Digital Film Part 2: Good Grays and Continuous Colors

Thor Olson
Management Graphics Inc.
Minneapolis MN

Presented at the
134th SMPTE Technical Conference
November 10-13, 1992
Metro Toronto Convention Center, Toronto, Canada
Preprint no 134-105

Abstract

The digital dangers related to making artifact-free smooth shaded images on film output are discussed. Principles and data from the visual science field are applied to determine the perceptual limits of smoothness. A method is devised to determine the number of steps needed to implement a perceptually smooth ramp from black to white. Further, the number of bits of precision needed to avoid excessive roundoff error is determined. A 1024 step, 12-bit scale is shown to successfully implement a perceptually smooth ramp for most popular scales except for linear-luminance devices (or data). The device color look up table, which implements the scale, can also be used for matching tonal responses between source and output images and to compensate for changes in contrast range in a perceptually acceptable way. Establishing the performance requirements to make good grays and smooth colors will help reach our goal of artifact-free digital film.

1.0 Introduction: motivations

In the spirit of part 1 of this presentation (Digital Film: Hiding the Raster, 133rd SMPTE Technical Conference) we continue to explore what is needed to produce convincing digital imagery on film and avoid the artifacts that tell a picture's computer origins. This paper will discuss more digital dangers, in particular those related to making smooth shaded objects (both color and black and white) with good tonal range.

This is a particularly important topic at this time because the motion picture industry is starting the transition toward fully digital compositing. Equipment is being designed, algorithms developed, systems put together, all with the purpose of making digital image manipulation, and in particular film compositing, a practical and economically feasible process. This makes it more important than ever to understand the technical requirements of the job.

A common requirement of compositing today is to merge synthetic objects into real scenes. Yet it is extremely difficult to do this and end up with a believable shot. Even suspending disbelief at whatever amazing object it is, or what it's doing, there remains the nagging feeling that the lighting wasn't right, or someone screwed up on the matte. Something, we're not really sure what, isn't consistent between the background and the object that was spliced in.

Well, one major thing that causes this nagging feeling is a mismatch between the tonal range in the background and that of the object. This could be the overall contrast range between black and white, or the distribution of the tones between. A truly obvious blunder happens when objects photographed under studio lights are cut and pasted into outdoor scenes.

The synthetic objects involved are often the product of a software rendering program. The modeling assumptions made by the program, and the resolutions and word sizes used, all impact the appearance of that object. If the tonal range used is highly mismatched from that of the output device, one can observe contouring artifacts, especially in the shadow details where it seems that much of this subliminal information is obtained.

The situation gets still worse when any amount of image processing is applied. Computation on image data results in roundoff errors. These will creep into the picture, especially if the pixel color data is represented with too few bits.

The problems are not restricted to synthetic images. Scanned images are also prone to tonal errors. The response of the scanner needs to be matched to the response of the film recorder, or the same sort of artifacts can result. Worse, if the output film has a different dynamic range than the source medium, matching the gammas will *not* result in duplication of the image. Further compensation is required.

Having established the motivation, we will explore the requirements of a continuous tone digital picture. Some of the difficulties outlined above can be solved by proper control of the film recorder output device, in particular the loading of its color look-up table (LUT).

2.0 Some color concepts: on the way to device independence

The perception of color is a three dimensional phenomenon, and full-color operations must be done using vectors in 3-space. Devices can be fully characterized for their color, though it is a difficult task, requiring that the entire 3-dimensional color space of the device be sampled and calibrated. Once it is done though, you have the information that describes the *gamut* of the device, that is, the entire range of colors that can possibly be generated.

Knowing the color characteristics of a device is a prerequisite to reaching *device-independent color*. The other information needed is the color space in which the digital image resides, and the transformation that will map it appropriately into the color space of the device. This is an active area of research in computer graphics today as part of the challenge of making realistic pictures, or at least the "best" picture possible, on any output device.

Making the best possible picture involves what used to be called color correction, though this term is becoming less meaningful as the reasons for needing color correction become better understood. Color correction embodies any changes made to the color content of an image in order to result in a more pleasing picture. This will consist of a simple (read "well understood") 3-D transformation to the color space of the output device. This brings most of the colors in the image into their proper place in the device gamut, but there are almost always some colors that get mapped outside of the range that can be properly reproduced by the device. Here is where the science stops and artistic judgement begins. Various strategies on how to handle these colors, and their relationship to the in-gamut colors need to be devised.

There is another purpose for color correction beyond the technical function just outlined: editorial color changes. There is a fundamental human need to change things, including the colors found in Nature. The artistic desires of the person making the picture need to be accommodated. This is hardly something that can be analyzed and codified; beauty is contained in the eye of the beholder.

Nevertheless, we can make mechanisms that meet these needs, and some work has been done toward the general purpose color corrector. It is basically a large 3D look-up table, loaded with the resultant of the required transform and any additional desired warpings of the color space. Implementing such a device, and managing it properly, will provide a big step toward device independent color.

3.0 Our (smaller) problem: how to make smooth shades

Our immediate problem is far less ambitious than the goal of device independent color. We just want to be able to render the subtle shadings in continuous tone digital images without incurring artifacts, especially the contouring effects that result when there are two few colors available, or not enough digital precision to represent those colors.

So what is needed to do smooth shading? Consider a ramp ranging from black to white. In a digital representation of that ramp, we are forced to represent it in a finite number of steps. If there are too few, we will see the individual steps. We need to find the minimum number so that *perceptually*, the ramp appears to be continuous. So the problem now becomes finding the perceptual limits of "smoothness". We need to prescribe the maximum step size such that two adjacent shades are digitally different, but perceptually indistinguishable.

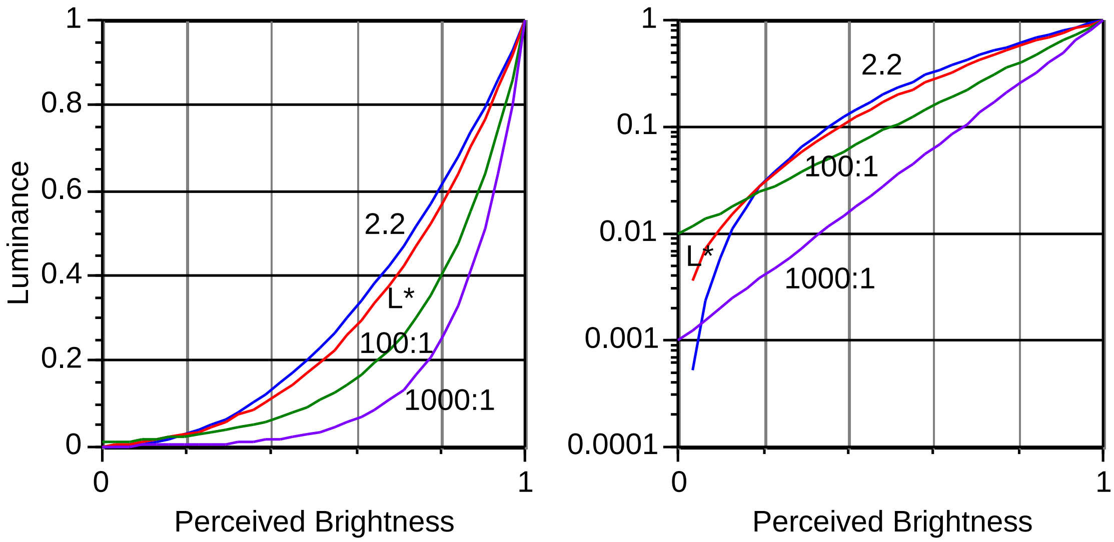

Figure 1: Luminance of some candidate uniform brightness scales.

