JOINT SOURCE-CHANNEL DECODING BY CHANNEL-CODED OPTIMAL ESTIMATION (CCOE) FOR A CELP SPEECH CODEC

Norbert Görtz

Institute for Circuits and Systems Theory
University of Kiel
Kaiserstraße 2, D-24143 Kiel / Germany
Tel.: +49 431 77572 406, Fax: +49 431 77572 403
E-Mail: ng@techfak.uni-kiel.de

ABSTRACT

A new algorithm called Channel-Coded Optimal Estimation (CCOE) is adapted to a CELP speech codec. The CCOE-algorithm performs joint source-channel decoding by optimal estimation using bit-reliability informations (soft-bits) and source statistics. The basic algorithm for a simple transmission scenario with a single scalar quantizer has been recently stated by the author. In this paper the basic idea is extended for the use in a CELP speech-codec that was developed for enhanced speech transmission in the GSM mobile radio channel. Simulation results based on informal listening tests are given which show that the new algorithm achieves substantial gains over separate error correction with hard decisions prior to source decoding and classical techniques for error detection and bad frame handling as widely used in the present mobile-radio systems.

1. INTRODUCTION

In mobile-radio systems the speech quality at the output of the receiver suffers from strong distortions due to noise on the channel even though channel coding is applied to correct bit errors prior to speech decoding. Many of the remaining bit errors are detected and concealed by bad-frame handling techniques. Still strong distortions occur which is partly due to separate source and channel decoding with hard decisions for the data-bits after channel decoding. New algorithms that exploit bit-reliability informations and source statistics for joint source channel decoding are able to improve the speech quality.

The paper is organized as follows: In section 2 the basic principle of Channel-Coded Optimal Estimation is restated and simulation results for a simple transmission system are given.

In section 3 a CELP codec developed for enhanced speech transmission in the GSM-system is described.

In section 4 the time-based dependencies of the parameters of the CELP speech codec are measured and modeled by first-order Markov-sources such that Channel-Coded Optimal Estimation is applicable.

In section 5 the performance of the CCOE-algorithm applied to the CELP speech codec is compared with presently used algorithms, e.g. error detection and parameter extrapolation.

2. PRINCIPLE OF CHANNEL-CODED OPTIMAL ESTIMATION (CCOE)

Channel-Coded Optimal Estimation (CCOE) is a technique for joint source channel *de*coding [5]. Bit-reliability informations of the received source-encoder bits and the redundancy bits of a channel code are combined with the source-signal statistics to calculate the a-posteriori probabilities for all possibly transmitted codewords. The probabilities are used to estimate the transmitted signal at the receiver.

2.1. Formulation as Estimation Problem

Figure 1 shows a model of a transmission system. The input signal $u(k)$ (e.g. a parameter of speech-codec or a waveform sample) is source-encoded, and the result-

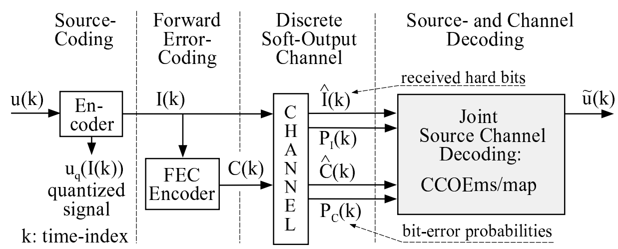

Figure 1: Simplified model of a transmission system

ing N_I bits are grouped into the bit-vector $I(k)$. Any systematic[1] channel-code (FEC) can be used to calculate the redundancy bit-vector $C(k)$. The codeword is $V(k) = \{I(k), C(k)\}$. A hypothesis for the codeword at the receiver is denoted by $V^{(\lambda)}(k) = \{I(k) = \lambda, C^{(\lambda)}(k)\}$ where $\lambda \in \{0, 1, \ldots, 2^{N_I} - 1\}$ is the index corresponding to the hypothesized bit-vector, and $C^{(\lambda)}(k)$ is the vector of redundant bits which is calculated from the bit-vector $I(k) = \lambda$ by a channel encoder located at the receiver. The a-posteriori probabilities

$$P_{AP}\big(\lambda, k\big) \doteq P\Big(V^{(\lambda)}(k) \mid \hat{V}(k), \hat{V}(k-1), \ldots\Big) \qquad (1)$$

of all possibly transmitted codewords $V^{(\lambda)}(k)$ conditioned on all received bit-vectors $\hat{V}(k), \hat{V}(k-1), \ldots$ up to the current time-instant are required to calculate an estimation $\tilde{u}(k)$ for the transmitted signal $u(k)$. For a maximum a-posteriori estimator (MAP) the index λ is searched for which $P_{AP}\big(\lambda, k\big)$ is maximum. The mean square (MS) estimator will possibly yield better results [4] especially if the coded signal $u(k)$ represents a "waveform". The MS estimator is given by

$$\tilde{u}(k) = \sum_{\lambda=0}^{2^{N_I}-1} \underbrace{u_q\big(I(k) = \lambda\big)}_{\text{quant. levels}} \cdot P_{AP}\big(\lambda, k\big) \quad . \qquad (2)$$

