THE POSSIBILITIES ARE INFINITE FUJITSU

Containing Sound and Light
Wireless Video Coming Up Net Using MPEG-4

TECHNOLOGY
BACKGROUNDER

Introduction

Many of us over the age of 30 will remember Dick Tracy's watch, the one from the 1950s that allowed the comic strip detective to make phone calls from his wrist. Sure, he had a limited number of people he could call, but anyone watching him do it was impressed. Now, of course, cell phones are ubiquitous, so the younger generations don't think much of Detective Tracy's cool watch. Today's equivalent might be a device allowing you to watch video streamed wirelessly to a palm-sized device from anywhere on the planet. This capability, as it emerges, will make handheld, real-time video conferencing a reality.

So here we are 50 years after Dick Tracy and billions of telecommunications dollars later and we've solved many of the challenges of audio - but what about wireless video? Most of the technology pieces are in place. The only major issues remaining are consumer demand and an economic model that will make it feasible for companies to develop the solution. One of the most important technology pieces is the MPEG-4 video standard. MPEG-4 is ideally suited to the task of wireless mobile device video, as will be discussed below. First, a bit of the technology background that has led us to this point.

The technology for streaming audio by itself is acceptable – just consider music and phone conversations. Video, however, is different and, by the way, requires audio. In the 1920s silent movies gave way to movies with sound, but only after the technology advanced enough to make sound viable. Today, silent video is considered one shade shy of "useless" in the mind of the consumer. The two key challenges with adding sound to movies were "synchronization" and "amplification." In a streaming video environment (whether wireless or over a wired network), the basic problem of audio/video synchronization is still with us. We can tolerate some occasional glitches in video, but we are extremely sensitive to audio discrepancies such as stuttering or "out-of-sync" audio. The human ear can detect audio errors as small as a few milliseconds, so accurate audio and video synchronization is critical to successful video transmission.

To ship audio and video together either over a wired or wireless network, we must first employ techniques to put the video and audio in a container and keep it together before, during and after shipping. While much has been spoken and written concerning the techniques of encoding and decoding audio and video, the technologies involved in placing these encoded audio and video streams in containers for shipping over wired and wireless networks is less understood. It is our goal in this article to shed a bit of light on this technology aspect of audio and video transmission.

A/V Containers

So, what is an A/V container? Simply put, an A/V container is an information framework where encoded audio and video samples reside. (See Figure 1.) Another term to describe A/V containers is a "file format." A container is NOT an encoding/compression method. Many A/V containers are actually encoding/compression method agnostic. They simply describe how A/V samples coexist inside of a "file." AVI, MOV, ASF, MPEG-1 System Streams, MPEG-2 Program Streams, MPEG-2 Transport Streams, and MPEG-4 System Streams are all A/V containers.

The basic concept of an A/V container is that it is simply a "box" into which audio and video streams are placed for shipping to a destination. "Destinations" include a hard drive as a file, a CD-ROM, a DVD, a remote location over a LAN or via the Internet. In some cases, the receiver has no idea what is in the container until it is opened. All container formats have "headers" that inform the receiver as to the contents of the container. (See Figure 2.)