3.1 What are the perceptual limits of "smoothness"?

There has been more than 100 years of work in the visual science field that addresses this issue. Weber's Law states that the "just noticeable difference" between two adjacent samples is proportional to their luminance. Different investigators have obtained varying data, but their results indicate the discrimination level is around 1%. We will use this number, though it is our experience that in a noise-free setting, the threshold is actually about half of that. Using 1% as the difference between steps, we can start at the white end of the scale and work down towards black until we reach the blackest level attainable on the specific output device. The number of steps needed to reach black will depend on the contrast range available.

The number of steps required is the solution to:

$$0.99^{N} = \frac{1}{C}$$

$$N(C) = \frac{-\log(C)}{\log(0.99)}; \quad N(100:1) = 458; \quad N(1000:1) = 687$$

The gray scale that results is an exponential scale. An excellent discussion of this scale is in [DEM91]. In it, the author points out the impact of dynamic range on the choice of scales, an issue that will be elaborated upon later in this paper.

There has been other visual research, aimed at establishing a perceptually uniform scale. The purpose of the scale is to relate the measurable quantity, *luminance*, to a perceptual quantity: *brightness*. Such a scale would seem to be composed of equally discriminable steps, and indeed some of the candidates are exponential as suggested above. Others however, are power laws, using values of gamma ranging from 2 to 3 depending on background and other environmental factors. The CIE uniform perceptual lightness scale, L*, uses a power of three over most of its range. Figure 1 plots some of the various relationships, including the exponential scales for contrast ranges of 100:1 and 1000:1, representative of video and film.

It is unlikely that there is a "true" scale that embodies the information we seek. Rather, it is likely that different scales are correct for various individuals at different times and conditions. We can take these curves and use them to establish the discriminable step size over the range from black to white. This is a "differential discrimination" function, obtained from the derivative of the uniform perceptual scales.

$$d(B) = \frac{1}{L}\frac{dL}{dB}$$

The discrimination function prescribes the *relative* amount of luminance increase associated with an incremental amount of brightness. These functions are plotted in figure 2.

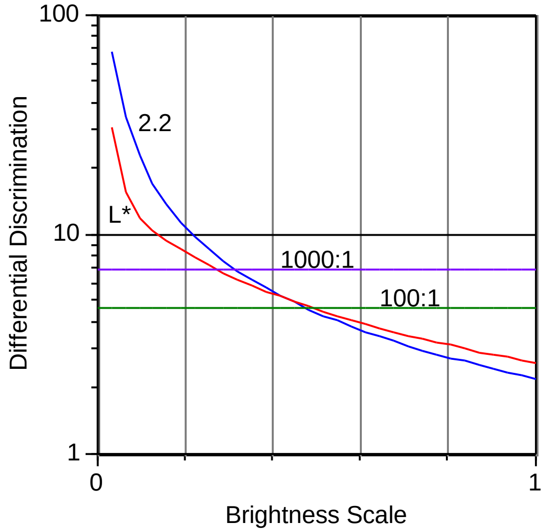

Figure 2: Differential discrimination function of the candidate uniform brightness scales, plotted against brightness.

Since each uniform perceptual scale relates brightness to luminance, each scale's discrimination function can be plotted as d(L), against luminance instead of brightness. This is done in figure 3. We see that the exponential scales result in discrimination levels which are constant, a result of their basis in Weber's Law.

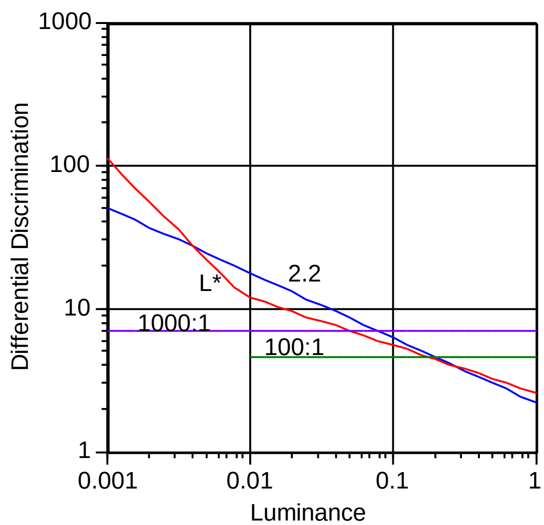

Figure 3: Differential discrimination function of candidate uniform scales plotted against luminance.

Weber's Law is actually an approximation to measurements of the sensitivity of the eye to small luminance differences. Above moderate luminance levels, the sensitivity is fairly constant.

As the luminance decreases however, the ability to discriminate a small change starts to fall off. A plot of the measured ability to detect small luminance differences is shown in figure 4 for a three decade range of photopic vision. This plot is the "step discrimination" function for the human vision. It can also be considered a step "tolerance" function, T(L) since it specifies the maximum step in luminance that would remain undetected.

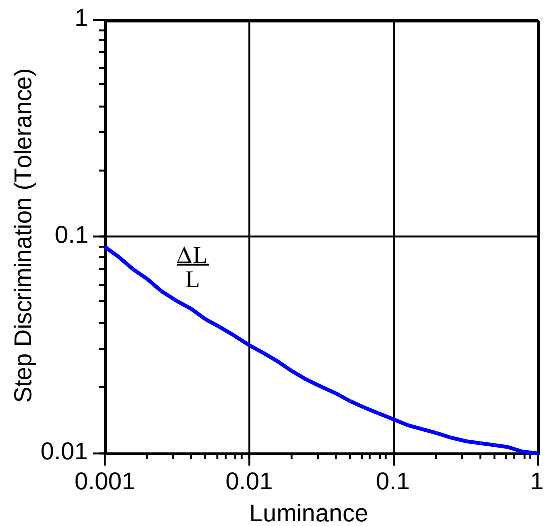

Figure 4: Human visual step discrimination, or tolerance T(L), of small luminance differences

We can conduct an experiment to compare the differential discrimination of a particular gray scale with the step tolerance of human vision. By reducing the number of steps used to represent the black to white ramp, we will eventually detect the individual steps in the ramp. We will know that the scale is perceptually uniform if the steps are equally distinguishable all the way across the ramp. If the steps at one end, say the dark end, appear to merge while those at the white end remain distinct, we can make a statement about how the discrimination of the scale relates to our actual tolerance, namely, that the scale has less discrimination at the dark end than it needs to be detectable (the step between adjacent dark shades is much smaller than the step tolerance amount that can be detected). Similarly, one can state that at the light end, the scale has a higher discrimination level than the step tolerance, because the steps are easily discerned.

