OPTIMAL AND SUB-OPTIMAL DECODING FOR VECTOR QUANTIZATION OVER NOISY CHANNELS WITH MEMORY

Mikael Skoglund

Department of Signals, Sensors and Systems
Royal Institute of Technology, S-100 44 Stockholm, Sweden
Email: skoglund@s3.kth.se Fax: +46 8 790 7329

ABSTRACT

This paper considers optimal decoding for vector quantization over a noisy channel with memory. The optimal decoder is *soft* in the sense that the unquantized channel outputs are utilized directly for decoding, and no decisions are taken. Since the complexity of optimal decoding is high, we also present an approach to sub-optimal decoding, of lower complexity, being based on Hashimoto's generalization of the Viterbi algorithm. We furthermore study optimal encoding and combined source–channel coding. Numerical simulations demonstrate that both optimal and sub-optimal soft decoding give prominent gain over decision-based decoding.

1 INTRODUCTION

The study of combined source-channel coding has become a major field of research, partly motivated by the increasing importance of wireless communications. The field is, however, also interesting from a more fundamental point of view: Implicit in Shannon's work is the fact that the source and channel coding can be separated without loss of optimality [1]. However the positive coding theorems of information theory only show such separability in the limit of infinite codeword length and hence infinite delay, thus justifying the study of combined source-channel coding, for example, when delay is a limiting factor (such as in two-way communications). Much of the existing work has concerned vector quantization (VQ) for noisy channels [2–12]. Most previous work, such as [2–4], has concentrated on discrete memoryless channels, with an emphasis on the binary symmetric channel. Some work, for example [5, 6, 11], has however studied robust VQ over waveform channels using *soft decoding*. In soft VQ decoding the operation of the decoder is not defined by a look-up in a finite decoder codebook, instead all of the received soft information is utilized for decoding and the decoder, in effect, has an infinite output alphabet. Such decoding was studied for the AWGN channel in [5, 6, 8, 11], and for Rayleigh fading channels in [7, 9, 11]. Soft decoding has also been employed for combined multiuser and VQ decoding [13].

In the work mentioned so far, a memoryless channel was assumed. Most practical communication channels are, however, not memoryless. There are a number of ways to handle the obstacles introduced by channel memory. For example, for channels with intersymbol interference (ISI), channel equalization algorithms are employed (see, e.g., [14, 15] and [16]). In this paper we take on the approach of designing optimal VQ encoding/decoding for a given channel with memory (or ISI). Previous work on VQ for channels with memory can be found in [10, 17].

2 PRELIMINARIES

The investigated communication system is depicted in Figure 1. The purpose of the system is to transmit source vectors, $\mathbf{X}_n$, to a destination by the use of VQ. The VQ data is transmitted over a noisy channel with memory. Estimates, $\hat{\mathbf{X}}_n$, are then formed by the decoder based on the received vectors $\mathbf{R}_n$. This section presents the different parts of the system, and states the basic assumptions made.

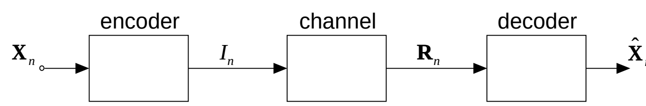

Figure 1: System model.

The source, $\{\mathbf{X}_n\}$, is assumed stationary. The encoder maps d-dimensional source vectors into integers according to $\mathbf{X}_n \in \mathcal{S}_i \Rightarrow I_n = i$ where the sets $\{\mathcal{S}_i\}_{i=0}^{N-1}$ form a partition of $\mathbb{R}^d$. We assume that $N = 2^L$, where L is an integer. Let $P(i) = \Pr(\mathbf{X}_n \in \mathcal{S}_i)$, and define the encoder centroids as $\mathbf{c}(i) = E[\mathbf{X}_n | I_n = i]$.