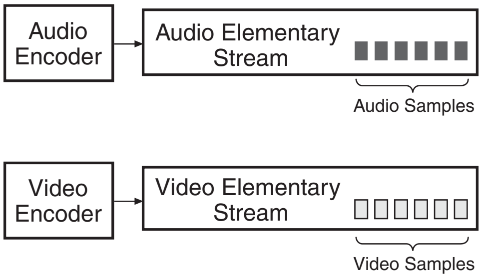

Figure 2 - Encoder

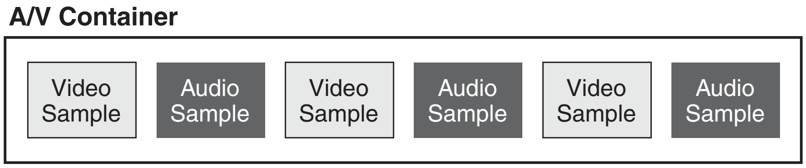

Figure 1 - A/V Container

AVI Files

An early PC-based A/V container was AVI, which stands for Audio/Video Interleaved. AVI files contain a header identifying the file as an AVI file, and then the audio and video samples. In the header of an AVI file is a four-character code (FOURCC) that identifies the type of video stream contained within the file. This FOURCC indicates to the receiver of the file what video decoder is needed in order to view the file. A list of FOURCCs is available at http://www.fourcc.org/codecs.php. AVI files were never designed for streaming across a network, either wireless or wired. In fact, AVI files actually predate common PC networking entirely. Figure 3 shows the typical layout of an AVI file with one video steam and one audio stream.

While it might seem possible to stream AVI files (since the video and audio can be interleaved), AVI files provide no mechanism for embedding time stamps within the audio and video streams to handle resynchronization when packets are lost on the network. Additionally, since the index usually comes at the end of the file, most players won't even attempt to initiate an AVI file until the index is received, that is, once the entire file has been transferred. Notice how it is possible to use virtually any compression method for both audio and video within an AVI file.

Due to the simplicity and flexibility of AVI files, it remains one of the most commonly used and popular A/V container formats. For example, A/V captured from digital camcorders is contained within AVI files under Windows. DivX. The popular public domain MPEG-4 codec contains itself within AVI files under Windows. So AVI will be with us for years to come. But it cannot and will never be able to stream. It is the prime example of a "downloadable" container format.

So let's now take a look at another extremely popular container format, MPEG.

MPEG

Many people don't think of MPEG as a container format but rather as an encoding or compression technology. This is reasonably accurate. MPEG does specify powerful methods for encoding video and audio, but it also specifies exactly how the encoding video and audio should be placed within a bitstream or container. MPEG uses the term "system" to describe how MPEG elementary streams (audio and video) are to be placed in a container. With MPEG it is not "legal" for an elementary stream to stand alone. When you add system information to the elementary video bitstream, you have then created a valid MPEG bitstream. For MPEG-1, the system information (or "layer" in MPEG speak) is straightforward. The relationship between elementary streams and system stream is strong and well-defined to make up a complete MPEG stream. This relationship is even stronger in MPEG-2 and considerably stronger in MPEG-4.

So what is the purpose of the system stream/layer in MPEG? Simply put, to keep track of what is happening within the elementary streams. The stream layer provides timestamps, synchronization, packetization, and other features to allow for the playback of elementary streams reliably and accurately and, in the case of video and audio together, in sync.

MPEG-1 was designed for a relatively error-free transmission environment such as CD-ROM. MPEG-2 added the capability for transmission over networks where errors are much more prone to occur. This meant that the system layer for MPEG-2 needed to be more sophisticated to handle such an environment; and it had to retain backward compatibility with MPEG-1. Figure 4 illustrates the scope of the MPEG systems layer with regards to MPEG-2.

MPEG-1 was popularized by a subset of its capabilities known as "Constrained Parameters Bitstream" or CPB. CPB was designed to store the same amount of video and audio on a

