A Novel Wavelet-Based Flicker Metric for Variable Refresh-Rate Displays

Hamid Reza Tohidypour*, Anahita Shojaei-Hashemi*, Frank Seto, Panos Nasiopoulos***
*** Department of Electrical & Computer Engineering, University of British Columbia, Vancouver, Canada**
**** Samsung Display America Lab, San Jose, CA, USA**

Abstract
We propose a wavelet packet-based flicker metric for variable refresh rate (VRR) displays, addressing limitations of existing metrics. Our method, integrating temporal contrast sensitivity and wavelets, outperforms current approaches by 17.88% in subjective evaluation across luminance and technology, achieving a Pearson correlation of 89.05%.

Author Keywords
Variable refresh rate; Flicker measurement; Human visual system; Wavelet transform; Temporal contrast sensitivity function; Subjective tests; Display metrology.

1. Introduction
The emergence of variable refresh rate (VRR) signifies a recent norm driven by advancements in video technologies and opportunity for energy conservation through dynamic changes in screen refresh rates. The primary goal of these advancements is to enhance user experiences, such as reducing lag or saving power; however, achieving this without introducing flickering remains a challenge, as there is currently no effective metric to objectively measure VRR flickering, forcing manufacturers to either be overly restrictive or to rely on subjective evaluation to identify and address the issue. However, relying on such methods are inefficient, resource intensive, and often happens too late in the design phase. Therefore, it is crucial to develop an objective measure of VRR flicker, enabling improvements to be made earlier in the design cycle for display makers. This metric would enable them to predict perceived flicker levels earlier during the design process, minimizing the reliance on human subjective evaluations at the end of the production chain.

The existing methods for measuring flicker were originally developed for earlier display technologies, where the refresh rate is fixed [2,3]. These methods can be categorized into time-domain and frequency-domain metrics. Time-domain based standards for flicker measurement include the LCD Contrast Min/Max and LCD Contrast RMS methods [2]. On the other hand, the Japan Electronics and Information Technology Industries Association (JEITA) metric, as well as another variation developed by Video Electronics Standards Association (VESA), are frequency-based metrics and are commonly used to assess flicker by the display industry [2,3]. Both of these metrics utilize the temporal contrast sensitivity function (TCSF), which allows them to accurately predict display flicker, at least for fixed refresh rates. However, these metrics fail to achieve similar accuracy levels for the case of VRR. This is mainly due to the fact that they rely on the fast Fourier transform (FFT), which tends to represent each coefficient as the average of a specific frequency across the entire signal, thus failing to track dynamic changes in time [4].

To address these shortcomings, in this paper we introduce a novel metric designed specifically for measuring VRR flicker. Our proposed metric is based on the wavelet packet transform, which allows it to effectively capture both the high-frequency transient components and the low-frequency slow-varying components of the VRR luminance signal—an essential feature in VRR signal analysis. For the analysis of luminance signals in the wavelet domain, we employed a wavelet packet transform that allows us to analyze and weight the signal according to TCSF to accurately reflect the VRR flicker level that correlates to human perception. To evaluate the performance of our metric across various display technologies and a wide range of luminance and flicker levels, we conducted a comprehensive subjective test in our Digital Multimedia Lab (DML) at the University of British Columbia (UBC). Performance evaluations revealed that our metric achieves a Pearson correlation of 89.05% with the mean opinion scores (MOS) from subjects, reflecting an improvement of 17.88% over JEITA, the current flicker measurement method in the display industry.

The rest of this paper is organized as follows: Section 2 explains our proposed VRR metric. In Section 3, we evaluate the performance of our metric and analyze the results. Finally, Section 4 presents our concluding remarks.

2. Our proposed VRR metric
Our objective is to design an accurate flicker metric for VRR displays that analyzes the captured luminance signal received from the display and assesses the perceived flicker levels. Our method aims to address the shortcomings of existing approaches that fail to consider both the time and frequency aspects of the signal. These shortcomings are related to the use of the FFT for tracking time-frequency changes in VRR displays, as this transform averages a specific frequency across the entire signal, which can lead to the cancellation or accumulation of distinct frequencies at different time intervals [4].

In our approach, we utilized the wavelet transform to analyze the signal, since it effectively captures both temporal and frequency information. From the wavelet transform family, we chose the wavelet packet transform (WPT) as it allows us to analyze the signal in accordance with TCSF.

