Digital Coding of High Quality TV

Staffan Ericsson and Eric Dubois*
INRS-Télécommunications
3 Place du Commerce, Verdun, Canada H3E 1H6

1. Introduction

The study and development of systems for higher quality television has seen tremendous growth in recent years. This has been motivated by the fact that existing television systems give inadequate picture quality, especially for viewing on large screen displays. The main impairments which have been identified are scanning artifacts (line stucture visibility and flicker), limited resolution, and luminance/chrominance cross effects. The approaches which have been proposed to reduce these impairments are high-definition television (HDTV) and various enhanced quality television (EQTV) schemes. HDTV involves scanning the original scene with an increased number of lines per frame (usually more than 1000), and with correspondingly greater horizontal bandwidth. This signal is directly transmitted and displayed on a HDTV monitor. EQTV systems have been proposed as a possible alternative to — or as an intermediate step towards — HDTV. In EQTV the transmission is compatible with the NTSC (PAL or SECAM) standard, at least in scan format, but the quality is improved by signal processing in the transmitter and receiver to reduce the visibility of scanning artifacts and luminance/chrominance cross effects. In this paper, we are concerned with techniques to reduce the visibilty of scanning artifacts. In this form of EQTV system, HDTV cameras and monitors are used. The signal is down-converted to the NTSC scanning standard in the transmitter, and up-converted in the receiver for the HDTV display. EQTV offers a picture quality that is somewhat inferior to HDTV; however, it gives a dramatic improvement on standard NTSC while maintaining full transmission compatibility.

Television signals can be distributed in either analog or digital form. Of course analog transmission is predominant at this time, and the existence of large numbers of NTSC channels and receivers is the main motivation behind EQTV. If analog transmission of HDTV is foreseen, the analog bandwidth will be at least four times that of the NTSC channel, and generally more depending on the aspect ratio and color coding format used. Digital transmission facilities are becoming increasingly available and will largely replace analog transmission (except perhaps in over-the-air broadcasting). Digital transmission allows for the use of sophisticated digital source coding techniques to make efficient use of the available channel. With such coding, the HDTV signal does not require the large increase in digital transmission capacity that is indicated by the analog bandwidth [1]. This is due to several factors:

This work was supported by the National Sciences and Engineering Research Council under Strategic Grant G0845.

* Present address: PicTel Corp., Peabody MA 01960.

- The HDTV signal has higher correlation between adjacent samples due to the finer sampling grid; hence, it contains more statistical redundancy which can be removed by digital coding.
- The quantization noise in the HDTV signal is distributed over a larger spatial bandwidth. Due to the characteristics of the human visual system, the distortion is significantly less visible in an HDTV system than in an EQTV system at the same signal to distortion ratio.
- The up-converter in the EQTV receiver generally amplifies high spatial frequencies to improve the contrast; however, the coding error is also amplified. Hence, the visibility of quantization errors is further increased, imposing more stringent requirements for coding of EQTV signals.

The goal of this paper is to study digital image coding performance as a function of sampling density. It should be noted that the theoretically optimum coding scheme that uses a higher sampling frequency can always give at least as good quality as an optimum coder operating at a lower sampling frequency with the same rate in bits per second — at the higher sampling frequency there is more freedom to make a trade-off between resolution and quantization noise.[1] Another important question addressed is the required bit rate for transparent transmission at different resolutions. In this case, we want to transmit a picture at a given sampling frequency in such a way that there is no visible degradation compared to the *sampled* original.

Section 2 of this paper presents an analysis of theoretical coding performance for different sampling densities. The tool used for this analysis is rate distortion theory. Both distortion with respect to an unsampled continuous picture, and with respect to the sampled original, are derived. The distortion measure takes into account the response of the visual system by using a frequency weighted mean square error criterion.

In section 3 we present the coder structure used for our studies. A theoretically optimum coding scheme makes a trade-off between resolution (i.e., bandwidth) and in-band noise (quantization). This trade-off can also be achieved in a transform coder to some extent; this is the reason that a transform coding scheme can generally operate at lower bit rates (per pel) than a predictive coder. At high rates they should theoretically be equivalent; both coders are transmitting the full available bandwidth, and no trade-off can be made between resolution and quantization noise. Our coding experiments were performed using transform coding. Since we want to compare results at different sampling densities, it is important to have a coding scheme capable of working over a wide range of rates in bits per pel. Intraframe adaptive transform coding was used. Higher compression ratios can be achieved with interframe coding; however, we believe that it is necessary to use adaptive techniques such as motion compensation [2] to get a significant gain over intraframe coding, especially on critical broadcast scenes such as panning.

---

[1] Nevertheless, in a digital image transmission system the choice of sampling frequency is not primarily governed by the available bit rate, since the image quality can be expected to depend on the number of bits *per unit area* rather than the number of bits per pel. Instead, the reason for not choosing an excessively high sampling frequency is to limit the processing requirements, usually a linear function of the sampling frequency.

Section 4 outlines the simulation methods we used to compare digital coding for HDTV and EQTV and presents some preliminary results. Conclusions are given in section 5.

2. Rate Distortion Analysis of Coding at Different Resolutions

In this section we give an estimate of the required bit rate to code an image at different resolutions. The rate *per pel* for a given distortion will be lower at higher resolution due to increased correlation between adjacent pels and decreased visibility of quantization noise. In the analysis, we will determine the rate distortion function [3] for a simple image model. This is a lower bound for the performance of any coding scheme. Several simplifications are made in order to get a tractable problem; the most restrictive is the assumption that the source signal can be modelled as a stationary Gaussian source. Neither the stationarity nor the Gaussian assumption are valid for typical images; however, it is interesting to see whether bit rate predictions based on a simple model can be extrapolated to the practical case.

We start by determining the rate distortion function for a continuous image source; the rate is expressed in bits per unit area. Then, the effects of sampling are analyzed. A weighted square error criterion is incorporated to account for the properties of the human visual system (HVS).

2.1 R(D) for a spatially continuous image source

The MSE (mean square error) rate distortion function of a time-continuous stationary Gaussian source with power spectrum $\Phi(f)$ has the parametric representation

$$\begin{aligned} D_\theta &= \int_{-\infty}^{\infty} \min[\theta, \Phi(f)]\,df \\ R(D_\theta) &= \frac{1}{2}\int_{-\infty}^{\infty} \max\left[0, \log\frac{\Phi(f)}{\theta}\right] df \end{aligned} \tag{1}$$

where D_θ is the MSE for a certain choice of the distortion parameter θ, and $R(D_\theta)$ is the corresponding rate. $R(D)$ represents the lowest rate which can be obtained by any coding method with distortion less than or equal to D. Fig. 1 illustrates the situation. The regions of the spectrum where $\Phi(f) \leq \theta$ will never be transmitted, while the other parts are quantized with Shannon quantizers that give the distortion θ, requiring the rate $\frac{1}{2}\log(\Phi(f)/\theta)$. Hence, we get an error spectrum $\Phi_N(f) = \min[\theta, \Phi(f)]$. The generalization of (1) to multi-dimensional signals is straightforward.

When the MSE is computed, it is assumed that the error generated by deleting regions of the spectrum (filtering error) is equivalent to quantization error; however, it is straightforward to modify the procedure for a distortion criterion giving a higher weight to quantization error than to filtering error.

A frequency-weighted MSE,

$$D = \int |W(f)|^2 \Phi_N(f)\, df \tag{2}$$

is easily incorporated by filtering the source signal with a filter having a frequency response equal to $|W(f)|$. Then, the rate distortion function is computed for the filtered signal with

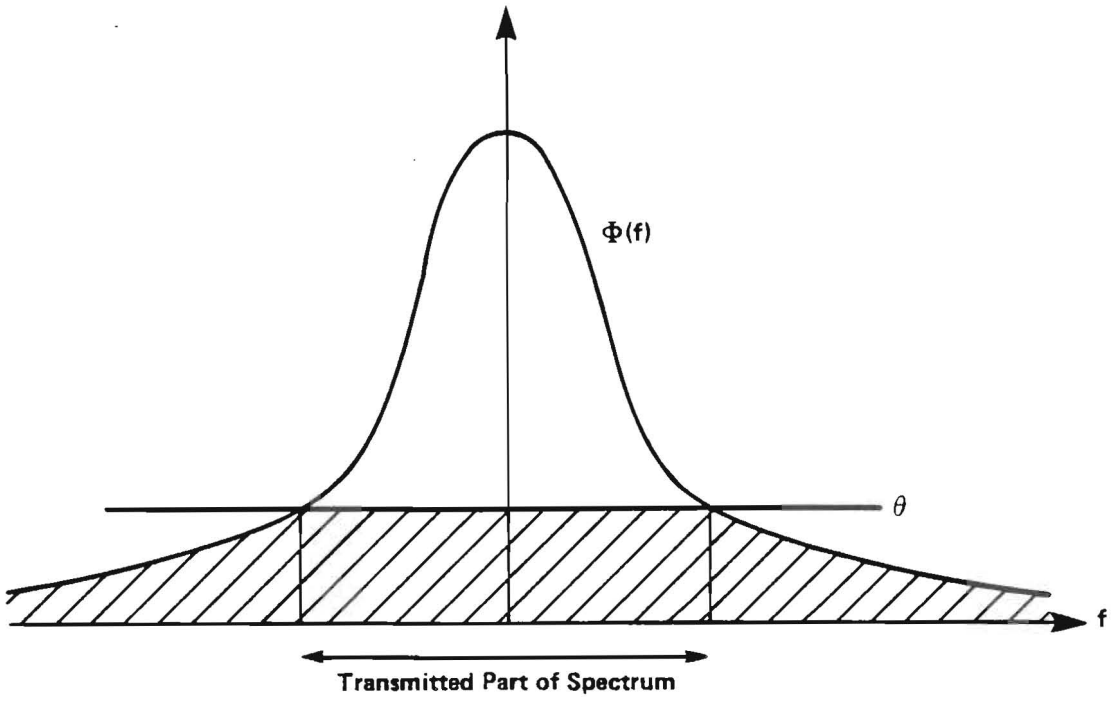


Fig. 1 Power spectrum of stationary Gaussian source and error under optimal source coding.

spectrum $\Phi(f)|W(f)|^2$. In the receiver, an inverse filtering is performed, and we obtain the error spectrum

$$\Phi_N(f) = \min\left[\frac{\theta}{|W(f)|^2}, \Phi(f)\right]. \tag{3}$$

Two commonly used models for still image statistics are a separable and an elliptic covariance model. A separable two-dimensional autocorrelation function (acf) that is exponential in both dimensions is given by

$$\phi_{sep}(x_1, x_2) = \sigma^2 \exp(-\alpha_1|x_1| - \alpha_2|x_2|). \tag{4}$$

The elliptic acf is defined as

$$\phi_{ell}(x_1, x_2) = \sigma^2 \exp(-\sqrt{(\alpha_1 x_1)^2 + (\alpha_2 x_2)^2}). \tag{5}$$

For $\alpha_1 = \alpha_2$ the elliptic model gives an isotropic acf. The separable acf is preferable if the images are dominated by horizontal and vertical edges. If there is a significant content of diagonal structures, the elliptic model is better.

