JOINT SOURCE-CHANNEL CODING FOR CODED SPEECH TRANSMISSION

Norbert Görtz

Institute for Communications Engineering (LNT)
Munich University of Technology (TUM)
80290 Munich, Germany
norbert.goertz@ieee.org

Abstract: Coded speech transmission over noisy channels can be interpreted as a delay-constrained encoding/decoding problem for a set of continuous-amplitude signals; the latter are the parameters extracted from the actual speech signal by advanced source coding schemes. The optimization problem for the encoder/decoder-design is formulated and it is shown, that its direct solution is intractable in most relevant cases. Therefore, several approaches for the practical system design are discussed.

1 Introduction

Consider the problem of encoding and transmitting a source signal vector[1] that consists of N_u successive speech samples. Due to delay constraints the vector dimension N_u is limited; depending on the application (e.g., telephone) the limitation is possibly strong. In advanced speech codecs such as CELP (Code Excited Linear Prediction, [1]) encoding is carried out in two steps: first, each vector u of source samples is decomposed into a set of parameter-vectors[2] x^1, x^2.... A parameter vector consists, e.g., of the LPC-coefficients[3] but the mean power of u could also be a (scalar) parameter. In the second step the vectors x^j are quantized and the output bit vectors i^j are generated. This scenario is depicted in Figure 1. Usually, a channel code is used to improve the performance of the system in presence of noise on the channel. The channel codewords z are computed from the bit vector i that is generated from the output-indexes i^1, i^2, ... of the source encoder by the multiplexer.

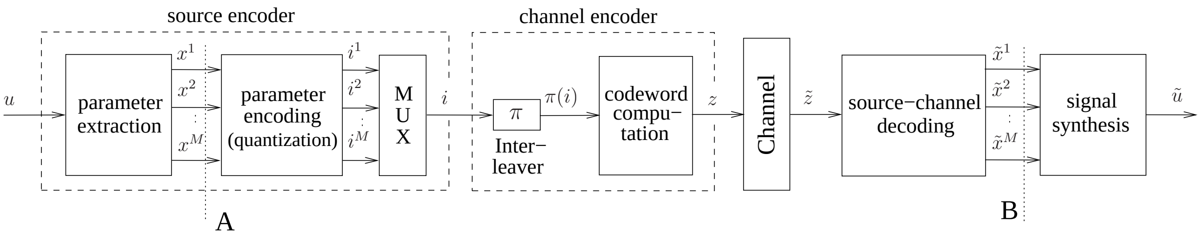

Figure 1 - Model of a speech transmission system

2 System Model

For the analysis below, we will consider the basic model of a communication system that is depicted in Figure 2. The goal is to transmit the input "source" signal, the N-dimensional source vector x, by K channel uses to a destination, where we want to obtain a reproduction $\tilde{x}$ with the highest quality possible. The source vector components x_l, $l = 0, 1, ..., N-1$, have continuous amplitudes and we will use the mean squared error

$$d(x, \tilde{x}) = \frac{1}{N} \sum_{l=0}^{N-1} \left(x_l - \tilde{x}_l\right)^2 \tag{1}$$

[1]For simplicity, the vector time-index is omitted.
[2]Superscripts are used to indicate the number of a source-encoder parameter and its corresponding quantizer index.
[3]The LPC-coefficients (Linear Predictive Coding) describe the spectral shape of a source vector.

Figure 2 - Simplified system model

as a quality measure for the reproduction (distance measure). This model is plugged in between the points “A” and “B” in Figure 1 where, for brevity of notation, we will only consider one parameter signal (therefore we will omit the superscripts in the following). This means, that the mean-squared error of the speech codec *parameters* is used a quality measure and not the mean-squared error of the reconstructed speech signal itself[4]. The reason is that the source codecs are complex algorithms that are hard to describe analytically. Hence, it is unknown how (and probably impossible) to optimally estimate the output speech signal directly from the channel outputs. The use of the mean-squared error in the parameter domain leads to algorithms (e.g., [2]) with acceptable complexity that perform quite well, although they are inherently suboptimal. The block length N of the parameter vectors is determined by the source codec (e.g., $N = 10$ if the LPC coefficients of a narrow-band speech signal are quantized).
In what follows, we will assume that the vector dimensions N and K are fixed and known and our goal is to find a deterministic encoder-decoder pair that minimizes the expected system distortion

$$D \doteq \mathrm{E}\{d(X, \tilde{X})\} \; . \tag{2}$$

When there is no risk of confusion, we will omit the notation of the random variables below, e.g., the probability density function (pdf) of the input vector will be denoted by $p(x)$ instead of $p_X(x)$.

