A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution Supplementary material

GYORGY DENES, University of Cambridge
ALIAKSEI MIKHAILIUK, University of Cambridge
AKSHAY JINDAL, University of Cambridge
RAFAŁ K. MANTIUK, University of Cambridge

This document contains supplementary material for the main manuscript.

ACM Reference Format:
Gyorgy Denes, Aliaksei Mikhailiuk, Akshay Jindal, and Rafał K. Mantiuk. 2020. A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution Supplementary material. *ACM Trans. Graph.* 39, 4, Article 133 (July 2020), 5 pages. https://doi.org/10.1145/3386569.3392411

S.1 ENERGY MODELS AND JNDS

In the main paper we argue that the difference in perceived quality can be expressed as the difference in CSF-normalized energy of two stimuli (Sections 5.3 and 5.4). In this section we demonstrate how this approach can be explained by detection models.

In Equation 13 in the main paper we normalized amplitudes of the Fourier coefficients $m(\omega;\sigma)$ by the CSF to account for the visibility thresholds:

$$\tilde{m}(\omega,\sigma) = \mathrm{CSF}(\omega)\, m(\omega;\sigma)\,, \tag{1}$$

It should be noted that such a normalization applies mostly to low amplitudes because of contrast constancy [Georgeson and Sullivan 1975]. Probability of detecting a single frequency is typically modeled with a psychometric function:

$$\mathbb{P}_{\mathrm{det}}(\omega) = 1 - \exp\left(\ln(0.5)\left(\frac{\tilde{m}(\omega;\sigma)}{\tilde{m}_{t,b}}\right)^{\beta_b}\right). \tag{2}$$

The function is constructed in such a way that $\mathbb{P}_{\mathrm{det}}(\omega) = 0.5$ when $\tilde{m}(\omega;\sigma) = \tilde{m}_{t,b}$. The probability of detecting a complex stimulus consisting of multiple frequencies is modeled with a probability summation [Daly 1992]:

$$\mathbb{P}_{\mathrm{det}} = 1 - \prod_i (1 - \mathbb{P}_{\mathrm{det}}(\omega_i)). \tag{3}$$

Authors' addresses: Gyorgy Denes, Department of Computer Science and Technology, University of Cambridge, gyorgy.denes@cl.cam.ac.uk; Aliaksei Mikhailiuk, Department of Computer Science and Technology, University of Cambridge, am2442@cl.cam.ac.uk; Akshay Jindal, Department of Computer Science and Technology, University of Cambridge, aj577@cl.cam.ac.uk; Rafał K. Mantiuk, Department of Computer Science and Technology, University of Cambridge, rafal.mantiuk@cl.cam.ac.uk.

0730-0301/2020/7-ART133 $15.00
https://doi.org/10.1145/3386569.3392411

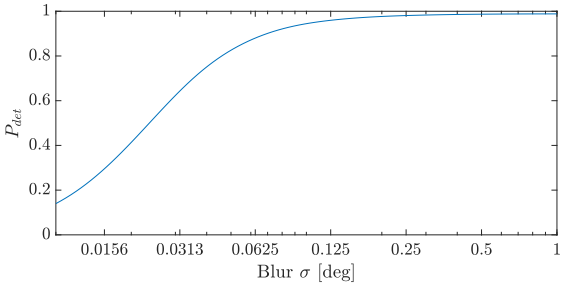

Fig. 1. Probability of detection as a function of blur amount σ.

However, when we combine both equations, we get:

$$\begin{aligned}\mathbb{P}_{\mathrm{det}} &= 1 - \prod_i \exp\left(\ln(0.5)\left(\frac{\tilde{m}(\omega_i;\sigma)}{\tilde{m}_{t,b}}\right)^{\beta_b}\right)\\ &= 1 - \exp\left(\ln(0.5)\sum_i\left(\frac{\tilde{m}(\omega_i;\sigma)}{\tilde{m}_{t,b}}\right)^{\beta_b}\right)\\ &= 1 - \exp\left(\ln(0.5)E_b(\sigma)\right)\,,\end{aligned} \tag{4}$$

where $E_b(\sigma)$ is the energy from Equation 13 in the main paper. The probability of detection is plotted as a function of blur amount σ in Figure 1.

