INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.240

(06/2004)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Measurement of the quality of service

Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform

ITU-T Recommendation J.240

ITU-T Recommendation J.240

Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform

Summary

This Recommendation presents an effective framework for remote monitoring of video quality for contribution and primary distribution of digital television transmission. In this framework, image features are extracted at each link point in the transmission chain using spread-spectrum and orthogonal transform. The extracted image features, i.e., coefficients, are transmitted to the central monitoring room by data circuit that is separated from the video transmission circuit and picture quality is estimated by comparing the coefficients.

A framework for extracting the coefficients is presented below. An example of coefficient extraction suitable for PSNR estimation is described in Appendix I. The theoretical background for PSNR estimation and the simulation results showing the effectiveness of this framework are also described in Appendix I.

Source

ITU-T Recommendation J.240 was approved on 29 June 2004 by ITU-T Study Group 9 (2001-2004) under the ITU-T Recommendation A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

ITU-T Recommendation J.240

Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform

1 Scope

This Recommendation presents a framework for automatic remote picture quality monitoring. The proposed framework extracts transform coefficients of the transmitted picture and transmits them to the monitoring operator by an additional data circuit. As the coefficients extraction and transmission is performed separately from the mainstream video transmission, this framework does not affect the quality of transmitted video and thus this is applicable for monitoring video transmission that requires high quality such as contribution and primary distribution.

2 References

2.1 Normative reference

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

– ITU-T Recommendation J.143 (2000), *User requirements for objective perceptual video quality measurements in digital cable television*.

2.2 Informative references

– ITU-T Recommendation J.144 (2004), *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*.

– ITU-T Recommendation J.147 (2002), *Objective picture quality measurement method by use of in-service test signals*.

– ITU-R Recommendation BT.656-4 (1998), *Interfaces for digital component video signals in 525-line and 625-line television systems operating at the 4:2:2 level of Recommendation ITU-R BT.601 (Part A)*.

3 Definitions

This Recommendation defines the following term:

3.1 **node**: Link point in the transmission chain.

4 Abbreviations

This Recommendation uses the following abbreviations:

DEC Decoder

ENC Encoder

OT Orthogonal Transform

PSNR Peak Signal-to-Noise Ratio

SS Spread Spectrum

WHT Walsh-Hadamard Transform

5 Picture quality monitoring based on transform coefficients extraction

5.1 Configuration of transmission chain

Configuration of the assumed transmission chain is shown in Figure 1. It is assumed that baseband signal is present at each edge (nodes 0 and 1) of the video transmission circuit in which coding and decoding equipment are included.

The video signal branches away from the mainstream at nodes 0 and 1 and the branched signals are put into a feature extractor. The extracted feature information is transmitted to a PSNR estimator via a data circuit that is separated from the video transmission circuit. The PSNR estimator measures the transmission quality PSNR using feature information at each node.

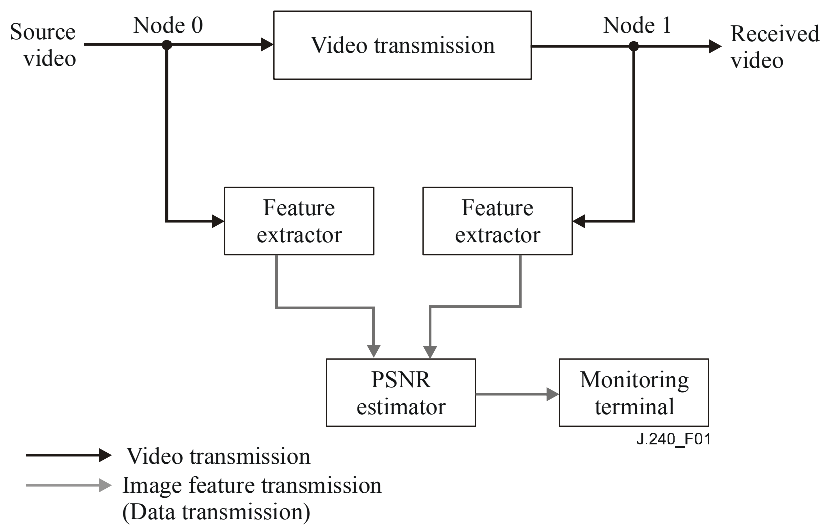

Figure 1/J.240 – Basic configuration

Although Figure 1 is configured for monitoring a single link, this can be extended to monitoring of a transmission chain by cascading the single link as shown in Figure 2. In Figure 2, feature extractors are connected at each node of the transmission chain and the feature information is transmitted to the PSNR estimator. At the PSNR estimator, feature information of two nodes between which transmission quality is monitored is selected and PSNR is estimated using the selected feature information.