2.1 Channel

The transmission is carried out over a discrete-time channel, which is used K times to transmit the N source/parameter vector components; thus, the channel input signal in our model is the K-dimensional vector z. Where it is necessary, the input alphabet Ω_Z and the output alphabet $\Omega_{\tilde{Z}}$ of the channel are specified below. As an example, we may think of a K-bit channel input and a K-dimensional continuous-valued output from an additive white Gaussian noise channel. The model, however, is not restricted to channels with binary input. We will assume that the noise on the channel is stationary, memoryless, and independent of all other signals and the components of the system. The conditional probability density function $p(\tilde{z} \mid z)$, which describes the probability density of the output vector $\tilde{z}$, given the input vector z, is assumed to be known.
Usually, the input power of the channel is limited, when the channel input alphabet is continuous. For instance, the power per channel-use, averaged over all vectors $z = \varepsilon(x)$, may be limited to $P_{\max}$, i.e.,

$$\frac{1}{K}\,\mathrm{E}\{\|Z\|^2\} \leq P_{\max} \; . \tag{3}$$

Such a power limitation is quite weak, because it does not limit the peak power for a channel-use. In other setups, the average power for each channel input vector may be limited, i.e., $\frac{1}{K}\|z\|^2 \leq P_{\max}$, or the power limit may apply for each of the K channel-uses, i.e., $\|z_l\|^2 \leq P_{\max}$.

2.2 Encoder and Decoder

The encoder, which is described by the unique deterministic mapping

$$z = \varepsilon(x) \; , \tag{4}$$

with $x \in \Omega_X = \mathbb{R}^N$ and $z \in \Omega_Z$, is a device that maps the input source-signal vector x into the input z of the channel. The encoder output z may have to fulfill a constraint imposed by the channel.

[4]Even if it was possible minimize the mean squared reconstruction error of the speech signal, this would still not lead to the best possible quality in terms of human perception, due to auditory masking and cognitive effects of human hearing, which could be (but are not fully) exploited to design better coding algorithms. The modeling of these effects lies far beyond of the scope of this paper.

The decoder is a deterministic device that maps the output $\tilde{z}$ of the channel to the decoder output signal $\tilde{x}$, which should be a good estimate of what has been transmitted. The decoder mapping is denoted by

$$\tilde{x} = \varrho(\tilde{z}) \ . \tag{5}$$

with $\tilde{x} \in \Omega_X = \mathbb{R}^N$ and $\tilde{z} \in \Omega_{\tilde{Z}}$. Since we want to obtain a good reproduction $\tilde{x}$ of the input x in the minimum-mean-squared-error-sense, the output alphabet of the decoder equals that of the encoder input.

3 Minimization of the System Distortion

The expected system distortion is defined by

$$D(\varepsilon, \varrho) \doteq \mathrm{E}\{d(X, \tilde{X}\} = \int\limits_{\Omega_{\tilde{Z}}} \int\limits_{\Omega_X} d\big(x, \tilde{x} = \varrho(\tilde{z})\big) \ \cdot \ p(\tilde{z}, x) \, dx d\tilde{z} \ , \tag{6}$$

where $p(\tilde{z}, x)$ denotes the joint pdf of the random variables $\tilde{Z}$ (channel output) and X (encoder input). Clearly, both the encoder mapping ε and the decoder mapping ϱ are required to compute (6). In what follows, our goal is to choose both mappings such, that the expected distortion D is minimized. For this, it is useful to introduce an alternative representation of D: we may write $p(\tilde{z}, x) = p(x \mid \tilde{z}) \cdot p(\tilde{z})$ and obtain

$$D(\varepsilon, \varrho) = \int\limits_{\Omega_{\tilde{Z}}} D_D(\tilde{z}, \varepsilon, \varrho) \ \cdot \ p(\tilde{z}) \, d\tilde{z} \ , \tag{7}$$

with

$$D_D(\tilde{z}, \varepsilon, \varrho) \doteq \int\limits_{\Omega_X} d\big(x, \varrho(\tilde{z})\big) \ \cdot \ p(x \mid \tilde{z}) \, dx \ . \tag{8}$$

The quantity defined in (8) is the conditional expected distortion, given a particular channel output $\tilde{z}$.