We study discrete-time channels, modeled according to

$$\mathbf{R}_n = \mathbf{s}(\mathbf{I}_{n-M}^n) + \mathbf{W}_n, \tag{1}$$

where $\mathbf{R}_n$ is the K-dimensional channel output, $\mathbf{s}(\mathbf{I}_{n-M}^n)$ is a known deterministic function of the $M+1$ transmitted indices $\mathbf{I}_{n-M}^n$ (where $\mathbf{I}_{n-M}^n = (I_{n-M}, \ldots, I_n)^T$), and $\{\mathbf{W}_n\}$ is iid noise with known pdf, p_W. Physically, for $M > 0$ such a model is valid for signaling over a stationary channel that introduces memory according to the mapping $\mathbf{s}(\mathbf{I}_{n-M}^n)$. An important special case, that will be considered in the simulations, is a binary-input Gaussian channel with memory modeled using a finite impulse response linear filter. The output of this channel, at "bit-level" time m, is given by

$$R_m = \sum_{l=0}^{M_b} h_l b_{m-l} + W_m, \quad m = 1, 2, 3, \ldots \tag{2}$$

where h_l, $0 \le l \le M_b$, is the (real-valued) impulse response of the channel and $\{W_m\}$ is AWGN of known variance σ_W^2. Here, the bits, $b_m(I_n) \in \{\pm 1\}$, of the index I_n at "block-level" time n, are transmitted as $b_m(I_n) = b_{(n-1)L+m}$, $m = 1, 2, \ldots, L$. In this model R_m corresponds to whitened samples from the receiver matched filter in coherent BPSK signaling over a linear filter channel with AWGN [16].

Decoding, at time n, is based on the observation $\mathbf{R}_1^n = \mathbf{r}_1^n$, (where $\mathbf{R}_1^n = (\mathbf{R}_1^T, \ldots, \mathbf{R}_n^T)^T$). The decoder produces an estimate, $\hat{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n)$, of the vector $\mathbf{X}_{n-\delta}$ transmitted at time

$n-\delta$, where $\delta \geq 0$ is the decoding delay. Since $\mathbf{R}_n$ is the "unquantized" channel output, the decoder is *soft* in the sense that the estimates are not chosen from a finite codebook. On the contrary, the decoder output alphabet is infinite. Expressed differently, the decoder is *estimator*-based rather than *detection*-based (c.f., [5, 6, 11]).

We define the *distortion*, $\bar{D}$, of the system as

$$\bar{D} \triangleq \bar{E}\|\mathbf{X}_{n-\delta} - \hat{\mathbf{X}}_{n-\delta}(\mathbf{R}_1^n)\|^2,$$

where $\bar{E}[X_m] \triangleq \lim_{M\to\infty} M^{-1}\sum_{k=1}^{M} EX_k$, and "optimal" will throughout the paper refer to the minimization of $\bar{D}$. We assume that the limit exists in all cases of interest.

3 OPTIMAL DECODING

In this section we study optimal soft VQ decoding, assuming that the source, the encoder and the channel are given. A more thorough account for derivations and proofs can be found in [18].

The optimal decoder is the MMSE estimator [10, 18]

$$E[\mathbf{X}_{n-\delta}|\mathbf{r}_1^n] = \sum_{\mathbf{i}_1^n} E[\mathbf{X}_{n-\delta}|\mathbf{i}_1^n]P(\mathbf{i}_1^n|\mathbf{r}_1^n). \quad (3)$$

Thus, optimal decoding at time n involves a sum over N^n terms (increasing with time!), making decoding based on (3) impractical. To overcome this difficulty we assume that the source can be modeled as a Markov process. We also assume that the resulting index process $\{I_n\}$ is Markov. Under these assumptions, it can be shown that (c.f. [18])

$$E[\mathbf{X}_{n-\delta}|\mathbf{r}_1^n] \approx \sum_{\mathbf{i}_{n-\delta-1}^n} \mathbf{z}(\mathbf{i}_{n-\delta-1}^n)P(\mathbf{i}_{n-\delta-1}^n|\mathbf{r}_1^n) \triangleq \tilde{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n)$$

where $\mathbf{z}(\mathbf{i}) \triangleq E[\mathbf{X}_{n-\delta}|\mathbf{I}_{n-\delta-1}^n = \mathbf{i}]$. In the rest of the paper, we will focus on the decoder $\tilde{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n)$. We see that $\tilde{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n)$ is formed as a sum over the vectors $\{\mathbf{z}(\mathbf{i})\}$, weighted according to $\Pr(\mathbf{I}_{n-\delta-1}^n = \mathbf{i}|\mathbf{r}_1^n)$. Hence, the vectors $\{\mathbf{z}(\mathbf{i})\}$ are used as "codevectors" and the decoding is based on the a-posteriori most probable vectors. We refer to the vectors $\{\mathbf{z}(\mathbf{i})\}$ as the *multi-centroids* of the encoder, since they are generalizations of the "single-index" centroids $\{\mathbf{c}(i)\}$, based on $\delta+2$ indices instead of only one (c.f. also [19]).

