VIDEO PLAYBACK VISUAL PERFORMANCE EVALUATION

Overview: This document provides an overview of the visual playback characteristics which define the end user's experience. This document is intended for individuals unfamiliar with evaluating visual performance.

1. Introduction

There are four major characteristics which define the end user's visual experience: frame rate, temporal linearity, frame quality and lip synch (audio/video lip synchronization). Frame rate is the visual metric most commonly used when discussing video playback performance. It is common for viewers who are not familiar with evaluating visual quality to mistake degraded temporal linearity or frame quality for degraded frame rate. The purpose of this document is to enable accurate visual evaluation by delineating the four major visual characteristics.

2. Frame Rate

Frame rate is the number of video frames per second which are displayed to the viewer. Movie content is typically mastered at 24 frames per second (fps). NTSC (i.e. U.S. television content) is mastered at 60 fields per second, which is equivalent to 30 fps. Playback at full frame rate results in smooth motion. If the frame rate is degraded, lower than 24 fps for movie content or 30 fps for NTSC content, motion is perceived as jerky and unnatural. Note that degraded frame rate does not result in a slow motion effect. Movement is perceived as being at a normal rate of time even though it appears to jump and lurch.

3. Temporal Linearity

Temporal linearity refers to how evenly the video frames are displayed and has an equal effect on the viewer's perception of the naturalness of motion. Ideally, all of the frames should be displayed with exactly the same amount of time between each frame. For example, if 24 fps movie content is being displayed, each frame should ideally be displayed at 1/24 second intervals. As a rule of thumb, most viewers cannot perceive a motion distortion if the frames are displayed within half of the optimal time interval (i.e. if the optimal time interval between frames is 1/24 second, most users will not perceive jerky motion if the frames are rendered within 1/48 to 1/12 of a second of each other.

4. Visual Quality

Visual quality defines how realistic the picture images are. Visual quality is composed of 3 major attributes.

- Clarity or sharpness of image
- Color
- Consistent image display

4.1 Clarity or Sharpness

Visual clarity or sharpness of image is an important factor when evaluating visual quality. Visual clarity or sharpness is determined by evaluating how closely objects are displayed relative to reality. The following are some common artifacts (flaws in the visual image) found when visual quality is degraded:

Ringing

Ringing is an artifact where an object's boundary appears to shimmer, or there appears to be multiple boundaries to an object. Ringing usually manifests itself at high contrast edges. For example, the border between a sheet of white paper placed on a dark tabletop, or a white shirt's collar against a dark jacket.

Aliasing

Aliasing is an artifact where diagonal lines are displayed jagged rather than smooth. Aliasing is particularly obvious in areas where a dark line is pictured against a light background. For example, an aliasing effect would render the edges of a lamp shade as jagged instread of having a smooth and straight outline.

Mosquitoes

The mosquito effect is an artifact where there are fluctuations around an object's boundary such that it appears that there is a swarm of mosquitoes surrounding the object.

Banding

Banding occurs where smooth regions in the video are marred by bands of pixels (appears as small square blocks) which are grouped together. For example, a plastic ball should have a smooth and rounded surface. A banding effect would render the ball with patches or bands of squares on the ball's surface.

Blocking

Blocking is an artifact where there are discontinuities in the image. Part of the image appears to be fractionally offset (either to the right, left, up or down) from the rest of the image.

4.2 Color

Color generates visual sensations which greatly influence the viewer's visual experience. The perception of color usually involves three quantities, known as *hue, saturation,* and *lightness*. *Hue* distinguishes among colors such as red, green, purple and yellow. *Saturation* refers to how far color is from a gray of equal intensity. Red is highly saturated; pink is relatively unsaturated. Royal blue is highly saturated; sky blue is relatively unsaturated. Pastel colors are relatively unsaturated. Unsaturated colors include more white light than do the vivid, saturated colors. *Lightness* defines a viewer's perception of an object's ability to reflect light. Although displays of hue, saturation and lightness are dependent on the content being displayed, a successful video player should be able to render a full range of color hue, saturation and lightness.

In addition to displaying a full range of hue, saturation and lightness, the visual image should have clear color boundaries. Color bleeding is a visual artifact where one color is smeared from one side of a color boundary to another.

4.3 Consistent Image Display

Consistent image display is a necessary part of a realistic video display. A given image must be rendered the same from frame to frame, assuming no change is intended by the video content. For example, if every fourth frame renders an object differently from the other frames, the object will appear to pulsate or vibrate. A still background object, such as a lamp, will appear to pulsate even though it should be static.

5. Audio/Video Lip Synchronization

Audio/Video Lip Synchronization (Lip Synch) measures how closely the visual display of people speaking matches the audio. Visual images of people speaking should match the sound of their voices. Mismatched lip synch is disorienting to viewers and can be perceived as a degradation of visual quality (i.e. “things don’t look right”).

6. Summary

The four visual characteristics discussed above, frame rate, temporal linearity, visual quality and lip synch, define a viewer’s visual experience. Individual viewers will place different emphasis on different visual characteristics. However, even though viewers’ “gut response” to a visual experience may vary from individual to individual, in order to guarantee the best possible visual experience for the majority of people, all of the visual characteristics mentioned above must be successfully addressed.