3.1 Optimal Decoder for a Given Encoder

In a first step of the system optimization, our goal is to find the optimal decoder mapping $\varrho^{\circledast}$ for a given encoder mapping ε. Since the distance measure $d(x, \tilde{x})$ in (6) is non-negative for any combination of x and $\tilde{x}$, the distortion $D(\varepsilon, \varrho)$ is minimized, if $D_D(\tilde{z}, \varepsilon, \varrho)$ is minimized for each particular channel output vector $\tilde{z}$. Thus, the optimal decoder for a particular channel output $\tilde{z}$ is given by

$$\varrho^{\circledast}(\tilde{z}, \varepsilon) = \arg\min_{\varrho} D_D(\tilde{z}, \varepsilon, \varrho) \ . \tag{9}$$

With the mean squared error as a distance measure in the source signal space, the solution of (9) is the well-known minimum mean-square estimator

$$\varrho^{\circledast}(\tilde{z}, \varepsilon) = \mathrm{E}\{X | \tilde{Z} = \tilde{z}\} = \int\limits_{\Omega_X} x \, p\big(x \mid \tilde{z}\big) dx = A \int\limits_{\Omega_X} x \cdot \underbrace{p\big(\tilde{z} \mid z = \varepsilon(x)\big)}_{\text{cond. channel pdf}} \cdot \underbrace{p(x)}_{\text{source pdf}} \ dx \ , \tag{10}$$

with

$$\frac{1}{A} \doteq p(\tilde{z}) = \int\limits_{\Omega_X} p\big(\tilde{z} \mid z = \varepsilon(x)\big) \cdot p(x) \, dx \ . \tag{11}$$

For any deterministic choice of the encoder mapping ε, (10) and (11) define the optimal decoder, which minimizes the expected system distortion for each particular channel output $\tilde{z}$.

3.2 Optimal Encoder

Now that we found the optimal decoder $\varrho^{\circledast}(\varepsilon)$ for any encoder ε, we will use the result and try to optimize the encoder mapping; this way, the optimal system would be determined. Since our goal is to minimize the expected system distortion, the general optimization problem is given by

$$\varepsilon^{\circledast} = \arg \min_{\varepsilon\text{: constraint}} D(\varepsilon, \varrho^{\circledast}(\varepsilon)) \ . \tag{12}$$

In most practically relevant situations, the solution of (12) is intractable, even by numerical methods. One problem is that, as we vary over the possible choices for ε, we have to vary the optimal decoder $\varrho^{\circledast}(\varepsilon)$. Hence, for simplification we may try to find an optimal encoder $\varepsilon^{\circledast}(\varrho)$ *for a given fixed decoder* ϱ; the corresponding optimization problem would read $\varepsilon^{\circledast}(\varrho) = \arg \min_{\varepsilon\text{: constraint}} D(\varepsilon, \varrho)$. Although this optimization problem is still intractable in most cases, it is easier to solve than (12), because the decoder does not change within the optimization. In [3], the problem is solved for a system that directly transmits source samples by 8-PSK modulation.

3.3 Special Cases

In this section, some special cases are briefly discussed, for which optimal and realizable solutions of the joint source-channel coding problem are known. Further results on this topic arose in the recent literature [4]; in contrast to these results, we will restrict ourselves to the practically relevant situation that a source and a channel are given, and a transmission system is to be designed, that transmits the source signal over the channel with minimum mean-squared error by use of deterministic mappings at both the encoder and the decoder.

