REPORT FROM THE OPEN STUDIO ROUNDTABLE

BY CHRIS ALLAIN

As the year draws to a close it's time to wrap-up unfinished business. One such item is a report on VIDEOGRAPHY's seventh Open Studio Roundtable, held on the evening of August 6 during this year's SIGGRAPH conference, in Los Angeles. This group discussion on the state of digital media-authoring tools attracted dozens of equiptment manufactures, software developers, and video professionals to the Omni Hotel. This report covers as faithfully as possible the topics discussed that evening

The Open Studio Interoperability Award

The Roundtable began with the presentation of the first annual Open Studio Interoperablility Award to Apple Computer for QuickTime 3.0. The award, planned as an annual event, will be presented to the company or product that's done the most to further successful collaboration among users of digital media-authoring tools. The award is what the Open Studio initiative is all about--advancing efforts for the sharing and exchange of digital-media assets among different systems on diverse platforms and across diverse networks.

Apple was the recipient of the first annual award because of the Open Studio Advisory Committee's belief that QuickTime--more than any other product or tool--and Apple --more than any other compnay--are responsible for spawning the digital media-authoring industry. We look forward to reviewing the work of other companies in the future, and to recieving your nominations and comments regarding future recipients (send via E-mail to **videography@psn.com**).

VIDEOGRAPHY Editor Brian McKernan presents Mitchell Weinstock, Apple Computer's MultiPlatform Product Marketing Manager, QuickTime Technology, with the first annual Open Studio Interperability Award for QuickTime 3.0,

Video Bit Depth and Color Space

Discussion began with Tim Schaff of Apple's QuickTime team addressing the news made the previous day on the East Coast during MacWorld Boston. That news, of course, was: the investment by Microsoft of $150 million in non-voting Apple stock to be held for three years; a guarantee that for five years Microsoft will provide the same number of major releases of Office for the Macintosh as for Windows; that Internet Explorer will become Apple's default brower; that there will be Java compatability between the two companies; and--perhaps most intriguing and least understood--the cross licensing of all patents currently held by, or granted to, the companies in the next five years.

The group moved on to discuss the need for a higher bit-depth video standard and the advantages of one colorapace over another. Special effects producers who work digitally on film projects often work at higher bit depths and many serial digital video devices work a 10 bits. Users asked what obstacles were preventing the standardization on a format with more bits.

QuickTime architect Peter Hoddie explained that his team was looking at the issue and had dealt with it somewhat in version 2.5. That version reads and displays Photoshop files created in 16 bits per color further support in version 3.0. They talked to Silicon Graphics (SGI) about their format when considering a standard for 16-bit-per-channel RGB files, believing that QuickTime should be compatible with something, rather than inventing a new, non-compatible implemention. The issue of higher bit-depth YUV video offers less consensus, so they are waiting for something closer to a standard to emerge and thereby avoid an independent solution. Users turn to QuickTime, in large part, for exchange of data so Apple is eager to support these formats.Bernard Lamborelle, Technical Marketing Manager for Matrox, brought up color conversions from RGB to YUV. He asked if storing in YUV and having applications process in YUV could provide improvements by avoiding the color space conversion.

Fast's Electronic's Meinrad Zeller called for Adobe to perform its processing in YUV color space rather than in RGB. He suggested that they could save 50% of the processing time by avoiding colorspace conversions. He supports the initiative by Microsoft to get manufacturers to perform colorspace conversion in hardware, for speed, but would prefer that all effects be done in YUV.

The issue of using 10 bit YUV versus 8bit RGB processing also came up. Microsoft's Steve Gabriel described 10 bit YUV processing as a red herring. It is a packed format designed for transmission and storage. Resampling requires precision beyond 10 bits — 16 bits per component would work well, he thought. Birkmaier added that the 10 bits were for the luminance channel only and included values outside of 100 IRE so that not all 1024 values were even available for recording the image.

