Elbaz

IEEE

THE SECOND ISRAEL SYMPOSIUM ON CIRCUITS, SYSTEMS AND CONTROL

ISCSC '88

May 30 - June 1, 1988
Hotel Dan Accadia/Daniel Tower - Diplomat Hall

Switched-Prediction Interframe-DPCM Image-Sequence Coding for Video-Conference Applications

M. Elbaz and D. Malah

Electrical Engineering Department
Technion - Israel Institute of Technology
Technion city, Haifa 32000, Israel

Abstract - This paper presents an improved image-sequence coding scheme for Video-Conferencing transmission at 1.544 Mbits/sec. The coding scheme is based on DPCM with switched Interframe/Intraframe prediction. The predictions used in the scheme are motion-compensated prediction (Interframe), and spatial prediction (Intraframe). The switching is done in a forward manner, on a block by block basis. The switching decision law applied in the scheme selects the prediction type which leads to better overall performance for the corresponding block. Adaptive scalar quantization is used in the scheme based on dynamic bit allocation and normalization of the prediction error blocks. The performance of the presented scheme is compared to a motion-compensated Interframe-DPCM scheme and is proved to be better. It is also shown that this scheme produces an improved transient response as well.

I. INTRODUCTION

Image-sequence coding for Video-conferencing applications is being increasingly investigated in recent years [1]. Straightforward digitization of the image-sequence (PCM) requires about 50 Mbits/sec. The ultimate goal in the design of a video compression system is finding efficient techniques which gives acceptable image quality under a given channel rate constraint. Compression is possible due to the vast amount of redundancy which exists between consecutive frames in the image-sequence, and between pixels in the same frame.

In this paper we present a new coding scheme which utilizes switched prediction to exploit the two redundancy types mentioned above. The channel rate under consideration is the standard rate of 1.544 Mbits/sec, which corresponds to 0.236 bits/pixel.

Interframe-DPCM schemes with motion-compensated prediction [1,2] are very effective in exploiting the vast amount of redundancy between consecutive frames but completely ignore the redundancy existing between neighboring pixels in the current frame. This kind of redundancy can be exploited by spatial (intraframe) prediction.

The presented coding scheme utilizes the two prediction approaches by switching between them on a block by block basis. The prediction approach which yields a lower *quantization error*, on the average, in a given block is selected for this block. Thus yielding, for the whole image, a better performance then the one obtained by applying any of these prediction approaches separately.

The idea of switched prediction is not new. Interframe/Intraframe switched prediction is a very effective method [4,5,6,7]. The interframe predictor combined with the spatial predictor can be either simple one [5,7] or a motion-compensated one. If a motion-compensated predictor is used, motion-vectors can be obtained either by a Pel-Recursive-Algorithm (PRA) [6] or by a Block-Matching-Algorithm (BMA) [1,2]. The spatial predictor can be of 1st order or of higher orders [3]. The switching can be done on a pel basis in a backward manner [4,6], i.e. the switching is based on information available at the receiver, therefore preventing the need for transmitting the switch position. Another method is block-by-block switching in a forward manner [5,7], for which the switch position selected must be transmitted. Quantization in such a coding scheme can be either simple [5] or adaptive, in various forms [7].

The coding scheme presented here differs significantly from the schemes mentioned above, mainly because this coding scheme is to work at a very low rate - 1.544 Mbits/sec - as compared with 30 Mbits/sec in [5,7]. The main differences are in the way the predictors are combined, the switching decision law applied, the way the decisions are transmitted, and the quantization method.

The paper is organized as follows : In section II, motion-compensated prediction is discussed. Section III discusses spatial (intraframe) prediction. In section IV we present the overall coding scheme. Simulation results are reported in section V.

II. MOTION-COMPENSATED PREDICTION

The simplest interframe prediction that can be utilized in the coding scheme is the one in which the previous reconstructed frame is taken as a prediction of the current frame. The prediction equation is therefore:

$$\tilde{I}_K(i,j) = \hat{I}_{K-1}(i,j) \qquad (1)$$

where $\tilde{I}_K$ denotes the predicted current (K^{th}) frame and $\hat{I}_{K-1}$ denotes the previous reconstructed frame.

The reconstructed frame is used here to ensure that the prediction image at the receiver will be the same as the one at the transmitter . The drawback of such a predictor is that the prediction error in moving areas can be very large. This error can be significantly reduced by using motion-compensated prediction [1,2].

