

Professional Audio and Video

A guide for Media Students

Dr David Ellis; Dr Paul Otterson



Dr Paul Otterson & Dr David Ellis

Professional Audio and Video

A guide for Media Students



Professional Audio and Video: A guide for Media Students

1st edition

© 2017 Dr Paul Otterson & Dr David Ellis & bookboon.com

ISBN 978-87-403-0994-2

Peer reviewed by Dr. Karl O. Jones, School of Engineering, Liverpool John Moores University

Contents

Foreword	9
Introduction	11
Part 1 Acquisition	13
1 Image Acquisition	14
1.1 Light	14
1.2 Lenses	17
1.3 Shutter Speed	25
1.4 Exposure	27
1.5 Sensors	29
1.6 Frames	31
1.7 Timecode	35



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang

2	Audio Acquisition	36
2.1	Sound	36
2.2	Microphones	42
2.3	Systems (Balanced and Unbalanced)	46
2.4	Connectors	46
2.5	Speakers	48
3	Scene and Studio Lighting	49
3.1	Three Point Lighting	51
3.2	Types of Lighting Fixture	52
3.3	Ancillary Equipment	55
3.4	DMX Lighting Control	63
	Part 2: Processing	65
4	Power, Signals and Measurement	66
4.1	Power	66
4.2	Audio levels	68
4.3	Video Levels	70
4.4	Measurement and Calibration	72
4.5	Synchronisation	77



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

5	Video Sampling	78
5.1	Bits and Bytes	78
5.2	Composite and Component	80
5.3	Sampling	83
5.4	SDI interface	94
5.5	Digital Video Compression	96
5.6	Profiles and levels	110
5.7	MPEG -4	111
6	Audio Sampling and Compression	113
6.1	Pulse Code Modulation (PCM)	113
6.2	Frequency	114
6.3	Digital Audio Standards	120
6.4	Digital Audio Interface	122
6.5	Audio Compression	123
6.6	Psychoacoustic Masking	126
6.7	Compression Systems	127

The advertisement features a night photograph of the Apollo Hotel 1 building. A large red lightbulb icon is on the left, with the text "CISO Conference" and "Produced by Inspired" next to it. On the right, a white box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and "Dec 5th 2019". Below the main image is a white banner with the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right is the "Inspired" logo.

Part 3: Delivery	132	
7	Media Streaming and Storage	133
7.1	Stream Creation	133
7.2	Network Connectivity	135
7.3	Media Streaming Servers	137
7.4	Storage	139
8	TV Transmission	143
8.1	Overview of Broadcast Signal Chain	143
8.2	Signal Routing & Compatibility	144
8.3	Effects: Electronic Graphics, Colour Keying	155
8.4	Distribution	159



Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

Dedication

The authors would like to dedicate this book in memory of Tony Moore, a friend and colleague at Liverpool John Moores University, with whom they worked for 15 years and whose long-standing experience at BBC Research & Development (and elsewhere) helped grow and sustain the university's Media Engineering group.

Foreword

The pace of change in the media market seems to have increased exponentially. From the first demonstrations of television to the introduction of regular colour broadcasts took around 40 years. New technologies such as teletext came along only around 5 years later. With analogue multichannel television in the 1980s and the move to digital television in the 1990s the growth in TV channels has been enormous. In the late 90s most people in the UK only had 4 TV channels and a few radio stations available. Now dozens of channels are standard and many hundreds not unusual. Already, this is being considered “old media” with the advent of Internet-based streaming channels. With YouTube, Netflix and BBC iPlayer, to name a few, the method by which people consume content is in a process of significant change.

It has famously been pointed out that “content is king”. The change in delivery methods does not change that reality. Consumers are more interested in what they can watch, rather than how it is delivered. However, the channel choice offered by each move forward in technology means that more content is required and better production values are expected.

The driver for each leap forward is technology. Without advances in digital compression technologies, for example, the requirement for more content and improved production techniques would not exist. With the proliferation of channels and delivery mechanisms, the need for people (both engineers and content creators) in the broadcast industry who understand the underlying technology continues to increase. Without such qualified people, programmes cannot be made and content will not be distributed.

I have been working in broadcast engineering since the late 1990s and in this time we have moved from analogue television, to digital TV, through advances and experiments in interactive TV to high definition. All of that happened in a little over 5 years and was driven by continuing advances in broadcast engineering. The future looks to offer ultra high definition and more services delivered over the Internet to devices as small as a mobile cell phone up to huge televisions.

At the same time, the way content is captured and edited is changing. Production facilities that were once installed in huge air conditioned rooms and operated by men in white coats, are now shrunk to the size of a laptop, allowing for a much quicker and simpler turnaround of programmes. For all of this to happen, the production chain needs to be refreshed and then staffed with suitably qualified people who can make sense of these changes on behalf of those around them. Those that understand everything from the basics of computer networking through to the workings of the latest studio hardware will be highly valued, whether their primary interest is in delivering the technology or in utilising it as a vehicle to create and deliver their potentially award-winning content.

I trained in Broadcast Engineering at Liverpool John Moores University and was fortunate enough to study under Tony Moore, to whom this book is dedicated. Tony was an ex-BBC research engineer who had been involved in pioneering work in digital television, before many people even had a colour television. He was a true genius, and yet a very humble man. His lectures were a delight, especially for those of us who shared his dry sense of humour. The best thing was his enthusiasm to give advice at any time. He had a genuinely open door policy which I would take advantage of, often for hours at a time causing me to miss subsequent lectures as we would digress into so many areas of discussion. He is sorely missed.

Dave Ellis, co-author of this book, also delivered lectures for the course. His hand waving and somewhat eccentric style of delivery made even the most dull subject areas interesting. I always enjoyed lectures in the large lecture theatre on the 5th floor, where Dave would wander around the room, pausing mid sentence to walk through the old projection room at the back, and then continue as if nothing had happened a few seconds later as he emerged. His passion for ensuring the next generation of production teams and engineers are well informed and ready to meet future challenges has led to the creation of this incredibly helpful and easy to follow book.

I started my career with Sky Television before moving to the BBC in 2001. In that time I've worked on everything from the launch of Freeview, digital switchover and now in the BBC World Service, the launch of TV programmes in multiple languages, looking at new ways to distribute our radio programmes and I regularly face the challenges of running projects in countries with very different technical expectations. During this time I've had great training and been involved in training others.

Working in the broadcast world means constantly learning and developing new skills. This keeps it interesting and means that you always have to be one step (or more) ahead of what is coming to the consumer and even ahead of those in the content creation world.

Paul and Dave's time and effort in writing this book is a great help in ensuring the broadcast industry will continue to be equipped with skilled people for years to come. I hope that in reading this book, you will find the knowledge you need that will allow you to be part of the next revolution in the world of professional media.

David James
Distribution Technology Manager, BBC World Service

Introduction

All too often media studies extend from a background which has, primarily, been based in the arts faculties. Media-based subjects, as taught in schools and colleges, are often seen as non-technical and purely 'artistically' creative. Those who have worked 'behind the scenes' will soon disabuse them of such notions. Creativity in rapidly solving technical problems is vital; a good artist knows his or her tools and understanding them helps to wield those tools effectively.

Many students progress with their studies, developing techniques in analysis and the ability to write well-structured essays to open-ended questions which might examine various artistic aspects of film or radio that are felt to be appropriate to those who want to work in the media industry. Whilst the development of research and thinking skills such as these is to be encouraged and commended, it is a sad reflection on a certain viewpoint which leads people to believe that this is all that is necessary for those wanting to break into the industry. As one of the authors of this book has often told his own students; "*Guess what – if you do get to work in this fun and fascinating industry, the chances are that nobody there will ever ask you to write an essay! You need to know how and why to use which kit and how to use it properly!*"

Whilst it is true that students may wish to enter a field in producing, directing, lighting, camera or audio work, it is equally true to say that, no matter how brilliant their artistic intent might be, it is absolutely essential to their success that they know and understand both the limitations and the opportunities that are presented by the appropriate selection and correct technical operation of the equipment, to bring their project to life. The finest and most brilliant artistic intent can be enabled through appreciating the inextricable link that must be built to the technical equipment which can enable their vision or, conversely, can dash their dreams on the rocks of insufficient technical understanding, leading to a failure to work to the required standards and consequent rejection of their content by the client at the technical acceptance testing stage.

The purpose of this book is to introduce the typical media student to some of the basic technical aspects and issues of media production. Media is a wide subject area, so we have taken the viewpoint that students will primarily be focused on the technology that underpins audio-visual productions, namely video, TV, and radio. This book will, therefore, be of use to college and university students studying a variety of media subjects including audio production, videography, television production and media streaming.

The book is split into three parts, Acquisition, Processing and Delivery. Within acquisition the student is introduced to the processes behind audio-visual media capture and the basics of lighting. Processing covers the basics on compression, (both audio and image), as well as a basic understanding of power, signals and measurement. Finally delivery covers the storage and emission of media content.

Throughout, attempts have been made to keep the maths content to a minimum. Where possible, processes are described and illustrated with diagrams and graphs to try to explain the basic concepts that will enable students to understand how to get the best from their technical equipment.

Between them, the authors have over 50 years' experience in designing, running and lecturing on degree courses which have included such topics as Broadcast Engineering, Media Production, Audio & Music Production, Electronics, e-Business & Technology Management. The authors are also involved in the running of their own video production company, and have also produced specialist video and audio material ranging from dramatic to factual genres for organisations as diverse as the BBC, telecommunications companies and local history groups as well as scoping specialist high-speed video acquisition systems for industrial research units.

In the last 10 years they have been particularly focusing on programmes blended between engineering and arts faculties such as Broadcast & Media Production and Audio & Music Production, on which their knowledge of the levels of understanding and abilities of students arriving with a typical media background are based. Many of their former students now work in well-known broadcasters such as the BBC, SKY and ITV, amongst many others.

We would encourage students to use this book as a compendium primer and to read further into each topic area as required.

Dr. Paul Otterson

Dr. David Ellis

February 2017

Part 1 Acquisition

1 Image Acquisition

In this chapter, we cover how an image is focused, exposed and captured within a camera or camcorder.

Learning Outcomes :

- To explain how a lens functions and how it is specified by measurement of focal length.
- To explain the relationship between the focal length, aperture of the lens and exposure to light at the sensor.
- To give an outline of the image acquisition and control devices typically found on cameras and camcorders.

1.1 Light

Definition: Light is the electromagnetic radiation of photons which are in a range visible to the human eye. Classically, light can be measured as a wave, having amplitude (brightness or luminance) and frequency or wavelength (hue), the shorter wavelengths (higher frequencies) being seen as blue and longer wavelengths (lower frequencies) as red.

Light is either radiated (emitted) from an object such as a lamp or the sun (sometimes referred to as incident light), or it is reflected from an object that is being illuminated by the incident light. Consequently an image in a field of view may be composed of objects which have differing light levels since the amount being reflected will always be less than the incident light.

'Ambient light' is a term used to describe the light level without the inclusion of artificial lighting being added to the scene by the photographer or videographer.

1.1.1 Colour Temperature

In photography and videography, the colour of the radiated light is measured by comparison with the temperature to which a perfect black body radiator would need to be heated so as to glow with light matching that of the light source in question. The Kelvin scale (see figure 1-1) is used and, for video acquisition purposes, has daylight defined as being at 5600K. The higher the temperature the 'cooler' (or more blue) the light so a 'daylight fluorescent light' is 6500K, this being the same colour temperature as the display monitor. The lower the colour temperature the 'warmer' (or more red) the light so a room with standard interior video lighting may be down to 3200K. As the colour temperature changes, the use of colour correcting filters and additional lighting may be required to make the image colour true to the subject or mood desired.

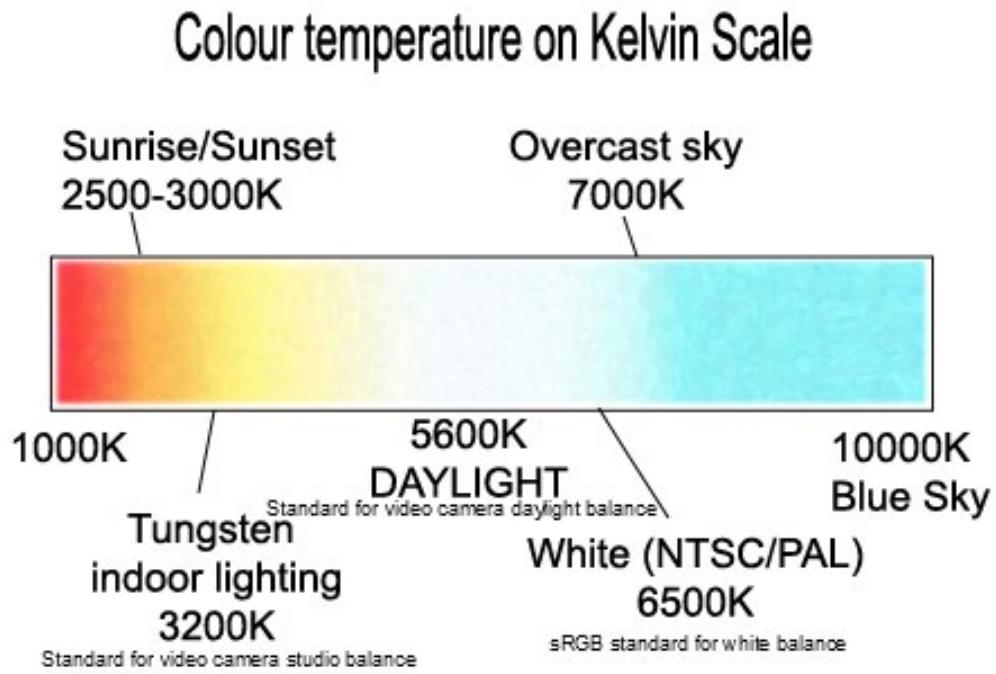


Figure 1-1: Kelvin scale



MTHøjgaard

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang

1.1.2 Light Measurement

Reflected light is measured by the illumination of the object on a Lux scale, Lux being a measure of Lumens per square metre (Lumens/sq.m.). A full moon in a cloudless sky gives around 1 Lux. Often a camera/camcorder has a specification of the lower limit Lux value it can capture. This is useful to know as it determines the low-light sensitivity of the camera and so how likely it is to need additional lighting, or a lens with wider opening aperture for a dark environment.

A point of caution however; no matter what figure is quoted by the manufacturer as the lowest Lux level a camera can cope with, a poorly illuminated subject will invariably result in a ‘noisy’ image. It is wise to treat the figure merely as an indication of how sensitive the camera is relative to other models, rather than as a useful indication of capability.

Light Value (LV) (Approx.)	Lux Value (Lx)	Typical Conditions
(-15)	0	Total Darkness
(-8)	0.0001	Starlight (no moon)
(-5)	<=1	Full Moon (no cloud)
7-8	400	Sunrise / Sunset (no cloud)
7-8	350 to 500	Office Lighting
12	1000	TV studio lighting
15	10000 to 25000	Daylight
17-18	32000 to 100000	Direct sunlight (no cloud)

Figure 1-2: Lux and LV, an approximate comparison.

Lux should not be confused with light value (LV) – see figure 1-2. Light Value is a photographic term for a measure of the luminance light level within a scene. LV(0) is defined as light level required for a 1 second exposure at an aperture of $f1$ on a sensitivity setting of ISO-100. LV is the ‘absolute’ value of light level reflected from the scene as viewed by the camera. On the other hand, Exposure Value (EV) is a ‘relative’ scale which enables us to relate different combinations of shutter speed and aperture settings in the camera (i.e. the exposure) to take account of the sensitivity of the sensor, for any given value of LV. LV = EV at ISO-100.

Note: Since Exposure Value is a function of shutter speed, aperture and sensitivity (ISO), a variety of these settings will achieve the same EV value. For instance, for the same setting of ISO-200, 1/250 sec at $f8$ gives the same EV as 1/500 at $f5.6$.

1.2 Lenses

Definition: a 'lens' is usually an optical quality glass disc that is ground into a convex shape with a refractive quality so it can bend a ray of light to a precise point without distortion. When we commonly talk about 'lenses' we are usually referring to a compound set of glass lenses, adjustable barrels and rings that form a single device, such as a zoom, a wide-angle, or a telephoto lens.

1.2.1 Focal Length

Definition: This is the measure of the distance (usually in millimetres) between the optical centre of a lens and the focal plane. The focal plane is at the point where the image is captured in focus, and so it is where the sensor or film is situated within the camera or camcorder.

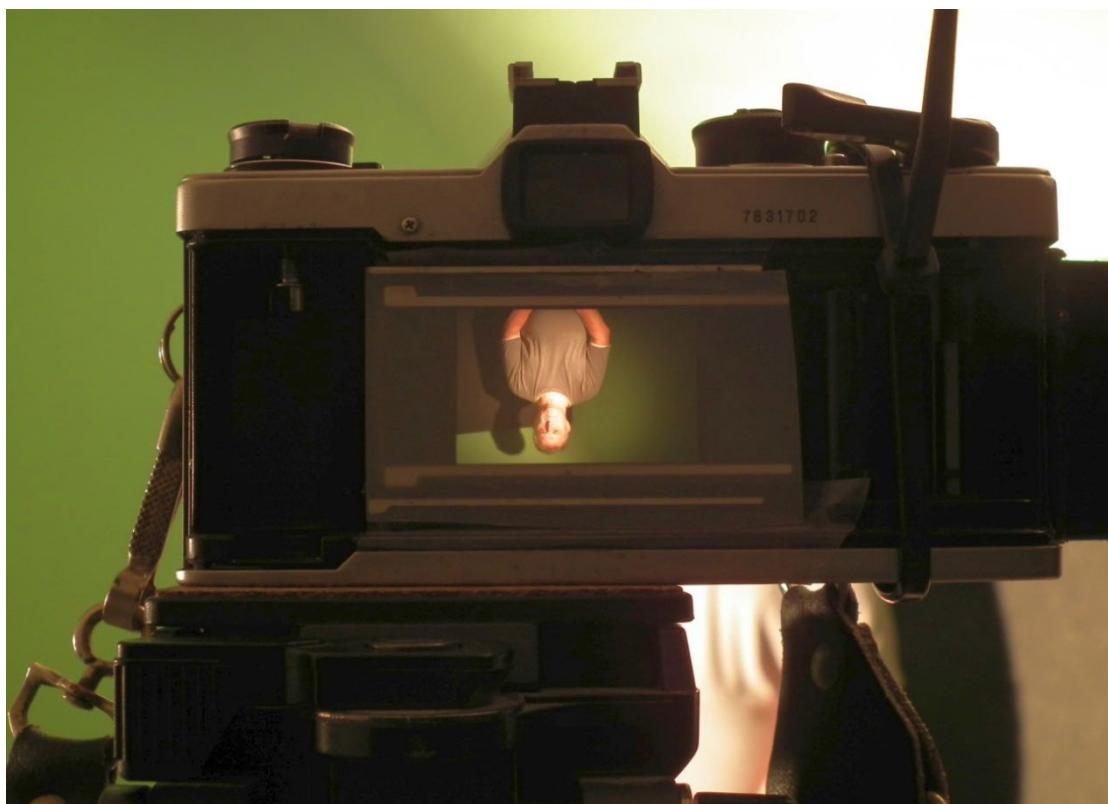
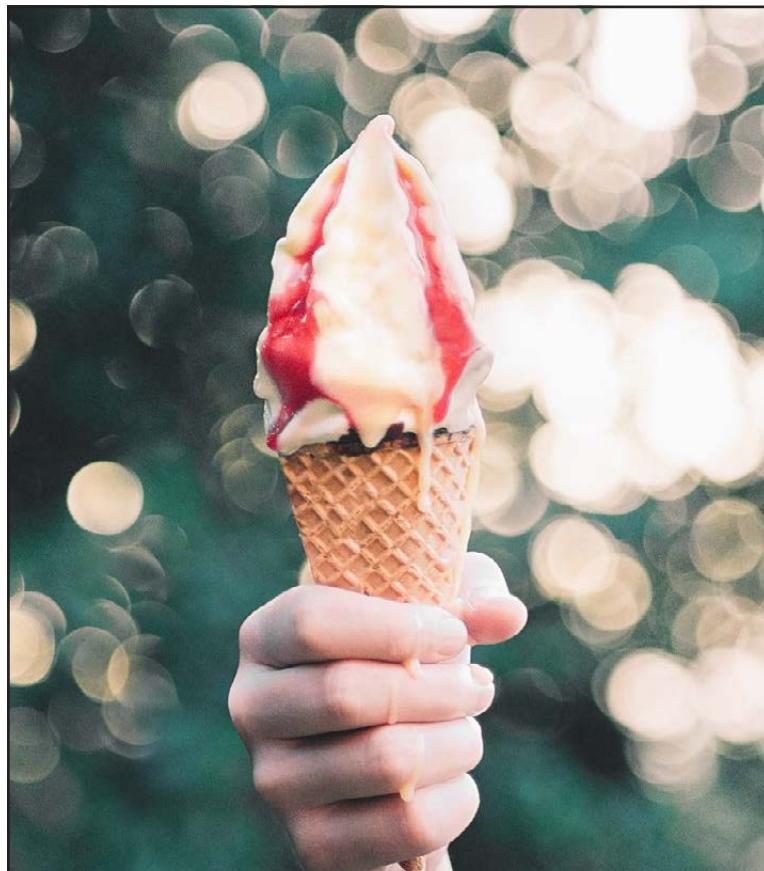


Figure 1-3: Focal plane shown on a camera where the sensor/film would be situated

Lenses come in two basic forms, ‘fixed focal length’ (prime, normal) and ‘variable focal length’ (zoom). Strictly speaking a ‘telephoto’ lens has a physical length shorter than its focal length; it means ‘long focus’ but common usage usually refers to a lens with a focal length greater than a standard lens and, hence, a narrower angle of view. A ‘standard’ lens is a fixed focal length lens with a diameter that is close to its focal length (i.e. a ‘fast’ lens – one where a wide aperture is possible), the focal length generally being close to the diagonal dimension of the sensor, or a little larger. (e.g. on a traditional 35mm film camera (for a frame of 24×36mm, giving a diagonal across the frame of 43mm), the standard lens was 50mm providing a field of view similar to that of a human.). A ‘prime’ lens is a fixed focal length lens with better optical qualities than a zoom as it doesn’t have to deal with focal length variation. A ‘wide angle’ lens has a focal length smaller than a standard lens so it sees a wider scene. ‘Macro’ lenses magnify objects for close-up imaging. A ‘tele-conversion’ lens (a.k.a. tele-converter) is an additional lens fixed to an existing lens to reduce the angle of view and, hence, increase the magnification of a distant object (or a very near object if attached to a ‘macro’ lens). Similar ‘wide angle converters’ are available. In both cases, the combined optical performance of a lens plus converter is rarely as good as a lens designed to achieve the same focal length as the combined pair, though the technique can be useful in situations where no suitable lens is available and an important shot would otherwise be missed.

As a basis for comparison, lenses are often compared with those developed for a full-frame sensor camera having the equivalent image size to that found in 35mm film of older cameras. We will deal with sensors, sensor size and crop-factors later in the chapter.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

Why an object is inverted on a focal plane

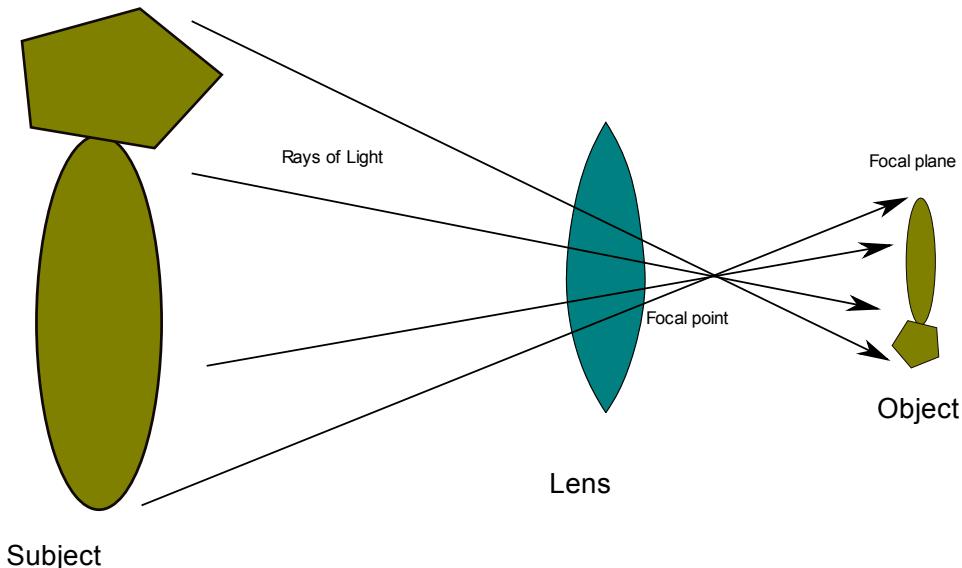


Figure 1-4: why an object is inverted on a focal plane

In a simplified diagram (figure 1-4), an image is seen by reflection of rays of light which are refracted (bent) when passing through a convex shaped lens. The point at which the image carried by these rays forms a focused image is the ‘focal plane’, and so this is where the capture device (sensor or film) is situated. Traditionally lenses are specified by their focal length; consequently a 50mm lens on a full-frame sensor camera refers to its focal length (not the diameter of the lens).

1.2.2 Aperture (Iris)

Definition: Aperture means an opening, in this case one which allows rays of light through it to illuminate the focal plane. The size of the aperture is controlled by means of an ‘iris’, which effectively masks part of the lens (see figure 1-5). An ‘iris’ is a mechanical device which acts in a similar way to the iris in your eye.



Figure 1-5: Iris front

1.2.2.1 *f*-stop

The aperture is measured in relation to the geometric size of the opening made by the iris and the focal length of the lens; this is known as the 'f-stop'. The *f*-number(*N*) is given in *f*-stop notation by the following equation.

$$N = \frac{f}{D} \text{ where } f \text{ is the focal length and } D \text{ is the diameter of the aperture.}$$

Example: if the focal length is 100mm and the aperture diameter is 25mm then the *f*-stop is *f*4.

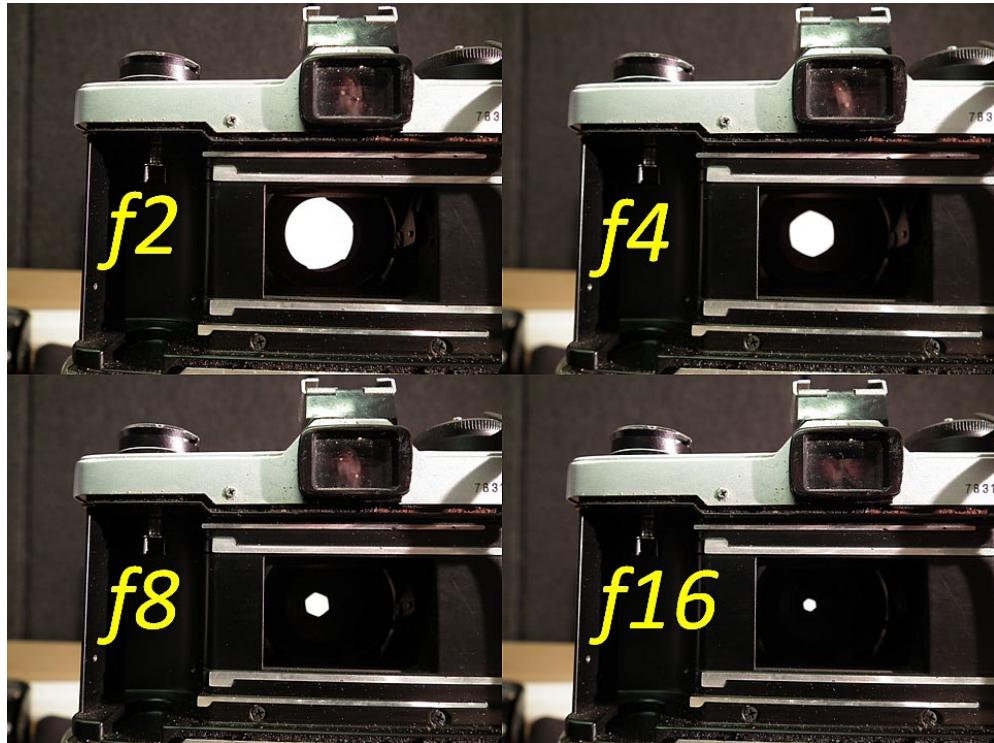


Figure 1-6: f-Stops

The advertisement features a large image of the Apollo Hotel building at night. The hotel's name is visible on the sign above the entrance. Overlaid on the image are several text elements and logos. On the left, there is a red lightbulb icon followed by the text 'CISO Conference' in white and 'Produced by Inspired' in red. On the right, there is a white rectangular box containing the text 'Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL' in black, and 'Dec 5th 2019' in blue. At the bottom, there is a call-to-action text 'Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders' in white, and the 'Inspired' logo with a blue lightbulb icon.

Simply put, *f*-stops are set at the points where the aperture halves in size; these are major stop points. As the aperture is an *area* of a circle then $A = \pi r^2$ and each stop is the square root of 2 raised to the next power.

$x=0$	$x=1$	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$	$x=7$	$x=8$	$x=9$	$x=10$
$N = 1/(\sqrt{2}^0)$	$N = 1/(\sqrt{2}^1)$	$N = 1/(\sqrt{2}^2)$	$N = 1/(\sqrt{2}^3)$	$N = 1/(\sqrt{2}^4)$	$N = 1/(\sqrt{2}^5)$	$N = 1/(\sqrt{2}^6)$	$N = 1/(\sqrt{2}^7)$	$N = 1/(\sqrt{2}^8)$	$N = 1/(\sqrt{2}^9)$	$N = 1/(\sqrt{2}^{10})$
$N = 1/1$	$N = 1/\sqrt{2}$	$N = 1/\sqrt[4]{4}$	$N = 1/\sqrt[8]{8}$	$N = 1/\sqrt[16]{16}$	$N = 1/\sqrt[32]{32}$	$N = 1/\sqrt[64]{64}$	$N = 1/\sqrt[128]{128}$	$N = 1/\sqrt[256]{256}$	$N = 1/\sqrt[512]{512}$	$N = 1/\sqrt[1024]{1024}$
$N = 1$	$N = 1/1.4$	$N = 1/2$	$N = 1/2.8$	$N = 1/4$	$N = 1/5.6$	$N = 1/8$	$N = 1/11.3$	$N = 1/16$	$N = 1/22.6$	$N = 1/32$
$f1$	$f1.4$	$f2$	$f2.8$	$f4$	$f5.6$	$f8$	$f11$	$f16$	$f22$	$f32$

Figure 1-7: f-stops table

1.2.2.2 T-Stop

The ‘T-Stop’ (Transmission Stop) is an adjusted *f*-stop. The coating and glass used in a lens will have some effect on the amount of light it can transmit, so less than 100% of the light that enters the lens will emerge from it. The T-stop for a lens will be the *f*-stop modified by a function of the transmittance value of the lens, and the adjustment will need to make the aperture larger to compensate to the same exposure value. As T-stops are rarely encountered except on high-end cinematographers’ cameras, they are mentioned here merely for completeness.

1.2.3 Lens Speed

This is a relative measure of how quickly the desired exposure value is reached within the lens. Simply put, lenses with a larger aperture diameter will be ‘fast’ in comparison to other lenses such as zoom lenses which will tend to have a smaller maximum aperture. A fast lens enables a stills photographer to use a shorter exposure time or ‘fast shutter speed’. However, as we will see, for video, a shorter exposure time can create problems with moving images so the advantages of a ‘fast lens’ are associated more with the ability to shoot in lower light situations and/or with shallower depth of field. A prime lens of 50mm at *f*/1.4 would be considered a fast lens.

1.2.4 Depth of Field (DOF)

Definition: The depth of field is a distance range that forms an acceptably sharp image. The range covers the distance extended both in front of and behind the object of focus within the field of view.

Depth of field relates the aperture, focal length and distance over which objects are clearly imaged.

A ‘shallow’ or ‘narrow’ DOF (figure 1-8) has a close-up object in the field of view but blurs the background. This is done with a wide aperture and a short distance to the object in the field of view, e.g. portraits. The ‘Bokeh’ effect is often used to create an abstract out-of-focus background (in conjunction with a sharp focused foreground subject) or completely abstract picture pleasing to the eye. Beware, however, of ‘jumping on the bandwagon’ of using Bokeh as a frequently-used technique. As with all special effects, it should be used very sparingly rather than frequently and only if there is a specific reason to do so.



Figure 1-8: Shallow DOF

A ‘long’ or ‘deep’ DOF (where everything in the picture is in focus – figure 1-9) uses a small aperture and a longer distance between the camera and the object in the field of view eg. landscapes. DSLRs in auto mode and camcorders having smaller sensors typically default to long DOF through the need to ‘stop down’ in order to avoid over-exposure, unless alternative techniques are used to avoid this.



Figure 1-9: Long DOF

A long DOF is particularly useful when shooting multiple objects distanced apart. The pre-digital (35mm film camera) lens method for obtaining an appropriate depth of field, was to focus on the farthest subject first and note its distance and then on the nearest subject next. The mid-point on a scale between these two distances gives the focal setting and lines engraved on the lens barrel would show which aperture must be selected in order to ensure that objects at both distances remain in focus. It will also be realised that careful manipulation can be achieved so as to set the farthest point at infinity by setting the focus slightly short of infinity. This will achieve the greatest depth of field for a given f-stop and the distance to the closest point which is sufficiently sharp is then referred to as the 'hyperfocal distance'.

When taking a still picture on DSLRs with an autofocus, DOF is controlled via a focus lock. The method is to point at the distant object, half press the shutter button, and keeping the finger on the button, recompose the picture with the nearer subject the point of focus, then fully press the shutter. Effectively the camera will use an exposure which chooses an aperture that allows for both near and far subjects to be in focus. This should not be confused with the auto-exposure lock which allows for the same exposure settings to be used across shots. As is usually the case, having equipment that is sufficiently professional to enable manual selection of settings (and knowing what to do with them) produces a more controllable, higher quality result, especially so when it comes to video. The size of the image sensor also plays its part in DOF. Camcorders have traditionally had small sensors and were, therefore, good at long DOF, but recent entries into the market with swappable lenses sometimes have larger sensors, similar to those of DSLRs, making 'shallow' DOF a possibility.

1.3 Shutter Speed

Definition: This is the proportional time the film or sensor is exposed to the light coming through the aperture. Traditionally the shutter in a film camera was a mechanical device that masked the light entering via the aperture before it reached the focal plane.

In order to maintain a constant exposure (see below), the principle of 'reciprocity' is used. This means that, as the area of the aperture is doubled (i.e. opened up by one stop), the length of time over which the light strikes the sensor (or the electronic sensor is active) must be halved.