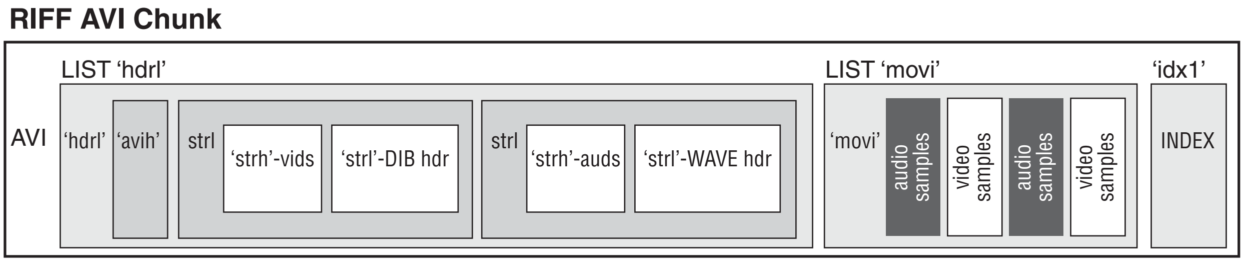

Figure 3 - RIFF AVI Chunk

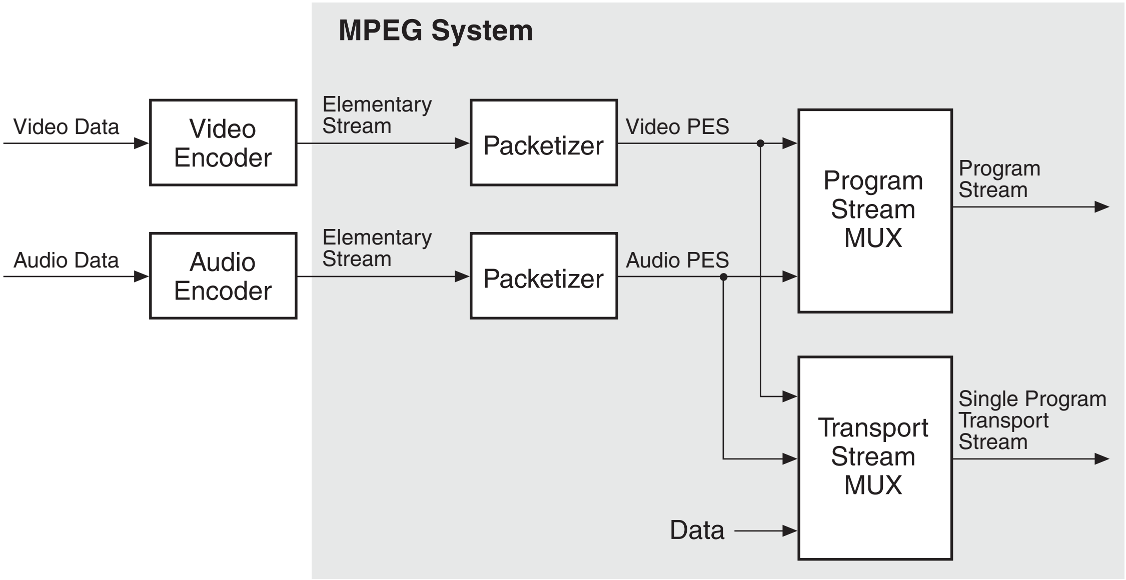

Figure 4 - MPEG System

CD-ROM as the Redbook audio standards would allow. This means that on a 74-minute CD-ROM, you can store a full 74 minutes of video and audio. This standard became known as White Book or Video CD (VCD). The video data rate was fixed at 1.15Mb/s and the audio at 224 Kb/s at 44.1kHz and the video size was "locked" at 352x240 30fps NTSC and 352x280 25fps PAL (also known as CIF).

This works nicely but it is only one fourth of the quality of "standard" television. To a television watching public, it is noticeable and somewhat annoying. Still, at the time it was an amazing technological achievement. At around 1.5 Mbit/s (125 Kbytes/s) today's cable modems can easily deliver a MPEG-1 CPB stream but wireless cellular technology networks still have a long way to go before doing this reliably. Additionally, MPEG-1 lacks the capability to correct errors in network transmission and resynchronization. MPEG-2 allows this type of functionality, but by the time the technology caught up to practical implementation, MPEG-4 and other streaming formats had moved onto the scene.