In order to implement $\tilde{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n)$, the a-posteriori probabilities have to be computed from the received data. The following presents a recursion for updating these probabilities in time as more data arrives: For ease of notation define

$$\mathbf{j}_n = \begin{cases} \mathbf{i}_{n-M}^n = (i_{n-M}, \ldots, i_n)^T & ; 0 \leq \delta \leq M-1 \\ \mathbf{i}_{n-\delta-1}^n = (i_{n-\delta-1}, \ldots, i_n)^T & ; 1 \leq M \leq \delta \end{cases}$$

Then we have that

$$\tilde{\mathbf{X}}_{n-\delta}(\mathbf{r}_1^n) = \sum_{\mathbf{j}_n} \mathbf{z}(\mathbf{i}_{n-\delta-1}^n)P(\mathbf{j}_n|\mathbf{r}_1^n). \quad (4)$$

The conditional pdf, describing the channel output in the model (1), is $p(\mathbf{r}_n|\mathbf{i}_{-\infty}^n) = p_W(\mathbf{r}_n - \mathbf{s}(\mathbf{i}_{n-M}^n))$. Regarding this pdf we note that; (i) $p(\mathbf{r}_n|\mathbf{i}_{-\infty}^n) = p(\mathbf{r}_n|\mathbf{i}_{n-M}^n)$, and; (ii) $p(\mathbf{r}_n|\mathbf{r}_1^{n-1}, \mathbf{i}_{n-M}^n) = p(\mathbf{r}_n|\mathbf{i}_{n-M}^n)$. Utilizing these properties of the pdf, it can be shown (c.f., [15]) that a recursion for the updating of the a-posteriori probabilities exists, and can be formulated as a *prediction* step

$$P(\mathbf{j}_n|\mathbf{r}_1^{n-1}) = P(i_n|i_{n-1}) \sum_{i_{n-1-K}=0}^{N-1} P(\mathbf{j}_{n-1}|\mathbf{r}_1^{n-1}), \quad (5)$$

where $K = M$ for $\delta \leq M-1$ and $K = \delta+1$ for $M \leq \delta$, and a *filtering* step

$$P(\mathbf{j}_n|\mathbf{r}_1^n) = \frac{P(\mathbf{j}_n|\mathbf{r}_1^{n-1})p(\mathbf{r}_n|\mathbf{j}_n)}{\sum_{\mathbf{j}_n} P(\mathbf{j}_n|\mathbf{r}_1^{n-1})p(\mathbf{r}_n|\mathbf{j}_n)}. \quad (6)$$

Similar recursions to these have been presented before in the context of maximum a-posteriori detection for channels with ISI (see, e.g., [15]).

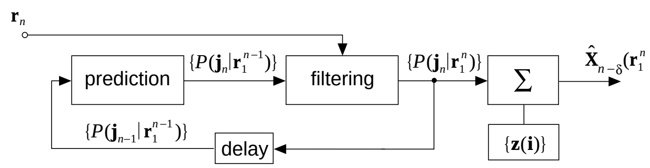

Figure 2: Recursive implementation of the optimal decoder.

The operation of the decoder is illustrated in Figure 2: One new channel output vector, $\mathbf{r}_n$, is received. The probabilities are updated according to (5) and (6), and the codebook, $\{\mathbf{z}(\mathbf{i})\}$, is then utilized for decoding as described in (4).

4 GVA-BASED DECODING

Assume that $\delta \geq M$, for simplicity, and consider the expression (4) for $\tilde{\mathbf{X}}$, being a sum over $N^{\delta+2}$ terms. It is reasonable to assume that terms in the sum corresponding to low values of the probabilities $P(\mathbf{i}_{n-\delta-1}^n|\mathbf{r}_1^n)$ can be neglected. Based on this observation, we formulate an approach to sub-optimal soft decoding. The proposed method utilizes Hashimoto's generalized Viterbi algorithm (GVA) [20]. In the following we provide a short description of this scheme; more details can be found in [18].

