

# Apparent Super Resolution by Vibrating Displays

Jaclyn Pytlarz

JPytlarz@stanford.edu

Stanford University

EE367 Winter 2019-2020

## Abstract

*Augmented reality (AR) and virtual reality (VR) displays are reaching a point where hardware constraints are in the way of increasing the resolution. This paper presents a novel algorithm using a vibrating display at sub-pixel resolution to produce apparent higher resolution imagery. It accounts for the critical flicker frequency dependence on Luminance adaptation to dynamically adjust the image on a per-pixel basis. This algorithm reduces flicker while increasing perceived resolution.*

*A psychovisual experiment is conducted that compares this method against a non-vibrating low resolution display as well as a vibrating display without flicker compensation. It is found that the vibrating display with flicker compensation leads to significantly improved image quality for those images that contain bright saturated content. It is also found that this algorithm results in a significant reduction in flicker.*

## 1. Introduction

The focus on resolution of virtual reality headsets has been the forefront of this technology for the last number of years. Today headsets can reach around 2000 by 2000 pixels per eye with few reaching 4000. When designing a display, a dream goal is to produce an image so lifelike that it is indistinguishable from reality. To this end, you need the display to exceed the capabilities of the human visual system in every way. Considering resolution, the human visual threshold is around 60 cycles per degree (cpd). At the current resolution of headsets today assuming 90 degree field of view, they are only achieving around 11 cycles per degree. This pails in comparison to a typical 4K television which is already achieving over 60 cpd. This resolution limitation in AR/VR causes pixel artifacts to be easily visible when utilizing this small screen technology. It has become a road block in the advancement and adoption of virtual reality. Improving resolution of AR/VR headsets is vital to its success in the long term.

Even more, building arrays of tighter packed and smaller pixels is nearing hardware constraints. New avenues are being explored, but the likelihood of a new hardware solution in the near future is low. In the mean time, the industry is in need of an interim solution to get perceived high resolution while still being limited to the lower resolution displays.

## 2. Related Work

The methods to solve the physical resolution limit of today's AR/VR displays have spanned a wide field, from physical optical based solutions, to screen vibration, to sub-pixel rendering. One of the earliest looks at increasing resolution on a fixed resolution display was done using two projectors [1]. The projectors were overlapping by half a pixel. Methods were employed to improve calibration and to account for warp due to projection offset. Then the images sent from each projector were optimized so that when visually integrated, the image had apparent higher resolution. Although the two display method does not directly relate to AR/VR solutions, it sparked work in similar directions.

In 2005, work was done to emulate the two display method using a single vibrating display [2]. The display was vibrated at sub-pixel resolution using a weighted motor as shown in Figure 1. The point spread functions (PSFs) generated during the rotation for each frame were used during the image optimization. They were able to run the display at 120 frames per second (fps). With the four images being displayed, it was a per location frame rate of 30 fps. The results were very promising (aside from difficulties of calibration), but the main drawback was image flicker. The human critical flicker frequency (the point at which image flicker fuses into a smooth image due to temporal integration) is typically thought to be around 30 Hz, but actually depends heavily on adapting Luminance. I will come back to this shortly.

Since this work in 2005, majority of the solutions have used a more static methodology. Some have explored optical methods [3] or trying to improve the lens array itself [4]. In 2018, a sub-pixel solution was explored that utilized the fact that every pixel is made up of red, green, and blue sub-

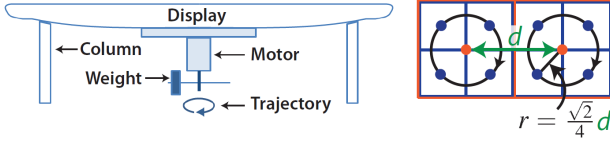


Figure 1: Model of Vibrating Display from [2]