Our analysis will be concentrated on the elliptic model. The power spectrum is given by

$$\begin{aligned}\Phi_{ell}(f_1,f_2) &= \frac{2\pi\sigma^2/(\alpha_1\alpha_2)}{(1+(2\pi f_1/\alpha_1)^2+(2\pi f_2/\alpha_2)^2)^{3/2}} \\ &= \Phi_{iso}(\omega_r) = \frac{2\pi\sigma^2/(\alpha_1\alpha_2)}{(1+\omega_r^2)^{3/2}}, \qquad \omega_r = 2\pi\sqrt{(f_1/\alpha_1)^2+(f_2/\alpha_2)^2}.\end{aligned} \tag{6}$$

The rate distortion function for the elliptic model can be shown to have the parametric representation [4]

$$\begin{aligned}D_{\omega_{cut}} &= \sigma^2\frac{1+\frac{3}{2}\omega_{cut}^2}{(1+\omega_{cut}^2)^{3/2}} \\ R(D_{\omega_{cut}}) &= \frac{3\alpha_1\alpha_2}{16\pi}(\omega_{cut}^2-\log(1+\omega_{cut}^2)).\end{aligned} \tag{7}$$

The parameter ω_{cut} is the normalized cut-off frequency; the region outside the ellipse

$$\left(\frac{2\pi f_1}{\alpha_1}\right)^2+\left(\frac{2\pi f_2}{\alpha_2}\right)^2 = \omega_{cut}^2 \tag{8}$$

is not transmitted. The region inside the ellipse is transmitted with an error $\theta = \Phi_{iso}(\omega_{cut})$. The rate distortion function is plotted in Fig. 2.

2.2 R(D) for a sampled image source

To study the effects of sampling we will assume that an ideal low-pass filtering is performed before sampling to avoid aliasing. After low-pass filtering, sampling and reconstruction is a transparent process. Hence, the effects of sampling are equivalent to the effects of low-pass filtering.

Denote the passband by B_p,

$$B_p = \{(f_1,f_2):\ -f_{p1} < f_1 < f_{p1}\ \wedge\ -f_{p2} < f_2 < f_{p2}\}. \tag{9}$$

This passband applies for rectangular sampling. For different sampling structures, other passbands are more appropriate [5]. Ideal low-pass filtering gives the spectrum Φ_{LP}:

$$\Phi_{LP}(f_1,f_2) = \begin{cases}\Phi(f_1,f_2), & (f_1,f_2)\in B_p \\ 0, & (f_1,f_2)\in \overline{B}_p\end{cases} \tag{10}$$

The rate distortion function is denoted $R'(D')$, where the distortion D' is the MSE compared to the *unfiltered* original. The rate distortion function for the low-pass filtered process is denoted $R_{LP}(D_{LP})$, where D_{LP} is the MSE relative to the low-pass filtered process. For a given value of the parameter θ the relation is

$$\begin{aligned}D'_\theta &= D_{filt}+D_{LP\theta} \\ R'_\theta &= R_{LP\theta}\end{aligned} \tag{11}$$

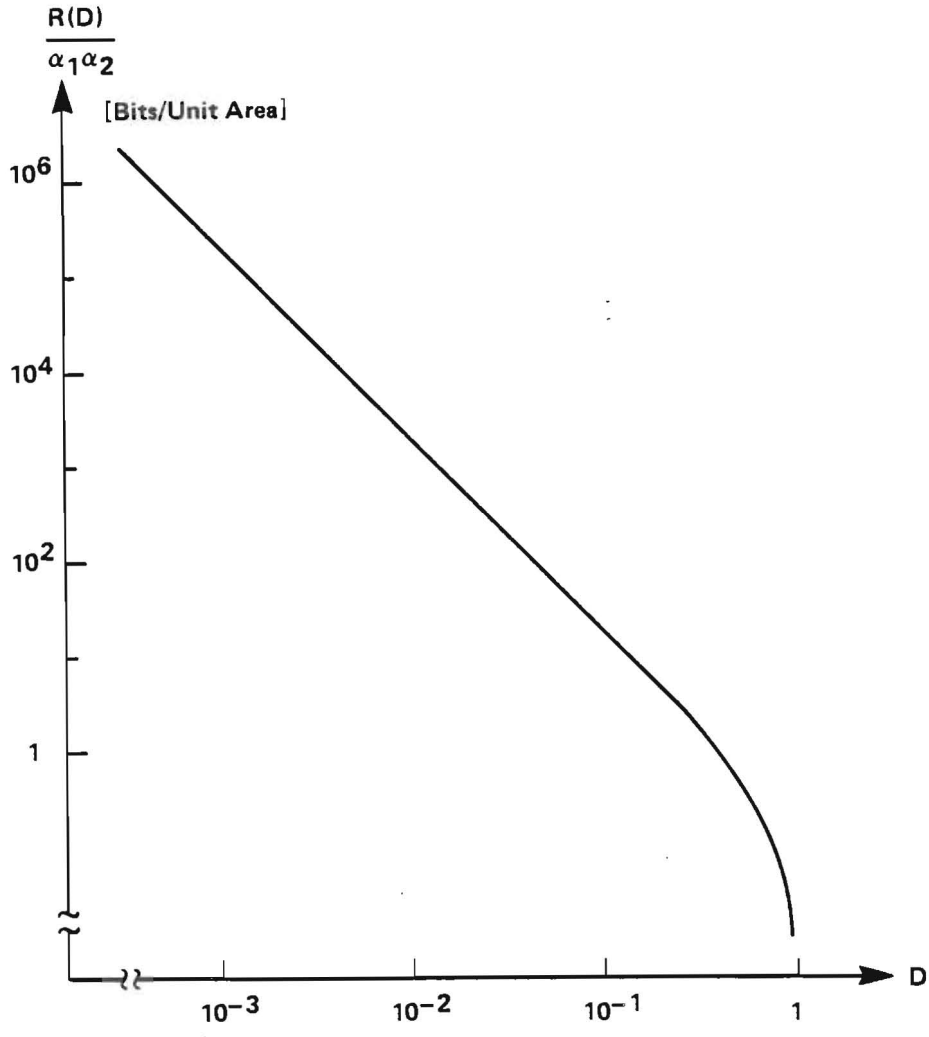


Fig. 2 Rate distortion function for elliptic covariance model under an unweighted MSE distortion criterion expressed in bits per unit area. The rate is plotted on a logarithmic scale.

where the error due to low-pass filtering is

$$D_{filt} \triangleq \int\int_{\overline{B}_p} \Phi(f_1, f_2) df_1 df_2. \tag{12}$$

For distortion levels above $D_{\theta crit}$, no frequencies outside B_p are transmitted even for the unfiltered signal. Hence, $R'(D')$ is equal to the unfiltered case:

$$D'_\theta = D_\theta, \qquad R'_\theta = R_\theta \tag{13}$$

if

$$\theta \geq \theta_1 \triangleq \max_{(f_1,f_2)\in\overline{B}_p} \Phi(f_1, f_2) \tag{14}$$

For the elliptic model the critical threshold is $\theta_1 = \Phi_{iso}(\omega_{crit})$, where

$$\omega_{crit} = 2\pi \min_{i=1,2} \frac{f_{pi}}{\alpha_i}. \tag{15}$$

Assuming a passband tailored to the signal statistics,

$$\frac{f_{p1}}{\alpha_1} = \frac{f_{p2}}{\alpha_2} = \frac{\omega_{crit}}{2\pi} \tag{16}$$

the rate distortion function $R'(D')$ is given by eq. (7) for rates lower than R_1.

$$\begin{aligned} R_1 &= \frac{3\alpha_1\alpha_2}{16\pi}(\omega_{crit}^2 - \log(1+\omega_{crit}^2)) \\ &\approx \frac{3\pi}{4} f_{p1} f_{p2} \quad \text{nats/unit area}, \qquad \omega_{crit} \gg 1. \end{aligned} \tag{17}$$

Sampling at the Nyquist limit,

$$f_{si} = 2f_{pi}, \qquad i = 1,2 \tag{18}$$

we get the rate

$$R_1 \approx \frac{3\pi}{16} \text{ nats/sample} \approx 0.85 \text{ bits/sample}, \qquad \omega_{crit} \gg 1. \tag{19}$$

For $\omega_{cut} \in (\omega_{crit}, \sqrt{2}\omega_{crit})$ the region of transmitted frequencies is as indicated in Fig. 3. If $\omega_{cut} \geq \sqrt{2}\omega_{crit}$, then the whole region B_p will be transmitted.

For an arbitrary power spectrum, all frequencies within the passband are transmitted when

$$\theta \leq \theta_2 \stackrel{\triangle}{=} \min_{(f_1,f_2)\in B_p} \Phi(f_1, f_2). \tag{20}$$

In that case

$$\begin{aligned} D'_\theta &= D_{filt} + D_{LP\theta} = D_{filt} + 4\theta f_{p1} f_{p2} \\ R'_\theta &= R_{LP\theta} = \frac{1}{2}\int\int_{B_p} \log \frac{\Phi(f_1,f_2)}{\theta} df_1 df_2 \\ &= R_2 + 2f_{p1}f_{p2} \log \frac{\theta_2}{\theta} \end{aligned} \tag{21}$$

where

$$R_2 \stackrel{\triangle}{=} \frac{1}{2}\int\int_{B_p} \log \frac{\Phi(f_1,f_2)}{\theta_2} df_1 df_2. \tag{22}$$

For the elliptic model, sampled at the Nyquist limit

$$R_2 \approx \frac{3}{4}\left(3 - \frac{\pi}{2}\right) \text{ nats/sample} \approx 1.55 \text{ bits/sample}, \qquad \omega_{crit} \gg 1. \tag{23}$$

The function $R'(D')$ is plotted in bits/sample in Fig. 4 for the sampling frequency $f_{si} = 25\alpha_i$. The rate distortion function for the lowpass filtered process is also indicated.