$\mathbb{P}_{\mathrm{det}}$ represents a theoretical probability of the detection mechanism, and it is not the probability of selecting A over B in a pairwise comparison experiment. This is because an observer makes a random guess when the differences between two stimuli are invisible. To account for the random guess, we model the probability of selecting blur σ_A as visible with respect to almost sharp stimulus $\sigma_B{\approx}0$ as:

$$P(Q(\sigma_A) < Q(0)) = 0.5{+}0.5\mathbb{P}_{\mathrm{det}} = 1{-}0.5\exp\left(\ln(0.5)E_b(\sigma)\right)\,. \tag{5}$$

Given such a probability of selecting one condition over another, we can convert it to JND units using a cumulative normal distribution, as explain in Section 2.6 of the main paper. Those steps, however, are unnecessary as the inverse of the cumulative normal distribution is very similar to the psychometric function from the equation above. In fact, when we plot JNDs obtained that way, these are very close to the original energy function $E_b(\sigma)$ (subject to a scaling factor), as shown in Figure 2. Therefore, we can express the quality given a blur factor σ as the linear function of energy $E_b(\sigma)$.

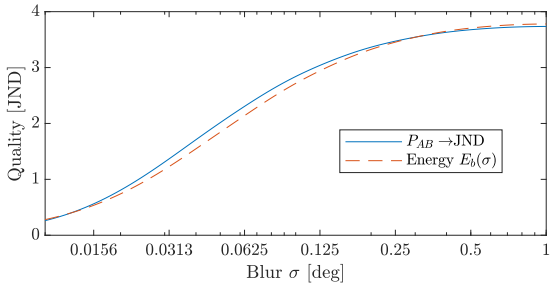

Fig. 2. The JND of blur can be found both by converting the probabilities of selecting A over B into JNDs; or directly using the energy of the blur $E_b(\sigma)$.

Our energy formula can be also seen as a special case of the Stevens psychometric scaling function, which assumes that a power function can map physical quantity (blur energy) into perceived magnitude of a stimulus.

S.2 FURTHER EYE TRACKING RESULTS

The eye tracks moving objects with smooth pursuit eye motion (SPEM; see Section 2.3). Such tracking SPEM is imperfect; the differences between object and gaze motion results in retinal blur (b_E). For our visual model, it is imperative that we can estimate the amount of this blur.

In Section 6.1, we assumed that eye blur is mainly influenced by object velocity and the type of motion; however, refresh rate does not have a significant impact. Figure 14 in the main manuscript reveals a linear relationship between velocity and eye blur. In this section, we present evidence that refresh rate does not significantly impact the quality of SPEM tracking.

We analyzed the recorded traces with metrics from [Suh et al. 2006], excluding the first five seconds of each trial and time periods of blinks. Velocities were obtained by discrete differentiation; saccades were then filtered out using a threshold method when either eye velocity exceeded 40 deg/s or acceleration exceeded 9000 deg/s^2. We verified that results of the threshold method corresponded to manual labeling. Table 1 shows the qualitative results of SPEM tracking averaged across trials: (1) number of saccades (2) eye position error defined as the average difference between target and gaze location measured in visual degrees, (3) eye position variability defined as the standard deviation of target-gaze difference measured in visual degrees, (4) eye gain defined as the ratio of target and gaze velocity, and (5) delay between gaze and target object, identified as the delay that gives the highest cross-correlation score of the target and gaze velocities. Note, that this definition of velocity gain differs from that of Daly et al. [Daly 1998], and as such, it is not comparable to the frequently-quoted gain value of 0.82.

There is a significant difference in all metrics between predictable and unpredictable motion; however, no such difference can be observed between different refresh rates.

S.3 FURTHER MARRR VALIDATION

We repeat our Experiment 5 described in Section 7.5 of the main manuscript with a few more refresh rate/resolution pairs including the one predicted by Debattista et al.'s model (75 Hz) [Debattista et al. 2018]. 6 people (aged 20-40), all with normal/corrected-to-normal vision, participated in the experiment. We used the same procedure and stimuli as before, however, the Eyelink II eye-tracker was replaced with a Pupil Core eye-tracker sampling the gaze location at 200 Hz. The results are reported in Figure 3. Similar to previous results, participants show a clear preference for MARRR compared to all constant-resolution-and-constant-refresh-rate rendering conditions.