Wavelet analysis decomposes the signal into approximation coefficients and detail coefficients [5]. Figure 1 shows the filter-bank representation of the $(s+1)^{th}$ stage of decomposition of the WPT applied to the current coefficients $C_{s,j}[n]$ *(to-be-decomposed)*. $C_{s+1,2j}[n]$ and $C_{s+1,2j+1}[n]$ represent the obtained approximation and detail coefficients of the $(s+1)^{th}$ stage decomposition, respectively. Approximation coefficients represent the low-frequency components of the signal, capturing the overall trend or smooth features, while detail coefficients represent the high-frequency components, capturing the rapid changes or fine details. Note that the filter-bank implementation of the first stage, which corresponds to s=0 and j=0, decomposition of WPT is the same as that of the first stage of decomposition of the Discrete Wavelet Transform (DWT).

Unlike the DWT, wavelet packets extend this concept by allowing both approximation and detail coefficients to be further decomposed, resulting in a hierarchical structure that provides a multi-resolution analysis and enhances the representation of the signal. Additionally, wavelet packet analysis provides greater

ISSN 0097-996X/25/5601-0484-$1.00 © 2025

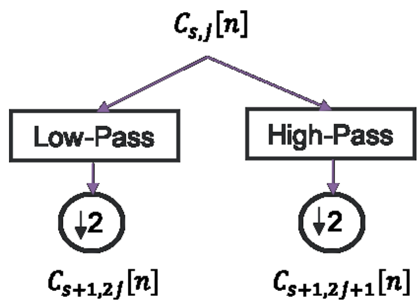

Figure 1. The filter-bank representation of the (s+1)th stage of decomposition of the WPT.

flexibility in selecting the decomposition level, enabling the identification of transient features that other methods may overlook. This capability is crucial in detecting VRR flicker level, as it helps highlight significant variations and anomalies within the data.

The sampling rate of the device that we used to measure display luminance was fs=3000 samples/sec, while the desired frequency range for analysis was 0-60 Hz, values that are in accordance with human perception as verified by the TCSF functions [6]. Considering these factors and the Nyquist theorem, we can determine the number of wavelet decomposition levels needed to obtain an approximation sub-band covering the specified frequency range. Figure 2 illustrates that four (4) decomposition levels are required to acquire an approximation sub-band that covers the desired range of 0-60 Hz. Note that the 5th stage of decomposition only covers 0-46.875 Hz. Thus, we determined that only four decomposition levels are needed for covering the range of 0-60 Hz. However, the sensitivity with respect to temporal frequency varies across different frequencies. This means that the approximation sub-band must be weighted in accordance with the TCSF.

To accurately weight the approximation sub-band, we need to consider TCSF's properties and the shape of its curves. TCSF describes the sensitivity of the visual system to contrast changes over time at different temporal frequencies (rates of light modulation). Sensitivity is generally high at low temporal frequencies (slow changes) and peaks at intermediate frequencies, typically around 5–20 Hz. Beyond this peak, at very high frequencies, sensitivity declines as the visual system has difficulty perceiving rapid changes. Figure 3 shows a TCSF curve for an example luminance level.

In order to more precisely weight the luminance signal according to the TCSF, we further decomposed the approximation sub-band of the 4th stage up to the 8th stage of decomposition using WPT. Unlike before, we included both approximation and detail (high-frequency) sub-bands in this process, resulting in a total of 16 sub-bands. The approximation coefficients at different stages of decomposition are computed as follows:

$$c_{s+1,2j}[n] = \begin{cases} \sum_{k=-\infty}^{\infty}\left(c_{s,j}[k]g[2n-k]\right), where \\ \quad 0 \le s \le 3, j = 0 \\ \sum_{k=-\infty}^{\infty}\left(c_{s,j}[k]g[2n-k]\right), where \\ \quad 4 \le s \le 7,\ 0 \le j \le 2^{s-4}-1 \end{cases} \tag{1}$$

where $g[n]$ represents the wavelet's low-pass filter. Note that $c_{0,0}$ represents the input luminance signal. On the other hand, the detail coefficients at different stages of decomposition are computed as follows:

$$c_{s+1,2j+1}[n] = \begin{cases} \sum_{k=-\infty}^{\infty}\left(c_{s,j}[k]h[2n-k]\right), where \\ \quad 0 \le s \le 3, j = 0 \\ \sum_{k=-\infty}^{\infty}\left(c_{s,j}[k]h[2n-k]\right), where \\ \quad 4 \le s \le 7,\ 0 \le j \le 2^{s-4}-1 \end{cases} \tag{2}$$

where $h[n]$ represents the wavelet's high-pass filter. Each of the resulting 16 sub-bands at the 8th stage of decomposition covers a different frequency range and varies in importance according to