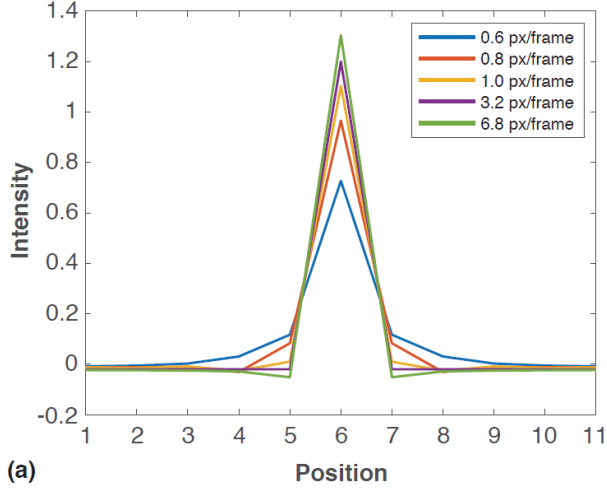


Figure 2: Fourier sub-pixel filters from [5]

pixels [5] as seen in Figure 2. They were aiming heavily for real-time application and used a bank of Fourier filters to optimize each image. Unfortunately this solution does not work well for still imagery as it relies on the smooth pursuit of our eye during tracking. Since 2018, some work has commenced back in the field of vibrating displays although to this authors knowledge, still no solutions exist to create apparent high resolution still imagery without inducing flicker.

### 3. Theory / Design

This paper will focus on improvements to the vibrating display. Simulations of a vibrating display was accomplished by shifting single pixels on a high resolution display. All images were rendered with four pixels corresponding to a single viewable pixel. As a baseline, gradient descent was used to mimic the method presented by [2]. This was accomplished by solving for the system of equations in Equation 1 through 100 iterations where  $\alpha$  was set to 0.5. In this equation,  $x$  represents the four images we are solving for, and  $b$  is the high resolution reference. The matrix  $A$  does the subsampling and temporal fusing (simulating the sub-pixel shift).

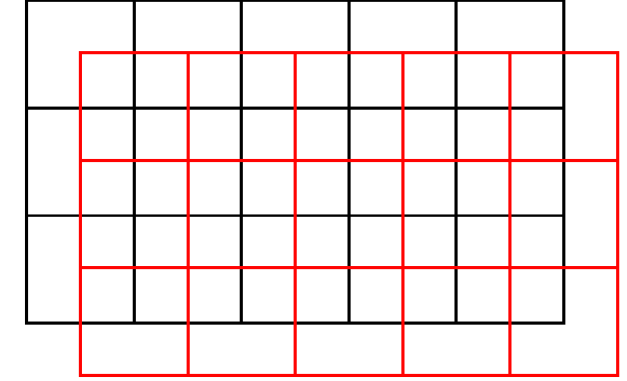


Figure 3: Sub-sampling technique solved with gradient descent and sub-pixel offsets

$$x = x - \alpha A^T (Ax - b) \quad (1)$$

The goal was to generate four images that when temporally fused would produce an image as close to the original high resolution image as possible. The image would be presented at sub-pixel accuracy (on the high resolution display) in the order of top right, bottom right, bottom left, and top left in a circular motion. The sub-pixel shift on the high resolution display is illustrated in Figure 3. Here the black squares represent the top left rendering and the red outlines represent the bottom right rendering.

The PSNR improvement during gradient descent is shown in Figure 4. After optimization, the four images are repeated in rotational manner intended for playback at 120 fps. They are combined using the FFMPEG tool with minimal compression. A static low resolution version was also created for comparison. The downsampling for the low resolution reference was accomplished with bilinear interpolation.

To solve the flickering, a relationship between the human critical flicker frequency (CFF) and Luminance needed to be derived. Luckily there has already been much research in parallel areas. One such paper [7] written in 1966 relates the critical flicker frequency to log retinal illuminance