Table 1. Quality of SPEM tracking. Aggregated eye tracking data for predictable (blue, top 5 rows) and unpredictable (green, bottom 5 rows) with average object velocity of 20 deg/s. Metrics described in text.

(Hz)	No. saccades	pos err.(°)	pos var.(°)	gain	delay (s)
16.5	21.8 ± 6.8	0.70	0.50	0.60 ± 0.10	-0.01 ± 0.02
27.5	19.9 ± 4.0	0.58	0.40	0.63 ± 0.12	0.01 ± 0.01
55.0	30.6 ± 2.1	0.62	0.45	0.57 ± 0.10	0.01 ± 0.01
82.5	24.0 ± 4.2	0.70	0.58	0.59 ± 0.10	0.00 ± 0.02
165.0	20.6 ± 6.7	0.65	0.48	0.64 ± 0.11	0.01 ± 0.01
16.5	80.1 ± 12.3	1.73	1.24	0.23 ± 0.05	0.11 ± 0.02
27.5	73.6 ± 13.1	1.43	1.05	0.27 ± 0.05	0.09 ± 0.01
55.0	69.4 ± 10.0	1.33	1.01	0.30 ± 0.05	0.09 ± 0.01
82.5	72.5 ± 12.5	1.33	0.98	0.31 ± 0.06	0.08 ± 0.01
165.0	70.8 ± 12.5	1.34	1.04	0.31 ± 0.06	0.09 ± 0.00

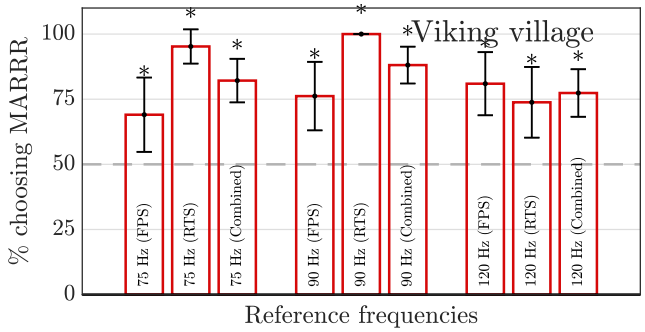

Fig. 3. Additional results for Experiment 5. Results show a clear preference for MARRR over all constant-resolution-and-constant-refresh-rate rendering conditions, including the one predicted by [Debattista et al. 2018] (75 Hz).

S.4 QUALITATIVE ABLATION STUDY

As described in the main manuscript the relative complexity of the proposed model can be justified by analyzing how each individual model component contributes to predicting the quality curves measured in Experiment 1. In Section 6.4 we present RMSE errors to show that the best fits are achieved when all model components are utilized (judder, blur, predictability of motion).

We argue that an additional benefit of using a model rooted in visual science is that predictions even outside the trained domain are plausible. This is in contrast with an empirical model with minimal mathematical complexity, which is unlikely to generalize well.

In this section, we plot the model predictions of each component within, below, and above the measured refresh rate domain of 50-165 Hz in Figures 4, 5 and 6 respectively.

S.4.1 50-165 Hz

As shown in Figure 4, a blur model (Q_P), or an empirical log model can capture the shapes of the quality curves reasonably well. The judder model on its own (Q_J) fails to distinguish between refresh rates above 100 Hz, which results in a poor overall fit. Note, that qualitative differences exist between the cruves (as discussed in Section 6.4), with the complete model yielding the best fit.

S.4.2 Above 165 Hz

Increasingly higher refresh rates will be possible on the next generation of displays. However, the qualitative value of increasing refresh rates is diminishing — current models predict a critical refresh rate of 250–700 Hz [Kuroki et al. 2006, 2007; Noland 2014]. As Figure 5 reveals, the diminishing returns is not captured by the simple empirical log model, whereas our proposed visual model produces more plausible predictions. To further analyze individual model components, we argue that the judder model reaches a plateau unexpectedly soon (around 100-150 Hz), which is inconsistent with previous findings. The blur model (Q_P) raises slightly steeper than the final proposed model.

