

**COMMITTEE T1  
CONTRIBUTION**

Document Number: T1A1.5/95-103

\*\*\*\*\*

**STANDARDS PROJECT:** VTC/VT Performance Standards Project

\*\*\*\*\*

**TITLE:** An In-depth Analysis of the P<sub>6</sub> Lost Motion Energy  
Parameter

\*\*\*\*\*

**ISSUE ADDRESSED:** Objective and subjective testing and correlation  
analysis

\*\*\*\*\*

**SOURCE:** NTIA/ITS - Stephen Wolf

\*\*\*\*\*

**DATE:** January 9, 1995

\*\*\*\*\*

**DISTRIBUTION TO:** T1A1.5

\*\*\*\*\*

**KEYWORDS:** Video Performance Testing, Objective Video Quality,  
Subjective Video Quality

\*\*\*\*\*

## 1. Introduction

This contribution provides an in-depth analysis of the behavior of the  $p_6$  performance parameter presented in contribution T1A1.5/93-152. This parameter can be described in engineering terms as measuring the motion energy that is lost when a video signal is transmitted through a Hypothetical Reference Circuit (HRC). The perceptual impairment that results from lost motion energy is unnatural motion due to freezes in the video that result from frame repetition. When the video is finally updated after a period of frame repetition, a sharp increase in the amount of scene motion results in the perception of jerkiness.

To perform this in-depth analysis, it was useful to examine three different levels of subjective and objective data aggregation. Let  $x_{ijkl}$  represent the subjective data from HRC  $i$ , scene  $j$ , viewer  $k$ , and laboratory  $l$ . Then, the three levels of subjective data aggregation examined in this contribution are denoted  $x_{ij..}$  (the MOS of an HRC x scene combination computed using all three laboratories data),  $x_{i...}$  (the HRC main effect computed using all three laboratories data), and  $x_{.j..}$  (the scene main effect computed using all three laboratories). Here, the dot represents averaging over the variable. The subjective data used in this contribution is described in section 3 of contribution T1A1.5/94-152. The HRC main effect and the scene main effect have been broken out separately since these effects are the main contributors to the MOS of a HRC x scene combination (the other less important contributor is the HRC x scene interaction). By examining the HRC and scene main effects in addition to the behavior of the parameters for certain HRC x scene combinations, one can obtain important insights as to the behavior of the objective measures.

This contribution proposes an additional form for the  $p_6$  parameter that can be computed using the fundamental temporal information (TI) frame by frame features given in prior contribution T1A1.5/93-152. This new parameter extracts additional information by using a new temporal collapsing function on the basic TI features. Recall that a temporal collapsing function specifies the method of calculating performance parameters from the basic time history of the TI features that are measured from each frame of video. These functions are necessary since viewers rate segments of video (in the case of the T1A1.5 tests, the video segment was 9 seconds long) whereas the TI features are measured for each frame of video.

## 2. Observations Regarding Temporal Information (TI) Features

### 2.1 The Non-linear Effects of Reduced Frame Rate

In Figure 10a of T1A1.5/94-152, one can see an area of points where the model consistently predicts lower scores than the actual subjective MOSs for given HRC x scene combinations (the points above the 45 degree dashed line). This area of points causes a noticeable non-linear response of the model when plotted against the subjective data. An examination of the temporal information values of the destination video ( $TI_D$ ) for these HRC x scene combinations has revealed that the majority of these outliers are 10 to 15 frames per second (fps) video. Figure 1 gives a plot of the first 50 frame samples of  $TI_D$

for the scene smity2 and HRC 10 (here, the model predicted a MOS of 2.41 quality units whereas the subjective MOS was actually 3.74 quality units). In the figure, the frames that contain motion have high values of  $TI_D$  whereas the repeated frames have low values of  $TI_D$ .  $TI_S$  for the source video, which gives the motion present in the source video, has been plotted with a solid line in the figure.