3.3.1 Gaussian Source and Gaussian Channel

Assume that we want to transmit an uncorrelated discrete-time Gaussian source signal with the variance σ_x^2. We may use the channel exactly once per source sample, i.e., $N = K$, and the average input power of the channel (per use) is limited to $\frac{1}{K} E\{\|Z\|^2\} = P_{\text{max}}$. The channel adds uncorrelated zero-mean Gaussian noise samples, which have the variance σ_w^2. Figure 3 shows the optimal system [5]: the input samples x are individually scaled by $\sqrt{P_{\text{max}}}/\sigma_X$ so that the power constraint is fulfilled (encoder mapping) and each channel output sample is individually scaled by $\sigma_X/\sqrt{P_{\text{max}}}/\big(1 + \sigma_w^2/P_{\text{max}}\big)$ (decoder mapping, which directly follows from (10)). Surprisingly, this system is optimal for any choice of N,

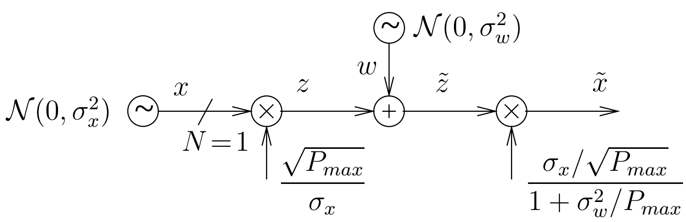

Figure 3 - Optimal encoder-decoder pair for the transmission of a Gaussian source over a Gaussian channel with average input power constraint.

even for $N = 1$. No amount of source and channel coding of long blocks ($N \to \infty$) could improve the system [5]. If, however, *any* of the requirements for the source and the channel is not fulfilled ($N \neq K$, source or channel with memory, different power constraint), then the system is no longer optimal and complex coding schemes with long block lengths are required.

3.3.2 Binary Symmetric Channel

If we assume a binary symmetric channel we have N continuous-amplitude source vector components that have to be mapped to K bits by the encoder. Hence, the number of possible channel inputs is limited to $M = 2^K$. This means that the source signal has to be quantized, i.e., the source signal space Ω_X

has to be partitioned into M disjoint regions $\Omega_X^{(j)}$, $j = 0, 1, ..., M-1$. If the partition regions $\Omega_X^{(j)}$ are known, which we will assume for the moment, the encoder mapping is simply given by

$$z = \varepsilon(x) = j \quad \text{if} \quad x \in \Omega_X^{(j)} \quad . \tag{13}$$

The binary representation of the number j of the region $\Omega_X^{(j)}$ that contains the current source vector is transmitted over the channel. At its output, a K-dimensional bitvector is received (denoted by $\tilde{Z}$), i.e., the channel-output alphabet equals the input alphabet, but due to "channel-noise" some bits may have been flipped, leading to a different bit vector at the channel output; the corresponding "index" transition probabilities are denoted by $P(\tilde{Z} = k \mid Z = j)$. The application of (10) leads to the optimal decoder

$$\tilde{x}^{\circledast}(k) = \varrho^{\circledast}(k, \varepsilon) = \frac{\sum_{j=0}^{M-1} P\big(\tilde{Z} = k \mid Z = j\big) \cdot P(j) \cdot y_j}{\sum_{j\prime=0}^{M-1} P(\tilde{Z} = k \mid Z = j\prime) \cdot P(j\prime)} \;, \quad k = 0, 1, ..., M-1 \tag{14}$$

with

$$P(j) = \int_{\Omega_X^{(j)}} p(x)\, dx \qquad \text{and} \qquad y_j = \int_{\Omega_X^{(j)}} x \cdot p(x)\, dx \Big/ P(j) \;, \tag{15}$$

where $P(j)$ is the probability that a source vector lies within the partition region $\Omega_X^{(j)}$ and y_j is the centroid of the partition region $\Omega_X^{(j)}$. Since the optimal decoder produces only a limited number of outputs $\tilde{x}^{\circledast}(k)$, the optimal decoding results (14) can be precomputed and stored in a table—the so-called codebook; thus, if some k appears at the channel output it can be decoded by a simple table-lookup.

The implementation of the encoder (13) is a problem in practice because the mathematical descriptions of the partitions $\Omega_X^{(j)}$ are usually hard to handle. Therefore, encoding is equivalently reformulated as a nearest neighbor search-problem in which the precomputed estimates (14) are used as the codevectors in a distance computation that includes the index transition probabilities due to bit errors. The codevectors are optimized by an iterative algorithm [6] that works similar to vector quantizer codebook training.