S.4.3 Below 50 Hz

Our model does not currently target refresh rates below 50 Hz; however, as shown in Figure **??**, its predictions still look plausible. Notably, JND differences are expected increase rapidly in this region, as judder artifacts suddenly become prominent. As such, the blur model (Q_J) probably captures the expected trends the best, with the complete model coming second. Once again, we argue that low refresh rates reveal an obvious issue with the empirical log model: the extrapolated steepness is too shallow. It is worth noting, that below 60 Hz, flicker artifacts can also become objectionable; our model would probably benefit from an added flicker component to improve predictions in this range.

REFERENCES

S. J. Daly. 1992. Visible differences predictor: an algorithm for the assessment of image fidelity. (1992), 2. https://doi.org/10.1117/12.135952

S. J. Daly. 1998. Engineering observations from spatiovelocity and spatiotemporal visual models, Bernice E. Rogowitz and Thrasyvoulos N. Pappas (Eds.). 180–191. https://doi.org/10.1117/12.320110

K. Debattista, K. Bugeja, S. Spina, T. Bashford-Rogers, and V. Hulusic. 2018. Frame Rate vs Resolution: A Subjective Evaluation of Spatiotemporal Perceived Quality Under Varying Computational Budgets. *Computer Graphics Forum* 37, 1 (feb 2018), 363–374. https://doi.org/10.1111/cgf.13302

B Y M A Georgeson and G D Sullivan. 1975. Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of Physiology* 252, 3 (1975), 627–656.

Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura. 2006. 3.4: Improvement of Motion Image Quality by High Frame Rate. *SID Symposium Digest of Technical Papers* 37, 1 (2006), 14. https://doi.org/10.1889/1.2433276

Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura. 2007. A psychophysical study of improvements in motion-image quality by using high frame rates. *Journal of the Society for Information Display* 15, 1 (2007), 61. https://doi.org/10.1889/1.2451560

K Noland. 2014. The application of sampling theory to television frame rate requirements. *BBC Research & Development White Paper* 282 (2014).

M. Suh, R. Kolster, R. Sarkar, B. McCandliss, and J. Ghajar. 2006. Deficits in predictive smooth pursuit after mild traumatic brain injury. *Neuroscience Letters* 401, 1-2 (jun 2006), 108–113. https://doi.org/10.1016/j.neulet.2006.02.074