To illustrate, figures 5 and 6 (following the text of this paper) show the black to white ramp for two scales, though it is unlikely that they are perfectly preserved in the reproductions here. Even so, it can be seen that the distribution of grays are different between them. Each figure shows five versions of the ramp using an increasing number of steps. The top ramp has 16 steps, which doubles with each successive ramp. There will be a scale, probably near the top, in which the steps are still visible, while the next one down appears smooth. Depending on which scale is being examined, the steps will seem easier to discriminate on one end of the ramp than the other.

Visual experiments on my colleagues using film media with 1000:1 contrast range, suggest that the gamma 2.2 scale is too discriminating at the bright end. In other words, there are more light shades of gray used by this scale than are really needed to achieve smooth ramps. This comes at the expense of the number of dark shades available.

The exponential curve has the opposite problem. For a division of this scale into a discrete number of steps, too many of them are dark compared with the number of available light shades to make a uniformly smooth ramp.

3.2 How many steps are needed? What size are they?

In both of the above cases, the number of steps required to make a smooth ramp from black to white is determined by the relationship of the scale's discrimination function to the visual step tolerance. We first compute a value for ? B, the size of a brightness step that does not exceed the tolerance function. This is found by:

$$T(L) \equiv \frac{\Delta L}{L} = \frac{\frac{dL}{dB}}{L} \Delta B$$

$$\Delta B = \frac{T(L)}{d(L)}$$

This just noticeable brightness step will vary along the scale depending on the differential discrimination of the scale and the visual tolerance function over the range of luminance (figure 7).

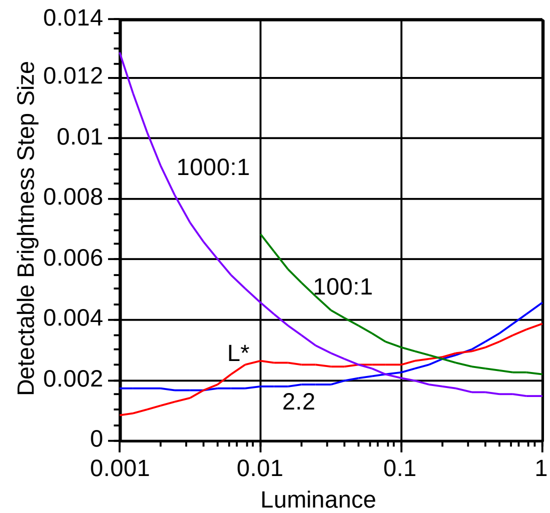

Figure 7: Just noticeable brightness step for four scales as a function of luminance

The reciprocal of ? B is an index of "detectability" along the scale. The units of this function are significant in that they represent the number of steps the scale must have in order to not perceive the jump to the neighboring step at that point in the scale.

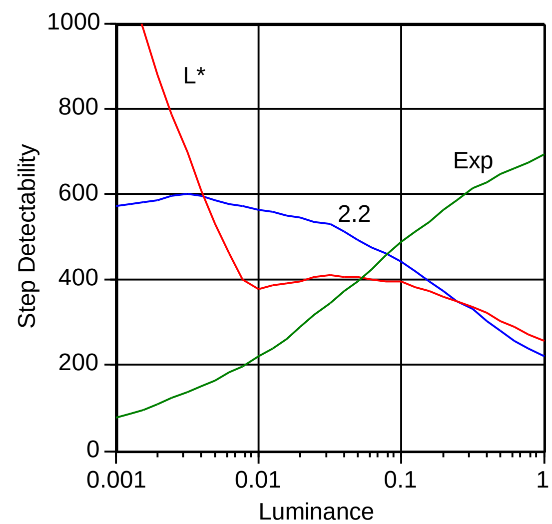

Figure 8: Step "detectability" d(L)/T(L) of three gray scales.

For the exponential scale, the highest detectability occurs at full white with a value of 687. If a black to white ramp is made with this many steps, the steps at the white end will be exactly at the perception threshold. All steps below white will be assured to fall within the step tolerance, because even though the scale's discrimination level was constant, the tolerance for luminance step size grows as one moves from white to black.

The number of steps indicated for the exponential scale matches the number predicted at the beginning of this section using the Weber's Law approximation for the visual step tolerance. It may seem that using this number of steps is very appropriate for making smooth ramps containing light shades, but it is overkill for the darker tonal ranges.

This suggests that a better scale could be devised where the detectable step size was a constant across the scale. This would mean that when the scale were actually divided into discrete steps, that the ability to discriminate between them would be the same everywhere along the scale. Such a scale would be optimal in that a perceptually smooth ramp from black to white could be made in a minimum number of steps.

The exponential scale for 1000:1 dynamic range media such as film has a detectable step size that increases substantially at the dark end, and therefore would be considered "wasteful" of this part of the scale. The 100:1 exponential is better in this regard; in figure 7 it was closer to the ideal flat line we seek, and so would be an OK choice for print media (and indeed, scales with equal density steps are frequently used for prints). But a better choice all around might be one of the other candidate scales.

The other two scales, gamma 2.2 and L*, have detectable step sizes that become smaller at the black end instead of larger. This explains the behavior of the lighter steps merging into a smooth ramp before the darker shades on these scales. The L* scale comes the closest toward the desired flat line, but even this scale has the unfortunate characteristic that it becomes too coarse in the deep shadows.

There is enough here however to hint at where the level might be for the optimal gray scale. It seems that a ? B level of 0.002, just below the flat section of the L* scale in figure 7, could be achieved. Not that the world needs another scale, but for the purposes of testing these concepts, a new scale was derived, based on an approximation to the visual step discrimination data. This new scale will be referred to as B# ("B-sharp") and its derivation is described in the appendix. Its purpose is to provide an approximation to an "equidiscriminant" scale where adjacent steps are equally detectable everywhere along the scale.

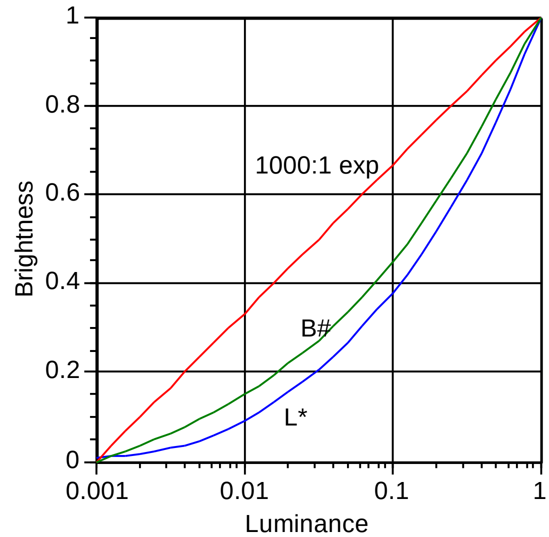

Figure 9: Brightness index of B# compared to two other scales as a function of luminance.

Figure 9 shows the relationship between brightness and luminance of B#. It can be found in the middle ground between the L* scale and the exponential scale. The detectable step size of the B# scale (figure 10) is seen to have the desired flat section without the sudden drop at the dark end. The minimum step size is 0.0025, implying that about 400 steps will successfully generate a perceptually smooth ramp from black to white. Though it may not reproduce properly, figure 11 is the set of gray scales for B# which show it to be a scale which has uniform detectability, the property for which it was designed. It also seems to be a candidate for a uniform brightness scale, in which equal steps are not just equally detectable, but also seem to represent equally spaced shades of gray;

Here then are the answers to some often-asked questions: how many steps are needed to represent the data in digital images, and what size are they? At least 400 steps are needed, and they must be steps in the equidiscriminant B# (or better) scale. Other scales will require more, never fewer, steps to achieve artifact-free smooth shades on digital film. This explains the difficulty in achieving smooth ramps on 8-bit per primary systems. Without even concerning ourselves with the precision of the steps, there are simply not enough of them.