Since the  $p_6$  parameter in the model integrates the lost motion energy between the curves  $TI_S$  and  $TI_D$  shown in Figure 1 (see T1A1.5/93-152 and T1A1.5/94-110 for detailed methods of measurement for  $TI_S$  and  $TI_D$ ), 15 fps will produce large values of  $p_6$  relative to 30 fps. However, for video teleconferencing, viewers do not seem to object strongly to 15 fps video if the picture clarity is good (i.e., the spatial distortions are small). This contribution proposes a method to linearize the  $p_6$  parameter with respect to perceived jerkiness in video teleconferencing by performing perceptual prefiltering on the TI values before the lost motion energy is computed. Section 3 of this contribution describes a new parameter,  $p_{14}$ , that uses the 3 point maximum filter described in prior contribution T1A1.5/93-152 as a perceptual prefilter on the TI frame samples. This filter smooths short periods of lost motion by replacing the current TI value with the maximum of itself and its two nearest neighbors (the TI value earlier in time and the TI value later in time). The 3 point maximum prefilter appears to be effective in linearizing the  $p_6$  parameter for the T1A1.5 video teleconferencing data. It is conceivable that different video applications might require different prefilters on the TI feature. For instance, viewers might object more strongly to 15 fps entertainment video than to 15 fps video teleconferencing video.

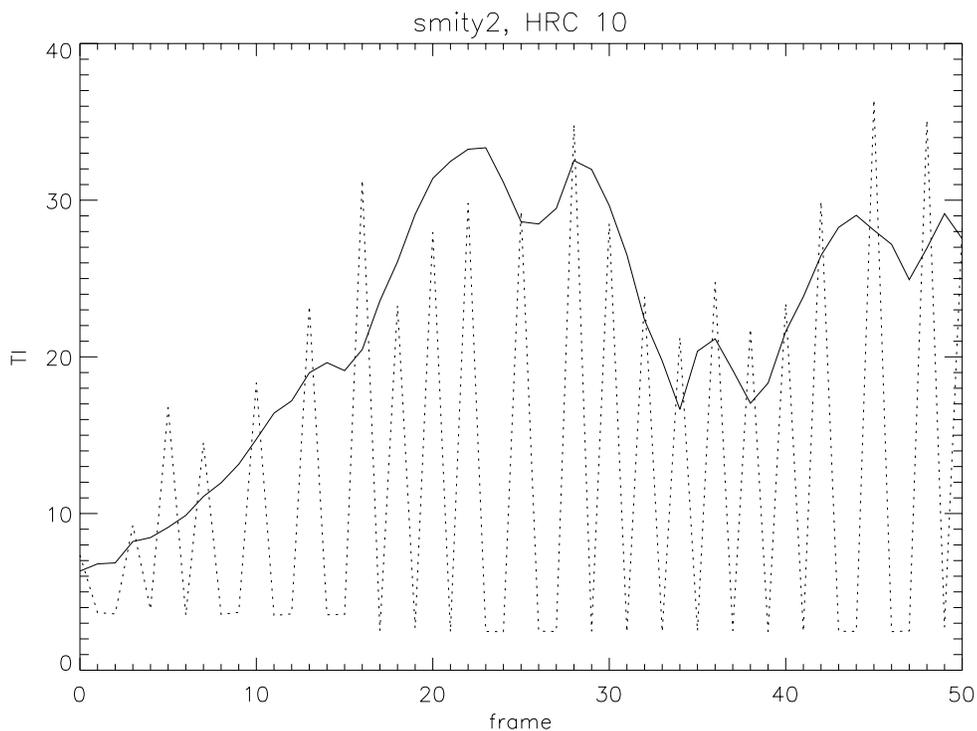


Figure 1  $TI_D$  (dashed line) and  $TI_S$  (solid line) for HRC 10, Scene Smity2