The scheme described above is called Channel-Optimized Vector Quantization (COVQ) [6]. Due to the complexity of encoding and the memory requirements for the codebook, COVQ is only applicable, if the number K of bits is small, e.g., $K < 10$. Therefore, COVQ cannot be directly used, e.g,. for speech, audio and image coding, because at a "typical" bit rate of 1 bit per sample the maximum block-length would only be $N = 10$ samples: unfortunately, source coding at that rate with sufficient quality is only possible with larger block lengths. In our setup, COVQ could be used to encode the parameters of a source codec. One drawback is, however, that for each value of the error probability a new COVQ codebook is required. In [7] a simplification of COVQ is introduced, which allows to retain a great deal of the performance gains of COVQ while only one codebook and an channel-adaptive scaling factor for the codevectors is required. Moreover, the extension of COVQ to channels with continuous outputs is possible. COVQ is an interesting concept, because it is the only non-trivial scheme, in which the encoder mapping is *not* divided into a cascade of partial mappings. At the same time, COVQ is the optimal solution for the delay-constrained problem, as long as we assume that the training procedure for the codevectors finds the global minimum of the distortion (which is an issue in practice).

4 Practical Approaches to Joint Source-Channel Coding

4.1 Separation of Source and Channel Coding

The foundation for the structure of today's systems is Information Theory [8]: roughly speaking, it guarantees that (possibly by use of infinitely large block lengths) one may replace the encoder and the decoder mappings in Figure 2 by cascades of mappings with binary interfaces, at which independent and uniformly distributed bits are exchanged—without any loss in performance compared with the best possible direct mapping. The basic notion is to apply channel coding to achieve error-free bits at a bit rate that equals the channel capacity. Then, a source code is applied to reduce the number of bits required to represent the source signal to the amount that is available on the channel; clearly, this reduction is only possible at the price of some distortion that is imposed on the source signal. It should be noticed, that the problem stated in Section 3.3.1, with its simple solution depicted in Figure 3, could also be solved by the separation into source and channel coding—at the price of infinite block length and complexity. Thus, the blindfold adoption of asymptotic results from Information Theory can turn out to be a bad choice in practice. Nevertheless, with some modifications discussed below, source and channel coding are separate processing steps in all practical system. The reason is that the separation allows to construct partial mappings, e.g., channel encoders/decoders, that, compared with asymptotic results, sometimes may have only moderate performance,[5] but their complexity is tolerable.

If the separation principle is applied in a system with limited block lengths (which are usually imposed by constraints for delay and complexity) the following problems occur: (i) it is impossible to realize channel codes that have a residual bit error rate of zero, (ii) practical source encoders don't produce independent and uniformly distributed bits, i.e., the coded bits contain residual redundancies, and (iii) the distortion at the source decoder output depends on which bit is in error.

4.2 Unequal Error Protection and Error Concealment

Figure 4 shows some system modifications that account for the drawbacks of the separation principle stated above. Implicitly, a channel with discrete input alphabet (e.g., by application of a digital modula-

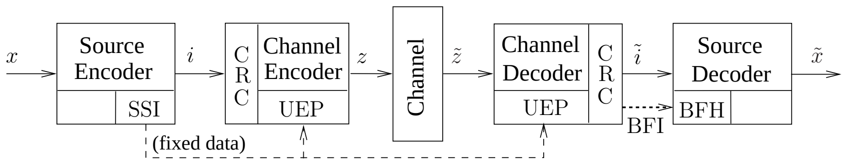

Figure 4 - Unequal error protection and error concealment

tion scheme) is assumed, because the channel codeword z—a bit vector—is directly fed into the channel. As the output bits of the source encoder are not equally sensitive to bit errors some of the data bits cause stronger quality degradations than others if they are in error but are used for decoding anyhow. Source significance information (SSI) can be derived from that and be used by the channel-coding scheme to apply stronger error protection to the more sensitive bits, e.g., by puncturing of those output bits of a low-rate channel code that protect the less sensitive bits.