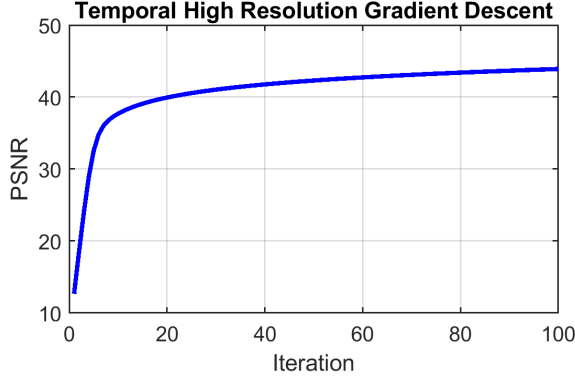


Figure 4: Temporal High Resolution PSNR with Gradient Descent

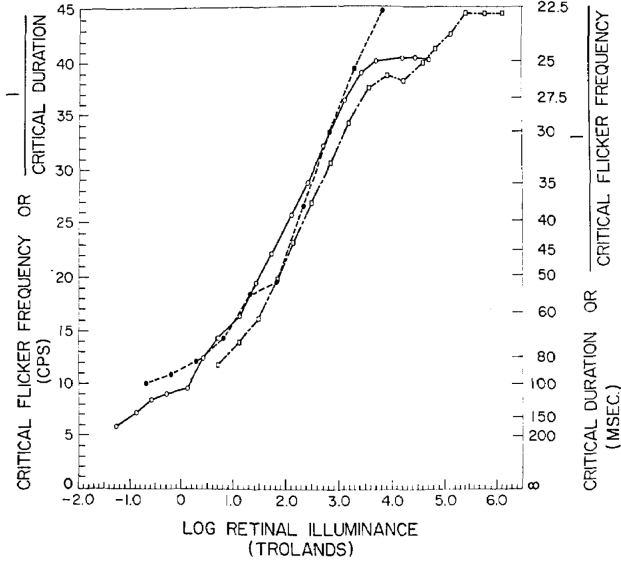


Figure 5: CFF vs Trolands from [7]

(Trolands). The CFF is the threshold at which the frequency of repeating bright and dark patches just fuses into a single gray image. The resulting data is shown in Figure 5. Trolands are equal to the product of Luminance and pupil area in  $mm^2$ .

$$T = L * P$$

Since the pupil area changes with Luminance, we now need to know the relationship between adapting luminance and pupil diameter. Fortunately this has also been researched heavily over the past 80 years. In 2012, Watson and Yellott [9] created a unified formula which encompassed all of the data as shown in Figure 6. This unified formula takes into account both age and number of eyes viewing the sample. For the purposes of this paper I have set the age to 28.58 (which significantly simplifies the equa-

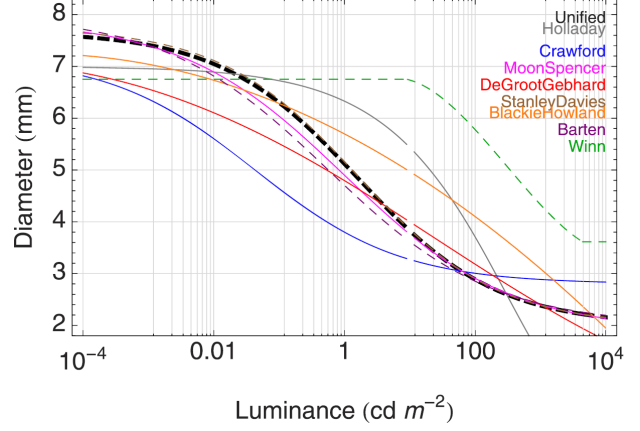


Figure 6: Pupil diameter vs Luminance from [9]

tion), and number of eyes to 2. For the area of the patch in degrees,  $a$ , I have set it to typical TV height in degrees squared of  $18.4^2$ . The unified formula with my simplifications for age and eyes is given below in Equation 2. Trends to note are that if you are only using one eye for viewing, the pupil diameters will tend to decrease. Likewise, as the age increases, the dynamic range the pupil will span with adapting Luminance decreases.

$$D_{SD}(L, a) = 7.75 - 5.75 \left( \frac{(La/846)^{0.41}}{(La/846)^{0.41} + 2} \right) \quad (2)$$

With this model and the CFF relationship from [7], we can now derive the relationship between Luminance and CFF. The result of this is shown in Figure 7. For the x-axis, Luminance is encoded in a perceptual quantizer (PQ) [8] typically used for HDR encoding. It was found that this encoding provides a linear relationship between CFF and Luminance.

We can now use this relationship to alter our optimization. The best case no-flickering option is to have a static image. So that will serve as our baseline. This will be accomplished with a 2x2 blur kernel simulating stepping further away from the screen. The gradient descent optimization will then be calculated twice: once for the static imaging, and once for the dynamic high resolution image. We will then blend between these renderings based on the CFF relationship we had previously derived. The PSNR during gradient descent for the blurred image is shown in Figure 8. The overall PSNR values are higher which is expected because this blurred image has more similarity between pixels.