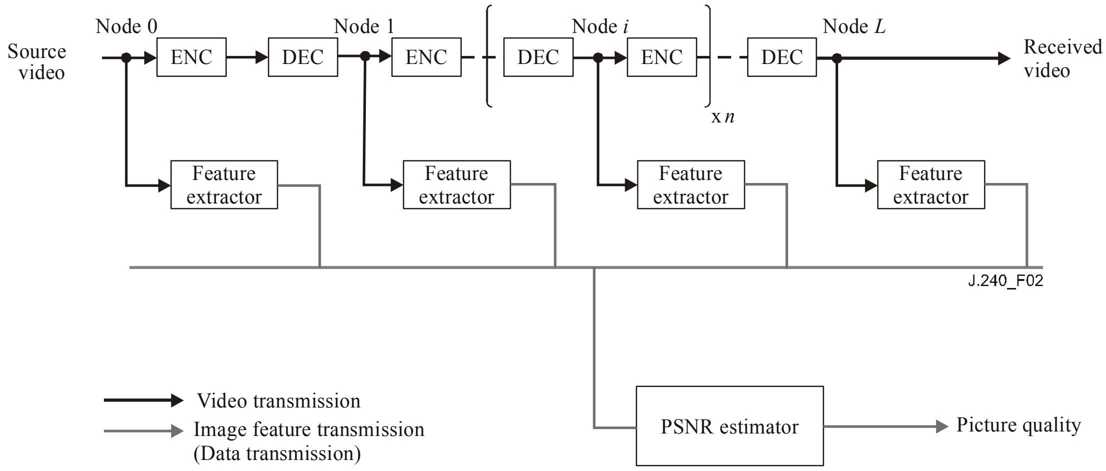

Figure 2/J.240 – Transmission chain monitoring

5.2 Extraction of image feature

The image feature is extracted in each pixel block after dividing the input image into blocks of size $N_x \times N_y$. When the number of pixels in horizontal and vertical direction is not equal to the multiple of N_x and N_y respectively, additional pixels with a middle pixel value are padded to make the number of pixels equal to N_x and N_y.

Figure 3 shows the operation for extracting the image feature. The image feature is obtained by the combination of the spread spectrum and orthogonal transform. Let $x_f^{(b)}(n)$ and $s_{\mathrm{PN}}{}^{(b)}(n)$ denote the input signal of block b in frame f and PN sequence of block b respectively, where n is the index of pixels. Then, the image feature $R_f[b]$ is obtained as follows using orthogonal transform OT:

$$R_f[b] = \mathrm{Amp}\left\{X_f^{(b)}[k_0]\right\} \quad (1)$$

$$X_f^{(b)}[k] = (SSOT)\left\{x_f^{(b)}(n)\right\} \equiv OT\left\{x_f^{(b)}(n) s_{PN}{}^{(b)}(n)\right\} \quad (2)$$

where *SSOT* stands for Spread Spectrum and Orthogonal Transform, *Amp[Z]* for the amplitude component of Z, and k_0 for the position of the extracted coefficient.

$R_f[b]$ can also be defined as follows applying inverse orthogonal transform OT^{-1} after multiplying another PN sequence s_{PN2}:

$$R_f[b] = \mathrm{Amp}\left\{y_f^{(b)}[n_0]\right\} \quad (3)$$

$$y_f^{(b)}[n] = (SSOT_2)\left\{x_f^{(b)}(n)\right\} \equiv OT^{-1}\left\{X_f^{(b)}(k) s_{PN2}{}^{(b)}(k)\right\} \quad (4)$$

where n_0 is the position of the extracted coefficient (in this case, pixel). This second implementation strengthens the effectiveness of spread spectrum by randomizing the spatial domain in addition to the frequency domain.