Since exposures on traditional film-based stills cameras would often be quite short (e.g. 1/250 sec), timings on film cameras were based around halving or doubling of this time period, for example 1/125, 1/250, 1/500. However, modern video cameras have to take account of the frame and field rate (25 frames/sec, 50 fields/sec) so sometimes tend to use a scale which matches more usefully with a halving of those values, e.g. 1/100, 1/200, 1/400 etc. The principle remains the same, however.

The term 'shutter' used to refer to a mechanical device that opened and closed to expose the film. Strictly speaking this isn't necessary with a digital camera, but, in principle, it could still be used for 'still' photographs on DSLRs, though is rarely encountered. Camcorders and DSLR cameras (in video mode) mimic the function of a shutter by electronically controlling the time over which the sensor is electrically active to the reception of light and so captures the image by scanning the sensor.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

For stills cameras, the faster the speed of the shutter (i.e. the shorter the exposure time) the greater the freezing effect on the action within the image being captured, thereby reducing blur. High speed action required a faster setting, low speed action a slower shutter setting (to the point of requiring a tripod or other method of camera stabilization to reduce blur from shaking or vibration). However, for video recording of motion, the use of a high shutter speed (short exposure) will result in an increase in the observation of distracting flicker, as movement of the subject or background (relative to the fixed screen frame) increases. Thus, for video applications, the shutter speed is inextricably linked to the video frame rate. Ideally, for maximum image quality, a frame rate higher than the eye's temporal response would be used, along with a shutter speed which is the reciprocal of that value (e.g. 100 frames/second could use a shutter speed of 1/100s). However, whilst experimental development of ultrahigh definition and high frame-rate systems continues, the traditional values of 25 or 50 frames/second are the most likely to be encountered by most users. Hence, the fastest shutter speed that would normally be selected is 1/50 second. (A speed well within the temporal response of the eye, such as 1/25 sec, would result in a very 'smeary' result and should be avoided.) Users should beware of some 'domestic' camcorders where automatic control of the shutter speed cannot be over-ridden. Bright situations will then cause the camera to choose a faster shutter speed, leading to an increasingly flickery result.

A common problem with video in earlier CMOS cameras was called 'rolling shutter' (see figure 1-10), where if a panning movement was too fast for the scanning process (or an object traverses a static field of view too quickly), it resulted in an image which looked slightly skewed when played back. CMOS sensors differ from CCD in the way they read information off the sensor. CCD use a global shutter; the image is captured on the sensor instantaneously then read off. A CMOS sensor uses a rolling shutter which is akin to allowing the camera to read each line of the sensor array sequentially. So by the time the last line is capturing its part of the image, it is slightly later than when the first line was recorded and the object has moved. It is this delay that causes skew or distortion of objects moving more rapidly than the speed of the shutter can cope with. Rotational devices can be particularly problematic, leading to 'spatiotemporal aliasing'. This can be seen as a strobing effect where (for example), propeller blades appear to break up into strips. This is a more complex form of aliasing than the simple 'wagon wheels appearing to turn backwards' on cinema film but both forms stem from a compromised frame rate which is too low to be able to capture the reality of the motion with sufficient accuracy.

Rolling Shutter Effect

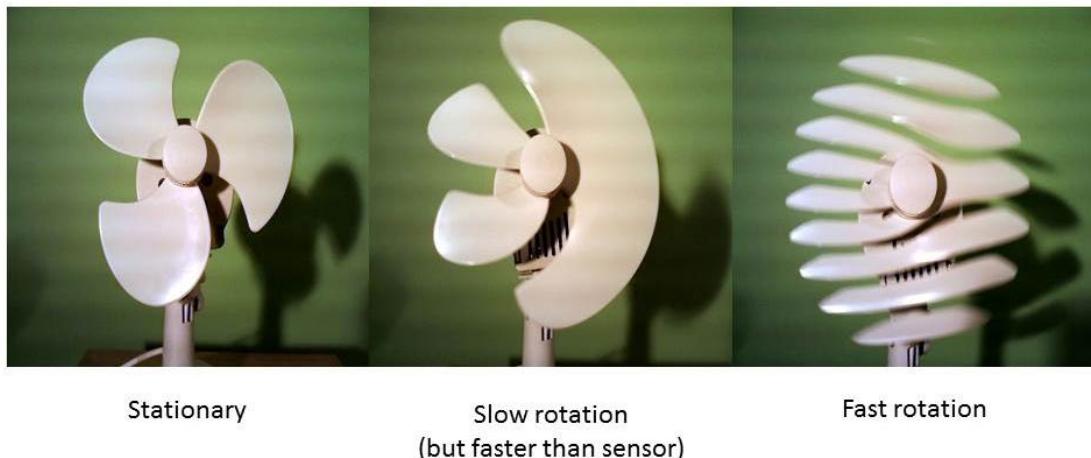


Figure 1-10: Rolling Shutter

1.4 Exposure

1.4.1 Sensitivity (ISO)

Definition: ISO is a setting used to calibrate the sensitivity to light of a film or a sensor within the camera. The International Organisation for Standardization creates and publishes harmonised technical standards throughout the world, with the prefix ISO.

The ISO settings on a camera consist of a geometric scale that doubles; the higher the ISO number the more sensitive to light it is. The settings originate with film where the sensitivity was rated by ISO, ASA or DIN scales (the International, American and German standards bodies, respectively). These have given way to just the ISO scale that is based on a norm of 100 but can result in values both above and below this reference.

With digital capture through sensors, the familiar ISO standard has been kept, but the lowest ISO setting on the camera refers now to the sensitivity of the sensor before any gain is applied to the signal. Effectively, as light on the sensor is converted to an electrical signal it can then be amplified (gain), this amplification doubling each time the ISO setting is incrementally raised. Consequently a change up to a higher ISO setting does not allow any more light onto the sensor (for example, by opening the iris further); instead it artificially enhances the captured image which will allow the camera to work in darker conditions but will increase the noise in the image. (This is the equivalent of using a 'higher speed' film in a film camera where there was more sensitivity to light but at the price of more visible 'grain' in the film.) When a signal is amplified, so too is the electrical background noise which has the effect of bringing grain and aberrant pixel colours into the image.

1.4.2 Gain

The sensitivity of the image capture chip in a camcorder is fixed during manufacture. As mentioned above, the signal level is electronically enhanced by altering the 'gain'. To increase the sensitivity to double the ISO equivalence (i.e. the electronic equivalent of 'opening up a stop'), then a gain of +6dB is added, which doubles the voltages. Thus, switching to +18dB will result in an otherwise inadequately low voltage coming from the sensor (owing to poor lighting) being boosted by a factor of 8 times, which may then result in satisfactory video levels, though the noise performance at such high gain would be noticeably poor. Gain adds noise to the signal which shows as 'grain' on the image. The decibel scale is logarithmic and is often used in electronics as it is a measure of the relative levels involved, rather than absolute values. Thus, a +6dB gain will always be a doubling of voltage, whatever the starting voltage may be. It can also be used to express the measure of the signal to noise ratio or SNR. We want the SNR to be as high as possible i.e. the wanted signal to be much bigger than the tiny noise and, again, this is a useful figure to look out for when specifying a camera.

1.4.3 Neutral Density Filters

Neutral Density (ND) filters are used where there is a need to reduce exposure in very bright conditions, but where the alternative strategy of stopping down the iris or using a higher shutter speed (i.e. shorter exposure), would produce an image with an overly long depth of field or with flicker, respectively. Conversely to the case with 'gain' (where it was necessary to artificially increase the electronic signal so as to compensate for poor lighting), to reduce over-exposure a 'Neutral Density filter' is applied. A true ND filter consists of glass having a reduced optical transmission but no colour cast and can often be dropped in (or omitted) between the lens and the sensor within the camcorder's body. Additional ND filters can be screwed on to the front of the lens, if required (as can effects filters). On professional video cameras, a small range of ND filters will be fitted as standard, allowing for several steps equivalent to stopping down the iris over perhaps 4 stops.

Applying an ND filter of $\frac{1}{2}$ (ND2) will reduce the transmittance of light by 50% effectively acting as one *f-stop*.

The same effect is achieved in some cameras by electronically attenuating (reducing) the signal coming from the light sensor, this being the opposite of gain. However, this relies on the sensor having sufficient 'dynamic range' to be able to cope with the extremely bright incoming light without suffering from 'white crushing'. Where such saturation occurs, the video levels may still be within allowable electrical limits but the image will suffer from severe distortion (i.e. items which are supposed to be moderately bright will appear as completely white). Optical filters are, therefore, the preferred method and tend to be found on the more professional cameras.

1.5 Sensors

Definition: The sensor is the device that has taken the place of film in a digital camera or camcorder. The two types of sensor are the older 'charge-coupled device' (CCD) or 'complementary metal-oxide semiconductor' (CMOS). The sensor actually consists of millions of tiny picture elements (or 'pixels') laid as an array, each of which converts light into a small electrical charge (electrons) which is read out as voltage before being converted to a digital value. The 'array' pixel count is usually in Mega-pixels (millions of pixels) and describes the definition of the still image.

DSLRs and some of the more cine-like modular camcorders now on the market tend to have one large sensor array, with pixels sensitive to the primary colours (red, green, blue) arranged in a pattern to provide adequate resolution of the colour image. Professional video cameras, however, tend to rely on 3 separate sensors for the red, green and blue channels and the light path must be directed using a trichroic splitter block, which separates out the incoming light into three separate paths as well as filtering them into the three primary colours using a succession of dichroic mirrors. Thus, 3 identical sensors permanently attached to the appropriate faces of the prism will receive 3 versions of the original image, one red, one green and one blue, each of which is converted to the appropriate electronic signal. In either case (single sensor or 3-chip), the result is 3 separate signals, representing the image in primary colours, which can then be processed into a variety of signal forms for onward transmission.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang

1.5.1 Sensor Size

The size of the sensor is referenced in comparison to a ‘full frame’ which is the equivalent size of a 35mm film camera frame (circa 24mm × 36mm). The 35mm frame is the standard used to measure sensor size equivalence. Sensors that are smaller than a full frame are called a ‘cropped sensor’, this means if a lens from a full frame camera was used with a cropped sensor then the image captured would be cropped accordingly. Optical magnification (optical zoom) is by the use of lenses and so the image integrity is maintained. Digital magnification (digital zoom) enlarges by duplicating pixels from a portion of the image sensor and, thus, is of inferior quality, resulting in ‘blocky’ images. Digital zoom should be avoided and optical zoom preferred.

1.5.2 Crop Factor

The crop factor is determined by the size ratio of the ‘cropped sensor’ to a full frame (see figure 1-11). A simple formula (crop factor × focal length) will give the equivalent size of lens focal length required on a ‘cropped sensor’ camera to match a lens on a ‘full frame camera’. For example a 50mm lens from a full frame camera would produce the equivalent view of an 80mm lens if used on a 1.6 crop factored camera.

Sensor Size and Crop Factor

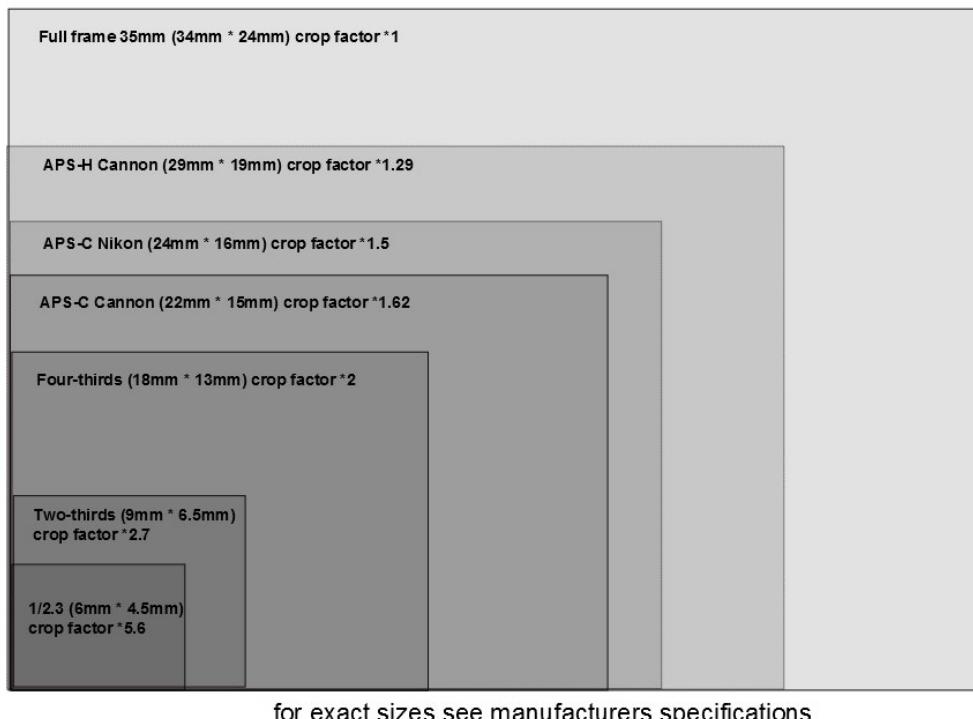


Figure 1-11: Sensor Size and crop factors

1.6 Frames

Definition: All video is a sequence of still images. The frame was a small part of the film exposed to light within the camera that was then processed as a celluloid negative, each frame being exposed in turn to form a sequence. In video, the principle of a frame remains but is made up from a series of discrete lines which are scanned from left to right, progressing down the screen. The frame consists of the entirety of both the visible part of all the scanned lines making up one still image in the sequence of images, plus parts of the frame which have picture detail blanked out, in effect overlapping the screen with an invisible border all the way round. The part of the frame containing picture information is often referred to as 'active video' while the surrounding parts are often referred to as 'blanking', both 'line blanking' (either side) and 'frame blanking' (above and below).

1.6.1 Frame Rate

When we look at a cinema film, we appear to be watching moving pictures. In fact, we are looking at a sequence of still pictures that are being shown so rapidly that they use the property of 'persistence of vision' in the eye/brain to fool us into thinking that the pictures are moving.

The speed of the sequence can vary from simple animation of around 12 frames per second to 25–30 frames per second (fps) for typical digital video recording. When using video recording of live action the 25–30fps is used and so playback is also set at the same speed as the recording fps. Assuming playback at normal frame rate, a slower fps recording speed causes speeded-up movement at playback (known as 'under-cranking' from the days of hand cranked film cameras). The reverse is also true; a faster fps recording speed causes a slower-motion at playback (called 'over-cranking'). The faster recording speed frame rate in comparison to the playback speed is the basis of 'slow-motion'. Cinema film on celluloid was traditionally shot at 24fps, this being the slowest speed that was barely acceptable for optically-recorded soundtracks placed alongside the frames for projection, whilst being the fastest that the mechanical processes could manage without ripping the film too often. It was (and remains) a compromise. There is little point in shooting video at 24fps as no TV systems are based on it. Frame rate for most of the world's TV systems are 25fps, whilst for the Americas and some Pacific-based countries, approximately 30fps is the norm. Before choosing a frame rate, it is essential that you consider carefully where you intend to use your material, choosing both wisely and consistently with all the material to be created for your programme. Ideally, your frame rate on all shots captured will always match the eventual frame rate for the region in which it is to be shown. For material to be shown on the web, the situation is less critical since the quality of web-based viewing platforms is significantly less controlled and most computer-based players tend to be 'agnostic' about the frame rate, handling all rates equally.

1.6.2 Frame Size

The frame size of an image depends on the aspect ratio used to capture it. There are 3 typical ‘aspect ratios’ commonly found in cameras and camcorders; 4:3, 16:9, and 3:2. An image captured in an aspect ratio that matches the print (or recording media) requires little adjustment or cropping. All the ratios are useable for still images, but 3:2 is closest to the original 35mm film, 4:3 for older analogue TV (NTSC and PAL) and computer formats (VGA, SVGA), and 16:9 for all current productions in HDTV. Modern computer screens cater for HD in 16:9 resolutions. Late Standard Definition productions (from roughly 2000 onwards) were often also at 16:9, especially in the UK, which adopted widescreen standard definition as a ‘stepping stone’ before High Definition services started.

For digital video there are 2 ratios used; traditionally 4:3 was used for standard definition (and there are still some countries which use this so, again, check the emission standard for your intended market), and 16:9, which is always used for high definition (HD). Legacy 4:3 standard definition video is now often converted to a 16:9 playable size, for modern HD device playback. This can be done either using aspect ratio conversion (ARC) in which the top and bottom of the frame will be lost when viewed on the 16:9 monitor, or the image is framed within a ‘pillar box’ format, with blank left and right borders.

A photographic still image from a camera can have a pixel definition which depends on the pixel array of the sensor, given in ‘mega-pixels’. Video pixel definition is dependent on both aspect ratio and the image definition used at capture.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

Width	Height	Usage	Display Aspect Ratio
720	480	NTSC Standard Definition – rectangular pixel shape – Width < Height	4:3
720	486	NTSC Professional digital video (Betacam, D1 and D5)	4:3
720	576	PAL Standard Definition – rectangular pixel shape – Width > Height	4:3
1280	720	High Definition video format – square pixel	16:9
1440	1080	HDV tape – rectangular pixel	16:9
1920	1080	High Definition video format – square pixel	16:9
3840	2160	4K Ultra High Definition TV	16:9
4096	2160	4K Digital Cinema Initiatives (for film projection)	256:135

Note : there are other 4K formats being used as this technology is in an early phase of development and standardisation of format is yet to happen.

Figure 1-12: Video Frame Sizes

1.6.3 Pixel Shape

Photographic stills, computer screens and High Definition video/TV use a square pixel for display.

Standard definition formats (PAL/SECAM/NTSC) use a rectangular (non-square) pixel. PAL/SECAM use a longer width pixel, whereas NTSC uses a taller height pixel.

HDV/HDCAM systems (1440×1080 a 4:3 aspect ratio) use a non-square pixel to fit a 16×9 display.

When creating graphics or stills for standard definition use, such as for DVD, care needs to be taken to ensure that distortion doesn't occur, such as ovals for circles and rectangles for squares. For example, a PAL DVD requires 720×576 frame, so it is a 4:3 aspect ratio only when played out as it uses a non-square pixel. The same applies to NTSC DVD but with a 720×480 frame.

When incorporating still pictures or graphics (bitmaps) a PAL frame equates to a 768×576 bitmap picture (4:3 aspect ratio – square pixel). Obviously the bitmap picture will then need to be saved in a 720×576 format (so it would look slightly squeezed from left to right), then it can be incorporated into the video and look proportional. NTSC would require a 720×534 bitmap picture, which will need saving at 720×480 (looking squeezed top to bottom). Standard definition PAL video played on wide screens (16:9 aspect ratio), will require a 1024×576 bitmap picture, again saved to a 720×576 format. NTSC would require an 864×480 bitmap picture. (For the sake of completeness, all of these values apply to the active (i.e. visible) part of the frame and take no account of the blanked parts of the frame which surround the visible part of the picture. Nor do they need to do so under normal circumstances.)

When playing an SD video footage through a computer display the pixel aspect ratio will need modification to fit a square pixel display (typically done by the computer software playing the video).

1.6.4 Anamorphic Recording

'Anamorphic lenses' were designed to utilise the whole of a 35mm film frame which would otherwise have been cropped when shooting widescreen, the aspect ratio is 2.39:1. It can be used on other film or video frame formats where a widescreen recording aspect ratio differs to the recording media aspect ratio. Anamorphic films need appropriate scaling when projected to revert to the true proportions of the image.

The advertisement features a night photograph of the Apollo Hotel 1 building. A large red circular logo with a white lightbulb icon is on the left. Text on the building reads 'APOLLO HOTEL' and 'CISO Conference Produced by Inspired'. To the right, a white box contains the text 'Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL Dec 5th 2019'. Below the main image is a white banner with the text 'Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders'. To the right is the 'Inspired' logo with a blue lightbulb icon. At the bottom right is a green oval containing a hand cursor icon pointing to the text 'Click on the ad to read more'.

'Anamorphic widescreen' or 'full height anamorphic' is a technique that utilises traditional 4:3 standard definition equipment to carry a 16:9 format image. It was popular in the early days of standard definition wide-screen video as it enabled the existing infrastructure to be utilised for widescreen production.

1.6.5 Interlaced and Progressive Video Format

A sensor is scanned horizontally across the number of pixels in the frame one row at a time.

1.6.5.1 Progressive

In 'progressive' mode the scanning takes place in sequential rows, so scans all rows in every scan cycle. As each scan cycle forms a complete image within the frame, then it must perform this at least 25 times per second to have a typical video output. (Note: this is entirely dependent on video timings used within geographic regions – USA/Japan system is approximately 30fps while most of the rest of the world is 25fps). Consequently the normal notation for a video recording format is given by size and frame rate e.g. 1920×1080p/25. (1920×1080p/30 in USA) or 1280×720p/50 However this is commonly shortened to 1080p/25 or 720p/50.

1.6.5.2 Interlaced

In 'interlaced' mode, the scanning cycle refreshes the odd rows first and then the even rows, effectively making two partial images for one complete frame. These part images are called 'fields' and traditionally interlaced formats were specified by the no of fields e.g. 1080i/50. There is a move to standardise calling all formats by the no of frames e.g. 1080i/25 to avoid confusion between fields/sec and frames/sec. Standard definition, for example, would be 576i/25.

It is essential to know the intended final output in order to select the correct scanning system, as choosing the wrong scanning frequency and mode will completely compromise the final output as it is usually impossible to convert without an obvious loss of quality.

1.7 Timecode

A video or audio timecode is necessary for clip synchronisation. The SMPTE (Society of Motion Picture and Television Engineers) time code is a standard code in HH/MM/SS/FF format (Hours: Minutes: Seconds: Frames). It is particularly useful when dealing with clips shot on different media and/or multiple cameras (or audio devices). Multicam editing can utilise time codes to re-group (re-synchronise) clips to multiple points making editing easier. Timecodes on the capture media traditionally start at 01.00.00.00, which then allows for the later inclusion of colour bars and tone, titles and any pre-roll. This differs from the time code on an editing software timeline, which typically defaults to 00.00.00.00. The tradition for final delivery to a UK broadcaster, however, was (and remains) for the programme itself to start at 10:00:00:00. As before, bars and tone would be supplied in a prescribed sequence for a specified time before this (e.g. from 09:59:30:00, or 09:58:00:00 when on tape). As always, it is wise to check the specifications and settings in use at all stages of production to avoid confusion and to ensure that delivery will be to the correct technical standard required by the broadcaster.

2 Audio Acquisition

In this chapter, we cover how sound is captured and made audible.

Learning Outcomes :

- To explain what is sound and how it is captured for recording purposes.
- To give an outline of the devices typically found for capture, and playback

2.1 Sound

Definition: Sound stems from a vibration which travels and propagates through a medium, resulting in a pressure wave which is measured from the displacement of that medium. An audible sound wave stems from the displacement vibration being within a frequency range perceptible to the human ear.

Because sound is a naturally occurring phenomenon perceptible to humans, it is analogue (i.e. a continuous wave) in origin, perception and (until computerisation) storage. It needs conversion to digitised formats for manipulation and storage on digital devices, the process of which is discussed in the ‘digitisation’ section. It needs conversion from digital to analogue format for human consumption.

2.1.1 Waveforms

The analogue form of a sound wave can be represented by a complex addition of a variety of sinusoids, each of which corresponds to the form of a sine wave. For example, a ‘test tone’ of 1KHz would appear as a single sine wave if measured on an oscilloscope. There are two variables within a waveform, amplitude and frequency, each of which can be varied. Unlike the complex waves of ‘real’ sounds that we normally want to record, the test tone (or alignment tone) consists of energy at a single frequency and constant amplitude.

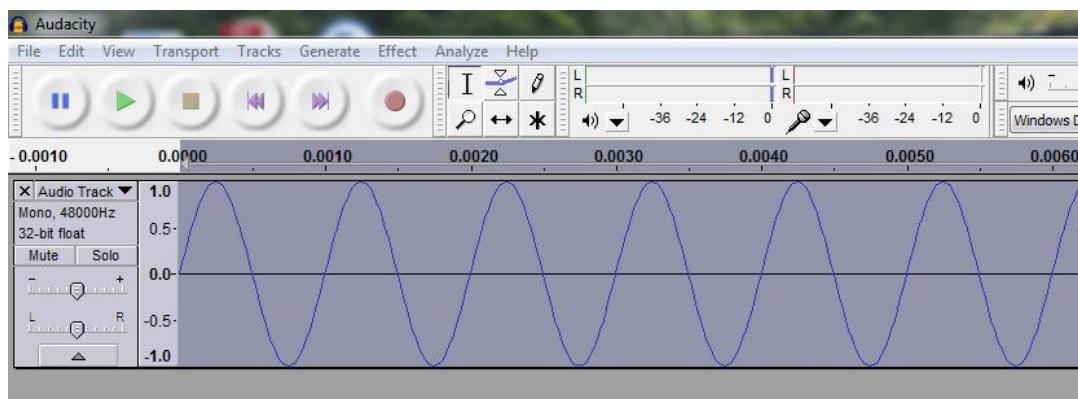


Figure 2-1: Test Tone

The final waveform as seen on an audio track is an aggregation of lesser waveforms occurring at the same moment in time and often emanating from a variety of sources simultaneously. Although a variety of diverse audio feeds can be recorded onto different tracks, and played on different speakers, the final perception is within the human ear and brain so will naturally be an aggregation.

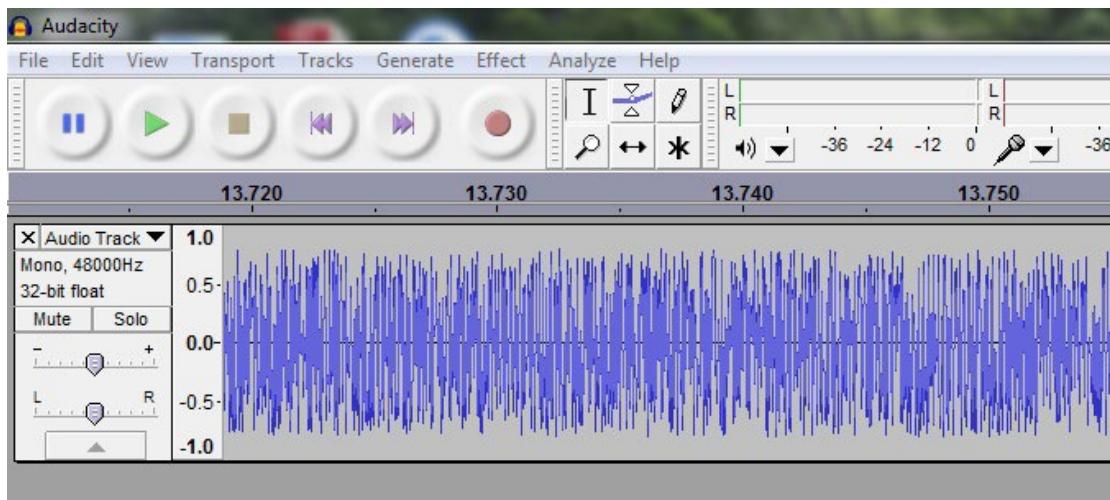


Figure 2-2: Waveform

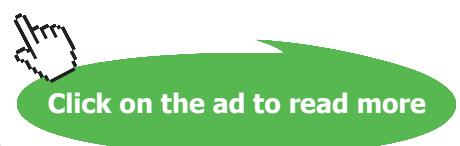
Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com

A photograph of a man with dark hair and a beard, wearing a white striped shirt, sitting at a desk and holding his head in his hands in a gesture of stress or despair. In the background, there is a laptop and some papers. The overall mood is one of frustration or overwork.



2.1.2 Amplitude

Within the waveform, the range of amplitude may be measured in various ways. When measured from the zero axis to the height of the peaks, this is known as the 'peak' amplitude. When measured from the height of the peaks to the depth of the troughs (a negative peak), it is known as 'peak-to-peak' amplitude. As sound decays over distance, or may not be 'loud' enough in origin to be at an acceptable audible level, it is typically amplified by boosting with an electronic device (i.e. its amplitude range is artificially increased). In measures of amplitude for amplification purposes, two widely used are Root Mean Squared (RMS) and Peak Music Performance Output (PMPO).

2.1.3 Frequency

The number of oscillations (complete cycles) within a given time period (usually one second) is the frequency, measured in Hertz. A waveform is often composed of a complex collection of aggregated frequencies. Some disaggregation of the waveform frequencies (i.e. breaking down the complex waveform into a set of simpler waveforms) is possible by a mathematical method called Fourier Transformation. The perceptible frequency range of the human ear is typically between 20Hz and 20KHz (20,000 Hz). Two typical frequencies seen most often by video and audio students are 44.1KHz and 48Khz. These are sampling rates and will be discussed in the sampling section.

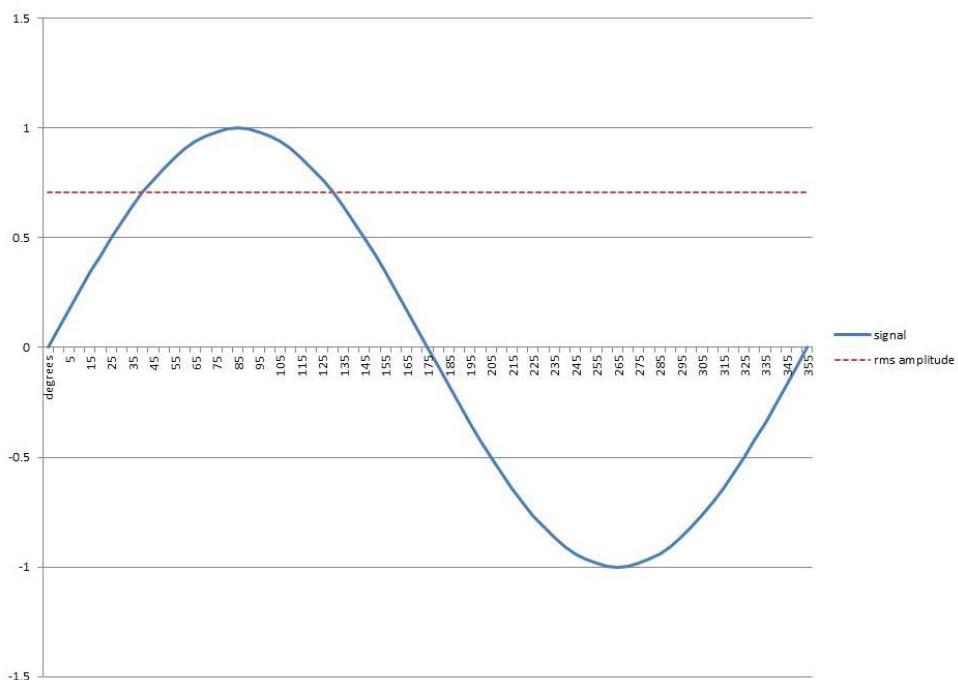


Figure 2-3: Frequency and Amplitude

2.1.4 Loudness

Loud is a correlation between the strength of a sound (amplitude) and the auditory capacity of the person hearing it. Consequently 'loudness' is a subjectively relative term based on a person's psychological determination of their experience of an auditory sensation.

2.1.4.1 Decibel Scale – Amplitude

Perceived loudness is relative. Consequently, it is important to use the instrumentation available correctly so as to ensure that the levels in use will result in a recording that is neither too quiet (leading to the need for excessive amplification of the captured signal which, in turn, will amplify the unwanted acoustic environment of the room (e.g. echoes)), nor too loud (resulting in distortion). The levels of the electronic signals indicated on modern digital recorders and cameras differ from those of older analogue ones but, in both cases, use the Decibel scale. In all cases a 'reference' level of 0dB is used and all other values are compared against it.

For digital equipment, the loudest sound which can be captured (the 'Full Scale') is the most important aspect and it is this which is used for the reference. In this case (to denote that the reference level is our 'Full Scale'), it will often be shown as 0dBFS and values using this reference can be similarly marked to give an *absolute* value (e.g. -10dBFS means 10dB below Full Scale) or can use simplified marking to show a *relative* value (e.g. -10dB meaning 10dB below whatever other value has just been quoted). The scale is a logarithmic progression, which is useful as it enables a vast range of levels to be represented in a manageable number range.

2.1.4.2 Decibel Scale – Power

In general, our meters measure the *voltage* of the incoming signal, since that is a straightforward electronic measurement to capture. However, we are normally interested in the *power* of the signal, since that is, in effect, a measure of the 'useful work' achieved by the signal. If we provide +6dB amplification (gain) to a signal, the *voltage* is doubled but the *power* is quadrupled. If we provide gain of +20dB, the *voltage* is 10 times higher but the *power* is 100 times higher! The use of decibels allows for simple addition, rather than multiplication. Thus a +3dB gain doubles the *power*. A further +3dB gain doubles that value again. Thus $+3\text{dB}+3\text{dB}=+6\text{dB}$ = a quadrupling of the power. This is true no matter what the *absolute* value of the signal may have been, to start with. Similarly with attenuation (reduction) of the signal, a -3dB attenuation of a signal halves the original power, a -6dB attenuation reduces it to one quarter of the original power, a -10dB attenuation reduces it to 10% of the original power, while a -20dB attenuation reduces it to 1% of the original power.

dBu ⁽¹⁾	mV ⁽²⁾	Voltage	μ W ⁽⁴⁾	Power	dBFS ⁽⁵⁾	PPM ⁽⁶⁾	Description
scale		Ratio ⁽³⁾		Ratio ⁽³⁾	scale	scale (UK)	
20	7750.000	10x	100000.000	100x			
18	6156.044	7.94x	63095.734	63x	0		Theoretical Max Coding Level
17	5486.580		50118.723	50x	-1		Hard Limit 'True Peak' value (TV) ⁽⁷⁾
15	4358.145		31622.777	32x	-3		Recommended Max 'True Peak' value ⁽⁷⁾ (TV)
10	2450.765	3.16x	10000.000	10x	-8		
8	1946.712		6309.573	6x	-10	6	(Approx) Permitted Max Level (Radio)
6	1546.328	2x	3981.072	4x	-12		
4	1228.292		2511.886	2.5x	-14	5	
3	1094.717		1995.262	2x	-15		
0	775.000	1x	1000.000	1x	-18	4	Alignment level (test tone)
-3	548.658		501.187	0.5x	-21		
-4	488.992		398.107	0.4x	-22	3	
-6	388.420	0.5x	251.189	0.25x	-24		(Approx) Average 'loudness' ⁽⁸⁾ (TV)
-8	308.533		158.489	0.16x	-26	2	
-10	245.077	0.316x	100.000	0.1x	-28		
-12	194.671		63.096	0.06x	-30	1	
-20	77.500	0.1x	10.000	0.01x	-38		
-30	24.508		1.000	0.001x	-48		
-40	7.750	0.01x	0.100	0.0001x	-58		

Notes:

- 1- dB scale used on analogue equipment
- 2- millivolts = volts/1,000 (rms values)
- 3- (Rounded) Relative to alignment level
- 4- μ W = microWatts = Watts/1,000,000, assumes a load resistance of 600 ohms
- 5- dB scale used on digital equipment (European standard, EBU R 68, IEC60268-18)
- 6- Peak Programme Meter (as per UK Radio, IEC60268-10 IIa)
- 7- As per Technical Delivery Standards of the UK Digital Production Partnership
- 8- Actually -23LUFS as per EBU R 128-2014, measured according to ITU-R BS.1770-3

Figure 2-4: Decibel scale – power



 MT Højgaard

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

When recording, it might be thought that the sensible thing to do would be to get as close to 0dBFS as we can. In practice, this is a bad move, since it does not allow for any unexpectedly loud sounds. Since 0dBFS is the loudest level that our digital systems can capture, anything which tries to go above that level will result in severe audible distortion when the top of each cycle of the waveform is 'clipped'. In cheap (and in older) equipment, it could even lead to the potentially disastrous 'digital wraparound' where the signal value jumps from the maximum level to the minimum level up to several thousand times per second, resulting in a deafening 'splat' of noise, which can damage both loudspeakers and ears! Therefore, professional practice is to capture recordings at a much lower level to keep well away from 0dBFS. Typical values for TV now aim to *average* about -23dB below 0dBFS. Radio (in the UK) will continue to aim to *peak* at PPM6 on the traditional Peak Programme Meter (which would correspond with -10dBFS if fed with a constant level tone) for the time being but it is anticipated that this will fall more into line with the newly-developed TV standards, in due course. In all cases, an alignment test tone is used at -18dBFS to adjust any interconnected equipment such that all scales should show -18dBFS when the tone is switched on. This ensures that only one set of meters need be monitored, whilst knowing that all the interconnected equipment is working to the same standard. (The alignment level for USA differs and is usually -20dBFS, in that case corresponding with +4dBu, there being a different relationship between the dBu and dBFS scales in that part of the world and enabling a maximum coding level of +24dBu.)

