Combined Source-Channel Coding : Panorama of Methods

CNES Workshop on Data Compression : 13-14 Nov. 1996, Toulouse.

Seyed Bahram ZAHIR AZAMI, Pierre DUHAMEL and Olivier RIOUL
École Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris Cedex 13, France
Tél: 01 45 81 76 77, Fax: 01 45 89 00 20, email: zahir@com.enst.fr

Abstract

This paper reviews a number of different approaches aiming at introducing some robustness into source coders/decoders when the coded bit-stream is to be sent through a noisy channel.

Various methods are presented with reference to a scheme in which all tasks that have to be completed in sequence are explicitly shown. Depending on the assumptions on which these methods rely, some blocks are merged, and have to perform a more complex task which is then to be optimized for minimum distortion under noisy channel conditions.

Past and current studies are outlined in this context.

1 Introduction

Roughly speaking, *source coding* is a data compression process that aims at removing as much as possible redundancy from the source signal, while *channel coding* is the process of intelligent redundancy insertion which creates some kind of protection against the channel noise. In this aspect, these two processes seem to act in opposition.

The joint source-channel coding theorem of Shannon consists of two parts [26]: a direct part that states that if the minimum achievable source coding rate of a given source is below the capacity of a channel, then the source can be reliably transmitted through the channel, considering that the sequences of source samples are appropriately long; and a converse part stating that if the source coding rate is strictly greater than channel capacity, the reliable transmission is impossible. This theorem yields that source coding and channel coding can be treated separately without any loss of performance for the overall system. In other words, the source and channel coding functions are fundamentally separable [20].

Hence, in the majority of the design algorithms, the basic design procedure consists of selecting a source encoder which changes the source sequence into a series of independent equally likely binary digits followed by a channel encoder which accepts binary digits and puts them into a suitable form for reliable transmission over the channel.

This separability holds if the communication is point-to-point, i.e. single channel [25]. However, this hypothesis is not realistic in a broadcasting communication with multi-path fading. In [29] a specially defined source-channel pair is given where the source is transmissible through the channel (with zero error), yet its minimum achievable source coding rate is twice the channel capacity. So a probable overcome in performance is also conceivable for certain specially defined source-channel pairs.

Moreover, a tandem source-channel coding may, in practice, necessitate very long blocks of source symbols and very complex coders. The following example illustrates a simple (artificially defined) case, where even though Shannon's solution is optimal, but there exists another optimal solution which is better, from an *economic* point of view.

Suppose that in a simple transmission scheme, we have a *binary symmetric channel* (BSC), with the transition probability, $\epsilon = 0.10$. It is required to transmit a signal from a *binary symmetric source* (BSS) via this channel, with average distortion, $D \leq 0.1$. Considering Shannon's joint source-channel coding theorem, one must first design a source code for the BSS with average distortion ≈ 0.1, and then design an appropriate channel code for the BSC with very small error probability.

However, there is a simpler system that yields the same performance without source coding nor channel coding. One can transmit the source signals, directly without any coding [21, problem 5.7]. That is, connecting directly the source to the channel. This occurs because the source and the channel are matched to each other in the sense that the transition probabilities of the channel solve the variational problem defining the *rate distortion function* ($R(D)$) and the letter probabilities of the source drive the channel at its capacity [3, page 73]. This simple example shows that despite its optimality, Shannon's separation theorem doesn't result in necessarily the best economic solution.

The objective of combined source-channel coding is to include both source and channel coding modules in the same processing block in order to reduce the complexity of the overall system, compared to the tandem scheme. It is however important to be noticed that the cost to be payed for this reduction of complexity, is the loss of flexibility [20]. If one opts for a jointly coded system, he/she can no longer easily adapt his/her system later to a different source (or channel).

In this paper, we give a general presentation of the problem. Next, vector quantization as a data compression tool is examined. The concept of channel coding is the next subject to be discussed. The first source-channel coding approach that we consider consist of a hierarchical protection of the bits. Then, we investigate the index assignment approach and the simultaneous optimization schemes that optimize the quantizer and the index assignment. Next, we examine the algorithms that bypass the binary representation step and provide directly the constellation points. Finally, we consider the Rate/Distortion approach.

1.1 Communication Model

Here we present the transmission block diagram. So we try to have a very general presentation. Although the presented model is not the most general one : for a more general model, one can put a fading channel instead of AWGN, for example. Figure (1) depicts this general model : the message emitted from the source is first passed from a transformation block, as is very common for audio signals; source compression is performed in order to eliminate more redundancy; *index assignment*[1] (IA), is then used for giving a good bit pattern to each codevector; the resulted source coded bits are then protected by a channel encoder; the modulation permits us to transmit the signal in the physical channel; the channel is assumed to be noisy, so the output of the channel is the sum of its

[1] Also known as *Labeling*.

input and noise which is usually modeled as Gaussian noise; the demodulator and the hard limiter are used to regenerate a binary sequence; finally, a series of operations including channel decoding, inverse index assignment, codebook search and inverse transformation are applied to recover the original message.

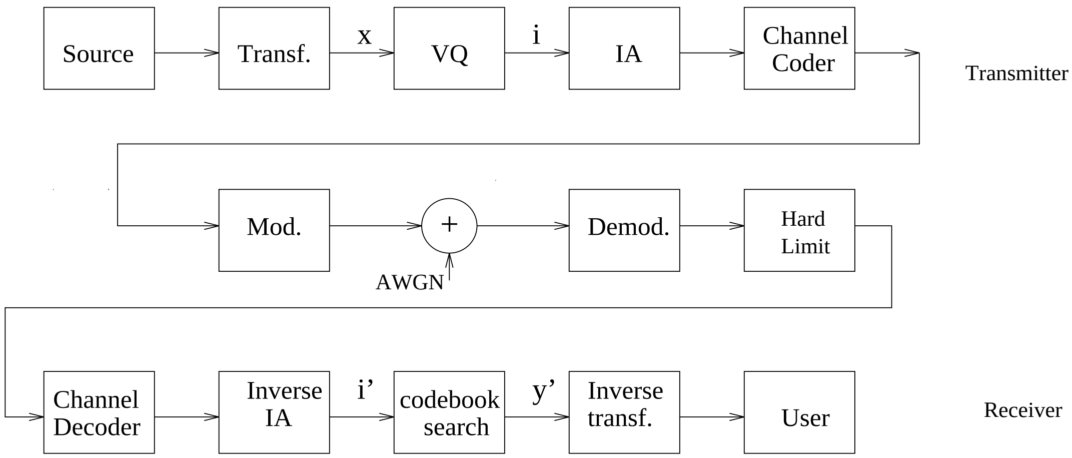


Figure 1: *Block diagram of the transmission system.*

This general model can be simplified in different ways. In fact, each method described in this paper, makes its own assumption on the model and combines some of the blocks in figure (1) into a single block and/or easily omits some of the blocks. For example, a BSC, simply models the modulator, channel noise, demodulator and hard limiter set as in one block.

