Signal Processing for Multimedia
J.S. Byrnes (Ed.)
IOS Press, 1999

Robust Image and Video Communication for Mobile Multimedia

Tor A. Ramstad

Norwegian University of Science and Technology (NTNU)
Department of Telecommunications
O.S. Bragstads pl. 2
N-7034 Trondheim
Norway
tor@tele.ntnu.no

ABSTRACT. What characterizes mobile multimedia is multiple signal sources, a multitude of applications, and transmission over channels with limited bandwidth and power, plus varying noise and attenuation. To cope with these variabilities new and improved communication systems should be developed. Joint source-channel coding may be one contribution towards this end. This article suggests an efficient communication principle where the sampled analog signal is first decomposed and thereafter mapped to the channel space using dimension-reducing nonlinear mappings. The article discusses the optimality of such an approach for artificial sources, but also shows simulation results for practical still image and video sources. Graceful degradation as a function of channel deterioration is shown to result from the proposed method.

1. Introduction

A slogan in the communication industry is "Communication everywhere at anytime to anybody". To achieve this end the user terminal must be mobile. When we also include multimedia services, we get the term *mobile multimedia* systems. Multimedia communication will contain signals of very different origin and characteristics. There will be analog sources like speech, music, still images, and video, and numerical sources like text, measurement data etc. Traditionally, all these signal types are transmitted over channels using the same modulation method and the same error correction type irrespective of the signal origin and its application. This is simple from an overall system perspective as the channel does not have to worry about the origin of the source, nor its application. On the other hand, the source coder will see a virtually error-free channel which alleviates all concerns about bit-errors. This strict distinction between source and channel coding for analog sources is implied by Shannon's work [9, 10], where it is shown that these operations can indeed be optimized separately in the ideal case. This, however, requires infinite system delay and infinite complexity. So the question remains, is the situation different for practical solutions? Also, Shannon's results do not give any indications about what happens when the channel changes and the required error protection no longer works.

With the advent of multimedia services for which the signals are to be transmitted over a variety of channels, including mobile and other wireless media where bandwidth and power are scarce, we might have to reconsider the traditional system philosophy and opt for more flexible modem designs which are adaptive to the input signal and the application. This calls for new system optimizations which take the signal and channel coding into consideration simultaneously.

FIGURE 1. Signal chain for the transmitter part of the image communication system. $H_1(s)$ and $H_2(s)$ are anti aliasing and imaging component removal filters, A/D and D/A are the conversion circuits, and DSP is the digital processing unit.

The bottom line in this discussion is that when we allow lossy coding of e.g. images or video, there is no need to transmit the compressed signals without bit errors. It will be advantageous to allocate some of the allowed noise to the compression operation, but spend the rest on the transmission part. Of course, it is imperative that we design the systems to minimize both noise contributions. The question is how much of either type should be allocated to get the overall best performance.

The above is characteristic of a situation where the channel is completely known. Then some optimal distribution between noise sources may be obtained. In wireless scenarios, in particular when mobile transmitters or receivers are involved, or in broadcasting situations where the different receivers encounter different channel conditions, robustness becomes the significant consideration. Variations in the channel should not be disastrous, but graceful degradations could be tolerated when the channel deteriorates.

When several data types coexist on a transmission medium it is necessary, under the above assumptions, that different signals are transmitted using different channel alphabets. In contrast to the above, certain types of data cannot tolerate bit errors. This may be numerical data or even image data obtained by lossless compression. With this scenario two different alphabets could be used. Assuming that C codewords can be applied in some modulation scheme to obtain an acceptable channel noise for the analog source signals, then a smaller configuration, say with an alphabet of size αC ($\alpha < 1$), can be used for transmission of numerical data. The number α must be set to a value small enough to keep the error rate within acceptable limits for numerical data. This mode will certainly also contain error protection bits. Similarly, there is usually some side information generated from the source encoder that needs special attention. This type of information must not be destroyed and should therefore also be transmitted in the secure mode.

As the numerical data must be transmitted in a traditional way, we focus our attention on the transmission of analog signals.

2. Model and performance limits for analog source transmission

The transmission system under consideration is shown in Figure 1. It is an analog-to-digital-digital-to-analog system, or considered as an end-to-end system, an analog-to-analog system, hopefully having some signal compression ability to be exactly defined later.

To simplify the description the following modeling assumptions are made:

- The signal can be ideally represented by uniform sampling at the rate f_s, where $f_s = 2B$, and B is the signal bandwidth.
- The channel is an ideal Nyquist channel that can transmit at a rate f_c over a channel with bandwidth $f_c/2$.

This means that a sampled signal of bandwidth B can be transmitted over a channel of bandwidth B. Consequently, the time discrete signal requires the same bandwidth as the original analog signal. We shall use this simplification below for defining system *compression* or *efficiency*.

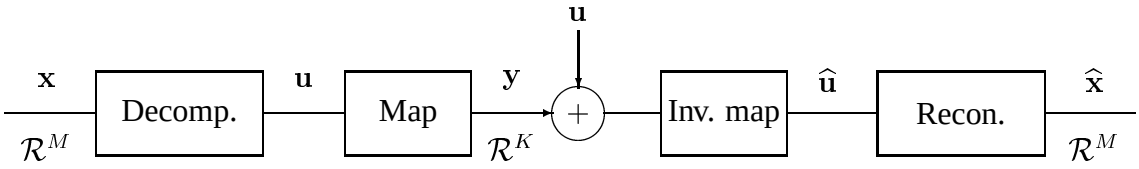

FIGURE 2. The idealized image communication system including signal decomposition, mapping from signal parameters to the channel representation, channel with additive noise, inverse mapping, and reconstruction.

We can, in practice, relax these idealized assumptions by incorporating some oversampling of the signal and introducing a roll-off factor for the channel filters to make them implementable.

A model composed of what is called the digital processor in Figure 1, the channel, and the inverse digital processor is shown in Figure 2.

On the transmitter side there are basically two parts:

- A signal decomposition unit.
- A mapping device.

As will become clearer below, the role of the decomposition unit is to cluster signal components into blocks with similar statistics. Compression is obtained through the mapping device, which takes a block of M source samples (collected into the vector $\mathbf{u}$) and maps it to the vector $\mathbf{y}$ consisting of K channel samples, where we require that, on the average, $K < M$ to obtain reduced bandwidth. This can be obtained in two ways:

- Discarding samples.
- Combining samples.

The methodology for sample combination will be given in Section 4.

2.1. *A measure for compression at a system level*

The system objective is to minimize the overall noise, as e.g.

$$\min E[||\mathbf{x} - \widehat{\mathbf{x}}||^2], \tag{1}$$

subject to a channel constraint, as e. g.

$$E[||\mathbf{y}||^2] \leq K\sigma^2_{c,max} \tag{2}$$

where $\sigma^2_{c,max}$ is the maximum average *energy per channel sample*. We have chosen to limit the channel power, although other possible limitations could be imposed as e. g. maximum amplitude or some combination of amplitude and power constraints.

The following observation is crucial for understanding why the suggested source-channel coding principle works: When the number of samples is reduced, two mechanisms alter the noise level in the received signal:

1. The sample reduction will inevitably generate approximation noise.
2. The channel noise per channel sample is the same as without compression. After decompression the number of samples is increased resulting in less channel noise per signal sample.