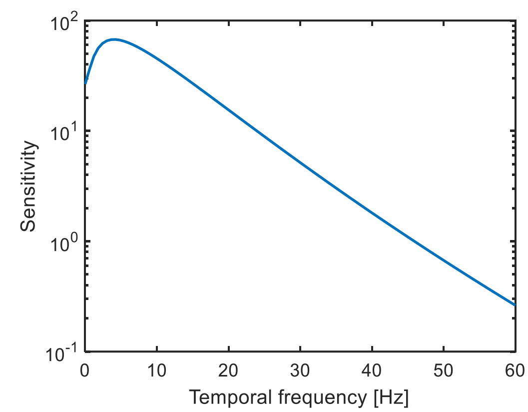

Figure 3. TCSF for an example luminance level.

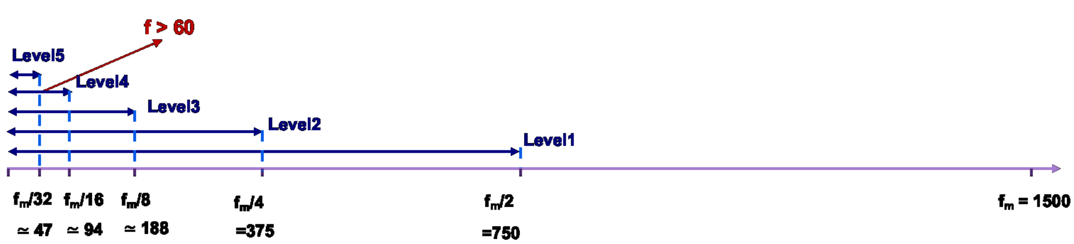

Figure 2. Illustration of the number of decomposition levels needed for obtaining an approximation sub-band that covers 0-60Hz.

the TCSF. Therefore, we decided to apply different weights to each sub-band based on TCSF. More specifically, for each of the sub-bands, we identified the corresponding frequency interval on TCSF and used the average TCSF values for that interval as the weight for each sub-band.

We applied the inverse wavelet transform to the weighted 16 sub-bands and reconstructed the weighted signal up to the 4th stage of decomposition to analyze the weighted signal covering the 0-60 Hz range (for flicker detection). This reconstructed sub-band represents the approximation sub-band of the 4th stage. $c_{4,0}'$ represents the reconstructed approximation sub-band of the 4th stage.

We designed our VRR metric based on the LCD contrast RMS metric [2] to track changes in wavelet coefficients. The original RMS calculation method computes the AC value of the luminance signal, which is then normalized by the DC level, defined as the average (mean) of the signal. In our implementation, we modified the original RMS by replacing the mean function with another measure of central tendency, the median function. This is because median is less sensitive to outliers, and our initial analysis showed that it outperformed the mean in terms of VRR flicker measurement performance. As a result, our proposed VRR flicker metric is as follows:

$$flicker_{WPT\ rms} = \frac{\sqrt{\frac{1}{L}\sum_{n=0}^{L-1}\left(c_{4,0}'[n]-median(c_{4,0}'[n])\right)^2}}{median(c_{4,0}'[n])}*100\% \quad (3)$$

3. Results and Discussion

To evaluate the performance of our metric, we need to assess its correlation with human perception. To this end, we created a diverse square shaped VRR sequences, representing a wide range of flicker levels and luminance levels across different display technologies [7]. These sequences cover different flicker categories, namely no flicker, mild flicker, moderate flicker, and severe flicker. Figure 4 shows a sample generated VRR signal. In each of our sequences, the refresh rate periodically switches between two luminance values within a fixed period, with the duty cycles set according to the frequencies supported by the displays. The flicker is generated in an objective and repeatable way, based on these specific settings. We conducted a comprehensive subjective test to measure the perceived flicker level for each of our generated sequences, during which the subjects were asked to evaluate the sequences in terms of their flicker levels. Details of our subjective test and comprehensive dataset will be presented in a forthcoming paper.