The idea of motion-compensated prediction in video-conferencing is based on a model which assumes that each pixel from the current frame originates from some other pixel in the previous frame by translatory motion.

Based on this model, a prediction image can be obtained by estimating the motion-vectors between the previous and current frames, and then moving elements from the previous frame according to the estimated vectors. The prediction equation in this case is:

$$\tilde{I}_K(i,j) = \hat{I}_{K-1}(i-D_x, j-D_y) \qquad (2)$$

where (D_x,D_y) is the motion-vector of the element (i,j) in the current frame.

The estimation of the motion-vectors is done by a Block-Matching [1,2] approach (BMA). By this approach, the current image is divided to sub-blocks of size M×N with $BI_K^{i_0,j_0}$ denoting the $(i_0 j_0)$ block in the K^{th} image. As recommended in [2], a block size 8×8 is used. All the pixels within any given block are assumed to have the same motion-vector. Therefore, a single motion-vector $[D_x,D_y]_{i_0,j_0}$, corresponds to each block $BI_K^{i_0,j_0}$. The motion-vector for each block is found by a matching process, i.e. finding the block in the previous reconstructed frame which best matches it. The degree of matching can be measured by cross-correlation, but for complexity reduction it is usually done by minimizing an error function of the form:

$$E(k,l) = \sum_{m=1}^{M} \sum_{n=1}^{N} f\left[BI_K^{i_0,j_0}(m,n) - \hat{I}_{K-1}(Mi_0+m-k, Nj_0+n-l) \right] \qquad (3)$$

and the (k,l) values which minimize the error function are the coordinates of the motion vector $[D_x,D_y]_{i_0,j_0}$.

Commonly used error functions are :

1. For a Mean Square Error (MSE) criterion : $f(x)=x^2$.
2. For a Mean Absolute Difference (MAD) criterion : $f(x)=|x|$.

where the MAD option is less complex but gives slightly lower performance. In our simulations the MAD option is the one used.

To reduce complexity, the search for the best matching position is done in a limited search area. We assume that each motion-vector coordinates are limited in the range [−P,+P]. This gives $(2P+1)^2$ possible motion-vector element values. A typically used value for P is P=6 [2].

Calculating the error function for each of the 169 possible motion-vectors (if P=6 is used) is still too much for real-time implementation. Several fast motion-vector estimation algorithms where suggested in the literature [1,2]. In the following simulations we decided to use the one called the "Conjugate-Direction Algorithm".

Since the predicted image at the receiver must be the same as the one at the transmitter, the motion-vectors are transmitted too. Straightforward binary code requires 8×4096=32768 bits (since there are 4096 motion-vectors and 169 possible motion-vector values). A very effective coding method in this particular case is the Huffman errorless variable-length code. The efficiency in using this code is high, since the motion-vectors probability density function is highly non-uniform - thus, yielding low entropy (2-3 bits/vector).

The structure of the code is also to be transmitted since the decoding of the received bit sequence depends on this structure. This is done by transmitting the 169 code-words lengths using 5 bits per one code-word (it can be shown that the maximum code-word length in this case is 18 bits). In our simulations, the number of bits required to transmit the Huffman coded motion-vectors, N_1, was found to be in the range of 8000-12000 bits.

III. SPATIAL PREDICTION

The other prediction type used in our scheme is spatial prediction. The prediction of the gray-level of a certain pixel is based on the gray-level values of neighboring pixels in the current frame which were already coded (i.e. causal prediction) [3].

The prediction equations in this case (for 1st and 2nd predictor orders) are :

$$\tilde{I}_K(i,j) = \begin{cases} c_1+w_{11}\hat{I}_K(i,j-1) & 1^{st}\text{ order} \\ c_2+w_{12}\hat{I}_K(i,j-1)+w_{22}\hat{I}_K(i-1,j) & 2^{nd}\text{ order} \end{cases} \qquad (4)$$

$$\text{for all } (i,j)\in A$$

where A is the region in which the prediction is to be computed, $\tilde{I}_K(i,j)$ denotes the spatially predicted (i,j) element (for convenience, we chose the same notation here as for interframe prediction); $\hat{I}_K(i,j)$ denotes the reconstructed (i,j) element in the current frame, and c_r, w_{qr} are constants and weighting coefficients, respectively, estimated separately to minimize the spatial prediction error globally (over large regions) or locally (over small regions).