At a certain compression ratio these two mechanisms balance. The two contributions in combined compression/modulation are equal to the noise without any compression, *i.e.*, pure PAM transmission of the source samples. If this compression ratio is $\mathcal{K}$, this number can be defined as the system *advantage* or *compression*.

What implications do we foresee from the overall noise minimization? To reduce the impact from the channel, it seems like a good idea to make small channel errors result in small errors in the reconstructed signal. This is equivalent to saying that neighboring symbols in

the channel signal configuration must correspond to neighboring symbols in the source signal representation. The opposite, however, is not necessary.

3. Performance limits for the complete image communication system

To assess the merits of the suggested techniques, it is convenient to establish the theoretical performance limits of a complete image communication system including the channel. This can be based on *channel capacity* and the *rate-distortion function*.

Channel capacity

Assume that the channel is of the Nyquist type. Then the signal samples, separated by a distance $1/2B$, can be transmitted without inter-symbol interference if the channel has an available bandwidth B. Assume that the channel samples have a maximum power level σ_c^2 and are disturbed by additive white Gaussian noise with variance σ_n^2. Then the channel capacity, C, which is the maximum amount of information that can be transmitted without errors, is given by

$$C = \frac{1}{2}\log_2\left\{1 + \frac{\sigma_c^2}{\sigma_n^2}\right\} \quad \text{bits per channel symbol.} \tag{3}$$

σ_c^2/σ_n^2 is the channel signal-to-noise ratio (CSNR).

The rate-distortion function

Rate-distortion functions of the type $R = R(D)$, where R is the rate and D is the allowable distortion, are available for certain simple statistical models. A Gaussian source of independent and identically distributed samples has the simple R-D function

$$R = \frac{1}{2}\log_2\left\{\frac{\sigma_x^2}{\sigma_d^2}\right\} \quad \text{bits per source sample,} \tag{4}$$

where σ_x^2 is the signal power and σ_d^2 is the accepted distortion. The ratio between the two is called the signal-to-noise ratio (SNR).

OPTA

By combining the channel capacity with the rate-distortion function, we get what is called the OPTA curve (OPTA=optimal performance theoretically attainable).

Assume M/K source samples per channel sample are transmitted, implying a sample compression of M/K. The following relation must then hold between the rate and the channel capacity

$$C = R\frac{M}{K}. \tag{5}$$

Inserting 3 and 4 into this equation and solving for the signal-to-noise ratio, we get

$$\frac{\sigma_x^2}{\sigma_d^2} = \left(1 + \frac{\sigma_c^2}{\sigma_n^2}\right)^{K/M}. \tag{6}$$

Plots in a decibel-decibel scale are shown in Figure 3 with K/M as parameter.
The OPTA curves give the best possible performance as a function of the CSNR for different sample compression factors. As an example, assume that an SNR of 20 dB gives an acceptable quality of the received signal. Without compression ($K/M = 1$) a CSNR of 20 dB is needed. With a 3/2 compression 30 dB CSNR is necessary, while at a compression of 2/1 40 dB is needed.

It should be noted that the above example is valid only for uncorrelated Gaussian signal and noise, but the principle will be the same for other kinds of statistics.

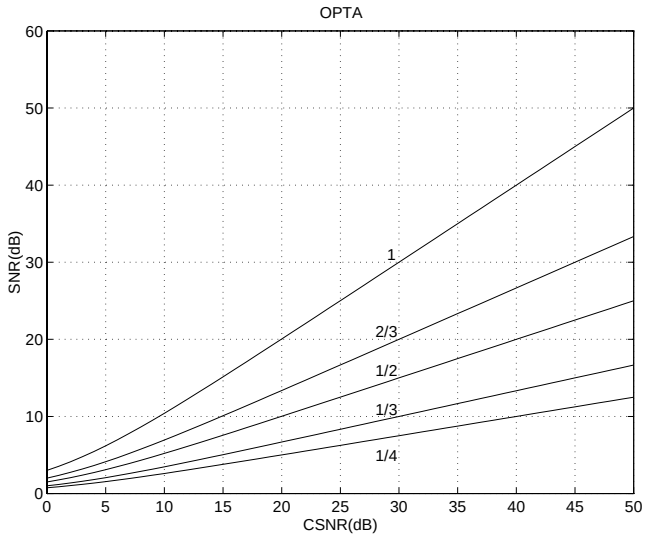

FIGURE 3. OPTA curves with K/M as parameter

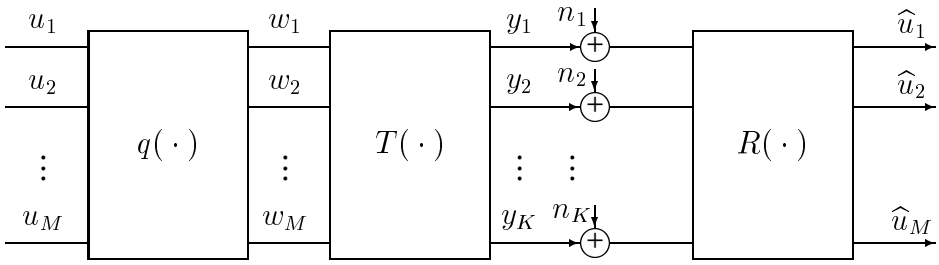

FIGURE 4. Mapping from higher to lower dimension (when $K < M$) as a two-stage process. $q(\cdot)$ indicates approximation while $T(\cdot)$ represents the mapping between dimensions. Included also are the additive channel noise sources and the dimension expanding mapping $R(\cdot)$ in the receiver.

4. Source-to-channel mapping

In this section the principles for combining samples leading to performance close to OPTA are presented.

Assume that M samples with identical distribution are collected in a vector $\mathbf{u}$. These samples may be a result of signal decomposition, as discussed in Section 4.1. As stated earlier, channel compression is obtained by reducing the number of samples.

The mapping operations are illustrated in detail in Figure 4, where the symbol reduction mechanism is split in two stages.

In the figure $q(\cdot)$ represents the approximation where any point in $\mathcal{R}^M$ is mapped to a point in a subset of the same space. If this were standard *vector quantization* (VQ), the subset would be distinct points in $\mathcal{R}^M$ (called representation points or codebook vectors). For the general case the subset will be represented by super-surfaces. $T(\cdot)$ is an invertible transform that maps all point of this super-surface to a space with dimension equal to the dimension of the super-surface. As an example, all points on some two-dimensional surface in a three-dimensional space can be mapped to a two-dimensional channel configuration.

The channel adds noise, indicated by the noise vector $\mathbf{n}$ in Figure 4, to the signal characterized by the CSNR. If the channel were not power or amplitude constrained, the channel signal could be made large enough to overcome the channel noise problem, and also the approximation noise problem, as it would be possible to make a dense set of representation values in $\mathcal{R}^M$.