Some methods make a single block from two or three other blocks and apply some optimization routines to it. As an example, in *coded modulation* (CM), the channel coder and the modulation block are put together. As another example, in modulation organized vector quantization, all the blocks : vector quantization, index assignment, channel coding and modulation are merged together and optimization is made for this merged block.

1.2 Vector Quantization

A common tool for data compression is *vector quantization* (VQ). It is a redundancy removal process that makes effective use of four interrelated properties of vector parameters : linear dependencies (correlations), nonlinear dependencies, shape of the *probability density function* (pdf) and vector dimensionality itself [19].

Let $\mathbf{x} = [x_1 x_2 ... x_{\mathrm{N}}]^T$ be an N-dimensional vector whose components $\{x_k, 1 \leq k \leq N\}$ are real-valued continuous amplitude random variables[2], $\mathbf{y}$ the output of the VQ (another real-valued, discrete-amplitude, N-dimensional vector). We write $\mathbf{y} = q(\mathbf{x})$, where q is the quantization operator.

The value of $\mathbf{y}$ is to be taken from a finite set of L elements : $\mathbf{Y} = \{\mathbf{y}_i, 1 \leq i \leq L\}$, which is called a *codebook*. The design of a codebook consists of partitioning the N-dimensional space of the random vector $\mathbf{x}$ into L non overlapping regions or cells $\{\mathcal{C}_i, 1 \leq i \leq L\}$ and associating with each

[2] Also usually assumed to be of zero-mean, stationary and ergodic.

cell $\mathcal{C}_i$ a vector $\mathbf{y}_i$. A well designed VQ is such that it minimizes a given error criterion. The most usual criterion is the Euclidean distance. Our aim is to minimize the average Euclidean distance, D, over a very large number of samples, M:

$$D = \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} d[\mathbf{x}(n), \mathbf{y}(n)] \tag{1}$$

$$d[\mathbf{x}, \mathbf{y}] = \sum_{k=1}^{N} (x_k - y_k)^2 \tag{2}$$

which simplifies, assuming ergodicity and stationarity, to :

$$D = \sum_{i=1}^{L} p(\mathbf{x} \in \mathcal{C}_i) E[d(\mathbf{x}, \mathbf{y}_i) | \mathbf{x} \in \mathcal{C}_i] \tag{3}$$

$$D = \sum_{i=1}^{L} p(\mathbf{x} \in \mathcal{C}_i) \int_{\mathbf{x}\in\mathcal{C}_i} d(\mathbf{x}, \mathbf{y}_i) p(\mathbf{x}) d\mathbf{x} \tag{4}$$

A well known algorithm for VQ design is the *Lindé-Buzo-Gray algorithm* (LBG) [18]. This algorithm is also known as *generalized Lloyd algorithm* (GLA) or *K-means algorithm* and is based on an iterative use of two concepts:

1- Each input vector shall be encoded into its closest codevector.

2- The optimum codevector assignment for each cell is the centroid of all input vectors being encoded to that cell.

1.3 Channel Coding

Channel coding[3] consists of various methods that add some protection to the message, once passed from the source coding process. This is done by adding some redundancy to the message which will be used later in the channel decoder to detect and to correct the errors due to the channel noise.

There are two main groups of channel coders : the block coders and the convolutional coders. A binary block channel coding, denoted by (n, k, d_{min}), is a collection of 2^k *code words*, each consisting of n binary elements. Roughly speaking, among 2^n possible code words, just 2^k information words are allowed to be transmitted. The channel decoder can receive any of the 2^k code words but is supposed to extract the best information word if the received code word is not an information word, itself. The *code rate* is defined as the ratio : $R_c = \frac{k}{n}$.

The minimum *Hamming distance* of any two code words in a channel coder defines the *minimum distance* (d_{min}). The number of bit errors that a given channel coding is capable to correct, can be obtained from the following relation :

$$t \geq \frac{d_{min}}{2} \tag{5}$$

[3] Also known as *error protection* and *error correcting coding* (ECC).

A code word $(\mathbf{C})$ can be obtained from an information word $(\mathbf{X})$ using the *generator matrix* $(\mathbf{G})$ [21] :

$$\mathbf{C} = \mathbf{X} \times \mathbf{G} \tag{6}$$

$$[c_1 c_2 ... c_n] = [x_1 x_2 ... x_k] \times \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{bmatrix} \tag{7}$$

The decoding is done using a *parity check matrix* $(\mathbf{H})$. If the received block is the same as the transmitted block, the product $\mathbf{S} = \mathbf{C}' \times \mathbf{H} = \mathbf{C} \times \mathbf{H} = \mathbf{0}$; otherwise the decoder has to search for the best information code that minimizes the expectation of error. This task can be quite difficult, specially in the case of non *perfect codes* [21]. In fact, channel decoding acts very similar to the vector quantization. Both method search for the optimum "code", among some restricted ones in the total "code space".

It is necessary to be noticed that some coding systems (MORVQ, for example) do not have a separate channel coding block. In fact, in these systems the role of channel coding is played by an appropriate structure of source coder indices.

2 Hierarchical Protection

One way to maintain the performance in the noisy environment transmission is to better protect the more sensitive information bits which are suspected to contribute to greater errors. This method is known as *unequal error protection* (UEP) in the literature. Another use of the hierarchy of information will be discussed in section 5.2, page 12.

As an example one can mention an LPC vocoder. The human auditory system is more sensitive to pitch and voicing errors than the errors in the other LPC parameters. In the LPC-10 algorithm [24, page 268], pitch and voicing are encoded so as to prevent single-bit transmission errors from causing gross pitch and voicing perturbations, while no channel coding is provided for the other parameters.

As another example, in one realization of the CELP vocoder, the most significant bits of the binary representations of the codevectors are more sensitive to channel errors than the least significant bits. This property has been used to protect only the most significant bits [23].

Of course, one can imagine a progressive use of channel coders : use the very simple channel coders (even none at all) for the least sensitive bits and the stronger channel coders for more sensitive bits. This approach can be employed in networking problems where many types of data with different sensitivities to noise are to be transmitted. In [11] an example of such a system is explained : for each bit, a factor of sensitivity to channel error is defined. Using this factor, the optimal error rate allowed for each bit that minimizes the effects of channel noise, is estimated. Finally, a UEP coder is used to achieve different levels of protection.

