

**Topic no. 76****Film & Video Conversion**

**Television standards conversion** is the process of changing one type of TV system to another. The most common is from NTSC to PAL or the other way around. This is done so TV programs in one nation may be viewed in a nation with a different standard. The TV video is fed through a video standards converter that changes the video to a different video system.

Converting between a different numbers of pixels and different frame rates in video pictures is a complex technical problem. However, the international exchange of TV programming makes standards conversion necessary and in many cases mandatory. Vastly different TV systems emerged for political and technical reasons – and it is only luck that makes video programming from one nation compatible with another.

**History**

The first known case of TV systems conversion probably was in Europe a few years after World War II – mainly with the RTF (France) and the BBC (UK) trying to exchange their 441 line and 405 line programming.

The problem got worse with the introduction of PAL, SECAM (both 625 lines), and the French 819 line service.

Until the 1980s, standards conversion was so difficult that 24 frame/s 16 mm or 35 mm film was the preferred medium of programming interchange.

**Overview**

Perhaps the most technically challenging conversion to make is the PAL to NTSC.

- PAL is 625 lines at 50 fields/s
- NTSC is 525 lines at 59.94 fields/s (60,000/1,001 fields/s)

The two TV standards are for all practical purposes, temporally and spatially incompatible with each other.

Aside from the line count being different, it is easy to see that generating 60 fields every second from a format that has only 50 fields might pose some interesting problems.

Every second, an additional 10 fields must be generated seemingly from nothing. The converter has to create new frames (from the existing input) in real time.

**Hidden Signals: not always transferred**

TV contains many hidden signals. One signal type that is not transferred, except on some very expensive converters, is the closed captioning signal.

Tele text signals do not need to be transferred, but the captioning data stream should be if it is technologically possible to do so.

With HDTV broadcasting, this is less of an issue, for the most part meaning only passing the captioning data stream on to the new source material. However, DVB and ATSC have significantly different captioning data stream types.

## **Role of information Theory**

### **Theory behind systems conversion**

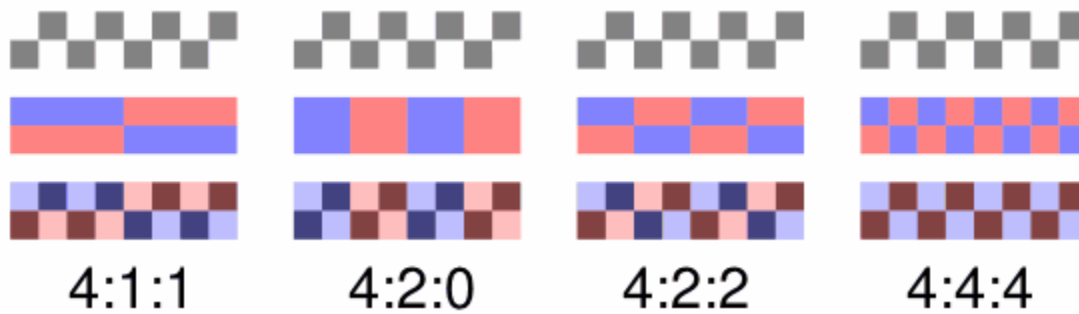
Information theory and the Nyquist–Shannon sampling theorem imply that conversion from one television standard to another will be easier providing:

- one is going from a higher frame rate to a lower frame rate (NTSC to PAL or SECAM, for example)
- one is going from a higher resolution to a lower resolution (HDTV to NTSC)
- one is converting from one progressive source to another progressive source (interlaced PAL and NTSC are temporally and spatially incompatible with each other)
- interframe motion is limited, so as to reduce temporal or spatial judder
- signal to noise ratios in the source material are not detrimentally high
- the source material does not possess any continuous (or periodic) signal defect that inhibits translation

### **Sampling systems and ratios**

The subsampling in a video system is usually expressed as a three part ratio. The three terms of the ratio are: the number of brightness ("luminance" "luma" or Y) samples, followed by the number of samples of the two color ("Chroma") components: U/Cb then V/Cr, for each complete sample area.