## 2.2 The Effects of a Noisy Source Scene

Comparing the scene main effect of the 3 parameter model with the scene main effect computed from the subjective data (see Figure 12a in T1A1.5/94-152) reveals that scene u (filter) has the largest difference. In this case, the scene main effect of the objective model is about .43 quality units lower than the corresponding subjective rating. The cause of this difference has been determined to be the behavior of the  $p_6$  parameter for test scenes that contain high levels of background noise. Recall that the  $p_6$  parameter uses the difference between the temporal information of the source ( $TI_S$ ) and the temporal information of the destination ( $TI_D$ ) to calculate relative lost motion energy. Figure 2 shows a plot of  $TI_S$  for the scene filter (solid line) and the corresponding  $TI_D$  for HRC 5 (dashed line). HRC 5 is the HRC where the objective model over-penalized the filter scene by the greatest amount (here, the model predicted a MOS of 2.48 quality units whereas the subjective MOS was actually 3.93 quality units). The high level of background noise in the scene filter shows up as motion energy so that even in periods of little hand motion (e.g., frame 110 to 160), the  $TI_S$  value is still quite large. By examining the corresponding  $TI_D$  for HRC 5, one can see that this HRC has eliminated a large portion of the background noise. The  $p_6$  parameter treats this eliminated background noise as lost motion energy and penalizes the HRC for eliminating (not passing) this noise. The viewers may in fact prefer a cleaner video picture where the background noise has been reduced.

The  $p_6$  measure has been shown to be strongly correlated to bit rate (see T1A1.5/94-148). In general, the lower the bit rate, the more background noise the HRC must threshold out to save bits for coding the moving foreground. This is one reason why the draft VTC/VT performance standard (T1A1.5/94-107) has a specification for the allowable noise on the input signal. Since  $p_6$  detects this eliminated background noise, it tends to also be highly correlated with bit rate. Unfortunately, what is true in general is not necessarily true in specific cases in that a high bit rate coder could threshold out more noise than a low bit rate coder.

To compute a lost motion energy parameter that does not penalize or detect lost motion due to eliminated background noise, the background noise level of  $TI_D$  and  $TI_S$  is subtracted out before calculation of the lost motion energy. Parameter  $p_{14}$  in section 3 of this contribution includes one possible method for subtracting this background noise level before computation of the lost motion energy.

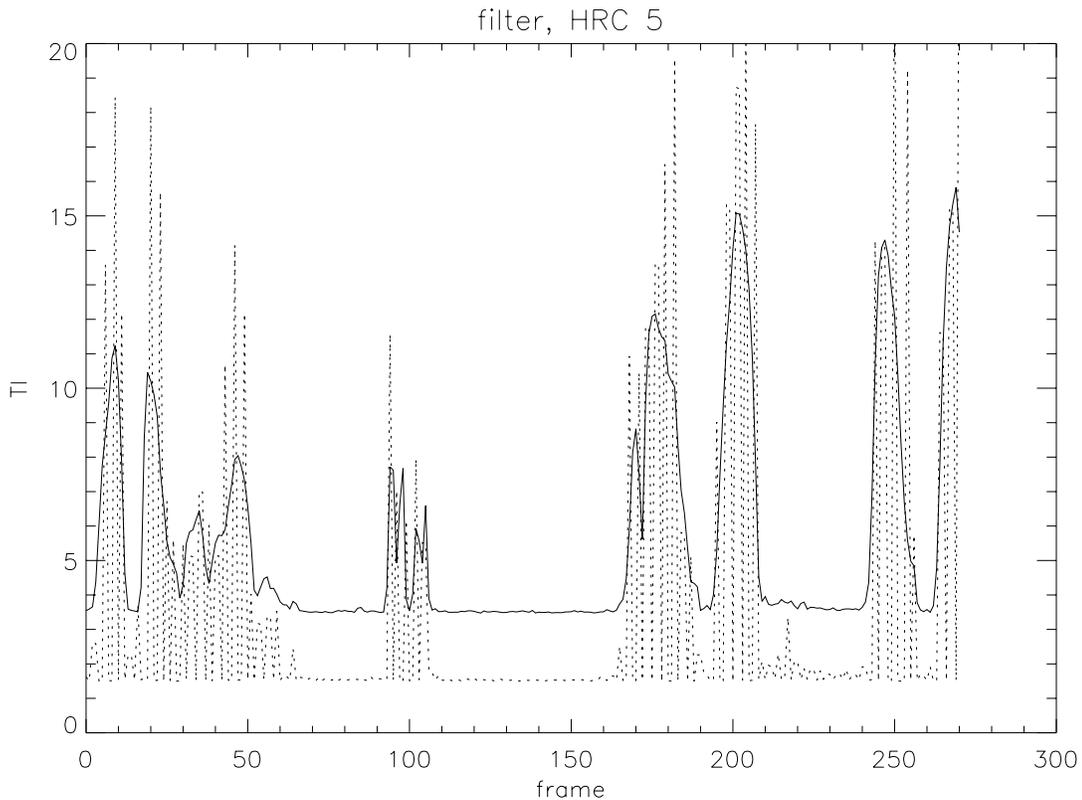


Figure 2  $TI_D$  (dashed line) and  $TI_S$  (solid line) for HRC 5, Scene Filter