3 Index Assignment

Index assignment of the codevectors does not affect the average distortion, in the absence of channel noise, while in the presence of channel noise, this assignment plays an important role in determining the overall VQ performance. Basically, LBG does not provide any protection against channel noise because any change of bit can redirect one codevector to any other one in the codebook. So, even a low *bit error rate* (BER) can heavily distort the signal if no index assignment strategy is used.

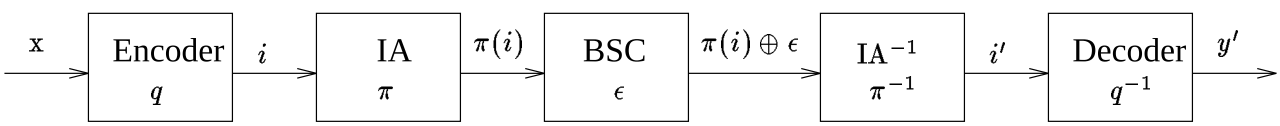
In effect, a VQ used in real circumstances, where noise exists, has to be reinforced. One way of such reinforcing is to provide some structure to the codebook where the codevectors that are more supposed to be confused represent close codevectors. This is called *pseudo Gray coding* in the literature[33] and can be achieved by a good index assignment.

It must be noticed that the IA is an *non polynomial* (NP)-complete task since there are $\frac{(2^b)!}{2^b \times b!} = \frac{(2^b-1)!}{b!}$ possible distinct combinations to assign $L = 2^b$ codevectors to L codewords. The 2^b and the $b!$ factors in the denominator eliminate respectively the symmetric cases and the bit permutation cases. This results 8.3×10^{499} distinct possible combinations for $b = 8$ bits.

Farvardin has observed [8] that when the *splitting*[4] technique [18] is used for VQ training, the resulting codebook has a *natural* ordering that can somehow protect the signal in the presence of channel errors. This is due to the splitting mechanism which makes *sister* codevectors behave similarly. However, this is not entirely efficient because if an error occurs on the first split bits, the resulting distortion can be much greater.

A general solution to the IA problem is to perform the VQ design first and then permute the indices in such a way that the resulting codebook becomes more robust against channel noise[5]. It is shown in [6] that a non negligible reduction in distortion can be obtained through a well designed IA rather than a random one.

The problem can be formulated simply as explained in figure (2):



π : Permutation Function.

$$\mathbf{y}' = q^{-1}\{\pi^{-1}[\pi(q(\mathbf{x})) \oplus \epsilon]\}$$

Figure 2: *Block diagram of the VQ based coding system used over a noisy channel.*

Two methods will be discussed in this cadre : simulated annealing and binary switching algorithm.

---

[4]That is to begin the training with a few (possibly just one) codevectors and dividing each codevector into two sister codevectors gradually, with the small perturbations.

[5]The other method consists of simultaneous optimization for source and channel, where the IA is included in the encoding process and will be discussed in the next section.

3.1 Simulated Annealing

Since IA is an NP-complete problem, Farvardin used *simulated annealing* (SA) to solve it [8]. SA is a Monte Carlo algorithm which has been widely used to solve combinatorial problems [14]. It imitates the physical process of annealing which finds a lower energetic equilibrium for the crystallization of steel.

An appropriate *temperature* variable, T, is to be defined. This variable is initialized to a high value[6], T_m, in the beginning of the process and is decreased progressively until a sufficiently small value[7], T_f, is reached. A high value of T signifies a high degree of randomness while a low value of it means that nothing is left at random. A high value of T at the beginning of the process, permits to avoid many local optima.

The SA algorithm can be written as follows [14]:

Initialization :

- An initial state for IA is given : $S = S_0$.
- Temperature is initialized to the melting temperature : $T = T_m$.

Iteration :

- Randomly choose another state, S', as a perturbation of the last state, by changing the assigned codewords of two codevectors (both codevectors are chosen randomly).
- Let $\Delta D_c = D_c(S') - D_c(S)$.
- If $\Delta D_c < 0$ then $S = S'$.
- Otherwise replace S by S' with probability $e^{-\Delta D_c / T}$.
- Slightly decrease T.

Termination :

- If $T = T_f$, or a stable state is reached, end.
- Otherwise continue the iterations.

The SA algorithm can theoretically give the global optimum solution, unconditionally on the initial state, provided that the initial value, T_m, and the schedule of decreasing T, are chosen appropriately. Unfortunately, this is difficult to achieve and therefore good optima from SA might be difficult to obtain in most practical cases.

As an example, Farvardin has reported a *signal-to-noise ratio* (SNR) of about 8.95 dB for SA, compared to 8.87 dB for a naturally organized LBG with splitting. The test parameters were : $\epsilon = 10^{-2}$, $N = b = 8$ bits for a first order Gauss-Markov source with the correlation coefficient, $\rho = 0.9$.

3.2 Binary Switching Algorithm

Another algorithm for an improved IA was proposed by Zeger and Gersho [33]: *binary switching algorithm* (BSA). In BSA, to each codevector $\mathbf{y}$ is assigned a cost function $C_\pi(\mathbf{y})$. This cost function

[6] Melting temperature (10.0 in the example of [8]).
[7] Freezing temperature (2.5×10^{-4} in the example of [8]).

is a measure of the contribution to the total distortion due to the possible channel errors when $\mathbf{y}$ is decoded, assuming a certain permutation, π. Then the codevectors are sorted in decreasing order of their cost values. The vector that has the largest cost, say $\mathbf{y}$, is selected as a candidate to be switched first.

A trial is conducted : $\mathbf{y}$ is *temporarily* switched with each of the other codevectors to determine the potential decrease in the total distortion $D_\pi = \sum_{k=0}^{L-1} C_\pi(\mathbf{y}_k)$, following each switch. The codevector which yields the greatest decrease in D_π when switched with $\mathbf{y}$ is then switched *permanently* with it. The algorithm is then repeated for the next highest cost and so on.

Although a global optimal IA is not necessarily obtained by BSA, good locally optimal solutions have been reported [33]. Simulation tests have been made with a first order Gauss-Markov source as well as an *independent identical distribution* (iid) and speech waveform. As an example, for $\epsilon = 10^{-2}$, $N = 4$, $b = 8$ bits, 1.5 dB gain has been achieved compared to the initial state.

4 Simultaneous Optimization of VQ and IA

We mentioned previously what the state of the art is to assign indices to the codevectors of a given vector quantizer when they are to be transmitted over a noisy channel. We now turn our attention to the direct design of a vector quantizer which is intended for use over a noisy channel. Figure (3) illustrates the block diagram for this situation.