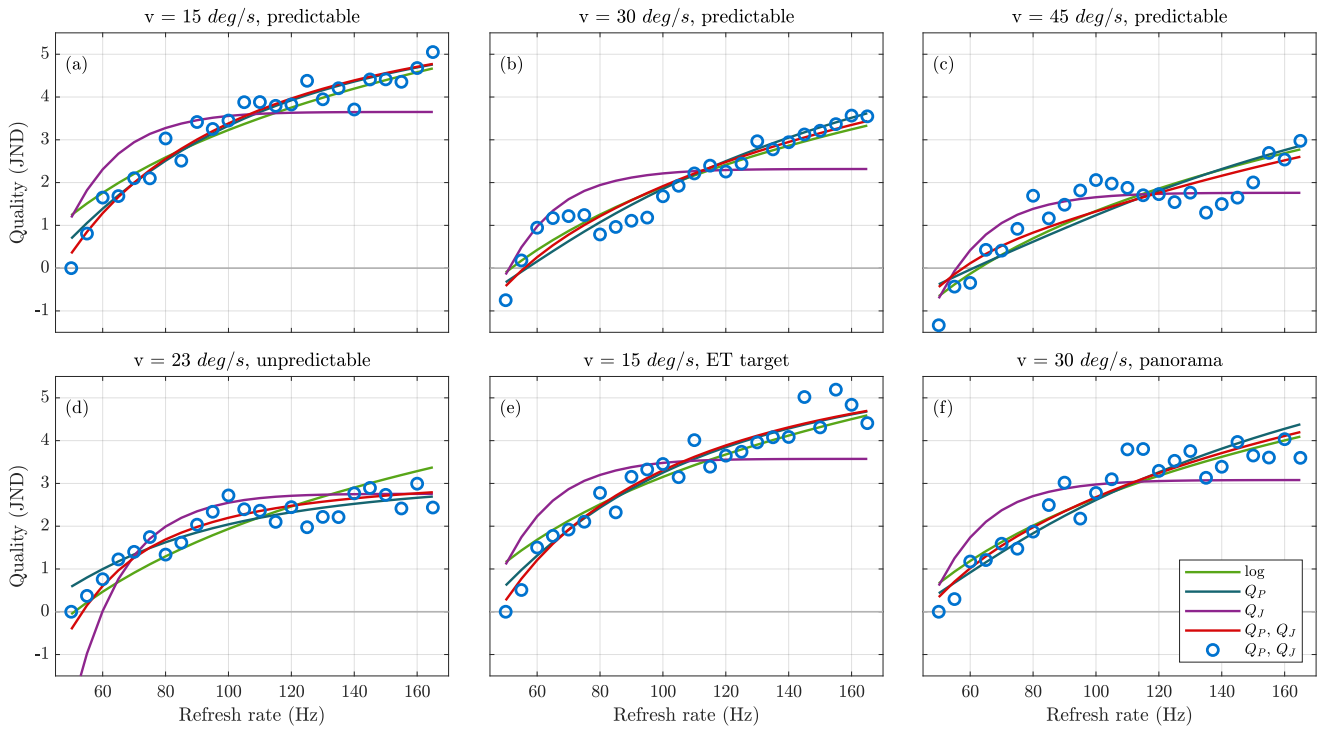

Fig. 4. Predictions of different model components for the target refresh rate range (from 50 Hz to 165 Hz). With the exception of the judder-only model (Q_J), all models provide reasonable predictions.