### 2.3 Normalizing the TI Waveforms for Unit Variance

Except in cases of very high frame repetition, the  $TI_S$  and  $TI_D$  time waveform envelopes normally have very similar shapes (i.e., the motion in the source and destination video are very similar). This property has been successfully applied to measure the video delay of the HRC x scene combination (Note: the video delay of the HRC is not fixed with respect to scene -- typically scenes with small amounts of motion produce smaller video delays than scenes with larger amounts of motion). Although the shapes of the  $TI_D$  and  $TI_S$  envelopes may be quite similar, they usually differ in their spread (i.e., variance) of motion values. This can be clearly seen from Figure 2. Normalization of the  $TI_S$  and  $TI_D$  waveforms allows for closer tracking of the two motion envelopes before computation of the lost motion energy. One method of performing this normalization is to divide each time sample of the TI waveform by the standard deviation (over time) of the TI waveform. This will produce a TI waveform with unit variance. Caution should be observed for TI waveforms from still scenes. Here, the standard deviation is very small and is due mainly to scene background noise and quantization noise. Section 3 discusses one method of adding a small constant  $k$  to the standard deviation (over time) to lower bound the divisor for still and low motion scenes.

### 3. A Lost Motion Energy Parameter (P<sub>14</sub>)

The above discussions regarding the computation of lost motion energy described in sections 2.1, 2.2, and 2.3 can be summarized as follows:

(1) Square the TI<sub>S</sub> and TI<sub>D</sub> waveforms given in T1A1.5/93-152; these will be denoted TI<sub>S</sub><sup>2</sup> and TI<sub>D</sub><sup>2</sup>. The reason for squaring the TI waveforms is that the minimum over time of the TI<sub>S</sub><sup>2</sup> and TI<sub>D</sub><sup>2</sup> waveforms will be used to estimate the background noise power levels. Under the assumption that the background noise is not correlated with the motion, the TI<sup>2</sup> waveform will be DC shifted by this background noise power level.

(2) Pre-filter the TI<sub>S</sub><sup>2</sup> and TI<sub>D</sub><sup>2</sup> waveforms; these will be denoted Fil\_TI<sub>S</sub><sup>2</sup> and Fil\_TI<sub>D</sub><sup>2</sup>. For the purposes of this contribution, the filtering function will be one pass of the 3 point maximum filter described in T1A1.5/93-152 and section 2.1 of this contribution. This filter replaces the TI<sup>2</sup> value with the maximum of itself and its two nearest neighbors (the TI<sup>2</sup> value earlier in time and the TI<sup>2</sup> value later in time).

(3) Normalize the Fil\_TI<sub>S</sub><sup>2</sup> and Fil\_TI<sub>D</sub><sup>2</sup> time samples; these will be denoted by Norm\_Fil\_TI<sub>S</sub><sup>2</sup> and Norm\_Fil\_TI<sub>D</sub><sup>2</sup>. As mentioned in sections 2.2 and 2.3, this normalization subtracts the background noise power level and allows for closer tracking of the two motion energy envelopes. This step is represented in equation form as

$$\text{Norm\_Fil\_TI}_S^2(t_n) = \frac{[\text{Fil\_TI}_S^2(t_n) - \min_{time}(\text{Fil\_TI}_S^2)]}{[\text{std}_{time}(\text{Fil\_TI}_S^2) + k_1]} \quad (1)$$

$$\text{Norm\_Fil\_TI}_D^2(t_n + d_v) = \frac{[\text{Fil\_TI}_D^2(t_n + d_v) - \min_{time}(\text{Fil\_TI}_D^2)]}{[\text{std}_{time}(\text{Fil\_TI}_D^2) + k_1]} \quad (2)$$

where k<sub>1</sub> is a small constant to lower bound the divisor for still scenes (a value of k<sub>1</sub>=0.5 was used for this contribution), t<sub>n</sub> denotes time sample n, and d<sub>v</sub> denotes the video delay (see contribution T1A1.5/93-152). The functions std<sub>time</sub> and min<sub>time</sub> are the standard deviation and the minimum functions, respectively. Figure 3 gives a plot of the resulting TI<sub>D</sub><sup>2</sup> and TI<sub>S</sub><sup>2</sup> waveforms of Figure 2 after squaring (step 1), prefiltering (step 2) and normalization (step 3).