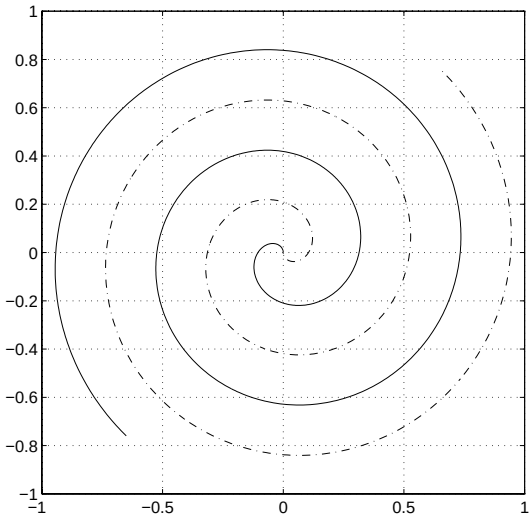

FIGURE 5. Approximation in two-to-one dimensional mapping.

The receiver part of the system consists of $R(\cdot)$, which is the mapping from dimension K back to dimension M. At high CSNR, $R(\cdot) \approx T^{-1}(\cdot)$. Often this approximation is also used at low CSNR, but this is not optimal. The received signal block of M samples can be expressed as

$$\widehat{\mathbf{u}} = R(\mathbf{n} + T(q(\mathbf{u})). \tag{7}$$

As an example, consider a mapping from two to one dimensions. The approximation must map all points in the two-dimensional space to a one-dimensional curve as shown in Figure 5. The mechanism is that each point is approximated by the closest point on the given curve.

The second stage is the mapping of the curve to the channel. In traditional systems using VQ, each representation point has an associated index which uniquely specifies the point, and the decoder knows both the representation points and the numbering system. The index is then transmitted as a digital symbol. The mapping from the index to the channel can be done in a variety of ways. A well known technique which accounts for channel effects is the so-called *index assignment procedure* [2]. The assignment is done in such a way that typical bit errors cause transitions between spatially close vectors. The mapping of the curve to the channel needs to have similar properties as in the index assignment procedure. As a mapping example, we choose for the channel representation *continuous pulse amplitude modulation* (CPAM). The CPAM amplitude can be chosen proportional to the length of the trajectory between the origin and the approximation point along the representation curve in the two-dimensional space. If no transmission errors affect the signal, any amplitude can be put back on the proper place on this curve by the receiver which, of course, possesses a replica of the mapping operator. With transmission errors a somewhat modified amplitude will be received which, mapped back to the two-dimensional space, will encompass a signal error. Because typical errors are assumed to be small, there are no disasters: there will be minor changes in one or both of the source signal dimensions.

Another important aspect of the double spiral approximation shown in Figure 5 is that it runs through the origin and covers the plane in a symmetric manner for the negative and positive channel amplitudes. This implies that the channel representation will be symmetric (provided that the source samples are rotation symmetrically distributed), and the transmitted power will be low because the probability density function is peaky at the origin where the signal will be represented by small channel amplitudes.

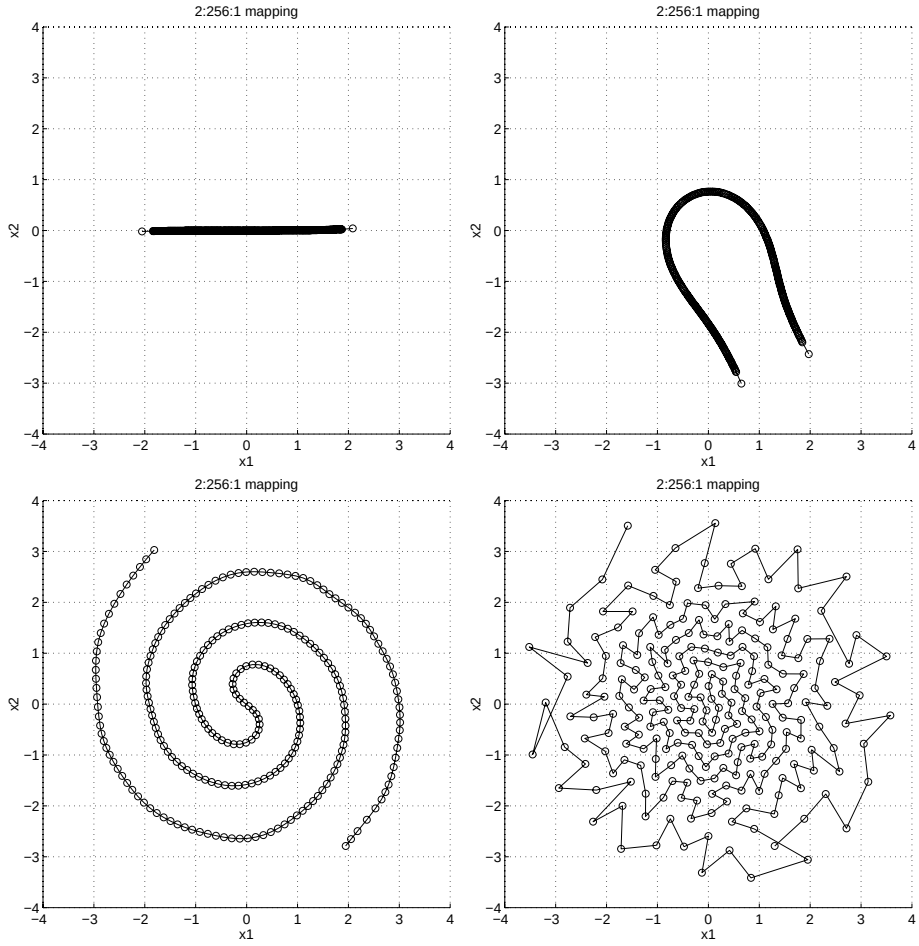

FIGURE 6. Four characteristic approximations for mapping from two-to-one dimensions. The channel signal-to-noise ratios are 0 dB, 7 dB, 25 dB, and 53 dB [3].

4.1. *Mapping optimizations for Gaussian signals and channels*

Section 3 introduced the OPTA curve for the case of Gaussian signal and channel noise. In this section we present mapping optimization results for this case.

Mappings for different CSNRs have been optimized for the 2:1 case based on the minimization of the noise expressed in Equation 1 by inserting $\widehat{u}$ from Equation 7 with $R(\,\cdot\,) = T^{-1}(\,\cdot\,)$ and $q(\,\cdot\,)$ representing linear stretching, under the constraint in Equation 2 [3]. Four characteristic results are shown in Figure 6. The mappings are designed as vector quantizers with size 256 codebooks, rather than a continuous representation, but including the channel constraints. The circles show each vector and straight lines are drawn between them to indicate the continuous approximation case.

It is interesting to note that for very low CSNR the approximation is given by a straight line in one of the dimensions, and therefore one sample is simply discarded through the approximation. When increasing the CSNR, the next phase is a horse-shoe approximation, which does not go through the origin. In the next figure, we see a double spiral which runs through the origin. And finally, at very high CSNR, the smooth spiral breaks up into something which may look like fractal behavior (if we extend to an infinite number of points and infinite CSNR).

The system performance is given in Figure 7 together with OPTA and performance for smaller codebooks.

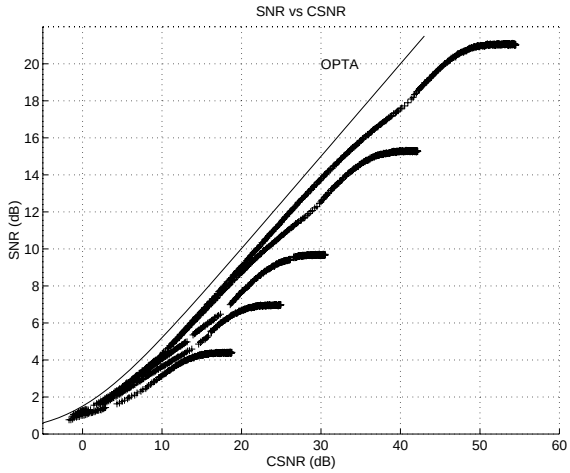

FIGURE 7. Performance of different size codebooks as a function of channel noise. The curves are (from top to bottom): OPTA, optimized coders for codebook sizes: 256, 64, 16, 8, and 4.