[1] systematic channel codes are only chosen for convenience of notation

2.2. Calculation of the a-posteriori probabilities

In [4] a recursive formulation for the calculation of the a-posteriori probabilities is given for a source modeled by a first-order Markov-process, if no channel coding (i.e. $V(k) = I(k)$) is applied (OE-technique). It was stated in [5] that a formally similar solution holds if channel coding *is* applied (CCOE-technique). Then the a-posteriori probabilities are given by

$$P_{AP}\big(\lambda, k\big) \quad = \quad M \;\cdot$$

$$\underbrace{\overbrace{P\Big(\hat{C}(k) \mid C^{(\lambda)}(k)\Big)}^{\text{see fig. 1: calc. from } P_C(k)} \;\cdot\; \overbrace{P\Big(\hat{I}(k) \mid I(k) = \lambda\Big)}^{\text{calculated from } P_I(k)}}_{\text{channel-probability of hypothesized index } \lambda} \;\cdot$$

$$\sum_{\nu=0}^{2^{N_I}-1} \underbrace{P\Big(I(k) = \lambda \mid I(k-1) = \nu\Big)}_{\text{source-index transition-probabilities}} \cdot P_{AP}\big(\nu, k-1\big)\;. \quad (3)$$

The constant M normalizes the sum of the $P_{AP}\big(\lambda, k\big)$ over $\lambda = 0, 1, ..., 2^{N_I} - 1$ to 1. The channel noise is assumed to be "white" and independent of the coded bits.

The source-index transition probabilities $P\big(I(k)=\lambda \mid I(k-1)=\nu\big)$ in (3) can be measured and stored in advance since the Markov-source is assumed to be time-invariant. The recursion is initialized by the unconditioned probability distribution of the source-encoder indices, i.e. $P_{AP}(\lambda, 0) = P(I = \lambda)$.

The channel-dependent terms in (3) are calculated by the following equations:

$$P\big(\hat{I}_m(k) \mid I_m(k) = \lambda_m\big) = \begin{cases} 1 - P_{I,m}(k)\,, & \hat{I}_m(k) = \lambda_m \\ P_{I,m}(k)\,, & \hat{I}_m(k) \neq \lambda_m \end{cases}$$

$$P\Big(\hat{I}(k) \mid I(k) = \lambda\Big) = \prod_{m=1}^{N_I} P\big(\hat{I}_m(k) \mid I_m(k) = \lambda_m\big) \quad (4)$$

The vector $P_I(k)$ contains the bit-error probabilities $P_{I,m}(k)$ of the N_I hard-decided bits (indexed by m), which are issued by the discrete soft-output channel.

2.3. Simulations and Performance

A Gaussian random signal was correlated by a low-pass filter with the transfer function $H(z) = \frac{(1-a)\cdot z}{z-a}$, $a = 0.9$, to generate the signal $u(k)$. A 5-bit optimal quantizer was used as a source-encoder and a 1-bit parity-check as a channel code. The codewords were transmitted over a discrete AWGN-channel and the OE-technique and the CCOE-technique were applied, both with MAP- and MS-estimators. The CCms/map-techniques can be derived from CCOEms/map by *not* exploiting the source-statistics, i.e. by setting

$$P\Big(I(k)=\lambda \mid I(k-1)=\nu\Big) = P\big(I=\lambda\big) = \frac{1}{2^{N_I}}\ \forall\,\lambda\,. \quad (5)$$

The CCmap-technique is equivalent to classical maximum-likelihood channel decoding.

For each algorithm the SNR-value $(u(k), \tilde{u}(k))$ over E_b/N_0, the ratio of the energy E_b per transmitted *data*-bit and twice the noise power spectral density $N_0/2$ on the AWGN-channel, is plotted in figure 2. For the CC-algorithms and the CCOE-techniques a rate correction was carried out by adding $10\log_{10}(\frac{1}{R}) = 0.79$ dB to the E_b/N_0-values to allow a fair comparison (with respect to the transmission power) of the algorithms that use the parity bit and those which do not.