Steve Gabriel worked at Ampex during the early days of digital systems participating in decisions that shaped the industry. He recalled this history and reminded us that we inherited the digital formats we use today from tape machines, designed originally with numbering systems barely adequate to digitize a signal once. Engineers developed 8-bit YUV, and 8 bit RGB on computers, allowing little room for processing.

"At Ampex, on the AVA frame buffer," Gabriel recalled, "we were trying to save money on hardware with only 4k DRAM's available, so we designed a frame buffer that was YUV instead of RGB.... I thought processing got harder, some things were easier, but the original representation was in RGB....Both YUV and Gamma are analog forms of image compression. The Gamma curve was for transmission ease and we use it today in DCT compression, JPEG and MPEG. It is kept, because it is a perceptual equalization, but it is not good for processing."

Gabriel described 16 bit RGB with linear gamma as an ideal system because it represents energy and models what the light is doing. A 16 bit RGB Gabriel described 16 bit RGB with linear gamma as an ideal system because it represents energy and models what the light is doing. A 16 bit RGB system would have sufficient bits for processing. He suggested that 10 bit YUV had come about because it was just sufficient for representing 8 bit RGB.

Gabriel thought that any processing advantages of YUV over RGB would be minimal and outweighed by the advantages of working with a linear RGB energy representation. "We think of digital as perfect -- it is not. We are so constrained by having 8 bits or ten bits that the generation losses get very discrete. We start having contouring and other errors. You need more levels in the digital representation so that it behaves more like analog. So that generations lead to a graceful addition of noise instead of more apparent artifacts. The same is true for both video and audio."

Fast's Zeller granted that RGB was the initial acquisition format at the CCD, but added that 4:2:2 YUV was an elegant storage format that would be difficult to replace in all acquisition equipment.

One attendee described 10 bit YUV as a good way to send a signal down a wire and to store it on disk, as a packed format not a processing format. For intermediate processing however, it is best to blow it up to provide room to over shoot and under shoot. The question was, he thought, how many colorspace conversions the image could tolerate.

Zeller asked, why do additional colorspace conversions if you can process in YUV? What can RGB do that YUV can't?

Gabriel responded that the only drawback or advantage to 4:2:2 YUV is that you throw away chroma samples, so samples don't line up in space. Once you've sub-sampled chroma you have processing problems. But, since 10 bit YUV transports best, we will continue to live with conversions.

Roy Edenson of Texas Instrument's Semiconductor Group described problems in processing with insufficient bit resolution. In digital color space conversion, calculations frequently result in values that cannot be precisely represented. Look up tables may not directly correspond to discrete levels available within a range. If you can't get between the bits, you have to select one bit above or below. Further downstream processing often results in compounding of errors.

Digital processing requires more values than that needed to represent a continuum of values within the dynamic range. The range must accommodate the widest swings without forcing the extreme values to assume a single bit value at the top or the bottom of the quantization range. When a film segment contains artistic information down in the bottom end, 24 bits of color total doesn't allow enough resolution to separate the various values. Processing problems arise when you have to work with only those few bits. For example, if artistic material uses only the bottom 8 levels, television people, and some equipment, may try to spread that picture out to assume an average picture level of about 50%. When processing digital signals in post production, operators should take care not to over process a signal, leaving too few bits within a dynamic range. This causes problems during downstream manipulation and display.

> "Why do additional colorspace conversions if you can process in YUV? What can RGB do that YUV can't?"

Texas Instruments builds DLP displays that provide linear light response and remove CRT gamma correction for maximum scene fidelity. At low brightness levels, insufficient bit information may lead to level contouring through no fault of the display. The signal may not use the display's full dynamic range.

Matrox's Lamborelle described the manufacturer's pressure to deliver a cost effective solution. Most digital video users process and store in 8 bit YUV 4:2:2 economically, while maintaining good picture quality. Why must they lose quality through color conversions, if software can process in YUV?