MPEG-2 is intended for higher data rates, in the 6 to 10 Mbit/s range. MPEG-2 at 720x480 between 6Mbit/s and 9Mbit/s is ideal for DVD and was therefore chosen for such work. At those bit rates, however, network delivery of content becomes much more difficult. The video quality, however, is on an order with that expected by those used to watching TV. MPEG-2 also has the packetization, error tolerance and recovery, "trick-play" capabilities, and other features in its system layer necessary for WAN broadcasting. So is MPEG-2 suitable for wireless/mobile device streaming video? The network channel must be able to deliver data on the order of one megabyte per second without fail in order to make MPEG-2 video reliable, playable and watchable. So the search continues for the best way of streaming video formats. MPEG-4 is the answer.

Why MPEG-4?

The requirements: a high compression ratio, great video quality, and a powerful systems layer. MPEG-4 provides these features, making it ideal for wireless mobile device video. (Note: there was no MPEG-3; the standard moved from MPEG-2 directly to MPEG-4. MP3 audio, which is often confused for MPEG-3, is actually MPEG-1 Layer 3 audio.)

With MPEG-4 it is possible to achieve MPEG-2 720x480 quality in the 1MBit/s range. If you scale down the video to MPEG-1 CPB, you can get to the 350KBit/s data rate range with reasonably acceptable video quality. Scaling down the video even further, data rates in the 100KBit/s range are possible. Of course, due to the degradation of quality, the video isn't worth watching. We consider 352x240 (also referred to as the Common Interchange Format or CIF) to be the minimum acceptable resolution for watching video. QCIF is interesting and on some network channels necessary. See Table 1 for a list of commonly used video-resolution terms.

Table 1.

Term	Resolution
D1	720 x 480 *NTSC* 720 x 576 *PAL*
CIF	352 x 240 *NTSC* 352 x 288 *PAL*
QCIF	176 x 120 *NTSC* 176 x 144 *PAL*
VGA	640 X 480
QVGA	320 x 240
QQVGA	160 x 120

Regardless of video resolution the question of whether MPEG-4 is suitable for wireless transmission of video comes into play. That question is answered by examining two aspects of MPEG-4: its video-compression capabilities and its systems layer. Today's mobile phones using 2.5G technology with General Packet Radio Service (GPRS) can achieve a network data transmission rate in the neighborhood of ~112kbps. Not a very speedy neighborhood and not a terribly reliable one either, depending upon where you are. 3G promises network data-transmission rates in the neighborhood of 2Mbps. This is a significant move up. While 2.5G w/GPRS is capable of some form of "streaming" video, it won't be until 3G is rolled out that we see video streaming becoming ubiquitous on mobile phones. And of course, the fatter the data pipe, the better the video quality.

To put some of these numbers in context, let's look at them a bit more closely. Most network bandwidth data rates are stated in bits per seconds. Since there are 8 bits in a byte, a speed that we can relate to more easily is bytes per second or bps divided by 8. So 2.5G w/GPRS is a 14 thousand byte per second channel (14kBps. Note the large B indicating bytes). 3G, by contrast, is a 250kBps channel. A CD-ROM drive operating at 1X is transferring data at a speed of 150kBps. Since 3G has bandwidth of 250kBps we should - theoretically - be able to stream MPEG-1 CPB video and audio directly to a 3G terminal without any problem. We would, of course, be consuming 60 percent of the channel for our video, but it would look pretty good. With considerably better compression than MPEG-1 could we get to DVD-type quality over 3G? With MPEG-4 it appears the answer is yes.

MPEG-1 would require a 600kBps (4.8Mbps) channel to achieve full D1 resolution video. This would clearly bust the bandwidth budget for 3G. DVDs (which currently use MPEG-2 video compression) are often compressed in the range of 6Mbps. So if we would like a DVD-type video experience with 3G we need a compression technology other than MPEG-1 or MPEG-2. It remains unclear as to whether a DVD video experience will be possible over first-generation 3G implementations, but if it is possible, MPEG-4 and its associated compression technologies will lead the way. So what will the "high end" of video performance over 3G in its current form be? We contend that it will indeed be CIF. As we have said, CIF at 30fps is very "watchable" and is currently the benchmark. So if we have 2Mbps of channel and MPEG-1 CPB VCD will fit over the channel, why not just use that? The answer is simply that with MPEG-4 we can get almost three times better compression for the same bit rate, and perhaps even better under certain conditions. That means that at CIF resolution at 30fps a MPEG-4 stream at 300-400 kbps will be the same quality as MPEG-1 CPB VCD at 1.2Mbps.