Although the performance is close to OPTA, the gap observed in the figure must have a mathematical explanation. If we again compare with the more well-known vector quantization, the problem with dimensionality comes to mind. To get optimal sphere packing, high dimensional vectors must be employed in VQ. The same effect can be obtained in this case by e.g. mapping from $2M$ to M dimensions to obtain 2 to 1 compression. However, the gain is a slowly increasing function of M. Presently no known method for optimizing for high dimensions is known.

5. Signal decomposition

Any signal that is meaningful to a human observer has some kind of structure, which for images can be categorized in terms of objects and textures. This implies that the signal has statistical dependencies among spatially close values, but also that the local statistical properties vary. Any efficient signal representation must cope with the variabilities, and remove redundancies originating from the statistical dependencies. A first step, before the source-channel mapping described above, should be to make some changes to the signal for clustering bits and pieces into classes, where each class contains statistically independent components with equal statistical properties. This is usually done by linear transforms, which include filter banks.

Linear transforms can, in principle, render full *decorrelation* of any stationary process. This will facilitate use of simpler mappings for close to optimal *performance*. We hasten to add that nonlinear techniques also can be used, and are, of course, more general, because they are capable of removing higher order statistical dependencies as well.

5.1. *Series expansion as a unifying decomposition engine*

Assume that the decomposition is separable, so that we can work with one-dimensional signals. One-dimensional signals related to horizontal and vertical scan lines of images are of finite length.

Any sequence $x(l)$ of finite length, L, can be exactly represented by a series expansion using L linearly independent functions $\boldsymbol{\phi}_i = \{\phi_i(0), \phi_i(1), \ldots, \phi_i(L-1)\}^T$, $i = 0, 1, \ldots, L-1$, as

$$x(l) = \sum_{i=0}^{L-1} u_i \phi_i(l), \quad l = 0, 1, \ldots, L-1. \tag{8}$$

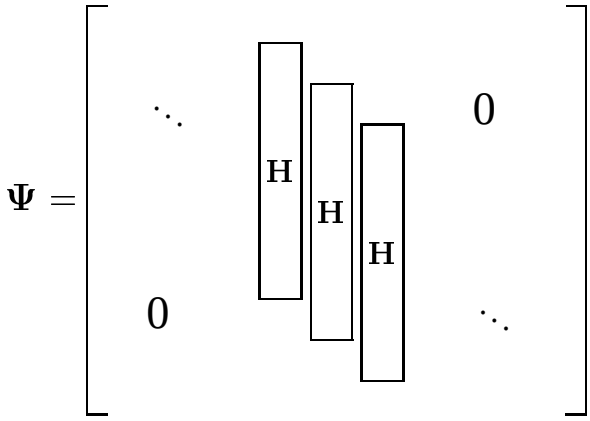

FIGURE 8. Transform matrix of FIR filter bank. The sub-matrices **H** contain the unit sample responses of the subband filters.

This equation represents the reconstruction from the decomposed components

$$\{u_l, l = 0, 1, \ldots, L-1\}$$

If the analysis-synthesis system is a *perfect reconstruction* (PR) system, the expansion coefficients can be found from

$$u_i = \sum_{l=0}^{L-1} x(l)\psi_i^*(l), \quad i = 0, 1, \ldots, L-1, \tag{9}$$

where the *reciprocal* or *dual* basis is given by

$$\boldsymbol{\psi}_j = \{\psi_j(0), \psi_j(1), \ldots, \psi_j(L-1)\}^T, \quad j = 0, 1, \ldots, L-1.$$

The mutual relationship between the two bases is then given by

$$\boldsymbol{\phi}_i^H \boldsymbol{\psi}_j = \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where H indicates Hermitian conjugate. In other words, the vectors in the two sets are *mutually orthogonal* if they do not share indices, and their inner product is equal to 1 for equal indices. If also the basis functions form an orthonormal set, the reciprocal basis is equal to the basis. We then say we have a *unitary system*.

In matrix form, Equation 9 can be written $\mathbf{u} = \boldsymbol{\Psi}^H \mathbf{x}$, or explicitly

$$\begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{L-1} \end{bmatrix} = \begin{bmatrix} \psi_0^*(0) & \psi_0^*(1) & \ldots & \psi_0^*(L-1) \\ \psi_1^*(0) & \psi_1^*(1) & \ldots & \psi_1^*(L-1) \\ \vdots & \vdots & & \vdots \\ \psi_{L-1}^*(0) & \psi_{L-1}^*(1) & \ldots & \psi_{L-1}^*(L-1) \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(L-1) \end{bmatrix}.$$

This equation tells us that every coefficient is based, in principle, on all the signal components. This is the most general linear transform. However, in images the correlations have predominantly local support. The local statistics of meaningful images vary. So, if our goal is to decorrelate the signal, each transform coefficient should depend only on a part of the input signal. Mathematically, this implies that the matrix $\boldsymbol{\Psi}$ should be a sparse band matrix. Even with that constraint, the flexibility of this formulation is enormous. Special cases of this formulation are square sub-transforms and filter banks. For both these cases the band matrix consists of repeated sub-structures, as shown in Figure 8. If we make the filter bank adaptive, the sub-structures will be unequal.

5.2. *Filter banks*

Typical analysis filter banks split the input signal in contiguous and slightly overlapping frequency bands, called *subbands*. If the analysis filter bank were able to *decorrelate* the signal completely, the output signal would be white. For all practical signals, complete decorrelation requires an infinite number of channels.

By performing critical decimation in each of the N channels, the total number of samples is conserved from the system input to decimator outputs. With appropriate channel arrangement the decimators also serve as *demodulators*. That is, the subband signals will, through this operation, be converted to baseband representation.

The synthesis filter bank has an inverse structure.

Throughout the last two decades an extensive literature on filter banks and filter bank structures has evolved. Perfect reconstruction (PR) is often considered desirable in subband coding systems. It is not a trivial task to design such systems, due to the downsampling required to maintain a minimum sampling rate. Certain filter bank structures inherently guarantee PR.

It is beyond the scope of this chapter to give a comprehensive treatment of filter banks. We shall only present different alternative solutions at an overview level.