Let $\mathbf{v}(n) = \mathbf{i}_1^n$ (corresponding to the bit-sequence $\mathbf{b}_1^{nL}$, where $b_{(n-1)L+m} = b_m(I_n)$), $\mathbf{u}(n) = \mathbf{i}_{n-\delta-1}^n$ and $\mathbf{s}(n) = \mathbf{b}_{nL-\beta+1}^{nL}$ (with $\beta < L(\delta+2)$). The vector $\mathbf{v}(n)$ is said to have the *label* $\mathbf{s}(n)$ if the corresponding bit-sequence, $\mathbf{b}_1^{nL}$, ends with $\mathbf{s}(n)$. One step of the algorithm can now be described as follows:

(0) As initial condition at time n the algorithm retains, for each $\mathbf{s}(n-1)$, a number of S surviving paths, $\mathbf{v}(n-1)$, each having the label $\mathbf{s}(n-1)$. For each path the value of the joint pdf $p(\mathbf{v}(n-1), \mathbf{r}_1^{n-1})$ is stored. Also stored, is the end part $\mathbf{u}(n-1)$ of $\mathbf{v}(n-1)$;

(i) One new vector, $\mathbf{r}_n$, is received. For each of the $S2^\beta$ saved paths, and each value of i_n, the probability $p(\mathbf{v}(n), \mathbf{r}_1^n)$ is calculated, using the stored probabilities, $p(\mathbf{v}(n-1), \mathbf{r}_1^{n-1})$, as $p(\mathbf{v}(n), \mathbf{r}_1^n) = p(\mathbf{r}_n|\mathbf{i}_{n-M}^n)P(i_n|i_{n-1})p(\mathbf{v}(n-1), \mathbf{r}_1^{n-1})$ and the new paths, $\mathbf{v}(n)$, are then classified according to their label;

(ii) For each $\mathbf{s}(n)$ the paths $\mathbf{v}(n)$, having label $\mathbf{s}(n)$, corresponding to the S largest values of $p(\mathbf{v}(n), \mathbf{r}_1^n)$ are found and stored;

(iii) The last step at time n is the soft decoding, based on the approximation

$$\tilde{\mathbf{X}}_{n-\delta} \approx \frac{\sum \mathbf{z}(\mathbf{u}(n))p(\mathbf{v}(n), \mathbf{r}_1^n)}{\sum p(\mathbf{v}(n), \mathbf{r}_1^n)},$$

where the sums are taken over all saved probabilities and paths.

The complexity of the algorithm is dominated by the computation and comparison of new candidates in steps (*i-ii*). As a function of δ, β, S and L, the complexity of one step of the algorithm is $O(S2^{\beta+L})$ operations [20]. Hence, the tradeoff between performance and complexity can be set by the choice of β and S. Note that the complexity is significantly lower than for optimal decoding. We refer to GVA-based decoding, with parameters β and S as GVA(β,S) decoding.

5 OPTIMAL ENCODING

Here we consider optimization of the encoder. The optimal encoder regions $\{\mathcal{S}_i^*\}$ for a given (but arbitrary) decoder, $\hat{\mathbf{X}}$, and a *memoryless* source are given by

$$\mathcal{S}_i^* = \left\{ \mathbf{x} : i = \arg\min_{i'} \bar{E}\left[||\mathbf{x} - \hat{\mathbf{X}}_{n-\delta}(\mathbf{R}_1^n)||^2 | I_{n-\delta} = i' \right] \right\}.$$