It can be shown easily that if the average of I_K in the corresponding region A is zero, c_r takes then the value zero too. Computing the remaining parameters is done by a set of equations called "the autocorrelation equations" [3].

The steps taken to perform the spatial prediction in our scheme are therefore as follows :

1. The frame plane is divided into sub-regions $\{A_g\}_{g=1}^{G}$. The region size used here is 8x8 pixels.
2. The average M_{A_g} for each region A_g is approximately given by :

$$M_{A_g} \approx \frac{1}{|A_g|}\sum\sum_{(i,j)\in A_g} I_K(i,j) \qquad (5)$$

3. The autocorrelation values, which appear in the autocorrelation equations, are estimated using the unbiased image (the source images after subtracting the averages). This is done only *once* using a training sequence.

Given the autocorrelations estimations, the weighting factors are evaluated using autocorrelation equations. In Table 1 the weighting factors for 1st-4th order predictors that we obtained are given.

Table 1 : Weighting factors for spatial predictors of different orders.

	1st order	2nd order	3rd order	4th order
	w_{i1}	w_{i2}	w_{i3}	w_{i4}
i=1	0.9900	0.4370	0.8853	0.9579
i=2		0.5620	0.9106	0.6312
i=3			-0.7977	-0.8758
i=4				0.2882

4. Based on equation (4), the prediction is constructed using the following equation (for a 2nd order predictor) :

$$(i,j)\in A_g \rightarrow \tilde{I}_K(i,j) = M_{A_g}+w_{12}\left\{\hat{I}_K(i,j-1)-M_{A_g}\right\}+w_{22}\left\{\hat{I}_K(i-1,j)-M_{A_g}\right\} \qquad (6)$$

IV. OVERALL CODING SCHEME

The overall coding scheme is shown in figure 1. This scheme includes only the main inputs and outputs for each building-block.

In this scheme two prediction images are computed at the transmitter :

1. Based on the previous reconstructed frame, $\hat{I}_{K-1}$, a motion-compensated prediction image, $\tilde{I}_{K_1}$, is generated by the building-block P_1.
2. Based on already quantized elements from the current frame, a spatial prediction image, $\tilde{I}_{K_2}$, is generated by the building-block P_2.

A switching law is then activated on a block by block basis (by building-block SW) to select one of the two prediction images. The average quantization error (the error between the quantizer input and its output) for each block is computed and the predictor which gives a lower error is selected by the switch for that block.

In principle, the resulting prediction image (after switching), denoted by $\tilde{I}_K$, is substructed from the current original image, I_K, and the prediction error image ($E_K=I_K-\tilde{I}_K$) is then quantized by building-block Q. However, since the quantization has been performed for each block as part of the switching unit processing, it needs not to be performed again.

To obtain the same switched prediction image at the receiver, the switching decisions are also transmitted as described in fig. 1. The received information is used to determine which of the two predictors to activate.

Following quantization, the prediction image, $\tilde{I}_K$, is added to the quantizer output, EQ_K, to produce the current received image (assuming an error free channel). This image will be used to build the motion-compensated prediction for the next stage.

A detailed description of the building-blocks now follows :

1. Building-block P_1 is described in figure 2. This unit generates the motion-compensated prediction image. First, the motion-vectors are estimated in the block ME (Motion Estimation), using the current frame, I_K, and the previous reconstructed frame - $\hat{I}_{K-1}$. These motion-vectors are coded and transmitted using N_1 bits.
   Then, the motion-compensated prediction image, $\tilde{I}_{K_1}$, is computed in the block denoted by MCP (Motion-Compensated Prediction), using $\hat{I}_{K-1}$ and the estimated motion vectors.
2. Building-block P_2 is described in figure 3. This unit generates the spatial prediction image. First, in block AVE (AVErage) the averages of sub-blocks (8×8 pixels) in the current frame are estimated using equation (5). To avoid the need for coding and transmitting these averages, they are estimated using the motion-compensated prediction image $\tilde{I}_{K_1}$, assuming that the statistical characterization of the two images is sufficiently close.
   The averages and the casual elements from the current

reconstructed frame, $\hat{I}_K$, are then used to compute the spatial prediction which is done in block **SP** (Spatial Prediction).