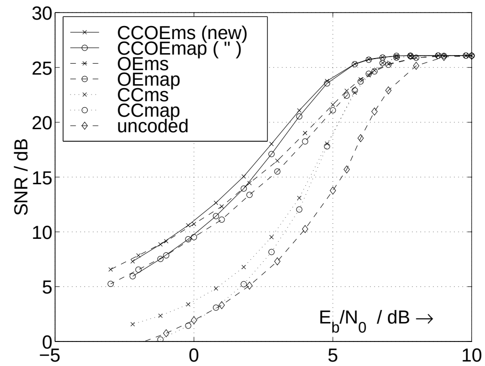

Figure 2: Performance of the transmission system for a strongly correlated source signal ($a = 0.9$) coded by a 5-bit optimal quantizer. Channel coding is realized by a single-bit parity check (code rate R=5/6). The coded bits are transmitted over an AWGN-channel with coherently detected BPSK-modulation

Figure 2 shows that the CCOE techniques work as good as or better than all other algorithms for all channel conditions if the same type of estimator is used.

For "good" channels the CCOE-algorithms rely on the parity bit, i.e. channel errors are corrected mainly by the channel code. Therefore the performance of CCOEms/map is similar to CCms/map, and it is slightly superior to the OE-algorithms that do not exploit channel coding.

If the channel quality is bad, the CCOE-algorithms put more weight on the information that is based on the correlation of the source-signal. Consequently the performance of CCOE is similar to that of the OE-algorithms. The quality is much better than for the CC-algorithms that do not exploit the source statistics.

The strongest gains of the CCOE-algorithms over the others are achieved for moderately corrupted channels (E_b/N_0-values around 3..6 dB).

All algorithms work better and mostly much better compared to a system where only hard decisions are taken for the data bits at the channel output (curve labeled "uncoded"). Qualitatively, the results hold for source signals with moderate and low correlations too.

3. CELP SPEECH-CODEC

The CELP encoder that is depicted in figure 3 processes frames of 160 narrow-band speech samples (20 ms), which are divided into 4 subframes. The total number of bits per frame is 214, i.e. the bit rate is 10.7 kbps.

3.1. Frame Processing

First, the mean power of the signal is calculated. It is logarithmically quantized by a 5-bit table, similar to [1]. Then a 10^{th}-order LPC analysis is carried out. The LPC coefficients are converted to Line-Spectrum Frequencies (LSF). They are split-vector-quantized according to the method in [2], resulting in 3 codebook indices for the 10 LSF, with 9+8+8=25 bits.

3.2. Subframe Processing

The search for the best components of the excitation vector ex, which is the input signal of the LPC-synthesis filter, is performed sequentially with the adaptive-excitation vector b first. While searching for the best

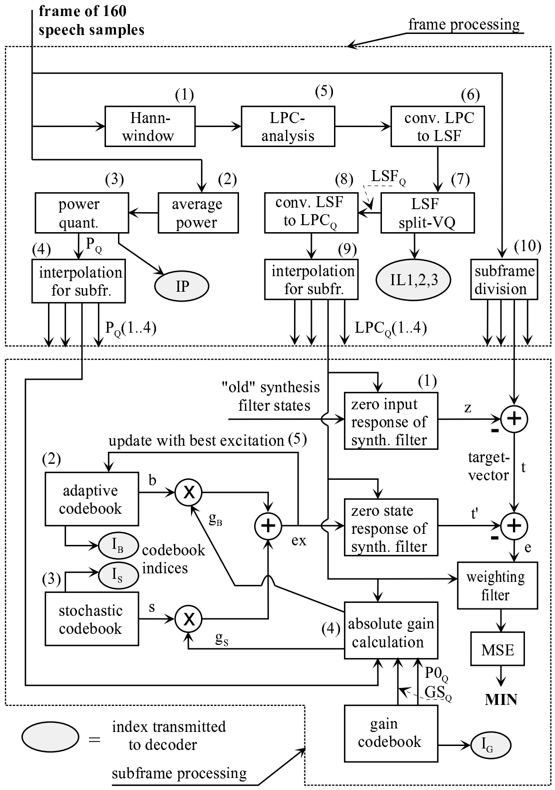

Figure 3: CELP speech-encoder

vectors b and s out of the codebooks the optimal unquantized gains are used. When the best adaptive- and stochastic-excitation vectors have been found, the best gain codevector for those excitation vectors is searched "closed-loop" in the gain codebook.

The adaptive codebook is searched "closed-loop" over a "lag"-range of 20..141. The best lag (number of samples back in the past where the adaptive excitation starts) is coded by 8 bits (index I_B). Up to 5 fractional lags between two integer samples are possible.