Note: The min<sub>time</sub> function provides a good estimate of the background noise power level only if the test scene contains periods of little or no motion.

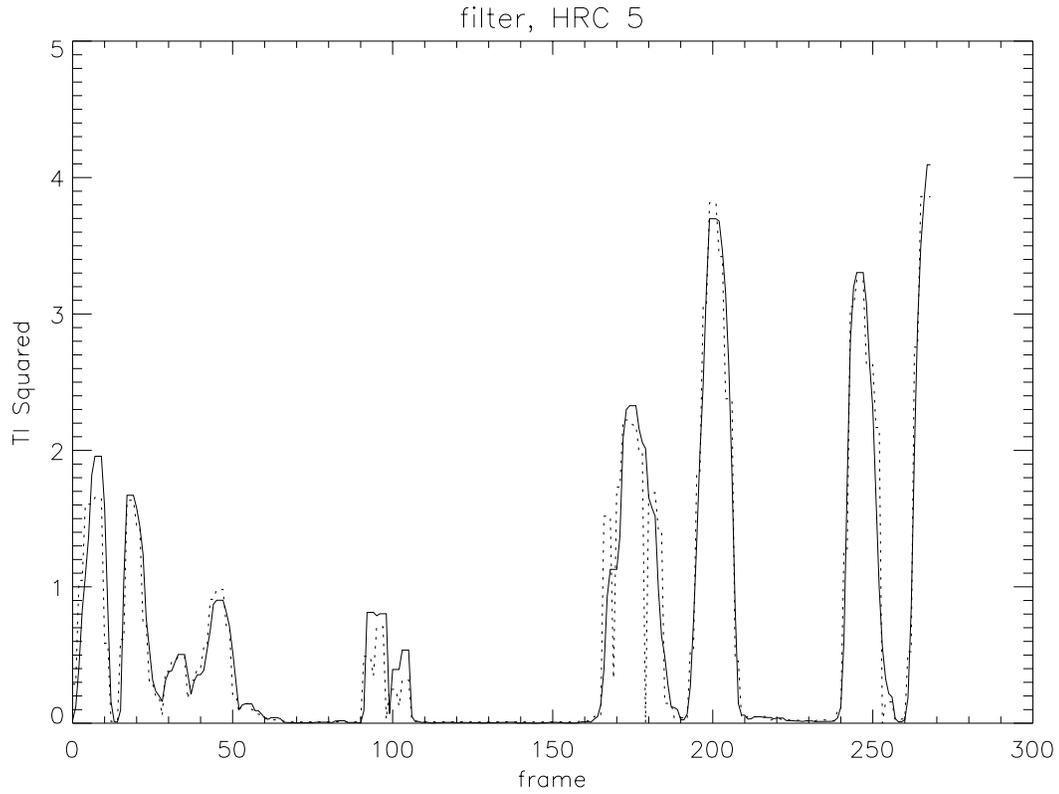


Figure 3 The  $TI_D^2$  (dashed line) and  $TI_S^2$  (solid line) Waveforms of Figure 2 after Squaring (step 1), Prefiltering (step 2), and Normalization (step 3)

(4) The lost motion energy parameter ( $p_{14}$ ) is computed in a nearly identical manner to the original  $p_6$  parameter given in contribution T1A1.5/93-152 except that the filtered, and normalized  $TI^2$  values from equations (1) and (2) are used, i.e.

$$p_{14} = rms_{time} \left[ \max \left( \frac{[\text{Norm\_Fil\_} TI_S^2(t_n) - \text{Norm\_Fil\_} TI_D^2(t_n + d_v)]}{[\text{Norm\_Fil\_} TI_S^2(t_n) + k_2]}, 0 \right) \right] \quad (3)$$

where  $k_2$  is a small constant to lower bound the divisor in areas of low motion (a value of  $k_2=0.5$  was used for this contribution). Recall that the  $\max(., 0)$  function in equation (3) limits the rms summation to include only areas where motion energy has been lost (i.e.,  $TI_S^2$  is greater than  $TI_D^2$ ).