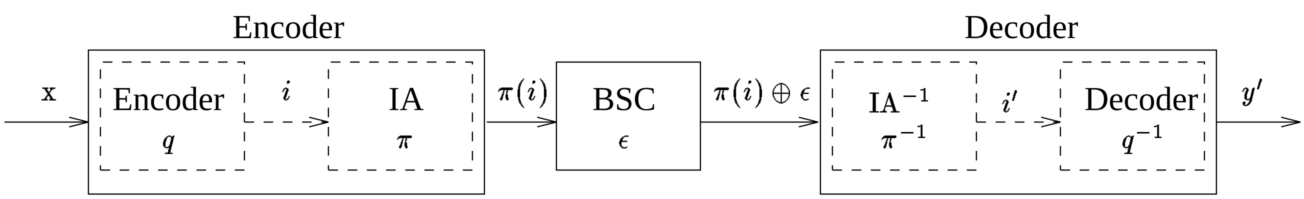


Figure 3: *Block diagram of the VQ based coding system used over a noisy channel. Here, the IA is included in the encoding process.*

We will examine two methods here that are based on simultaneous optimization of quantizer and IA : "channel optimized vector quantization" which is a generalization of LBG for the noisy channel transmission and "self organizing hyper cube" which is a generalization of Kohonen map into higher dimensions.

4.1 Channel Optimized Vector Quantization

Farvardin proposed a joint optimization for the source and the channel coders [9, 10]. It is in fact a generalization of the same methodology used for optimum scalar quantizer design in [7]. This optimization is performed by presenting a modified distortion measure involving both the quantization error and the error due to channel perturbation [2]:

$$D = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \sum_{j=1}^{L} p(j|\pi(i)) \times d[\mathbf{x}(n), \mathbf{y}_j(n)],\ \mathbf{x}(n) \in \mathcal{C}_i \quad (8)$$

This distortion measure leads to a simultaneous optimization for source and channel. The resulting algorithm is very similar to the LBG algorithm and is named *channel optimized vector quantization* (COVQ). The cells, $\mathcal{C}_i$, are updated according to the following equation [10]:

$$\mathcal{C}_i^* = \{\mathbf{x} : \sum_{j=1}^{L} p(j|\pi(i)) \times ||\mathbf{x} - \mathbf{y}_j||^2 \leq \sum_{j=1}^{L} p(j|\pi(l)) \times ||\mathbf{x} - \mathbf{y}_l||^2, \forall l\},\ i \in \{1, ..., L\} \quad (9)$$

That is, each input vector $\mathbf{x}$ is classified into the cell with the least *expectation of distortion* : this equation is referred to as *generalized nearest neighbor.* The codevectors, y_j, are updated according the following equation:

$$\mathbf{y}_j^* = \frac{\sum_{i=1}^{L} p(j|\pi(i)) \int_{\mathcal{C}_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\sum_{i=1}^{L} p(j|\pi(i)) \int_{\mathcal{C}_i} p(\mathbf{x}) d\mathbf{x}},\ j \in \{1, ..., L\} \quad (10)$$

The term $\mathbf{y}_j$ represents the centroid of all input vectors that are decoded into $\mathcal{C}_j$, even if the transmitted index, i, is different from j. So this equation is called *generalized centroid.* It can be noted that these two equations can be simplified into the LBG learning equations by simply assuming that:

$$p(j|\pi(i)) = \begin{cases} 1 & : \quad j = \pi(i) \\ 0 & : \quad j \neq \pi(i) \end{cases} \quad (11)$$

This way, LBG can be regarded as a special case of COVQ when transition probability, $\epsilon = 0$.

It is shown that the obtained optimum encoding cells are convex polyhedrons and that some cells might vanish thus creating *empty cells* [9]. This means that the system trades quantization accuracy for less sensitivity to channel noise. Figure (4) shows an example of COVQ for a two-dimensional (N=2), three-level (L=3) VQ and a *discrete memoryless channel* (DMC) with the parameters as in the following Table.

Transition matrix $P(i|j)$ in the DMC example.

$i\|j$	1	2	3
1	$1-2\epsilon$	ϵ	ϵ
2	2ϵ	$1-4\epsilon$	2ϵ
3	ϵ	ϵ	$1-2\epsilon$

This figure illustrates that the higher the channel noise is, the greater is the risk that some cells vanish. Assuming that there are L' nonempty encoding cells, $L' \leq L$, only L' codewords need to be transmitted. Of course, any of L binary codewords may be received and therefore the codebook must remain of size L. It is interesting to observe the analogy that exists between the presence of empty cells (codevectors with no corresponding input vector) and the added redundancy in channel coding.