Fujitsu has added its MPEG-4 core to its new line of mobile media processors or MMPs. (See http://www.fujitsu.com/us/services/edevices/microelectronics/otherassps/mobilemedia/.) The power-frugal Fujitsu MPEG-4 core is capable of bit rates up to 2Mbps and encoding/decoding to CIF at 30fps. (See Figure 5.) It is suited to current video-phone quality of QCIF at 15fps. There are two noteworthy features of this MPEG-4 core. First is its motion-estimation circuit, which employs an adaptive motion vector technique that yields results of one-twentieth the computational load for traditional motion detection. The other is its autonomous clock-gating control, which supplies clock only to the functional blocks necessary for the current encoding/decoding demand. In short, the Fujitsu MPEG-4 core is the ideal chip for existing 2.5G wireless phones using GPRS as well as the next generation 3G phones, where video quality will be in high demand.

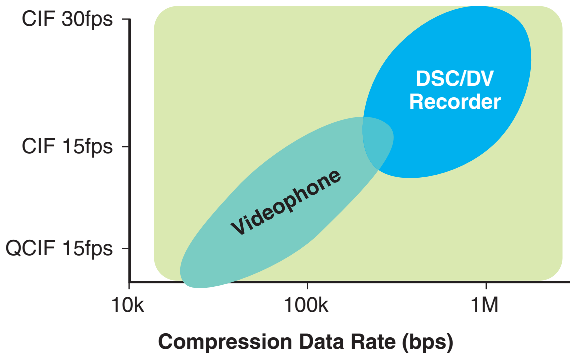

Figure 5 - Compression Rates

MPEG-4 System Layer

A major advancement of MPEG-4 over MPEG-1 and MPEG-2 is its “systems” layers. A cursory look at MPEG-4 systems can yield a better understanding of the importance of MPEG-4 now and in the future. Many think that MPEG-4 is simply a better compression standard for both audio and video. While this is certainly true, MPEG-4 has many other features that exist within the standard that will be relevant to the video broadcast industry for years to come. In addition to audio and video compression, MPEG-4 adds the ability to code animations, 3D environments and images, provide for stream interactivity, provide digital rights management, and allow for content distribution on a variety of networks, including IP and MPEG-2 transport streams (DVB). Hundreds of companies have participated in the creation of the MPEG-4 standard and, in fact, MPEG-4 encompasses 20 International Standards Organization (ISO) specifications, as well as several industry organization standards such as 3GPP (http://www.3gpp.org) and the Internet Streaming Media Alliance or ISMA (http://www.isma.tv). For a visual representation of the MPEG-4 standard, see Figure 6.

According to the MPEG-4 working group (http://www.chiariglione.org/mpeg/), MPEG-4 “systems” were developed to “provide the necessary facilities for specifying how audiovisual objects can be composed together in an MPEG-4 terminal to form complete scenes, how a user can interact with the content, as well as how the streams should be managed for transmission or storage.”

Also according to the working group: “Much of the functionality that MPEG-4 provides comes from the systems part. As systems take care of (among other issues) streams management and scene description, it acts as a ‘wrapper’ to the source coding technology.”