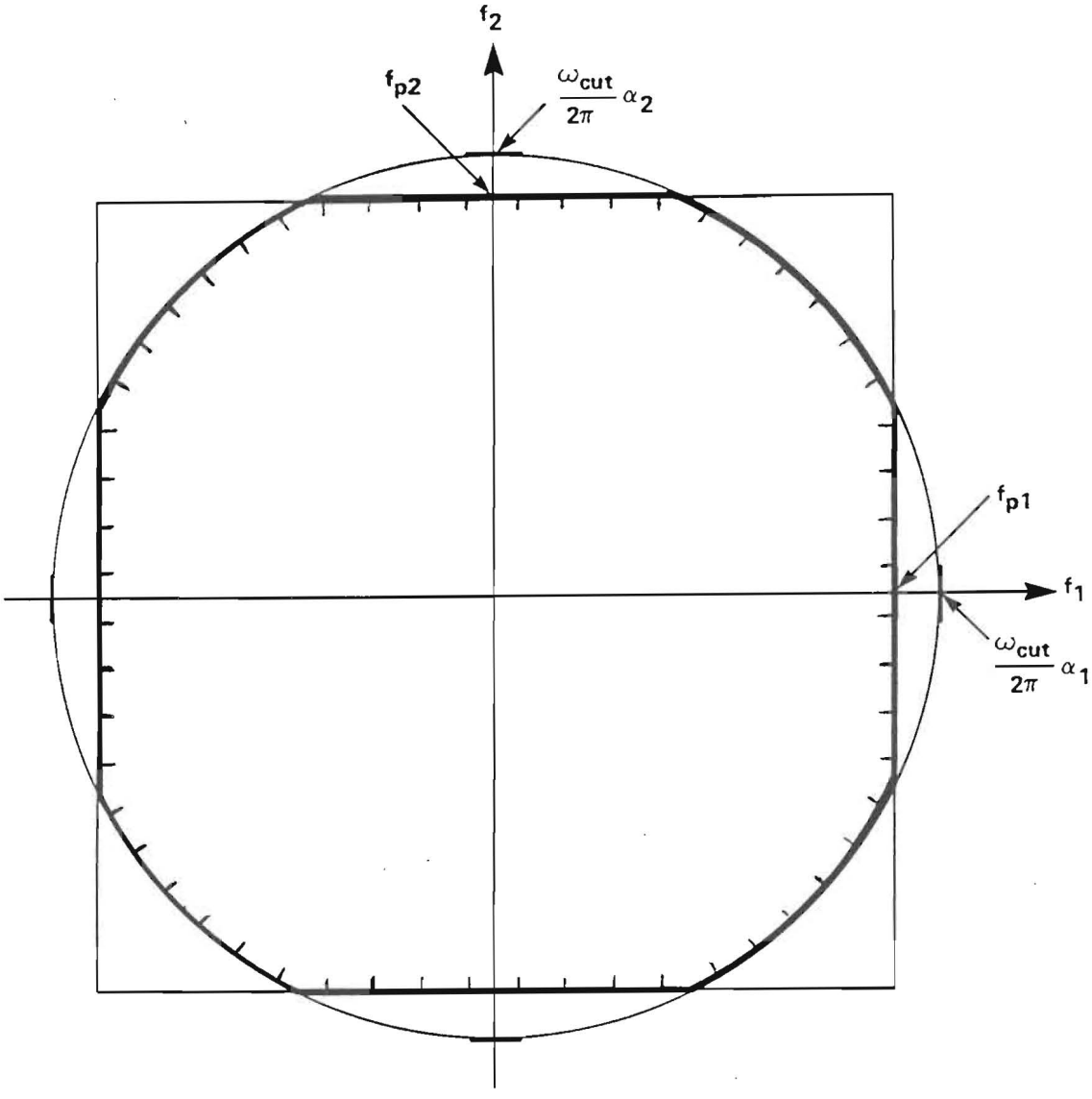


Fig. 3 Transmitted region of spectrum for rates $R_1 < R < R_2$.

2.3 Frequency weighted MSE

It is well known that the contrast sensitivity of the HVS has a bandpass characteristic. In the low spatial frequency range, the HVS acts as a differentiator due to lateral inhibition; at high spatial frequencies the response is limited by the optics of the eye and the density of the receptors and neural network of the retina. If we assume that the high-frequency cut-off is due to diffraction, we obtain a simple model for the contrast sensitivity,

$$C_{nb}(f_r) = e\frac{f_r}{f_0}e^{-f_r/f_0}C_{nbmax} \tag{24}$$

where f_r is radial frequency; the function obtains its maximum at $f_r = f_0$. The index nb

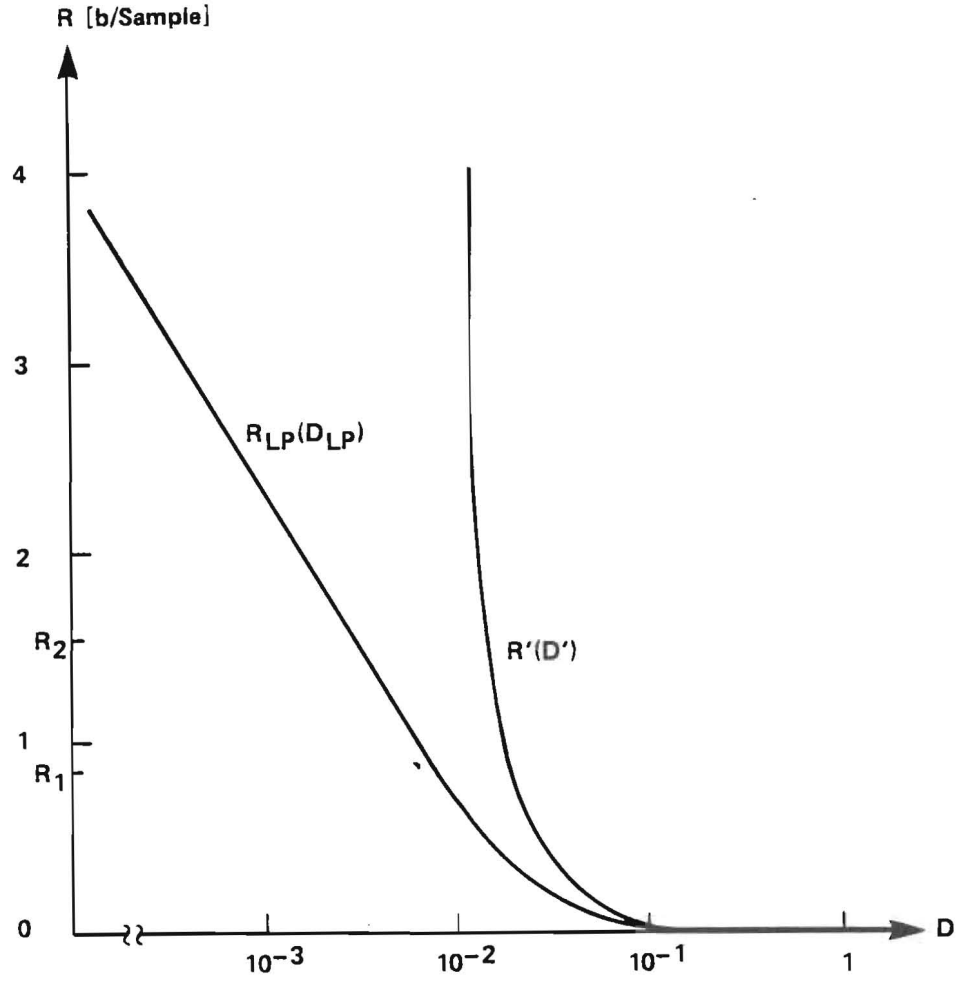


Fig. 4 Rate distortion function in bits per sample for elliptic covariance sampled at $f_{si} = 25\alpha_i$.

indicates that narrowband stimuli are considered.

Most measurements of the spatial frequency characteristics have utilized sinusoidal gratings. Narrow-band noise was used by Sakrison [6]; the contrast sensitivity curve had about the same shape as for sinusoidal gratings. A reasonable fit to most published curves is obtained by eq. (24) with $f_0 \approx 4$ cycles/degree, although the high-frequency roll-off is somewhat less abrupt than $\exp(-f_r/f_0)$. The anisotropy of the HVS has not been incorporated, although it is known that the sensitivity to diagonal gratings is about 3 dB less than to horizontally or vertically oriented gratings [7].

A weighted squared error has the form

$$D = \int\int \Phi_N(f_1, f_2)|W(f_1, f_2)|^2 df_1 df_2 \tag{25}$$

where $\Phi_N(f_1, f_2)$ is the noise power spectrum, and $W(f_1, f_2)$ the weighting function. Now,

would the contrast sensitivity function give a useful weighted MSE? The answer depends on the kind of distortion that is being considered. It should be appropriate to predict the visibility threshold for narrowband stimuli, but it is probably less useful for suprathreshold distortion levels or broadband noise, e.g., quantization noise.

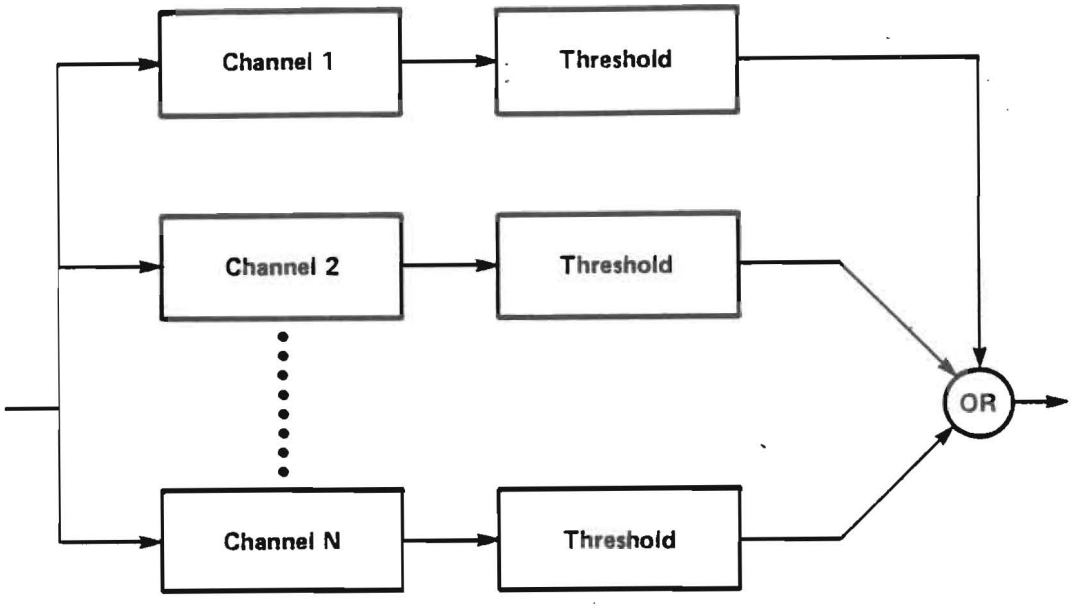


Fig. 5 Detection model for the human visual system with frequency-selective channels.