2.1.4.3 Sound Pressure Level

The values outlined in the decibel power scale are for the electronic signals and measure the voltage of those signals related to our reference voltage. However, there is also a need to consider the sound pressure level (SPL) of the actual sound in its natural environment.

Example	dB (SPL)
Shotgun @ 1m	140
Jet aircraft @ 25m	130
Threshold of pain	120
Nightclub PA speakers @ 1m	110
Pneumatic drill @ 1m	100
Fast noisy lorry @ 10 m	90
Street traffic noise @ 5 m	80
Household appliance @ 1 m	70
Conversational speech @ 1 m	60
Normal household	50
Quiet library	40
Background in Light Ent TV studio	30
Background in radio studio	20
Rustling leaves in distance	10
Hearing threshold	0

Figure 2-5: Example dB – sound pressure level

Again, the dB scale can be used but, since we are now talking in terms of sound pressure level, the reference used must be an absolute measure of sound pressure level (measured in Newtons/metre² (or Pascals)), rather than of voltage. Unlike our electronic signals (where there are several different reference levels that have been used as we moved from analogue to digital in different parts of the world), the absolute reference value of 0dB for SPL remains at 0.00002Pa (which is just perceptible at the threshold of hearing) and this is often assumed rather than quoted. As is often the case with the dB scale (which is used in dozens of applications), the reader must take care to read the context of any quoted measurement so as to ensure use of the (often implicit) reference level against which the relative dB values are being used. In order to avoid damage to hearing, care should be taken to avoid prolonged exposure to SPL higher than 85dB above the reference, which is still a surprisingly low level, akin to having to raise one's voice to only a moderate degree, to be heard clearly in a noisy room for example. Remember, also, that every +3dB doubles the power available to do damage to your ears and many apparently mundane locations may place you into a potentially harmful situation without your realising it.

2.2 Microphones

Definition: A microphone converts mechanical movement (vibration) of a sensitive diaphragm caused by sound pressure waves, to an electrical signal via a transducer. These signals are then electronically amplified.

2.2.1 Types by Construction

There are 5 common types of microphone used within audio and video recording; Electret Capacitor, Studio Capacitor, Dynamic, Piezoelectric and Ribbon.

2.2.1.1 Electret (Electret Capacitor)

The 'electret capacitor' microphone works by sound waves vibrating a diaphragm (in this case, an electrically charged material – dielectric) close to an electrostatic plate (the charge for this plate does not require power as it is built-in during manufacture). The vibration causes a change in proximity which results in a small voltage change which then needs amplification. Electret microphones are cheap to produce and so very common. The power for these microphones is purely for amplification and can usually be provided either by battery or by phantom power. (Phantom power for microphones is explained in a later section of this chapter.)

2.2.1.2 Condenser (Studio Capacitor)

The 'Condenser' microphone (sometimes referred to as a 'Studio Capacitor' or 'Studio Condenser' microphone) is also based on a capacitor (a device to store energy). The condenser microphone works in a similar way to the electret, i.e. a diaphragm vibrates close to an electrically charged plate, which in this case has an external voltage supplied to it to provide the charge. The vibrational movement causes a capacitance change, used to modify (amplitude modulate) the electrostatic charge. These minute changes are then amplified. Phantom power is required for both generating the signal and amplification.

2.2.1.3 Dynamic

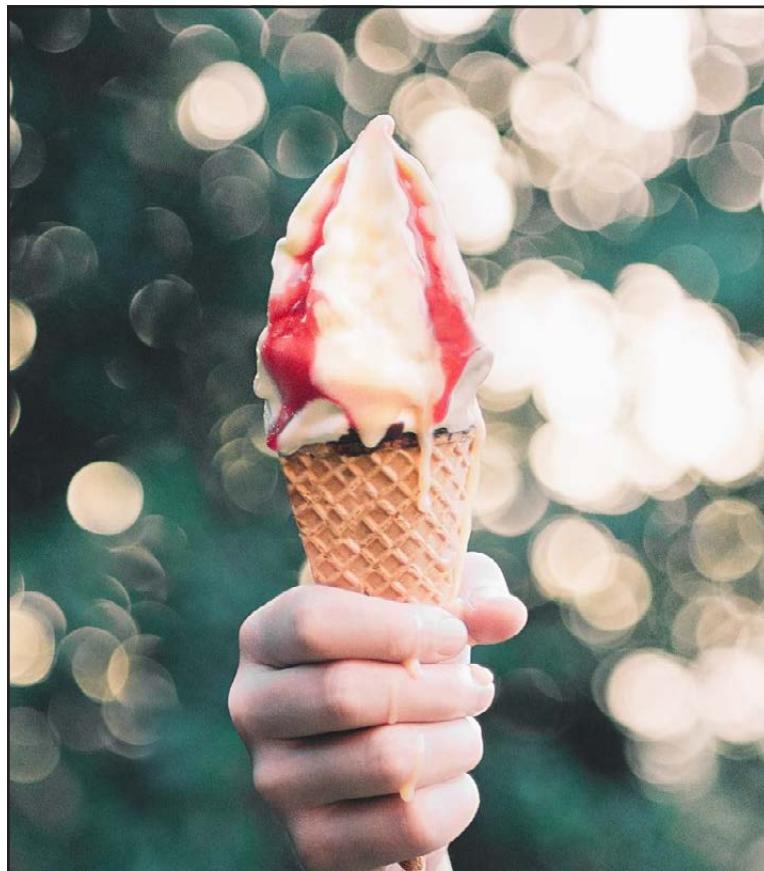
The dynamic microphone uses a magnet within a coil to create an audio signal. As the diaphragm vibrates the coil that is wrapped around the magnet generates a small electrical current which is sent down the wires; this relies on a principle known as electromagnetic induction. This type of microphone requires no power to work, and is effectively a loud speaker in reverse.

2.2.1.4 Ribbon

The ribbon microphone works by the movement of a thin metallic strip within a magnetic field. The strip is connected by wires to send the signal, which is caused by magnetic induction as the strip vibrates. This requires no power to generate the signal but may use power for amplification.

2.2.1.5 Piezoelectric

Typically used as contact microphones (e.g. attached to a musical instrument). They utilise the property of a piezoelectric crystal which, if subjected to pressure, generates a small electric charge. They require no power to generate a signal but will need power for amplification. For instance an acoustic guitar may use a piezoelectric pickup, but a typical electric guitar uses a magnetic coil pickup. (See dynamic microphones.)



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more



Figure 2-6: Microphone types

2.2.2 Polar (Pick Up) Patterns

The polar pattern refers to the shape of the area within which the microphone will pick up sound.

2.2.2.1 Omni-directional

The omni-directional pattern picks up sound from a 360 degree radius around the microphone. A typical example of this pattern in use is for recording background sound (eg. ambient noise)

2.2.2.2 Bi-directional

This pick-up pattern is also known as a figure of eight. It picks up sound from both in front and behind the microphone. A typical example of this pattern in use is in a radio studio, where one microphone is situated suspended between presenter and guest.

2.2.2.3 Cardioid

The cardioid pattern resembles a heart or apple shape. It predominately picks up from a front facing direction and less so from the rear face. As a unidirectional pattern it is commonly used for stage microphones by vocalists, particularly if the microphone is handled it would pick up less handling noise than an omni-directional microphone.

2.2.2.4 Shotgun

A highly directional pick-up pattern, it picks up little from side and rear but forms a longer thin pickup field to the front. Most used by videographers who wish to pick up the audio from the direction to which the camera is pointing. Alternatively, a shotgun can be mounted on a boom pole and held out of picture shot but enables a clear recording to be made as the microphone can be held closer than it would otherwise be when mounted on the camera.

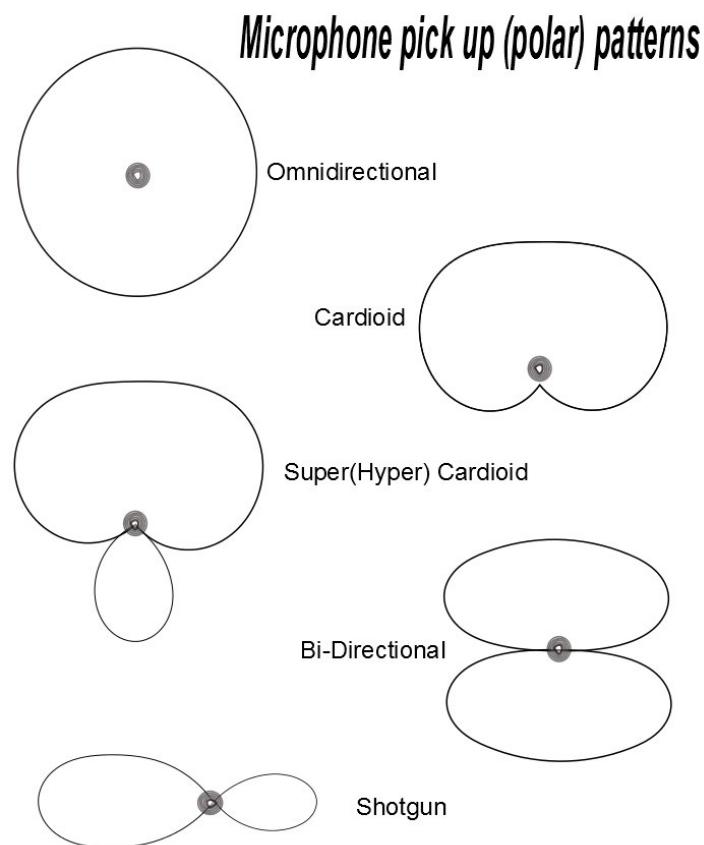


Figure 2-7: Polar Patterns

2.2.3 Phantom Power

Phantom power refers to an electrical voltage sent to a type of microphone that requires it, (see condenser or studio capacitor microphones). Typically, (but there are variants of 12v, 24v and 10v for digital microphones) phantom power is +48v direct current (DC) and sent via the XLR connection to the microphone. Some electrets may use ‘phantom’ power, or power from batteries if they require far less than 48v. The term phantom relates to the fact that the dc power supply *to* the microphone is cleverly carried by the same wires that carry the ac signal *from* the microphone. Thus the power supply is there but you can’t actually see any separate cables carrying it there.

2.3 Systems (Balanced and Unbalanced)

All systems that pass electrical signals are susceptible to noise (interference) that will degrade the quality of the signal.

2.3.1 Unbalanced

An unbalanced system refers to one in which the return path for the current being supplied by the device acting as the source of the signal is shared with other return paths and may include electrical noise which is being shunted to earth. Since any current flowing in the shared return path is part of the overall circuit loop, the unwanted noise is passed into the next stage of the electronic signal path. The input of the next stage takes in the difference between the two conductors but, since the 0v ('ground') conductor may now be varying with the shared noise signal, the difference between them no longer represents just the intended signal relative to 0v so the varying noise signal will be heard at the output. This system is typically used by domestic equipment (e.g. Phono (RCA) leads).

2.3.2 Balanced

A balanced system uses three wires, two wires each carrying the signal so any noise (interference) arising from stray electric fields will most likely affect both wires equally. The audio signal is the same in both wires but in anti-phase to each other, so that one carries a positive voltage while the other carries an opposite equally strong negative voltage. The third wire is the ground (i.e. 0 volts) and is connected to the outer screening braid of the cable, as usual, thus helping to further improve noise rejection.

It is the difference between voltages on the two signal wires that is used to calculate the signal strength. Since any imposed noise will affect both signal wires equally (shifting the voltages up or down together), the difference between the two remains unchanged and correctly represents the original wanted signal. In effect, the interference is 'cancelled out'. This system tends to be used in professional equipment.

2.4 Connectors

2.4.1 XLR

XLR was an abbreviation by ITT-Cannon for a connector using its own x-type connector (X), a latch (L), and a rubber (R) compound around the female connection. The XLR is also used as an abbreviation to name the 3 pins ground (X), Live (L) is the in-phase signal also called 'hot' denoting the positive signal, Return (R) is the out-of-phase also called 'cold' denoting the negative signal. However, some power connections use a 4 pin XLR and other systems use 5 pins, so this only applies to 3 pin connectors.

XLR connectors are usually indicative of a balanced system.

2.4.2 TRS

TRS is an abbreviation of Tip, Ring, Sleeve, the component parts of the connector. TRS may be balanced or unbalanced. In a balanced system the tip is hot (positive), ring is cold (negative) and the sleeve is ground (0 volts). TRS may also be used for unbalanced stereo signals using the tip and ring for left and right channels and, again, the sleeve for ground, connected to the cable's braided screen. The normal size for TRS connectors is $\frac{1}{4}$ inch (6.35mm) diameter.

2.4.3 Phono (RCA)

The RCA 'phono' type connector is unbalanced, the middle pin carrying the signal (Hot) and the outer ring carrying the ground (Cold). The coaxial cable has the middle wire sheathed from the outer braided screen conductor. The RCA phono connector was used widely for audio (typically coloured white and red) and composite video signals (Yellow), although also for component video (Red, Green, Blue) in domestic systems. These largely replaced the old European 5-pin DIN audio connectors in recent times.

2.4.4 Line-in (PC jack plug)

Rather like miniature versions of a TRS, jack plugs usually come in 3.5mm size. These are unbalanced connections typically used to connect headphones, speakers, microphones etc to computers and audio-video (A/V) equipment such as domestic camcorders. Occasionally, 2.5mm will be encountered.

A photograph of the Apollo Hotel building at night. The hotel's name is visible above the entrance. On the left side of the building, there is a large red circular logo containing a stylized lightbulb icon, next to the text "CISO Conference" and "Produced by Inspired". On the right side, a white rectangular box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and "Dec 5th 2019". Below the main image, a white banner contains the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders" and the "Inspired" logo.

2.5 Speakers

A speaker is an audio output device, which works rather like a microphone in reverse. The cone is a diaphragm that is vibrated using a magnetic coil which produces an air pressure wave which recreates the recorded sound. It is the size and shape of the cone which gives the speaker its sound quality and the frequency range it can output.

2.5.1 Hi-fi speakers

High-Fidelity (Hi-Fi) was a term coined in the 1950s for equipment that met a reproduction standard. The audio signal input is amplified (so requires an Amplifier – often built into the unit) before being sent to the loud speaker. Hi-Fi systems use a range of speakers to cover the output frequency range, each being a dynamic speaker (see dynamic mic) and so are referred to as drivers. A typical range of drivers are, subwoofer (very low frequency range), woofer (low frequency), mid-range driver (as the name suggests), and tweeter (high frequency range).

2.5.2 'Monitor' Studio Speakers

Ordinary Hi-fi speakers are not considered good enough for accurate monitoring of sound. Different speakers can produce a slightly different sound quality for the same signal input as the manufacture, size and shape of the speakers are different. Monitor speakers ought to be able to accurately reproduce a sound and so will have been calibrated to a range of test conditions.

2.5.3 Foldback Monitors

Simply a name given to monitor or stage speakers that face the stage or studio players not the audience; these are typically used by bands to listen to their music as they are playing. These are needed as the sound will distort in an auditorium as it is likely to be 'absorbed' by the audience or sound absorbing materials, reflected off hard surfaces and initially the speaker is facing away from the originating source e.g. the band. The use of wedge-shaped monitors pointing upwards from the feet of the performers can also be enhanced through a separate monitoring mixing desk, enabling each performer to receive a personalised mix of the band, according to specific needs. (e.g. the drummer does not need to hear a foldback signal of the drums, as these are loud enough already but failure to provide sufficient signal from the guitarist or vocalist to the drummer's monitor would make it impossible to hear them as they are already in front, and projecting forwards.)

2.5.4 Amplifiers

An amplifier works by boosting the input to a level suitable for the subsequent stage of a process. There are many types of amplifier including power, video signal, op-amp (operational amplifier), instrument and audio power amplifier. It is the audio power amplifier that provides the final change in level to drive the speakers in a Hi-Fi system.

3 Scene and Studio Lighting

In this chapter, we cover light, basic lighting techniques and typical devices

Learning Outcomes:

- To explain lighting for video and studio purposes
- To give an outline of the devices typically found for lighting

Any light is usable to illuminate a scene, the problems arise with colour, intensity and flicker.

Colour: As mentioned in Chapter 1, light has a ‘correlated colour temperature’ and so a colour cast. It rather depends on the type of bulb and the power it consumes to reach its working temperature. For example a typical domestic lighting has a yellow cast as it has a tungsten filament and will often be around 2700K. Halogen lamps from a building site will also be a light yellow at about 3200K, so often quite suitable for use with a video camera when that is set for studio lighting. Low-pressure sodium road lighting is a monochromatic orange colour and reproduces the range of colours in the original scene very poorly indeed.

Intensity: Contrast is important but, just like the eye, the ability of a camera to resolve detail in the shadow areas whilst simultaneously reproducing detail in the highlights will require careful control of the lighting, as the range of illumination which can be accommodated simultaneously is limited. If the eye adjusts to accommodate a bright area, the rest of the scene will appear to be too dark and may miss detail in shadows. In much the same way that the eye adjusts, the iris on the camera will also adjust (if on auto) and uncontrolled lighting that varies in intensity will cause the camera to constantly adjust, which rapidly becomes noticeable and distracting to the viewer. As with so many other settings, a professional approach is to control the situation manually. On studio cameras and camcorders ‘Zebra’ bars are used to indicate where highly exposed areas are at risk of saturation, thus enabling action to be taken so that the exposure can be kept within bounds. Dimmers might be used if the light is too intense but this causes a problem of colour shift when using traditional bulbs, since running the bulb below temperature will cause a shift towards the red-end of the spectrum. An alternative is to move the lighting further back, if possible, since illumination level is a function of distance squared.

Flicker: One effect of using artificial lighting that is connected to the mains (an alternating current – so it effectively turns the power to the bulb on and off in phase with the supply), is the flickering effect caused by the phase of the lighting being ‘seen’ because of any slight disparity between the mains supply and the frame rate of the camera. This is particularly problematic with ‘gas discharge’ lighting, such as domestic ‘fluorescent tubes’, which can seem to grow dim and then bright even at standard frame rate, so should be avoided if possible. More modern office light fittings drive the tubes at a high frequency, so these tend to be less of a problem, though the difficulty of dealing with a discontinuous spectrum and the effects on colour reproduction remain.

A lot of videographers now shoot high frame rate footage for slow-motion playback. This tends to require additional lighting as the faster the frame rate, the lower the exposure time and consequently the footage is darker. It is not unusual to see light flicker when shooting slow motion if the videographer has no control over the type of lamps used at the venue, eg. sporting events. Lamps that ‘glow’ longer, hotter, have a higher frequency (such as specialist fluorescent tube studio lights), or are connected to a battery or power-pack (a direct current source – so no phase problems), are to be preferred.

Similarly, the choice of ‘shutter speed’ (integration time) has a crucial effect on the appearance of video and is inextricably linked to lighting level. Assuming a normal 1:1 recording and playback speed (i.e. not slow motion), the playback rate will normally be based on 50Hz (60Hz USA) systems. If a shorter than necessary shutter speed is used, this will have two effects:

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

1 – it will ‘freeze’ the action of each field or frame, meaning that fine detail with sharp edges (which would otherwise blur in a desirable way during movement) will become obvious in each screen refresh, leading to visible judder or flicker.

2 – it will reduce the overall exposure time for each frame, meaning that more light is required to capture the footage.

There are also complex technical consequences when attempting to compress the recorded video if too much high-frequency signal (i.e. fine detail) is present in each frame. It can lead to severely reduced image quality (blockiness). Therefore, it is recommended that the shutter speed should be manually set to 1/50s (1/60s) and locked. If the scene becomes too bright and the camera would normally choose a faster shutter speed then Neutral Density (ND) filters should be deployed to reduce the amount of light entering the camera’s sensor. Many cameras have ND filters built-in but they can also be fitted in front of the lens.

3.1 Three Point Lighting

Three point lighting is the basic method of illuminating a subject (actors, interviews etc.). Each point is a light that illuminates the subject from a different angle (see figure 3-1).

3.1.1 Key light

The Key light is the primary source of illumination. If you only have one light available then it becomes the ‘key’ light. Cameras often have inbuilt lights or external light mounts on them. The problem with these are low power and position. A direct lighting source on the camera rather acts like an inbuilt flash or photo-booth, it ‘flattens’ out the subject. It is always better to elevate the light source and move it to half way between the front and side of your subject. This reproduces the position of the sun and gives the subject a 3 dimensional feel.

A key light is often a spot light i.e., it illuminates a particular area by focusing the light to a point of interest on the subject.

3.1.2 Fill Light

If two lights are available, the second light becomes a ‘fill’ light. A fill light is so called as it fills the shadows on the subject caused by the ‘key’ light and, thus reduces the contrast range to manageable values that allow both highlight detail and shadow detail to be captured. Obviously if you have two lights of equal intensity at similar angles in front of the subject the effect would be to ‘flatten’ out (similar to one light problem) so the fill light can be moved further away. The fill light is then less intense, so detail is seen in the shadows, but the shadows still exist. A fill light may also act as a flood (illuminating a larger area) and should be more diffuse, which can be achieved by putting a mesh (or ‘scrim’) over a traditional lamp.

3.1.3 Back Light

A back light illuminates the subject from the rear. Not only does this give the subject 'separation' from the background, it gives the subject a 3-dimensional feel. A common effect with a back light is to give the subject a 'halo' or soft glow against a contrasting background.

The back light should not be within the shot (as it will affect the overall balance) so will be to the side and rear, or hidden behind the subject.

Three Point Lighting

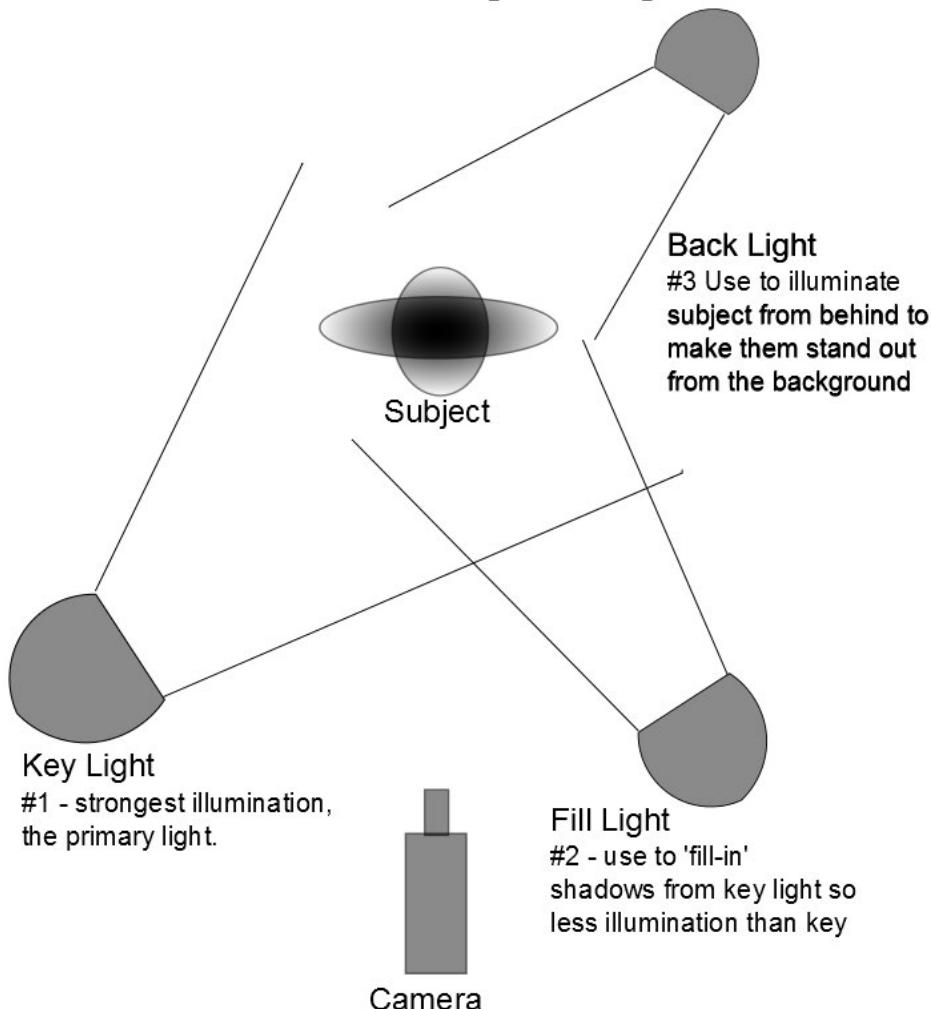


Figure 3-1: 3 point lighting

3.2 Types of Lighting Fixture

Strictly speaking a 'light' is anything that emits light but we tend to call it by a variety of names when dealing with lamps and lanterns. A lantern, lighting instrument or luminaire is the complete housing whereas the lamp is the light emitting component within.

3.2.1 PAR CANS

Parabolic Aluminized Reflector (PAR) is a ubiquitous lamp used in many situations such as music stage venues. The lamp is housed in a can, sized in units of 1/8th of an inch across the diameter eg. a PAR16 is a two inch wide lamp (often referred to as a 'birdie'), whereas a PAR64 is an eight inch lamp. Being parabolic it reflects the rays in a concentrated beam, these beam widths also define the type of lamp. Extra Wide Flood (EXG), Wide Flood (WFL), Medium Flood (MFL), Narrow Spot (NSP), and Very Narrow Spot (VNSP).

LED par cans use the same size notation, some using Red, Green and Blue (RGB) LEDs in an array to form a 'white' light. However, great care will need to be taken with LED lamps for use with video. Most are not intended for such use and can generate extremely obvious flicker at certain colours, whilst the colourimetry of such lanterns is often very poor when used with video cameras. Only lanterns designed specifically for video are really suitable but these tend to be at the top end of the price range. Cheaper units intended for dancefloor use are fine for the intended application but often hopeless for video.

3.2.2 Cyclorama

Cyclorama lamps are strip lights to illuminate backgrounds, often using gels or glass filters to change colour. RGB LEDs are now common for this. Again, the same care in selection of LED units will be required for use with video.



The image shows a woman with long blonde hair wearing a black VR headset, looking upwards with a surprised or excited expression. To her right is a red rectangular advertisement for MT Højgaard. The ad features the company logo (a stylized 'M' and 'H') and the text 'MT Højgaard'. Below this, in large white letters, is 'BEDRE LØSNINGER'. A block of Danish text follows: 'I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?' At the bottom of the ad is the website 'mth.dk/vorestilgang'. A green button at the bottom right of the page has a hand cursor icon pointing to it, with the text 'Click on the ad to read more'.

3.2.3 Floods and Spots

Floods: A ‘flood’ is a basic lamp and reflector housed in a lantern, used to illuminate backgrounds. Floods are not focused as they have no lens, but some (eg Blondes) can be adjusted in relation to their reflector for a wider or narrower beam. Three typical flood lights found in a studio are Blondes, Redheads and Peppers.

Blondes are high power floods usually around 2000 watt (2K).

Redheads are the most common in small video studio work, typically 800 watt.

Peppers are small floods around 200–400 watt.

Arc lamps emit a much more intense brightness by incandescence of a chemical element within a gas, rather than a filament within a bulb. Consequently arc lamps are more common on a TV or film set, eg HMI (Hydrargyrum Medium-Arc Iodide). Early arc lamps were limelight (calcium oxide in hydrogen/oxygen mix) and later carbon rods. They usually require a special power supply unit.

Spots: Spot lights are used for areas of specific focus, as they have a lens.

A ‘Fresnel’ (pronounced ‘Frennel’) spot light has graduated rings on its lens (a step lens) to allow focusing of the light when the lamp is moved in relation to its lens within the lantern. At the focal point the rays of light are parallel; moving the lamp closer to the lens (inside focal point) gives a divergent beam and away from the lens (outside focal point) a convergent beam.

A ‘Profile’ is a clearly defined spot light usually used with a ‘gobo’ (a metal sheet with a cut-out pattern which can be focussed on to a backdrop or floor). It may also have an iris to reduce the spot diameter.

The ‘Followspot’ is a sophisticated profile spotlight, with multiple lens and iris. A Followspot is used to follow a subject around a stage. Some arc lamps were used as Followspots and so the term ‘limes’ still remains.

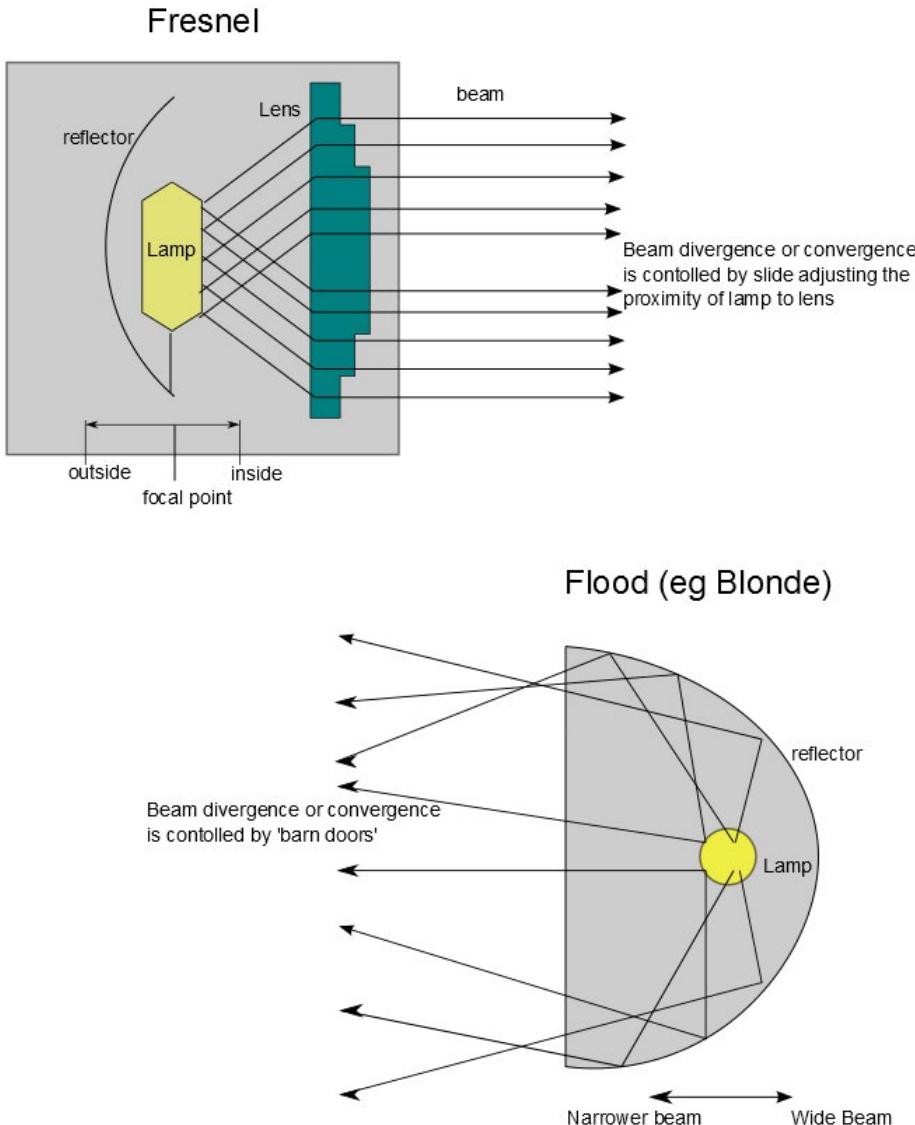


Figure 3-2: Fresnel versus Flood

3.3 Ancillary Equipment

3.3.1 Gels

Gels are transparent coloured films, usually a polycarbonate sheet, that filter a light source by absorbing colour. In history coloured liquids, silk, gelatine (aka gelatin in U.S.A. – hence the name ‘gel’) and polyester, have been used to the same purpose.

Typically the gel films you see today are polycarbonate sheets, but these don't have a very long life span. Permanent gels are made of glass with a ‘dichroic’ film coating. Dichroic – i.e. a coating that reflects back light other than the colour it allows through. Warning – always ensure that gels are designed for hot operation. Using the wrong type with a hot lantern could result in showering people with molten plastic and/or a fire being started!

3.3.2 Reflectors

Anything that reflects light can be used as a reflector (caution – do not use mirrors or tinfoil to reflect sunlight), but more often it refers to a reflective material made from a cloth with metallic coating. Gold is used to give a warm reflection, silver for a bright light (without altering the light source's colour), blue to add a cool tone, and white for additional white light (not as bright as silver) which is useful for shooting white material without altering its tone (e.g. weddings).

As paint can reflect, it is wise to think about the ambient colour of a room and its effect on the 'subject' being filmed, particularly moving between rooms in different scenes. The shot may need post-production colour correction. Also the colour of the room in which you edit may also have an effect on your eyes and inadvertently influence post-production colour correction. Better to use a neutral grey or darkened room, ensure your editing monitors are calibrated using a colour meter. However if just using a non-calibrated standard computer monitor then use the same colour profile as your camera (sRGB). A note of caution – colour meter apps are now available but often require you to have the device's sensor a couple of feet away from the source being measured, and so can be potentially affected by other light sources in the room.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

3.3.3 Filters, Diffusers, Gobos

3.3.3.1 Filters

Colour filters do much the same as gels, except they alter the quality of the light entering the iris of your camera. Beyond colour there are, UV, polarised, effects and graduated filters which add an effect to the shot. UV filters are often used as clear glass protectors for lenses (far cheaper to replace if scratched than a lens), as film is less rarely used and ultraviolet light less of a problem in digital cameras and camcorders. Polarised filters are used for glare, darken a scene to increase contrast (e.g. cloud definition in a light sky) and reflections. Depending on the direction of sunlight relative to the angle of the subject, they can also increase the apparent colour saturation in a way that is often far more visually satisfying than can be achieved by an attempt to make corrections in post-production. (This is usually the case for most things. Better to get it right when shooting than to try to 'fix it in post' afterwards!) Filters are often housed in a Matte Box.

