Gamma (γ)

3D applications have a Gamma option and this control causes a lot of confusion. Knowing what it really is helps the artist to decide when to use it and when better not.

Light

Every real object is defined by its form and colour on its surface and when it is lit, we can see it. This is the same for virtual objects in a 3D computer graphics image: without light, colour, texture and form cannot be seen. We expect that when the output of a light source is half of what it was before, the object will get darker, actually half as bright. How the human eye sees this light does not matter here because it works the same whether we look at nature or on a computer screen (besides, that topic is a bit more complicated than gamma).

Display

What does matter is how the apparatus that displays the image behaves. The first computer screens were CRTs (cathode ray tubes) and when such a device gets the double of power from a pixel in a picture, it cannot display it double as bright over the full brightness range. Dark pixels are displayed brighter than they actually are and very bright pixels less bright. In other words, CRTs cannot display light in a linear fashion. This nonlinearity varied between the different brands of CRTs.
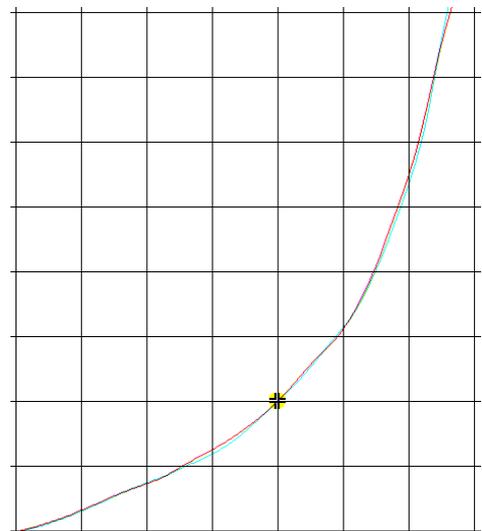
Pictures

The missing linearity was not a big deal when there were only 16 colours. With "True-Color" it became an issue. The pictures that contain the three colours red, green and blue (RGB) with 256 brightness shades each can contain about 16 Million colours, hence "True Color".

To compensate for the nonlinearity of CRTs, the pixels in the pictures were subjected to the inverse nonlinearity and when displayed, they showed right. Computer manufacturer chose theirs own compensation. Macintosh computers saved the PICT files with a pre-gamma of 1.8, PCs saved the BMPs with 2.2; other manufacturers used 1.6. A converted picture from either a Macintosh or a PC seen on the other computer had either too much contrast, being a bit dark; or was with washed-out contrast and too bright.

Meanwhile, pre-gamma for pictures is standardised, flat screens display pixel values linear and the graphics card introduces the gamma compensation. This is true for all 8/24 bit pictures file formats (JPG, BMP, TIFF, PNG, etc.) that have 8 bit (256 values) for each of the three primary colours red, green and blue.

At right is a camera curve from a modern FX (full frame) digital camera. The curve is not linear, it is bent. It was plotted from several RAW 14/42 bit exposures made into an HDRI (high dynamic range) 32/96 bit picture. Pixel values are on the abscissa (horizontal) and the exposure on the ordinate (vertical). There are no values, they do not matter here. The curve looks close to the standard pre-gamma of 2.2 and we can assume it is.

The dot represents the cross-over point where pixel value and exposition match.

Gamma

So what does gamma do to the pixel values? In a conventional picture file, every pixel has an integer value in the range of 0 to 255. The gamma calculation needs these values in the range of 0.000 to 1.000 as real or floating point values, which is done by dividing each pixel value by 255.

Each pixel value is now raised to the power of gamma, i.e. 2.2: gamma value = $px^{2.2}$. This is what each graphics program does when saving the picture and how every camera exports the photograph. This is the red curve at right.

The CRT display or the graphics card in your computer that drives the flat screen introduces the inverse gamma. Each pixel value is raised to the power of one divided by gamma and the result multiplied by 255; $px^{1/2.2}$ or $px^{0.454545}$ (green curve above). The result you see on the screen is linear (blue straight line) and that is what you expect.
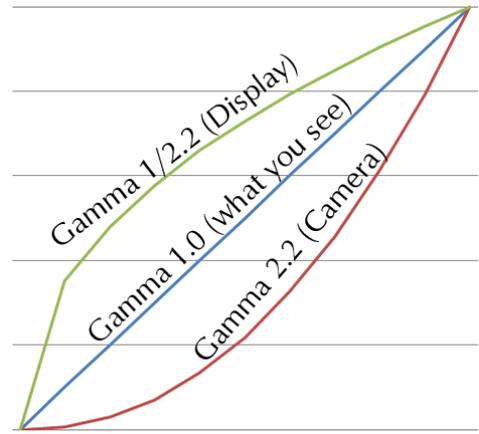
Pictures with more than 8/24 Bit

The above applies to all conventional pictures you get from your camera or computer graphics or 3D program. There are also pictures that have 16/48 or 32/96 bit per pixel and these are different and cannot be directly displayed on a computer screen because their brightness range and colour resolution exceed the capabilities of the displaying device.

There are two 16/48 bit variants. One (TIFF) uses unsigned integer values in the range of 0 to 65,535 or signed -32768 to +32767 and they may, or may not, have gamma applied, mainly not. The other one uses half precision floating point reals, like the OpenEXR file format. This is a high dynamic range variant and the pixel data are stored in a linear fashion without any gamma.

The 32/96 bit formats store the pixel data as single precision floating point reals without any gamma. TIFF and PFM are HDRI formats. The well-known RGBE radiance format is an HDRI format without gamma as well. It also stores the pixel values as single precision reals but uses only the mantissa for red, green and blue and stores the highest exponent of the three colours only and thus compresses 96 bit into 32 (hence RGBE: red, green, blue, exponent).