Experiments with detection of compound gratings consisting of two frequency components have led to a model with parallel processors, the independent channels model, shown in Fig. 5. The processors are assumed to be spatial-frequency selective and a stimulus is detected when the activity in any of the frequency channels exceeds a threshold. The structure of a single channel has been studied by Mostafavi and Sakrison [8] (the results are reviewed in [6]). Experiments indicate that the relative bandwidth of each channel should be approximately constant [9]; Sakrison's work indicates a radial bandwidth of about one octave, and angular bandwidth of $\pm 10°$. To account for the different absolute bandwidths in high and low frequency channels, we multiply $C_{nb}(f_r)$ by f_r. Hence, a simple model for the sensitivity to wide-band stimuli gives the sensitivity function

$$C_{wb}(f_r) = \frac{e^2}{4}\left(\frac{f_r}{f_0}\right)^2 e^{-f_r/f_0} C_{wbmax} \tag{26}$$

which attains its maximum value at $f_r = 2f_0$.

The weighted MSE does not accurately model the OR operation of the channel model; however, considering the large radial bandwidth of the channels we can expect a reasonably good prediction of the detection threshold for isotropic noise. In particular, a noise spectrum

inversely proportional to the weighting,

$$\Phi_N(f_r) = \frac{a}{|C_{wb}(f_r)|^2} \tag{27}$$

should give simultaneous detection in all channels for a certain value of a. Hence, for optimum encoding according to rate distortion theory a weighting function proportional to $C_{wb}(f_r)$ should be used.

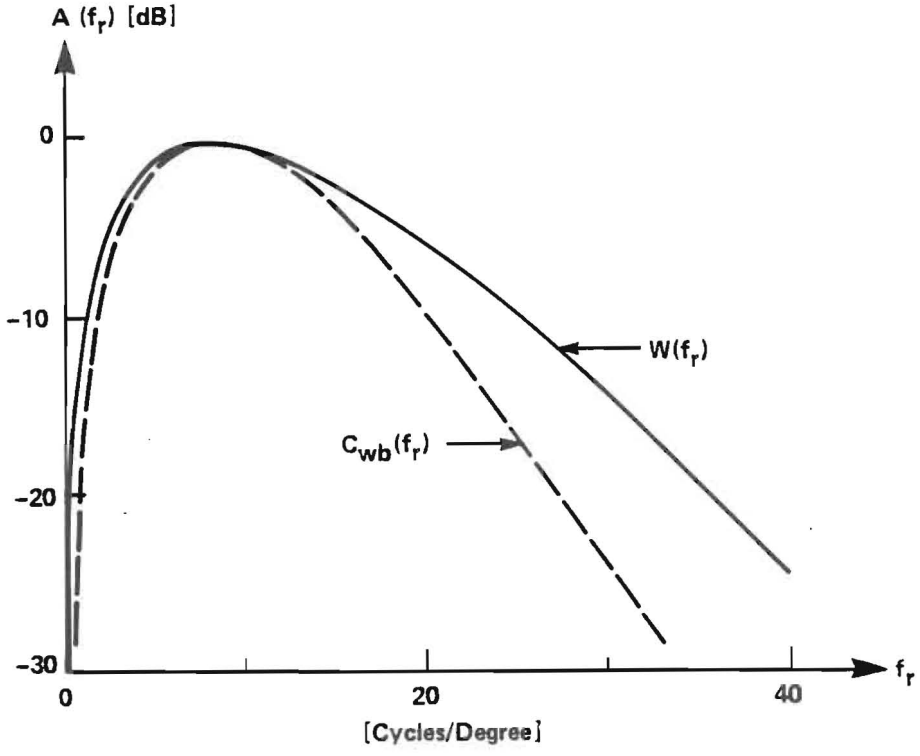


Fig. 6 Experimentally obtained weighting function [10] compared to a weighting function derived from a simple model of the HVS.

Mannos and Sakrison simulated optimum encoding of still pictures using different frequency weighting functions [10]. The best subjective quality was obtained with

$$W(f_r) \approx 2.6(0.0192 + 0.114 f_r) \exp\left[-(0.114 f_r)^{1.1}\right] \tag{28}$$

where f_r is the radial frequency in cycles/degree. The function is outlined in Fig. 6 together with our theoretically derived function $C_{wb}(f_r)$. As seen, $W(f_r)$ also obtains its maximum at 8 cycles/degree, one octave above the maximum for narrow-band stimuli. The high-frequency roll-off is slower—the same tendency as when comparing measured contrast sensitivity functions to $C_{nb}(f_r)$. The low-frequency attenuation is significant: 26 dB at $f_r = 0$.

The rate distortion function was computed numerically for the weighted MSE distortion measure defined in eq. (25) and the weighting function (28). An isotropic source is assumed;

the spectrum is given by eq. (6) with $\alpha_1 = \alpha_2 = \alpha$. The rate distortion function for the continuous image source is shown in Fig. 7. The effect of low-pass filtering is also shown for cut-off frequencies $f_{p1} = f_{p2} = 12.5$ and 25 cycles/degree.

The rate in bits per sample is shown in Fig. 8. We have also indicated the weighted MSE for a coder optimized under a non-weighted MSE criterion. The distance between the two curves is the improvement that can be expected by taking the HVS spatial frequency characteristics into account in the coding process. Fig. 8 can be used to estimate the difference in bit rates for coding EQTV and HDTV at the same weighted MSE. We assume that $f_s = 25$ corresponds to EQTV and $f_s = 50$ corresponds to HDTV. Suppose that the HDTV signal achieves the desired distortion level at 1 bit/pel. The same weighted MSE is obtained at about 2.2 bits/pel for the EQTV signal. Assuming a 4:1 ratio of sampling frequencies, the HDTV bit rate is only 1.8 times higher than the EQTV rate.

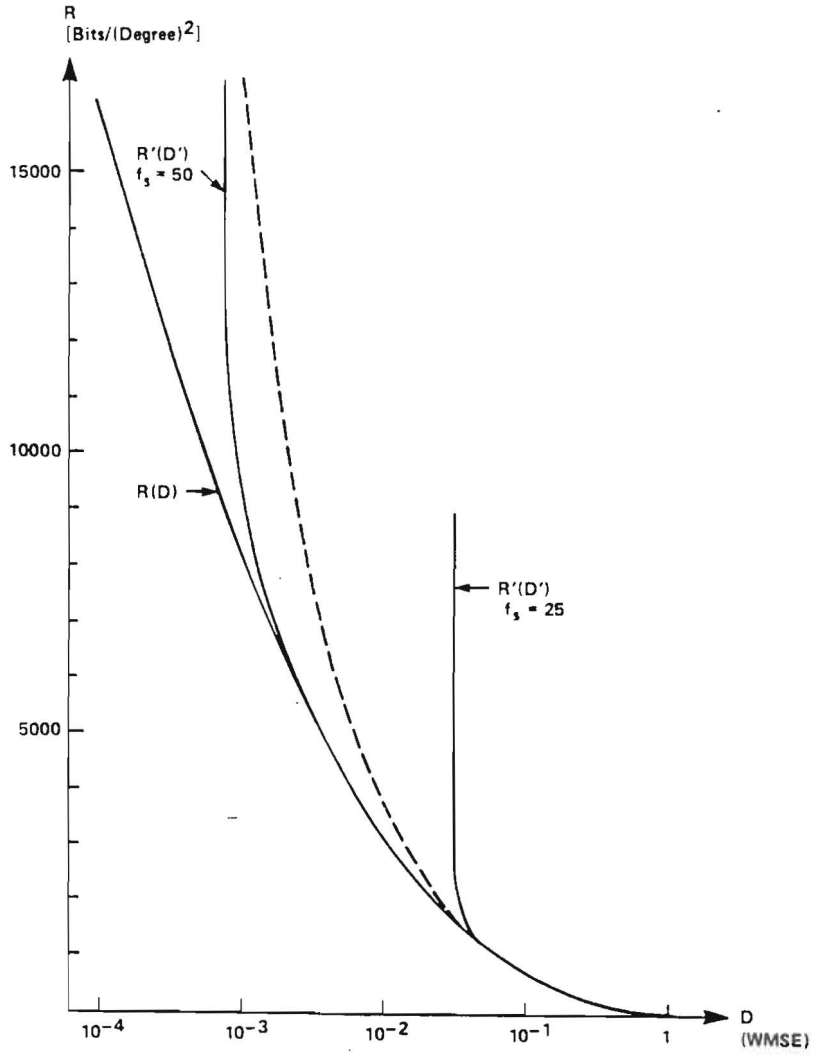


Fig. 7 Rate distortion function for isotropic model under a weighted MSE criterion; $\alpha_1 = \alpha_2 = 2$ cycles/degree. The dashed curve indicates the weighted MSE for a coder that is optimum under a non-weighted MSE criterion.