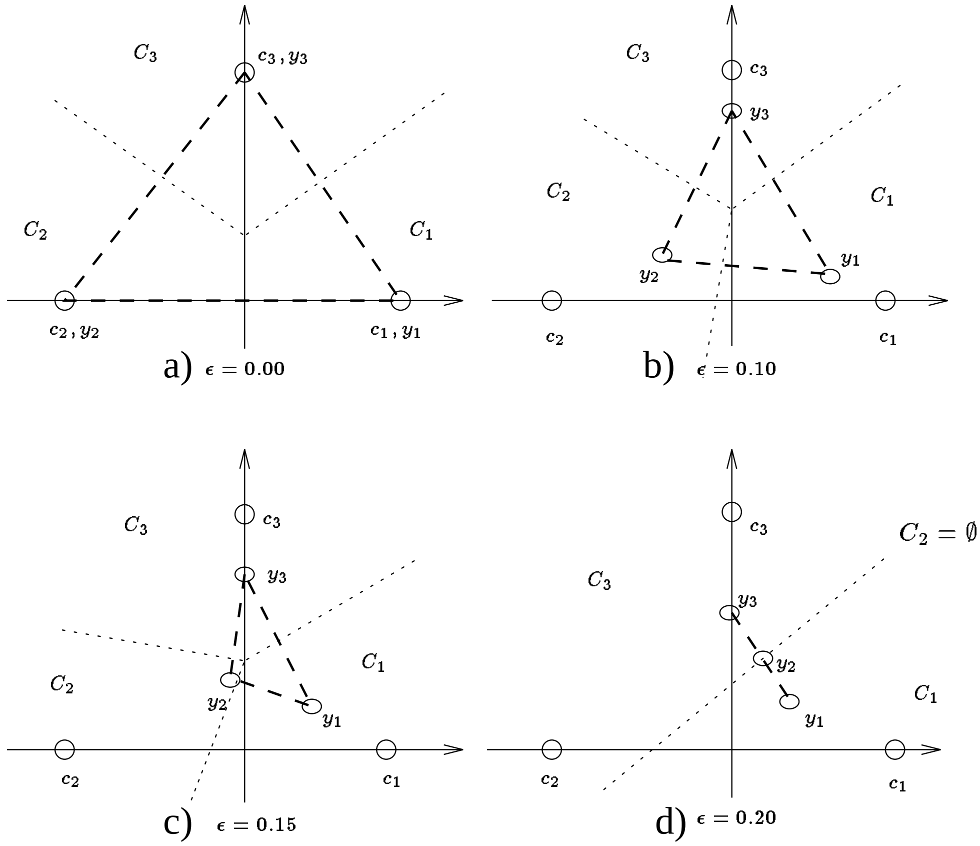


Figure 4: *Figures (a), (b), (c) and (d) show the quantization cells for* $\epsilon = 0.00, 0.10, 0.15$ *and* 0.20, *respectively for a simple DMC. The codevectors get closer when* ϵ *increases and finally one of the cells,* $\mathcal{C}_2$, *vanishes for* $\epsilon = 0.20$. *The* c_i *are the codevectors for a non noisy environment.*

Simulations have been reported [9, 10] for first order Gauss-Markov sources, as an example, for $\rho = 0.9$, $\epsilon = 10^{-2}$, $N = b = 8$. COVQ and naturally organized LBG with splitting have resulted in 9.70 dB and 8.87 dB, respectively. COVQ has had $L' = 26$ empty cells (out of $L = 256$), in this example.

4.2 Self Organizing Hyper Cube

A direct mapping from the input space to the Hamming space has been proposed in [31]; this mapping is roughly a b-dimensional generalization of the 2-dimensional Kohonen map [15, 16], hence it is named as *self organizing hyper cube* (SOHC). Kohonen map is also known as *self organizing map* and *competitive map*, because it is an iterative algorithm to design a VQ, which keeps a topological similarity of the input space in the coding space.

SOHC is trained with an algorithm similar to the Kohonen algorithm with some modifications: the codevectors are arranged in a b-dimensional cube (instead of a 2-dimensional map); the neighborhood function is defined in the hyper cube and Hamming distance (d_H) is used as the distance measure of the binary representations of the indices (instead of Euclidean distance in the Kohonen map).

As a result of such a definition of distance between the codevector indices, in SOHC, there is almost no difference between the quantized bits. In other words, the least and most significant bits have no sense in SOHC. Roughly speaking, the effect of noise on each bit is almost the same.

Examining figure (5), if we consider that the chosen codevector to be transmitted is 0000, a single bit of error can commute it to either of 1000, 0100, 0010 or 0001. Since all these codevectors are the first order neighbors of 0000 (with $d_H = 1$), this commutation does not contribute a gross error.

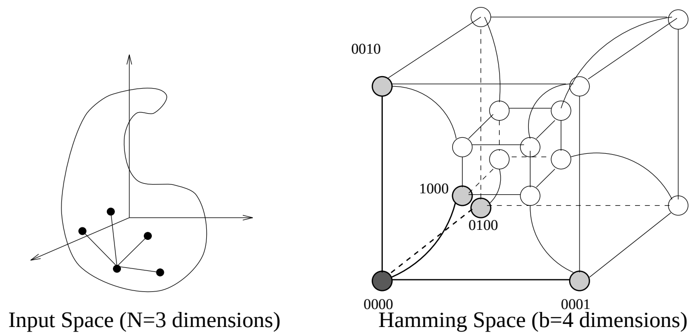


Figure 5: *An example of SOHC. Left : input space. Right : SOHC. Codevector 0000 and its first order neighbors are highlighted in both spaces.*

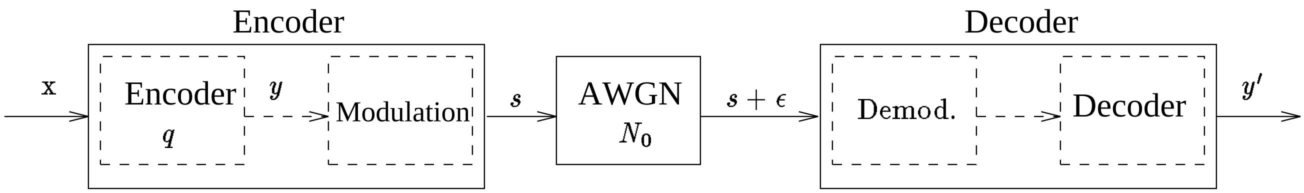
Adding the splitting technique to SOHC, improves further its performance [32]. In SOHC with splitting, each time that the codewords are split, the dimension is increased, too. SOHC has been tested for quantizing and transmitting *log area ratio* (LAR) [30] parameters of speech, over a BSC. Better objective results were reported, compared to naturally organized VQ and Kohonen map, specially for high transition probabilities.

For instance, with a transition probability $\epsilon = 10^{-2}$, $N = 10$, $b = 8$, the *spectral density distortion* (SD) [12] measure for SOHC, Kohonen map and naturally organized LBG with splitting were about 3.3 dB, 3.4 dB and 3.5 dB, respectively. With SOHC, a further protection is also possible, using some classic error control coding technique, since SOHC provides the bit patterns in which all the bits are (almost) equally likely to cause error.

5 Direct Modulation Organizing Scheme

Another possible source-channel configuration is the direct modulation organization. In this configuration, the encoder includes the modulator and benefits directly from the flexibility that is naturally present in a constellation. As shown in figure (6), the channel is considered with an *additive white Gaussian noise* (AWGN).

Several works have been done in this field. To mention some, we can indicate a competitive learning algorithm which gives aft direct mapping from input space to the signal space is presented [27]; the hierarchical modulation, in which the constellation points are located to minimize the error expectation is explained [25, 5]. There exists some other works that we will not extend in this paper : joint optimization of three blocks (source coder, channel coder and modulator) [28]; Trellis coding and Lattice coding which are special kinds of covering the signal space by the constellation points [1, 17].



$$s \in S = \{\text{Set of all Constellation Points}\}$$

Figure 6: *Block diagram of the Direct Modulation VQ based coding system, used over an AWGN channel. The source encoder, channel encoder and modulator are represented in one block.*

5.1 Modulation Organized VQ

Withdrawing any binary representation, Skinnemoen proposed the *modulation organized vector quantization* (MORVQ) [27]. This method uses a quantizer which maps the codevectors directly into the constellation plane. It makes efficient use of the Kohonen learning algorithm to map the N-dimensional input space to the 2-dimensional signal space, in such a manner that the close codevectors in the modulation space, are assigned to the close points in the input space. This property is obtained by proper use of a neighborhood function [15, 16] and the resulting codebook has some organized structure. Having this structure, most little changes due to channel noise make the output codevector to be one of the neighbors of the source vector and so the distortion will not be very important.