We can distinguish between different filter banks based on several properties. In the following, five classifications are discussed.

i. FIR vs. IIR Although IIR filters have an attractive complexity, their long unit sample responses and nonlinear phase are obstacles in image coding. The unit sample response length influences the *ringing problem*, which is a main source of objectionable distortion in subband coders. The nonlinear phase makes the *edge mirroring technique* impossible [7] for efficient coding of images near their boundaries.

ii. Uniform vs. nonuniform filter banks This issue concerns the spectrum partitioning in frequency subbands. It is the general conception that nonuniform filter banks perform better than uniform filter banks. There are two reasons for that. The first is that our visual system also performs a nonuniform partitioning, and the coder should mimic the type of receptor for which it is designed. The second is that the filter bank should be able to cope with slowly varying signals (correlation over a large region) as well as transients that are short and represent high frequency signals. Ideally the filter banks should be adaptive, but without adaptivity one filter bank has to be a good compromise between the two extreme objectives cited above. Nonuniform filter banks can give the best tradeoff in terms of space-frequency resolution.

iii. Parallel vs. tree-structured filter banks Parallel filter banks are the most general, but tree-structured filter banks enjoy a large popularity, especially for obtaining octave band (dyadic) frequency partitioning. The popular sub-class of filter banks called *wavelet filter banks* or *wavelet transforms* belongs to this class.

iv. Linear phase vs. nonlinear phase filters There is no general consensus about the optimality of linear phase. In fact, the traditional wavelet transforms cannot be made linear phase. There are, however, at least three arguments in favor of linear phase. 1) The noise in the reconstructed image will be anti-symmetrical around edges with nonlinear phase filters. This does not appear to be visually pleasing. 2) The mirror extension technique [7] cannot be used for nonlinear phase filters. 3) Coding gain optimizations give better results for linear than nonlinear phase filters.

v. Unitary vs. nonunitary systems A unitary filter bank has the same analysis and synthesis filters (except for a reversal of the unit sample responses in the synthesis filters with respect to the analysis filters to make the overall phase linear). Because the analysis and synthesis filters play different roles, it seems plausible that they, in fact, should *not* be equal. Also, the gain can be larger, as demonstrated in Section 5.3, for nonunitary filter banks as long as straightforward scalar quantization is performed on the subbands.

Several other issues could be taken into consideration when optimizing a filter bank. These are, among others, the actual frequency partitioning including the number of bands, the length of the individual filters, and other design criteria than coding gain to alleviate coding artifacts, especially at low rates. For example, blocking artifacts can be avoided by ensuring that the different phases in the reconstruction process generate the same noise: the noise should be stationary rather than cyclo-stationary. This will impose requirements on the norms of the unit sample responses of the polyphase components [1].

5.3. *Optimal filter banks*

The *gain* in subband or transform coders depends on the detailed construction of the filter bank as well as the quantization scheme. Here the gain for uniform filter banks in combination with scalar quantization is considered.

Assume that the analysis filter bank unit sample responses are given by $\{h_n(k),\ n = 0, 1, \ldots, N-1\}$. The corresponding unit sample responses of the synthesis filters are required to have unit norm ($\sum_{k=0}^{L-1} g_n^2(k) = 1$).

The coding gain of a subband coder is defined as the ratio between the noise using scalar quantization (PCM) directly on the input signal and the subband coder noise when quantizing the subbands using optimal bit-allocation as explained in [5, 8]:

$$G_{SBC} = \frac{\sigma_x^2}{\left[\prod_{n=0}^{N-1} \sigma_{x_n}^2\right]^{1/N}}. \tag{11}$$

Here σ_x^2 is the variance of the input signal while $\sigma_{x_n}^2$ are the subband variances given by

$$\sigma_{x_n}^2 = \sum_{l=-\infty}^{\infty} R_{xx}(l) \sum_{j=-\infty}^{\infty} h_n(j) h_n(l+j) \tag{12}$$

$$= \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) |H_n(e^{j\omega})|^2 \frac{d\omega}{2\pi}. \tag{13}$$

From these equations it is evident that the subband variances depend both on the filters and the second order spectral information of the input signal.

For images the gain is often estimated assuming that the image can be modeled as a first order Markov source (also called an AR(1) process) characterized by

$$R_{xx}(l) = \sigma_x^2 0.95^{|l|}. \tag{14}$$

(This model is valid only after removal of the image average).

We consider the maximum gain using this model for three special cases. The first is the transform coder performance using the *Karhunen-Loève transform*. This is an important reference as all image and video coding standards are based on transform coding. The second is for unitary filter banks, for which optimality is reached by using ideal filters. The third case is for nonunitary filter banks, often denoted *bi-orthogonal* when the perfect reconstruction property is guaranteed. In the nonunitary case *half whitening* is obtained within each band.

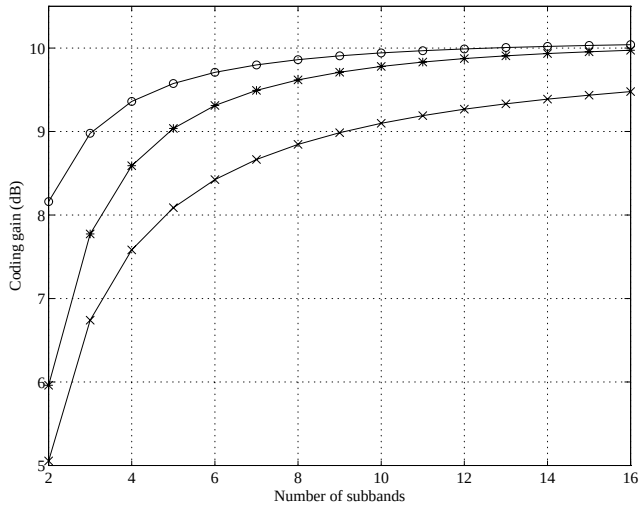

FIGURE 9. Maximum coding gain as a function of the number of channels for different one-dimensional coders operating on a first order Markov source with one-delay correlation $\rho = 0.95$. Lower curve: Karhunen-Loève transform. Middle curve: Unitary filter bank. Upper curve: Unconstrained filter bank, nonunitary case.

Mathematically this can be seen from the optimal magnitude response for the filter in channel n:

$$|H_n(e^{j\omega})| = \begin{cases} c\left[\frac{S_{xx}(e^{j\omega})}{\sigma_x^2}\right]^{-1/4} & \text{for } \omega \in \pm[\frac{\pi n}{N}, \frac{\pi(n+1)}{N}] \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

where c is a constant that can be selected for correct gain in each band.

The inverse operation must be performed in the synthesis filter to make completely flat responses within each band.

In Figure 9 the optimal coding gains as a function of the number of channels are given.

Notice that the above optimization is done under a bit constraint. The results are therefore not directly applicable to the resource allocation case in the next section, although we suspect optimality will not change much.

6. Efficient use of resources

When optimizing the complete image communication system, the channel imposes important constraints. Here we assume that the channel has limited bandwidth and power. These resources must therefore be assigned to the signal components where they are most efficiently used.