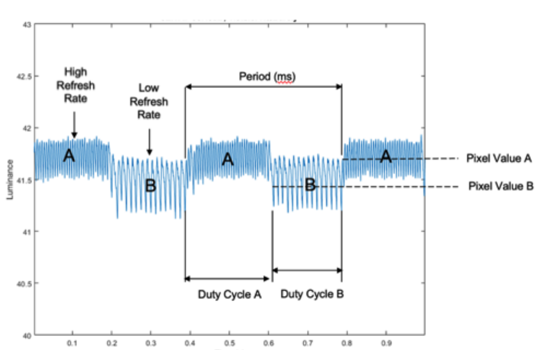

Figure 4. Sample VRR signal used for our subjective tests.

Table 1. PCC of our proposed method vs other methods.

Pearson Correlation			
Our proposed metric	Min/Max	RMS	JEITA
89.05%	- 2.13%	6.14%	71.17%

We evaluated the performance of our metric using our dataset. In this regard, we employed Pearson correlation coefficient (PCC) to examine how closely the values of the metric align with the subjects' scores [8]. Table 1 presents the PCC values achieved by our metric, Min/Max, RMS and JEITA. We observe that our metric outperforms the existing metrics, yielding a PCC value of 89.05% with MOS. It is worth noting that the Min/Max and RMS, which do not analyze the signal in the frequency domain, showed very low correlation, yielding PCC values of -2.13% and 6.14%, respectively. Finally, the JEITA metric achieved a PCC of 71.17%, which is 17.88% lower than that achieved by our proposed metric.

To demonstrate the superiority of our proposed metric in differentiating between various flicker levels—no, mild, moderate, and severe flicker—we plotted the distribution of our metric for each MOS flicker category (see Figure 5 a). In this figure, the median value of the metric for each flicker level is represented by the line inside the box. The box itself spans from the 25th percentile to the 75th percentile, showing the interquartile range (IQR) where most of the data is concentrated. The whiskers extend to the minimum and maximum values within

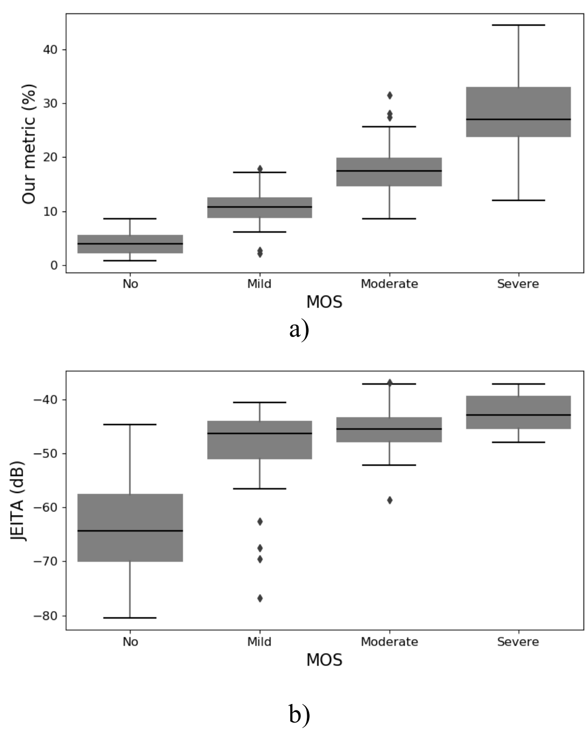

Figure 5. Distribution of a) our wavelet based metric and b) JEITA metric values for each MOS flicker category.

1.5 times the IQR, while individual points outside the whiskers indicate outliers, which are values significantly higher or lower than the rest of the data. Figure 5 b shows the same analysis for JEITA in order to highlight our method's advantages. We observe that JEITA exhibits significant overlap in its distribution across the no, mild, medium, and severe flicker levels. This overlap is extensive, indicating that JEITA can only loosely classify flicker into two broad categories: no flicker and flicker, with possibly significant misclassification. In contrast, our metric shows less overlap between flicker levels, enabling a clearer distinction across four categories. Note that some overlap between adjacent levels is expected, as human perception may struggle to differentiate between them, and the dataset could contain noise at these boundaries. Furthermore, unlike JEITA, our metric demonstrates a linear correlation with the MOS, with its value increasing as the perceived flicker intensity rises. These advantages can provide valuable insights for display makers during their design and optimization processes.