(A proof can be found in [18].) For sources with (Markov) intervector *memory* this expression gives a good approximation to the optimal regions under the condition that $p(\mathbf{i}_1^n \setminus i_{n-\delta} | \mathbf{x}_{n-\delta}, i_{n-\delta}) \approx p(\mathbf{i}_1^n \setminus i_{n-\delta} | i_{n-\delta})$ (where $\mathbf{i}_1^n \setminus i_{n-\delta} = (i_1, \ldots, i_{n-\delta-1}, i_{n-\delta+1}, \ldots, i_n)^T$). Intuitively, this condition says that $\mathbf{X}_{n-\delta}$ does not contain (much) more information about $\mathbf{I}_1^n \setminus I_{n-\delta}$ than does $I_{n-\delta}$. Consequently $p(\mathbf{i}_1^n \setminus i_{n-\delta} | \mathbf{x}_{n-\delta}, i_{n-\delta}) \approx p(\mathbf{i}_1^n \setminus i_{n-\delta} | i_{n-\delta})$ is a reasonable assumption when i_n represents $\mathbf{x}_n$ closely.

The expression for the optimal encoder regions can be employed in an iterative joint design of the encoder-decoder pair, giving a combined source-channel coding approach according to the principle of channel optimized vector quantization (COVQ) (c.f. [2–4] and [18]).

6 SIMULATIONS

In this section we investigate the performance over two different binary channels: *Channel 1*: of length $M_b = 2$ and with impulse response $\mathbf{h}_0^2 = (0.407, 0.815, 0.407)$, and *Channel 2* of length $M_b = 4$ with $\mathbf{h}_0^4 = (0.227, 0.460, 0.688, 0.460, 0.227)$ (both taken from [16] p. 616). Both channels have spectral zeros on the unit circle. We consider first-order Gauss-Markov sources with correlation a, modeled as $X_m = aX_{m-1} + U_m$ where $\{U_m\}$ is iid and Gaussian. The corresponding vector source, $\{\mathbf{X}_n\}$, is obtained as $\mathbf{X}_n = (X_{(n-1)d+1}, \ldots, X_{nd})^T$. Performance is measured in terms of the output signal-to-noise ratio, $E||\mathbf{X}_n||^2 / \bar{E}||\mathbf{X}_n - \hat{\mathbf{X}}_n||^2$ (abbreviated "SNR" below), versus channel signal-to-noise ratio (CSNR). For the binary channels the employed definition of CSNR is $\sigma_W^{-2} \sum_{m=0}^{M_b} h_m^2$.

We compare the introduced soft decoders to soft decision Viterbi equalization (maximum likelihood sequence detection), according to [14], plus table-look up VQ decoding. This two-stage approach is referred to as the *Viterbi decoder* for simplicity. We use a large fixed delay, δ^{V}, in the implementation. The decoder codebook is defined by the encoder centroids in the RVQ results, and by the optimal decoder vectors (see, e.g., [4]) in the COVQ results. The transition matrix of the discrete channel given by the concatenation of the channel and the Viterbi detector was estimated. Employing this transition matrix, good index assignments (IAs) were obtained using simulated annealing (c.f., [3]), and COVQ design was carried out according to, e.g., [4]. For simplicity, we use the notation "V$-\delta^{\mathrm{V}}$" for the discrete channel (at a particular CSNR), where δ^{V} is the window-size. In all results for soft decoding it is assumed that the decoder knows

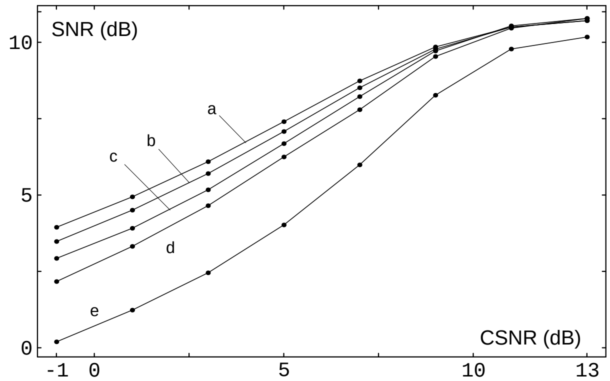

Figure 3: Performance for channel 1. Gauss-Markov source with $a = 0.9$, VQs with L=4 and d=4. (a) Optimal decoding; (b) GVA(4,4); (c) GVA(2,4); (d) GVA(1,2); and (e) Viterbi decoding. For all soft decoders $\delta = 1$ (4 bits), and for Viterbi decoding the delay is 40 bits (channel time-units).