Assume that the signal (or rather the subband components) can be split into B classes with different statistics, where each class has a probability P_i (which is equal to the relative occurrence of this class). The classification will be based on the signal variances. Assume furthermore that there is one mapping device available for each class, which uses $\{r_i = (K/M)_i, i = 1, 2, \ldots, B\}$ channel symbols per signal sample and consumes a channel power $\sigma_{c_i}^2$. These parameters must be chosen to balance the available channel power, σ_c^2, and the available rate r. The rate constraint is given by

$$\sum_{i=1}^{B} P_i r_i \leq r, \tag{16}$$

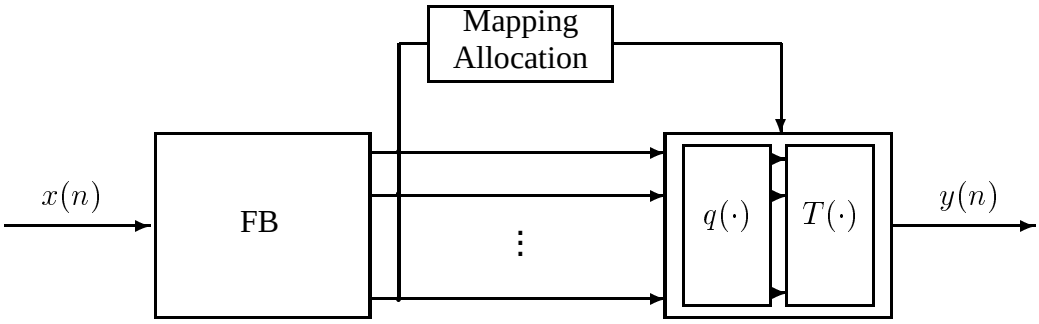

FIGURE 10. Signal communication system including signal decomposition and mapping allocation.

while the power constraint can be expressed as

$$\sum_{i=1}^{B} P_i r_i \sigma_{c_i}^2 \leq r\sigma_c^2. \tag{17}$$

Both r_i and $\sigma_{c_i}^2$ must be nonnegative real numbers for physical reasons.

The optimal power and bandwidth allocation can now be found from

$$\min_{\{r_i,\sigma_{c_i}^2\}} \sum_{i=1}^{B} P_i D_i(\sigma_{c_i}^2, r_i) \tag{18}$$

subject to the bandwidth and power constraints, where $D_i(\sigma_{c_i}^2, r_i)$ is the class distortion incorporating both approximation (mapping) noise and channel noise.

For the special case of Gaussian signal and channel, an explicit minimization problem exists:

$$\min_{\{r_i,S_i\}} \sum_{i=1}^{B} P_i (1 + \sigma_{c_i}^2/\sigma_n^2)^{r_i} \sigma_{u_i}^2. \tag{19}$$

The solution to this problem shows that each non-zero channel symbol is allocated an equal amount of power and the rate for each source symbol is given by [6]

$$r_i = \begin{cases} 0, & \text{if } \sigma_i^2 \leq \sigma_L^2 \\ \frac{1}{2C} \log_2 \frac{\sigma_i^2}{\sigma_L^2} & \text{if } \sigma_i^2 > \sigma_L^2. \end{cases} \tag{20}$$

σ_L^2 is a parameter which must be chosen to balance Equation 16, and the channel capacity is given by Equation 3.

In practice it is necessary to limit the number of mappings to a small number. The mappings also need to be simple in the sense that K and M must be small numbers. The selection of representative mappings will therefore be essential.

7. System model

The main ingredients for putting together a complete practical coder have now been introduced. Based on this, a complete source-channel encoder is shown in Figure 10.

The filter bank decomposes the input signal, resulting in close to decorrelated subband signals. Each subband typically has different statistics, and due to the variability in images and video signals, each subband also has fluctuating statistics. The subband signals are therefore split into blocks. The block energies are calculated and used as indicators as to which member of a set of available mappings should be used. The dimension ratio M/K is the compression ratio. The more important samples are compressed the least, which is achieved by a direct mapping without compression, or even by dimension increase. For the less important samples the ratio should be higher. The least important samples are simply discarded.

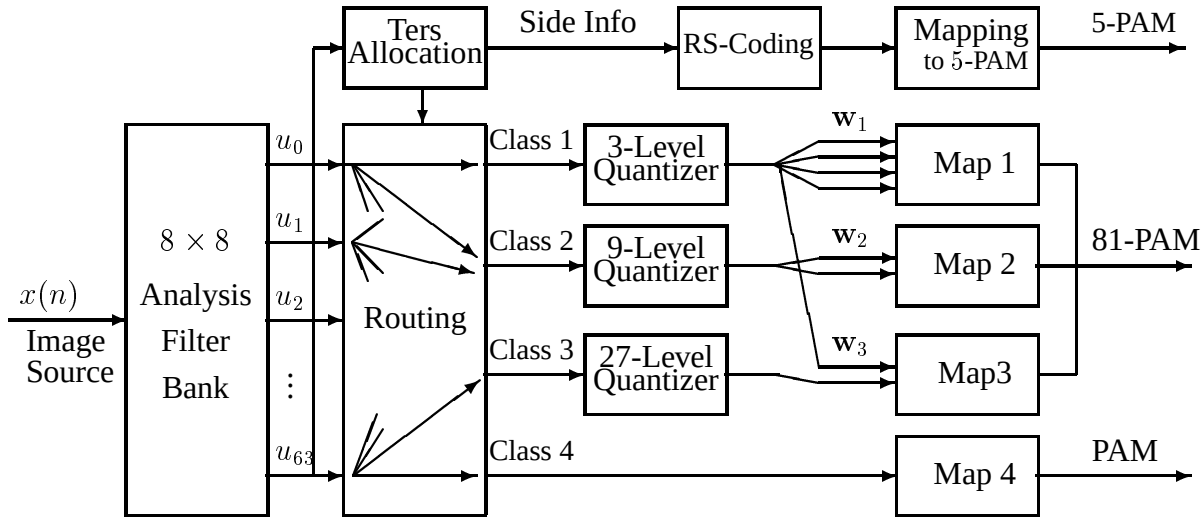

FIGURE 11. Subband coder combined with 81 PAM.

These operations should be performed according to the resource allocation described in the previous section.

In the following we describe two experiments using the mapping principles in conjunction with subband decomposition. The first example concerns still image communication, while the second example deals with transmission of video information.

7.1. *Still image coding combined with 81 PAM*

The following still image coder combines subband coding, allocation of *scalar* quantizers, and mappings from the quantized symbols of different resolution to an 81 pulse amplitude modulation scheme. The signal is transmitted, as usual, over an ideal Gaussian channel. There are several simplifications in this example compared to the optimal conditions described above. But, as will become apparent, the obtained performance is still remarkable.

The complete block diagram of the system is shown in Figure 11. The figure contains decomposition by an 8×8 filter bank followed by a selection of 4 different quantizers of 3, 9, 27, and 81 levels. The "ters (ternary symbol) allocation" block picks the right quantizer for each block of samples based on the power levels. This is similar to bit allocation. The only difference is that three instead of two levels are used. The most interesting part of the system is the bank of mapping devices. This combines the multilevel symbols into 81 level symbols. The combinations used are

- Mapping 0: Discard samples
- Mapping 1: $3 \times 3 \times 3 \times 3$
- Mapping 2: 9×9
- Mapping 3: 3×27
- Mapping 4: 81.

One of the mappings is illustrated in Figure 12. The figure also shows the approximations used. Because the quantizers are scalar, the representation values are located on a rectangular grid, formed by the Cartesian product between two pdf-optimized scalar quantizers. It should be noted that VQ, and in particular continuous VQ, would perform better.

The mapping allocation has to be transmitted as side information. Errors in the side information would be disastrous, as it would cause loss of synchronism between transmitter and receiver. To avoid side information errors, RS codes in combination with 5 PAM are used for this part of the code.