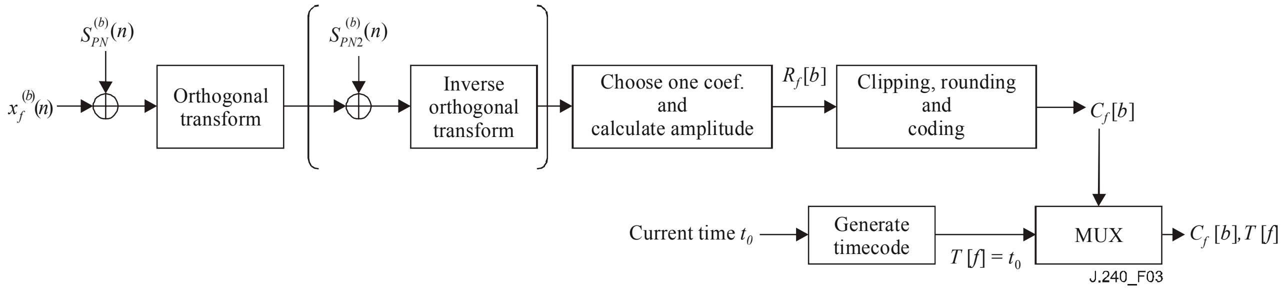

Figure 3/J.240 – Operation in feature extractor

As for PN sequence, different sequences can be used for each block. But the same PN sequences must be used at all nodes in the transmission chain for each block in the picture. This is a mandatory requisite for synchronization between nodes, which is described in 5.3.

5.3 Clipping, rounding and coding of coefficients

The extracted coefficient $R_f[b]$ is outputted to the data circuit. As $R_f[b]$ is generally a real number, clipping and rounding are applied to $R_f[b]$ in order to express the coefficient by limited bit length. Compression coding may be applied optionally for data compression provided that the probability distribution of $R_f[b]$ is biased and thus the amount of information is expected to be reduced. All coded coefficients in a frame or field are packed and data is outputted.

5.4 Additional information

As additional information for the PSNR estimator, time information $T[f]$ can be outputted at the feature extractor. Time information describes the correspondence of the frame or field number with the output time of the coefficients, which is multiplexed with coded coefficients $C_f[b]$ at the final stage of the feature extractor. When a timecode such as LTC and VITC is available, the timecode can be transmitted instead of output time. This is used for delay adjustment of image features in the PSNR Estimator.

5.5 PSNR estimation

The operation in the PSNR Estimator is shown in Figure 4. First, multiplexed information sent from two nodes which are located at each edge of the monitored link is demultiplexed and coded coefficients $C_{0f}[b]$, $C_{1f}[b]$ and time information $T_0[f]$, $T_1[f]$ are obtained. When the order of arrival of the coefficients and the time information does not conform to the order of output at the feature extractor, the received information have to be sorted to ensure that the coefficients and the time information are arranged in frame number order.

Next, in case the coefficients are coded, decoding them is applied and coefficients $D_{0f}[b]$ and $D_{1f}[b]$ are obtained. Then delay between two nodes is adjusted using the coefficients and the time information $T_0[f]$ and $T_1[f]$.