Skinnemoen observed a great difference between explicit error protection and the structure of a codebook. He states that any transmission system (with or without error protection) has a BER working threshold. Above that limit, the system's performance breaks down. The role of MORVQ is to increase this threshold. This is the great advantage of MORVQ; however, in MORVQ, no more channel coding can be added since it does not produce any intermediate bit pattern which can be processed by the channel coding.

Good numerical results have been reported in quantizing first order Gauss-Markov sources and *line spectrum pairs* (LSP) [13] parameters of speech spectrum in an AWGN channel. As an example, for quantizing LSP parameters with $N = 10$ and $L = 256$, SD was 2.11 dB and 7.82 dB, respectively for MORVQ and LBG, for a highly noisy channel. Also it is observed that for MORVQ, the degradation curve by increasing channel noise is rather smooth while for LBG there is a threshold above which the system performance drops rapidly.

5.2 Hierarchical Modulation

Ramchandran et al. have proposed in [25] a Multi-Resolution broadcast system. One basic idea in their proposition consists in partitioning the information into two parts : the coarse information and the refinement or the detail information[8]. The coarse information is to be received correctly

[8]Many transformations, like subband and wavelet coding have a natural multiresolution interpretation.

even in a very noisy transmission environment, while the detail information is mostly destined to the receivers whose channels have better qualities (graceful degradation)[9].

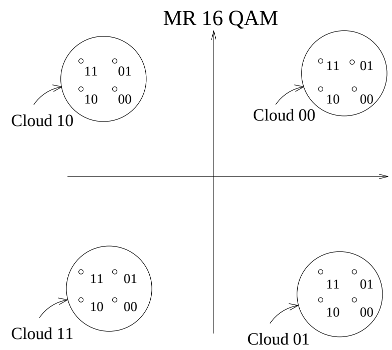


Figure 7: *An example of Multi-Resolution Constellation. Each set of close points constitutes a cloud with four satellites points surrounding it. The detail information is presented in the satellites while, the course information is represented in the clouds. So there is 2 bits of coarse information and 2 bits for detail. Note also that the Gray code is used for numbering the satellites (and the clouds) in such a way that the codewords with Hamming distance equal to 2 are far from each other. This is like the application of Karnaugh map in digital design and can be used for larger constellations, too.*

As an efficient end-to-end broadcast have its transmission constellation matched to its source coding scheme, they proposed a multi-resolution constellation as depicted in figure (7). The coarse information is carried by the clouds, while inside each cloud, the mini-constellations or *satellites* provide the details. The loss of coarse information is associated with the receiver inability to decipher correctly which cloud was transmitted while the loss of detail information occurs when the receiver confuses one intra-cloud signal point for another. Of course, many other configurations could be thought which yield the same property and one has to choose the best configuration according to the presented problem. The same idea has been used in conjunction with *Trellis modulation coding* (TMC) as well as with embedded channel coding [25].

Combelles et al. [5] have used the same idea of multi resolution coding, in conjunction with Turbo code [4], which was used to protect the coarse and detail information with $\frac{1}{2}$ and $\frac{3}{4}$ rates, respectively. They achieved 4 dB better performance for the coarse information while 2 dB degradation for the detail information, compared to a single resolution system using Turbo code, with the same overall spectral efficiency to obtain the same error rate (10^{-4}), while for a Rayleigh fading channel their simulation shows 5 dB of better performance for the coarse information and 3 dB degradation for the detail information.

In [25] an example is given where with a multi resolution system, the broadcast coverage radius (64 km) is much greater than for a single resolution system (45 km) while for the multi resolution system, the radius of full data availability is a little smaller (38 km).

[9]This classification can be made more precise, making several classes of importance.

6 Rate/Distortion Source-Channel

A special case is considered : a memoryless source with uniform pdf is to be coded and transmitted over a BSC. The uniformity of source, does not permit too much for source coding (except the dimensionality [19]). However, we will see that since different bits have different contributions to the total error, it is rather reasonable to send different bits with variable compressing and/or protection rate.

The originality of the work to be presented here is that it considers each sample, bit by bit and performs the compression and the protection operations on them separately. To do so, the blocks of n_1 *most significant bits* (msb) are grouped together; the blocks of n_2 bits from the next row, until n_N *least significant bits* (lsb).

This hierarchical combination of source and channel coders is depicted in figure (8). Each block is either a source coder or a channel coder (it can also be just a close connection or an open one, as the extreme cases of coder). As the importance of bits augments (for the most significant bits, for example), one expects that a stronger channel coding scheme is to be employed. Contrarily, for the less important bits (the least significant bits, for example), a powerful source coding is used.

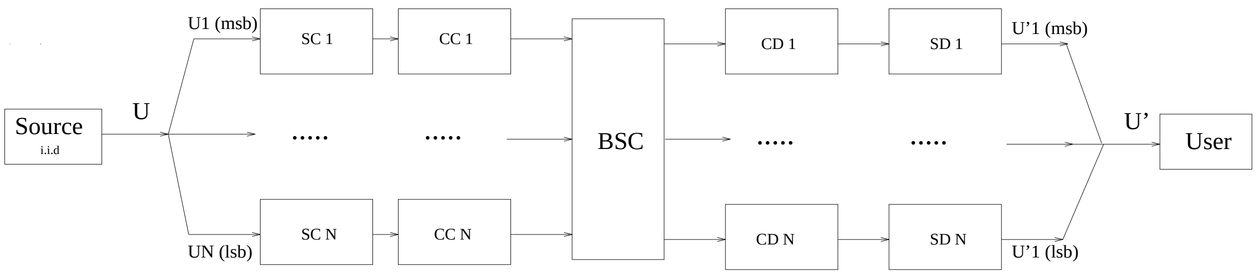


Figure 8: *Source-Channel coder combination. Each line of bits (msb, ..., lsb) is treated separately.*

Although the source coding and the channel coding in this structure are performed separately, we consider it as a combined source-channel coding since the optimization constraints are applied simultaneously to both coders.

This system was optimized for a limited number of possible source and channel decoders. To simplify the implementation, we considered just *Hamming* and *repetition* codes as the channel encoder and the *inverse Hamming* and *Majority vote* (inverse of repetition) codes as source encoder.

Seeing the duality that exists between channel coding and source decoding, also between source coding and channel decoding, we have considered the same channel decoders as source coders and the same channel coders as source decoders.