#### 4. A Two Parameter Model ( $P_{14}$ and $P_7$ )

With the  $p_{14}$  parameter, we find that the “best” two parameter predictor of subjective score is a  $p_{14}$ ,  $p_7$  model where  $p_7$  is described in prior contribution T1A1.5/93-152. This model is given by

$$\hat{s} = 4.932 - 3.564 \cdot p_{14} - 2.620 \cdot p_7 \Big|_1^5 \quad (4)$$

where  $\Big|_1^5$  means that the output of the linear model has been clipped at 1 and 5. Figure 4 gives a plot of the subjective MOS vs. the objective predictions for this model for the 625 HRC x scene combinations in the T1A1.5 data set. The statistics for the plot of Figure 4 are:  $\rho = 0.882$ ,  $\rho^2 = 0.778$ , RMSE = .513, Maximum Difference (subjective-model prediction, positive or negative) = -1.6, and the Number of Differences  $> 1 = 33$ . This two parameter  $p_{14}$ ,  $p_7$  model explains an additional 6.5% of the variance in the subjective data compared to the *non-linearized* 3 parameter model presented in T1A1.5/94-152 ( $\rho^2 = 0.714$  for the *non-linearized* 3 parameter model shown in Figure 10a of T1A1.5/94-152). The two parameter  $p_{14}$ ,  $p_7$  model produces slightly better performance than the *linearized* 3 parameter model presented in T1A1.5/94-152 ( $\rho^2 = 0.771$  for the *linearized* 3 parameter model shown in Figure 10b of T1A1.5/94-152).