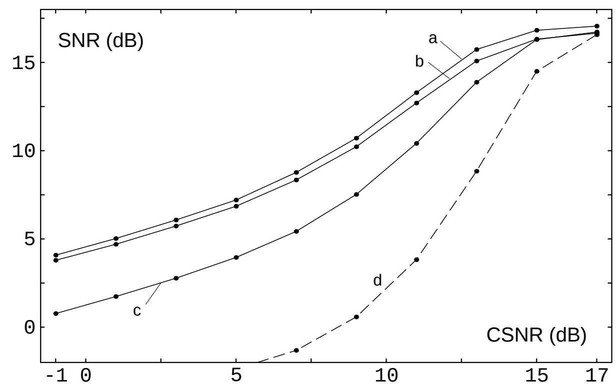

Figure 4: Channel 2, Gauss-Markov source with $a = 0.9$. VQs with $L = 7$ and $d = 3$ (rate 7/3). (a) multi-centroids; (b) centroids; (c) Viterbi decoding, and (d) random IA, Viterbi decoding. The soft decoders have $\delta = 1$ (7 bits), and the Viterbi decoder $\delta^{\mathrm{V}} = 70$.

the CSNR. In practice the decoder has to use (non-perfect) estimates of the CSNR computed at the receiver.

Figure 3 illustrates the performance of optimal decoding and GVA-based decoding. The source is Gauss-Markov with $a = 0.9$, and the encoder is defined by a VQ trained for a noiseless channel. The IA of the encoder is equal in all cases and was obtained for the discrete V-40 channel at a CSNR of 5 dB. Hence, the IA is optimized for Viterbi decoding. We use the same encoder in all cases to investigate the impact on the performance from different decoders. As can be observed, the gain of optimal soft decoding over Viterbi decoding is prominent. For example at an SNR of 7 dB, the gain is 4.1 dB in CSNR. We also note that the GVA-approximations perform well compared to Viterbi decoding. The gain of optimal soft decoding over Viterbi decoding is mainly due to two facts: (*i*) soft decoding gives a gain over decision-based decoding; and (*ii*) the soft decoders utilize knowledge of the source statistics for error protection, since the a-priori (Markov) probabilities are part of the decoder expression (4). This means that the soft decoders can utilize intra- and intervector source redundancy to counteract channel noise and distortion (c.f. [21]).

Figure 4 shows the performance for VQs of higher rate (rate 7/3) than of those employed in Figure 3. The performance is investigated for channel 2. The plot shows the performance of an encoder optimized for a noiseless channel,

Table 1: COVQ performance (SNR in dB) over channel 1. Results for rate 2, 1 and 0.5, 4-dimensional VQ for Gauss-Markov and iid Gaussian sources, for soft and Viterbi decoding. For the soft decoders $\delta = 1$. The Viterbi decoders all use $\delta^{\mathrm{V}} = 40$.

	Gauss-Markov		iid Gauss	
	Soft	Viterbi	Soft	Viterbi
CSNR [dB]	2 dimensions and 4 bits			
9	13.1	11.2	8.31	7.91
7	11.7	9.37	7.24	6.66
5	10.4	7.73	6.12	5.48
3	9.33	6.46	5.17	4.44
1	8.59	5.51	4.31	3.60
CSNR [dB]	4 dimensions and 4 bits			
9	10.1	8.98	4.11	4.00
7	9.39	7.81	3.54	3.29
5	8.32	6.60	2.87	2.58
3	7.46	5.62	2.31	1.99
1	6.93	4.80	1.89	1.57
CSNR [dB]	8 dimensions and 4 bits			
9	7.74	6.95	1.96	1.92
7	7.09	6.14	1.66	1.56
5	6.46	5.36	1.36	1.21
3	5.85	4.59	1.08	0.92
1	5.20	4.13	0.86	0.72

with GVA(4,4) and Viterbi decoding. In curve a the decoder of (4) has been employed, and in b the multi-centroids $\{\mathbf{z}(\mathbf{i})\}$ have been approximated by the single-index encoder centroids; $\mathbf{z}(\mathbf{i}) \approx E[\mathbf{X}_n | I_n = i]$. The IA of the encoder was optimized for the V-70 channel at a CSNR of 5 dB. Also shown, for reference, is the performance of Viterbi decoding for an encoder with random IA. As can be observed, there is a large gain of soft decoding over the Viterbi decoder. For example at an SNR of 10 dB, the gain of optimal decoding over Viterbi is about 1.9 dB in CSNR. Also, as expected, the random IA plus Viterbi decoding performs very poorly, illustrating the importance of a good IA.