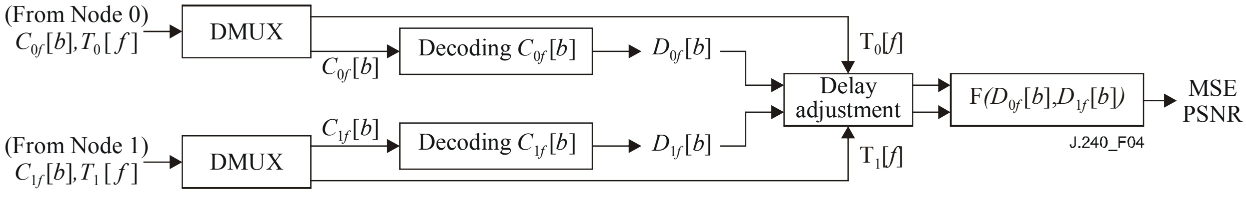

Figure 4/J.240 – Operations in PSNR estimator

The necessity for delay adjustment is shown in Figure 5. There is a video transmission delay d_V between two nodes and data transmission delay d_{C0} and d_{C1} between each node and the PSNR estimator. The coefficients of the frame "F1" that is displayed on time t_1 on the transmission side arrives at the PSNR estimator on time $t_1 + d_{C0}$ from Node 0 and $t_1 + d_V + d_{C1}$ from Node 1 respectively. Thus, arrival time of the coefficients of two nodes is generally different and thus it is not guaranteed that the coefficients for the same frame arrive at the PSNR estimator simultaneously from two nodes. Further, some transmission procedures of the data circuit do not ensure that the order of the coefficients reception is the same as the transmission order (i.e., frame number order). Therefore, it is meaningless to compare the coefficients that arrive at the PSNR estimator at the same time from two nodes. The PSNR estimator should synchronize the coefficients from two nodes at the "delay adjustment" block after sorting the coefficients in transmission order at DMUX output. Time information $T_0[f]$ and $T_1[f]$ can be utilized for this purpose.

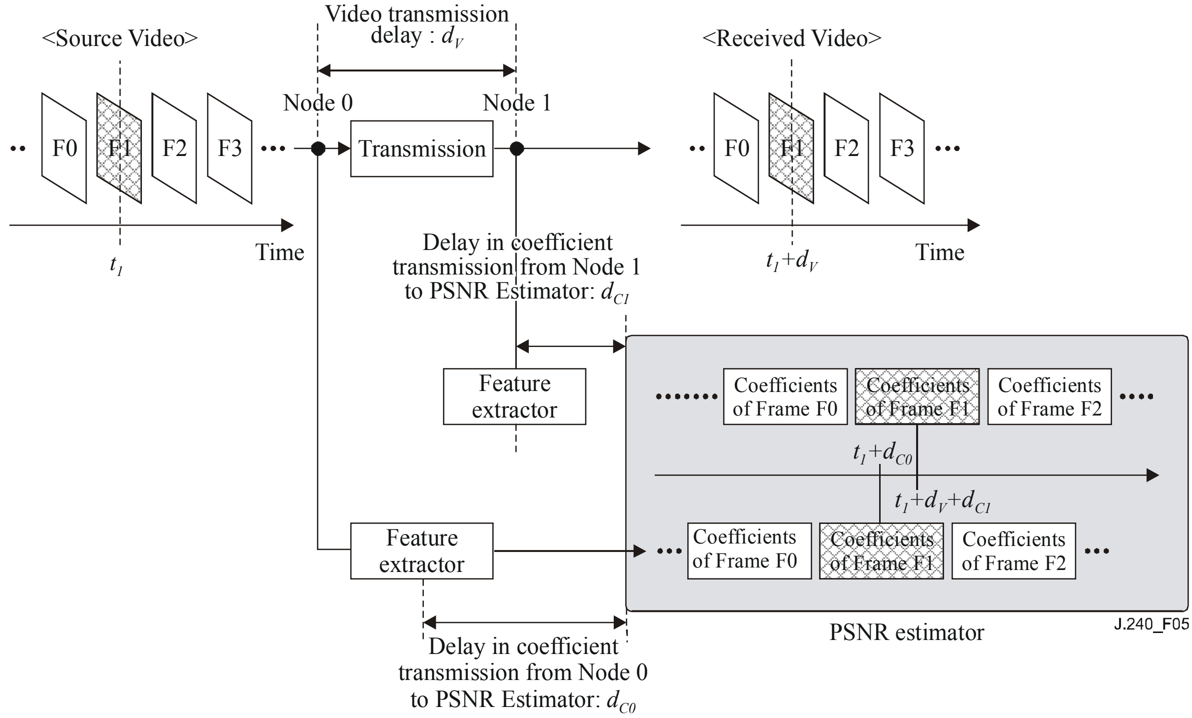

Figure 5/J.240 – Delay adjustments between nodes

After synchronizing the coefficients, the degradation of picture quality caused by video transmission is calculated using the coefficients $D_{0f}[b]$ and $D_{1f}[b]$. Let MSE denote the degradation. It can be expressed as follows:

$$\text{MSE} = Average\left(D_{0f}[b] - D_{1f}[b]\right)^2 \tag{5}$$

PSNR is derived from MSE as follows:

$$\text{PSNR} = 20\log_{10}\sqrt{\frac{255^2}{MSE}} \tag{6}$$

By repeating the above procedure in all frames, the picture quality of the link between Node 0 and Node 1 can be obtained. When PSNR is to be used for subjective quality estimates, it is important to use cropped PSNR, i.e., PSNR calculated using video regions that are active in both the source and the processed video. It is the cropped PSNR that will correlate most highly with the subjective quality ratings.

Appendix I

Implementation example

I.1 Feature extraction

Walsh-Hadamard Transform (WHT) is applied as an orthogonal transform for image feature extraction. The parameters for the feature extractor is $N_x = N_y = 8$, and the inverse transform is included. Thus, the extracted coefficient $R[b]$ is derived as follows:

$$R[b] = x_{SSS}^{(b)}(n_0) \quad \text{(I-1)}$$

$$x_{SSS}^{(b)}(n) = WHT^{-1}\left[WHT\left\{ x^{(b)}(n)\, s_{PN}^{(b)}(n) \right\} s_{PN\,2}^{(b)}(n) \right] \quad \text{(I-2)}$$

I.2 Coding of coefficients

Details of coding the coefficients are presented below. When the feature extraction scheme shown above is applied to 8-bit/pixel input signal, the coefficient $R[b]$ finally has 15-bit length as shown in Figure I.1 (bit length theoretically increases by $\log_2 \sqrt{N_x N_y}$ for $N_x \times N_y$ WHT because transform coefficients are scaled by $\sqrt{N_x N_y}$ when $N_x \times N_y$ WHT is applied).

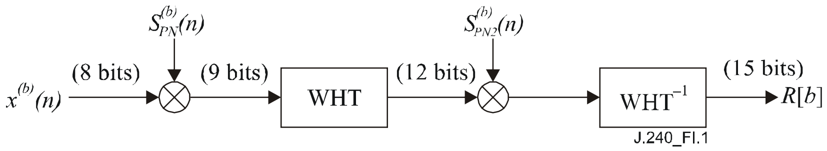

Figure I.1/J.240 – Increase of bit length in feature extractor

Assuming that the video signal format is 720 × 480, 30 fps, the required bandwidth for the reference path will be $\frac{720 \times 480}{8 \times 8} \times 15 \times 30 = 2430000$ [bit/s]. However, this is not a practical rate and thus quantization is applied to reduce the bit length of $R[b]$. For example, when $R[b]$ is quantized to 8 bit, the bit rate of the coefficients becomes $\frac{720 \times 480}{8 \times 8} \times 8 \times 30 = 1296000$ [bit/s], which is almost half the bit rate than when using 15 bits. If further bit-rate reduction is required, enlargement of block size for feature extraction (i.e., N_x and N_y) and spatial decimation of coefficients can be applied.

I.3 MSE calculation

Equation I-4 only defines the use of a certain function using coefficients from two nodes. However, in a typical application, MSE is derived by the following equation:

$$\text{MSE} = \sum_{b=0}^{N_b - 1} \left(D_i[b] - D_j[b] \right)^2 / N_b / scale \quad \text{(I-3)}$$

where N_b and *scale* denote the number of blocks in the frame and the scale factor that depends on the block size for feature extraction and the number of quantization bits for $R_f[b]$.

I.4 Theoretical backgrounds

In this clause, the theoretical model of PSNR estimation based on the comparison of coefficients is shown below. Note that the inverse orthogonal transform is not included in this clause for simplicity.