The CFF vs PQ Luminance equation doesn't yet relate to an alpha value we can use for blending. We know that at lower Luminance the flickering is not visible, so we can safely use the dynamic version. Then at high Luminance, we are more susceptible to flicker so we'd like to use more

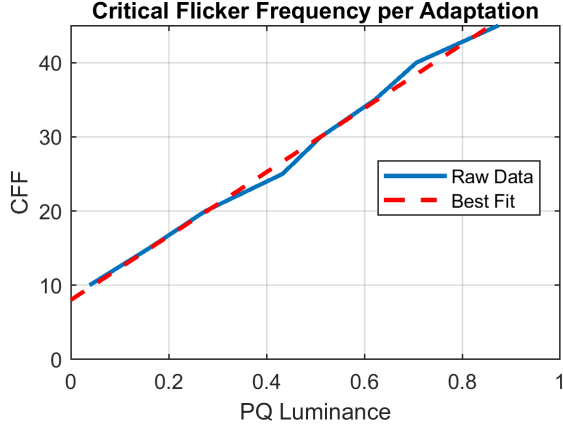


Figure 7: Derived critical flicker frequency vs PQ Luminance where  $y = x * 43.07 + 8.02$

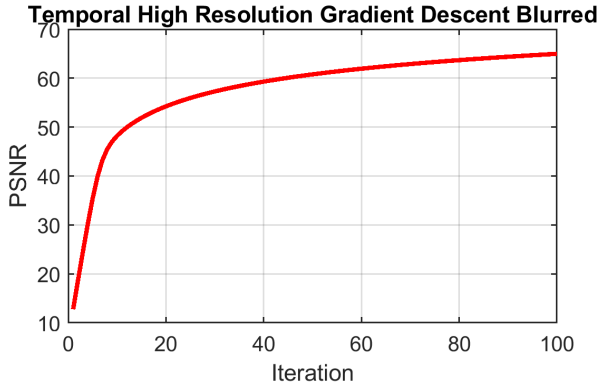


Figure 8: Blurred Image PSNR with Gradient Descent

of the static version. Since this is a linear relationship, we can simply normalize the equation.

$$\begin{aligned}
 CFF &= PQ * 43.07 + 8.02 \\
 \alpha &= \frac{CFF}{45} \\
 \alpha &= \frac{PQ * 43.07 + 8.02}{45}
 \end{aligned} \quad (3)$$

The PQ Luminance value will be taken from the 2x2 blurred image to allow for some spatial integration during interpolation. This helps limit additional induced flicker. The PQ value can be calculated by Equation 4. Display luminance for standard dynamic range television is normally between 100 and 400  $cd/m^2$ . For the purposes of this paper, 300  $cd/m^2$  was used. The RGB to Y matrix derivation can be found in [6]. The linear to PQ equation can be found in ITU-R BT.2100 [8].

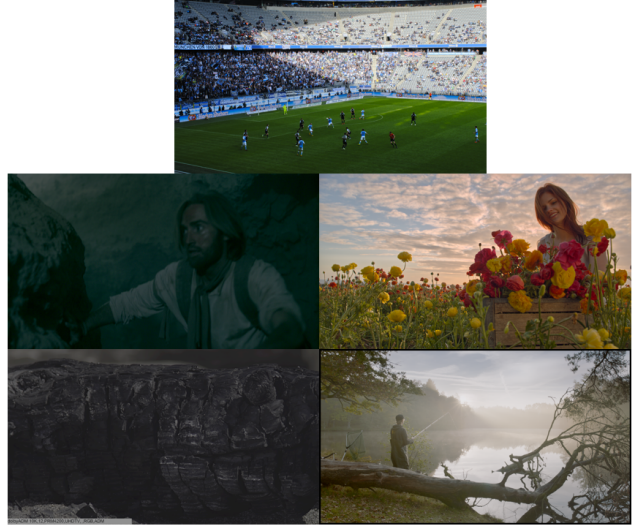


Figure 9: Images used in the psychovisual experiment

$$\begin{aligned}
 RGB &= im^{2.4} * displayLuminance \\
 Y &= R * 0.2126 + G * 0.7152 + B * 0.0722 \quad (4) \\
 PQ &= LinearToPQ(Y)
 \end{aligned}$$

The  $\alpha$  values are calculated per pixel and the output image is a linear blend between the blur optimized versions and the sharp versions as show in Equation 5. This output is then repeated as before in a clockwise manner and is encoded at 120 fps.

$$\begin{aligned}
 &\text{for } j = 1 : 4 \\
 &\quad out(j) = \alpha * blurrIm + (1 - \alpha) * sharpIm \quad (5) \\
 &\text{end}
 \end{aligned}$$

## 4. Evaluation Method

For evaluation, a psychovisual experiment was conducted. Observers were asked to rank the quality of the video they were watching. There were 5 images total which are shown in Figure 9. These images were chosen because they spanned a wide range of Luminance levels and level of detail.