Neutral density filters (ND filter) are a series of graduated filters that darken an over-bright scene. Individual ND filters can be externally mounted onto the lens, and they are typically graduated on an ND scale. On some camcorders and cameras these filters are integrated into the camera itself giving a more limited graduated choice.

3.3.3.2 Diffusers

Diffusers are anything that gives light a 'softer quality' by evening out the light source's beam. The effect of diffusing a light source is to reduce the contrast by increasing the effective area of the luminaire from a point of light into a larger disc. This also has the effect of reducing hard shadow edges and reduces the tendency to reveal fine but shallow details. For this reason, diffuse light sources are often used in portraiture work. In the three-point lighting (above), the key light tends to be a relatively 'hard' light (produced by a semi-focused point source) while the 'fill' light is usually a diffused lighting instrument, often set somewhat lower than the height of the key, thus softening both the hard-edged shadows (e.g. neck/chin and nose etc) while reducing the contrast by bringing up the illumination levels on the parts that would otherwise be too dark. Traditional glamour portraiture will tend to favour the use of more diffuse sources, since it is the hard key lights that tend to reveal wrinkles while 'soft' lights help to disguise them!

In addition to diffusion of the light source, diffuser filters can also be applied to the lens so as to produce a 'soft focus' effect, giving a less sharply defined look, and so are also popular in portraiture work. The effect differs from that of diffusing the light source, where the subject may still be in pin-sharp focus but evenly and softly lit, whereas a soft focus effect filter results in all fine detail and sharp transitions being lost. Oscar winning cameraman Jack Cardiff would breathe on his lens to get a foggy diffused look that would disappear gradually.

A cheap well-known alternative to a diffusion filter is to smear Vaseline™ on a UV filter. Also a soft diffused look was achieved by shooting through a fine gauze sheet, hence the famous quote of movie actress Tallulah Bankhead “they used to photograph Shirley Temple through gauze; they should photograph me through linoleum”.

Meshes such as gauze have been used in the theatre for a long time, the ‘shark-tooth scrim’ (woven material) is used for the special effect of revealing an object behind what looks like a solid wall. However, ‘Scrim’ is also the name used for a diffusing mesh attached to spotlights (especially those that cannot be electronically dimmed) or placed between light source and subject to combat ‘hot spots’ of light that would look ‘burnt out’ on shooting. Typical scrims attached close to spotlights would be made of metal wire.

3.3.3.3 Gobos

A gobo (GOes Between Optics) is a mask (often a metal template) that can be used to control the shape the light source throws, or for decorative effects.

3.3.4 Barn Doors – Light Shielding

Anything that prevents light transmission is effectively a light shield. In order not to allow light to spill in an uncontrolled manner, it is useful to shield off areas from the light sources. Spotlights and stage lights have flaps that perform this function called ‘Barn Doors’.

Matte Boxes act in a similar way to a camera lens hood by preventing unwanted light (e.g. sunlight from an angle) reaching the lens. Matte boxes can also double up as a multiple filter holder. An alternative (or supplement) to the matte box is a French Flag. This is simply the equivalent of one of the barn doors (though it can be larger and of different shape) which can be positioned using (for example) a swan-neck holder attached to the camera rails so as to block light from a particularly problematic source. Alternative flags can also be positioned using C-stands around the set so as to block specific lights from illuminating chosen areas, as required to control the overall lighting effect.

3.3.5 Calibration Cards

In order to provide acceptable image quality in the finished production, it is necessary to set-up both the camera (during shooting) and the video signals (in post-production) correctly. To assist with this aim, a videographer can use standardised calibrated test cards.

There are two functions to be performed before shooting. The exposure controls and the ‘white balance’ must both be set correctly.

The white balancing process takes account of the correlated colour temperature of the light sources. In order to avoid an unintentional colour cast, it is necessary to adjust the image processing circuitry in the camera to ensure that a subject that is completely neutral (white, black, grey, having no colour value) produces equal voltage signals from the red, green and blue channels. This can be achieved by filling the frame of view with a neutral test card which is arranged so as to reflect only the lighting which will illuminate the subject. Sophisticated studio cameras have a range of controls to adjust the gain and the black level ('sit') for each colour channel to a very fine extent. This is important since changing from one camera to another will reveal even slight differences, especially on skin tone to which the human eye is quite sensitive. Cameras intended for PSC (Portable Single Camera) shooting tend only to have a 'White Balance' button (WB) which can be pressed. At that point, the camera will make its own adjustment to balance the gains of the amplifiers for each of the red, green and blue signals.

The exposure controls enable the camera to record a range of values related to the lighting intensity at each point within the scene. Under normal circumstances, the aim is to ensure that subjects with reasonably high reflectance are not recorded with so high a level that they appear to have the same intensity as the light source such that all detail is lost in the highlights (white crushing). Similarly, subjects with fairly low reflectance should not be recorded at too low a level, resulting in loss of shadow detail (black crushing). This can be achieved by a variety of means.

A night photograph of the Apollo Hotel building in Amsterdam. The hotel's name is visible above the entrance. On the left side of the building, there is a large red circular logo containing a stylized lightbulb icon, next to the text "CISO Conference" and "Produced by Inspired". On the right side, a white rectangular overlay contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and "Dec 5th 2019". Below this overlay, there is a white banner with the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right of this banner is the "Inspired" logo, which includes a blue lightbulb icon.

Increasing [or reducing] the diameter of the iris (aperture) will allow more [or less] light to arrive at the camera's sensor, remembering that this adjustment will also affect the depth of field. If the light source is so bright that white crushing still occurs (or if it is not desired to 'stop down' which would increase the depth of field) then Neutral Density filters can be inserted in the light path to reduce the exposure. Conversely, if the subject is still too dark when the iris is fully open (or if it is not desired to 'open up' which would reduce the depth of field) then there are two possible solutions: Increase the lighting or increase the 'gain' of the electronic image processing circuits. The former is certainly preferable to the latter for image quality, though is more cumbersome than merely flicking the gain switch on the camera! However, using gain (i.e. a setting greater than 0dB) is the electronic equivalent of using a faster film, with similar consequences on producing a grainy or electronically 'noisy' image. The overall contrast ratio between the lightest and darkest parts of the image can only be controlled by careful use of light sources. If it is found that white crushing can only be avoided at the expense of losing the details in the shadows then additional lighting will be required.

There is one other potential solution to the problem which is to vary the shutter speed (equivalent in a cine camera to the shutter angle). However, this is not recommended since a shutter speed longer than about 1/50s (e.g. 1/25s) will produce a result that appears to be very 'smeary' whilst a shutter speed shorter than 1/50s (e.g. 1/100s and faster) will make the judder and flicker of the relatively slow frame rate much more visible, freezing irrelevant high frequency background information (fine detail) which becomes highly distracting to the viewer, as well as potentially producing problems for bit rate reduction (compression) systems (e.g. MPEG), which may be unable to cope with the additional data thus created, resulting in the final image quality becoming degraded simply in order to fit the available bandwidth of the transmission system.

As with all 'auto' settings on cameras, the use of auto exposure would be frowned upon in video circles since it will often 'hunt' around what the *camera* perceives as being the 'correct' settings to give an average value, when integrated across the whole frame. There are two problems with this: the levels will be seen to go up and down over time, which is highly noticeable; and the videographer may well have a particular highlight or shadow in which detail is required to be seen but the camera adjusts for the overall average, preventing correct capture of the specific subject. However, in the majority of cases where the overall contrast is sufficiently well controlled, most normal scenes will produce a reasonable range of video signal levels centred about an average value of reflectance, which enables the exposure settings to be assessed and the controls locked so that they do not vary.

A function of the human visual system means that what appears to be a mid grey level, half way between black and white, is in fact reflecting only 18.42% of the light intensity of the light source which illuminates it. The electronics within the camera also takes into account this phenomenon and, if the camera's exposure controls are set such that a scene having an average mid-grey level of about 18% reflectance will result in a signal level of half the available voltage range then, on average (and so long as the contrast of the scene is not excessive), the result will normally work well and be useable.

To aid the videographer, a standard calibrated 18% reflectance grey card can be used. This is inserted into the field of view at the location of the lit subject, preferably filling the frame, whilst being angled between the light source and the camera. Either adjusting the settings so that the exposure value is shown to be in balance on the camera's meter, or momentarily enabling the camera's auto-exposure then locking it at this value should produce a satisfactory result. On the reverse is usually a white side which has around 90% reflectance. A camera which has a 'zebra' facility (i.e. a setting which inserts moving stripes on the viewfinder wherever the signal level is very high, so as to alert the user to areas of potential overexposure) can use the white side of the card to check for correct operation of the over-exposure warning. The non-linearity of the eye's response means that the 90% reflectance should result in a 95% video luma signal level, which is often one of the options available to select for the zebra warning.

There is some debate ongoing about whether camera and light meter manufacturers are now issuing their equipment with meters that would produce the correct luma signal voltage level when the reflectance is about 12-14% rather than 18%. It is possible that they might be failing to take account of the discrepancy between the equation which is used to represent the performance of the eye and that which is used for Rec709 in video. To produce a luma signal at 50% of the available range, Rec709 would actually require a reflectance of 22%. Consequently some users are reporting that they adjust the iris by half a stop to compensate.

In practice, the non-linear relationship of Rec709 should take care of the issue further up the scale but, in any case, a good videographer would consider making use of an additional calibration card which has a range of known reflectance values ranging from near black (0%) to near peak white (100%). A semi-pro camera may have an inbuilt waveform monitor, which in combination with such a test card enables much more useful calibration of the camera to take place, for it can show at a glance whether white or black crushing is occurring, as well as averaged middle values. A lower-specification camera can still make use of this technique if the output of the camera can be fed to a laptop running high-end editing software (e.g. Avid®, Final Cut Pro®, Adobe Premiere® etc) in which waveform monitors and vectorscopes are often available. This is particularly effective when shooting a scene which is far from average (e.g. snow) where relying on the camera's meter to integrate an average value across the whole viewpoint would result in the snow being exposed so as to produce an average grey rather than white! Shooting a few seconds of the test card in position immediately in front of the main subject will also assist in the post-production process, where pixel values can easily be sampled and any fine adjustments made for both colour correction and exposure but, as always, it is far better to 'get it right on the set' rather than try to 'fix it in post'!

An alternative approach to that of measuring the reflected light is to measure the incident light falling directly on the subject. This is particularly useful when measuring the ratio of light intensity between two lighting instruments so as to control the contrast on the subject and it is fairly straightforward to take a relative reading between the two, observing the indicated difference in f-stops, remembering that the invercone (often a detachable white plastic dome) must first be placed over the sensor to compensate for the alternative measurement method and that the sensor is now to be pointed towards the light source rather than towards the subject. (When comparing two light sources, each must be alternately shaded from the sensor, of course.)

However, for an absolute (rather than relative) reading of overall intensity, it is necessary to enter into the meter the equivalent ISO value of the camera's sensor, which is often not known. This, along with the camera's shutter speed, is fed into the meter which returns a value for the f-stop to use on the lens.

In order to determine the camera's ISO setting so that this technique can be used, set the camcorder to 0db (i.e. no additional electronic gain), use a known mid-range f-stop (e.g. f4), light the 18% grey card to a level that causes the camera's output to read 50% of the available signal level range on the waveform monitor. Now use the light meter without its invercone in normal reflective light mode, pointing it towards the grey card and setting the shutter speed on the meter to that used by the camera and then altering the ISO setting on the meter so that it returns a reading the same as the iris on the camera (f4 in this example). Now the light meter will provide the baseline ISO equivalent of the camcorder and it can be used with the invercone in incident mode, as outlined above.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

3.4 DMX Lighting Control

Digital multiplexing (DMX) is a control protocol commonly used for lighting, but can be used for other non-safety critical devices. It is basically a means by which a control source (ie a lighting desk) can address a device and send to it a packet of data with instructions to operate in a given fashion. Actual implementation of the DMX system may vary depending on controller so this explanation is intentionally general.

3.4.1 Addressing

In a DMX universe (i.e. all the devices connected to the control source), each device needs an independent address. For example, in a DMX512 universe (a possible 512 addresses) if each device was a dimmable lantern with a single channel, then up to 512 devices could be addressed. In a more complex system, the number of different operations controllable on a device will each require a further address per channel. For example if each device were an LED lantern that had 3 controllable channels (red, green, blue), then the maximum number of devices is $512/3 = 170$. The device knows it is being controlled as its address is the same as its first channel.

The addresses are binary so in DMX512 this requires 9 bits ($2^9 = 512$, so effectively addresses between 0 and 511), the device has its start address set either by dip switches or internal memory. Any further dip switches on the device may be used as non-DMX controlled features dependent on device and manufacturer.

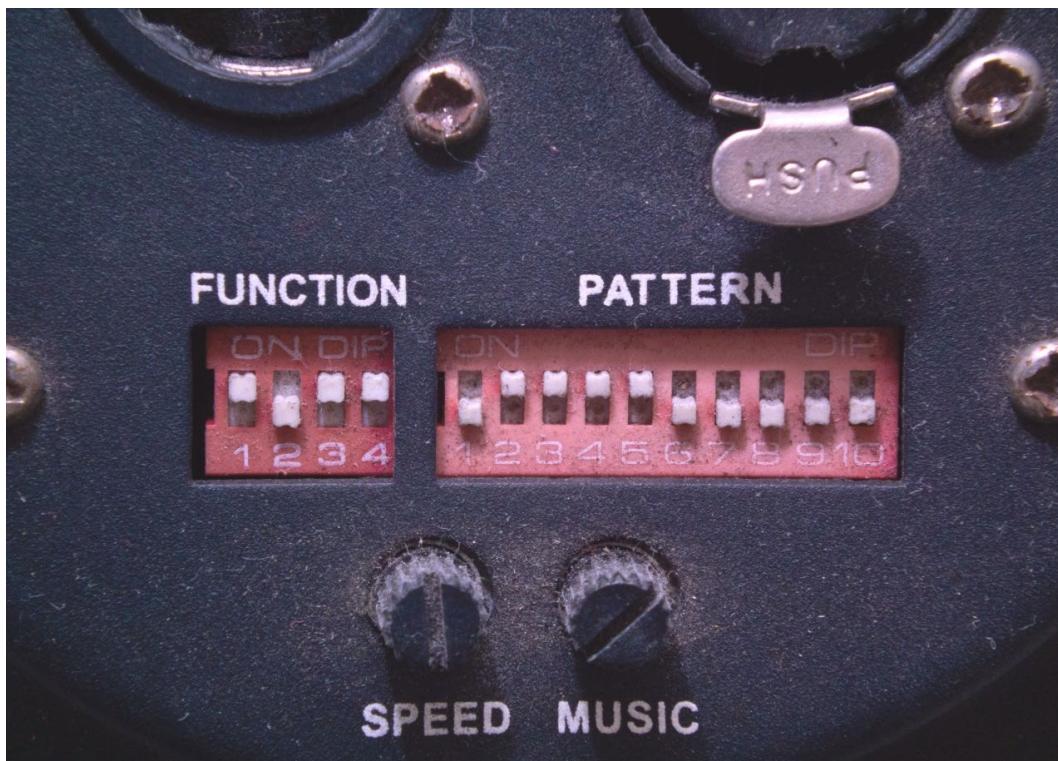


Figure 3-3: DMX LED dip switches

For each DMX channel, the data sent to control the device is an 8 bit binary string (i.e. $2^8 = 256$, so effectively any number between 0 and 255). This number may control the brightness setting on a single dimmer, or the amount of red in an RGB, or a gobo setting. It may also be used to ‘tilt and pan’ depending on the device.

The packet of data will consist of a series of binary digits for synchronisation, addressing and data. These packets will be passed to each device in turn (if ‘daisy’ chained) until matched. The number of channels addressable in one second will depend on the packet size and on the transmission rate. There will be an inevitable delay between command and execution if too many devices are ‘daisy chained’, which may be unacceptable.

However, the DMX protocol does not require unused channels to be included in the communication so a wise lighting operator will programme the desk to disable unused channels and, thus, reduce the total time to update all the fixtures.

It is important to note that, although DMX lighting uses XLR connectors, it is most inadvisable to use microphone cable to interconnect devices. The ‘characteristic impedance’ of the cable has a significant effect on the correct operation of the whole system so only cable marked for DMX use should be deployed. Similarly, it is highly advisable to terminate the cable chain with the correct characteristic impedance of 120 ohms. Failure to do so results in the digital signals ‘bouncing’ back along the line and interfering with the correct signal at various locations, resulting in unpredictable failures of lanterns to behave correctly. Termination plugs can either be bought or simply made using a spare XLR plug and soldering a 120 ohm resistor across the pins that connect to the signal wires.

Part 2: Processing

4 Power, Signals and Measurement

In this chapter, we explain the typical technical specifications found on electrical and electronic devices used within media.

Learning Outcomes

- To explain what is power and how it is measured
- To give an explanation of the signals used in audio and video, how they are measured and applied.

4.1 Power

In the connection of devices and systems it is important to match them correctly to avoid unwanted distortion or worse, e.g. burn-out and damage. This section is not designed to teach electrical theory but to give a simple understanding of the electrical details associated with typical media devices and systems found in their technical specifications.



MTHøjgaard

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

4.1.1 AC/DC

Alternating current (AC) refers to an electrical circuit that has a single cyclical phase that oscillates from positive to negative and back again. Three-phase systems have three AC conductors carrying a single phase each, having a uniformly offset phase relationship from each other. The main reason for three-phase is to balance the power loading to match the supply equipment and is likely to be found on major lighting systems within media applications but also within any major building. When correctly balanced, the load on the generator is uniform.

Smaller electronic devices (cameras etc.) need a lower level of constant uniformity of power, so use Direct Current (DC). DC current usually has a battery or power-supply source (linked to the mains). Unlike AC, direct current has no phase as it is not cyclical and does not vary with time.

4.1.2 Hz (Hertz)

The frequency of oscillation of the ac power system (i.e. the number of complete cycles per second) is measured in Hertz. In most of the world, the power system runs at 50Hz, though much of the Americas, parts of Japan and some other countries use 60Hz.

4.1.3 Volts

Volts are a measurement of the potential difference (pd) between a source and its reference level (usually but not always 0v). It provides the potential for current to flow around a circuit when connected by conducting wires. The higher the potential, the higher will be the current, for a given fixed value of load. When selecting equipment, it is important to ensure that the device is correctly matched to the ac mains supply for which it is intended (e.g. 240v or 120v etc). A device with a separate step-down power supply to a lower dc voltage will also need to match the correct voltage (e.g. 5v, 12v, 19v etc).

4.1.4 Amps (Amperes)

An Ampere is a measure of the electrical flow rate that is being carried by a conductor within a given time, so it can be envisaged as the ‘current’ flowing in the wire. A typical camcorder may use 1-2 amps.

4.1.5 Ohms (resistance and impedance)

Ohms are a measure of resistance (*resistance* is term used for DC circuits). If an electrical current is passed through a wire then the impurities and structure of the wire will impede the flow, so some of the current’s energy will be changed to heat in the wire. *Impedance* is a term that encompasses ‘resistance’ of the current in AC circuits, and ‘reactance’ as it also resists voltage change. Speaker specifications often refer to their impedance rating and power dissipation (Watt).

4.1.6 Watts

This is the unit of power dissipation. Consequently the greater the wattage rating of a device the more power it will dissipate, so the greater the power supplied to it must be in terms of source of potential energy (Volts) and/or the current which must flow through the circuit loop (Amps).

4.1.7 RMS power

RMS stands for Root Mean Square, it denotes the effective (complex) ‘average’ dissipation of the device – so is expressed in Watts. This average continuous consumption is often used in the technical specifications on speakers and other equipment.

4.1.8 PMPO (Peak Power)

‘Peak Momentary Power Output’ or ‘Peak Music Performance Output’ or ‘Peak Music Power Output’ (all these terms seem to be used for the same meaning), is a disputed measure of the maximum output peak, used for loudspeaker rating given in Watts. Its disputed accuracy means RMS is widely regarded as the better measure as it is scientifically accurate whereas PMPO is often used by marketing departments using unknown calculations in such a way as to favour their own product.

4.2 Audio levels

4.2.1 Audio Line-in, Line out and Mic inputs

These values will vary depending on whether domestic equipment or professional equipment is used and even vary between different types of professional equipment and different countries!

On domestic and consumer computer systems the audio line-in (light blue), line-out (light green), and mic input (pink) sockets are typically 3.5mm jack plugs. However their input levels are different and will either distort with overloaded signals or produce inaudible results if incorrectly applied.

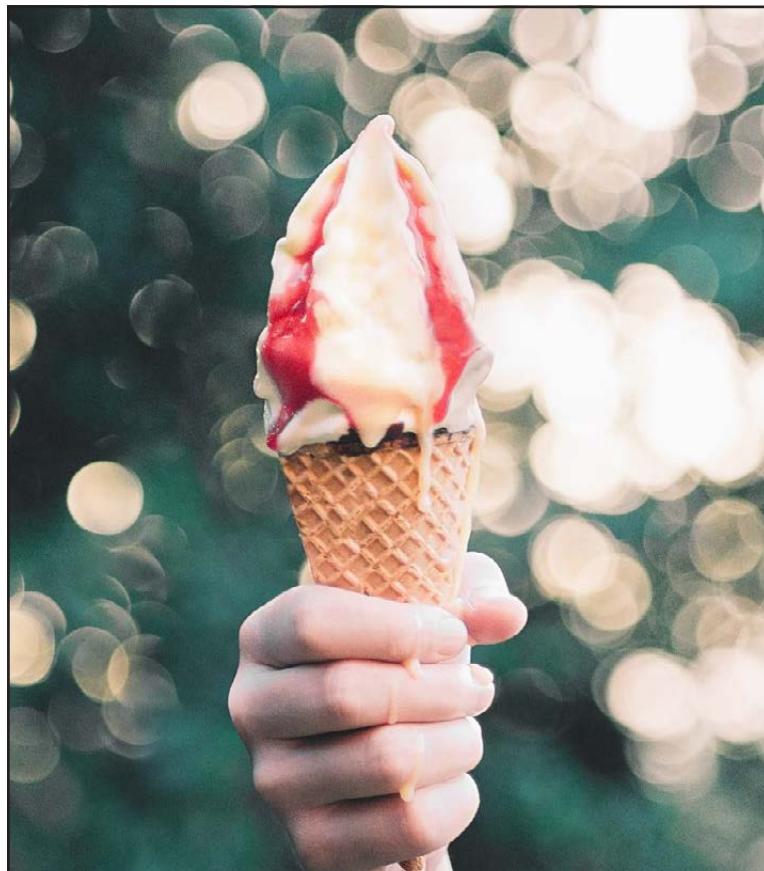
First, we need to distinguish between the different levels involved. Microphone inputs have very low voltage signals which will often range up to a maximum of a few hundred millivolts and much of the time down into microvolts (i.e. significantly less than 1v). The first stage of any device having a microphone input is a mic pre-amplifier. Clearly, if we were to try to insert a signal having a voltage that is already around 1v or more, then the amplifier will attempt to add yet more gain and the result will be a very large signal which will sound horribly distorted as the circuit will ‘clip’ the signal peaks and troughs.

The sort of signals that would already be at around 1v (or more) are those from the likes of CD players, electronic musical instruments (e.g. keyboards), radio tuners, mixing desk outputs etc. These are known as ‘line level’ (as opposed to ‘mic level’) signals.

Conversely, if we take a mic level input and try to feed it into a device expecting a line level input, the incoming signal is so small that even with maximum gain selected, the result will be barely audible and we will also end up amplifying the electrical noise in the circuit as well as the desired signal. Again, this is unsatisfactory.

We can use decibels (dB) to conveniently allow for quick calculation and reference but we must know which reference is being used, against which we are comparing the measured value. For domestic equipment, the reference is 1 volt. Hence, in absolute terms, we refer to the signal level in dBV. The alignment level of domestic equipment is -10dBV which is $316mV=0.316\text{Volt}$. It would be very unusual to get a microphone to work up to this value, except occasional peaks of very loud source material.

By contrast, professional equipment operates at a much higher level. The reference here varies depending on country and application. The reference value is referred to 0dBu. The u means that we disregard the impedance of the circuit (it is unspecified) but the voltage is 0.775Volt. For the broadcast industry, this is our normal alignment level and normal operation may see levels frequently peak as high as 10dB above this value, meaning voltages of 2.5V with occasional excursions up to a theoretical peak nearly 17dB above reference (5.5V) before the 'hard limiter' cuts in, on the new TV standards.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

To complicate matters a little, American standards for recording studios differ and many mixing desk manufacturers quote such standards. In those cases, the reference level is assumed to be +4dBu and the desk can handle up to 20dB above that. In other words, signals 24dB higher than 0.775V (i.e. 12.28V!) could theoretically be encountered. Compared with the alignment level of domestic equipment at 0.316V, it can now be seen why you can't just 'mix and match' professional with domestic but must use the correct equipment. Some professional mixing desks do have a few unbalanced connections for domestic equipment but domestic equipment is never designed to handle balanced professional audio. The correct way to deal with this situation is to use a proper balancing transformer which steps the voltages up or down accordingly, as well as isolating a correctly balanced system from the shared earth connection of an unbalanced system in domestic equipment.

4.2.2 Hi-Z and Lo-Z

Z stands for impedance, so devices are designed to be connected either Hi-Z to Hi-Z, or Lo-Z to Lo-Z. Optimum power transfer always occurs when the source impedance matches the load impedance. For example, trying to connect a low impedance load to a high impedance source will result in 'sinking' the voltage level as the load will be too large for the supplying device to cope, resulting in a severe mismatch in levels.

4.3 Video Levels

4.3.1 Analogue video

An analogue video signal (eg Y'PbPr) is measured in volts. The signal related to the monochrome (black and white) image is known as luma (Y) while the 'colour difference' signals convey only information about the hue and saturation of colours and are known as the 'chroma' signals, of which there are various types, Pb and Pr being one such pair.

In America, the luma voltages are converted to a representative scale known as IRE (Institute of Radio Engineers) and apply to standard definition images. The scale is from 0 – 100 IRE (black to white) but there are differences in implementation. USA starts at 7.5IRE as a legacy from NTSC systems to differentiate 'black' from 'video blanking', which consists of the parts of the image that are not visible on screen. Note, Japan starts at 0IRE even though it uses NTSC. Elsewhere in the world, IRE values are not used and the standards are simply quoted in millivolts. (1000mV=1V). The active 'payload' part of the luma signal ranges from 0mV (black level) to 700mV (peak white). For the chroma signals, such as Pb and Pr, they range from -350mV to +350mV. Any part of the image which consists of pure white, black or neutral grey conveys no information about colour and so the chroma signals at that point are at 0mV.

4.3.2 Digital video

Digital video signals convert to a simpler format of binary code just as you would expect on a computer system. The scale is from 0 to 255 (8bit representation for standard definition) and the luma set up is usually between 16 (black) to 235(white). Note some editing systems use sub 16 codes (1–15) for ‘sub-black’ and above 235 (236–254) for ‘super-white’, which allow for the undershoot and overshoot of rapidly changing ‘real life’ signals. Note 0 and 255 are never used as they are ‘reserved’ codes for synchronisation purposes.

Caution should be observed as going outside the limits of 16-235 can cause ‘black crushing’ or ‘white crushing’ in the image.

The chroma signals are, in effect, shifted up by 350mV so that they now range from the equivalent of 0mV to 700mV (i.e. the same range as the luma signal) and are coded so that the digital numbers range from 16 to 240. To distinguish a digital signal group from an analogue group, they are renamed as Y'CbCr

Again broadcasters will set safe limits for video material in their specifications and these should not be exceeded. This can be problematic when undertaking computer-based editing and special effects operations in R'G'B' since the signals, when converted to Y'CbCr, can be outside the allowable ‘gamut’ (range) of values and may need to be fed through a ‘legaliser’ which clips them to the allowable range but creates compromised results.

The advertisement features a night photograph of the Apollo Hotel building. A large red circular logo with a white lightbulb icon is on the left. The text "CISO Conference" is written in white, with "Produced by Inspired" below it. On the right, a white box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL Dec 5th 2019". At the bottom, a white box contains the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right of this text is the "Inspired" logo, which includes a blue lightbulb icon.

4.3.3 REC 601 and 709

When an RGB video signal is created in a camera, the luma signal needs to be calculated for conversion to TV systems that use Y'CbCr or Y'PbPr. The RGB signal has to be converted to a greyscale (Y') luma signal using a recommended formula. These are ITU-Rec 601 for SD TV standard, and ITU-Rec 709 for HDTV standard. The red, green and blue colours are weighted according to a formula based on the perceived difference in luminosity of the colours and these weightings are then applied to the RGB values.

$$\text{Rec 601 } Y' = 0.299R' + 0.587G' + 0.114B'$$

$$\text{Rec 709 } Y' = 0.2126R' + 0.7152G' + 0.0722B'$$

4.4 Measurement and Calibration

4.4.1 Audio

4.4.1.1 Test Tone

The reference test tone (alignment level) for audio equipment (including those used in video recording) is 1kHz at 0dBu (-18dBFS). It is normal to have this tone generated at the time of setting levels for audio equipment. Note: the tone can also have different regular interruptions (silences) to identify the different audio channels (stereo left/right). Surround sound mixes use a 'Blits Tone', consisting of a range of different frequencies which cycle around the specific loudspeakers. A test tone is often generated to accompany the video colour bars as per the specification of broadcasters at the start of recorded videos.

4.4.1.2 PPM (Peak Programme Meters)

Most domestic equipment (and many semi-professional recording studio desks) use VU (Volume Unit) meters. However, the analogue versions of these are relatively slow to respond and, therefore, under-represent transient peaks. LED equivalents, on the other hand, can be set to rise and fall very rapidly but are, therefore, much harder to read and can lead to eye-strain.

Professional broadcasters throughout much of the world rely on the PPM which is designed to rise very quickly to give a good indication of the peak levels (which are important to avoid 'overmodulation' of the transmitter in analogue systems such as AM and FM radio or distortion in recording devices) but then to hold the peak before slowly falling away. This makes them much easier on the eye and enables an experienced operator to achieve appropriate programme levels. They do not, however, show the true instantaneous peaks of extremely brief transients so would more correctly be known as quasi peak meters. True peak metering has become important for digital systems and true peak metering adopts some of the visual style of PPMs whilst capturing the true value of peaks for display electronically.

There are various designs of PPM, including multi-needle varieties and several scales, depending on location around the world. Some have two needles and can show two levels simultaneously. These varieties include two 'legs' A & B, which usually correspond to left and right and are shown as red and green respectively or 'Mid and Side' (M&S) or 'Sum and Difference' (white and yellow respectively), which show the level for a listener to a mono receiver plus the additional signal containing the 'width' of the stereo image. The white should normally be higher than the yellow. Reversal of connections to one channel will reverse the relative positions, helping to indicate a fault that cannot be observed by normal level meters. Normal operation would keep output below PPM6 (on the UK scale) but deliberate manipulation of signals (especially by advertisers) has led to the introduction of new average loudness metering (see below) for TV. PPMs remain a useful tool but cannot be used on their own to guarantee compliance to the new standards. At the time of writing, the existing PPM standards remain in use for UK radio.

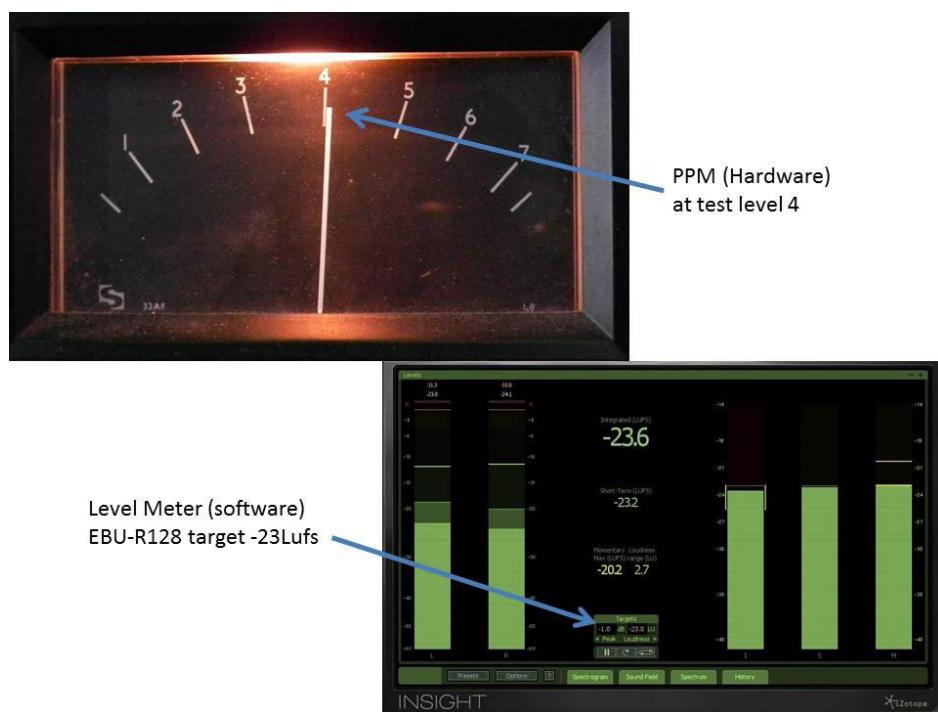


Figure 4-1: PPM and LUFs meters

4.4.1.3 LUFS (Loudness Units relative to Full Scale)

An EBU-R128 standard was introduced to European broadcasters to ensure the average 'loudness' (rather than peak level) of an entire programme (not merely an instantaneous or momentary value) remains within tightly specified limits and so the audio signal is normalised to an average value of -23LUFS where 1LU corresponds with 1dB but the LU value includes a clever 'gating' system which pauses the integrated measurement during quiet segments. This ensures that programme material that is too loud cannot benefit from long passages of silence or quiet ambient background and enables much more consistent audio loudness between different genres of material.

The new standards are intended to ensure that clever manipulation (dynamic range compression) applied by advertisers no longer results in deafening ad breaks between programmes or large shifts in level when changing channels. Material supplied for broadcast on TV will now need to meet this audio standard, which is being replicated around the world with corresponding standards such as ATSC A/85 in the USA, TR-B32 in Japan and OP-59 in Australia.