Let $x^{(b)}(n)$ and $x'^{(b)}(n)$ denote the pixel value of the source and received picture of block b respectively. Then, the MSE of a field in the received picture can be expressed as follows:

$$MSE = \sum_{b=0}^{N_b-1} \sum_{n=0}^{N_p-1} \left[x^{(b)}(n) - x'^{(b)}(n)\right]^2 / \left(N_p N_b\right) \tag{I-4}$$

where N_b shows the number of blocks in the field and N_p shows the number of pixels in a block, respectively. Next, let $s_{PN}^{(b)}(n)$ denote the PN sequence of the block and spread spectrum of $x^{(b)}(n)$ and $x'^{(b)}(n)$. Then the MSE can be written as follows because $[S_{PN}^{(b)}(n)]^2 = 1$:

$$MSE = \sum_{b=0}^{N_b-1} \sum_{n=0}^{N_p-1} \left[x^{(b)}(n) \cdot s_{PN}{}^{(b)}(n) - x'^{(b)}(n) \cdot s_{PN}{}^{(b)}(n)\right]^2 / \left(N_p N_b\right) \tag{I-5}$$

From Parseval's equation, Equation I-5 can be modified as follows:

$$MSE = \sum_{b=0}^{N_b-1} \sum_{k=0}^{N_p-1} \left|X_{SS}{}^{(b)}[k] - X'_{SS}{}^{(b)}[k]\right|^2 / \left(N_p N_b\right) = \sum_{b=0}^{N_b-1} \sum_{k=0}^{N_p-1} \left|E_{SS}{}^{(b)}[k]\right|^2 / \left(N_p N_b\right) \tag{I-6}$$

where $X_{SS}{}^{(b)}[k]$ and $X'_{SS}{}^{(b)}[k]$ denote the spread spectrum of $x^{(b)}(n)$ and $x'^{(b)}(n)$ respectively, and $E_{SS}{}^{(b)}[k]$ can be expressed as $E_{SS}{}^{(b)}[k] = X_{SS}{}^{(b)}[k] - X'_{SS}{}^{(b)}[k]$.

From Equation I-6, since $X'_{SS}{}^{(b)}[k]$ can be obtained from the received picture, it is confirmed that MSE can be calculated by transmitting frequency components of the spread spectrum of the pictures. However, in this coefficient transmission method, all frequency components cannot be transmitted because of the bandwidth restriction of the reference data link. Therefore, this implementation method assumes transmission of only one component per block. When we denote k_0 as the frequency component to be transmitted, the power of frequency component k_0 can be expressed as follows:

$$\left|E_{SS}{}^{(b)}[k_0]\right|^2 = \sum_{k=0}^{N_p-1} \left|E_{SS}{}^{(b)}[k]\right|^2 / N_p + D^{(b)}[k_0] \tag{I-7}$$

In the above equation, if the spectrum of the source picture is spread uniformly, we can obtain $D^{(b)}[k] = 0$. However, the actual distribution of $|E_{SS}{}^{(b)}[k]|^2$ has a certain dispersion from the mean power in the block since frequency components of the spread spectrum has randomness. Then, the following equation is derived by applying Equation I-7 to Equation I-6:

$$MSE = \sum_{b=0}^{N_b-1} \left(\left|E_{SS}{}^{(b)}[k_0]\right|^2 - D^{(b)}[k_0] \right) / N_b = \sum_{b=0}^{N_b-1} \left|E_{SS}{}^{(b)}[k_0]\right|^2 / N_b - \sum_{b=0}^{N_b-1} D^{(b)}[k_0] / N_b \tag{I-8}$$

Further, as a characteristic of the spread spectrum, we assume that the dispersion $D^{(b)}[k_0]$ is balanced between the blocks in the frame, i.e., $\sum_b D^{(b)}[k_0] \approx 0$. Then, MSE can be approximated as follows:

$$MSE \cong \sum_{b=0}^{N_b-1} \left|E_{SS}{}^{(b)}[k_0]\right|^2 / N_b \tag{I-9}$$

Equation I-9 shows that MSE can be estimated by transmitting only one frequency component using the characteristics of the spread spectrum. Finally, by applying $X_{SS}^{(b)}[k] = A^{(b)}[k]e^{j\theta[k]}$ and $X'_{SS}^{(b)}[k] = A'^{(b)}[k]e^{j\theta[k]}$ assuming θ ≈ θ' (the phase component is hardly affected by transmission degradation), Equation I-9 can be expressed as follows:

$$MSE = \sum_{b=0}^{N_b-1} \left| \left(A^{(b)}[k_0] - A'^{(b)}[k_0]\right)e^{j\theta[k_0]} \right|^2 / N_b = \sum_{b=0}^{N_b-1} \left| A^{(b)}[k_0] - A'^{(b)}[k_0] \right|^2 / N_b \qquad \text{(I-10)}$$

Equation I-10 implies that estimated MSE can be finally expressed by difference of the amplitude components of the coefficients. In case of WHT, no complex numbers are included, so that the theory becomes much simpler.