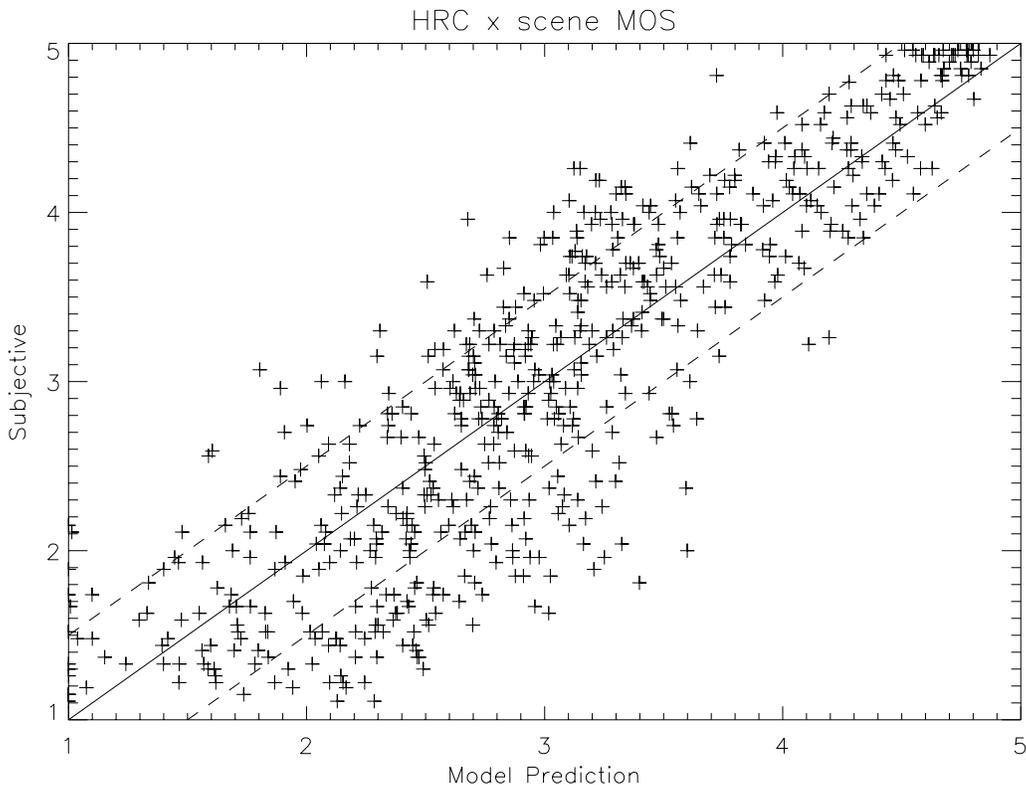


Figure 4 Subjective vs. Objective for  $p_{14}$ ,  $p_7$  Model

The HRC main effect ( $x_{i...}$ ) for the  $p_{14}$ ,  $p_7$  model is given in Figure 5 for the 25 HRCs.

The statistics for this plot are:  $\rho = 0.950$ ,  $\rho^2 = 0.903$ ,  $RMSE = .317$ , and Maximum Difference (subjective-model prediction, positive or negative) =  $-0.72$ . This two parameter  $p_{14}$ ,  $p_7$  model explains an additional 10.9% of the variance in the HRC main effect compared to the *non-linearized* 3 parameter model presented in T1A1.5/94-152 ( $\rho^2 = 0.794$  for the *non-linearized* 3 parameter model shown in Figure 11a of T1A1.5/94-152). The two parameter  $p_{14}$ ,  $p_7$  model produces slightly better performance than the *linearized* 3 parameter model presented in T1A1.5/94-152 ( $\rho^2 = 0.896$  for the *linearized* 3 parameter model shown in Figure 11b of T1A1.5/94-152).