For quality comparison, only the ratio between those values is important, so 4:4:4 could easily be called 1:1:1; however, traditionally the value for brightness is always 4, with the rest of the values scaled accordingly.



The sampling principles above apply to both digital and analog television.

### Telecine judder

The “3:2 pull down” conversion process for 24 frame/s film to television (telecine) creates a slight error in the video signal compared to the original film frames.

This is one reason why NTSC films viewed on typical home equipment may not appear as smooth as when viewed in a cinema. The phenomenon is particularly apparent during slow, steady camera movements which appear slightly jerky when telecined.

This process is commonly referred to as **telecine judder**.

PAL material in which 2:2:2:2:2:2:2:2:2:3 pull down has been applied, suffers from a similar lack of smoothness, though this effect is not usually called “telecine judder”.

In effect every 12th film frame is displayed for the duration of 3 PAL fields (60 milliseconds) – whereas the other 11 frames are all displayed for the duration of 2 PAL fields (40 milliseconds). This causes a slight “hiccup” in the video about twice a second.

Television systems converters must avoid creating telecine judder effects during the conversion process.

Avoiding this judder is of economic importance as a substantial amount of NTSC (60 Hz, technically 29.97 frame/s) resolution material that originates from film – will have this problem when converted to PAL or SECAM (both 50 Hz, 25 frame/s).

### Historical standards conversion techniques

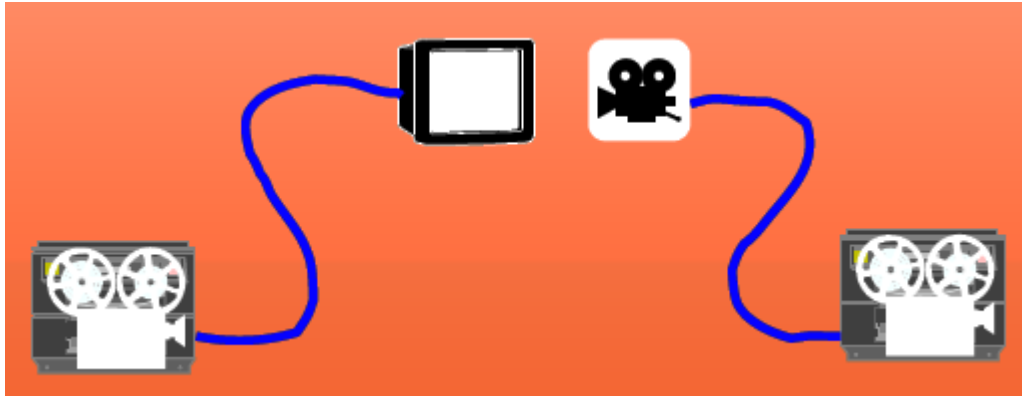
#### Orthicon to orthicon

This method was used by Ireland to convert 625 line service to 405 line service. It is perhaps the most basic television standard conversion technique.

RTÉ used this method during the latter years of its use of the 405 line system.

A standards converter was used to provide the 405 line service, but according to more than one former RTÉ engineering source the converter blew up and afterwards the 405 line service was provided by a 405 line camera pointing at a monitor!

This is not the best conversion technique but it can work if one is going from a higher resolution to a lower one – at the same frame rate.



The first video standards converters were analog. That is, a special professional video camera that used a video camera tube would be pointed at a Cathode ray tube video monitor. Both the Camera and the monitor could be switched to either NTSC or PAL, to convert both ways. Robert Bosch GmbH's Fernseh Division made a large three rack analog video standards converter. These were the high end converters of the 1960s and 1970s. Image Transform in Universal City, Ca used the Fernseh converter and in the 1980s made their own custom digital converter. This was also a larger 3 rack device. As digital memory size became larger in smaller packages, converters became the size of a microwave oven. Today one can buy a very small consumer converter for home use.

### **SSTV to PAL and NTSC**

The Apollo moon missions (late 1960s, early 1970s) used SSTV as opposed to normal bandwidth television; this was mostly done to save battery power. The camera used only 7 watts of power.