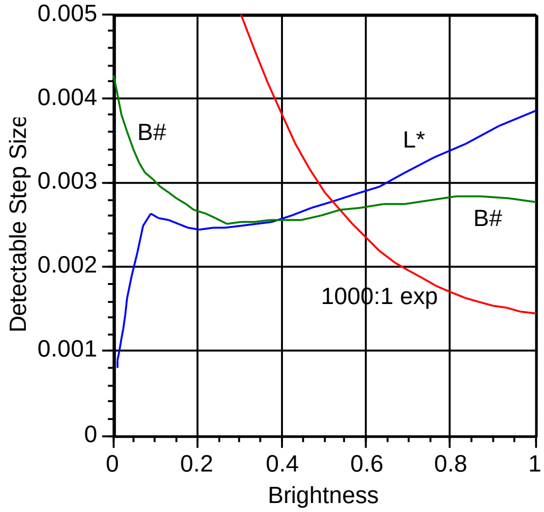

Figure 10: Detectable step size of the B# scale compared to two others over their brightness range.

There are, of course, some disclaimers to be applied to the 400 step number just quoted. First, we have been scrutinizing only the gray axis of the entire color space. If the human visual system is more sensitive to slight changes in a color other than neutral gray, then the step size must be further decreased (and total step count increased) to take this into account. Further, this discussion is based on specific luminance step detection experiments, which may or may not be of the correct configuration to evaluate perceptual steps in nearly smooth ramps. Our experience however shows that this number is very reasonable. Experiments with the gamma 2.2 scale indicate that the number of steps required on 1000:1 media is in excess of 512, entirely consistent with the prediction of 600 in figure 8.

There are some effects that work for us instead of against us however. Our sensitivity to luminance steps also depends on the spatial distribution. If a ramp is very steep, the requirements for small steps are lessened. Motion also reduces the perception of shading artifacts. Image noise is very effective at obscuring the discrete steps in shading. In fact it is so effective that noise is often deliberately added to images to improve the overall perceived quality in devices where a limited number of shades available.

3.3 What precision is needed to represent the steps?

Think of the perceptually smooth scale as being implemented by a table of numbers. The index into the table is the position along the scale. The contents of the table is the luminance level at that step. We now know how many steps (entries in the table) are needed to smoothly go from black to white. We also know what each step needs to be in terms of the luminance levels along the scale. How precisely should these luminance levels be represented? In other words, how accurately must we generate the luminance levels along the scale to avoid a step error that might be visible? Our interest is in finding out the number of bits required to represent the samples of the scale.

For the moment, lets assume that the scale has an infinite number of steps (so the individual luminance steps become insignificantly small), but only a discrete number of luminance levels are available. It's convenient to think of the luminance of a device as being controlled by a voltage (or other such signal) converted from the digital numbers in the scale. The range of the (integer) number sets the number of luminance levels available.

Light generating devices used in film recording usually follow some sort of natural law relationship between voltage and the resulting luminance. We could, for example, have a linear device with L(V) = V. The roundoff error between two successive steps in the brightness scale could be as high as 1 least significant bit (1/2 lsb in each direction). Assuming there are N_b bits used to represent the voltage signal, this error is used to estimate the luminance error:

$$L_{err} \approx \frac{dL}{dV} \Delta V$$

$$\frac{L_{err}}{L} = \frac{1}{L} \frac{dL}{dV} \frac{1}{2^{N_b}}$$

Usually L(V) has an inverse V(L) and it is ok to write:

$$\frac{L_{err}}{L} = \frac{1}{L} \frac{1}{2^{N_b}} \frac{1}{dV/dL}$$

which is the luminance roundoff error expressed in terms of L. Figure 12 shows this error (normalized to N_b=0) for several classes of devices (linear, power law, and $\sin^2$).

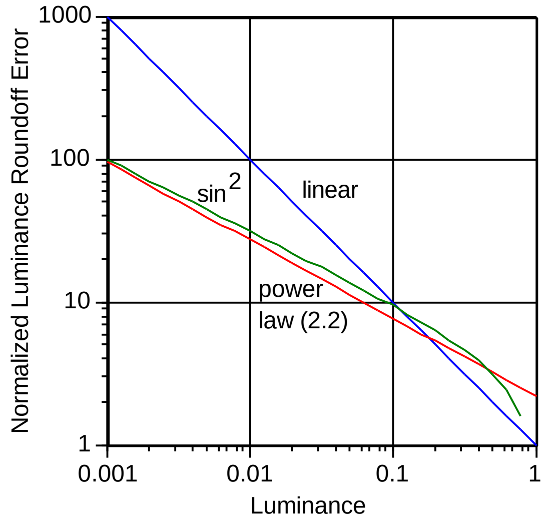

Figure 12: Luminance roundoff error for several device characteristics.

The luminance roundoff error may be directly compared to the human visual step discrimination. It can be seen that at least 10 bits are needed to represent the voltage for power law and $\sin^2$ devices in order to remain under or close to the visual threshold. Figure 13 shows the roundoff error for a 10-bit system.

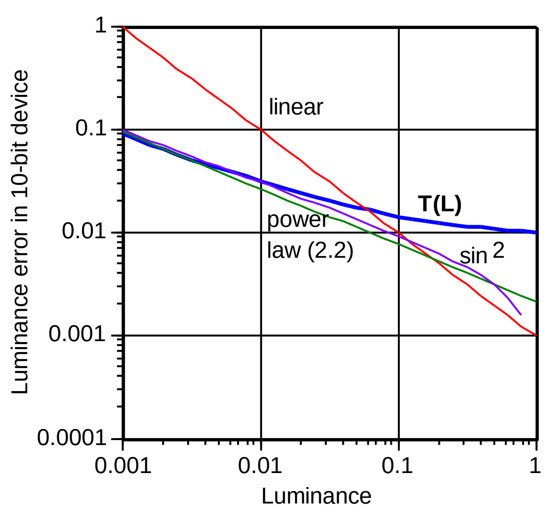

Figure 13: 10-bit device luminance errors compared to the human visual step tolerance, T(L).

The poor guy who has a linear luminance device needs at least 13 bits to avoid contouring artifacts in the shadows! This is one of the reasons that charge coupled device based scanners have a tough time. CCD sensors are intrinsically linear devices, measuring the luminance directly. It is a challenge to operate them over this dynamic range with 13 bits worth of signal to noise ratio.

Recall the step discrimination functions of the previous section. These too can be treated as errors. Figure 14 shows the luminance step sizes for several 512-step scales and compares them to the discrimination threshold.