Table 1 shows the performance of different COVQ schemes, employing optimal and Viterbi decoding for Gauss-Markov ($a = 0.9$) and iid Gaussian sources. The performance was measured at the same CSNRs as for which the COVQs were trained (perfect match). The same general conclusions as made in connection to the figures hold also for the COVQ performance. In particular, we note that the gain of soft decoding is larger for low CSNRs, and for correlated sources.

7 CONCLUSIONS

We have introduced the optimal (MMSE) soft decoder for vector quantization over a noisy channel with memory. Since the complexity of optimal decoding is high, we also presented a sub-optimal approach, of lower complexity, based on Hashimoto's generalization of the Viterbi algorithm. We, furthermore, considered optimal encoding and combined source–channel coding. Numerical simulations then verified the theoretical results, and demonstrated that the introduced decoders can give prominent gain over decision-based table look-up decoding.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[2] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electronics and Engineering in Japan*, vol. 67-B, pp. 39–47, Jan. 1984.

[3] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 799–809, July 1990.

[4] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 155–159, Jan. 1991.

[5] V. Vaishampayan and N. Farvardin, "Joint design of block source codes and modulation signal sets," *IEEE Transactions on Information Theory*, vol. 38, no. 4, pp. 1230–1248, July 1992.

[6] F. H. Liu, P. Ho, and V. Cuperman, "Joint source and channel coding using a non-linear receiver," in *Proc. IEEE International Conference on Communications*, Geneva, Switzerland, 1993, pp. 1502–1507.

[7] F. H. Liu, P. Ho, and V. Cuperman, "Sequential reconstruction of vector quantized signals transmitted over Rayleigh fading channels," in *Proc. IEEE International Conference on Communications*, New Orleans, USA, 1994, pp. 23–27.

[8] M. Skoglund and P. Hedelin, "Vector quantization over a noisy channel using soft decision decoding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Adelaide, Australia, Apr. 1994, pp. V605–V608.

[9] M. Skoglund, "A soft decoder vector quantizer for a Rayleigh fading channel—application to image transmission," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, USA, 1995, pp. 2507–2510.

[10] V. Kafedziski and D. Morell, "Vector quantization over Gaussian channels with memory," in *Proc. IEEE International Conference on Communications*, Seattle, 1995, pp. 1433–1437.

[11] M. Skoglund and P. Hedelin, "Hadamard-based soft decoding for vector quantization over noisy channels," *IEEE Transactions on Information Theory*, To appear.

[12] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Theory for transmission of vector quantization data," in *Speech coding and synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Elsevier Science, 1995.

[13] M. Skoglund and T. Ottosson, "Soft multiuser decoding for vector quantization over a CDMA channel," *IEEE Transactions on Communications*, vol. 46, no. 3, Mar. 1998.

[14] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[15] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proceedings of the IEEE*, vol. 58, no. 5, pp. 779–785, May 1970.

[16] J. G. Proakis, *Digital Communications*, McGraw-Hill, 3rd edition, 1995.

[17] M. Skoglund and T. Ottosson, "Joint equalization and soft decoding for vector quantization over channels with intersymbol interference," in *Proc. IEEE International Conference on Communications*, Dallas, USA, 1996, pp. 1025–1029.

[18] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *Subm. to IEEE Transactions on Information Theory*, March 1997.

[19] M. Skoglund and J. Skoglund, "On non-linear utilization of interframe memory in speech coding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, USA, May 1998.

[20] T. Hashimoto, "A list-type reduced-contraint generalization of the Viterbi algorithm," *IEEE Transactions on Information Theory*, vol. IT-33, no. 6, pp. 866–876, Nov. 1987.

[21] N. Phamdo and N. Farvardin, "Optimal detection of discrete markov sources over discrete memoryless channels - applications to combined source-channel coding," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 186–193, Jan. 1994.