Trish Meyer, a principal in CyberMotion, a project design studio in Sherman Oaks CA, described a project that illustrates a problem users face. Meyer received 3,600 frames originated on 35mm VistaVision film transferred to Cineon and bumped down to 8-bit. She processed the files over two weeks and converted the frames back out to SGI files. The client found problems with the project that, it turned out later, were mostly with the 35mm transfer. The film house blamed her "8-bit" system and the conversion down from 10 bit to eight. Users don't want to have to make excuses. So, she concluded, it may be more cost-effective to deliver 8-bit YUV solutions, although users need the option to work with higher end files.

Video professionals made compromises earlier because of inadequacies of CRTs, recording mechanisms, and other devices, Meyer explained. We patched around the original RGB to get it through a narrow pipe. We threw information away using analog compression. Now we can apply digital compression to very high-quality etc.

Audio Sync, DV Audio, and Audio Resolution

Chris Meyer of CyberMotion, one of the roundtable's most faithful members, brought up some old issues but with a new twist, DV. The fast growing DV format has raised new questions about audio sync and pixel aspect ratio.

Chris Meyer again reminded all manufacturers present that SMPTE provides a standard for syncing audio. SMPTE document 272m describes it. Meyer's position, "If you do professional audio, comply with the SMPTE spec." He also advocates standardization on 48 kHz sampling and, as a last resort, he wants manufacturers to publish the way they broke the rules so that users can figure out how to compensate when synchronization problems arise. He says that many software applications and digital media cards don't pack the correct number of samples per video frame

Chris Meyer asked manufacturers at the roundtable for clarification of the new twist added by DV, locked versus unlocked audio. Michael Brinkman of Panasonic broadcast division explained that his company had avoided the 44.1 kHz rate entirely in the DVC Pro format by using 48 kHz sampling exclusively. He couldn't comment on decisions made by the consumer division of his company.

Fast's Zeller offered comments on locked versus unlocked audio in DV. He said either would be in sync practically speaking, but that with unlocked there might be a few samples remaining at the last frame of an edit that would result in audio clicks at a hard-cut edit.

Chris Meyer pointed out that you usually do get a click at edits even if you do have the correct number of samples unless the audio is silent because there is usually a DC offset between the wave forms in two adjacent tracks. This is why users need a provision in editing software for a very quick fade up and down to remove the clicks.

Zeller brought up a problem with current video editing software. It does not support DV audio properly. QuickTime 3.0 will support audio in a compressed DV stream but, today's video editing software will not. Now, he explained, editors must artificially extract the audio out of the DV data stream and put it into a QuickTime file or the wave portion of an AVI file. This sometimes causes the editing software to have a two frame offset between audio and video. The DV format doesn't cause this. It is a limitation of current video editing software.

The dynamic range or bit depth of both audio and video came up repeatedly through the evening. Matrox's Lamborelle asked if there was any consensus on what the market needed for audio 20 bit, 24 bit, or 48 bit.

Chris Meyer distinguished between audio delivery and production formats. Higher resolution, with headroom, allows overshoot during the production process so that we can bring it back down later to fit within the range of the distribution format. In Meyer's opinion, for distribution, "sixteen-bit linear resolution is pretty darn good if you use it all up—highest to lowest. But, for production, you really want 24 bits."

Demonstrating that consensus can sometimes be a rare commodity, a user from a digital audio production house offered his thoughts.

"Twenty-four bit isn't always enough," he said. "We use floating point, which if you run the numbers works out to 24 bits with no specific top or bottom until you are ready to crush it back down to 16 bits for output." He wondered if there might be an opportunity for a similar approach to video, using a floating-point system as a production format. He also discussed the IEEE PCM format for 32-bit floating point audio defined by Microsoft in the Wave format. He suggested that users like himself would appreciate the ability to transmit that data stream to hardware for real-time playback.