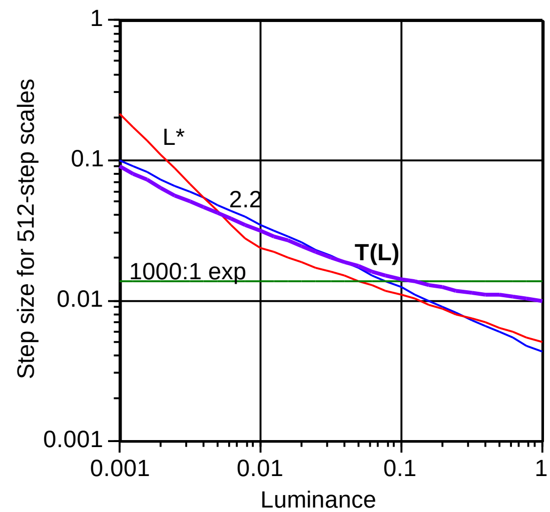

Figure 14: Step sizes for several scales expressed as DL/L, compared to visual step tolerance, T(L).

To summarize, when infinite precision to represent luminance levels exists (no roundoff error) then at least 400 steps are needed in a smooth gray scale (more if the scale is not equidiscriminant). If an infinite number of steps is available, the luminance must be represented with at least 10 bits of precision (unless the device is linear; then 13 bits are needed).

Obviously neither of these arrangements exists in a real digital device. There is both quantization of the luminance levels *and* the number of steps along the scale. This means that errors from both sources will add. Figure 15 shows the sum of the errors from both the scale steps, and the luminance roundoff, for a system with 1024 steps and 12-bit accuracy on a $\sin^2$ device. Comparing to the visual step threshold indicates that any of these gray scales can be successfully implemented with such a system.

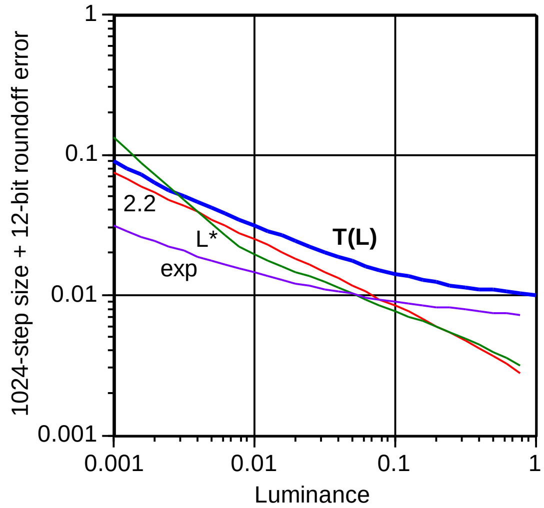

Figure 15: Combined luminance jumps for several 1024-step scales on a 12-bit sin-squared device, compared to visual step tolerance T(L).

4.0 The amazing look up table and what you can do with it

The control of a film recorder's output is accomplished by means of the famous color look-up table (LUT), a hardware translation mechanism that maps the incoming data stream to the appropriate device codes that generate the desired colors. The current technology level for film output, especially for high throughput motion picture work, relies on 1-dimensional corrections, that is, each primary is adjusted independently of the other two color components.

It is through this device that the previously discussed luminance scales can be implemented. We now know how many entries it needs (more than 400) and we know the precision of its contents (at least 10 bits for a CRT or other similar nonlinear device). Because this is such a general purpose device, we can consider using it to solve a number of imaging problems.

Digital images are usually obtained from two common sources: input scanners, and computer synthesis. Motion picture work today merges and intercuts between these sources, printing the composited result to a film recorder. Common artifacts result when the contrast (dynamic) ranges are mismatched, or the tonal curves are mismatched. These variables need to be properly managed if a seamless composite is to result.

The translation between contrast ranges and the matching of gammas can be done using the device look up table. Of course the LUT is also taking care of the internal nonlinearities of the film recorder device (and the film). All of these functions are lumped into the numbers stored in the LUT.

4.1 Using the LUT to match tonal curves

Assume that an original source picture has been scanned and its digital representation saved. In reproducing that image, the tonal scale that the film recorder generates must match that of the scanner or the picture won't be the same as the original. This seems obvious, but a mismatch is an extremely common occurrence. The most common mistake (and also the easiest to identify and correct) is obtaining linear luminance scan data and imaging it on a device with a gamma 2.2 (video) characteristic.

Many people, on entering the world of digital imaging, are unaware of the responses of their scanners and printers. Because the effects of mismatching them are often subtle, unlike the example above, they may be aware that something isn't quite right, but they're not sure what. The output is "washed out" or "too dark". This results in convictions being established that digital imaging technology doesn't work yet. This may be the right conclusion, but for the wrong reason!

Whatever response the scanner has, the printer should match. If the scanner is calibrated to a specific gamma, or other type curve, the film recorder should also be calibrated to that curve. Some systems are set up to do closed-loop calibration: a test target is scanned and the generated film output must match.

What about the case where the image data isn't from a scanner? Many computer rendered scenes are based on physical lighting models, which operate in linear luminances. Their output will be in a linear scale (gamma=1). To make these pictures look right when printed, either the film recorder must be calibrated to a linear luminance scale, or the image data needs to be converted to the gamma of the film recorder.

Unfortunately, device calibration is not always simple. It requires the proper equipment and is often a challenging, time consuming process, prone to mistakes. This disincentive causes programmers to write image conversion utilities to change image data currently in one scale to that in another. This is a good thing to do; it corrects the output, but it's usually a time consuming operation that must handle each and every pixel. With tight production schedules it would be nicer to avoid this delay. The device look up table offers a way to do this by placing the conversion in hardware, executed in the device.

A useful way of doing this is based on modifying the built-in LUT in a way that does not require knowledge of its exact contents. If the overall response function of the film recorder is known, the factory-set LUT can be uploaded and resampled. This has the benefit of not needing to know the internal device characteristics, only the total response from data to film. The LUT entries are redistributed according to the desired response. The resampled LUT is then downloaded back to the film recorder. Appendix 2 contains a sample conversion between one gamma response curve to another.

4.2 Moving between media of different dynamic range

The closed loop calibration between scanner and printer works well for systems where the output medium is the same as the input, but what about when they aren't the same? Scanning film positives and generating negatives, or scanning reflective prints and printing transparencies are good examples. In these cases we can't directly compare input to output. Some mapping of the luminance between these media is required.

Consider the process of printing a negative. The negative holds a "density-compressed" version of the image which is then expanded upon printing. For example, a density range of 1.5 say, on the negative ends up as a range of 2.0 on the print. The print paper has a gamma of 1.33 (2.0/1.5) to accomplish this. The function that this printing operation performs is to raise the luminance through the negative to the -1.33 power. This suggests that the appropriate mapping in this situation is to raise the luminance to the power of the density range (log contrast) ratios:

$$L_2 = L_1^{-(D_2/D_1)} = L_1^{-(\log C_2/\log C_1)}$$

In practice, all photographic media have an S-shaped response instead of the idealized straight-line case above. When calibrating negatives, the mapping between negative density and final print density needs to be taken into account. This is usually measured, rather than analytically modelled because, due to the printing process, it is not a simple or even unique relation. This mapping ends up being embodied by the data in the look up table, "folded in" along with the device control and desired scale being implemented.

Now consider what happens when you scan a reflective print having a contrast of 100:1 and then output the data to transparency film which has a contrast of 1000:1. One might expect this to be a benign operation, since the original print luminance range is uniformly spread out over the larger range of the transparency, but the result is an image with overemphasized shadows.