3. The quantization unit - **Q** - is described in figure 4. The quantization utilized here is based on a model which assumes that the prediction error elements in E_K are Laplace distributed with zero mean and with a space-variant variance.

   Based on this model, two adaptation approaches are utilized here. The first is dynamic bit allocation to the blocks of the prediction-error image, and the second is normalization of the data in each block before quantization.

   Since spatial prediction follows quantization (so that the prediction of the current pixel is based on the neighboring *reconstructed* pixels), and since no quantization is possible before determining the bit allocation, the bit allocation is based in our scheme only on the motion-compensated prediction error.

   Therefore, the motion-compensated prediction error image is first found ($I_K - \bar{I}_{K_1}$). This image is then divided into blocks (8×8 pixels), and the variance of each error block is estimated in the sub-unit **VAR$_1$** (**VAR**iance). These variances are logarithmically quantized (4 bits), Huffman-coded, and transmitted using N_2 bits. N_2 was found to be in the range of 4500-6000 bits.

   The variances are estimated using the expression :

   $$\hat{\sigma}_{[i_0 j_0]} = \sqrt{2} \sum_{i=8i_0}^{8i_0+7} \sum_{j=8j_0}^{8j_0+7} | I_K(i,j) - \bar{I}_{K_1} | \qquad (7)$$

   where $\acute{\sigma}_{[i_0 j_0]}$ is the estimated variance for the 8×8 block $[i_0 j_0]$.

   The estimated variances are then used for determining the bit allocation which is done in the block **ALL** (**ALL**ocation). The bit allocation is determined such that, for a given total number of bits, the total mean square error between E_K (the quantizer input) and EQ_K (the quantizer output) is minimized. An iterative algorithm for computing the bit allocation is used.

   Given the estimated variance and the corresponding bit assignment for each block, the corresponding prediction-error image block in E_K is quantized. For the motion-compensated prediction-error image, first, each block is normalized by its estimated coded variance obtained from **VAR$_1$**, then, quantizing the normalized values using a Max-Loyd optimal quantizer for $L[m=0, \sigma=1]$, and finally, denormalizing the quantizer output, using the same estimated coded variance. These operations are done in the block denoted by **NQ** (Normalized Quantization).

   For the spatial prediction error, the quantization steps are the same with one difference. the variance of the prediction error is computed recursively by :

   $$\hat{\sigma}(n) = \alpha\hat{\sigma}(n-1) + (1-\alpha)\sqrt{2}\,| \bar{I}_{K_1}(i,j) - \hat{I}_K(i,j) | \qquad (8)$$

   where $\hat{\sigma}(n)$ is the the new estimated variance, based on its previous value and the last reconstructed pixel (so that the same operation can be done at the receiver). This operation is done in the block **VAR$_2$**.

   This recursive estimation is initialized by $\hat{\sigma}(0)$ chosen to be the estimated variance obtained from the motion-compensated prediction error for the corresponding block.

   The parameter α is a decay-factor, and is usually in the range of [0.9,1). In our simulations we used α=0.95.

4. The last building-block is the one which controls the switching between the two predictions. This unit is denoted by **SW**.

   The switching is done by computing for each block the performance of the system with each of the two predictors and choosing the predictor which gives lower mean square error. For each of the predictors, the system's performance for the $[i_0 j_0]$ block is evaluated by :

   $$MSE[i_0, j_0]_p = \sum_{i=8i_0}^{8i_0+7} \sum_{j=8j_0}^{8j_0+7} \left[E_{K_p}(i,j) - EQ_{K_p}(i,j) \right]^2 \quad p=1,2 \qquad (9)$$

   where p denotes the prediction type (p=1 for motion-compensated prediction and p=2 for spatial prediction) and E_{K_p}, EQ_{K_p} are the prediction-error and the quantized prediction-error, respectively, for prediction type p.

   This unit also controls the computations required to obtain the two predictions for the corresponding block and the quantization of the prediction errors (as described in detail in 3 above). Note that the quantization is done twice - once for the motion-compensated prediction error and once for the spatial prediction error.

   Since the decision is based on the current source image, the switch position selected must be transmitted too. There are 4096 decisions (one per block) to transmit, and they are coded, again, by a Huffman code. N_3, the amount of bits required, was found to be in the range of 1500-2500 bits.