There were three choices during the experiment: static low resolution image, dynamic high resolution rendering, and the flicker compensated high resolution rendering. Out of these three, the observers were asked to pick the image he/she thought was the highest quality. The experiment had 15 observers all in the 20-35 age range. The low age range was targeted due to the curve fitting parameters chosen for the pupil diameter. There were 5 female and 10 male observers all with normal or corrected vision.

The display used to view the results was a calibrated LG C9 display in Technicolor reference mode. This mode was selected to ensure motion smoothing was disabled. A test image with sequential numbers at 120 fps was displayed and captured with a slow motion camera to ensure a true 120 fps sequence would be displayed without blended or dropped frames. The actual peak Luminance of the display was  $283 \text{ cd/m}^2$  and the calibrated gamma was 2.4. The images were viewed in a dark surround and results were recorded by a moderator. The observer was asked to sit three picture heights away from the screen for appropriate 1920x1080 resolution viewing.

As a follow up, observers were asked to rate the flicker of each processing method on the soccer image with a scale of 1 to 5, with 1 being no flicker and 5 being obvious and annoying flicker. This was done to see if the proposed solution met the goal of reducing flicker while improving image quality.

## 5. Results

The image quality ratings were analyzed first. The observers responded by selecting the image processing method they thought resulted in the best image quality (the order of the images was random, so the observers didn't actually know the algorithm they were rating). The average of the responses is given in Figure 10. The red bar represents the static version of the image which is simulating a non-vibrating low resolution display. The green bar showcases results of the dynamic vibration method similar to [2]. The blue bars are the results from the dynamic/static flicker reduction method as described in the previous sections. The results are split into the five images that was being rated. Trends to note are that the flicker reduction algorithm generally outperforms the other two methods. The cave image (dark with little detail) all three methods performed roughly equally. The image order was not randomized during the experiment for easier data handling.

The soccer image was further analyzed with an additional user study to discern improvements in the visibility of flicker (Figure 11). The amount of flicker is given as the average rating over all observers. A value of 1 means no flicker, and a value of 5 means annoying extreme flicker. The static image was never chosen to have any flicker, which is why the bar is down at exactly 1. The dynamic method was shown to exhibit large amounts of flicker. The flicker compensation method (labeled Ours) shows significant improvement. All results have less than one as the standard deviation.

## 6. Discussion

The main goal of the proposed method was to reduce the visibility of flicker in bright regions. That being the case,