The residual redundancies in the source-encoder output bits can be exploited to conceal bit-errors (bad frame handling (BFH)), e.g., by repetition of the old bits from the previous block. Clearly, this only makes sense, if the bits are correlated in time. For the initiation of the error concealment, an error detection is required, which is usually realized by an error-detecting channel code (Cyclic Redundancy Check (CRC)). The strongly sensitive bits of the source encoder are CRC-encoded (Class-1-bits) and afterwards the CRC codeword is channel encoded together with all remaining data bits. Thus, in combination with UEP, the bits of the source encoder are divided into several classes which are protected by individually adjustable amounts of channel-code redundancy. Fortunately, the more sensitive bits are typically the stronger correlated ones at the same time, i.e., unequal error protection and error concealment can be combined very efficiently. Hence, both schemes are frequently used, e.g., in every mobile radio standard.

4.3 Source-Controlled Channel Decoding

The basic scheme is depicted in Figure 5. The idea is to exploit the residual redundancies more systematically than above as a-priori information in the decision procedure for the data bits in channel decoding to reduce the bit error rate. This concept can be implemented very efficiently for the decoding of convolutional codes; moreover, the idea can also be extended to soft-output decoders (APRI-SOVA, [10]),

[5] It should be mentioned, that, in the "waterfall-region," Turbo Codes [9] and the developments based thereon brought the performance of channel coding schemes with fairly long block-lengths very close to the information theoretical bounds.

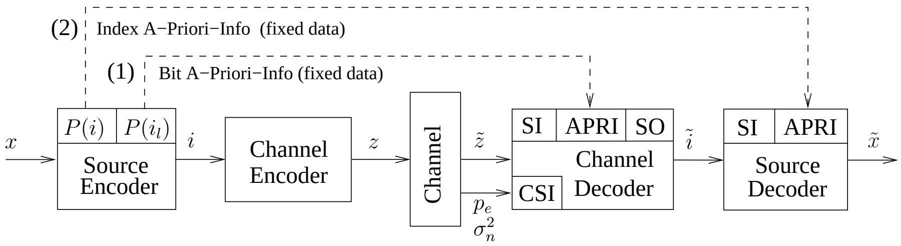

Figure 5 - (1) Source-controlled channel decoding and (2) estimation-based source decoding

which supply reliability information—and not only hard decisions—for the data bits. These reliabilities can be exploited by estimation-based source decoders described below. For a correct weighting of the a-priori information and the soft channel-outputs, a channel state information (CSI) must be available (e.g., the noise variance).

A drawback is that the redundancies are usually exploited on bit-basis[6] (because mostly binary channel codes are used), although the source encoder often emits correlated *indexes* that consist of several bits. Thus, a large part of the redundancies is removed if only the marginal distributions of the individual bits are used as a-priori information.

4.4 Estimation-Based Source Decoding

In conventional systems, simple procedures, e.g., quantizer table-lookups, are frequently applied for source decoding. If, however, source encoding is not perfect and residual redundancies—e.g., correlations in time—are left in the data bits, this a-priori knowledge may be exploited for better source decoding by performing (optimal) estimations of the source-encoder input [12–14]; such a system is also depicted in Figure 5.

Conceptually, the idea is similar to the one described in the previous section, but now it is possible to exploit the full correlations on *index*-basis. The soft-in/soft-out (SISO) channel decoder acts as a device that improves the reliability of the virtual channel "seen" by the source decoder. As in the previous section, bit-based a-priori information may be used to aid channel decoding, but it is not clear, how the a-priori information shall be optimally divided between the channel decoder and the source decoder without using it twice.

4.5 Iterative Source-Channel Decoding

In [15, 16] iterative source-channel decoding is introduced. Although the sketch of the system in Figure 6 is similar to the systems in Figure 5, the theoretical concept is different: the transmitter is interpreted as

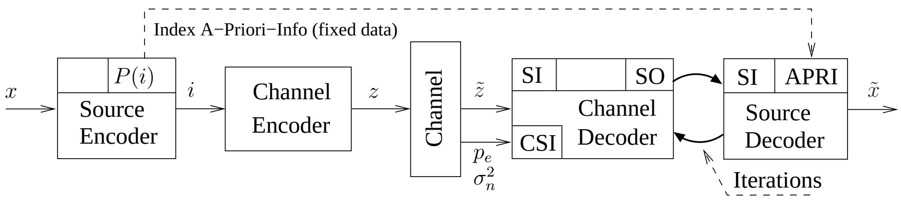

Figure 6 - Iterative source-channel decoding

a serially concatenated channel-coding scheme; the constituent "codes" are the implicit residual redundancies within the source encoder indexes and the explicit redundancy of the channel code. This calls for the use of the Turbo Principle [9, 17] for decoding: as in all iterative decoding schemes, decoders for both constituent "codes" must be available that are able to exchange extrinsic information on the data bits within the iterations. While such decoders for convolutional channel codes are well-known from