Those who will be responsible for submission of content to broadcasters are recommended to consult the “Practical Guidelines for Production and Implementation in accordance with EBU R 128” (<https://tech.ebu.ch/docs/tech/tech3343.pdf>)

4.4.2 Video

4.4.2.1 Colour Bars

Colour bars are used to allow calibration or matching of recorded video to edit and broadcast equipment to ensure the chroma and luma levels are correct. These can be generated by the camera, editing software or downloaded. Care should be taken to match the correct colour bar to the video format as per broadcaster specifications, (SMPTE for American, EBU for European transmission level or 100/0/100/0 – called 100% luma 100% chroma – for studio and digital recording levels in most of the world). To some extent, digital video no longer benefits to the same extent from the use of colour bars, since all levels supplied as numbers are, by default, unable to drift out of calibration. However, the bars are still useful when setting up displays, checking signals and demonstrating presence of a signal.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more



Figure 4-2 HD Colour Bars (still image screenshot taken from: Adobe Premier Pro CS6)

4.4.2.2 Test Patterns (Test Cards)

A test pattern, often referred to as the ‘test card’, is the television equivalent of the video colour bars. Again they differ for broadcasting (America vs rest of world) and for standard or high definition.



Figure 4-3: Test Card (Phillips PM5544)

4.4.2.3 PLUGE (Picture Line-Up Generation Equipment)

A suitable test signal exists to enable video monitors to be adjusted, called PLUGE and is generated by a test pattern generator. This consists of a zero chroma signal having four blocks of neutral grey up to peak white (110mV, 200mV, 450mV and 700mV), plus two black bars at +/- 2% of the peak video signal (+/- 14mV) centred about the black level. When fed to a monitor, this enables the contrast control to be set such that the peak white is measured at 80cd/m² (Candolas/square metre) and then the brightness control is adjusted so that the -2% black disappears into the 0% black that surrounds it, while leaving the +2% black bar just perceptible above 0% black. Non-broadcast flat panel displays may have a problem if they clamp the black level input so that the -2% black bar is incorrectly raised to 0mV, destroying the ability to differentiate between them.

4.4.2.4 Vectorscopes

The vectorscope (a standalone piece of equipment or found within video monitoring equipment or editing software) is a form of oscilloscope that simultaneously plots the chroma values of all of the pixels in a video frame on a graticule, i.e. a graphical display overlay showing a graduated colour distribution, rather like a ‘colour wheel’, where white, black or neutral grey are at the centre.. The hue of a colour is represented by the angle around the circle while the degree of saturation of the colour is denoted by the length of the vector from the centre. Note there is a difference between graticules depending on NTSC/PAL and HD as the colour bar limits are depicted on the graticule. The axes are Pb (or Cb for digital) on the horizontal scale and Pr (or Cr for digital) on the vertical scale.

Within editing, vectorscopes are used to look at the colour (chroma) of a single frame along with a variety of other signal waveform displays. The vectorscope should show a distribution of saturated colours around the centre – tightly packed at the centre suggests a lack of saturation which would be displayed as ‘washed out’ colours while an offset from the centre may mean the image has a colour cast or bias which might need correcting.

Other typical video editing scopes include:-

RGB Parade – shows the Red, Green and Blue channel levels – to check for colour ‘crushing’ and levels within each colour

Y CrCb Parade – shows the balance and difference between the chroma signals (red-related and blue-related) against the overall luminance.

Y-C Waveform – shows the overall luma and chroma levels within the image.



Figure 4-4: Scopes (still image taken from: Adobe Premier Pro CS6)

4.5 Synchronisation

4.5.1 Genlock

In general, many video signals are being worked on at the same time in a studio. This can only be undertaken if all these signals are synchronised. This means that the scanning spot reaches any given point on the picture (e.g. the top left) on all devices simultaneously. If the signals were out of sync, you tend to see a dark line appear horizontally on the screen that progressively moves vertically downwards. A typical aberration that comes from videoing a TV screen with a camera which lacks a synchronising genlock signal.

4.5.2 Framestore

An alternative to genlock is a framestore system, where unsynchronised inputs are presented to the vision mixer. Inside, digital memory holds (at any given moment) at least a frame of each input video signal but each frame is then ‘played out’ of memory synchronously with all the other inputs that are being treated similarly. In other words, it delays them all so that they can be played out ‘in sync’. However, this then results in at least one frame delay between ‘real life’ and the output, so the audio must also be delayed by the same amount to avoid loss of lip sync, which complicates matters somewhat. Sometimes, the traditional methods are better, though the self-sync mixer can be handy in portable applications as it means less equipment needs to be transported and wiring is simpler.

5 Video Sampling

In this chapter, we cover the sampling and compression of video signals

Learning Outcomes

- To explain what is sampling why it needs to be performed.
- To give an outline of typical methods of compression.

All digital video cameras convert an analogue signal into a digital form, by representing the value of the signal at measured intervals as a series of binary digits. The process is called digitisation and the binary value has the advantage of being reproducible without degradation but the disadvantage of losing some of the initial signal value accuracy within the quantisation process.

5.1 Bits and Bytes

Bits: A **binary digit** is called a bit. It can be used to represent something being present or not and, when applied to a binary scale, a value.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

Bytes: Eight bits make a byte. This has been a basic building block of computer systems since the earliest digital machines. In general, the number of possible values is 2^n where n= number of bits. The highest value that can be represented is 2^n-1 . A byte thus represents the values from 0 to 255 (making 256 values in total). It is the positions of the bits along the scale that create the value when processed. In binary this is $2^8 = 256$ possible values but remember that the first possible value is zero, hence the highest possible value in 8 bits being 255, not 256. Take an example of the binary value 10101010.

Binary weight	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Decimal weight	128	64	32	16	8	4	2	1
Bit presence	1	0	1	0	1	0	1	0
Decimal value	128	0	32	0	8	0	2	0
Total = 170 (128+32+8+2)								

Figure 5-1: Binary Example

Instead of having a time varying analogue voltage signal, the signal is sampled at regular intervals and numbers representing the magnitude (size) of the analogue voltage is stored in digital format. This being the natural language of computers, it enables the analogue waveform to be recreated with great repeatability. Now, instead of considering a line of video as just a voltage waveform, we can also consider it to be a series of numeric values.

Your computer monitor may use 24 bits (i.e. three sets of 8 bits to represent red, green and blue respectively – also known as ‘true-colour’) which is 2^{24} or $256^3 = 16.7$ million colours. Some Windows users may notice ‘true colour (32 bit)’ on their monitors. The last 8 bits are for the alpha channel used for translucency; it does not provide any more colours. Graphics cards may do more than 24 bit, some have 10 bits or more per colour channel. These are referred to as ‘Deep Colour’ of which the 30 bit implementation is common. You may wish to calibrate your monitor to use the sRGB setting which will more closely match any videos you are playing on a computer, as most DSLR camcorders/cameras use the sRGB colour profile. However, it should be borne in mind that, for broadcast video, a YCbCr colour space will normally be in use to Rec601 (for standard definition) or Rec709 (for high definition). The need to import, convert, display and output your material in the correct format will require some thought to ensure that you are utilising the correct colour space for your application and equipment.

5.2 Composite and Component

The signals representing voltage proportional to the quantity of light of each primary colour (red, green and blue) arriving at their respective sensors is usually converted to an equivalent monochrome signal (from black, through neutral greys, to peak white), known as luma or Y'. (This is often referred to incorrectly as 'luminance' but the signal is derived from the "gamma corrected" RGB signals (designated as R'G'B', the full details of which need not concern us here) and the term luma is used to differentiate the resulting signal from that which would otherwise occur in a system having a linear relationship, such as luminance.)

In order to convey colour, two other signals, known as the colour difference (or 'chroma') signals are also used. These are, basically, scaled versions of B'-Y' and R'-Y'. The purpose of this apparently pointless conversion is actually very significant, for it makes use of a property of human vision that enables benefits to be bestowed on electronic systems recreating images.

The human vision system is capable of resolving fine detail in content which enjoys high *contrast* (i.e. changes between very dark and very light). This, however, does not apply to fine detail in which only the *colours* change. Thus, we require a *monochrome* signal which conveys the information about fine detail. A video signal carrying finely detailed information requires high frequency signals to represent that detail and, hence, a full bandwidth luma signal needs to be generated in order to accommodate the high frequencies.

Conversely, the information being carried in the chroma signals (which convey only the additional information necessary to add the colour) does not need to carry the high frequency (fine detail) information signals, since the eye cannot resolve the fine detail changes of colour anyway. Thus, the bandwidth of the chroma channels can be limited and, hence, it is possible to provide all the information required for vision in double the bandwidth of a single channel, rather than three times the bandwidth of the single channel, as might otherwise be expected. (This, in turn, reduces storage space required and enables a greater number of channels to be carried by services.) Within a production environment, however, there is sometimes a desire to maintain the full bandwidth of all the signals, especially when visual effects such as computer-generated imagery (CGI) or colour separation overlay ('chromakey') are to be used, as the retention of the full information reduces errors which might otherwise become visible in the final image. The conversion to reduced bandwidth chroma signals then takes place after the image manipulation has been completed and the signals are sent on their way.

The methods by which these signals are stored and transmitted may be encountered in a variety of combinations. The basic terminology falls into two main groups: Component or Composite. The Component signals are those in which the chroma signals are kept separate from the luma signal. When these are all combined on to a single cable or channel (using some clever mathematical 'modulation' techniques), the process results in a 'composite video signal'.

Typical R'G'B' connections in a domestic environment include computer VGA and TV Scart cables. These have largely fallen out of favour since the advent of HD, so the High Definition Multimedia Interface (HDMI) is now more likely to be used. Arguably, these are a form of component video, though the common parlance rarely refers to them as such. To confuse the issue, a Scart connector will often also carry a composite version of the component signals and the settings of equipment in use must be checked to see which connections are actually being utilised.

The chroma signals ($B'-Y'$) and ($R'-Y'$) are scaled differently for different purposes. For one set of signals, they become known as U and V respectively. When combined with Y, the set of Y'UV signals are combined to make a composite analogue PAL signal. (Similarly for Y'IQ signals to create an NTSC signal in the Americas and Japan.) In practice, it is extremely unlikely that a media production student would encounter genuine Y'UV signals, since U and V are only ever created as an intermediate stage to generate an analogue PAL composite signal, so they usually only exist inside a piece of equipment. Since analogue composite signals such as PAL or NTSC are carried on the same cable and within the same bandwidth, some interference between the luma and chroma signals can occur, resulting in errors such as 'cross colour', recognisable as the distracting red/cyan and blue/yellow strobing patterns that occur on fine details such as checked shirts etc. Even though analogue composite signals are not usually created these days, there are millions of hours of legacy material that was recorded in that format and which may continue to be used for archive-based shows, so the more astute student will appreciate that this is still worth some consideration.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

If an analogue studio is used, the most likely signal set to be found would be when the chroma signals are scaled so as to give a voltage range of $\pm 350\text{mv}$ (i.e. a total range of 700mv , centred about 0v). This set is then known as Y'PbPr and is carried on three coaxial cables, using BNC connectors. The signals remain separate and, thus, component video is the designation, the separation ensuring that no cross-colour effects (and other distortions) are seen.

In a digital studio, the same principle and scaling factors apply as to Y'PbPr. However, the chroma signals are, in effect, shifted up by 350mv and thus range from the equivalent of $0\text{-}700\text{mv}$, like the luma (Y') signal. (The values employed in US/Japanese signals differ slightly in value but are close to these ranges.) This has the advantage that there is no need to worry about negative values in the chroma signals. To maintain differentiation from the analogue component signals, the designation for the digital component set becomes Y'C_bC_r rather than Y'PbPr. In practice, since the three signals are digitised separately and can be transmitted along a cable (or stored) as a carefully separated sequence of numeric values without interference to each other, this is still a form of component video, even though it is possible to convey all the information on one cable (or channel space), using the same physical infrastructure as was used by the analogue PAL composite signals that preceded it (75 ohm coaxial cable and BNC connectors). This method of transmission is generally known as SDI or Serial Digital Interface.

In the domestic environment, composite analogue PAL/NTSC/SECAM may be found on the AV cable from domestic camcorders and DVD players etc. and is often a simple single 'phono' (RCA) connector. However, like the R'G'B' set, since the development of HD, this is obsolescent so the High Definition Multimedia Interface (HDMI) is now more likely to be used.

Some domestic equipment also delivers Y'PbPr but the cables using simple phono (RCA) connectors for this signal group are often a triplet having plugs coloured red, green and blue providing opportunity for confusion with R'G'B' signals. It probably makes sense to use the blue connector for Pb, the red for Pr, leaving the green for the monochrome signal (Y').

The only remaining signal group that might be encountered in a domestic situation is also obsolescent. This is where both the chroma signals (Pb, Pr) are carried on one cable (designated 'C') while the luma signal is carried on a second cable (Y'). These often tended to arrive on a small multipin plug. Since the chroma signals are kept separate from the luma, no cross colour distortions are produced and this is a form of component video.

There is much ‘looseness’ in the terminology deployed, which can result in considerable confusion. Strictly speaking, PAL, NTSC and SECAM signals are only ever analogue composite signals. Although it is technically possible to digitise the already compromised analogue composite signals, this is never done in practice. Thus, there is no such thing as digital PAL or digital NTSC. (SECAM is, to all intents and purposes, dead and buried and will not be encountered these days anyway.) There is certainly no such thing as HD PAL or HD NTSC. However, these terms do crop up. Almost invariably, their purpose is to denote the refresh frequency of the system in use (50Hz for PAL countries, 60Hz (approx) for NTSC countries.) Strictly speaking the terminology should refer to The Americas/Japan rather than NTSC and Rest of World rather than PAL.

Similarly, the term Y'UV is often bandied around to indicate a non-composite signal. It may refer to genuine Y'UV (very unlikely these days) or more often to Y'PbPr or Y'CbCr. There is also confusion as to the numeric values used. Computer-based operators will assume a range from 0 for black to 255 for white in the Y' signal. Professional video engineers use 16 to 235 respectively, which allows for undershoot and overshoot, which happens in ‘the real world. (Similarly for chroma signals, 16-240). However, DSLRs recording R'G'B' signals (rather than Y'CbCr) often do use a range from 0-255 and an operator importing material to an edit station (for example) will need to check as to the colour space and value ranges in use so as to select the correct format to import. Finally, it is very common to encounter terminology in which the prime symbol to denote gamma corrected sources (') is missing. (Indeed, we have reverted to a number of these ‘shorthand’ notations throughout this book). As always, it is necessary to have an understanding of the concepts and principles involved so as to be able to extract the real intended meaning from the context!).

5.3 Sampling

5.3.1 Sample value bit size

For a binary number, the maximum number of steps to take us from zero to full scale is determined by the number of bits available. The relationship is $X=2^n$ where \times is the number of states (or levels of voltage, for example) and n is the number of bits. With just one bit $X=2^1 = 2$, so we can only record two states (0 or 1); with two bits, $X=2^2 = 2 \times 2 = 4$, we can record four states (00,01,10,11), with three bits, $X=2^3 = 2 \times 2 \times 2 = 8$ so we can record eight states (000, 001, 010, 011, 100, 101, 110, 111) etc. Thus, with 10 bits, we can record 2^{10} (=1024) different steps from black (0v) to white (700mV) of the luma signal. This ensures that there are so many steps from black to white that the eye cannot see any transition from one step to the next.

Although using a small number of bits results in smaller files (or a lower transmission rate), if we use too few steps, the result is a noticeable degradation of the signal – see figure 5-2. If we use too large a number of bits, we may be making larger files (or needing higher transmission rates) than we need, with no visible benefit.

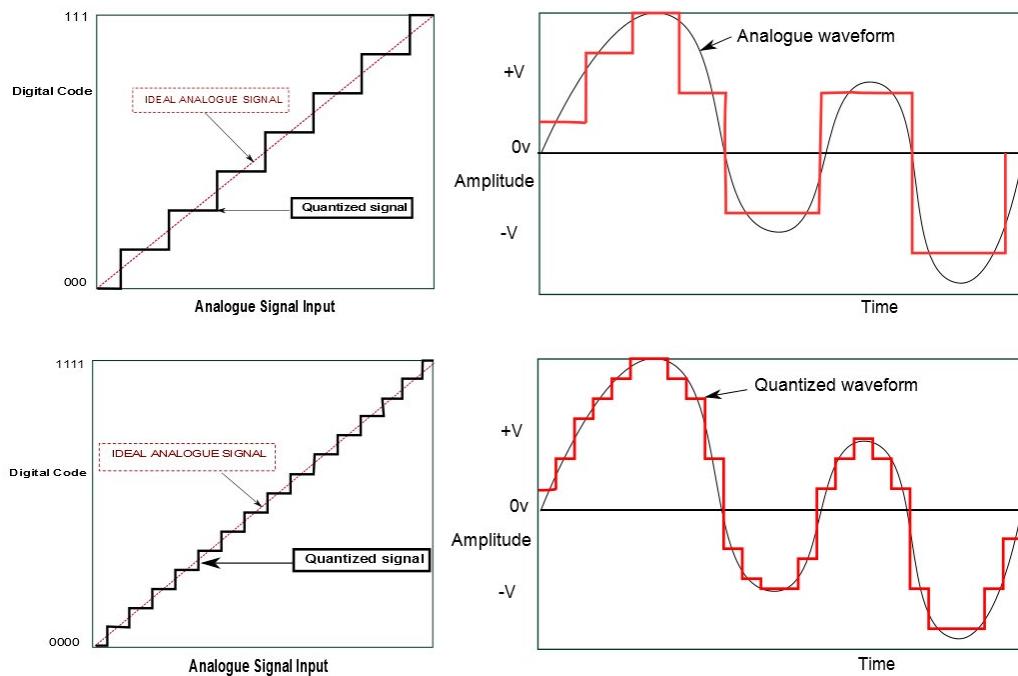
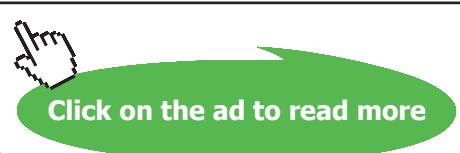


Figure 5-2: Quantization diagram

The advertisement features a large image of the Apollo Hotel building at night. On the left, there's a red lightbulb icon and the text "CISO Conference" followed by "Produced by Inspired". On the right, there's a white box containing the address "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and the date "Dec 5th 2019". Below the main image, there's a call-to-action: "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". At the bottom right is the "Inspired" logo.



On a video signal, insufficient quantisation levels might look like the pictures below (see figure 5-3):



Figure 5-3: Quantization Picture

Generally (for final viewing) 8 bits (giving 256 levels) will be sufficient for each of the three component signals but 10 are sometimes recorded, especially for high definition, to allow for digital manipulation of the signal. (For high end ‘digital intermediate’ processing (e.g. digitising some ‘film’ formats) up to 12 or even 14 bits may be used for more accuracy.)

The final output may then be converted down to 8-bits for final transmission, which allows each of the luma and colour difference (chroma) signals to be distinguished into one of 256 different levels. In practice, with each of the 3 signals having 8 bits, that allows the finished picture to have up to 2^{24} (= 16.7 million) different combinations! In practice, this is often referred to as 16.7 million colours. That is not strictly true as we are working in a form of “YUV” (not “RGB”) here. What it really means is that there are 65536 combinations of hue (colour) each of which has 256 brightness levels available. However, it needs to be borne in mind that not every possible combination of Y’CbCr which can be generated (or which results from manipulation of the image) will result in ‘legal’ values of “RGB” when converted back in the display device. Thus, when manipulating Y’CbCr values (e.g. colour correcting, grading or special effects), care must be taken to ensure that only values that will result in ‘legal’ “RGB” combinations are produced. Software and hardware to restrict signals within such values are available and are referred to as ‘legalisers’.

5.3.2 Number of samples per image

Each pixel of an image needs to be sampled so the number of samples will depend on the video image size eg. HD at 1920×1080 , or 1280×720 , or standard definition such as 720×576 .

In 1080 HD the total number of pixels is $1920 \times 1080 = 2,073,600$ i.e. 2.07 million pixels per frame.

If we take as an example European standard-definition video, we know the answer as we go from top to bottom: there are 576 visible lines, 625 lines in total.

As we go across the picture, it has been agreed that the number of samples (separate pixels) will be 864 on each complete line, of which 720 can be populated with visible video signal. (In fact, within that 720, the central 702 are where the traditional analogue signal would be placed – a slight adjustment was made when designing the move from analogue to digital signals but this is not hugely significant in most cases – it's just useful to be aware of it in case documentation is encountered referring to 702 samples.)

If we did not sample sufficiently, pixellation would become visible – see figure 5-4. (This is the equivalent on video cameras of using ‘digital zoom’, which is always to be avoided.)

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Figure 5-4: Pixelation

When sampling a waveform, we end up taking the samples at regular intervals, thus:

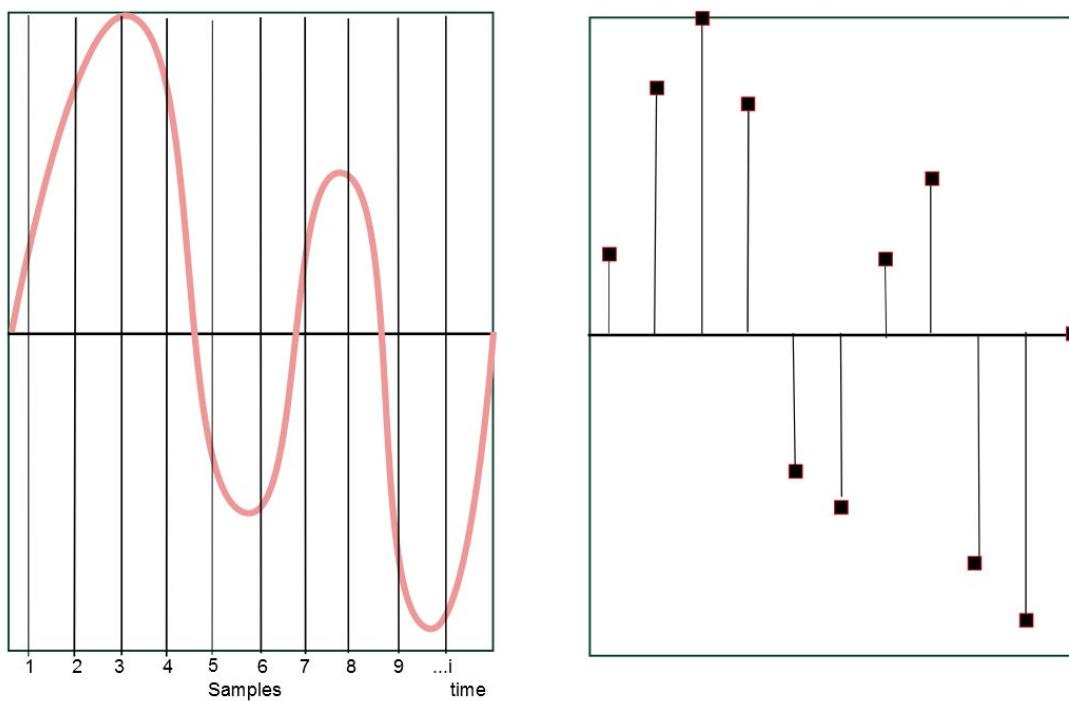


Figure 5-5: Sampling diagram

It should be possible to see that the impulses which show the moments at which the signal on the left was sampled have taken on the height which represents the magnitude of the analogue signal, as shown on the right. By recording a numeric value for each of the samples, it is possible to store numbers representing the height of the waveform at those moments. If we want to reproduce the original waveform at some time, we can use a digital to analogue converter, which takes in the numeric values and provides a voltage proportional to those values. Provided the samples are frequent enough, the output voltage will follow the original shape quite well.

However, if the samples do not occur often enough, the re-constituted waveform might not be correctly reproduced (known as ‘aliasing error’), as shown in figure 5-6:

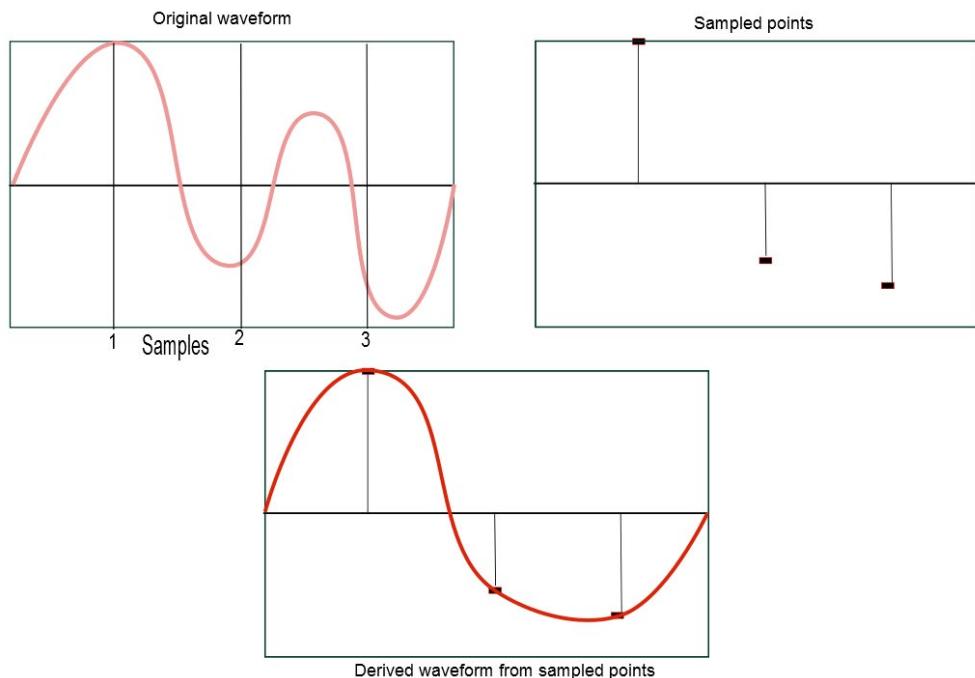


Figure 5-6: Aliasing error

In the top left picture, the waveform is sampled infrequently, at the times shown. In the top right, the height of the samples represents the magnitude of the waveform at those moments. In the lower diagram, it should be possible to see that the reconstituted waveform does not match the original at all but at the moments when the samples were taken, the heights do match! This is an example of aliasing error. For accurate reproduction, samples must always be taken with a frequency that is at least twice the highest frequency we want to be able to reconstruct (i.e. twice as often as the highest frequency’s cycle time). This is commonly known as the Nyquist theorem.

5.3.3 Sampling Rate

The video sampling rate determines the number of times a video line is sampled to measure its light intensity. For standard definition, the luma signal has a top frequency of around 5.5MHz (i.e. 5.5 million black-to-white-to black cycles per second can occur (in a UK signal)). [Note that some countries allow up to 6MHz.] Thus, our samples need to occur with a frequency of at least 11MHz [12MHz] if we are to avoid aliasing errors. In practice, a frequency of 13.5MHz was chosen, as this is the nearest frequency that allows both the European (625/50) and American (525/60) systems to share a common standard. In other words, 13,500,000 values corresponding to the luma level are taken every second, each sample corresponding with one pixel location. In HD, it should be fairly obvious that there are many more samples to be taken in each second and the luma Y channel is sampled at 74.25 MHz i.e. 74,250,000 times a second.

5.3.4 Sampling Ratio

The sample values need to be stored but, for general-purpose application, the ratio of sample storage can be reduced to save on storage space, as outlined above in terms of the bandwidths. The signals conveying information about colour (Cb, Cr) have lower maximum frequency (bandwidth) and, hence, do not need to be sampled as often as the luma (Y') signal.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

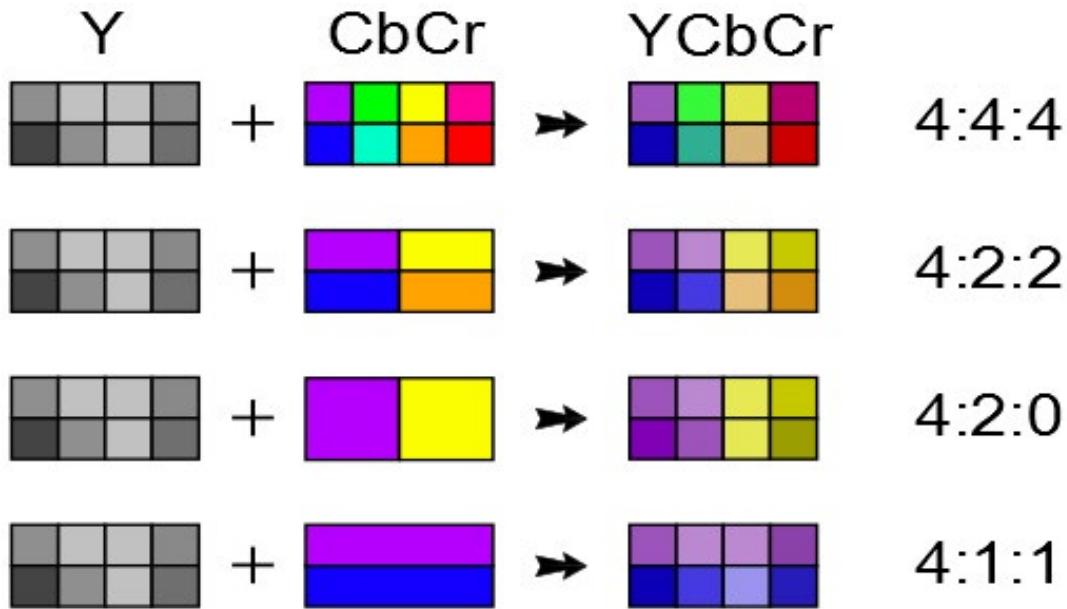


Figure 5-7: Chroma subsampling

5.3.4.1 Subsampling Ratio 4:4:4

This is a ratio for raw uncompressed footage. It means that for every 4 samples of Y there are also 4 samples of Cb and 4 samples of Cr stored as well. Since there is no advantage in transmitting all three Y'CbCr type signals at full bandwidth, 4:4:4 is normally only used for 'raw' R'G'B' signals in specialist studio or computer post-production areas. (Sometimes, 4:4:4 will be seen – this is where an additional 'alpha channel' or 'key' signal is being carried along with the R'G'B' signal for chroma-key or CGI purposes.)

In 4:4:4 sampling, every single pixel location has values for R', G' and B', so the final display can recreate the exact colour and brightness that was sampled at the time of the image's creation for every single pixel location.

5.3.4.2 Subsampling Ratio 4:2:2

We already know from previous sections that we don't need the same resolution for the C_B and C_R signals as we do for the Y'. For historical reasons that we don't need to worry about here, all the sampling rates for SD video are referred against a base frequency of 3.375MHz. Thus, the luma signal is normally sampled at 4x that reference frequency ($4 \times 3.375 = 13.5$ MHz) while the Cr and Cb samples are taken at 2x the reference frequency (= 6.75MHz). (This is more than adequate to sample the colour difference signals as it would mean that Cr and Cb bandwidths up to 3.375MHz would be possible.) Thus, the terminology for normal Y'CbCr digital video sampling frequencies is 4:2:2.

When sampling the Y signal, we have seen that 864 samples would occur over the full 64 μ s (European) line period. This means that each sample would occur $64\mu\text{s}/864 = 74.074\text{ns}$ (i.e. every 74 billionths of a second!). Now, $T=1/F$ or $F=1/T$ so that would correspond with a frequency of $1/0.00000074074 = 13,500,000\text{Hz}$. In other words, 13.5MHz, which matches what we would expect. Only 720 of the samples carry the **active** luma part of the transmitted signal – the rest contains the parts of the image that aren't seen (blanking) within which is positioned a synchronising pulse which ensures that all the video devices connected to the system start their scans at the same moment. (In practice, digital video doesn't need to transmit all the blanking and sync pulse samples since special numeric codes can be sent ahead and behind each line of the active video image to signify SAV (Start of Active Video) and EAV (End of Active Video).)

However, since the samples for the C_b and C_r signals only occur at half the frequency, this means that they are spread apart by twice the distance on the screen. (i.e. there is only one sample point of the chroma signals for every two samples of Y). Thus, 360 samples will represent each of the **active** chroma signals in the transmitted signal (the 'payload' area). This is often known as 'sub-sampling'. (There are those who argue that, since the actual samples were often taken at full frequency and then half of those were discarded, it should strictly be known as 'down-sampling', since 'sub-sampling' implies that only half the samples were originally taken in the first place. Either way, the result is the same!)

Normal transmission around a studio complex would be of Y'CbCr signals in 4:2:2 format. All these standards are defined in a universal standards document: ITU-R BT.601 (often known as Rec601). The implementation (via Rec656) is known as 'Serial Digital Interface' or SDI and enables all three signals to be carried on a single 75ohm coax lead, as previously used for analogue PAL composite signals but (unlike PAL) still enabling perfect separation of the three components, avoiding the problems such as cross colour which plagued composite signals such as PAL. Unlike PAL transmission, however, the signal will only transmit up to 100m or so before it needs to be re-timed and boosted.

5.3.4.3 Subsampling Ratio 4:1:1

For 'prosumer' and transmission of digital video to the end user, even more bandwidth can be saved by reducing the sampling still further.

In America (though sometimes encountered here in JVC's DVCPRO25 system), another standard only uses 180 samples per active line for the C_b and C_r signals. This is called 4:1:1 sampling. In this case, the 'averaging' process to reconstitute the chroma samples is getting pretty stretched and it will be realised that the errors in the chroma channels will be greater than in 4:2:2 or 4:2:0 (see below).