The superiority of our metric comes from the fact that, unlike JEITA, our method employs a wavelet function that averages frequencies over smaller time intervals, allowing us to focus on specific moments where transitions may occur, which is crucial for variable flicker measurement. This significantly enhances the superiority of our metric.

Overall, our metric demonstrates a strong correlation with MOS, making it an effective objective measure for VRR flicker. One challenge we faced comes from the fact that people perceive flicker levels differently. Future improvements to the flicker metric should address these challenges. Note that we tested sequences with average luminance range from 1 to 40 cd/m². Within this range, a limitation of our approach is that the metric ranges for different flicker levels at luminance levels below 2 cd/m² do not accurately match those of higher luminance levels. Further research in this area is needed, possibly by concentrating on an updated TCSF that better represents human sensitivity at low luminance levels. Another possible future research direction may be to check the performance of our metric for luminance ranges above 40 cd/m².

Recall that our current implementation relies on a capture sampling rate of 3 kHz. However, according to the TCSF, the cutoff frequency is much lower than this sampling rate. Future work could also focus on determining the minimum sampling rate and duration required to maintain performance.

4. Conclusion

In this paper, we proposed a novel wavelet packet based flicker measurement metric that effectively captures both temporal and frequency information—an essential feature in VRR signal analysis for VRR displays. We designed our wavelet tree based on TCSF to analyze the luminance signal perceptually. This approach allowed us to weight the signal according to TCSF, enabling accurate measurement of the flicker level. To evaluate the performance of our metric across different display technologies and a wide range of luminance levels, we conducted a thorough subjective test to gather human subjects' assessments to compute the correlation between our metric and human subjects. Performance evaluation revealed that our metric achieved a Pearson correlation of 89.05% with the subjects' mean opinion scores (MOS), representing a 17.88% improvement over JEITA.

The proposed metric has significant potential to impact the display technology industry by addressing the limitations of existing flicker evaluation methods, particularly for VRR displays. Its strong correlation with human perception, as demonstrated by its Pearson correlation performance, offers an accurate and objective alternative to resource-intensive subjective tests. By providing clear differentiation between various flicker levels and minimizing misclassification, the metric has the potential to assist display manufacturers in optimizing designs with greater precision. Moreover, its adaptability to varying luminance levels as well as the integration of wavelet analysis and temporal contrast sensitivity make it a robust and versatile tool for evaluating VRR flicker across diverse display technologies. Ultimately, adoption of an improved metric for VRR flicker would drive the industry towards an improved and smoother viewing experience for VRR displays.

5. Acknowledgements

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, Samsung Display America Lab and Samsung Display Co., Ltd.

6. References

1. Slavenburg G.A., Janssens M., Lucas L., Schutten, R.J. and Verbeure T. 46-1: Invited Paper: Variable Refresh Rate Displays: In SID Symposium Digest of Technical Papers, 51(1), 669-672; 2020.
2. Admesy Advanced Measurement System. Flicker measurement display & lighting measurement [Technical note]; 2017. Available from: https://www.admesy.com/documents/flicker/
3. VESA. VESA Adaptive-sync display certification test specification (Adaptive-sync display CTS) [Revision 1.0]; 2022. Available from: https://vesa.org/wp-content/uploads/2022/05/Adaptive-Sync-Display-CTS-r1.0.pdf
4. Proakis J., & Manolakis, D. Digital signal processing (4th ed.): Pearson; 2006.
5. Mallat S. A. wavelet tour of signal processing. 3rd ed. San Diego (CA): Academic Press; 2008.
6. Pappas T. N., Safranek R. J., Chen J. 8.2 - Perceptual Criteria for Image Quality Evaluation: Communications, Networking and Multimedia, Handbook of Image and Video Processing (Second Edition), Academic Press, 939-959; 2005.
7. Seto F. vrr_flicker (v5) [Computer software] Github [updated 28 December 2023]. Available from: https://github.com/frankseto/AdaptiveSyncTest/tree/vrr_flicker
8. Benesty J., Chen J., Huang Y., & Cohen I. Pearson correlation coefficient: Noise reduction in speech processing, Springer, 2, 89–90; 2009.