Since we seem to have reached consensus about needing different bit depths for audio production versus distribution, Radius' Mike Jennings asked, do we need different formats or can we have a single format like wave that can handle different resolutions? He asked what part QuickTime might play in that.

Apple's Peter Hoddie responded that they built support for floating point audio and for 32 bit and 64 bit audio into QT 3.0. Apple worked on this with SGI, who had already defined methods to handle it with AIFF files. High resolution audio works in QuickTime 3.0 with AIFF files, QuickTime movies, and AU files. They would like to have it working with WAVE files but have had difficulty getting specifications on that format.

"Apple…built support for floating-point audio and for 32-bit and 64-bit audio into QuickTime 3.0."

DV, Pixel Aspect Ratios, and DTV

Chris Meyer also asked for clarification on how to deal with the vertical resolution of the 720 by 480 DV format. Since DV uses 480 lines vertically it creates questions as to where the image starts in the 486-line raster used by typical serial digital systems. If you import a DV stream into Adobe After Effects, for example, do you place the 480-pixel image at the top of the raster, move it down by a set number of lines, or do you scale it? Scaling slightly alter the aspect ratio and could introduce artifacts; a vertical offset of an odd number of lines might result in inverting the field order, a devastating problem for interlaced images. If you output a 720 by 486 video stream to DV format, do you squeeze it down to 480 lines or do you crop lines off the top and bottom?

Panasonic's Brinkman said that he believes that DVCPRO devices crop out lines and don't change pixel aspect ratios. Truevision engineer Marshal Johnson, said that his company—along with Radius, Adaptec, and others—are working on the issue. He said that DVCPRO outputs SMPTE 259M using a pattern that omits two lines at the top and four lines on the bottom.

Microsoft's Steve Gabriel explained that much of the vagueness in formats such as D-resulted from the fact that they essentially digitized analog video. Blanking problems exist in digital systems because they were never well-defined in the analog word. He described compromises that we accept such as the 1/700 skew in all TV rasters and the half lines at the top and bottom that help to compensate for that skew. Most systems ignore the skew and treat the image as an orthogonal window. He pointed out that ATSC for digital broadcast and DVD both use 704 horizontal lines instead of 720, although some configurations support the 720-line resolution as an option. No one thing has destroyed the standards. The system is basically a mess.

Gabriel also made a few interesting comments regarding advanced television and the war between computer and consumer electronics manufacturers. Computer designers understand the need for generalized scaling hardware after demodulation and MPEG decoding. Computer systems now include scaling as part of current graphics hardware, because these systems must scale into different size windows.

If consumer electronics manufacturers would use this kind of scaling technology in their DTV sets, most conflicts would go away. Gabriel describes it as the trade off the computer industry would like to have. If computer DTV manufacturers have to put de-interlacing hardware into their systems, the consumer electronics manufacturers should have to put in scaling hardware. If manufacturers build a chip for TV sets following the ATSC spec that saves three or four gates by ignoring the full MPEG, they ignore the possibility of supporting a 240 by 360, or a 320 by 240 pixel image. These viable, smaller formats provide quality comparable to NTSC. VHS equates to a resolution of 160 by 200 and most people find it quite acceptable. Computer system vendors believe that the ability to have high levels of multiplexing provides value and the consumer

(continued on page 112)

electronics vendors are looking narrowly at the issue. The computer industry and its allies don't want to lock out the consumer electronicsmanufacturers. They are simply looking for more extensible and interoperable systems.

Gabriel would prefer to see the television industry make the transition to digital before moving up the resolution scale. Many of us believe that this would allow a focus on practical and economically viable improvements until higher resolution HDTV becomes affordable. Why shouldn't we, for instance, enjoy the benefits of 640 by 480 resolution progressive scan displays for motion video when VGA monitors costing $200 offer this resolution today? Most production professionals who deal daily with the idiosyncrasies of interlacing would wish it a long and happy death.