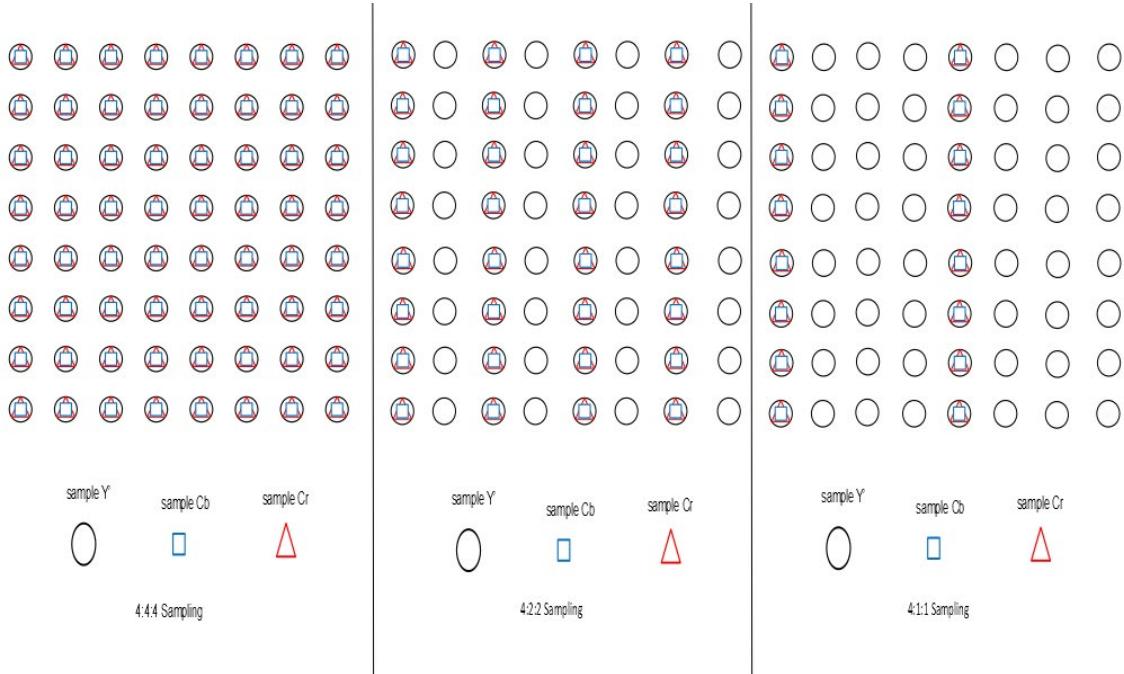


Figure 5-8: Sampling structure

5.3.4.4 Subsampling Ratio 4:2:0

Regions that did not use NTSC in analogue days tend to use a system somewhat eccentrically called 4:2:0. There are, in fact, several different variants of 4:2:0 sampling, depending on the function. The examples below show a progressive scan version in which chroma samples are recorded interstitially (i.e. providing a notional position for the sample which does not actually match the true position of the pixels that are used to create it) while the interlaced version shown is one of several that can exist, depending on the scanning standard used. The interlaced version shown is for European standard-definition TV. The aim of 4:2:0 is, like 4:1:1, to reduce the required data rate and, indeed, this is achieved to the same degree. However, unlike 4:1:1 in which full vertical resolution is maintained but the horizontal resolution is severely limited, 4:2:0 schemes attempt to spread the sampling in a more uniform way in both dimensions. Referring to the interlaced example below, it can be seen that the lines of the upper field have chroma samples that are Cr along one line and Cb along the next (i.e. only 180 chroma samples per line compared with 360 in 4:2:2 and 720 Y' samples in all the schemes). In the lines of the subsequent lower field, the same applies. Although at first glance there appear to be two lines of Cr followed by two lines of Cb chroma samples, it needs to be borne in mind that the lines of the upper field and those of the lower field do not occur simultaneously but are 1/50s apart, from a temporal point of view.

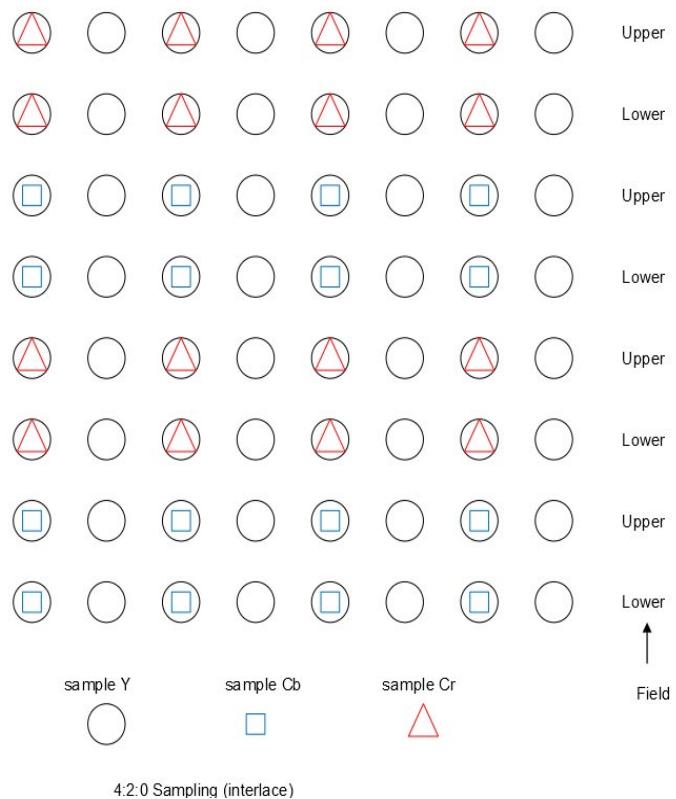


Figure 5-9: 4:2:0 Sampling structure



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

The transmission system or DVD writer (MPEG encoder in both cases) sets a flag which tells the decoder (set top box or DVD player, respectively) which method to use to reconstitute the chroma samples during upscaling, (often seen as a Field Order selection in Non-Linear Editors). Early DVD players often failed to implement this correctly, resulting in very poor quality images with jagged horizontal coloured stripes on diagonal edges when material that originated in progressive format (e.g. film) was decoded incorrectly using the interlace method.

Even when correctly decoded, it should be realised that, while it is possible to use the 4:2:0 signal to recreate a signal having values of Y', C_b and C_r at the same positions as a genuine 4:2:2 or 4:4:4 signal, the actual information that is available is already limited so it is inevitable that some errors are, therefore, introduced, resulting in a compromised image.

4:2:2 needs more samples (and hence more bits) than 4:2:0. On the other hand, 4:2:2 has more colour resolution. This could be critical if downstream processing had to be carried out on the signal.

4:2:2 sampling, whilst not quite as good as 4:4:4 is generally regarded as being sufficient for normal broadcast production and is the lowest standard that will normally be accepted by the major broadcasters such as the BBC, ITV, C4, SKY etc.

5.4 SDI interface

As we move into a digital world, we like to connect different pieces of kit together digitally. In a studio environment, the separate signals for Y'CbCr are being replaced with Serial Digital Interface (SDI) which is designed around standard definition 4:2:2 signals.

If 4:2:2 sampling is used, the overall number of samples generated per second is 27 million (13.5 million for Y', 6.75 million each for C_b and C_r). Allowing up to 10 bits to represent each of these values means the overall bit rate is 270 Mbit/s. This is fast, but is only normally used to connect kit in a studio. It is absolutely critical to use the correct 75ohm coax cable and to terminate the lines with 75ohm terminators, as it was with the analogue signals. However, with an analogue signal, incorrect termination results in too light or too dark an image – the error can be seen and corrected. With SDI, the high-frequency signals are blocked if there is any reflection on the line due to incorrect termination so there is simply nothing at the output – the system either works or it doesn't (the 'Digital Cliff' effect), which makes fault-finding more awkward.

Assuming we are using an 8-bit coding system, it might be thought that we could represent numbers ranging from 0 to 255, which is how computer scientists tend to treat signals. However, this does not mean that 0 represents black and 255 represents peak white (although some software does do this, which then has to be re-scaled afterwards). Instead, on the Y' signal, black is assigned the level 16 and peak white is assigned the value 235. This allows for minor overshoots, mis-adjustments and so forth, without the number folding round (remember that 255+1 would appear as 0 in an 8 bit system!).

As previously mentioned, a special offset coding system is used to represent the two ‘colour difference’ signals for CbCr. This means that the mid range that would have been 0v in PbPr (analogue) is now (in effect) 350mV in CbCr (digital) and is assigned the value 128 in an 8-bit digital regime. (128 is the half way value between 0 and 255). The maximum values for Cb and Cr in 8-bit digital systems are 240 while the minimum values are 16. (Note the different max values between Y’ (235) and Cr,Cb (240)! It doesn’t matter, since all equipment is designed for it but you need to be aware that it’s different, especially if you are manipulating the values in a computer (for example, during colour correction or SFX), as any scaling has to be re-applied differently.) If working in 10-bit then these values would be multiplied by 4.

In place of line blanking, special codes are used to indicate start of active video and end of active video. When the serial stream is formed, it goes through a digital scrambling circuit to provide reasonably frequent changes of the signal between zeroes and ones in the serial stream. This eases the design of handling circuits, which synchronise themselves to the signal using the transitions between zeroes and ones. (If there was a long period of just ones or just zeroes, the equipment would lose sync.) There is a corresponding de-scrambler at the SDI decoder.

This means the analogue sync pulse does not have to be in the range of the brightness signal, nor need it be transmitted in full. The time ‘saved’ by not transmitting the full sync pulse can be put to use carrying chunks of digital audio data that can be buffered at the receiving end, meaning that it is often possible to use a single coax lead to interconnect equipment with a component video and stereo audio signal set which would have required five cables in an analogue component environment.

5.4.1 SD-SDI (Widescreen)

So far, all calculations have been based on the 4:3 (=12:9) aspect ratio. For widescreen, there are two ways of dealing with the enlarged image.

One way would be to take more samples across the 16:9 image to maintain the same pixel size. In SDI, there is provision for this and the sampling rate would increase to 18MHz giving 960 active samples per line. The SDI transmission rate would, therefore, increase from 270Mbps to 360Mbps. However, it is quite rare to see this implemented.

The other option, which is far more common, is to use the ‘full height anamorphic’ technique. In this, the same sampling system as 4:3 aspect ratio is used (i.e. 720 samples per line) but this means that each sample is ‘stretched out’ horizontally to cover a wider pixel shape. The advantage is that traditional 4:3 production equipment can be used (although special effects involving circles in the live signal chain (e.g. vision mixer) will become oval unless also being viewed on 16:9 monitors) but it allows for greater versatility in the deployment of most equipment. When viewed on a 4:3 monitor, the image appears squeezed, horizontally, so people appear tall and thin but look fine on a 16:9 monitor.

5.4.2 HD-SDI

High definition Serial Digital Interface is now used in studios in preference to standard definition. For 720p/50 (720p/60 or 720p/59.94) and 1080i/25 (1080i/30 or 1080i/29.97), HD-SDI has a nominal data rate of 1.485Gbits/s. As SDI and HD-SDI are generally for high end professional cameras and licensed technology, HDMI is often provided on camcorders and DSLRs. High-end studio production also uses 1080p/50 (1080p/60 or 1080p/59.94) which has the advantage of easy down conversion to any of the lower standards but the equipment and signal routing costs are very high, as the data is now being transmitted and stored at around 3Gbps.

5.5 Digital Video Compression

5.5.1 Compression and De-compression (codecs)

Compression: is the reduction of the number of bits which allows for faster transmission and reduced storage space. There are two forms of compression; lossless (the original data is de-compressible and all the data is recovered) and lossy (once the data is reduced, decompression will not return all the original data). Compression techniques that work well on general data files (WinZip, 7Zip etc.) are not really suited to audio and video signals, but can compress some still image formats.

The advertisement features a night photograph of the Apollo Hotel building. A large red circular logo with a white lightbulb icon is on the left. The text "CISO Conference" is written in white, with "Produced by Inspired" below it. To the right, a white box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL Dec 5th 2019". At the bottom, a dark banner reads "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". On the right side, there is a logo for "Inspired" featuring a blue lightbulb icon.

There is a need to reduce file size for audio visual transmission, as there are restrictions on bandwidth and read/write speed considerations that make the use of 'raw'(uncompressed) footage undesirable. It is possible to reduce file size by changing the pixel count by transcoding to a smaller frame size (from HD to SD), frame rate reduction (from 50/60fps to 25/30fps), dropping audio from 48Khz to 44.1 or 32Khz), but most users would typically use a compression technique that is built into a change of file format e.g. from an 'uncompressed' AVI file to .mp4.

The term 'codec' is given to the compression/decompression software that is applied within the file format. Eg. H.264 is the compression found in a .MOV (High Definition Quicktime) file (amongst others). The key points to know about compression are what it does to your file, and to ensure the recipient of your file has the correct de-compression (codec) to play it. Proprietary codecs and formats are often the cause of problems for students trying to copy one file from one software package to another.

Compression of various sorts has been around since the start of television.

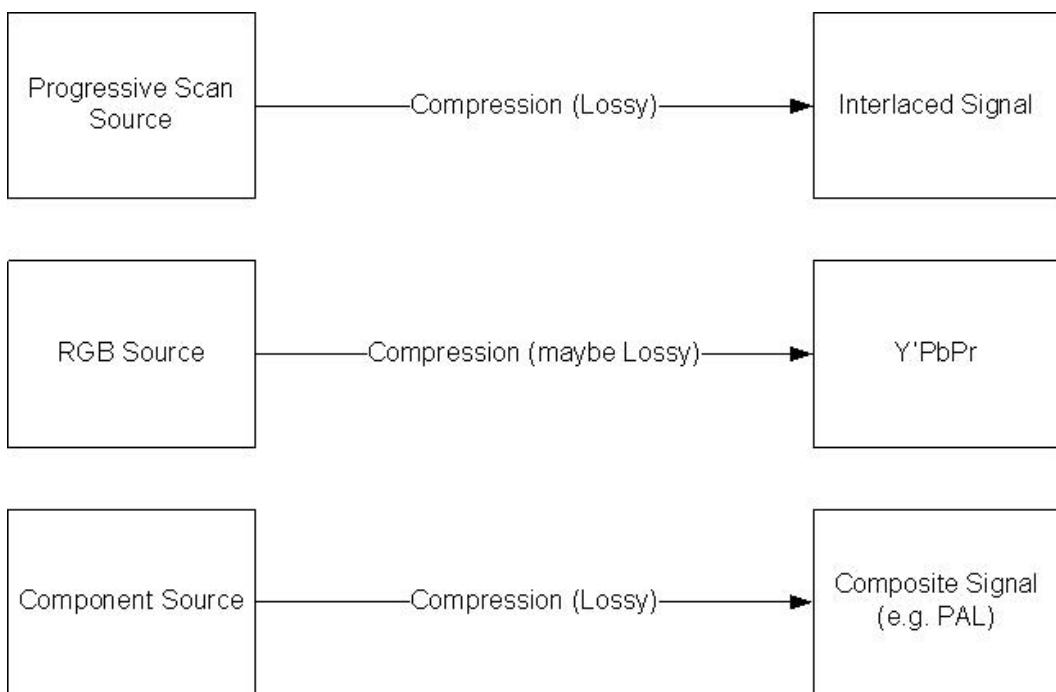


Figure 5-10: Compression

5.5.2 Lossless Compression

It is possible, using mathematical techniques (the full details of which we need not consider here) to reduce the quantity of data to be stored or transmitted. The basis of such reduction rests with the fact that there is often redundancy built into the information that is transmitted. For example, in the English language, we know that, normally, the letter u will follow the letter q. Therefore, if we wished to reduce the number of letters being transmitted in a message, we could omit the data for u following q and have the receiver re-insert it automatically so that the Queen is not reproduced as Qeen, for example. Another example of such a system would be the Zip program for reducing the size of computer files. When ‘unzipped’, the files must be exactly as they were before compression, otherwise they will not work and data will have been lost.

For media files, lossless compression techniques are used in image format (GIF and PNG) and in audio (FLAC, Apple Lossless). For video there are some lossless compressions (H.264 Lossless, VP9, Jpeg2000 Lossless).

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

5.5.3 Lossy Compression

In video terms, typically the amount of data that needs to be removed in order to provide an effective transmission system means that we have to use a lossy system. (A lossless system can only provide about 2:1 or 3:1 maximum.) One such system that is now in widespread use is MPEG, with MPEG-2 being the normal system used for TV transmission at standard definition. MPEG-2 is, in fact a multi-level system in which varying degrees of compression can be applied and it can also be found in camcorders and DVD players.

The MPEG system, used in digital terrestrial, satellite or cable transmission of domestic TV services, is ‘asymmetrical’ in that the coder (of which there only needs to be one at the transmission end) is ‘smart’ but expensive. It varies its actions dependent on the incoming signal. However, the decoder in a viewer’s set-top box or TV is ‘dumb’ but cheap. It always does exactly what it is told to do by the incoming bitstream.

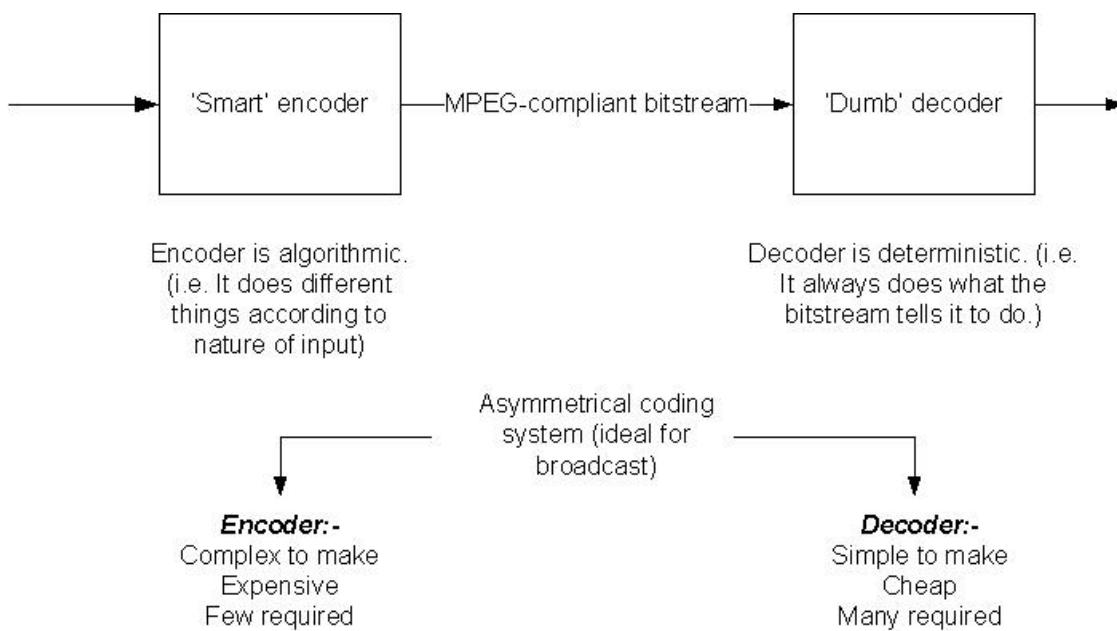


Figure 5-11: Encoding

Providing everything works ok, the coder-decoder pair (codec) results in an image that, while lossy, has jettisoned data in such a way that the viewer does not notice. In order to do this, the developers of MPEG had to understand how the human eye-brain system responds to a variety of stimuli and designed the compression system accordingly, enabling just those elements that “don’t matter” to be discarded. This is known as ‘perceptual encoding’ (or sometimes ‘perceptive encoding’). In a sense, one could describe the compression system as relying on both redundancy and irrelevancy. The lossless coding element relies on the ability to remove data that is not needed (redundant) since it can easily be re-inserted at the receiver, while the lossy coding element relies on the ability to throw away parts of the image that are irrelevant to our perception of the recovered signal.

It must be realised, however, that since the data emerging from the receiver are not identical to those entering the transmitter, any cascading (or ‘concatenation’) of such encode/decode processes (for example, trying to uncompress then edit a heavily compressed MPEG file, then recompress it for storage or transmission) results in unacceptable degradation of the signal.

5.5.4 Inter-Frame vs Intra-Frame Compression

Video compression systems fall into two main classes – *intra*-frame and *inter*-frame. What do these terms mean? First, we need to define that *intra* means *within* whereas *inter* means *between*. Thus intra-frame means within (or inside) one frame whereas inter-frame means between different frames.

5.5.4.1 Intra-frame

In pure intra-frame compression, each frame (picture) is coded independently of all the other frames. This makes editing easy, as individual frames (or sequences of frames) can be removed or inserted in the video editing process without affecting other frames. An example is DV coding which relies on spatial redundancy (i.e. it removes unnecessary data located within the picture) and spatially-based perceptual coding (i.e. data which you do not notice is not in the compressed picture, afterwards). Intra-frame compression is the sort of compression you can apply to a still photograph. (There are no other frames in a still photo!) When used in video, pure intra-frame compression is sometimes referred to as motion-JPEG (or MJPEG), though other types can exist. It is a mild compression, sometimes known just as ‘spatial compression’ and often used in high quality VTRs, camcorders etc.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

Intra-frame compression codes a picture in the same way as a JPEG picture is created. The picture is divided into blocks of 8 by 8 pixels. Then instead of sending the values of the actual luma or chroma samples (pixels) in the blocks, they are “transformed” by a process known as Discrete Cosine Transformation. The DCT process looks at each 8x8 block and produces another 8x8 block of values. However, these values actually relate to how much content there is in the block at certain frequencies.

Thus, the DCT in effect ‘looks like’ this:

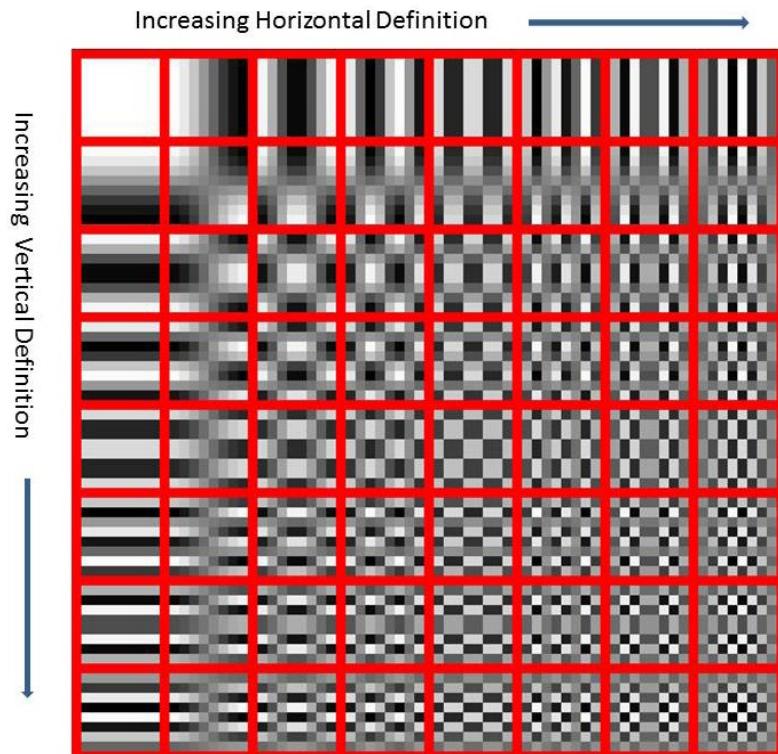


Figure 5-12: Discrete Cosine Transform

This picture shows a set of 8 by 8 representations of horizontal and vertical frequency. The DCT produces a grid of 8 by 8 numbers, whose values indicate the contribution to the original picture block made by each of the frequency components in this picture. The lowest frequency (constant dc or zero frequency) is in the top-left corner and represents the average level of luma in the block (or chroma, if the block represents a block of chroma samples). Horizontal frequency (i.e. dark-light cycles per picture width found in the original picture block) increase from left to right while vertical frequency (i.e. dark-light cycles per picture height found in the original picture block) increase from top to bottom.

Most of the things at which you point a camera are fairly smooth, however; the exceptions are edges and texture. This means that most of the significant numbers produced by the DCT will arise predominantly from components in the top left of the transform. (They tell us that the dc component plus low frequency signals (i.e. those that do not change very rapidly across (or down) the block of samples) are the dominant elements making up the picture signal.)

Secondly, the eye is less sensitive to slight changes of value in areas where there is already a lot of detail (where the pixel values change frequently), whereas it can easily notice a change within a large area of uniform tone. Therefore, the numbers corresponding to the high frequency detail do not require the same accuracy as the numbers corresponding to low frequency detail. When we round (or ‘quantise’) these DCT coefficients, many (or most) of the higher-frequency coefficients are already so small that they will become zero.

The combination of these two effects is that for many blocks in the picture, only a few low frequency coefficients need be used, thereby contributing to a reduction in the bit rate since we can avoid transmitting all the zeroes one after another.

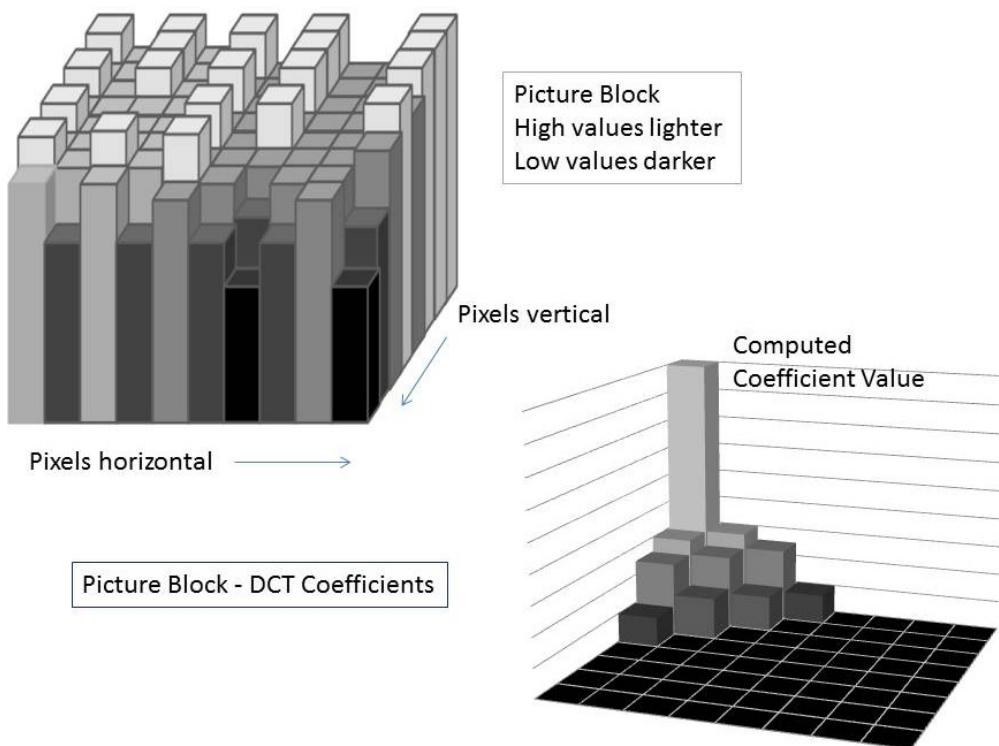


Figure 5-13: Picture blocks

By sending the coefficients in an order which favours the top left first, it is possible to send the significant coefficients then a simple code which says how many zeroes would follow, thus reducing the total data to be transmitted.

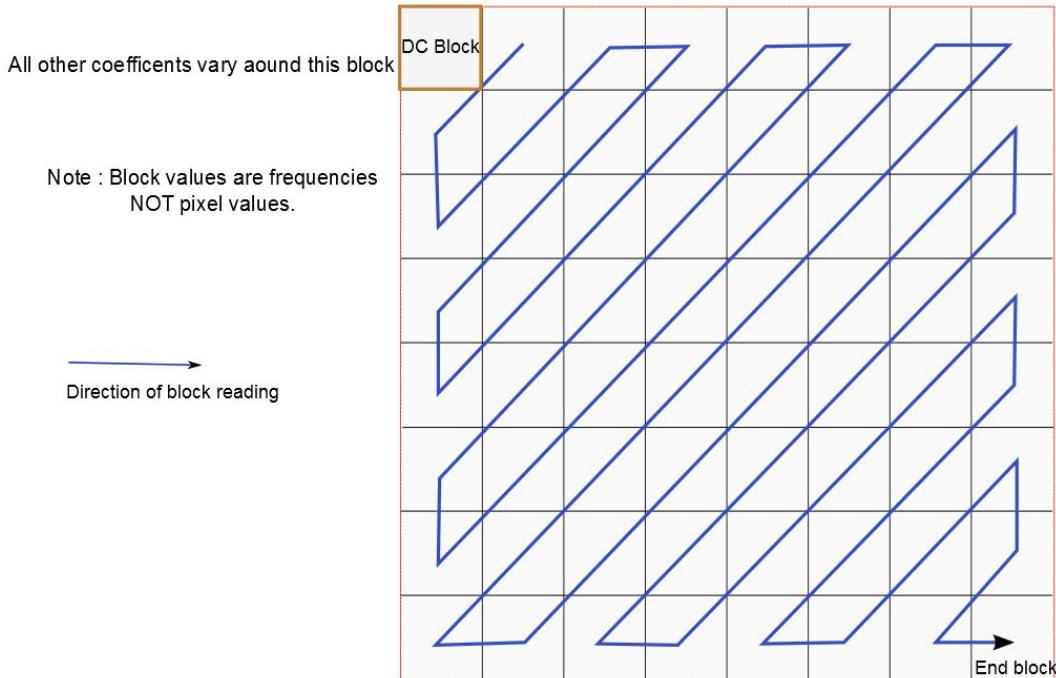


Figure 5-14: DCT Read order

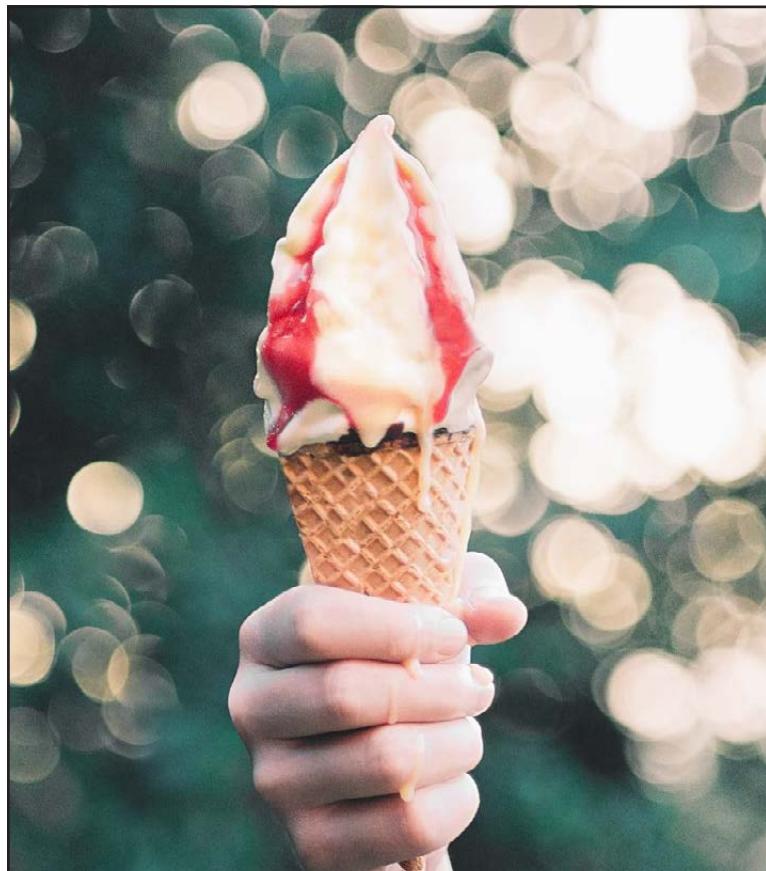
At the receiver the DCT values arrive and an inverse DCT operation is performed to recreate the original picture block. The process is repeated for all the blocks in a frame, saved in temporary memory. Once all the blocks are processed and re-arranged into the correct order, the memory contains a complete frame of data for a viewable picture and can be played out before the process repeats for the next frame. Note that some picture degradation is unavoidable as some inaccuracies were introduced when rounding (quantising) the values in the DCT coefficients. The trick is to make these invisible to the human eye. This is why fast shutter speeds can be a problem with moving objects – they result in fine-detail flicker which might need many of the high-frequency coefficients to be non-zero and, hence, not easily compressed, resulting in unpleasant artefacts becoming visible.

5.5.4.2 Inter-frame

Inter-frame compression exploits the high degree of correlation (similarity) between adjacent frames in any video sequence. (If there were no correlation, TV would appear as a very rapid succession of random pictures, which would contain no useful purpose and which the human perceptual system could not absorb). The clever thing about such systems is that planar motion (translation) can be identified and ‘motion vectors’ (data which identifies the movement from one frame to the next) can be created which allows us to reduce the amount of data which it is necessary to transmit.

In other words, we can take an intra-frame compressed image (i.e. a frame which contains enough information to create a useable picture, on its own) and then send a small amount of data which describes to the decoder how the next frame will differ from the current one. The decoder can then reconstruct the new frame and display it, without needing the full data. This is useful because the eye ‘latches on’ to an object moving from side to side, or up and down so the ability to accurately portray such motion without having to send the full data is a good way to reduce the total data that has to be sent without compromising the image quality too much. These types of coder are far more complex, but a lower bit rate is achievable for similar quality. Since (by definition) the inter-frame coding must take place using more than one frame and since the frames occur over a period of time, this type of compression is sometimes referred to as ‘temporal’ (i.e. time-based) compression.

A common compression system used in standard definition TV and which uses a combination of spatial (intra-frame) and temporal (inter-frame) coding is MPEG-2. Typically DV (intra-frame only) uses 25 Mbits/sec which is suitable for origination of news-gathering (ENG) ‘prosumer’ type cameras and editors, while MPEG 2 (intra- plus inter-frame) gives adequate quality for final transmission to the viewer at home using around 3 to 5 Mbits/sec. In Europe, DV compression is mostly applied to 4:2:0 coding of the video signal for camera recording, while MPEG-2 compression is applied to 4:2:0 coded video for final transmission of the signal to the viewer or the creation of DVDs.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

5.5.5 Mpeg-2

MPEG 2 is a more flexible system but is somewhat more complex than spatial compression alone. It can be used for standard definition signals from countries that previously used NTSC, PAL or SECAM and can also cope with high definition signals, though more efficient coding systems tend to be used for HD. (There isn't a separate MPEG-3 for HD, although there was initially, until it was realised that MPEG 2 could be made flexible enough).

Some frames are transmitted as 'stand-alone' frames. These are the spatially-compressed intra-frames (or I-frames) described above and typically might comprise one frame in 12 or 15. They need a reasonable number of bits but do mean that a complete and useable frame arrives approximately every half second and these are used as the basis to build the frames in between.

At the encoder, the current frame and the next frame are compared and another image can be created which contains data showing just the differences between the two frames.

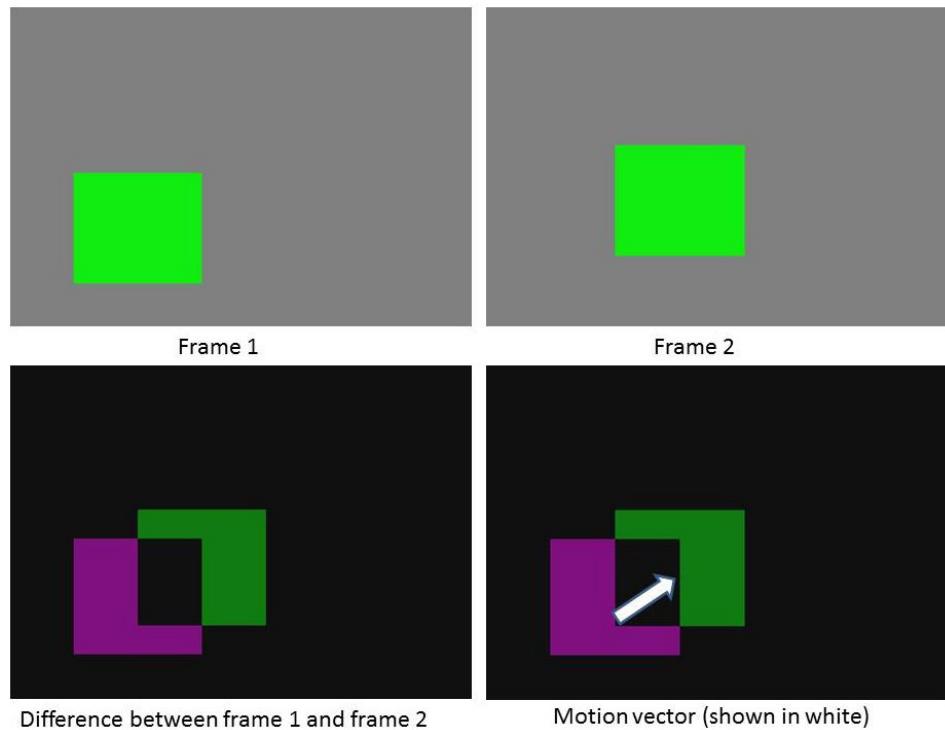


Figure 5-15: Motion vector

Between the I-frames are the remaining frames, which are created by predicting their contents, based on the information derived from the I-frames. In general, most images will have very little change from frame to frame (excluding scene changes, of course).