Considering five different repetition coders ($R_{3,1,3}$, $R_{5,1,5}$, $R_{7,1,7}$, $R_{9,1,9}$, $R_{11,1,11}$) and eight different Hamming coders($H_{7,4,3}$, $H_{15,11,3}$, $H_{31,26,3}$, $H_{63,57,3}$, $H_{127,120,3}$, $H_{255,247,3}$, $H_{511,502,3}$, $H_{1023,1013,3}$), the total number of possible combinations grows too rapidly with the number of used bits. Fixing the maximum number of bits to 10, there will be $((5+8+2)^{10})^2 = 3.32 \times 10^{23}$ possible combinations. This makes an exhaustive search very hard to be done.

The optimization was done with the use of the *bit allocation algorithm* (BAA) [22]. It fits a polyline on the R/D plane as shown in figure (9) which covers all the possible combinations. For a highly condensed cloud, BAA gives the envelope of the permitted region. It works with an iterative use of a subroutine which gives one point on the envelope corresponding to a given tangent to the envelope, λ. BAA searches for the optimum λ which yields an acceptable bit rate and maximum SNR, like a binary search.

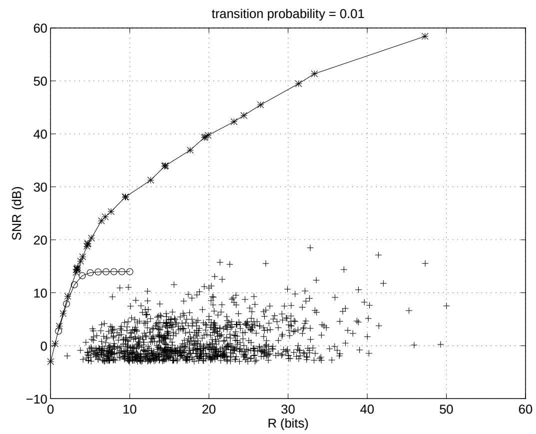


Figure 9: *The cloud of possible combinations. The curve with "*" shows the attainable bound and the curve with "o" shows the performance of a simple system without any SC nor CC for 1, 2, ..., 10 bits.*

This algorithm does not guarantee the global optimum for non condensed codes, however it gives very good results. For example, the exhaustive search and BAA resulted, respectively, 16.47 dB and 16.53 dB SNR, for $\epsilon = 10^{-2}$ and *desired bit rate*, $R_d = 4$ bits. The cloud of possible combinations is also illustrated in figure (9) for the same example. The points of the cloud are selected randomly with an independent uniform probability for each block. To draw the figure, 1000 sample configurations were used among many possible points and give an idea for the possible region in the R/D plane.

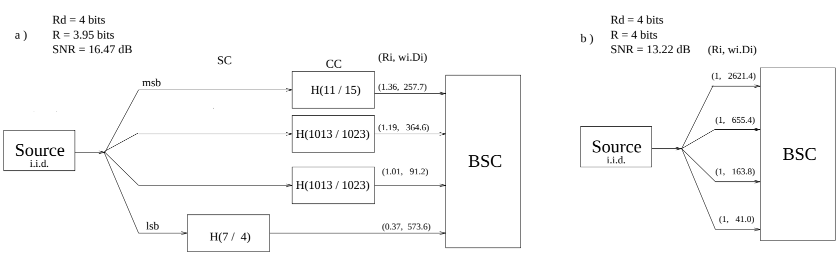


Figure 10: *The optimum system of coders for $\epsilon = 10^{-2}$ and $R_d = 4$. The pair of numbers in parenthesis show the (R_i, D_i), the bit rate and the contribution to distortion due to each line.*

Figure (10) illustrates the result of optimization for $\epsilon = 10^{-2}$ and $R_d = 4$. One can observe that there is a tendency to equalize the distortion due to each line, D_i. In fact, the space of used (both source and channel) coders is very sparse, otherwise, one could expect a much more equalized error contribution of all lines. The distribution of bit rate to each line, R_i, is inversely proportional to the line number i. So : $R_i \geq R_j \Leftrightarrow i \leq j$.

The performance of the proposed optimized system is compared to two other system performances in figure (11) : a system without any coding at all and a system with only channel coding and the same channel coding for all bits. All the systems run at six bit/sample rate.

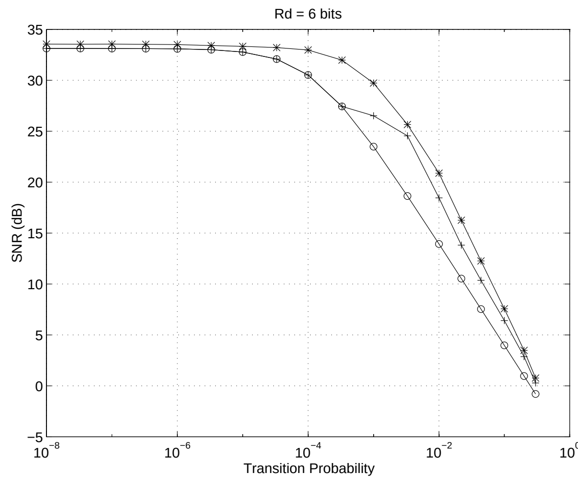


Figure 11: *The system performances : "*" rate/distortion optimized system; "o" without any coding; "+" with the same channel coding for all the lines.*

As shown in the figure, the two last systems have practically the same performances for transition probabilities below $\epsilon < 2 \times 10^{-4}$ while the performances of the proposed optimizations are always at least 1 dB above the two others and the maximum gain is about 6 dB. It is also noticeable that even for non noisy channels ($\epsilon \approx 0$) the proposed algorithm provides some gain.

It is also necessary to indicate that although the model used for this system is very simple, a real optimization is searched and further improvements for more complex models are to be done. In a special direction, the coder space can be also extended to more powerful codes which will lead a more effective optimization.

7 Conclusion

A survey of methods of combined source-channel coding is presented in this paper. The different methods have been classified : first, one that exploits the unequal sensitivity of different symbols and hence economize bit-rate for the most sensitive ones; second, those that rearrange a codebook by permuting its codevectors; third, those that are the generalizations of the optimum source coders in noisy conditions; fourth, those that exploit the flexibility offered in the modulation and match the modulation space with the source input space, and fifth, those that exploit the redundancy in different bit streams of a binary source. This last system is a novel contribution.

The following table summarizes some informations about these methods : the most important references used, the considered source and channel and finally the error criterion used.

Survey of the reviewed methods.