The stochastic excitation consists of ten +1/-1 pulses which are systematically placed in the excitation vector and coded by 30 bits (index I_S), similar to ACELP [3].

The excitation signals are scaled by two gains which are jointly quantized by an 8 bit codebook.

3.3. Decoder and Postfiltering

As usual in "analysis-by-synthesis" codecs, the operations to be performed in the decoder are similar to those already carried out in the corresponding encoder stages. A postfilter is employed to increase the speech quality at the decoder-output in terms of human perception.

4. APPLICATION OF CCOE TO CELP

4.1. Markov models for the indices

First the probability distributions $P(I=\lambda)$, $\lambda = 0,1,\ldots,2^{N_I}-1$ and the joint probability distributions $P(I(k)=\lambda, I(k-1)=\nu)$, $\lambda,\nu = 0,1,\ldots,2^{N_I}-1$ of the source-encoder indices were measured. As an example the joint probability distribution of adjacent indices of the mean power is depicted in figure 4. Logarithmic values are used for the plot to squeeze the range of the values that have to be mapped to gray-scales.

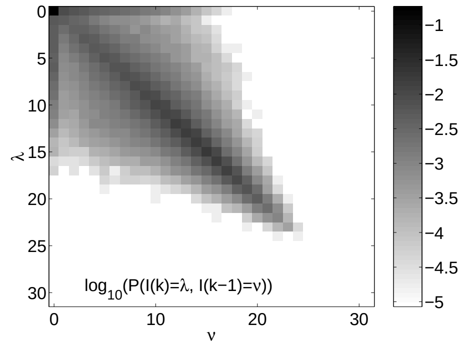

Figure 4: Logarithmic values of the joint probability distribution of adjacent indices $I(k)=\lambda$, $I(k-1)=\nu$ of the mean power ($\lambda,\nu \in \{0,1,\ldots,31\}$)

Parameter	Index	No. of CRC-bits per frame
LSF 1..3	ILSF1	2
LSF 4..6	ILSF2	1
LSF 7..10	ILSF3	1
mean power	IP	2
Lag	IB	$4 \cdot 1 = 4$
Gains	IG	$4 \cdot 1 = 4$
		SUM: 14

Table 1: Allocation of the CRC-bits to the codec indices

Figure 4 shows that there is a considerable dependency between adjacent indices of the mean power. The matrix that is depicted in figure 4 directly corresponds to the Markov-source index transition probabilities in (3).

4.2. Channel Coding

The speech codec was developed for enhanced speech transmission in the GSM full-rate channel. Now the codec shall be used for speech transmission in the half-rate channel where 228 bits per frame (of 20 ms) are available instead of 456 in the full-rate channel. For speech coding 214 bits per frame are required, so only 14 bits are left for channel coding. They were allocated according to table 1. The subjectively most important indices are that of the mean power and that of the first three line-spectrum frequencies. Therefore they are channel coded by CR-Checks with 2 parity bits each. The other indices are channel coded by a single-bit parity-check. Only the pulses of the stochastic excitation remain uncoded since there are no more bits available and the pulses are subjectively least significant.

5. SIMULATIONS AND PERFORMANCE

The channel codes, the Markov-models of the index-dependencies, and the channel outputs can now be used to calculate the a-posteriori probabilities by (3). Still unclear is the choice of the estimators (MS or MAP).

If an MS-estimator is used for the lag-values the problem of averaging between the true and multiples of the pitch-period occurs which would result in a bad performance since the average lag-value would not correspond to an appropriate codebook-entry. Therefore it is already clear from a theoretical point of view that a MAP-estimator should be chosen for the lag-values.

Informal listening tests revealed that there is no noticeable difference in the subjective speech quality if MS-

Quality decrease	Valuation
inaudible	0
hardly audible	1
little annoying	2
annoying	3
very annoying	4
extremely annoying	5
catastrophic	6

Table 2: Mapping of the values 0...6 to the perceived quality decrease

or MAP-estimators are used for the line-spectrum frequencies, the mean power, and the gains. Therefore MAP-estimators are used for all indices. The advantage is that joint source-channel decoding can be implemented as a preprocessor of speech decoding: so the speech decoder itself remains unaffected and it is only fed with the indices estimated by the new algorithms.

In the following the new algorithms are compared with standard techniques of error detection and handling which can be found in the implementations of presently used speech codecs, e.g. the full-rate GSM-system [7]. More sophisticated schemes of error detection and handling were published in [6] where channel errors are detected by a combination of forward error detections by CRCs and "zero-redundancy" error detections (no extra bits) based on the parameter dependencies. In case of an error the current corrupted parameter is replaced by an extrapolation procedure that exploits the correlation of adjacent parameters/indices (bad parameter handling, denoted by "BPH" in figures 5 and 6).