In the example, above, a green box moves in a north-easterly direction from frame 1 to frame 2. There is very little difference between the two frames, with most of the image remaining the same. The only difference between the two frames is shown in the bottom left hand frame. If this frame is ‘added’ to frame 1, the green part will replace what was plain grey while the magenta part will cancel out what was green, leaving plain grey there. All other pixels are unchanged (shown here as black). Thus it can be seen that only a small amount of new data in this (admittedly very simple) example needs to be transmitted, saving a substantial amount of data compared with transmitting a complete frame.

In fact, it's even more efficient than this. The encoder contains a ‘motion estimator’ which identifies and measures the distance and direction of motion of elements of the picture from frame to frame using ‘macro-blocks’ (which are just the blocks parcelled up to give a unit of 16x16 pixels). It then transmits very small amounts of data as ‘motion vectors’ (as shown in white on the bottom right hand diagram). These are then used by the decoder to predict what will be where in the next frame. Of course, a moving object might rotate, become shaded or better lit, or move behind other objects, from frame to frame, so there will still be some error. However, rather cunningly, the encoder then decodes its own output to see how accurate it was. It then compares this predicted next frame with the ‘real’ next frame and produces a difference image. If the motion estimator was good, the difference image will be virtually empty (i.e. there are no ‘residual errors’). Transmitting the residual error image to the receiver helps the receiver to correct the errors. By spatially-compressing this residual error image, even more data can be saved.

The advertisement features a night photograph of the Apollo Hotel building. A large red circular logo with a stylized lightbulb icon is positioned on the left. To its right, the text "CISO Conference" is written in white, with "Produced by Inspired" underneath. On the right side of the image, a white rectangular box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and "Dec 5th 2019". Below the main image, a white banner contains the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right of this banner is the "Inspired" logo, which includes a blue lightbulb icon.

Now, we have an encoder which can send an I-frame (spatially compressed) plus a combination of motion vectors and a residual error image (spatially compressed) which, together, enable a frame (that hasn't even been transmitted) to be predicted and, thus, recreated by the decoder! This is called a predicted frame or P-frame. (Of course, if we carried on trying to do this, the errors would build up over a number of frames. However, the insertion of I-frames 'resets' the process by giving the decoder a fresh image to start with, typically every 12-15 or so frames.)

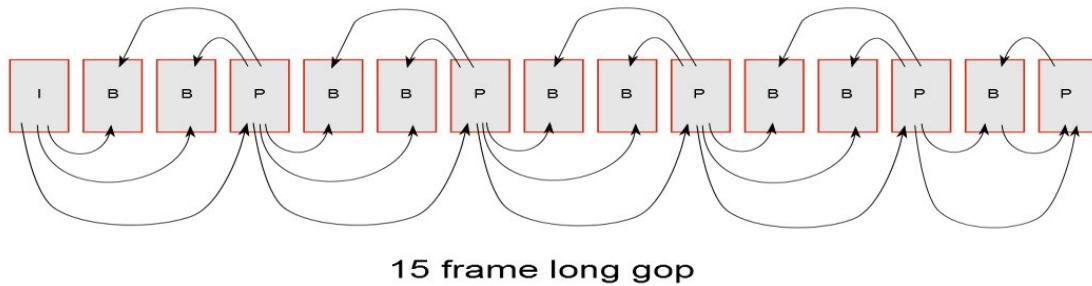
The process becomes even more ingenious. The prediction can be improved (and, hence the image quality improved and the errors reduced, saving even more transmission or storage, since the residual error image data needing to be sent will be smaller) through the use of bi-directional prediction. Frames that are created from information derived from both sides of them are known as B-frames.

However, how can we achieve all this since it relies on being able to predict a B-frame 'in the future' based on a P-frame that is even further 'in the future'? The simple answer is that if we introduce a delay in the encoder during which we temporarily store all the incoming frames, we can work on those so 'the future' in terms of output from the encoder is dealing with frames that have already been captured in the past. This is why all digital TV and radio services have a delay compared to analogue, which can be a genuinely live service. Digital is always delayed while the encoding and decoding processes take place.

5.5.6 Group of Pictures (GOP)

I-frames, P-frames and B-frames need to be grouped into a sequenced structure. GOP structure is defined using the distance between I or P frames, and the distance between I frames (complete image). The GOP pattern must have an I-frame in it, but the ratio of the B and P frames determines the GOP length. Short GOP is a pattern of I-frame only, whereas Long GOP refers to a pattern that has many B-P frames to one I-frame eg. MPEG-2 is 15 frames (PAL) 18 frames (NTSC). GOP applies to HD formats as well, 1080 HDV uses a long GOP of 15 frames, yet 720 HDV uses 6 frames. A 'closed' GOP starts with an I-frame for a set number of frames in a repeating pattern always ending in a P-frame (so may repeat a P-frame out of pattern), while an 'open' GOP can start with an I-frame or a P-frame.

Although it is possible to have almost any combination of I,P and B-frames, a reconstructed group of pictures (GOP) at the decoder will typically look something like this:



I = INTRA-CODED FRAME B = BI-DIRECTIONALLY PREDICTED FRAME P = PREDICTED FRAME

Figure 5-16: Group of pictures (long GOP)

A moment's thought will bring the realisation that a scene change can cause havoc with this process. A scene change that occurs where a B or P frame is located would result in a huge residual error image. In practice, a good quality encoder might choose to send a fresh I-frame if it has sufficient capacity to do so. This is why files that have already been heavily MPEG-compressed (e.g. DVD) cannot be adequately re-edited. The only suitable edit points would be I-frames and those are only about $\frac{1}{2}$ second apart. (This is also why there is always at least a $\frac{1}{2}$ -second delay when you try to change channel on a cable or satellite service.)

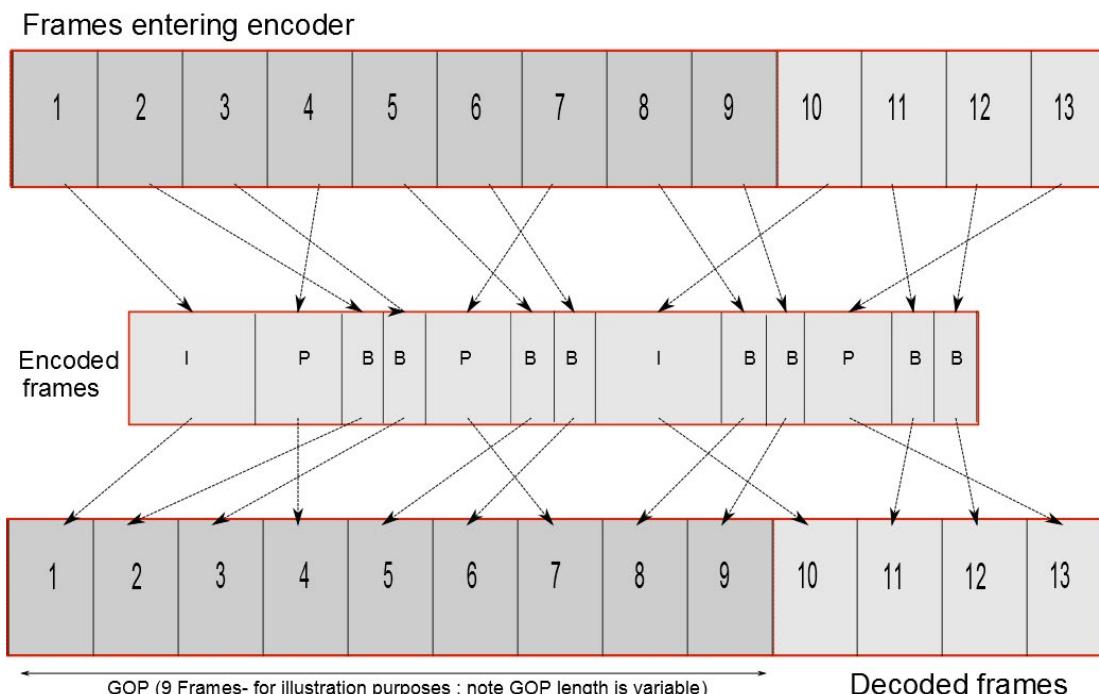


Figure 5-17: GOP encoding and decoding

To achieve bi-directional coding, it is necessary to re-order the frames in the encoder. A complementary reordering must take place in the decoder. However, even a scene change on a regular I-frame can still cause a problem. Imagine the last two B-frames in the GOP shown above. They are derived from both the preceding P-frame and what should be the succeeding I-frame. If that I-frame is a scene change then these B-frames will be of poor quality since they are derived from the preceding P-frame (scene A) and from the I-frame (scene B – totally unrelated). A good encoder will choose to turn those last two B-frames into P-frames instead which will not be derived from the following I-frame. However, the encoder may have trouble deciding whether to bother doing this. (Is this a scene change or merely a poorly-estimated motion? There comes a point where it's better to replace with a P-frame but how to decide?) Many editing and encoding packages for DVD creation (for example) enable the person doing the editing to mark up the files with special markers that 'tell' the encoding software when there is a scene change so that this process can be carried out reliably.

It is still not without difficulty, for it now means that as well as the forthcoming burst of data for the I-frame, it now has to encode more intensive P-frames instead of B-frames. However, provided the scene changes are not too frequent, the buffer used to assemble the data stream should be able to cope. If it can't, then the encoder goes into 'fallback' mode and the macroblocks become visible on screen as a mosaic-tiled effect. This is why you should avoid doing very rapid scene changes in succession – it may look wonderful on the editor but there's no guarantee that it won't look dreadful on the viewer's TV! This mosaic problem is also sometimes noticeable on dissolves, as the values of every single pixel are changing from field to field in a non-predictable way, making P-frames and B-frames difficult to create.

Provided the encode-decode process is working normally, the instantaneous data rate will look something like this:

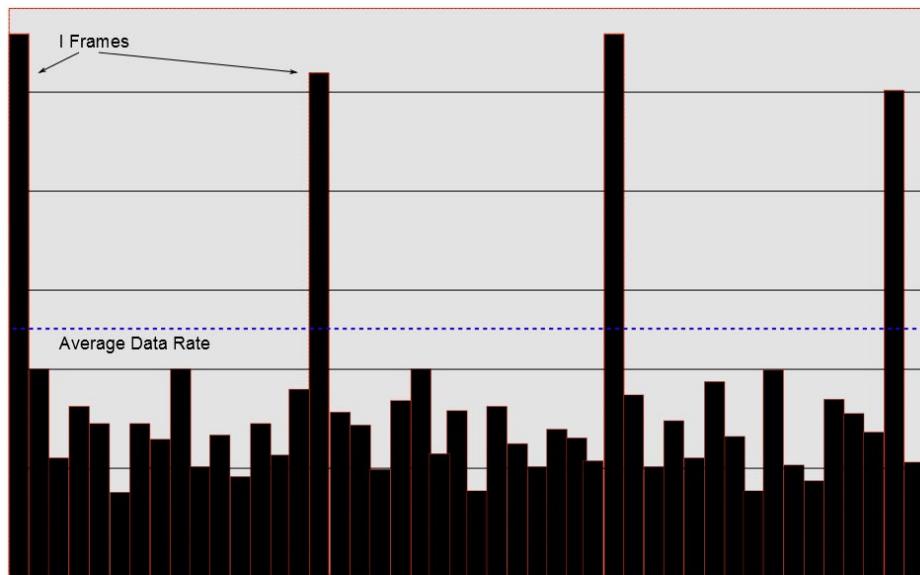


Figure 5-18: Encoding data rate

The I-frames can clearly be seen. The *average* data rate can be kept relatively low, however, since the additional data needed for the I-frames is spread over time and loaded into the buffer ready for use when required.

MPEG 2 is a storage and transmission format. It is generally assumed that the signal is just viewed or stored for transmission, i.e. no further processing will take place.

5.6 Profiles and levels

A profile refers to the set of algorithms used for coding and decoding. There is a main profile which uses 4:2:0 coding. A simple profile has no B-frames. An interesting profile is the 4:2:2 profile which (as the name suggests) has higher colour resolution. This might be appropriate when you want to do something like downstream colour keying.

A level refers to the standards of the input video. For example main level means 625 line TV (standard definition). When you get a main level MPEG decoder in your set top box, it makes no claims to decode HD.

Note: Digital TV uses MPEG and DVB (digital video broadcasting) standards for transmission in Europe and elsewhere. ATSC is an American system used in place of DVB.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com

Some examples of profiles and levels are shown below:

<i>Abbr.</i>	<i>Name</i>	<i>Picture Coding Types</i>	<i>Chroma Format</i>
SP	Simple profile	I, P	4:2:0
MP	Main profile	I, P, B	4:2:0
422P	4:2:2 profile	I, P, B	4:2:2

<i>Abbr.</i>	<i>Name</i>	<i>Frame rates (Hz)</i>	<i>Max horizontal resolution</i>	<i>Max vertical resolution</i>
LL	Low Level	24, 25, 30	352	288
ML	Main Level	24, 25, 30	720	576
H-14	High 1440	24, 25, 30, 50, 60	1440	1152
HL	High Level	24, 25, 30, 50, 60	1920	1152

5.7 MPEG -4

This is a more advanced compression system. It was designed to allow the information on your TV screen to arrive in separate streams (TV, Internet and so on) and to allow different streams to be handled separately (so you could switch out the weather forecaster to allow you to concentrate on the map).

Generally however, its main property is that it is a more efficient way of compressing video signals. So, compared to MPEG-2, you get better quality for the same overall bit-rate, or a lower bit-rate for the same quality. MPEG-4 is used for high definition transmissions on Satellite and Freeview.

The full title of the format is H.264/MPEG-4 AVC. Sony used XDCAM onto disc recording media, and Panasonic brought out the P2 format onto solid state (flash memory card) media.

AVC encoding has also been used for recording onto HD DVDs (more precisely, Blu-ray). Hard discs have also been used. Editing systems using AVCHD require powerful processing (as the MPEG-4 compression system is complex, as is the decoder).

AVCHD will often produce a bit rate of around 25 Mb/s, but there is a reduced figure of 18 Mb/s for Blue Ray discs and final transmission. The standard acceptable for broadcast acquisition is usually 50Mbps.

H264 Advanced Video Coding is based on the same basic principles as MPEG-2 but takes them much further so as to be able to achieve a much greater compression without significant loss of perceptual quality (most of the time!). However, as with MPEG-2, it is severely tested when detail changes from frame to frame such as with short shutter speeds, ‘wobbly’ camera shots, film weave, panning on detailed surfaces (e.g. grass, trees etc), random frame changes (e.g. flash guns) and random noise (e.g. film grain).

However, the results are generally much better than for MPEG-2 allowing HD signals to be coded within a moderate and useful bandwidth.

This is achieved through a combination of:

1. Use of a range of block sizes including smaller blocks (allows more accurate motion prediction, hence lower residual errors to transmit)
2. Use of spatial prediction in I-frames (block is predicted based on surrounding blocks) rather than just temporal prediction between B- and P-frames
3. Use of smaller 4x4 DCT-like transform (no longer need 8x8 since many low-freq co-efficients are handled locally at the receiver by the spatial prediction)
4. Use of a de-blocking (smoothing) filter at block transitions to reduce visible mosaic tiling
5. Use of more than one reference picture to make predicted frames (if a non-typical frame is found in sequence, e.g. flashgun or dissolve)
6. Use of different sizes and orientations of macroblock (useful at non-horizontal or non-vertical edges of a moving object)

Although these enhancements have greatly improved the possibilities for HD signal recording and transmission, they are not infallible and good image quality depends more than ever on good stable camerawork with sensible use of settings!

Again, editing requires the reconstruction of frames needing a fast computer, with the ensuing encode/decode/encode process and resultant picture degradation per edit cycle. The additional complexity of encoding and decoding the more complex signal is reflected in the price of cameras and editing equipment but is continually falling, with the prosumer market now starting to enjoy the benefits.

6 Audio Sampling and Compression

In this chapter, we cover audio sampling and compression

Learning Outcomes

- To explain how sound is sampled for recording purposes.
- To give an outline of methods typically found for audio compression.

The principle involved in converting analogue audio signals into digital audio signals are similar to those of digital video signals, though dealing with a simpler input.

6.1 Pulse Code Modulation (PCM)

An analogue signal (for example, the voltage waveform representing sound pressure as presented to a microphone) could be plotted against time on the x axis with the voltage plotted on the y axis. At regular intervals, we ‘take a snapshot’ of the amplitude (size) of the waveform. The amplitude of the impulses in the diagram on the right represents the amplitude of the original wave on the left hand side. So far, this process has resulted in ‘pulse amplitude modulation’ (i.e. the amplitude of the pulses has been altered (or modulated) by the audio wave being sampled). As soon as we quantise the signal in the right hand diagram (i.e. round it to discrete whole values in the y axis) and assign numbers representing the heights of the quantised pulses, we have ‘pulse code modulation’ or PCM.

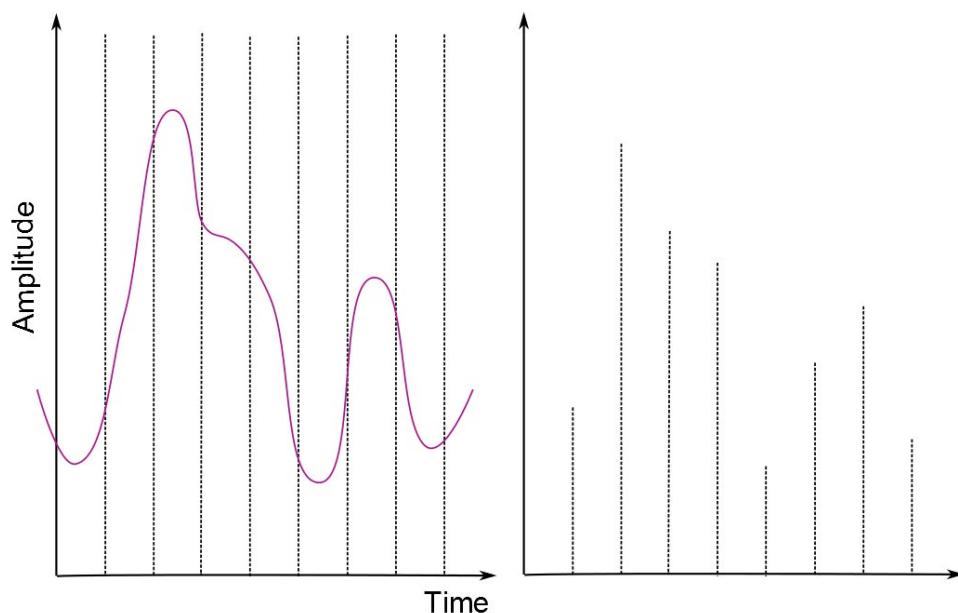


Figure 6-1: PCM

PCM is the ‘basic building block’ of digital audio, created through a process of ‘analogue to digital conversion’ (A to D), before any processing has been carried out. WAV or AIFF files, for example, are basically ‘parcelled up’ files full of PCM values and represent the highest quality (‘unmolested’) version of a digital audio file. Once we have a WAV file, it consists of lots of numbers representing the amplitudes of regular pulses which we can then manipulate in any way we like to create a new WAV file. However, when we look at what all this means in terms of the frequencies of the sounds involved, we can see some strange things going on.

6.2 Frequency

If we were to plot the frequency (Hz) along the x-axis (instead of time), we would get a spectrograph or ‘frequency response’ of the signal which would show how much energy there is for each frequency. For an audio signal, we might expect to see energy up to 20KHz, which we might refer to as the Bandwidth of the signal, with the highest frequency being denoted as B Hz.



The image shows a woman with long blonde hair wearing a black VR headset, looking upwards with a surprised or excited expression. To her right is a red advertisement for MT Højgaard. The ad features the company logo (a stylized 'M' and 'H') and the text 'MT Højgaard'. Below this, in large white letters, is 'BEDRE LØSNINGER'. A block of Danish text follows: 'I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?' At the bottom of the ad is the website 'mth.dk/vorestilgang'. A green button at the bottom right of the ad contains a white hand cursor icon pointing to the text 'Click on the ad to read more'.

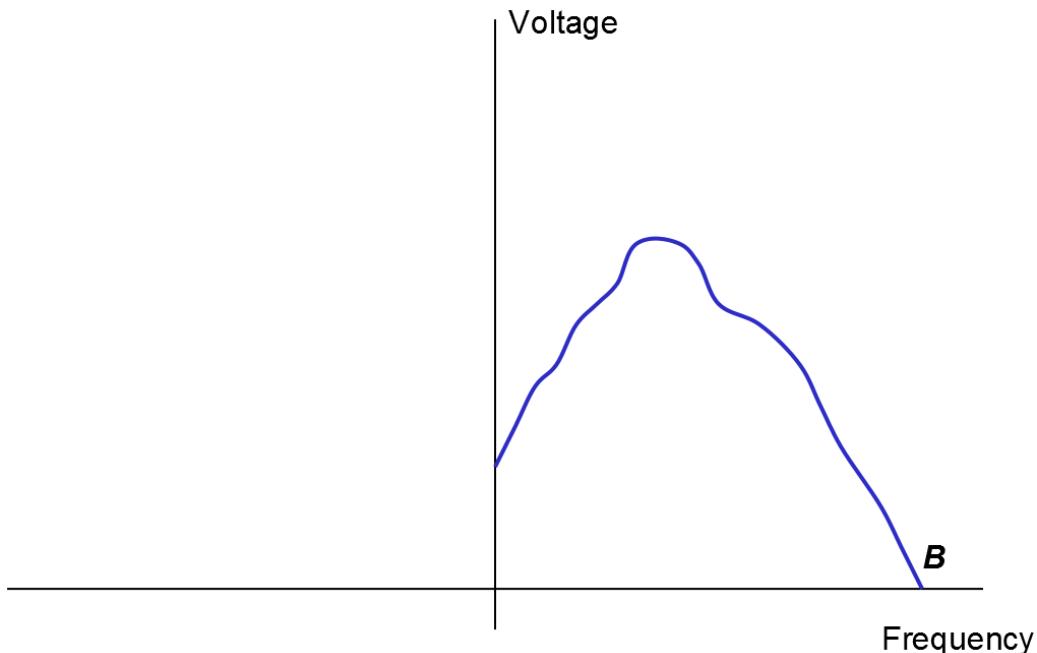


Figure 6-2: Frequency Plot

Strangely enough, although ‘in reality’ we would not normally consider it possible to have a ‘negative frequency’, for the purpose of sampling it is necessary to imagine that the real spectrum is mirrored about the 0Hz (DC) vertical axis. Consequently, the spectrum can be shown thus, with the ‘imaginary’ negative frequencies being shown dotted in the following diagram for clarity only. (Note that the graphs above and, hence, below can be any shape depending on the signal but what is shown might be a typical response.)

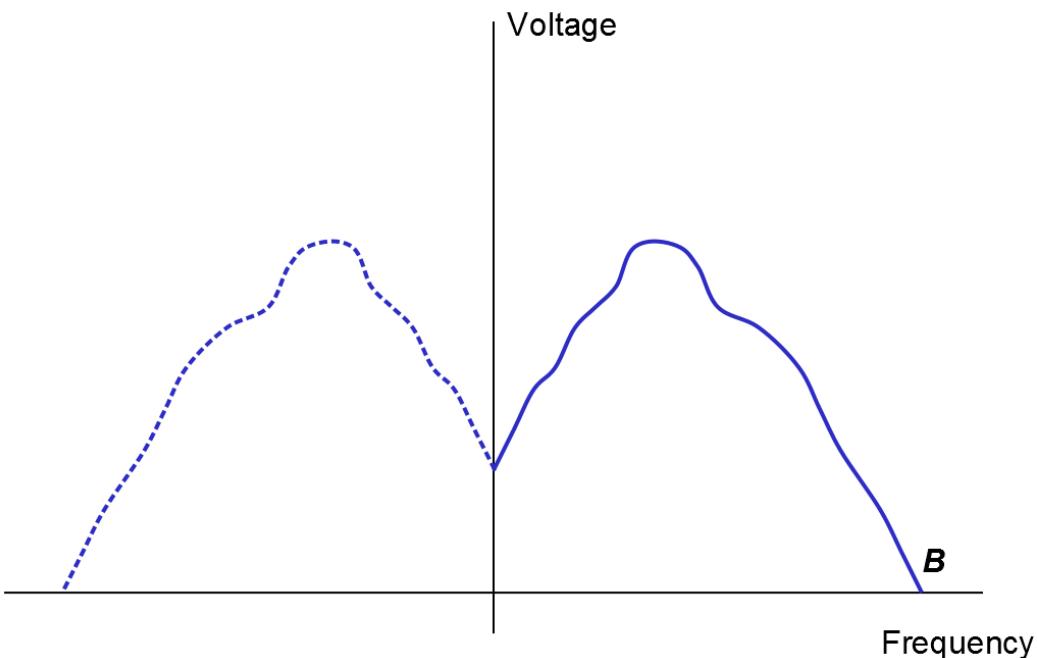


Figure 6-3: Mirrored Frequency

First let us consider the process of sampling again. Let us imagine that we take a regular sample of our input signal every T_s seconds. The frequency at which the sampling takes place will be $1/T_s = F_s$. For example, if we were to sample data every $1/10^{\text{th}}$ of a second (T_s), then we would be sampling 10 times a second so $F_s=10\text{Hz}$. Since audio is normally considered to start from 20Hz (the lowest frequency in the 20 – 20,000Hz range where the human ear can perceive sound), this would clearly not be often enough so let's now imagine that we're sampling with $F_s=10\text{KHz}$ (i.e. 10,000 times a second). If we were sampling audio on a telephone line which responds up to about $3\frac{1}{2}\text{KHz}$ (so we assume that our value for B is now this value) then we would have a 'real' signal that we *might expect* to look like this:

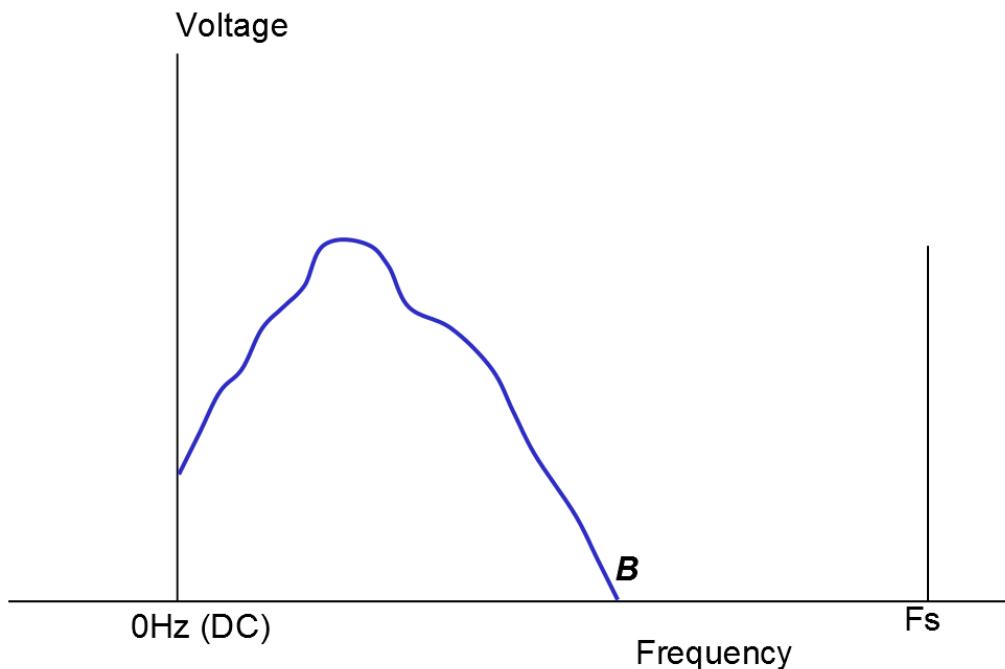


Figure 6-4: Telephone Audio Sample

However, this would *not* be the case. If only a single frequency was present at F_s then this would be a pure sine wave of frequency F_s , not a train of narrow pulses, which is what we want. In fact, a train of narrow pulses would appear along the frequency axis as a series repeating at F_s , $2F_s$, $3F_s$ etc. up to infinity. Mathematically, we know that an imaginary 'image' of this repeating pattern would also exist in the negative area to the left of 0Hz, at $-F_s$, $-2F_s$, $-3F_s$ etc. (Again, don't worry about why this happens – just accept that it does, as a weird mathematical side-effect.)

Furthermore, the spectrum of frequencies in the wanted 'baseband' signal (and its 'negative' image frequency) are also replicated about each of the impulse frequencies so the actual frequency spectrum would look like this:

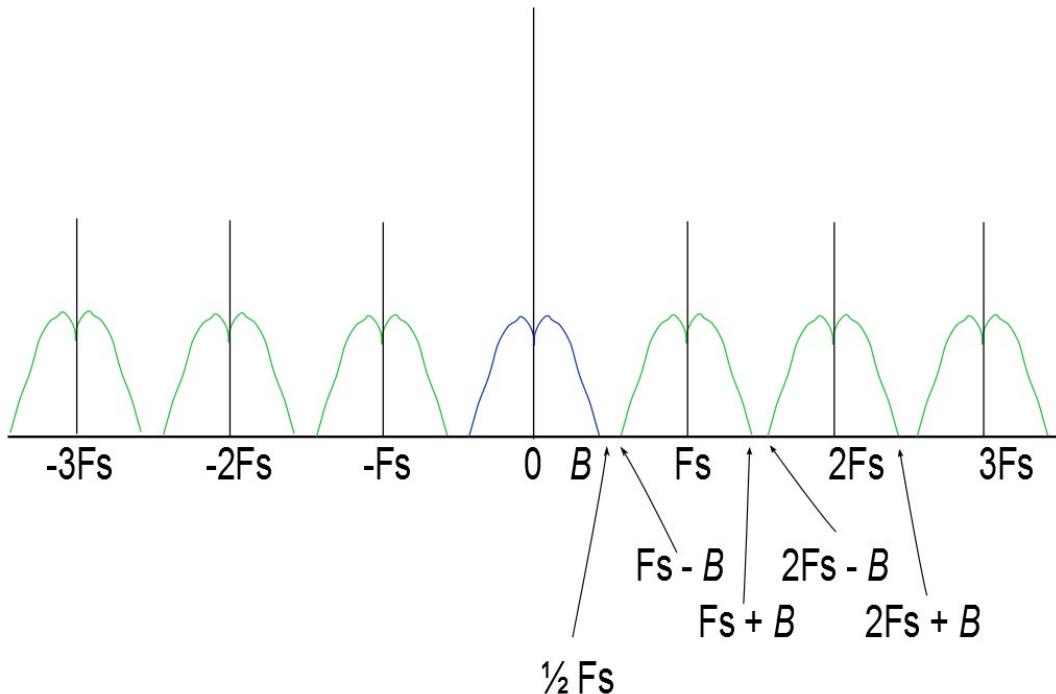


Figure 6-5: Actual Frequency Spectrum



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

It would carry on up to infinity (and backward to minus infinity). In practice, the part that we are interested in is the part from 0Hz (dc) up to bandwidth B . What this graph shows is that we also get unwanted ‘image’ frequencies from $F_s - B$ to $F_s + B$ and $2F_s - B$ to $2F_s + B$ etc. At first sight, this doesn’t appear to be a problem. If we want to play back our sampled data, we can see that there is a gap between B and $F_s - B$, centred around $\frac{1}{2}F_s$. If we filter out all the frequencies that we don’t want, letting through just those frequencies between 0 and B Hz then this would leave us with just the frequencies we want. We can do this with a ‘low pass filter’ on the output of our D to A converter (i.e. one that passes just the low frequencies that are below $\frac{1}{2}F_s$ and rejects all those above. This is represented with the magenta line in this diagram:

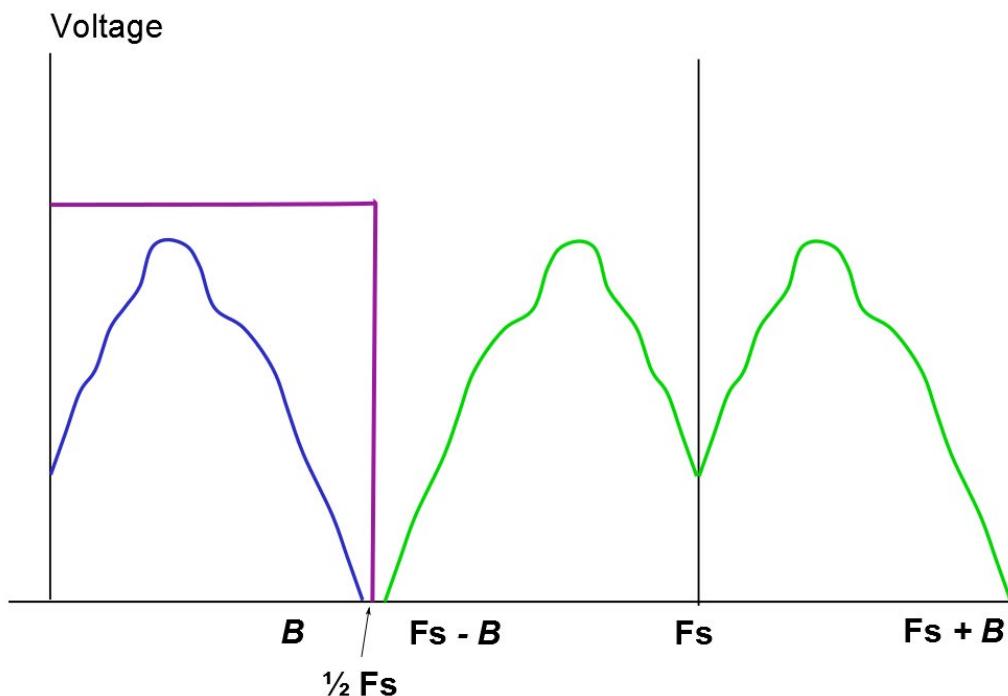


Figure 6-6: Low Pass Filter

Thus, we will no longer hear F_s (10KHz in the example above) or any of the image frequencies which extend down to $F_s - B$ when we convert our digitally sampled signal back to an analogue one for reproduction using an amplifier and loudspeaker.