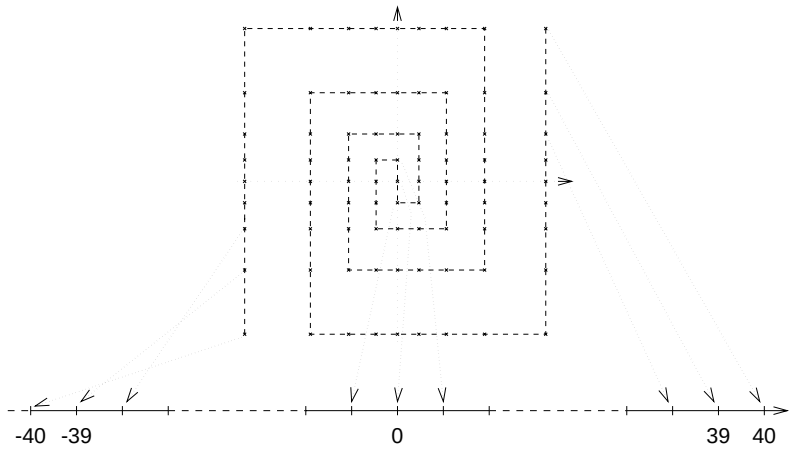

FIGURE 12. Scalar quantization and mappings to 81 PAM for 9×9.

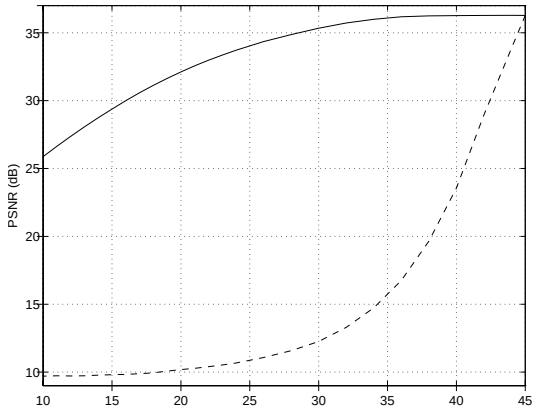

FIGURE 13. Simulation result for the "Lenna" image. Solid line: proposed system. Dashed line: random mappings. Horizontal axis: CSNR

Simulation results using this model on the test image "Lena" are shown in Figure 13 for 0.12 symbols/pixel.

The visual quality when transmitting over a noisy channel is demonstrated in Figure 14. The image to the left has been compressed using the same coder, but the mappings to the channel have been done at random, corresponding to what you might expect for a standard communication system. The channel has 35 dB CSNR. The image to the right has been processed by the proposed method and transmitted over a channel with 20 dB CSNR.

The complete image coding system is arranged in such a way that the approximation noise and the channel noise are additive. That is, both types of noise will generate visually equivalent distortions. If the approximation stage is designed to give pleasing artifacts, so will the channel distortion. This is very different from most signal transmission systems.

7.2. *Video communication system with QAM*

In the video encoder the time domain correlation is exploited through differential coding incorporating motion compensation (as in the H.263 video coding standard). The frame difference is coded in much the same way as in the still image coder. The complete system is shown in Figure 15.

The subband samples belonging to Class 1-3 are normalized and quantized using vector quantizers of three different rates. The blocks having the smallest mean squared value (Class 0) are discarded. The VQ output indices are mapped to the amplitude levels of the QAM channel symbols by the index maps. For the proposed system, the choices of source vector

FIGURE 14. Decoded images. Left: random mappings at CSNR=35 dB, right: proposed system at CSNR=20 dB.

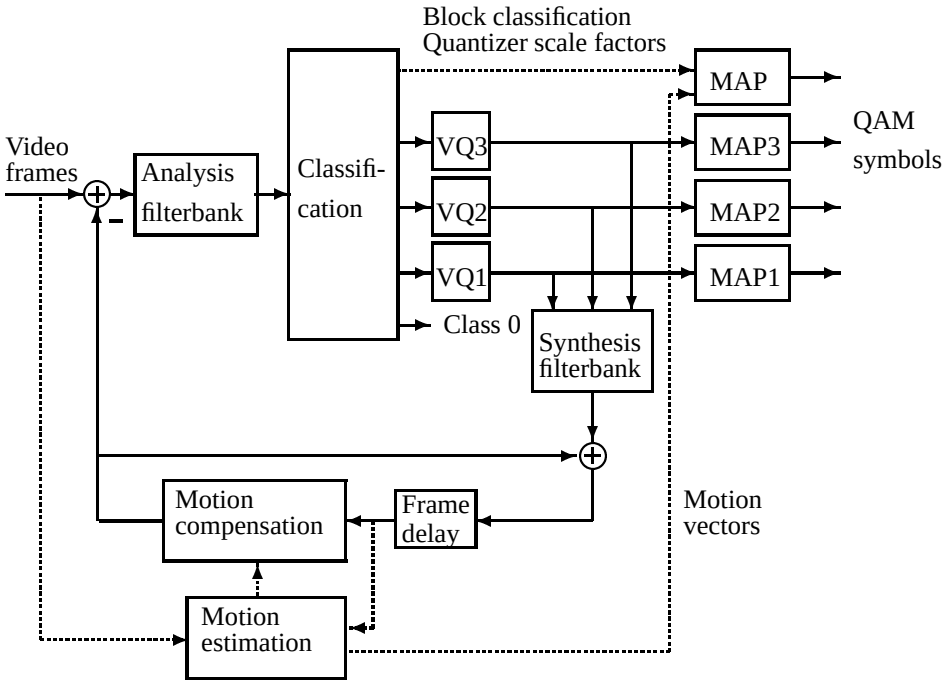

FIGURE 15. Proposed system (transmitter only). The normalization and re-normalization operations are assumed to be included in VQ1-VQ3.

Class	M	J	K
1	2	9	1
2	3	81	2
3	1	15	1

TABLE 1. Source vector dimension (M), codebook size (J), and channel space dimension (K) for each class. The case $K = 1$ corresponds to the real or imaginary axis of a QAM constellation.

dimension (M), codebook size (J), and channel space dimension (K) for each of the three classes are listed in Table 1.

The QAM signal constellations have an odd number of amplitudes in each dimension for power efficient transmission: codebook vectors having the highest probability are mapped to a channel symbol with zero amplitude. As an example, consider Class 1. The subband

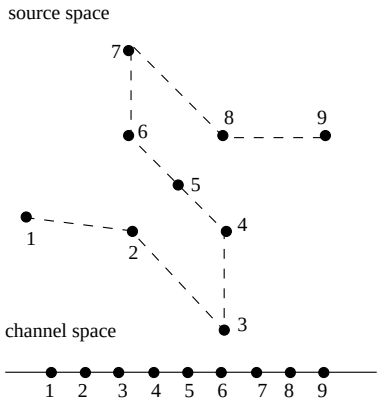

FIGURE 16. Typical configuration of MAP1. The numbers indicate corresponding points in the source and the channel space. The dashed line indicates points in the source space that are neighbors in the channel space.