We can see why this is from the plots of figure 17. The luminance curve of the 100:1 medium matches the 1000:1 medium curve having the same gamma at most luminance levels, but obviously must fall below it at the dark regions. This is a discrepancy between input and output response. We could solve it by only using part of the dynamic range of the output media, but apart from wasting the available contrast, it wouldn't look right either (even though the luminance levels are identical, our perceptual system wouldn't buy it).

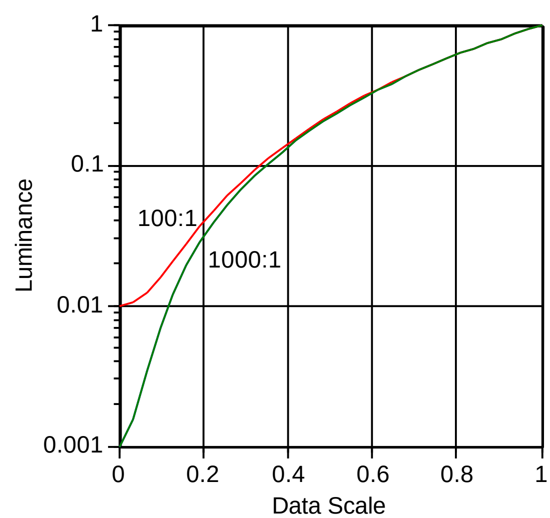

Figure 17: Luminances of the same (gamma 2.2) scale on different contrast media.

It is not clear that there is any "right" way to solve this. In effect we are adding information to the picture and we wish to do so in a perceptually acceptable way. We experimented with several scalings of the image data. Our initial approach was to map the source medium to a perceptually uniform scale, linearly expand (or shrink) the result, and then map it back to the destination response. Unfortunately, any expansion at all results in an apparent overall darkening of the picture. Worse, the contrast is increased in the highlights but reduced in the shadows, exactly the opposite of what is desired. Figure 18 illustrates the result.

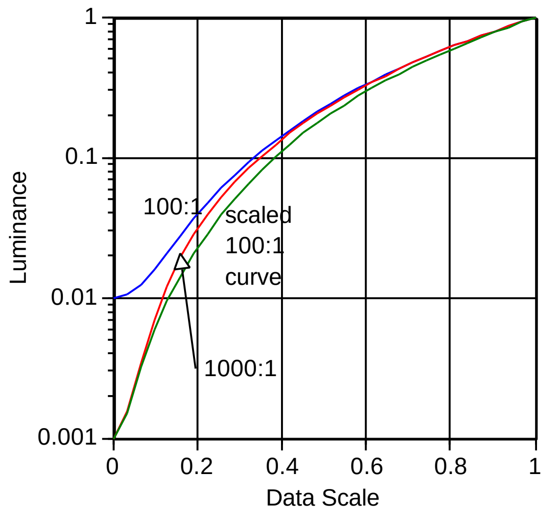

Figure 19: The effect of simple scaling of a low contrast response to a higher one.

The effect of this scaling is similar to changing the gamma of the curve to a higher value. The initial gamma was 2.2. It changed to 2.4, as measured in the region near white. This suggests that shifting gamma in the opposite direction while expanding the luminance range might be an improvement. The result is called gamma scaling. It results in a nice compromise in the shadow region where the errors are largest at the expense of a small shift over the remaining part of the curve (figure 20).

The amount that gamma is shifted is the ratio between the two *brightness* ranges (not luminance ranges). Our experiments used the L* brightness scale, which prescribed:

$$\gamma_2 = \frac{100 - L^*_{1min}}{100 - L^*_{2min}} \gamma_1 = \frac{91}{99.1} \cdot 2.2 = 2.02$$

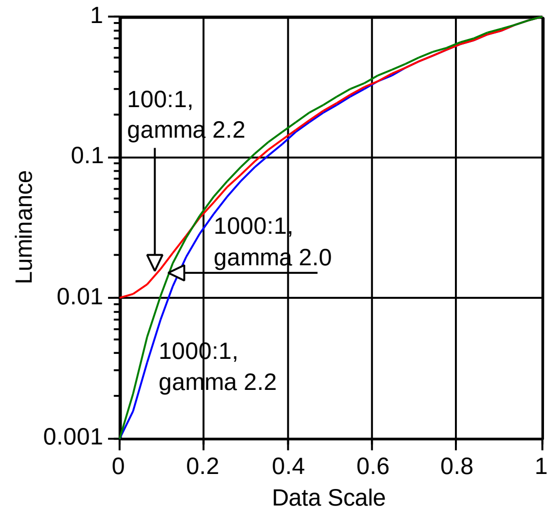

Figure 20: Scaling gamma between different contrast media.

While this empirical approach yielded a satisfying and simple result, I am sure there are even better choices and the issues related to imaging across different dynamic ranges deserves more study.

5.0 Conclusions and Summary

We have made but a start in digital imaging and are already at a point where the fundamental hardware and software components and tools of the business are being established. The rapid pace of this development is driven by competitive forces, but is sometimes hampered by the lack of available imaging science to help make informed engineering decisions and tradeoffs. This paper is an attempt to fill a small portion of that need.

There are a number of results that can help in the generation of high quality, artifact free digital film output. By applying principles of psychophysics and sensory measurements made over the last century, we have learned that no fewer than 400 steps are needed to smoothly traverse the distance from black to white (on 1000:1 contrast medium). The choice of tonal scale will impact the number of steps required, always making the number larger than this. Fortunately, most popular scales can be implemented in a perceptually smooth fashion within 1024 steps. This explains the frequent disappointment that results when a 256 step ramp which looked perfectly good on the (100:1) video monitor, results in a ragged appearance when output to film.

Of course, the 400 step lower bound only applies if the internal control of the exposing device has perfect precision. If a zillion-step scale could be used, we find that at least 10 bits of precision is needed or we will see the coarseness in the ramp. A suitable balance between these two requirements will be found in the vicinity of 1024 steps of 12-bit precision. An additional 3-bits of precision is required in the case of a device (or image data) which has a linear luminance response.

Making smooth-shaded ramps is only one of the problems for digital compositing. The matching of tonal scale between source images and output devices is also on the list. As are the artifacts introduced when transfering images across media having different contrast ranges. The device look up table can help us here by allowing us to calibrate the system between input and output. It can also be used for compensation and tonal correction providing hardware-speed translation between scales.

There are plenty of other challenges in making good grays and continuous colors on digital film. Someday, full 3D color gamut mapping will be an available tool. For now however, the single axis color LUT, when properly loaded, can take us a long way.

Appendix 1: Derivation of the B# Scale

This gray scale is based on an approximation to the human visual luminance step discrimination measurements shown in figure A1-1 along with the approximating function:

$$T(L) \equiv \frac{\Delta L}{L} \approx W_2 \ln^2(L) + W_0$$

$$T(L) \approx 0.001 \ln^2(L) + 0.01$$

Where L is the normalized luminance, having a value of unity at the white point. It may not seem that the approximating function is terribly good at its job, but it has some important redeeming characteristics. First, it has a minimum at the white point. This is where the eye will be most sensitive. The minimum at this point implies that the slope of the function will be zero here as well, a characteristic that generates Weber Law-like behavior at this end of the scale. The actual value of the function at white is set to W_0, the Weber level of luminance step discrimination. Other data may imply a different number should be used here. We like the value 0.01.