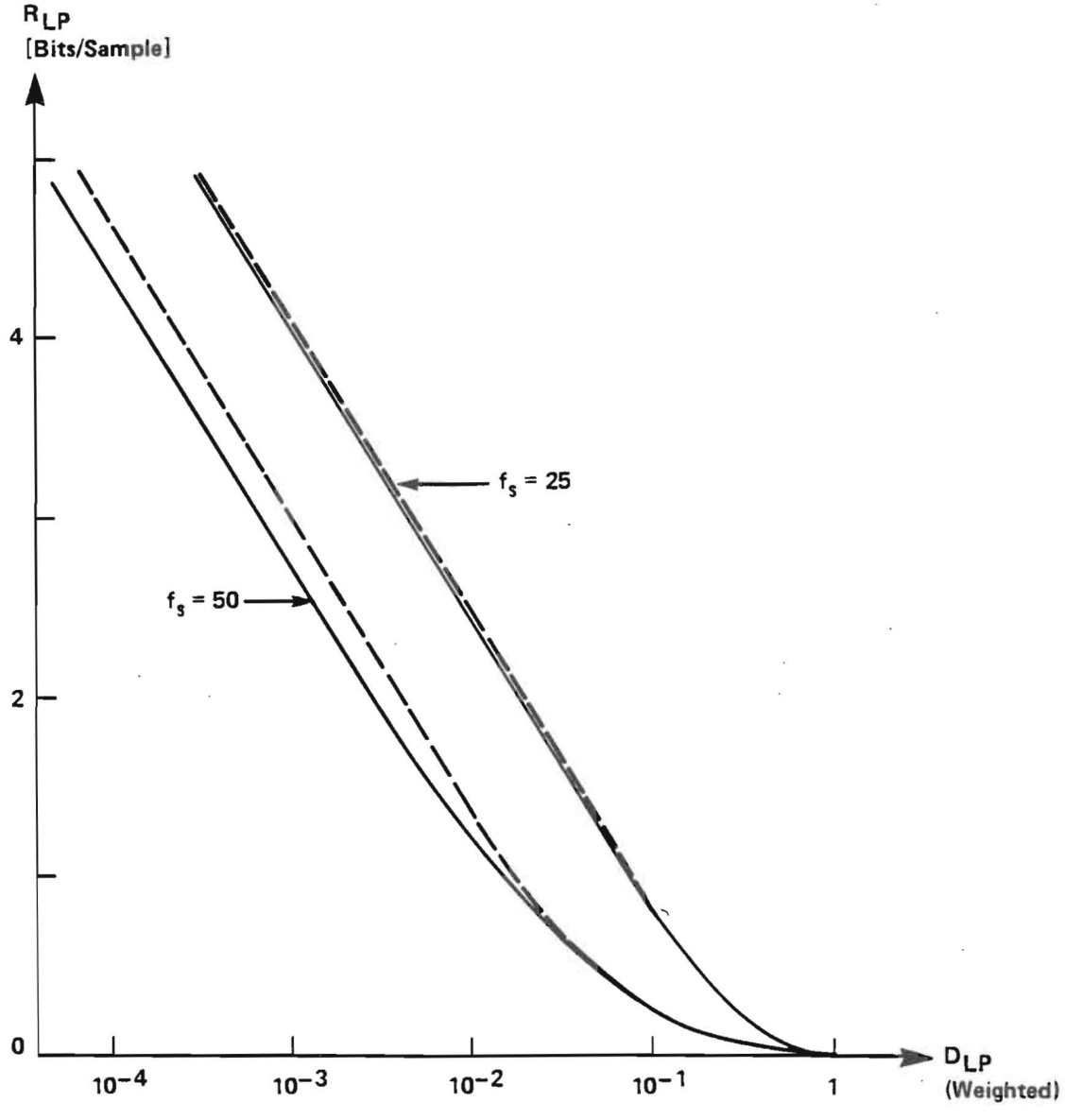


Fig. 8 Rate distortion function in bits per sample for a medium-detailed picture ($\alpha = 2$) and sampling frequencies $f_{s1} = f_{s2} = 25$ and 50 samples/degree. The dashed curves indicate the weighted MSE for a coder that is optimum under a non-weighted MSE criterion.

3. Frequency Weighted Transform Coding

In our coding experiments, an adaptive transform coder has been used. As illustrated in Fig. 9, transform coding consists of three steps: transformation, quantization, and variable length coding. In the receiver, the quantized values are reconstructed from the received data stream, and the inverse transform is performed.

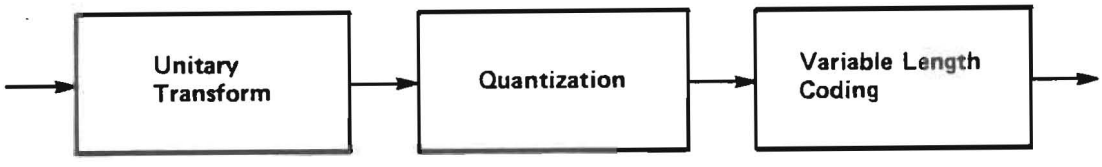


Fig. 9 Transform coder.

The coding scheme employed here has many characteristics in common with the Scene Adaptive Coder described by Chen and Pratt [11]. The main addition is that frequency weighting is employed to account for the different sensitivity of the HVS to the different transform coefficients. The weighting function (28) was determined for full frame Fourier transform; however, important modifications must be made for a block-wise transform. The frequency-weighted transform coding scheme presented here has also been used for interframe coding with good results [12].

Transform

The Discrete Cosine Tranform (DCT) is used, since its energy compaction performance is close to the theoretically optimal Karhunen-Loève transform for Markov sources (exponential covariance function) with correlation in the range 0.5 to 1 [13]. For a two-dimensional source with separable covariance, the performance will also be close to optimum if the correlation is exponential in both dimensions. In the previous section we focused our attention on an isotropic covariance function. Natarajan and Ahmed have shown that the energy compaction is close to optimum for the isotropic case also [14].

The two-dimensional DCT is defined as

$$X(u,v) = \frac{C(u)C(v)}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x(i,j) \cos \frac{(i-\frac{1}{2})(u-1)\pi}{N} \cos \frac{(j-\frac{1}{2})(v-1)\pi}{N} \tag{29}$$

$$C(u) = \begin{cases} 1 & \text{for } u = 1 \\ \sqrt{2} & \text{for } u = 2,3,\ldots,N \end{cases}$$

for $u, v = 1, 2, \ldots, N$.

A large blocksize is more efficient for decorrelation of a stationary source; however, when the statistics are non-stationary a smaller blocksize can adapt better to the local statistics.

Generally, a smaller blocksize gives better image quality at high bit rates (high quality), while a larger blocksize is superior at low rates [15]. Based on previous experience, a blocksize of 8×8 was chosen for the current experiments.

Quantization

Zonal coding and threshold coding are the two most commonly used quantization strategies. In a *zonal coder* a bit assignment is decided for each transform coefficient. More bits are usually assigned to low sequency (frequency) coefficients. Adaptivity can be achieved by choosing one out of several bit assignments. Each transform block is classified into one of M classes, and then quantized using the corresponding bit assignment.

Threshold coding is a procedure where all transform coefficients with a magnitude exceeding a certain threshold are quantized and transmitted. It is inherently adaptive; different coefficients are transmitted from block to block. A uniform quantizer has been used in several studies, e.g., by Tescher [16]. In his scheme, the transformed coefficients are scanned according to Fig. 10. Long runs of zeros are coded with a run-length code and a separate Huffman table.

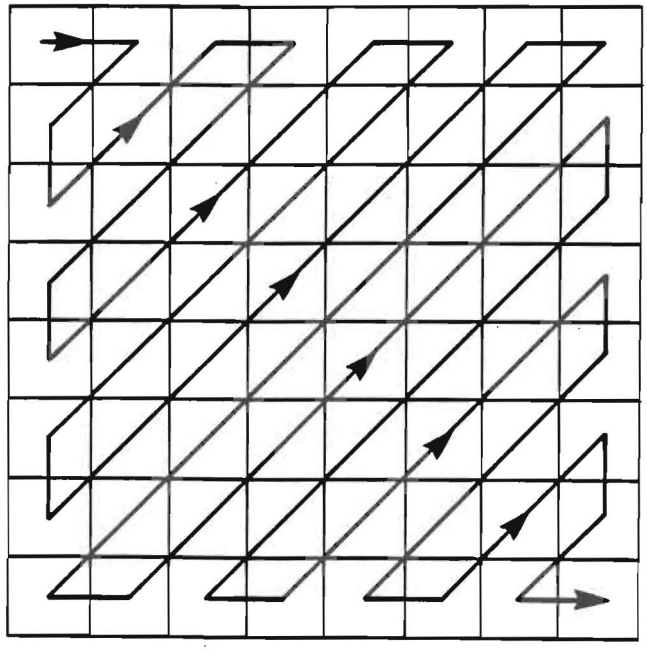


Fig. 10 Zig-zag scanning of transform coefficients (from [17]).

Chen and Pratt [11] designed an adaptive transform coder with many characteristics similar to Tescher's coder: zig-zag scanning of each block and run-length coding of zeros; however, the uniform quantizer characteristic has been modified, as outlined in Fig. 11. The zero level has been extended to increase the probability of a zero. Two Huffman codes are used, one for amplitudes and one to code runs of zeros.

The different sensitivity to quantization noise in different transform coefficients can be accounted for by weighting them. Then, the weighted coefficients are quantized such that

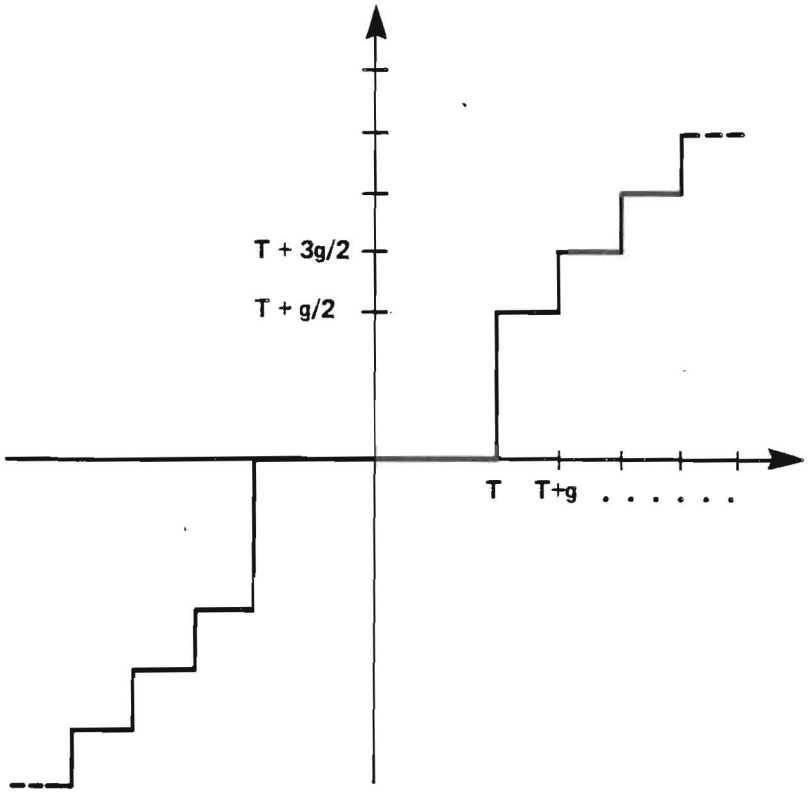


Fig. 11 Threshold coder quantization characteristic.

they get the same mean square error. An inverse weighting is performed in the receiver. Thus, the error in each coefficient depends on the weight.