V. SIMULATIONS RESULTS

Simulations done with 50 frames of a typical video-conference image-sequence are presented in this section. The simulations were carried out on a VAX-750 computer hosting an image display system (GOULD IP8500) with a Real-Time-Digital-Disk system. The images are of size 512×512 pixels, were each pixel is represented by 8 bits. The sequence contains a fair amount of head motion (up to six pixels/frame). In the simulations two major tests were carried out :

1. Steady state response - assuming that the first frame is known also at the receiver and the coding begins at the second frame, using the first frame as a basis to the motion-compensated prediction image construction.
2. Transient response - referring to the case in which the coding begins right from the first frame, thus yielding an initialization problem since there is no previous frame to be used as a basis for constructing the motion-compensated prediction image. To solve this problem, the average value of each block in the first frame is estimated, coded and transmitted. These averages are used to construct the 0^{th} frame in which each block element is set to the corresponding coded average value. This method was found to almost completely eliminate transient effect. In this work the averages were estimated for blocks of 8×8 pixels (4096 blocks), coded by a 4-bit uniform quantizer and then by a Huffman code. There is, of course, a need to substruct the amount of bits required for transmitting the coded averages from the total number of bits available for transmitting the first image (61760 bits).

The tests were carried out for two coding schemes. The first scheme is the one described in this paper (i.e. using switched prediction) with a 1^{st} order spatial predictor. The second scheme is a simpler one in which there is only a motion-compensated predictor. The quantization in both schemes is as described earlier.

In fig. 5, the MSE vs. the frame number for the steady-state response is presented for the two schemes. As shown in this graph, significant improvement is obtained when switched-prediction is used.

Fig. 6 refers to the transient-response. As shown in this graph, the method mentioned above reduces considerably the transient response and again, the switching prediction scheme is significantly better then the other simpler scheme.

VI. SUMMARY AND CONCLUSIONS

In this work we presented a new scheme for image-sequence coding, at 1.544 Mbits/sec, which is based on Interframe/Intraframe DPCM with switched-prediction. The predictors used are a motion-compensated predictor (based on the block-matching technique for estimating the motion-vectors) and a spatial linear predictor. This scheme was compared to a simpler scheme which utilizes motion-compensated predictor only and was shown to perform significantly better, both at steady-state and at transients.

REFERENCES

[1] H.G. Musmann, P. Pirch & H.J. Grallert : "Advances in Picture Coding.", Proc. IEEE, vol. 73, pp. 523-548, April 1985.

[2] A. Puri, H.M. Hang & D.L. Schilling : "An Efficient Block-Matching Algorithm for Motion-Compensated Coding." ICASSP, pp. 1063-1066, January 1987.

[3] W.K Pratt : "Digital Image Coding.", Wiley-Interscience, N.Y., 1978.

[4] K.A. Prabhu : "A Predictor Switching Scheme for DPCM Coding of Video Signals.", IEEE Trans. Comm. , vol. 33, pp. 373-379, April 1985.

[5] H. Yamamoto, Y. Hatori & H. Murakami : "30 Mbit/s Coder for the NTSC Color TV Signal Using an Interfield-Intrafield Adaptive Prediction." IEEE Trans. Comm. , vol. 29, pp. 1859-1867, December 1981.

[6] K.A. Prabhu & N. netravali : "Motion Compensated Component Color Coding.", IEEE Trans. Comm. , vol. 30, pp. 2519-2527, December 1987.

[7] B. Girod, H. Almer, L. Bengtsson, B. Christensson & P. Weiss : "A Subjective Evaluation of Noise-Shaping Quantization for Adaptive Intra/Interframe DPCM Coding of Color Television Signals.", IEEE Trans. Comm. , vol. 36, March 1988.

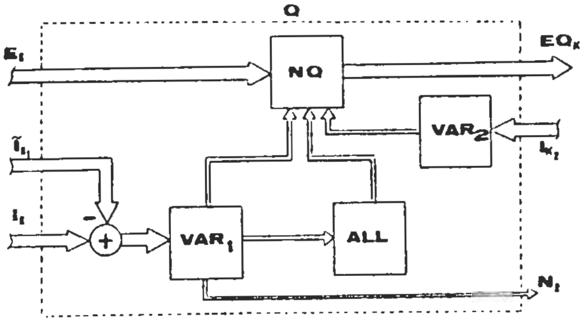

Fig. 4 - Quantizer building-block Q.