I.5 Performances

Six SDTV test sequences were coded by MPEG-2 TM5 at four bit rates and the PSNR of the decoded picture is measured. WHT is used as orthogonal transform and inverse transform is also utilized. Four bit rates for the reference data link were examined. Furthermore, as a comparison method, PSNR measurement without spread spectrum was also examined. The experimental conditions are presented in detail in Table I.1.

Table I.1/J.240 – Experimental conditions

Test Sequence	Cheerleaders, Flamingoes, Green Leaves, Marching in, Mobile and Calendar, Soccer Action
Video format	704 × 480, 30 fps, 4:2:2
Codec	MPEG-2 Test Model 5
Bit length for transmitted coefficient	10 bit/sample
Coding bit rate	45, 22.5, 11.25, 5.125 Mbit/s
Block size (corresponding bit rate for reference data link)	8 × 8 (1584 kbit/s), 16 × 8 (792 kbit/s), 16 × 16 (396 kbit/s), 32 × 16 (198 kbit/s)

Figure I.2 and Table I.2 present the experimental results. The PSNR estimation error shows the average estimation error of 24 sequences (6 titles × 4 bit rates). The method with a spread spectrum has over 100 times as small estimation error as the comparison method and these show that application of the spread spectrum is very effective for PSNR estimation.

Table I.2/J.240 – Experimental results

Bit rate of reference path [kbit/s]	PSNR estimation error [dB]	
	with SS	without SS
1584	8.33E-04	5.77E+00
792	1.36E-03	5.64E+00
396	1.91E-03	5.70E+00
198	3.05E-03	5.92E+00
106	1.52E-02	6.90E+00
53	3.14E-02	6.96E+00

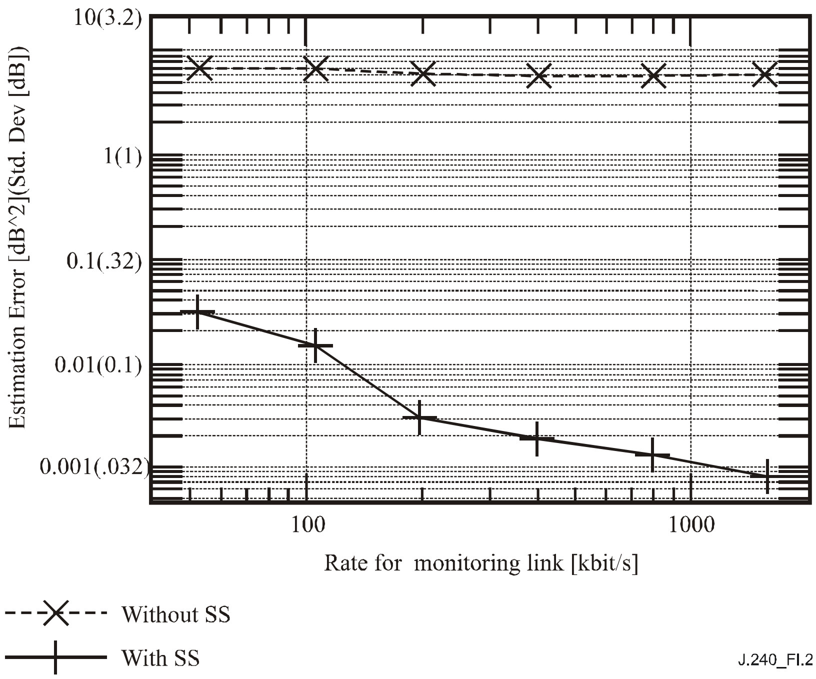

Figure I.2/J.240 – Comparison of estimation accuracy

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure, Internet protocol aspects and Next Generation Networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2004