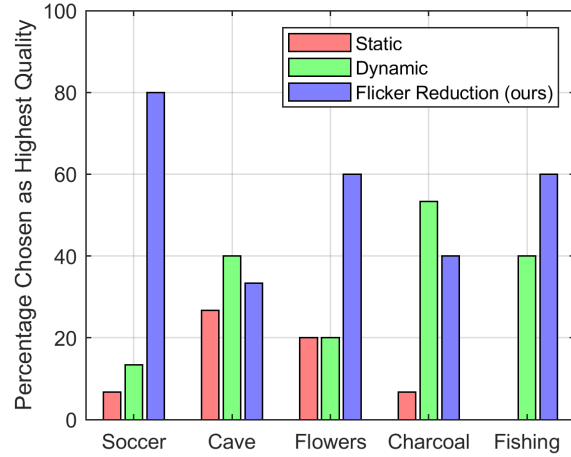


Figure 10: Image quality rating from the psychovisual experiment

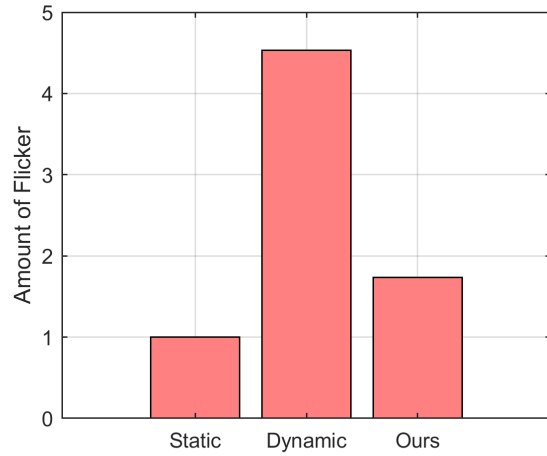


Figure 11: Amount of flicker in the soccer image

we will begin by discussing Figure 11. As expected, the static version of the image did not exhibit any flicker. As the image is not changing, any hint at flicker would have been caused by internal display processing, so it's good to see that it was consistently rated at 1. The dynamic version also as expected was rated to exhibit a substantial amount of flicker. This was easily visible in the bright regions of the stands and was obvious and annoying. The flicker reduction method results in a significant reduction of flicker. When reviewing the image and talking with observers after the experiment, it was determined that some of the remaining flicker came from edge artifacts around the screen. In the future it may be beneficial to handle the edges differently or to crop the viewing region. Aside from these edge

artifacts, the reduction in flicker was substantial.

In order to be a successful algorithm, the reduction in flicker should come with an increase in perceived image quality. Without this, it would be equivalent to the static method and would make the algorithm obsolete. The results of the quality experiment are given in Figure 10. At first glance, we can see that the flicker reduction seems to be improving the perceived quality of imagery. Starting with the soccer image, we see that the flicker reduction method is significantly higher rated than the dynamic or static method. Likely the significant increase over the dynamic method is due to the reduction of flicker in the bright scenes.

The cave scene is a very dark scene that is quite blurry overall (lacking small detail). The quality ratings show roughly equal preference to all methods. This is expected as we don't expect flicker to be an issue for a dark scene and with the lack of small details, the reduction in apparent resolution doesn't negatively affect the image quality.

Both the cave and the charcoal image are very dark images. For both of these images, the dynamic version is rated as higher quality. This may be because the flicker reduction is stronger than necessary at this low Luminance level. Since the flicker reduction is a linear relationship, at these low levels it's still applying a percentage of the static image where it may not be necessary.

For the images that have substantial dark detail, but bright areas as well (soccer and flowers), the flicker reduction method outperforms the other two. The fishing image has a number of small details in the tree branches. Likely the static method with the resolution reduction is more of a detriment than apparent flicker. Also because this image is fairly bright overall, the adaptation will likely be brighter where our sensitivity to flicker will be reduced.

## 7. Conclusion

Playback speeds of 120 fps with a vibrating display are too short for high quality (no flicker) renderings of bright images. In this paper I have proposed a method to compensate for flicker based on the relationship between the critical flicker frequency and the Luminance of each pixel. This method consists of blending between a static blurred version and the dynamic full resolution optimization. This novel method results in significant improvement to image quality for those scenes that exhibit small to medium bright regions. It also significantly reduces the amount of flicker in those image types. In the future, some avenues for improvement would be to potentially provide an adaptive solution based off of age or for a single eye. It would also be advantageous to account not only for the per-pixel adaptation, but also for the overall adaptation of the scene. This way, the algorithm could properly predict the amount of compensation for generally bright or dark scenes.

## References

- [1] W. Allen and R. Ulichney. Invited paper: Wobulation: Doubling the addressed resolution of projection displays. *SID Symposium Digest of Technical Papers*, 36, 2005.
- [2] F. Berthouzoz and R. Fattal. Resolution enhancement by vibrating displays. *ACM*, 2011.
- [3] G. et al. Super-resolution optics for virtual reality. *SPIE*, 10355, 2017.
- [4] G. Haas. Microdisplays for augmented and virtual reality. *SID*, 49(1):506–509, 2018.
- [5] D. P. Lee, H. Real-time apparent resolution enhancement for head-mounted displays. *ACM*, 2018.
- [6] B. Lindbloom. Rgb/xyz matrices. <http://www.brucelindbloom.com>.
- [7] L. Martin. Critical duration, the differential luminance threshold, critical flicker frequency, and visual adaptation: A theoretical treatment. *Journal of The Optical Society of America*, 1966.
- [8] I. T. Union. Bt.2100 : Image parameter values for high dynamic range television for use in production and international programme exchange, 2018.
- [9] Y. J. Watson, A. A unified formula for light-adapted pupil size. *Journal of Vision*, 2012.