transmitted as an extra parameter bit. At the receiver, it is not used for conventional error detection, but as additional soft information to enhance the estimation of the parameter *a posteriori* probabilities:

$$\Pr(\hat{U}=U|\mathbf{z}) \approx C \cdot p_U(\hat{U}) \cdot \frac{\Pr(\hat{u}_p=u_p|\mathbf{z})}{p_{u_p}(\hat{u}_p)} \cdot \prod_{i=1}^{M} \frac{\Pr(\hat{u}_i=u_i|\mathbf{z})}{p_{u_i}(\hat{u}_i)}. \qquad (4)$$

The enhancement mechanism can intuitively be understood by noting that the term $\Pr(\hat{u}_p=u_p|\mathbf{z})$ is close to one if the received parity bit fits to a certain parameter code and approaches zero if it does not. Thus the *a posteriori* probabilities of parameter values with correct parity bits are increased while the others are diminished. The extension of this linear precoding technique to more complex linear codes is straightforward.

It should be noted, that in a system with a constant gross data rate, the error protection by subsequent channel coding becomes weaker, as the additional parity bits increase the net rate before channel encoding. We implemented the complete transmission system with linear precoding for the *Line Spectrum Frequencies*

(LSF) codebook indices of a *Variable Rate* (VR)-CELP [4]. Informal listening tests proved that the degradation due to weaker error protection is clearly over-compensated by the gain attained by linear precoding.

4. NON-LINEAR PRECODING

Non-linear precoding may be seen as a generalization of linear precoding with the aim to adapt the code to both the parameter generating process and the channel characteristics with respect to a maximum end-to-end parameter SNR. While in linear precoding the valid codewords[2] are determined by the parity equations of the linear code (e.g. for the SPC only codewords with even number of ones), we will now allow any subset consisting of Q codewords taken from the set of 2^N N-bit patterns to be a potential code, where Q needs no longer to be a power of two. In this case we obtain in general a code which does not fulfill the algebraic linearity condition any more. The rate of this code consequently is $\frac{\mathrm{ld}(Q)}{N}$.

The code design has to solve the following task: Select a subset of Q codewords from the set of 2^N bit patterns and assign them to the possible parameter values. Both design steps, selection and assignment, must be performed such, that the end-to-end parameter distortion is minimized, i.e.

$$E\left\{|\hat{U}_{opt} - U|^2\right\} = \sum_{U} \int |\hat{U}_{opt} - U|^2 p(U, \mathbf{z}) \, d\mathbf{z} \doteq \min. \qquad (5)$$

This is a non-trivial task due to the immense amount of possible variations; in fact there are $\frac{2^{N!}}{Q!}$ possibilities. If we want to precode e.g. a 4 bit-parameter with a 7 bit-code, about 10^{200} assignments are possible. This simple example shows, that a full search for the optimal assignment is not feasible. However, the design task can be interpreted as a generalized *Index Assignment* (IA) problem. Therefore, it is possible to adapt the algorithms originally developed for the IA problem [8, 9] to the precoder design task. The basic idea of these algorithms is, starting from an initial codeword assignment, to select two parameter values and interchange their dedicated codewords. If this modified IA yields a reduced cost value (5), it is accepted, otherwise the previous one is recovered and the interchanging of two other codeword assignments is checked. This procedure is repeated until a local minimum of the cost function is reached. The resulting code will, in general, be non-linear.

Provided that the error events affecting single codeword bits are statistically independent, the joint pdf $p(U, \mathbf{z})$ in (5), which is often difficult to compute if channel coding is involved, can be approximated by the error probability of a simple *Binary Symmetric Channel* (BSC) model. Although being a rough approximation, our experiments showed, that it gives nearly as good results as a complex description. With p_e denoting the error probability of the BSC, $\mathbf{u}$ the codeword assigned to the actual parameter value U and $\hat{\mathbf{u}}$ the received codeword yielding the estimate[3] $\hat{U}$, we get the simplified cost function

$$\sum_{U} \sum_{\hat{U}} |\hat{U} - U|^2 p_U(U) \, p_e^{|\mathbf{u} \oplus \hat{\mathbf{u}}|} \, (1 - p_e)^{N - |\mathbf{u} \oplus \hat{\mathbf{u}}|}. \qquad (6)$$

The notation $|\mathbf{u} \oplus \hat{\mathbf{u}}|$ represents the Hamming distance of two codewords.

---

[2]In the sequel, we will use the term codeword for the bit pattern that is assigned to a parameter value by the precoding process

[3]As the BSC generates a countable number of different error events on the N code bits, the number of possible estimates in this case is also countable.

5. SIMULATIONS

To show the performance of the new precoding scheme we perform two experiments. In both experiments the source parameter is modelled by a white, Gaussian distributed process U with variance $\sigma_U^2 = 1$. In the first experiment we compare linear precoding to non-linear precoding. The parameter U is quantized by a *Lloyd-Max Quantizer* (LMQ) with $Q = 8$ quantization centroids, thus $M = 3$. For linear precoding we use an SPC and for non-linear precoding a rate $\frac{3}{4}$ non-linear code optimized at $p_e = 0.08$. In both cases blocks of 24 precoded parameters are channel encoded with a rate $\frac{2}{5}$ RCPC code [10] and transmitted over an AWGN channel for various values of E_s/N_0. The disturbed bit stream is soft decoded by a symbol-by-symbol MAP decoder [6] and the parameter values are recovered by estimation according to eq. (3).