The behavior of the function toward the lower luminance levels indicates a lower sensitivity to step changes (higher tolerance for them). The function increases, but not as much as the actual data suggests it could. This allows the function to be useful over a wide range of actual luminance levels and even with higher contrast media, successfully modeling the tolerance change but not overestimating it. This avoids the sudden drop in detectable step size found at the dark end of the L* scale. The constant W_2 sets the shape of the curve. We chose W_2 to make the curves equal at about the 1% white level, thinking this may be beneficial for use with 100:1 contrast media as well.

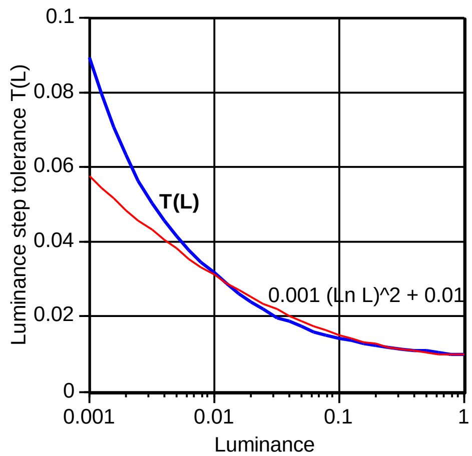

Figure A1-1: Luminance step threshold data and approximating function.

Finally, the form of the function allows us to analytically, rather than numerically, perform the next steps, resulting in a useful formula for B# when all is done. We start by observing that ΔL can be written using the differential discrimination of the brightness scale:

$$\Delta L = \frac{dL}{dB}\Delta B$$

We substitute this into the step discrimination function T(L) and solve for dB:

$$dB = \Delta B \frac{1}{L \cdot T(L)} dL = \frac{\Delta B}{W_2} \cdot \frac{1}{L \cdot \left[\ln^2(L) + W_0/W_2 \right]} dL$$

Integrate to solve for B(L):

$$B = \int dB = \frac{\Delta B}{W_2} \int \frac{1}{L \cdot \left[\ln^2(L) + W_0/W_2 \right]} dL = \frac{\Delta B}{\sqrt{W_0 W_2}} \left[\arctan\left(\frac{\ln L}{\sqrt{W_0/W_2}} \right) + C \right]$$

The integration constant C is obtained by evaluating the arctan function for the luminance of the black level and setting C to exactly cancel it, so the brightness becomes zero here. We can now determine ΔB, the size of the just-noticeable brightness step assumed by the scale. When L is at full scale (one), B is also equal to one, and:

$$\Delta B = \frac{\sqrt{W_0 W_2}}{C} = \frac{-\sqrt{W_0 W_2}}{\arctan\left(\frac{\ln L_{min}}{\sqrt{W_0/W_2}} \right)}$$

With our preferred values for W_0 and W_2, we christen the result the B# scale, a possible improvement over L*. Here are the functions for 100:1, 1000:1, and 10,000:1 contrast media respectively:

$$B_2^{\#}(L) = \frac{1}{0.97} \left[\arctan\left(\frac{\ln L}{\sqrt{10}} \right) + 0.97 \right]$$

$$B_3^{\#}(L) = \frac{1}{1.14} \left[\arctan\left(\frac{\ln L}{\sqrt{10}} \right) + 1.14 \right]$$

$$B_4^{\#}(L) = \frac{1}{1.24} \left[\arctan\left(\frac{\ln L}{\sqrt{10}} \right) + 1.24 \right]$$

The minimum number of steps implicit in these functions are: 307, 361, and 392 to span from black to white. It is seen that the number of steps needed is relatively constant. Even for the high contrast medium case, less than 500 steps would assure a perceptually smooth ramp.

The inverse function relates the luminance to brightness:

$$L(B^{\#}) = \exp\left[\sqrt{W_0/W_2} \tan\left(C \cdot B^{\#} - C \right) \right]$$

Appendix 2: Resampling the density distribution of a LUT

This will describe how to make a gamma change for a LUT which has been constructed at a specific gamma. The white point remains the same for both LUTs.

The mathematical relationship prescribed by the standard gamma curve is:

$$t(c) = t_{min} + (t_{max} - t_{min})c^{\gamma}$$

where t(c) is the transmission of the film resulting from an input code of c (c is normalized to range between 0 and 1). t_{min} is the minimum transmission resulting from the code 0, and t_{max} is the maximum transmission, occurring for the full scale input c=1.

To change the gamma (?) of the curve, solve for the old color index, c_1, in terms of the new one, c_2:

$$c_1 = c_2^{(\gamma_2/\gamma_1)}$$

The resampling then consists of indexing through each c_2 of the new LUT curve, computing a c_1 for each from the above formula. A convenient way to compute the c_1 is:

c1 = exp((gamma2 / gamma1) * ln(c2))

Look up the contents of the old LUT curve at index c_1 (interpolation between integer indices will be required), and place it at index c_2 of the new LUT.

References:

[DEM91] Demos, Gary, "The Use of Logarithmic and Density Units for Pixels", *SMPTE Journal*, 100:805-816, 1991.

[HUN87] Hunt, R. W. G., *Measuring Color*, John Wiley and Sons, New York, 1987.

[OLS91] Olson, Thor, "Digital Film: Hiding the Raster", Preprint no. 133-102, *133rd SMPTE Technical Conference*, SMPTE, White Plains NY, October 1991.

[STE61] Stevens, S. S., "To Honor Fechner and Repeal His Law", *Science*, v133:80-86, 1961.

[WYS67] Wyszecki, Gunter, and W.S. Stiles, *Color Science*, John Wiley and Sons, New York, 1967

Additional Figures:

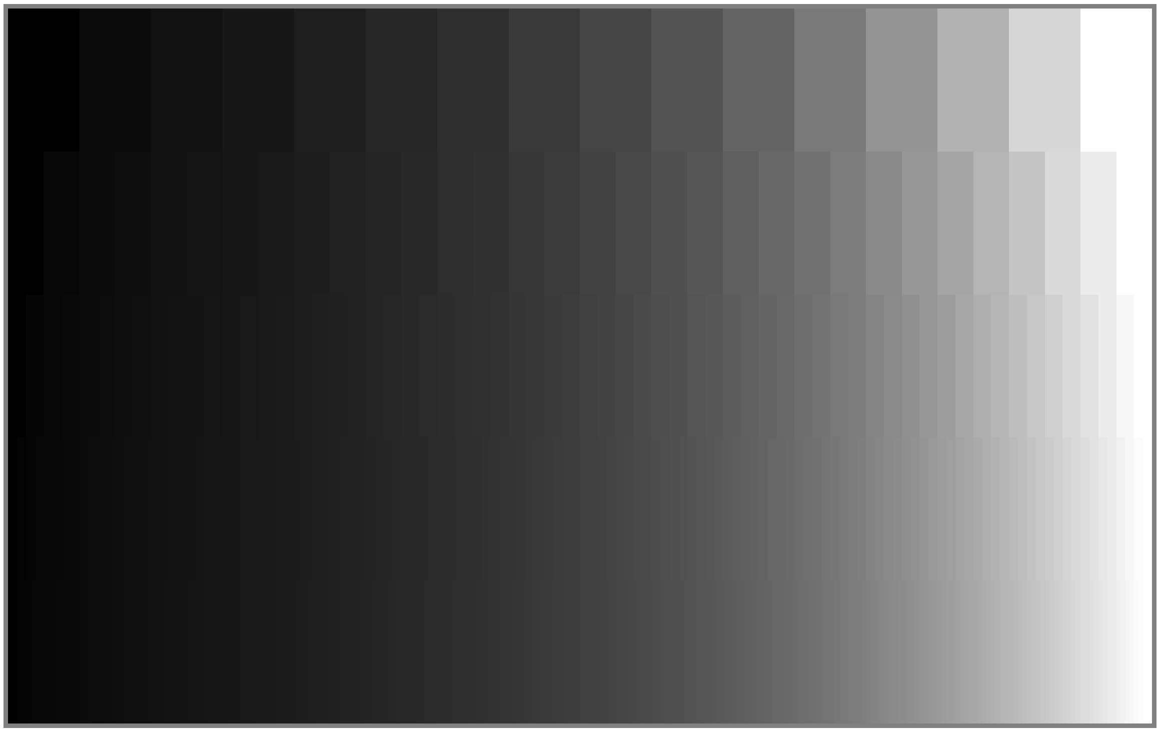

Figure 5: Progressively increasing samples of the linear density scale

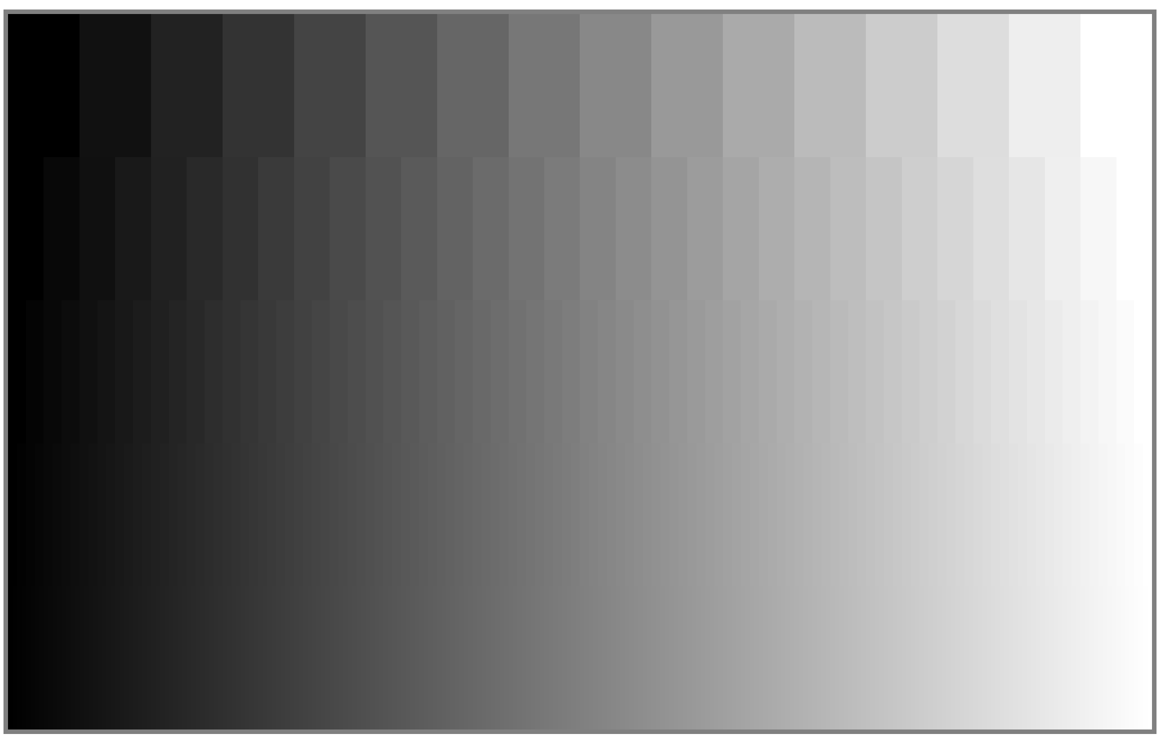

Figure 6: Progressively increasing samples of the gamma 2.2 scale

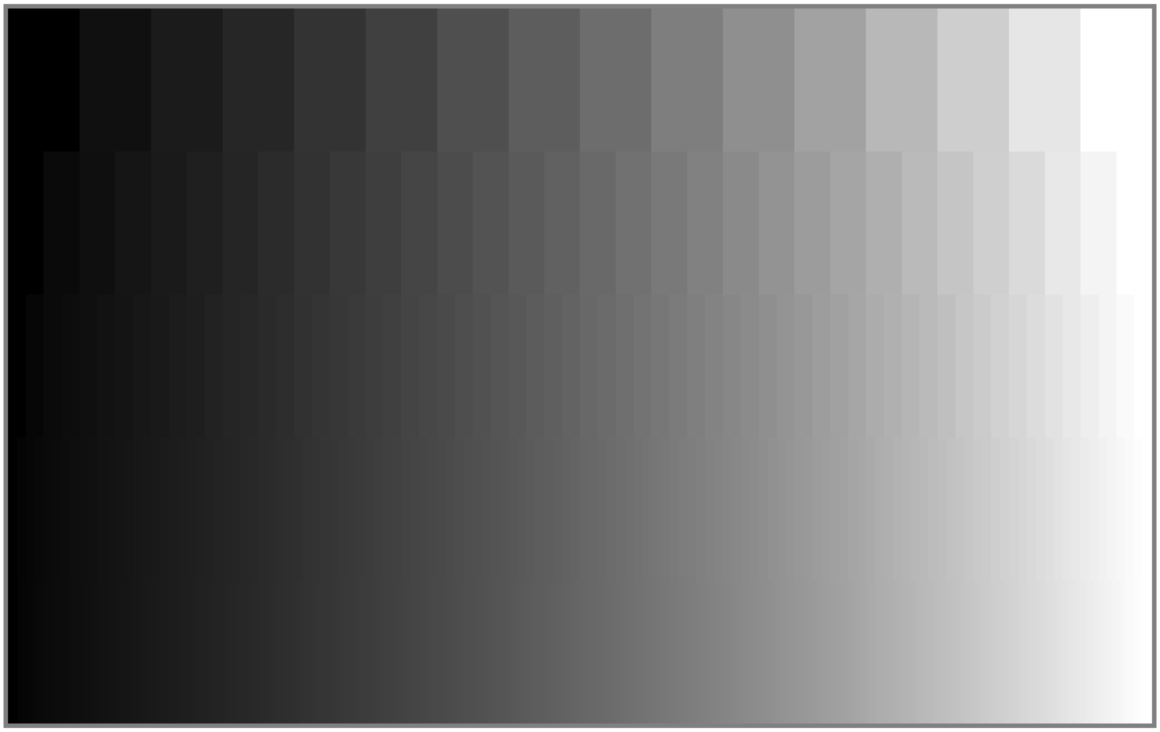
Figure 11: Progressively increasing samples of the B# scale