[6] In [11] a channel-decoding scheme for binary convolutional codes is given that is able to optimally exploit index-based a-priori information, if all the bits of an index are adjacently mapped into the input of the channel encoder. However, the method may not be applicable because usually bit-reordering/interleaving is employed at the channel-encoder input to account for different sensitivities of the bits or to improve the distance properties of the concatenated code.

literature [17], they can also be formulated on the basis of estimation-based decoding for the source-coding-part of the system. It can be shown [15], that the iterative decoding scheme is a well-defined approximation of the prohibitively complex optimal joint decoder. In a recent publication [18] we have introduced a procedure for the optimization of the quantizer bit mappings for the iterative decoding process, which leads to further strong gains in transmission power. Among the realizable decoding schemes, iterative source-channel decoding works best but at the same time it is the most complex approach.

5 Conclusions

The problem of designing encoder-decoder pairs for the transmission of continuous-amplitude (parameter) source signals with a delay-constraint was considered. The optimal solution turns out to be intractable in various practical cases both analytically and numerically. Therefore, the encoders and the decoders are split into cascades of partial mappings, which can be implemented with acceptable complexity; the concept is based on the separation principle from Information Theory. Modifications are necessary, however, that partially rejoin the separated source and channel coders to compensate for the imperfectness of the real-world system components. Some practical approaches for this were given; the most recent concepts concentrate on the close-to-optimum design of the decoder by use of the Turbo Principle.

References

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 25.1.1–25.1.4, Mar. 1985.

[2] N. Görtz, "Joint source-channel decoding by channel-coded optimal estimation (CCOE) for a CELP speech codec," in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH)*, vol. 3, pp. 1447–1450, Sept. 1999.

[3] N. Görtz, "Source-adaptive shifting of modulation signal points for improved transmission of waveform-signals over noisy channels," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 2517–2520, May 2002.

[4] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, pp. 1147–1158, May 2003.

[5] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2693–2723, Oct. 1998.

[6] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Transactions on Information Theory*, vol. 37, pp. 155–160, Jan. 1991.

[7] N. Görtz and J. Kliewer, "Memory efficient adaptation of vector quantizers to time-varying channels," *EURASIP Signal Processing*, vol. 83, pp. 1519–1528, July 2003.

[8] R. J. McEliece, *The Theory of Information and Coding*. Cambridge: Cambridge University Press, second ed., 2002.

[9] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, Oct. 1996.

[10] J. Hagenauer, "Source-controlled channel decoding," *IEEE Transactions on Communications*, vol. 43, pp. 2449–2457, Sept. 1995.

[11] S. Heinen, A. Geiler, and P. Vary, "MAP channel decoding by exploiting multilevel source a priori knowledge," in *Proceedings of the ITG-Fachtagung "Codierung für Quelle, Kanal und Übertragung"*, pp. 89–94, Mar. 1998.

[12] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels—applications to combined source-channel coding," *IEEE Trans. on Information Theory*, vol. 40, pp. 186–193, Jan. 1994.

[13] T. Fingscheidt and P. Vary, "Robust speech decoding: A universal approach to bit error concealment," in *Proceedings of the IEEE Intern. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 1667–1670, Apr. 1997.

[14] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Transactions on Information Theory*, vol. 45, pp. 1293–1307, May 1999.

[15] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1662–1670, Sept. 2001.

[16] R. Perkert, M. Kaindl, and T. Hindelang, "Iterative source and channel decoding for GSM," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 2649–2652, May 2001.

[17] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, pp. 429–445, Mar. 1996.

[18] N. Görtz, "Optimization of bit-mappings for iterative source-channel decoding," in *Proceedings of the International Symposium on Turbo Codes & Related Topics*, Sept. 2003.