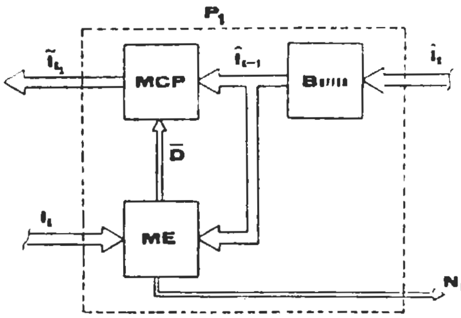

Fig. 2 - Motion-compensated prediction building-block P_1.

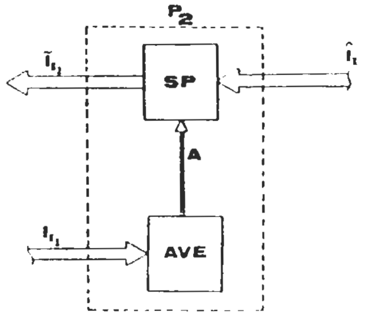

Fig. 3 - Spatial prediction building-block P_2.

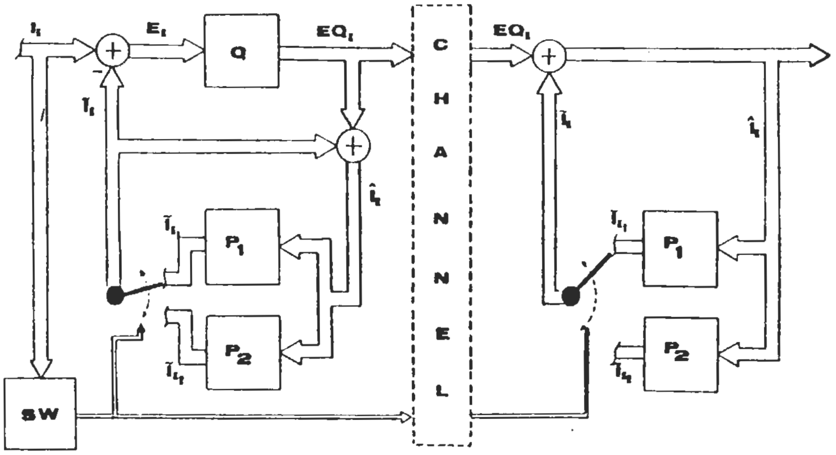

Fig. 1 - The overall coding scheme (transmitter and receiver).

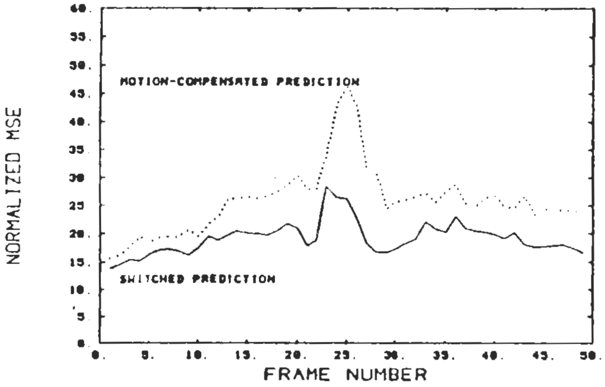

Fig. 5 - MSE vs. frame number for steady-state response.

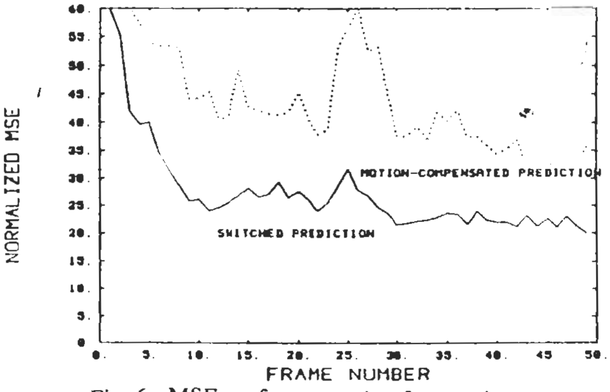

Fig. 6 - MSE vs. frame number for transient response.