Fig. 2 shows the performance in terms of end-to-end parameter SNR for linear precoding (dashed line) and non-linear precoding (solid line). While the linear precoding approach exhibits a slightly better SNR at good channel conditions it is outperformed by non-linear precoding at medium channel conditions by up to 0.8 dB parameter SNR.
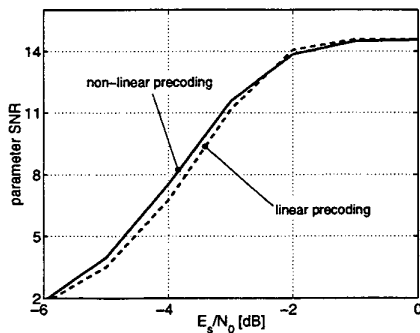

Figure 2: End-to-end SNR for linear and non-linear precoding

In the second experiment, non-linear precoding in combination with constant rate convolutional coding is compared to a UEP system. For the precoding approach we drop the restriction that $M = \mathrm{ld}(Q)$ has to be an integer value by applying an LMQ, whose number of centroids is not a power of two; actually we set $Q = 10$. Non-linear precoding at rate $\frac{\mathrm{ld}(10)}{4}$ fills up the fractional source coding rate to 4 bit/parameter. As in the first experiment, the precoded parameters are additionally protected by a $\frac{2}{5}$ convolutional code. In comparison, we consider an RCPC-UEP scheme with a 16 level LMQ and three different channel coding rates $\frac{1}{3}$ (MSB), $\frac{2}{5}$ (bit 2 and 1) and $\frac{1}{2}$ (LSB), thus achieving the same gross bit rate as in the system with precoding.

Fig. 3 depicts the simulation results. It can clearly be seen, that by non-linear precoding it is possible to compromise between quantization exactness and channel robustness in a very efficient manner. The advantage of this concept can be seen regarding the following example: Assume that due to system constraints it is sufficient to provide an end-to-end parameter SNR of 15 dB (dashed horizontal line in Fig. 3) for the parameter U. Then the non-linear precoding (solid line) preserves the required SNR until $E_s/N_0 = -2.1$ dB, while UEP with hard source decoding (dash-dotted line) fails at $E_s/N_0 = -0.4$ dB. Thus, in terms of E_s/N_0, a gain of 1.7 dB due to non-linear precoding is possible. Even compared to UEP in combination with soft source decoding according to eq. 3 (dotted line), a gain of 1.3 dB is obtained.
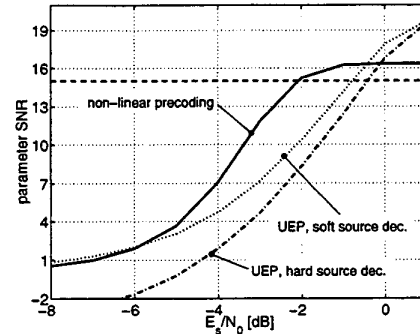

Figure 3: End-to-end SNR for non-linear precoding and for UEP

6. CONCLUSIONS

We presented a robust scheme for the transmission of source parameters over a noisy channel, which is based on soft bit source decoding and a new non-linear precoding technique. The approach can efficiently be used in combination with conventional FEC. Simulations show, that the new non-linear scheme outperforms linear precoding. In comparison to UEP with hard source decoding, a gain of 1.7 dB in terms of E_s/N_0 could be obtained. Non-linear precoding allows a very flexible exchange of error robustness and quantiztion precision due to its ability to exploit non-integer bit rates for error protection without entropy encoding. This makes it also a dedicated means to realize AMR systems.

REFERENCES

[1] T. Fingscheidt and P. Vary, "Robust Speech Decoding: A Universal Approach to Bit Error Concealment," in *Proceedings ICASSP*, vol. 3, pp. 1667–1670, Apr. 1997.

[2] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Trans. Information Theory*, vol. 36, pp. 799–809, July 1990.

[3] N. Görtz, "Joint Source Channel Decoding Using Bit-Reliability Information and Source Statistics," in *Proc. Int. Symp. on Information Theory, ISIT*, p. 9, IEEE, Aug. 1998.

[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*, (Phoenix, Arizona), Mar. 1999.

[5] C. Gerlach, "A Probabilistic Framework For Optimum Speech Extrapolation in Digital Mobile Radio," in *Proceedings ICASSP*, (Minneapolis), pp. II.419–II.422, Apr. 1993.

[6] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Information Theory*, pp. 284–287, Mar. 1974.

[7] J. Hagenauer and P. Höher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," *Proceedings GlobeCom*, pp. 1680–1686, Nov. 1989.

[8] K. Zeger and A. Gersho, "Pseudo-Gray Coding," *IEEE Trans. Communications*, vol. 38, pp. 2147–2158, Dec. 1990.

[9] P. Knagenhjelm and E. Agrell, "The Hadamard Transform - A Tool for Index Assignment," *IEEE Trans. Information Theory*, vol. 42, pp. 1139–1151, July 1996.

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