Alpha Bundling With Compressed Video

Radius' Mike Jennings brought up another problem that we've discussed previously at the Roundtables, the need to bundle alpha-channels with M-JPEG video. Three-D programs create alpha channels with the frames they render that allow compositing and keying of those images. M-JPEG, the compression used by most real time motion playback systems, does not support alpha channels. M-JPEG codecs ignore them.

Jennings describes the procedure for dealing with alpha channels now. Two QuickTime codecs "Apple Animation Millions Plus" and "Apple None" deal with alpha channels, but create very large files. Users rendering extremely long segments must extract the alpha channel using a utility and save it to a separate file, then reprocess the original RGB data into M-JPEG. That means more render time and more disk space. Jennings wants software developers to create a standard format where systems output alpha channels to a track that will survive when the RGB data is compressed with an M-JPEG codec.

He called for a standard way to handle alpha channels in QuickTime and for an evangelizing effort to 3D and compositing software developers.

Peter Hoddie responded by asking how the group would prefer the files be handled. In one scenario you store both the M-JPEG and the alpha channel as separate tracks in a movie; in another you would embed the alpha-channel information into the JPEG bit stream. Hoddie suggested that separate tracks might allow greater freedom to reprocess the alpha channel separately from the video, even including JPEG compression, if desired. Unfortunately, Hoddie says, it doesn't provide much of a solution since applications don't support it today. The alternative, embedding the alpha channel information into the JPEG bit stream, creates a fourth channel that QuickTime can render

"Most production professionals who deal daily with the idiosyncrasies of interlacing would wish it a long and happy death."

but that the codec will just ignore. You can embed the forth channel in a legal way so that applications that parse JPEG correctly will work.

Jennings preferred a separate track since he often needs to process the alpha channel separately for special effects. He hoped that the Open Studio Roundtable could provide a forum to develop a consensus on the issue and get such software developers as Radius and Adobe to agree on a standard. He suggested forming a committee of software developers to work on the problem.

An engineer in the group described huge benefits in supporting alpha channels in MPEG. He suggested that alpha channels compress efficiently in MPEG, providing an additional 2- or 3:1 compression. Every digital studio in the world has keyers that can use alpha channels, but—long range—he suggested, the issue has to do with compositing multiple streams of MPEG. His company has implemented MPEG-2 with alpha, called MPEG-plus, that

uses the standard semantics of MPEG-2 PES packets. They "co-opted" one of the three available user data types and implemented MPEG with alpha data. They avoided DCT compression on the alpha channel because it created severe artifacts, so they used modified run-length encoding, which is legal, he said, within the MPEG semantics.

Someone suggested that it didn't make sense to bundle an alpha channel with an MPEG stream since MPEG was only a distribution format. Several people disagreed.

The engineer pointed to user-defined formats such as JPEG-Plus, implemented by Storm Technology and others. Embedding additional data into the image that doesn't break the decoder, he suggested, could allow support for alpha channels and other metadata.

VIDEPGRAPHY'S Birkmaier further established the value of bundling meta-data-like alpha channels with MPEG

at the set-top level or at the PC level. Developers of MPEG-4 are considering a composition model that includes foreground and background sprites. Alpha channels allow compositing of sprites with capabilities superior to some other methods. Systems might, realistically, send alpha channel information to a destination and composite it locally at the receiver. We may eventually carry all this information through the production cycle and encode multiple streams of alpha channels and other metadata, suspending that composition all the way to the receiver. We are developing a pallet of tools for the ultimate delivery platform. Whether you are producing for WebTV or for products from Microsoft, Apple, or the consumer electronics industry, the infrastructure should support it.

These were the issues we discussed at the SIGGRAPH roundtable. Thanks to the sponsors and the participants who understand this industry so well for sharing their ideas — ideas that enable each of these events to advance the industry of digital media authoring.