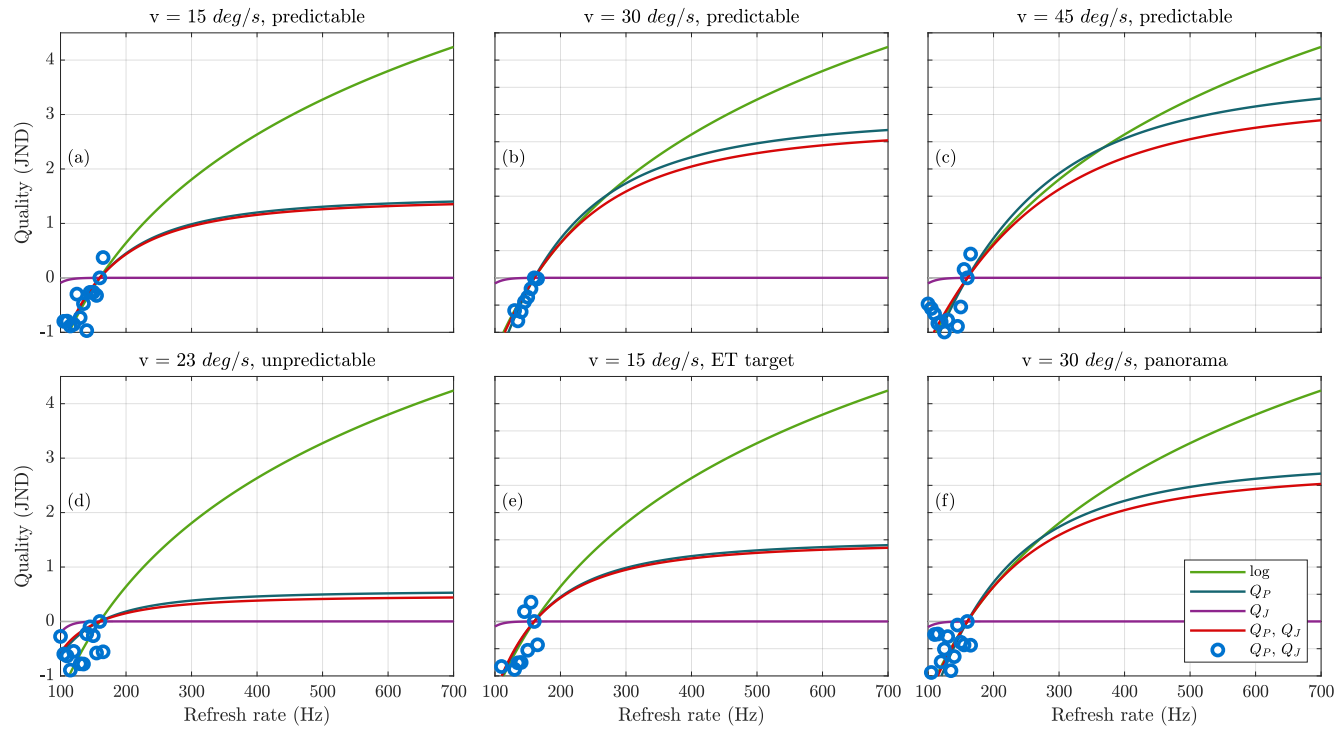

Fig. 5. Predictions of different model components above the target refresh rate range (>165 Hz). The complete proposed model (red) provides the most reasonable predictions; the log-model (green) is inconsistent with the visual system's diminishing ability to differentiate between high refresh rates; the judder model (purple) is inconsistent with existing measurements which show that humans can differentiate between refresh rates above 150 Hz.

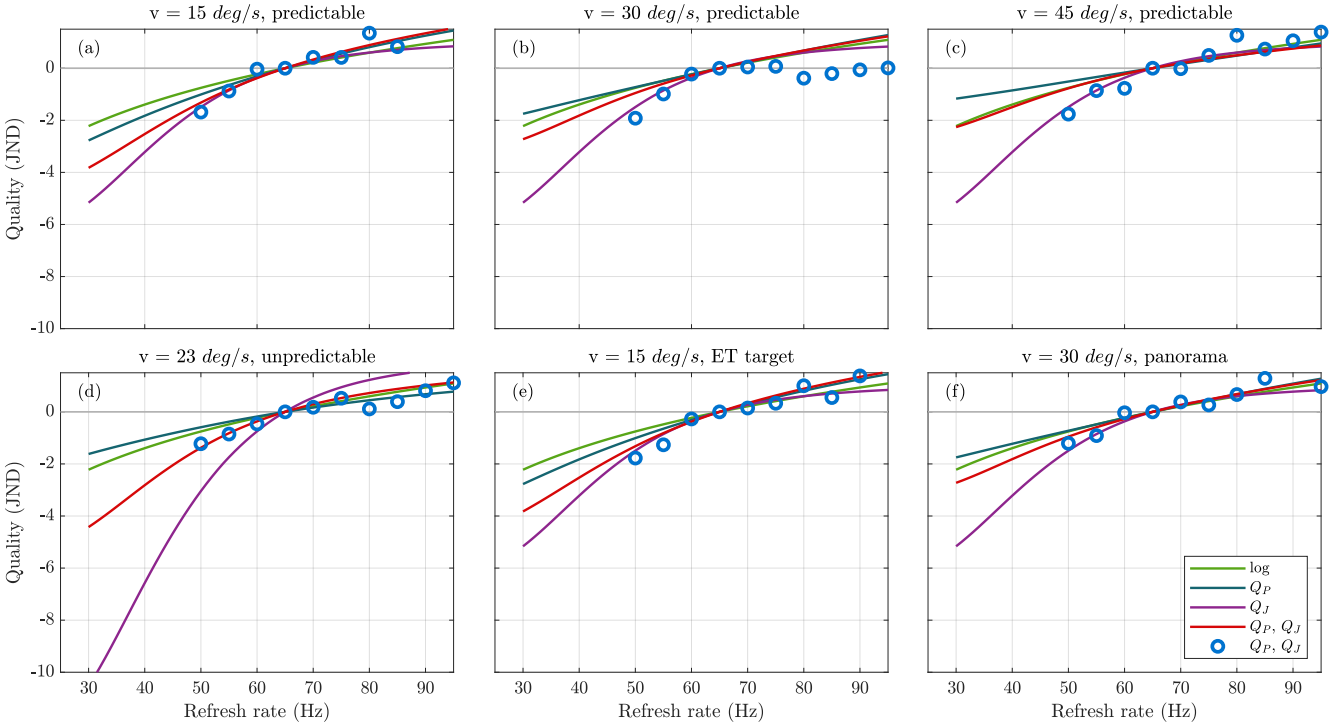

Fig. 6. Predictions of different model components below the target refresh rate range (<50 Hz). Witch such low refresh rates, the quality curve is expected to be reasonably steep. The judder-only model provides perhaps the most intuitive results, with the complete proposed model (red) coming second best. For more accurate predictions, flicker artifacts would also need to be considered.