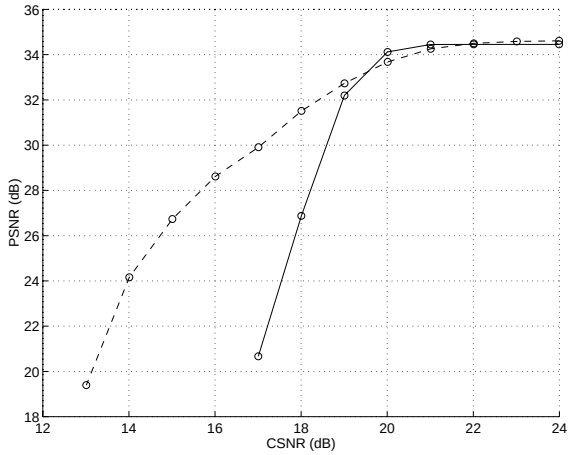

FIGURE 17. Left: PSNR versus CSNR for the reference system (——) and for the proposed system (- - -).

samples belonging to this class are divided into vectors of length $M = 2$ and quantized using a VQ codebook of size $J = 9$. The 9 codebook vectors are mapped directly to 9 points on the real or imaginary axis ($K = 1$) of an 81-QAM signal constellation (MAP1). Such a mapping is illustrated in Figure 16.

Finally, the side information parameters consisting of the motion vectors, the block class information, and the quantizer scale factors are transmitted by using a sparse subset of an 81-QAM signal constellation, resulting in larger robustness to channel noise.

To assess the performance of the proposed system a reference system based on the H.263 video coder is used. The H.263 coder is implemented according to [4] without any of the optional coding techniques, and without the chrominance information. In addition to the specifications given in [4], resynchronization based on the group of block (GOB) header as well as error detection and error concealment techniques are used in the decoder. Error detection is based on detection of illegal codewords, while error concealment is implemented by repeating information from the previous GOB or from the previous frame whenever an error is detected. In order to reduce the propagation of errors through subsequent frames, 3 intra-blocks are introduced systematically in each frame at the encoder side. The output bit stream is mapped to a 16-QAM signal constellation by traditional Gray coding. Note that, since the H.263 coder uses variable length coding, it is not possible to take the significance or the meaning of the individual bits into account when choosing the mapping.

The performance of the two systems is compared in Figure 17. The increased robustness

FIGURE 18. Left: Reference system, frame no. 200, CSNR = 17.0 dB. Right: Proposed system, frame no. 200, CSNR = 17.0 dB.

is significant but not as pronounced as for the still image example. The reason is that the dynamic range of the video signal is smaller and thus mapped to a smaller channel constellation. (81 QAM (corresponding to 9 levels per dimension) versus 81 PAM).

If we pick a channel with CSNR=17 dB, the result will be as shown in Figure 18 for the two systems.

Because H.263 is based on transform coding, the channel errors will, for this coder, cause block errors in the decoded image. These tend to appear in clusters as the variable length code will lose synchronism whenever an error occurs. Correct decoding is only recovered after new synchronization information is received correctly. We stress that the proposed system does not have any error protection or synchronization other than for the side information.

The channel configurations are selected to give close to the same performance at CSNR= 22 dB for both coders. As the H.263 coder uses 16 levels and the proposed coder 81 levels, this indicates that the first coder gives higher compression. The proposed coder needs a higher rate to get the same performance. This coder could definitely be improved, but as this is a fixed rate coder and the H.263 uses variable length coding, the standard has an inherent advantage in terms of rate distortion performance. Still, we obtain the same or improved quality at almost all CSNRs. This is the advantage of allocating noise to the transmission system as well as to the quantization.

8. System aspects

In the introductory chapter we argued that it is possible to transmit analog and digital source signals over the same physical channel while maintaining some sort of optimality for both sources. As an example, consider a still image coder which can generate both a lossless bit stream and a sequence representing mapped signal information. This coder could be adapted to a variety of applications. The transmission of these two signal types over a channel requires different solutions: error free transmission for the lossless case, and a mode which accepts channel noise for the lossy coding. In the latter case the signal does not need to be quantized at all, but can be transmitted as CPAM.

A model of the above system is depicted in Figure 19. In this example it is assumed that the same decomposition technique can be used for both lossless and lossy compression. In the lossless case, some type of entropy coding (EC) is used followed by insertion of error protection bits (FEC). The signal is transmitted using a size αC alphabet in a modem (M_1). For the lossy case, the compression is obtained by dimension reducing mappings as described above. The obtained signals are either transmitted as time discrete analog signals, or as signals with an alphabet of size C in a modem (M_2). The outputs from the two modems are multiplexed onto the same physical channel.

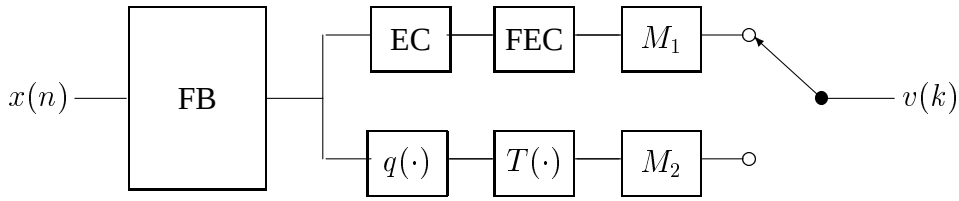

FIGURE 19. Optimal encoding and modulation system for lossless and lossy source coding

There are two reasons why the last system will be able to transmit a significantly larger number of sources. First of all, lossy compression will give a substantially lower rate even when the compression degradation is not noticeable. Secondly, the channel alphabet is larger so that more information can be sent per channel symbol.

It appears practical to make flexible systems like the above for efficient channel utilization. For mobile channels this may become an advantageous principle.

9. Discussion

In the designs proposed in this chapter we allow for distortions both in the compression and transmission phases. By including the channel in the mapping optimizations, very robust systems are obtained.

We have observed remarkable performance in the simulations. However, there are many challenges ahead. One major theoretical task is to find efficient modulation methods for very low CSNR channels. Practical channel models, including fading and other types of noise, should be included. The gap between the distortion bounds and the performance of the presented models should be bridged.

We believe that *variable length coding* should be avoided in the context of radio transmission, although this is presently the prevailing compression technique. To overcome this problem, compression and transmission specialists should work together. Then new generations of mobile multimedia systems could be developed with higher capacity, lower prices, and more graceful quality degradation.

References

[1] Ilangko Balasingham, Tor A. Ramstad, and John M. Lervik. Survey of odd and even length filters in tree-structured filter banks for subband image compression. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, April 1997.

[2] Nariman Farvardin. A study of vector quantization for noisy channels. *IEEE Trans. Inform. Theory*, 36(4):799–809, July 1990.

[3] Arild Fuldseth and Tor A. Ramstad. Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, volume IV, pages 3093–3096, 1997.

[4] ITU-T (CCITT), Study Group 15, Working Party 15/1, Expert's Group on Very Low Bitrate Videophone. *Video Codec Test Model, TMN5*, January 1995. Source: Telenor Research.

[5] N. S. Jayant and Peter Noll. *Digital Coding of Waveforms, Principles and Applications to Speech and Video*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.

[6] John Markus Lervik. *Subband Image Communication over Digital Transparent and Analog Waveform Channels*. PhD thesis, Norwegian University of Science and Technology, December 1996.

[7] S. Martucci. Signal extension and noncausal filtering for subband coding of images. In *Proc. SPIE's Visual Communications and Image Processing*, pages 137–148, November 1991.

[8] Tor Audun Ramstad, Sven Ole Aase, and John Håkon Husøy. *Subband Compression of Images – Principles and Examples*. ELSEVIER Science Publishers BV, North Holland, 1995.

[9] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, 1948.

[10] Claude E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, pages 142–163, March 1959.