SSTV was used to transmit images from inside Apollo 7, Apollo 8, and Apollo 9, as well as the Apollo 11 Lunar Module television from the Moon; see Apollo TV camera.

- The SSTV system used in NASA's early Apollo missions transferred ten frames per second with a resolution of 320 frame lines using less bandwidth than a normal TV transmission.
- The early SSTV systems used by NASA differ significantly from the SSTV systems currently in use by amateur radio enthusiasts today.

- Standards conversion was necessary so that the missions could be seen by a worldwide audience in both PAL/SECAM (625 lines, 50 Hz) and NTSC (525 lines, 60 Hz) resolutions.

Later Apollo missions featured color field sequential cameras that output 60-frame/s video. Each frame corresponded to one of the RGB primary colors. This method is compatible with black and white NTSC, but incompatible with color NTSC. In fact, even NTSC monochrome TV compatibility is marginal. A monochrome set could have reproduced the pictures, but the pictures would have flickered terribly. The camera color video ran at only 10 frame/s. Also, Doppler shift in the lunar signal would have caused pictures to tear and flip. For these reasons, the Apollo moon pictures required special conversion techniques.

The conversion steps were completely electromechanical, and they took place in nearly real time. First, the downlink station corrected the pictures for Doppler shift. Next, in an analog disc recorder, the downlink station recorded and replayed every video field six times. On the six-track recorder, recording and playback took place simultaneously. After the recorder, analog video processors added the missing components of the NTSC color signal: These components include...

- The 3.58-MHz color burst
- The high-resolution monochrome signal
- The sound
- The I and Q color signals

The conversion delay lasted only some 10 seconds. Then color moon pictures left the downlink station for world distribution.

### **Standards conversion methods in common use**

#### **Nyquist subsampling**

This conversion technique may become popular with manufacturers of HDTV --> NTSC and HDTV --> PAL converter boxes for the ongoing global conversion to HDTV.

- Multiple Nyquist subsampling was used by the defunct MUSE HDTV system that was used in Japan.
- MUSE chipsets that can be used for systems conversion do exist, or can be revised for the needs of HDTV --> Analog TV converter boxes.

#### **How it works**

In a typical image transmission setup, all stationary images are transmitted at full resolution. Moving pictures possess a lower resolution visually, based on complexity of interframe image content.

When one uses Nyquist subsampling as a standards conversion technique, the horizontal and vertical resolution of the material are reduced – this is an excellent method for converting HDTV to standard definition television, but it works very poorly in reverse.

- As the horizontal and vertical content change from frame to frame, moving images will be blurred (in a manner similar to using 16 mm movie film for HDTV projection).
- In fact, whole-camera pans would result in a loss of 50% of the horizontal resolution.

The Nyquist subsampling method of systems conversion only works for HDTV to Standard Definition Television, so as a standards conversion technology it has a very limited use. Phase Correlation is usually preferred for HDTV to standard definition conversion.

### **Framerate conversion**

There is a large difference in frame rate between film (24.0 frames per second) and NTSC (approximately 29.97 frames per second).

Unlike the two other most common video formats, PAL and SECAM, this difference cannot be overcome by a simple speed-up, because the required 25% speed-up would be obviously noticeable.

To convert 24 frame/s film to 29.97 frame/s NTSC, a complex process called "3:2 pulldown" is utilized, in which parts of some frames are duplicated and blended. This produces irregularities in the sequence of images which some people can perceive as a jitter/stutter during slow pans of the camera.

### **What is Telecine?**

**Telecine** is the process of transferring motion picture film into video and is performed in a color suite. The term is also used to refer to the equipment used in the post-production process.

For viewing native PAL or SECAM material (such as European television series and some European movies) on NTSC equipment, a standards conversion has to take place. There are basically two ways to accomplish this.

- The framerate can be slowed from 25 to 23.976 frames per second (a slowdown of about 4%) to subsequently apply 3:2 pulldown.

- Interpolation of the contents of adjacent frames in order to produce new intermediate frames; this introduces artifacts, and even the most modestly trained of eyes can quickly spot video that has been converted between formats.