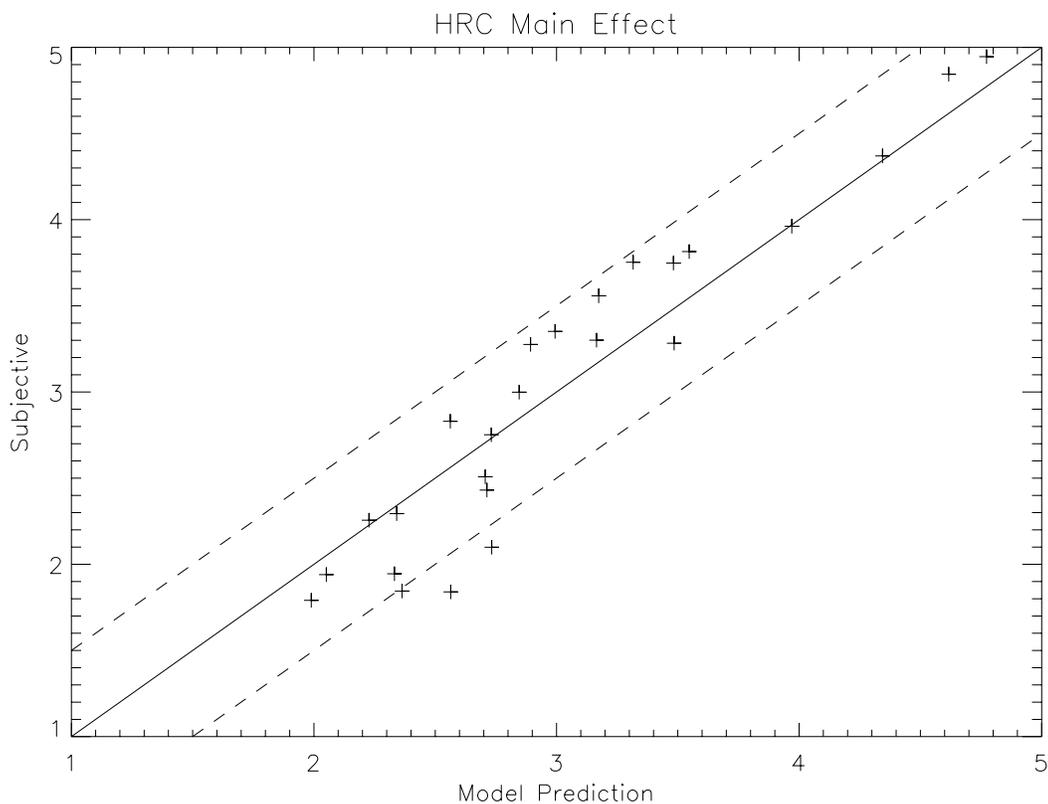


Figure 5 HRC Main Effect for  $p_{14}$ ,  $p_7$  Model

The scene main effect ( $x_{.j.}$ ) for the  $p_{14}$ ,  $p_7$  model is given in Figure 6 for the 25 scenes. The statistics for this plot are:  $\rho = 0.943$ ,  $\rho^2 = 0.889$ ,  $RMSE = .153$ , and Maximum Difference (subjective-model prediction, positive or negative) = 0.34. This two parameter  $p_{14}$ ,  $p_7$  model explains an additional 9.9% of the variance in the scene main effect compared to the 3 parameter model presented in T1A1.5/94-152 ( $\rho^2 = 0.790$  for the 3 parameter model shown in Figure 12a and 12b of T1A1.5/94-152).

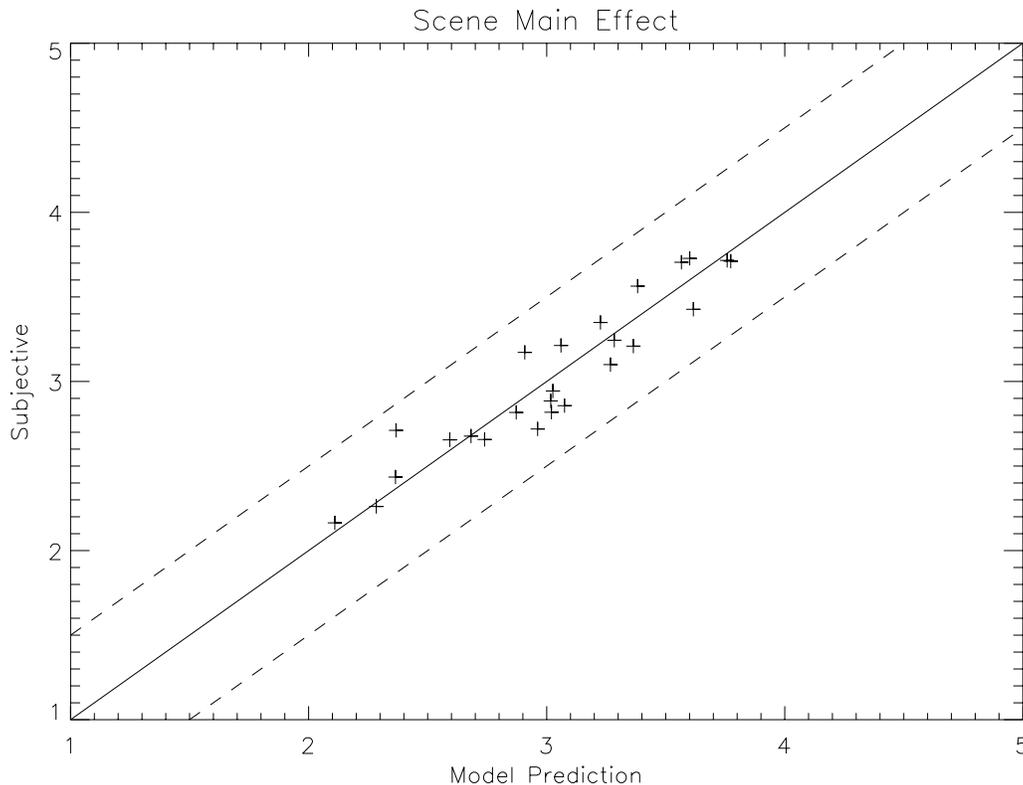


Figure 6 Scene Main Effect for  $p_{14}$ ,  $p_7$  Model

## 5. Conclusion

This contribution proposes an additional parameter,  $p_{14}$ , for calculating the perceptual effects of lost motion energy for noisy source scenes and video teleconferencing users. The  $p_{14}$  parameter uses the same temporal information (TI) frame samples as given for parameter  $p_6$  in prior contribution T1A1.5/93-152 but uses a different time collapsing function on these measured TI frame samples. The  $p_{14}$  parameter explains an additional 6% to 11% of the variance in the subjective data and produces a more linear model response for the T1A1.5 subjective test data.