The performance of the error-handling schemes and the joint source-channel decoding by CCOE and it's derivates was evaluated by informal listening tests where the decrease of the speech quality at the speech decoder output caused by the noisy transmission was compared with the "clear-channel" quality of the codec. The values 0..6 are mapped to the perceived subjective speech-quality decrease by table 2.

AWGN channel: First the system-performance was evaluated on an AWGN-channel. To be fair with respect to the transmission power, a rate correction was carried out as in section 2.3 by adding $10\log_{10}(228/214) = 0.275$ dB to the E_b/N_0-values for the systems using channel coding. The results of the listening

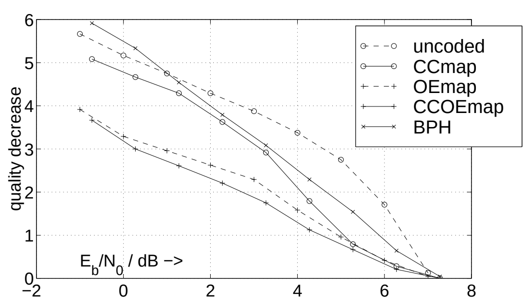

Figure 5: Performance of CCOE for a CELP speech codec on an AWGN channel

tests are depicted in figure 5. The CCOEmap-algorithm works best of all followed by OEmap and CCmap. The BPH-technique works significantly worse compared with CCOEmap and OEmap. The new CCOEmap-algorithm works significantly better than OEmap for moderately corrupted channels (E_b/N_0 around 3 dB).

GSM half-rate channel: The bit-rate required for the transmission of the channel-coded indices exactly maps into the GSM half-rate channel. In figure 6 the results of the listening tests are plotted for the carrier-to-interferer-ratios of C/I=4,7,10 dB. No rate-correction was carried out since the algorithms that do not exploit channel coding cannot take advantage from that because the bit rates are fixed in GSM. The CCOEmap-algorithm

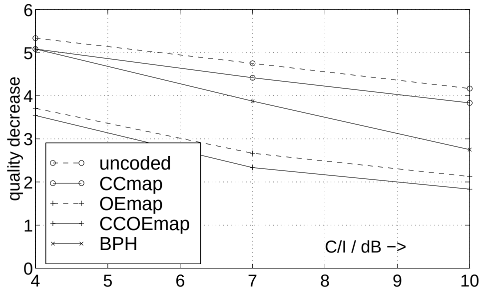

Figure 6: Performance of CCOE for a CELP speech codec on the GSM-half-rate channel

works better than OEmap and both algorithms work much better than all the others. In contrast to the AWGN channel the BPH-algorithm works better than CCmap. This is due to the bursty error-structure on the GSM-channel.

6. CONCLUSIONS

Channel-Coded Optimal Estimation (CCOE) was applied to a CELP speech codec. The performance of CCOE on moderately and strongly corrupted AWGN and GSM half-rate channels (evaluated by informal listening tests) is superior to classical error handling techniques and separate source and channel decoding (CCmap). The disadvantage of CCOE lies in the memory requirements and the computational complexity. Future work will focus on these issues.

7. REFERENCES

[1] Electronic Industries Association, Cellular Systems Dual-Mode Subscriber Equipment - Network Equipment Compatibility Specification, December 1989

[2] K.K. Paliwal and B.S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", IEEE Trans. on Speech and Audio Processing, Vol. 1, No. 1, pp. 3-14, January 1993

[3] J-P. Adoul, P. Mabilleau, M. Delprat, S. Morisette, "Fast CELP coding based on algebraic codes", Proc. ICASSP, pp. 1957-1960, 1987

[4] T. Fingscheidt and P. Vary, "Robust Speech Decoding: A Universal Approach To Bit Error Concealment", Proc. ICASSP, pp. 1667–1670, 1997

[5] N. Görtz, "Joint Source Channel Decoding Using Bit-Reliability Information and Source Statistics", Proc. IEEE International Symposium on Information Theory (ISIT) 1998, p. 9

[6] N. Görtz, "On the Combination of Redundant and Zero-Redundant Channel Error Detection in CELP Speech-Coding", Proc. EUSIPCO'98, pp. 721-724

[7] ETSI/GSM, Rec. 06.10, "Full-rate speech transcoding", March 1992, and ETSI/GSM, Rec. 05.03, "Channel Coding", February 1992