Part of the MPEG-4 systems layer is its file or “container” format. A MPEG-4 video elementary stream, such as produced by the Fujitsu MPEG-4 core, cannot stand alone. It must be placed inside of a “container” in order to be processed and decoded by a MPEG-4 decoder. This means that MPEG-4 video, prior to being transmitted from one mobile phone to another, must be placed inside of a container (referred to as a MP4 file). This MP4 file describes what is inside of the MP4 container and how the video and audio should be played by the receiver of the MP4 file. Without this framework, the receiver cannot identify the data it has received. A properly formatted MP4 file is critical to MPEG-4 interoperability between terminals. For more information on MPEG-4 systems see http://www.chiariglione.org/mpeg/faq/mp4-sys/mp4-sys.htm.

The Fujitsu MMPs produce a MPEG-4 video elementary stream (ES) also known as an .M4V. When coupled with a MPEG-4 audio ES, this video ES, which is sometimes called an .M4A file, constitutes a “watchable” video known as an .MP4 file. So where does the “coupling” take place and who does it? The answer for Fujitsu MMPs is the Fujitsu MP4 software library. The Fujitsu MPEG-4 CODEC core, coupled with the Fujitsu MP4 library, allows for the creation of MPEG-4

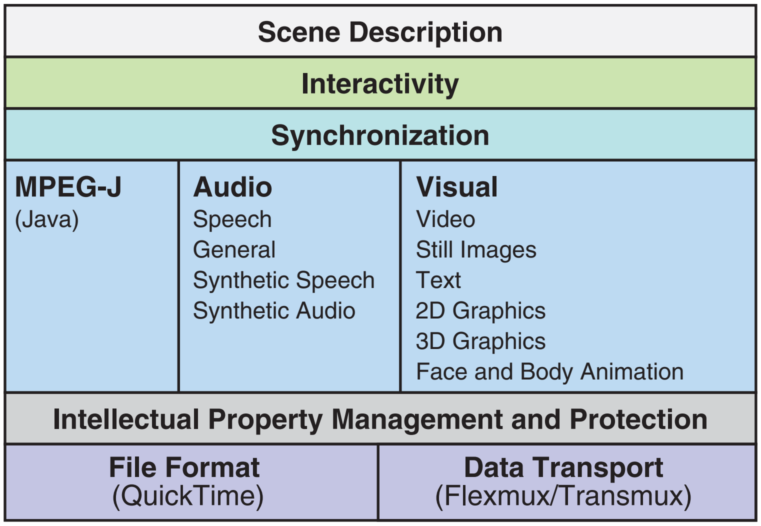

Figure 6 - MPEG-4 Standard

compliant files that are interchangeable across virtually all MPEG-4 terminals. MPEG-4 video elementary streams that are not encapsulated inside of MP4 files, however, cannot be guaranteed to play on all MPEG-4 compliant terminals.

There are a variety of custom MPEG-4 implementations on the market today. DIVX, the popular Internet video-compression technology, consists of MPEG-4 video elementary streams encapsulated inside of an .AVI file format. Microsoft stores its MPEG-4 video elementary streams inside of a custom container format called Advanced Systems Format or ASF. These are clear examples of companies seeking to take advantage of the improved video compression that MPEG-4 offers, but without making a commitment to the entire MPEG-4 standard. The advantage to using the MPEG-4 file format is the ability to transmit seamlessly and interchangeably across MPEG-4 terminals be they mobile devices, desktop devices or simply MPEG-4 TV terminals regardless of make or model. Other container formats, including Microsoft's Windows Media, cannot offer anywhere near the interchangeability of MPEG-4. (See http: / /all-streaming-media.com/ streaming-media-faq/ faq-mp4-streaming.htm.)

FUJITSU MICROELECTRONICS AMERICA, INC.
Corporate Headquarters
1250 East Arques Avenue, M/S 333, Sunnyvale, California 94085-5401
Tel: (800) 866-8608   Fax: (408) 737-5999
E-mail: inquiry@fma.fujitsu.com   Web Site: http://us.fujitsu.com/micro

Printed in U.S.A.
GDC-TB-21212-8/2006