Most work on frequency weighted transform image coding has been done using a full frame Fourier transform, e.g., Yan and Sakrison [18]. They applied the weighting function (28) to the transform coding in a two-component coding scheme. When smaller transform blocks are used, it is not possible to use the same weighting function due to edge effects. Because of the finite window, the spectra of the transform basis functions are not located at discrete frequencies, as illustrated in Fig. 12. In [12] it was shown that a modified low frequency portion of the weighting function (28) was necessary for small blocksizes. Thus, the weighting function $W'(f_r)$ was defined, where

$$W'(f_r) = \begin{cases} 1, & |f_r| < 8 \text{ cycles/degree} \\ W(f_r) & \text{otherwise} \end{cases} \tag{30}$$

Furthermore, the weights

$$\begin{aligned} w_{i_1 i_2} &= W'(f_r) \\ f_r^2 &= f_1^2 + f_2^2 \\ f_j &= \begin{cases} \frac{i_j - 1}{2N_j} f_{sj}, & i_j = 1, 2, \ldots, N_j - 1 \\ f_{sj}/2, & i_j = N_j \end{cases} \qquad j = 1, 2 \end{aligned} \tag{31}$$

were suggested for an $N_1 \times N_2$ DCT, where $N_1, N_2 \geq 8$. If the weighted transform coefficients are quantized with the same mean square error, the error spectrum will give a good fit to the desired spectrum (given in eq. (3)) for radial frequencies above 8 cycles per degree.

In the experiments, the transform coefficients were weighted according to eq. (31), and then quantized by the same quantizer. The weighting is equivalent to a scaling of the quantizer characteristic. The weighted transform coefficients were quantized by the quantizer outlined in Fig. 11, which is defined by a threshold T and a quantization step g. The ratio $T/g = 1.5$ was used in the experiments.

Quantizer overload is avoided by using many levels. Hence, infrequent large errors should not occur in the picture. The strategy is supported by our knowledge of subjective image quality; the subjective quality is based on the local error in a few critical areas rather than on a global average [19,20].

Variable Length Coding and Rate Control

The variable length coding was performed as described in [11]. The DC coefficient was coded with a fixed number of bits. The AC coefficients were scanned according to a zig-zag pattern as outlined in Fig. 10, and Huffman coded. The code table contains codewords for amplitudes, runs of zeros, and an End-of-Block word indicating that the remaining coefficients were quantized to zero.

In most applications, a fixed output bit rate is required. This was achieved by scaling the quantizer characteristic. The scaling factor was found for each frame by an iterative search algorithm. Although inefficient for real-time implementation, this method facilitates comparison of different coding parameters without going into the details of rate buffer control.

4. Experiments

The transform coding experiments were performed on a VAX-11/780, and the results were displayed on the HDVS (High resolution Digital Video sequence Store) of Bell-Northern Research/INRS Computer Laboratory. This system allows real-time acquisition and display of 525-line component video.

The experiments were performed on a sequence recorded by a broadcast quality TV camera. Only the luminance component, sampled at 13.5 MHz, was processed in the experiments described here. The active window consists of 256×212 pels per frame, which is approximately 15 % of the full screen area.

To be able to perform experiments without access to a high-definition video signal we consider the active window to be an even smaller part (3.7 %) of a system with twice the number of lines compared to NTSC, i.e., 1050 lines per frame, 2:1 line interlacing, and 30 frames/60 fields

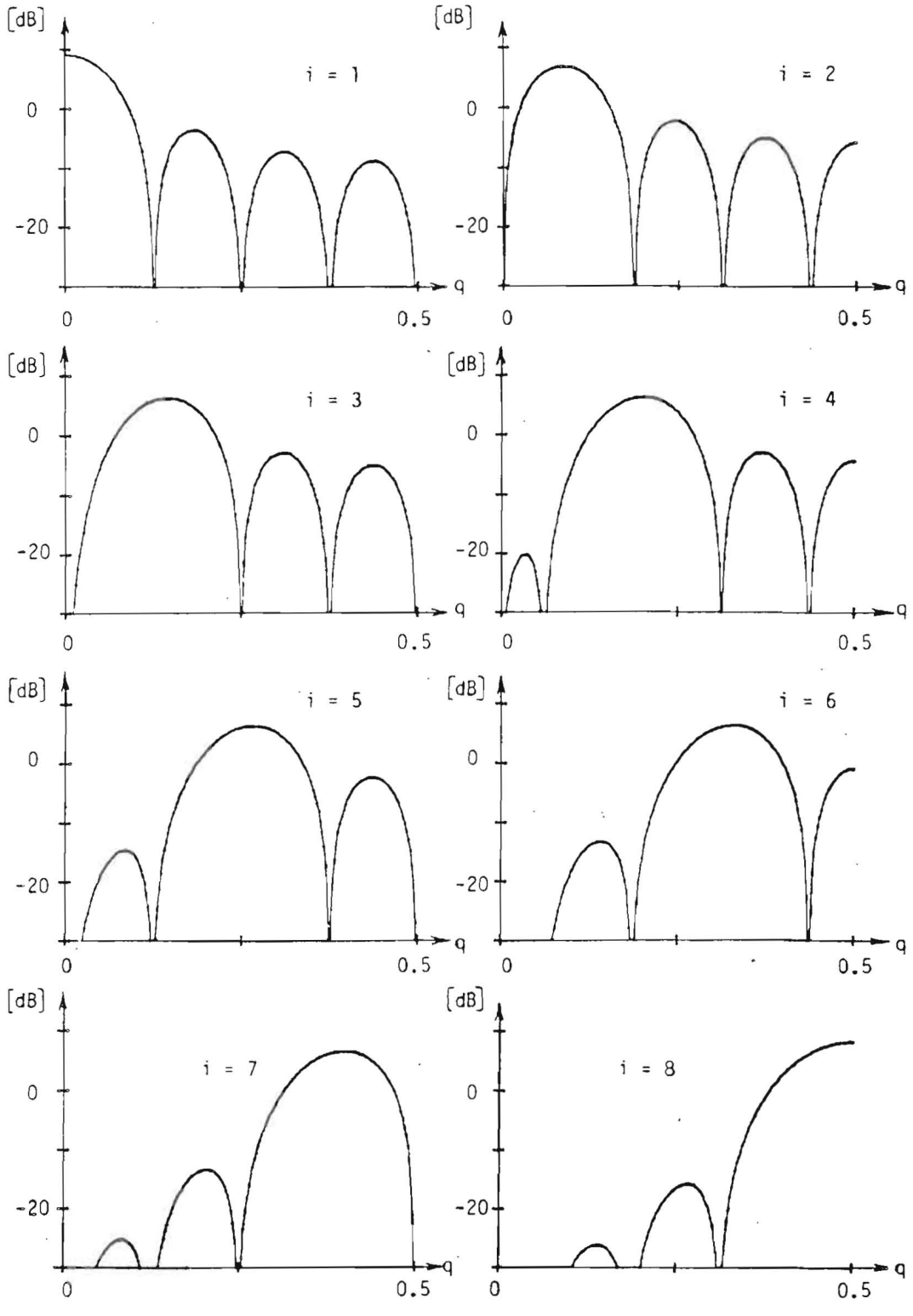


Fig. 12 Spectra of the 8-point DCT basis functions.

per second. The aspect ratio is 4:3 to be compatible with NTSC, and the Kell factor (defined here as the ratio of horizontal to vertical resolution) is chosen to be 0.5, the same value as in the NHK 1125-line system [21]. This corresponds to a bandwidth of 12 MHz, assuming the same ratio between active line period and horizontal blanking interval as in NTSC. All frequencies would be 4 times higher in such a system, e.g., the sampling frequency used by the DVS corresponds to 54 MHz in the 1050-line system.

Two cases have been studied: coding of a (supposed) "1050-line" HDTV signal, and coding of a 525-line signal in an NTSC-compatible enhanced TV (EQTV) system. The EQTV case will illustrate the stricter requirements that are posed on the 525-line standard when the receiver employs up-conversion and a high-definition monitor.

In the first case, an "HDTV original" was obtained by bandlimiting the luminance to $12/4 = 3$ MHz. The lowpass filter was an equi-ripple symmetric FIR filter of length 35. The "HDTV" signal was transform coded with different parameters and displayed for subjective evaluation.

The second case included down- and up-conversion. The HDTV original was down-converted to 128×53 pels per field by intrafield processing of every 256×106 input field. The spatial relationships between an HDTV frame and a down-converted frame are shown in Fig. 13. The output lines are obtained by linear interpolation in the vertical direction with the weighting factors 3/4 and 1/4. In the horizontal direction, a 2:1 subsampling is performed; no prefiltering is necessary.

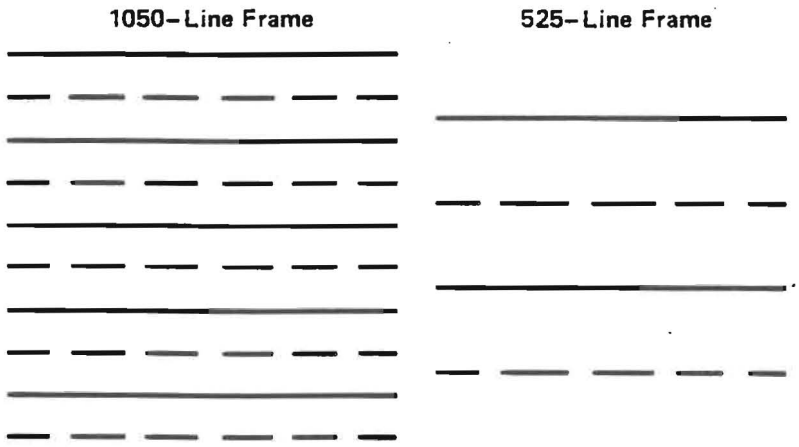


Fig. 13 Spatial location of lines in 1050-line frame and the down-converted 525-line frame. Solid lines indicate even field, hatched lines indicate odd field.

The bandlimitation introduced by the NTSC channel was simulated by applying a low-pass filter to the down-converted signal. A 25-point symmetric FIR filter was used. It is an equi-ripple design with 0.16 dB ripple below 3.75 MHz, and 34 dB attenuation above 4.75 MHz.

Then, the "NTSC" luminance window of 128×53 pels was transform coded, up-converted to 256×106 pels, and displayed on the HDVS. Following Tonge [22], the up-conversion process

is divided into two stages:

I. Conversion from 525-line/60Hz/2:1 interlaced to 525/60/1:1 non-interlaced format.

II. Conversion from 525-line non-interlaced to the desired 1050-line/60Hz/2:1 interlaced signal.

Stage I involves motion-adaptive processing, while Stage II is a straightforward digital interpolation.