Algorithm	ref.	Source	Channel	criterion
Hier. Prot.	[11]	Image	BSC	SNR
SA	[8]	Gauss-Markov	BSC	SNR
BSA	[33]	Gauss-Markov	BSC	SNR
		Speech (Wave)	BSC	SNR
COVQ	[8, 9, 10]	Gauss-Markov	BSC/DMC	SNR
SOHC	[31, 32]	Speech (LAR)	BSC	SD
MORVQ	[27]	Gauss-Markov	AWGN	SNR
		Speech (LSP)	AWGN	SD
Hier. Mod.	[5, 25]	Image (HDTV)	AWGN	Packet loss
Rate/Dist.	new work	BSS	BSC	SNR

A universal solution that can satisfy all the problems is not known. Nevertheless, for each problem, one can find a solution that is better matched to the application among the various proposed methods.

Acknowledgment We would like to thank Ch. Pepin and J.Y. Chouinard for their precious comments and suggestions on the primary version of this paper.

References

[1] J.P. Adoul, "Lattice and Trellis Coded Quantization for Efficient Coding of Speech", in *Speech Recognition and Coding : New Advances and Trends*, Springer 1995, pp. 405-422.

[2] E. Ayanoğlu, R.M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders", *IEEE Trans. on Information Theory*, vol. 33, no. 6, Nov. 1987, pp. 855-865.

[3] T. Berger, *Rate Distortion Theory : A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.

[4] C. Berrou, A. Glavieux, P. Thitimajshima, "Near Shannon Limit Error Correcting Coding and Decoding : Turbo-Codes", *Proc. of the International Conference on Communications*, Geneva, Switzerland, May 1993, pp. 1064-1070.

[5] P. Combelles, D. Castelain, C. Berrou, "Système de Codage Canal Hiérarchique pour la Télévision Numérique", *Quatorzième colloque GRETSI*, Juan-les-Pins, France, Sep. 1993, pp. 615-618.

[6] J.R.B. De Marca, N.S. Jayant, "An Algorithm for Assigning Binary Indices to the Codevectors of a Multidimensional Quantizer", *Proc. of the IEEE International Conference on Communications*, Seattle, WA., June 1987, pp. 1128-1132.

[7] N. Farvardin, V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels : An Approach to Combined Source-Channel Coding", *IEEE Trans. on Information Theory*, vol. 33, no. 6, Nov. 1987, pp. 827-838.

[8] N. Farvardin, "A Study of Vector Quantization for Noisy Channels", *IEEE Trans. on Information Theory*, vol. 36, no. 4, July 1990, pp. 799-809.

[9] N. Farvardin, V. Vaishampayan, "On the Performance and Complexity of Channel Optimized Vector Quantizers", *IEEE Trans. on Information Theory*, vol. 37, no. 1, Jan. 1991, pp. 155-160.

[10] N. Farvardin, "Speech Coding Over Noisy Channels", in *Speech Recognition and Coding : New Advances and Trends*, Springer 1995, pp. 387-404.

[11] K. Fazel, J.J. Lhuillier, "Application of Unequal Error Protection Codes on Combined Source-Channel Coding of Images", *Proc. of the International Conference on Communications*, 1990, pp. 320.5.1-320.5.6 (0898-0903).

[12] R.M. Gray, A. Buzo, A.H. Gray, Y. Matsuyama, "Distortion Measures for Speech Processing", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 28, No. 4, pp. 367-376, Aug. 1980.

[13] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals", *Journal of Acoustic Society of America*, Vol. 57, 1975.

[14] S. Kirckpatrick, C.D. Gelatt, M.P. Vecchi, "Optimization by Simulated Annealing", *Science*, Vol. 220, pp. 671-680, May 1983.

[15] T. Kohonen, *Self Organization and Associative Memory*, Springer-Verlag, 1984.

[16] T. Kohonen, "The Self-Organizing Map", *Proc. of the IEEE*, Vol. 78, pp. 1468-1480, Sep. 1990.

[17] R. Laroia, N. Farvardin, S.A. Tretter, "On Optimal Shaping of Multidimensional Constellation", *IEEE Trans. on Information Theory*, vol. 40, No. 3, July 1994, pp. 1044-1056.

[18] Y. Lindé, A. Buzo, R.M. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Trans. on Communications*, Vol. 28, No. 1, pp. 84-95, Jan. 1980.

[19] J. Makhoul, S. Roucos, "Vector Quantization for Speech Coding", *Proc. of the IEEE*, Vol. 73, No. 11, pp. 1551-1588, Nov. 1985.

[20] J.L. Massey, "Joint Source and Channel Coding", in *Communications Systems and Random Process Theory*, ed. Sijthoff, Noordhoff, Alpheh aan den Rijn, the Netherlands, pp. 279-293, 1978.

[21] R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, 1977, in French.

[22] N. Moreau, *Techniques de Compression des Signaux*, Masson, 1995.

[23] K.K. Paliwal, 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, Jan. 1993.

[24] T.W. Parsons, *Voice and Speech Processing*, McGraw-Hill, 1987.

[25] K. Ramchandran, A. Ortega, K.M. Uz, M. Vetterli, "Multiresolution Broadcast for Digital HDTV Using joint Source/Channel Coding", *IEEE Trans. on Selected Areas in Communications*, Vol. 11, No. 1, pp. 3-14, Jan. 1993.

[26] C.E. Shannon, "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27, no. 3-4, pp. 379-423 and 623-656, July and Oct. 1948.

[27] P.H. Skinnemoen, *Robust Communication with Modulation Organized Vector Quantization*, Thesis, Norwegian Institute of Technology, 1994.

[28] V.A. Vaishampayan, N. Farvardin, "Joint Design of Block Source Codes and Modulation Signal Sets", *IEEE Trans. on Information Theory*, July 1992, pp. 1230-1248.

[29] S. Vembu, S. Verdú, Y. Steinberg, "The Source-Channel Theorem Revisited", *IEEE Trans. on Information Theory*, vol. 41, No. 1, pp. 44-54, Jan. 1995.

[30] R. Viswanathan, J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 23, No. 3, pp. 309-321, June 1975.

[31] S.B. Zahir Azami, B. Azimisadjadi, V. Tabatabaii, H. Borhani, "Self Organizing Hyper Cube and Some of Its Applications", *Proc. of the First Iranian Conference on Electrical Engineering*, Tehran, Iran, May 1993, pp. 464-472.

[32] S.B. Zahir Azami, G. Feng, *Quantification Vectorielle pour des Canaux Bruités*, Report of DEA, INPG (Grenoble), June 1995, in French.

[33] K. Zeger, A. Gersho, "Pseudo-Gray Coding", *IEEE Trans. on Communications*, Dec. 1990, pp. 2147-2156.