Linear Workflow

Any industry standard 3D application submits to a linear workflow. This means no gamma. How can this work with the graphics card that subjects every picture to an inverse gamma for display? Display and render engine are two different things that work independent.

Lights

All lights behave linear like in the real world. Doubling light diffuse exactly doubles the light output. No matter what light source you use, increasing its diffuse must increase the light output by the same ratio.

An HDRI contains the lights linear and can be considered as just another light source. If a pixel value is half of another, its light output will also be half. The light adjustment of the HDRI must not be confused with the brightness of the backdrop. To adjust the global light emitted from the HDRI, the control must dim the light output to half if its value is halved, or boost accordingly when doubled.

Materials / Shaders

If a picture is used as a shader or material on an object, it comes into the program with gamma and the display compensates for that. You see it correctly. No need to apply or remove gamma.

The render engine strips the gamma from the picture texture and calculates how bright it gets under the light it is illuminated with. Light is linear, the picture texture is linear, the result is linear. Once the scene is rendered, the result is subjected to gamma before it is sent to the display and you behold it correct. If the rendered result is saved to an image file, it is automatically subjected to gamma so that it displays correctly in any graphics program.

Image Backdrop

A scene can be set up completely with objects covering everything the camera can see, or even populate the scene around the camera. A photographic background can be added to give the scene some additional grandeur. The photograph or image used for the backdrop is handled by the 3D program the same way as a photographic shader. After all, it is just another picture material on a flat surface.

A better option is to use a panoramic photograph as background, either cylindrical or spherical if objects in the scene have reflecting properties. The camera needs to be inside the cylinder or sphere. As far as gamma is concerned, the same applies as for photographs on a flat surface.

Using an HDRI panorama as backdrop has its advantages but is also a bit tricky to use. The HDRI provides light and specular and both properties should be adjustable separately. However, the visible backdrop usually does not look good when the brightness range is squeezed to fit the display. The background needs a tone-mapping means to look good. If background, light and specular is subjected to a common control, only one aspect can be adjusted as desired, the other two will be degraded. The simplest tone-mapping means is to apply additional gamma only to the backdrop. If it is applied to the light, it is not linear anymore, if it is applied to specular, gloss fades.

Export Render as HDRI

Some 3D programs can export the resulting render as a high dynamic range image (96 bit TIFF, RGBE, OpenExr) and these pictures will contain the pixel values exactly as the renderer calculated them: linear without gamma. Such pictures cannot be correctly displayed.

Saving or exporting a render as HDR image has advantages. Render engines use more bits per pixel when calculating the image than what you finally get when saving as conventional picture. The added precision in colour and brightness is not lost. If you want to post process the picture, you have more options and image data for adjustments before you reduce it to a conventional 8/24 bit picture. You need an HDRI capable graphics program to do so.

Using Gamma

A 3D computer graphics program is a tool and the more options and controls there are the more freedom the artist has to create his or her artwork; the catch may be that it becomes bewildering.

Photographs, pictures, cameras, display units and render engines all care for gamma and therefore gamma should be none of your concerns. There is no need for a gamma control and if there is one it has to be set to 1.0 by default and changes made consciously.

Attempting to create a photorealistic scene using even an unbiased render engine and engage gamma will utterly fail. There is no gamma in nature. Some 3D applications have some post production options that are directly applied to the rendered image and gamma is one of them. Some also have a separate gamma control for the picture shaders.

Gamma can be applied anytime later in a graphics application. Most feature a histogram and here gamma can be adjusted as shown in the example at right. A more sophisticated means is to export the render as an HDRI and then use a tone-mapping operator on it that has more controls.
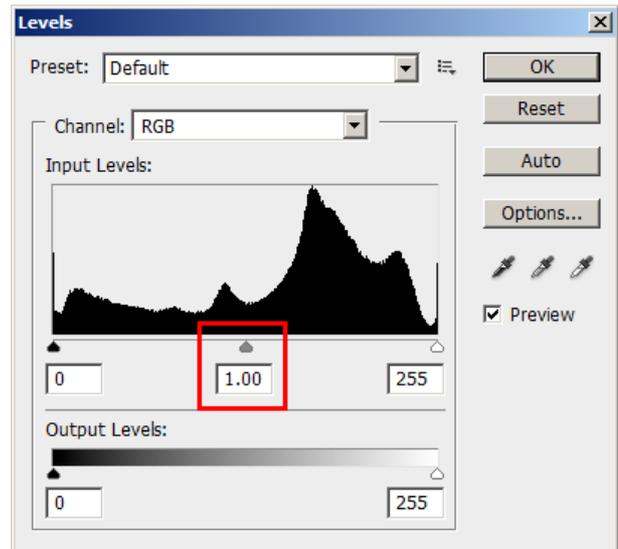
Conclusion

If the artist is convinced that the artwork he or she set out to create succeeded, it is impeccable. The artist is always right, no matter how the artwork was accomplished and what others may think — except the customer if it was a commissioned work, of course.

That said, be aware that the electronic tools you use care about gamma and that there is no such thing as gamma in the real world. If you go for photorealism, the gamma control has to be nailed at 1.0. You do not need to use an unbiased render engine if you plan to spoil the result by applying gamma anyway. The graph at right shows what you get if you add gamma 2.2 to an already applied gamma 2.2. This is not linear and not natural.

3D art is more than "just" photorealism. For other types of artwork, there may be good reasons to use gamma to create special effects. Applying gamma is not evil; it just has no place in photorealism.

Also remember that gamma goes both ways. If it is set above 1.0, things get low contrast because dark parts get proportionally brighter. If it is set below 1.0, contrast appears increased because dark parts get even darker.

I hope I could demystify gamma.