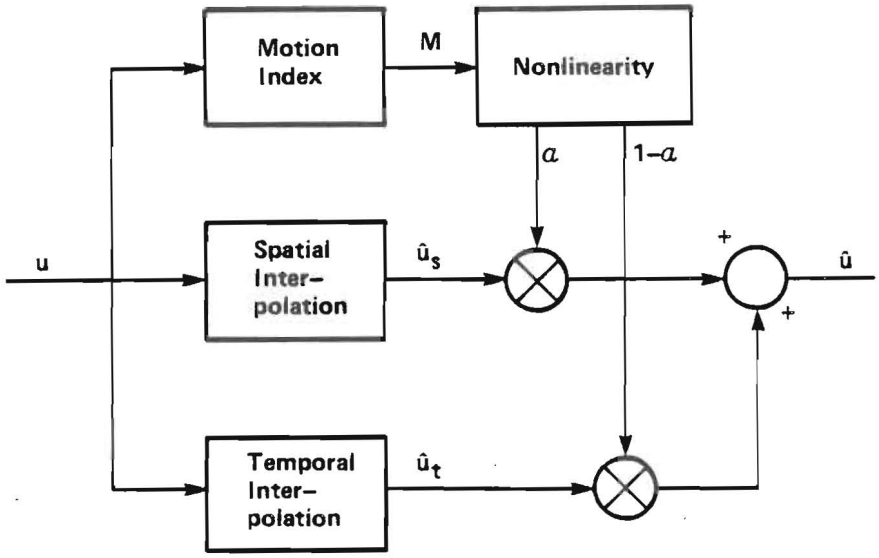


Fig. 14 Motion-adaptive interpolation system (from [5]).

A motion-adaptive interpolator is shown in Fig. 14. It weighs the outputs of a temporal and a vertical interpolator depending on a motion index α. The interpolators are implemented as the average in temporal and vertical direction respectively:

$$\begin{aligned} \hat{u}_t(n,m,k) &= (u(n,m,k-1) + u(n,m,k+1))/2 \\ \hat{u}_s(n,m,k) &= (u(n,m-1,k) + u(n,m+1,k))/2 \end{aligned} \tag{32}$$

The indices n, m, k indicate horizontal, vertical, and temporal direction, respectively.

The control parameter α can only take on two different values, i.e., either vertical or temporal interpolation is used. The decision is based on the temporal and vertical differences in a 3×3 window:

$$\begin{aligned} D_{temp} &= \sum_{i=-1}^{1} \sum_{j=-1}^{1} |u(n+j, m+i, k+1) - u(n+j, m+i, k-1)| \\ D_{vert} &= \sum_{i=-1}^{1} \sum_{j=-1}^{1} |u(n+j, m+1+i, k) - u(n+j, m-1+i, k)| \end{aligned} \tag{33}$$

Vertical interpolation ($\alpha = 1$) is used if the temporal difference is larger than the vertical difference.

$$\alpha = \begin{cases} 0 & \text{if } D_{temp} \leq D_{vert} \\ 1 & \text{if } D_{temp} > D_{vert} \end{cases} \tag{34}$$

By comparing two differences we avoid the problem of choosing a motion detector threshold, which would be dependent on the noise level. Hence, it should not be necessary to know the input SNR with the above criterion.

Stage II consists of two linear filters: a horizontal 1:2 interpolator and a vertical quarter-pel shifter. The horizontal interpolation filter is shown in Fig. 15 for different values of the HF boost parameter β_h. The value $\beta_h = 0.25$ was used in the experiments. The filter performing the fractional pel shift (1/4 pel) is shown in Fig. 16. This filter is used to shift the odd field downwards, while its mirrored version $h_{shift}(-n)$ is applied to the even field to shift it 1/4 pel upwards. The cascade of the down-converter and the quarter-pel shifter gives the transfer function outlined in Fig. 17. The vertical HF boost parameter was chosen to be $\beta_v = 0.75$.

Twelve frames of the test sequence were coded and displayed on the HDVS. The viewing distance corresponded to three times picture height for a 960-line display. For the HDTV original the visibility threshold was at an approximate bit rate of 1.2 bit per pel. The inclusion of frequency weighting gave a very marginal improvement.

The down-converted picture needed a bit rate of 2.8 b/pel to give a coding error at the visibility threshold after up-conversion. Thus, assuming a 4:1 ratio of sampling frequencies, the HDTV bit rate is 1.7 times higher than the EQTV rate. This is quite comparable to the theoretical result obtained in Section 2. Furthermore, the HDTV picture at the same bit rate per unit area as the EQTV picture, i.e., 0.7 b/pel, looked significantly better than the uncoded sequence that had passed the down- and up-converter.

5. Conclusion

A theoretical study of optimal coding performance as a function of resolution has been carried out using rate distortion theory. This predicts, as expected, that coding efficiency increases as the sampling density increases. Although this result has been known qualitatively for some time, few quantitative results have been available.

A transform coder operating at different sampling densities was used to determine how these theoretical results compare with actual coder performance for conditions approximating those of HDTV and EQTV. Results quite comparable to the theoretical predictions were obtained: the coded HDTV signal requires a bit rate only about 1.7 times greater than the EQTV signal, assuming the same aspect ratio, when both are coded near the threshold of visibility of coding distortion. This is much smaller than the 4:1 ratio of analog bandwidths or PCM bit rates. Furthermore, at the same bit rate, the coded HDTV signal gives a significantly better picture quality than the upconverted EQTV signal. These results should be considered in the planning of digital transmission systems for future television services.

6. References

[1] W.F. Schreiber, "Picture coding," *Proc. IEEE*, vol. 55, pp. 320–330, March 1967.

[2] S. Sabri and B. Prasada, "Coding of broadcast TV signals for transmission over satellite channels," *IEEE Trans. Commun.*, vol. COM-32, pp. 1323–1330, Dec. 1984.

[3] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice-Hall 1971.

[4] J.B. O'Neal, Jr. and T. Raj Natarajan, "Coding isotropic images," *IEEE Trans. on Information Theory*, vol. IT-23, pp. 697–707, Nov. 1977.

[5] E. Dubois, "The sampling and reconstruction of time-varying imagery with application in video systems," *Proc. IEEE*, vol. 73, April 1985.

[6] D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Trans. Commun.*, vol. COM-25, pp. 1251–1267, Nov. 1977.

[7] F. W. Campbell and J.J.Kulikowski, "Orientational selectivity of the human visual system," *J. Physiol. (London)*, vol. 187, pp. 437–441, 1966

[8] H. Mostafavi and D. Sakrison, "Structure and properties of a single channel in the human visual system," *Vision Res.*, vol. 16, pp. 957–968, 1976.

[9] L. Maffei and A. Fiorentini, "The visual cortex as a spatial frequency analyzer," *Vision Res.*, vol. 13, pp. 1255–1267, 1973.

[10] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 525–536, July 1974.

[11] W.-H. Chen and W. K. Pratt, "Scene adaptive coder," *IEEE Trans. Commun.*, vol. COM-32, pp. 225–232, March 1984.

[12] S. Ericsson, "Frequency weighted interframe hybrid coding," Report No. TRITA-TTT-8401, Telecomm. Theory, The Royal Inst. of Tech., Stockholm, Jan. 1984. Also presented at the *1984 Picture Coding Symposium*, Rennes, France, 3–5 July, 1984.

[13] A. K. Jain, "A sinusoidal family of unitary transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 356–365, Oct. 1979.

[14] T. R. Natarajan and N. Ahmed, "Performance evaluation for transform coding using a nonseparable covariance model, *IEEE Trans. Commun.*, vol. COM-26, pp. 310–312, Feb. 1978.

[15] A. G. Tescher, "Block size considerations for adaptive image coding," *Proc. NTC '82*, pp. E1.2.1–E1.2.4, 1982.

[16] A. G. Tescher, "A dual transform coding algorithm," *Proc. NTC '79*, pp. 53.4.1–53.4.4, 1979.

[17] J. R. Parsons and A. G. Tescher, "An Investigation of MSE Contributions in Transform Image Coding Schemes," *Proc. SPIE*, vol. 66, pp. 196–206, 1975.

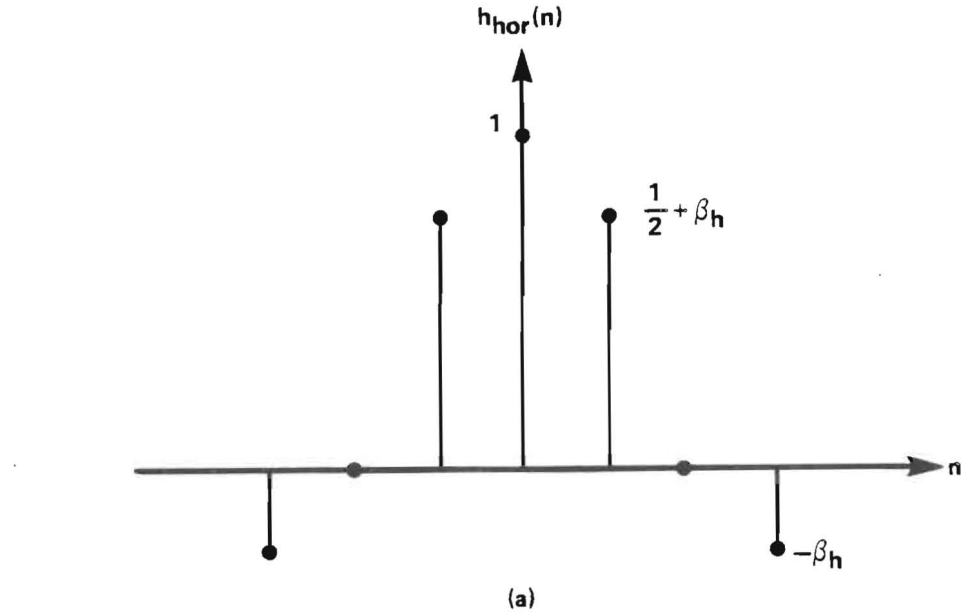
[18] J. K. Yan and D. J. Sakrison, "Encoding of images based on a two-component source model," *IEEE Trans. Commun.*, vol. COM-25, pp. 1315–1322, Nov. 1977.

[19] J. O. Limb, "Distortion criteria of the human viewer," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, pp. 778–793, Dec. 1979.

[20] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. COM-30, pp. 1679–1692, July 1982.

[21] T. Fujio, "High definition television systems: Desirable standards, signal forms, and transmission systems," *IEEE Trans. Commun.*, vol. COM-29, pp. 1882–1891, Dec. 1981.