### **Linear interpolation**

When converting PAL (625 lines @ 25 frame/s) to NTSC (525 lines @ 30 frame/s), the converter must eliminate 100 lines per frame. The converter must also create five frames per second.

To reduce the 625-line signal to 525, less expensive converters drop 100 lines. These converters maintain picture fidelity by evenly spacing removed lines. (For example, the system might discard every sixth line from each PAL field. After the 50th discard, this process would stop. By then the system would have passed the viewable area of the field. In the following field, the process would repeat, completing one frame.) To create the five additional frames, the converter repeats every fifth frame.

If there is little inter-frame motion, this conversion algorithm is fast, inexpensive and effective. Many inexpensive consumer television system converters have employed this technique. Yet in practice, most video features significant inter-frame motion. To reduce conversion artifacts, more modern or expensive equipment may use sophisticated techniques.

### **Doubler**

The most basic and literal way to double lines is to repeat each scanline, though the results of this are generally very crude. Linear interpolation use digital interpolation to recreate the missing lines in an interlaced signal, and the resulting quality depends on the technique used. Generally the bob version of linear deinterlacer will only interpolate within a single field, rather than merging information from adjacent fields, to preserve the smoothness of motion, resulting in a frame rate equal to the field rate (i.e. a 60i signal would be converted to 60p.) The former technique in moving areas and the latter in static areas, which improves overall sharpness.

### **Interfiled interpolation**

Interfiled Interpolation is a technique in which new frames are created by blending adjacent frames, rather than repeating a single frame. This is more complex and computationally expensive than linear interpolation, because it requires the interpolator to have knowledge of the preceding and the following frames to produce an intermediate blended frame. Deinterlacing may also be required in order to produce images which can be interpolated smoothly.

Interpolation can also be used to reduce the number of scan lines in the image by averaging the colour and intensity of pixels on neighboring lines, a technique similar to Bilinear filtering, but applied to only one axis.

There are simple 2-line and 4 line converters. The 2-line converter creates a new line by comparing two adjacent lines, whereas a 4-line model compares 4 lines to average the 5th. Again, the greater the complexity and resulting price tag!

Interfield interpolation *reduces* judder, but at the expense of picture smearing. The greater the blending applied to smooth out the judder, the greater the smear caused by blending.

### **Adaptive motion interpolation**

Some more advanced techniques measure the nature and degree of inter-frame motion in the source, and use adaptive algorithms to blend the image based on the results. Some such techniques are known as motion compensation algorithms, and are computationally much more expensive than the simpler techniques, thus requiring more powerful hardware to be effective in real-time conversion.

Adaptive Motion algorithms capitalize on the way the human eye and brain process moving images - in particular, detail is perceived less clearly on moving objects that.

Adaptive interpolation requires that the converter analyzes multiple successive fields and to detect the amount and type of motion of different areas of the picture.

- Where little motion is detected, the converter can use linear interpolation.
- When greater motion is detected, the converter can switch to an inter-field technique which sacrifices detail for smoother motion.

*Adaptive Motion Interpolation has many variations and is commonly found in midrange converters.* The quality and cost is dependent upon the accuracy in analyzing the type and amount of motion, and the selection of the most appropriate algorithm for processing the type of motion.

### **Adaptive motion interpolation + block matching**

Block matching involves dividing the image into mosaic blocks - say perhaps for the sake of explanation, 8x8 pixels. The blocks are then stored in memory. The next field read out is also divided up into the same number and size of mosaic blocks. The converter's computer then goes to work and starts matching up blocks. The blocks that stayed in the same relative position (read: there was no motion in this part of the image) receive relatively little processing.



- For each block that changed, the converter searches in every direction through its memory, looking for a match to find out where the "block" went (if there's motion, the block obviously had to have gone somewhere..).
- The search starts at the immediate surrounding blocks (assuming little motion).
- If a match isn't found, then it searches further and further out until it finds a match.
- When the matching block is found, the converter then knows how far the block moved and in which direction.
- This data is then stored as a motion vector for this block.
- Since interframe motion is often predictable owing to Newton's laws of motion in the real world, the motion vector can then be used to calculate where the block will probably be in the next field.
- The Newtonian method saves a lot of search and processing time.