[22] G. Tonge, "Signal processing for higher-definition television," *IBA Technical Review*, vol. 21, pp. 13–26, Nov. 1983.



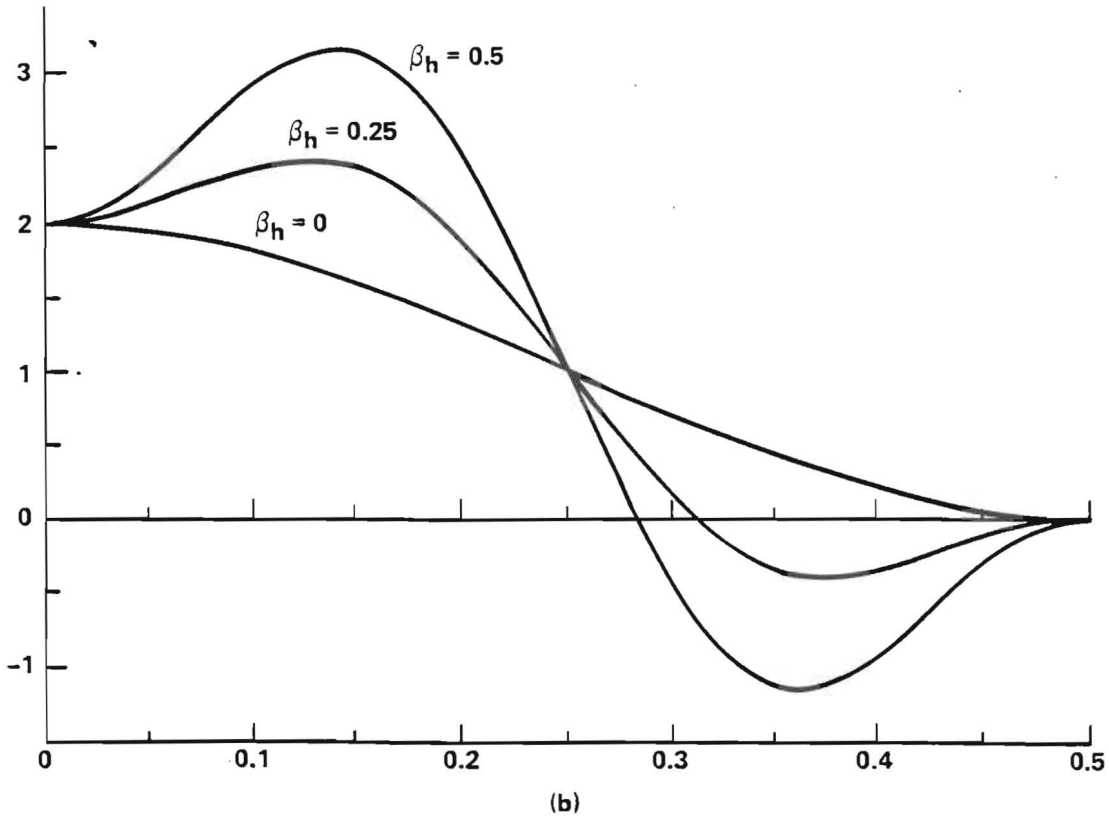
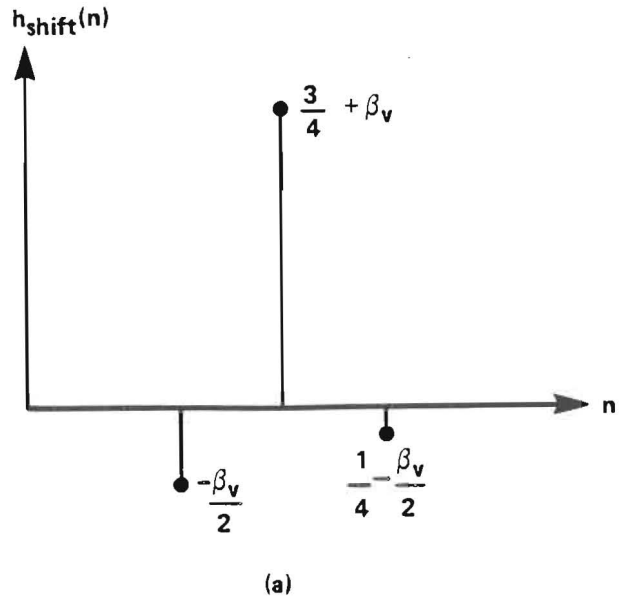
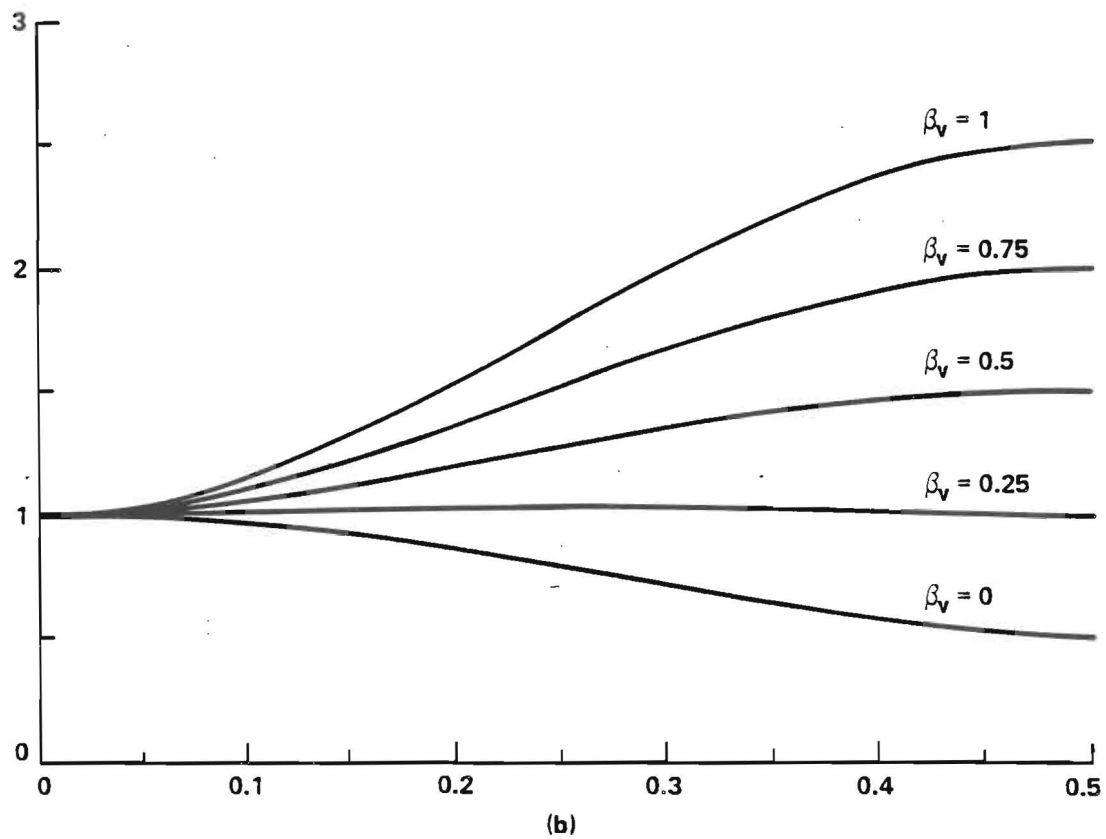


Fig. 15 Horizontal interpolation filter. (a) Impulse response. (b) Frequency response for different values of β_h.



(a)



(b)

Fig. 16 Quarter-pel shifter. (a) Impulse response.
(b) Amplitude response for different values of β_v; $f_s = 480$ c/ph.

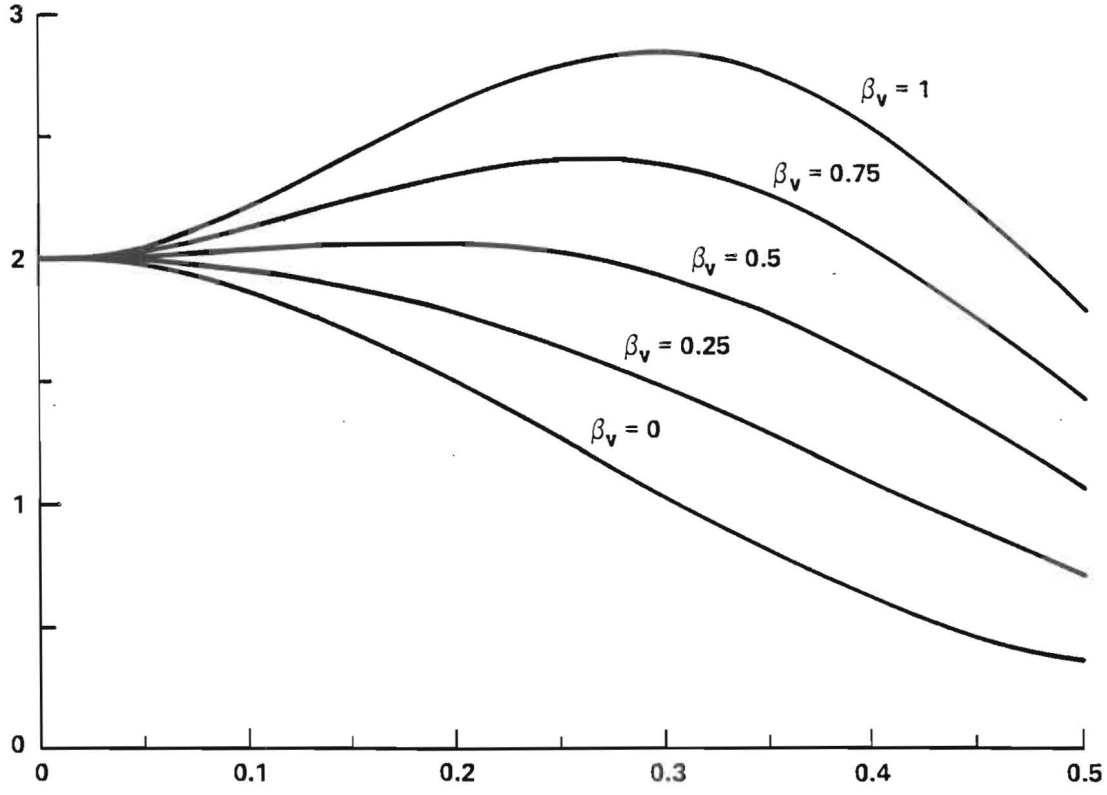


Fig. 17 Vertical amplitude response of down-converter cascaded with up-converter for different values of β_v; $f_s = 480$ c/ph.