When panning from left to right is taking place (over say 10 fields) it is safe to assume that the 11th field will be similar or very close.

- Block matching can be seen as the "cutting and pasting" of image blocks.

The technique is highly effective but it does require a tremendous amount of computing power. Consider a block of only 8x8 pixels. For each block, the computer has 64 possible directions and 64 pixels to be matched to the block in the next field. Also consider that the greater the motion, the further out the search must be conducted. Just to find an adjacent block in the next field would entail making a search of 9 blocks. 2 blocks out would require a search and match of 25 blocks - 3 blocks further distant and it grows to 49 etc.

The type of motion can exponentially compound the compute power required. Consider a rotating object, where a simple straight line motion vector is of little help in predicting where the next block should match. It can quickly be seen that the more inter frame motion introduced, the much greater the processing power required. This is the general concept of block matching. Block match converters can vary widely in price and performance depending on the attention to detail and complexity.

A weird artifact of block matching owes to the size of the block itself. If a moving object is smaller than the mosaic block, consider that it's the entire block that gets moved. In most cases, it's not an issue, but consider a thrown baseball. The ball itself has a high motion vector, but its background that makes up the rest of the block might not have any motion. The background gets transported in the moved block as well, based on the motion vector of the baseball, What you might see is the ball with a small amount of outfield or whatever,

tagging along. As it's in motion, the block may be "soft" depending upon what additional techniques were used and barely noticeable unless your looking for it.

Block matching requires a staggering amount of processing horsepower, but today's microprocessors are making it a viable solution.

### **Phase correlation**

Phase Correlation is perhaps the most computationally complex of the general algorithms.

Phase Correlation's success lies in the fact that it is effective with coping with rapid motion and random motion. Phase Correlation doesn't easily get confused by rotating or twirling objects that confuse most other kinds of systems converters.

Phase Correlation is elegant as well as technically and conceptually complex. Its successful operation is derived by performing a Fourier Transform to each field of video.

A Fast Fourier Transform (FFT) is an algorithm which deals with the transformation of discrete values (in this case image pixels).

When applied to a sample of finite values, a Fast Fourier Transform expresses any changes (motion) in terms of frequency components.

What is the advantage of using FFTs over simply trying to predict the motion vector on a pixel by pixel basis?

- Mathematically, it's far easier and faster to recognize and process frequency signatures from which very accurate motion vectors can then be calculated.
- *Rather than having to measure where every pixel goes from frame to frame the FFT rather results in representing just the changes from one frame to the next.*

Since the result of the FFT represents only the inter-frame changes in terms of frequency distribution, there's far less data that has to be processed in order to calculate the motion vectors.

- Unlike other motion vector calculating methods, the FFT technique is not easily fooled by objects that have rotational or spiraling motions.
- What results from the FFT is a three-dimensional frequency distribution represented mathematically by peaks in a three-dimensional wave pattern.
- The 3rd dimension in this coordinate system represents subsequent fields of video.

In summation: Objects in motion can be mathematically correlated to their peaks in the frequency distribution. Once the FFT is performed it becomes a computationally simple matter for the computer to track just the peaks and assign them the appropriate motion

vectors. This conversion technique is both elegant and computationally involved. Sophisticated software and large amounts of processor "horsepower" are required for these complex computations.

### **DTV to analog converters for consumers**

A **digital television adapter**, (CECB), or digital-to-analog converter (box), is a device that receives, by means of an antenna, a digital television (DTV) transmission, and converts that signal into an analog television signal that can be received and displayed on an analog television.

These boxes cheaply convert HDTV (16:9 at 720 or 1080) to (NTSC or PAL at 4:3). Very little is known about the specific conversion technologies used by these converter boxes in the PAL and NTSC zones.

Downconversion is usually required, hence very little image quality loss is perceived by viewers at the recommended viewing distance with most TV sets.