So far, so good but what happens if we try to use our 10KHz sampling on a signal that has a bigger bandwidth, extending up to, perhaps, 7Khz? This would result in B being greater than $\frac{1}{2}F_s$ so the lower part of the image frequencies (green) extending down to $F_s - B$ would now be overlapping with the baseband signal frequency up to B (blue) that we want to hear. Thus, we would be able to hear all sorts of strange tones, as we would no longer be able to rely on a low pass filter since there would be no separation between the wanted signal (blue) and its image frequency (green) where the cut off limit of the filter could be positioned.

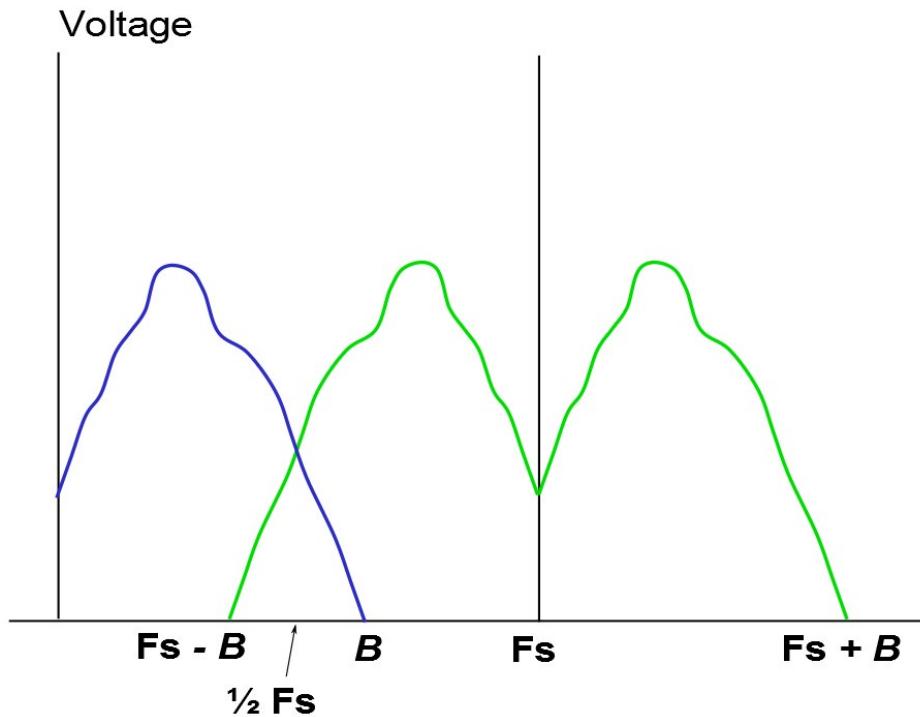


Figure 6-7: Alias Components

This process generates ‘alias components’. Thus, the only frequencies that could be recovered without aliasing would be those up to $F_s - B$, so we would have to low-pass filter the output of the D-A converter up to that frequency limit and lose the frequencies higher than that, significantly reducing the quality of our reproduced baseband signal.

For example, suppose there was a significant 6 kHz signal within the audio signal, which was then sampled at 8 kHz (which would give slightly better than telephone quality). Then the alias component would appear at (8–6) kHz, or 2 kHz. This would appear as an extraneous whistle unless we filtered the output so severely that only frequencies up to 2KHz were let through. Doing so would impair the quality so badly that it would be difficult to understand what was being said (rather like having the bass boosted to max and the treble cut to min on a hi-fi, only much more so).

There are a couple of ways to prevent this problem from happening. If we had placed a low pass filter at the input while recording, which allowed only the signals having frequencies up to $\frac{1}{2}F_s$ to enter the system, then we could at least reproduce the signal up to a frequency of $\frac{1}{2}F_s$ (which, as can be seen in the diagrams above, is fairly obviously higher than $F_s - B$ in this example). Limiting the bandwidth of the input signal like this would give reproduction where the lower end of the ‘green’ signal (at $F_s - B$) would just meet the upper limit of the ‘blue’ baseband signal at B . So long as they don’t overlap, we can successfully low-pass filter the output of the D to A on playback to let just the ‘blue’ signal through.

A better way would be to choose a sampling frequency (F_s) which is higher than twice the highest frequency of the baseband signal (i.e. $F_s > 2xB$ or $\frac{1}{2}F_s > B$) as this would allow reproduction of the full frequency range of the wanted signal. Even under these circumstances, it is good practice to low-pass filter the input at the recording stage so that no frequencies higher than $\frac{1}{2}F_s$ sneak through accidentally. However, a higher F_s means that there are more samples to take, record or transmit but constrictions on recording space or the transmission system may dictate that we cannot do this, so we may have to resort to the limitation of bandwidth of the input signal, as explained above.

Application of Nyquist theorem

Nyquist's theorem, put simply, says that the frequency of sampling ought to be at least twice that of the highest frequency in the signal. Given the explanation above, we can now think about what this means for good-quality reproduction of audio. We know that young people with good hearing can hear audio signals from about 20Hz to 20KHz. Therefore, our upper bandwidth limit (B) is 20KHz and, if we want to be able to reproduce all audio signals, we need to set F_s to at least twice this so the sampling frequency, F_s , needs to be at least 40KHz for perfect audio reproduction.

6.3 Digital Audio Standards

Probably the most common sampling frequency for digital audio signals is the compact disc sampling frequency, 44.1 kHz. On a compact disc, the audio has been sampled 44,100 times per second for each of the stereo tracks (left and right) separately.

The advertisement features a photograph of the Apollo Hotel 1 building at night. The hotel's name is visible on the roofline. In the foreground, a large red circular logo containing a stylized lightbulb shape is positioned next to the text "CISO Conference". Below this, it says "Produced by Inspired". To the right, a white rectangular box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL" and "Dec 5th 2019". At the bottom, a dark banner contains the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right of this banner is the "Inspired" logo, which includes a blue lightbulb icon.

This is slightly above 40KHz and, in theory, would mean that signals up to about 22KHz could be reproduced. In practice, there is a slight safety margin as the filters used are not perfect ‘brick wall’ filters. (i.e. there is no such thing as a filter that goes from passing a frequency at 19.999KHz to one that passes nothing at 20.001KHz. They are always slightly more gentle, depending on the type (the 4 classic analogue filters are Bessel, Butterworth, Chebyshev and Elliptic), so it is prudent to ensure that F_s is a little above the ‘passband’ (i.e. the *baseband* signal contains all the frequencies including those that need to be filtered out. Once a filter is applied the signal is called a *passband* as it contains only those frequencies the filter has allowed through, and so the filter is called a *bandpass* filter).

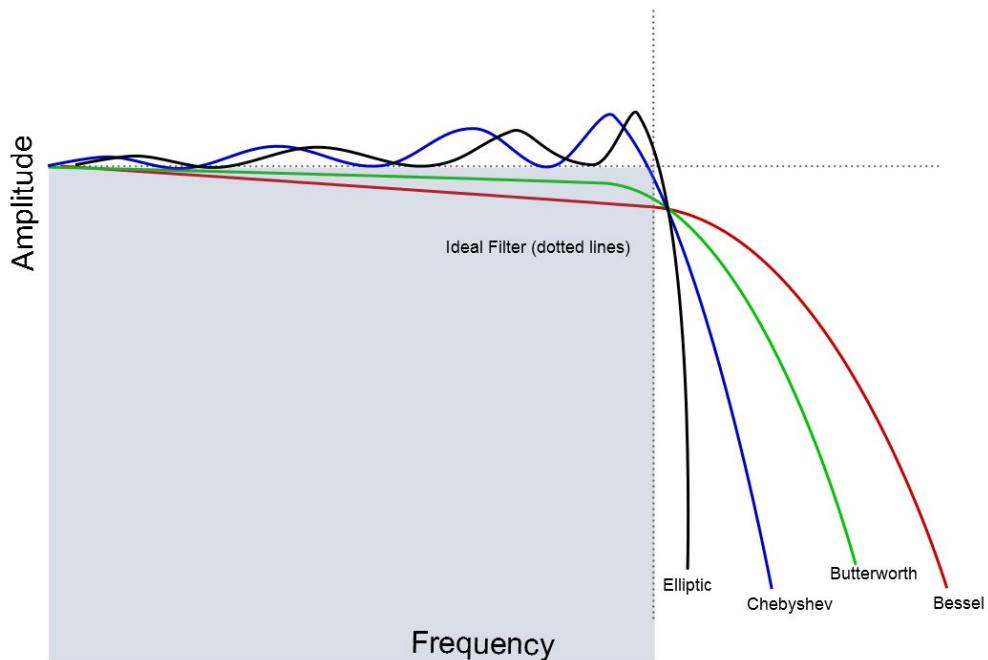


Figure 6-8: Bandpass filters

Note, also, that the steeper the transition zone, the less flat is the passband. The ripple in the passband is another reason why we need to leave so much ‘headroom’ between the permitted maximum level (approx -10dBFS) and maximum coding level (0dBFS).

In the media industry, 48 kHz is commonly used as the sampling frequency (such as in professional camcorders and video editors). Using any other frequency during acquisition will sometimes result in an edit station failing to work properly so you need to ensure that you have set your camera to 48KHz sampling before you start shooting – as you probably can’t go back and re-shoot that once-in-a-lifetime sequence afterwards. However, 32 kHz is also encountered for transmission (e.g. NICAM), as about 15KHz is more than adequate for most TV programmes (Equivalent to FM stereo radio too). It can also be found in digital television and radio but, again, would have to be filtered at recording to prevent any alias components being produced.

This, of course, means that conversion between audio at 32 kHz, 44.1 kHz and 48 kHz is often required. It is usually best to work at the higher frequency (48 kHz) throughout a project as it is always possible to lose signal fidelity at the last stage but it is never possible to put back in what has been lost or wasn't captured in the first place.

'High definition' audio is now available at sampling frequencies of 96 kHz and 192 kHz. It is difficult to justify these numbers on the basis of any audible improvement.

The other important issue is the number of bits required. 8 bits would allow 256 different amplitude ('volume') levels to be resolved for each sample, but the resulting sound would be noisy and unacceptable, so CDs use 16 bits per sample per channel. This allows straightforward handling by computer equipment (16 bits is 2 bytes).

Note that the total bit rate for CD audio standards is over 1.4 Mbits/sec. Sampling at the professional rate of 48 kHz generates over 1.5 Mbits/sec. Yet DAB uses figures such as 128 kbytes/sec. The reason, of course, is that the audio signals are compressed (bit-rate reduced) for digital broadcasting.

6.4 Digital Audio Interface

The most common digital audio interface for professional audio is the AES/EBU interface. This can be used to connect digital audio devices together, without having to convert to analogue and back again.

The AES/EBU interface carries a stereo pair, with allocations of up to 24 bits per sample (per channel). 48 kHz is the preferred sampling frequency, although others are supported. A sender should support one or more of 48 kHz, 44.1 kHz and 32 kHz; a receiver should be able to detect each of these.

The physical coding employed means that there is a balance between positive and negative voltages in the signal: this considerably simplifies input circuitry.

Each sample (per channel) actually uses 32 bits: the extra bits (above 24) are used for framing information (so that the receiver knows where the audio bits start) and other information such as a validity bit, a user data bit, a channel status bit and a parity bit.

The user data bit is part of a repeating pattern spread over 192 samples. 192 bits (using ASCII coding) supports $192/8 = 24$ characters. These can be used as an identifier, e.g. "{programme name}{studio ID}".

There is a domestic equivalent called the SPDIF interface. (Sony Phillips Digital InterFace). It does not have some of the extra bits mentioned above.

AES/EBU connectors can be XLR (using balanced cabling) or phono (using coaxial cable). SPDIF uses only the latter. Optical connectors can also be used.

6.5 Audio Compression

Audio compression takes two forms; dynamic range and bit rate reduction.

Compression of the dynamic range is used to control the sound amplitude.

Bit rate reduction is required as the uncompressed data rate an audio signal needs is too large for many applications.

6.5.1 Dynamic Range Compression

Dynamic range compression is undertaken to ensure that a listener in a noisy environment can still hear the quiet passages of music or speech and to ensure that unexpectedly loud transients do not lead to 'overmodulation' of the transmitters, by suppressing their amplitude. This process where loud sounds are compressed and quiet sounds are expanded is sometimes called companding. However, as we will see companding in the digital domain is also used but with a different meaning so caution is required.

'Overcompression' can sound absolutely awful with 'pumping effects' and modulation of one instrument by another. As most radio stations will apply some degree of dynamic range compression, it is important to ensure that our original material is not already treated like this as cascading (or 'concatenation') of such processes can lead to an unlistenable result for the end user.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com

What dynamic range compression achieves is simply a scaling of the analogue wave amplitude. However, the digital sample rate of a signal that has been treated in this way is still the same (e.g. still being sampled at 44.1kHz in 16-bits for CD after it has been dynamically range compressed in the analogue sense) so the same number of samples is still being generated and needs to be transmitted.

6.5.2 Bit-rate Compression

Bit-rate compression is used to reduce the number of bits that needs to be transmitted in the same way that we did with video so that we can either transmit more channels or transmit at a lower data rate, enabling transmission under more arduous conditions which may prevent high bit rates being transmitted with sufficient reliability. Similarly, bit-rate reduction enables smaller files to be stored and exchanged more easily.

There are two main classes of compression: lossless and lossy. Lossless compression exploits the redundancy within the signal, and is analogous to using WinZip for compression of computer files. The main lossless compression system for audio is FLAC (Free Lossless Audio Compression). Since it is lossless, FLAC is the only bit-rate reduction (data compression) system acceptable for the exchange of audio data files by the BBC, for example.

As we saw with video, lossless compression systems can only go so far by taking out all the redundancy that we can find. However, we still need to reduce the data rate further than this for final transmission so we also need to take out the ‘irrelevant’ information (i.e. the “bits” that we can’t actually hear).

There are several popular techniques used to achieve compression and many share common elements and techniques. The first implementation was the BBC’s development of NICAM 728 (Near Instantaneous Companded Audio Multiplex at 728kbps) which enabled an additional digital signal to be squeezed into the bandwidth already occupied by the PAL analogue video plus mono FM audio signals transmitted within each channel space. This allowed stereo sound to be made available to TV. With the digital switchover (in the UK) now complete, the system is obsolete for TV but its principles form the basis of much that is still current and it is still used for signal distribution of radio channels between studios and transmitter stations.

All lossy compression techniques employ a process of ‘masking’. The human auditory system, in a complex process referred to as ‘psycho-acoustics’, informs us of the source of a sound through its pitch, volume and direction (using phase differences between signals arriving at left and right ears). To do this, it works in the domains of time, frequency and space. Although we know that the best human hearing ranges in frequency from 20Hz to 20kHz, its sensitivity is not equal at all frequencies.

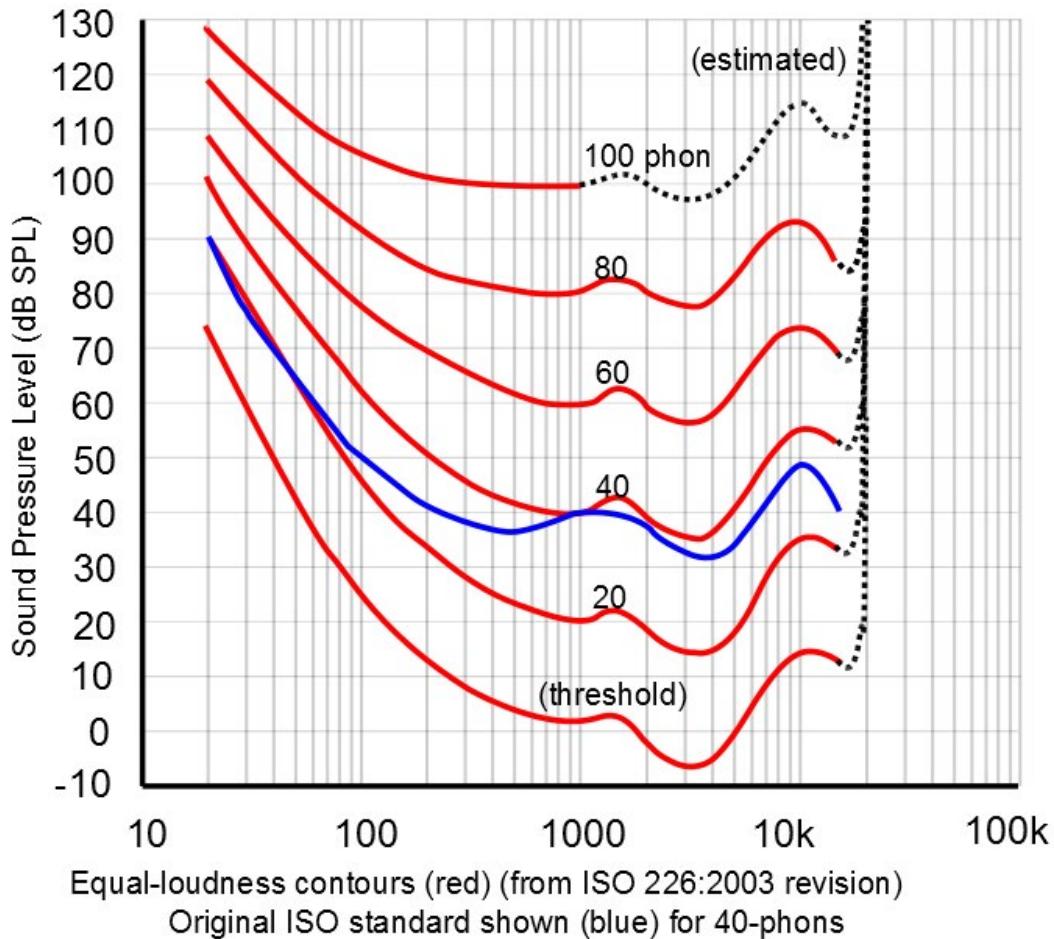


Figure 6-9: Equal Loudness

The graphs show the sensitivity of human hearing at a range of different sound intensities. Notice that the required sound intensity is lowest around 2-5kHz. This tells us that these are the frequencies where our hearing is most sensitive, with other frequencies (especially very low and very high frequencies) requiring a significantly raised intensity before we can detect them.

However, a hearing threshold curve becomes altered by the presence of an audio signal which could 'mask' the presence of one or more additional signals. There are several types of audio masking:

1. Frequency (or spectral) masking, where two or more frequencies occur simultaneously
2. Temporal (or time-related) masking, where two or more signals occur in close time proximity to each other
3. Noise (or intensity to noise ratio) masking, where the number of bits used is smaller than the total required to reproduce the full dynamic range but the bits that are discarded depend on whether the sound is quiet (near the noise floor) or loud (where the noise floor is masked by the loud sound and can, therefore, be disregarded).

6.6 Psychoacoustic Masking

In more modern systems, techniques can exploit the masking effect of a loud signal on another signal.

At a given sound pressure level, a curve of the psycho-acoustic property of human hearing might follow the curve below.

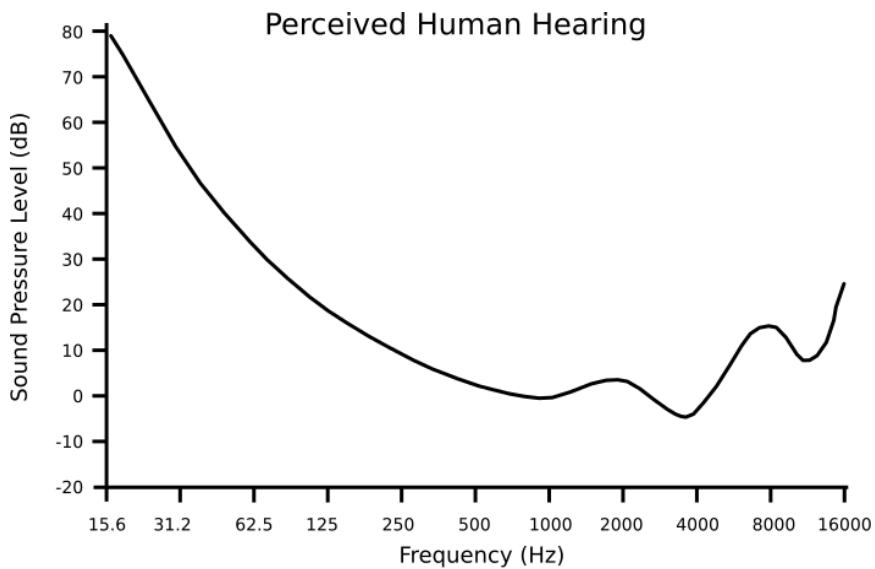


Figure 6-10: Perceived Human Hearing

A woman with long blonde hair, wearing a blue tank top, is shown wearing a black VR headset. She appears to be in a virtual environment, looking upwards with a surprised or excited expression. To her right is a red advertisement box for MT Højgaard.

 **MTHøjgaard**

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Click on the ad to read more

Thinking in terms of sine wave (pure tone) signals (which only contain energy at a specific frequency), then imagine there is a strong signal at 260 Hz, and a weaker tone at 130 Hz. The threshold of audibility is modified by the presence of the louder tone. Provided the 130 Hz signal is below a certain level compared with the level of the louder signal, it will not be heard (but would be heard if the stronger signal were removed).

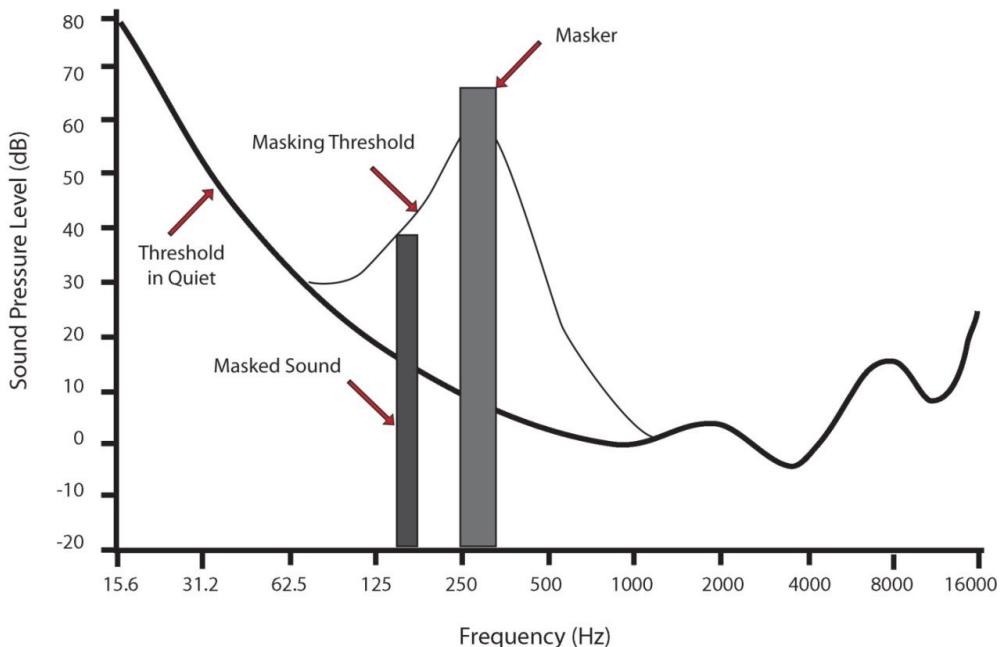


Figure 6-11: Audio Masking

Now, suppose the 130 Hz signal had its frequency changed to 65 Hz (which increases its frequency separation from the strong signal at 260 Hz). Now, the masking effect becomes weaker, so even though the weaker signal can still be masked when its frequency is 260 Hz, it is no longer masked even though the level remains unchanged when its frequency is 65 Hz. This is because the masking effect is reduced as the frequency difference between the masking and masked signals increased (or, to put it another way, the masking effect increases as the frequencies move closer).

6.7 Compression Systems

6.7.1 MPEG

This is one of the techniques deployed within the MPEG encoding systems. Different levels of sophistication and development have taken place over time since the original MPEG1 Layer 1 scheme for audio was developed. For near-hi-fi quality sound, the improving techniques have enabled more efficient coding to be deployed, leading to a lower bit rate for a given quality, typical values of which are:

MPEG 1 Layer 1: 192kbps per channel (384kbps for stereo)

MPEG 1 Layer 2: 128kbps per channel (256kbps for stereo)

MPEG 1 Layer 3: 64kbps per channel (128kbps for stereo) [= .mp3]

Obviously, for a compression system to work based on frequency dependent masking, the system has to take short blocks of the signal, do a frequency analysis, and drop any frequency ranges which would be inaudible. (Dropping components saves bits). The number of samples in a frame over which this technique is deployed varies depending on the MPEG model used. For layer 1 audio, 384 samples are used whereas for layer 2, 1152 are used. The time spans (at 48kHz) are, therefore $1/48000 \times 384 = 8\text{ms}$ for layer 1 and $1/48000 \times 1152 = 24\text{ms}$ for layer 2. The long window period of layer 2 is good for dealing accurately with relatively steady-state periodic signals whereas the short window is better for brief transient signals. Layer 3 (mp3) can use both the long and the short window, switching between them as required.

The resolution of the human auditory system consists of a number of critical frequency bands which are quite narrow at low frequencies and quite wide at high frequencies. MPEG audio encoding attempts to approximately model this by dividing the audio range into 32 equal-width frequency ‘sub-bands’ using a range of frequency filters and then it works on each sub-band separately. Each filter sub-band ‘bin’ enables an examination of the energy in the frequency band in question and, if the psycho-acoustic model suggests that the energy in that band is below the audible or masking threshold then it is discarded, saving bits (i.e. we don’t transmit that frequency if we can’t hear it). This is not a perfect system, however, since the critical bands in the human system are not of equal frequency width whereas those used in the sub-band coding are. This (and overlaps between bands) means that perfect reconstruction of the original signal during decoding is not possible. At the same time, noise audibility determines how finely (how many bits) the signals from the different frequency ranges are coded.

Layer 2 is the system which is used in DAB radio and DVB-T, DVB-S and DVB-C TV broadcasts. Layer 1 is largely obsolete but the practices developed for it are extended in layer 2 and both are shared in layer 3 (mp3), as explained above.

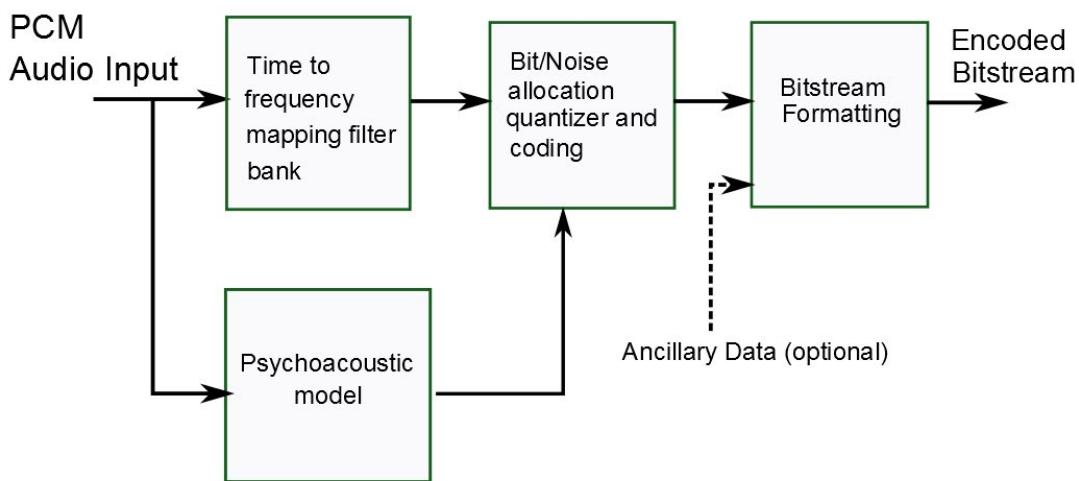


Figure 6-12: PCM to Encoded Bitstream

This approach is used in the well known MP3 (MPEG-1 layer 3) format which also uses modified DCT-based coding to try to compensate for the shortcomings of the filter bank mentioned above. The additional sophistication enables a much lower bit rate for a given audio quality.

At the receiving end the signal block is reconstructed from the parts of the signal which have been retained.

6.7.2 AAC – Advanced Audio Coding

Compression systems can also exploit time dependent masking, which relies on the fact that you can't detect a quiet sound just after a loud sound, or, indeed just before. Obviously the application of the technique depends on how much quieter than the loud sound the quieter signal actually is.

Some approaches do not treat the left and right signals in a stereo pair separately, but exploit the correlation between them by coding the sum and difference signals which saves more bits. Speaking of stereo, entertainment systems often now require more than two channels (e.g. 5.1).

Dolby's AAC (advanced audio coding) system can cope with a large number of channels and claims better performance than MP3 for a given bit-rate. Indeed, countries which recently began digital radio broadcasting use this system, but the UK was one of the first countries to broadcast a system (DAB) and appears to be lumbered with an out of date system. AAC is used for the audio in MPEG-4 systems also.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

Basically, MPEG-4 AAC uses more clever coding techniques including noise substitution and long-term prediction (like the predicted frames in MPEG-encoded video) to achieve a lower bit rate for a given quality. A significant technique is called spectral band replication (where the missing high frequency region of a previously low-pass filtered signal can be recovered based on the existing low-pass signal and a small amount of additional control data). In other words, you can omit some of the frequency range and guess it correctly in the receiver.

6.7.3 NICAM

NICAM is pretty much a legacy system but is still used for signal distribution of radio channels between studios and transmitter stations.

Noise masking is a relatively inefficient technique when used on its own. The technique is called companding (in the digital sense of the word, not as used with dynamic range compression) and is used in the well known NICAM system. The fewer bits you use, the more noise you create in the system. This noise (called quantising noise) sounds random, i.e. like white noise (which is spread uniformly across the audio band).

The crucial aspect of digital companding is this: you need to keep the noise below the hearing threshold. This determines the number of bits you need. It is possible, however, to have a level of noise which you can't hear when the music or speech is loud, but which you can hear when the music or speech is quiet. Systems could be designed around quiet content, which represents a waste when the content is louder.

However, when the content is quiet (even for brief periods) you don't need as many bits for the signal as the upper parts of the audio amplitude range are not used. For example if (for a short period) the level was less than 1/16 of the maximum allowed (positive or negative) then you would need 4 fewer bits than at a loud passage. The bits you don't need on quiet passages are the most significant bits since these are all 'leading zeroes' and can be ignored.

Conversely, on loud passages, the bits you don't need are the least significant bits as these represent the finest, tiny transition values which are swamped by the high signal level and are inaudible. So, you need a 'sliding scale' of bits, and the particular scale you choose depends on the maximum signal value in the time interval in question.

NICAM chooses short intervals of 1 millisec. (If you chose shorter intervals you would allocate a significant number of bits simply having to tell the decoder repeatedly which range you were in. A longer interval might be too crude, allowing variations in level within the interval to reveal either saturation of the signal on transient peaks or the inability to represent low level detail in quiet sections).

NICAM starts with 14 bits at a sample rate of 32 kHz, reduces it to 10 bits depending on the peak value within a 1ms timeframe. Each stereo track is treated separately. A parity (or checksum) bit is added to each sample, which helps the receiver to know whether it has correctly decoded the signal. Together with extra bits to indicate the range of the sliding scale (one of five positions) and framing bits (so that the decoder knows where the bits are in the bit stream) a total bit rate of 728 kbytes/sec was produced. By today's standard, this is not impressive. But it was the first broadcast of digital media into the home, and predated MPEG, MP3, Sky Digital, Freeview, broadband internet and DAB by about two decades.

NICAM was primitive because it did not exploit the masking effect of a loud signal on a quieter signal. It exploited the masking effect of the signal on the (quantising) noise produced by using a limited number of bits.

The advertisement features a night photograph of the Apollo Hotel building. On the left, a large red lightbulb icon contains a white silhouette of a person's head. To its right, the text "CISO Conference" is written in white, with "Produced by Inspired" in smaller red text below it. The hotel's name, "APOLLO HOTEL", is visible in red letters above the entrance. On the right side of the image, a white rectangular box contains the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL Dec 5th 2019". At the bottom, a white banner contains the text "Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders". To the right of this banner is the "Inspired" logo, which includes a blue lightbulb icon and the word "Inspired" in blue.

Part 3: Delivery

7 Media Streaming and Storage

In this chapter, we learn about media streaming techniques and typical devices used within streaming. Streaming is the process of media delivery via computer networks, most notably the internet.

Learning Outcomes

- To explain streaming and how media is transmitted via computer networks
- To give an outline of the encoding and storage of video material.

Media Streaming is the process of transmitting audio and video signals via computer networks, most notably the internet. It requires three parts, a source (to encode the stream), a server (to host the streaming service), and a browser or player (to view the stream).

Streaming is an increasingly important technology to learn about as content continues to be delivered to a variety of internet connected devices.

Internet Protocol Television (IPTV) is effectively streaming programmes (both TV and radio) and movies over the internet instead of terrestrial broadcast. The media streamed may be live (e.g. news) or on-demand (e.g. movies, programmes etc.). IPTV is usually over a ‘closed’ or ‘subscriber’ network e.g. VirginTV (UK) with a specified minimum Quality of Service. This should not be confused with Internet TV (a.k.a. Web TV) which is transmitted using the same protocols but primarily consumed via a web-browser on the ‘open’ internet (eg. BBC iPlayer).

7.1 Stream Creation

7.1.1 Capture

The media stream can be pre-recorded or a live feed that is ‘captured’ and run through an encoder. An audio live feed can be used for internet radio; it requires a sound card to capture the audio input. Sound capture devices can be internal (eg sound cards or integral computer motherboard device) or external devices (e.g. audio interface) though it would be wise to review the earlier chapter on professional versus domestic signal levels before assuming that a built-in motherboard can handle your incoming audio signal. Similarly, live video is via a camera (webcam – poor quality, or video camera – higher quality) but the capture device is typically by video capture cards, IEEE 1394 connection (aka Firewire) or for domestic quality, a USB device. Some USB plug-in devices carry both audio and video signals; it is the software in the encoder that is set to look for where the feeds are attached to the computer.

Current tablet and smart-phones can be used as internet streaming devices. These, along with some cameras, are already ‘cloud connected’ for storage and sharing while IP-Cameras (viewable and controllable on streaming sites) can be used for home security etc. It is possible that these will become more ubiquitous and have greater streaming functionality in the future.

7.1.2 Encoding

Encoding is performed by an encoding software package such as Adobe Flash Live Media Encoder. A ‘live’ feed or pre-recorded media must be compressed and fed into the stream at a suitable bit rate and in a format with which the media server can cope for the purposes of ingest and re-direction. It is necessary for any audio video compression process to have the right decompressor at the play-out destination. Compression (bit-rate reduction) at the transmission end can be in either one or two pass encoding and at a constant bit rate (CBR) or variable bit rate (VBR).

Live encoding must be done in real-time ‘on-the-fly’ so it is a one pass encoding i.e. the image data is analysed and compressed once. Pre-recorded data can be a multi-pass (usually two-pass) process so the quality of the encoding is higher, but is not used in live stream encoding.

Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Click on the ad to read more

Constant bit rate (CBR) is used for media streaming as the encoding quality and feed to the server remains at the same bit rate, so can be at the maximum level the process will allow. Variable bit rate (VBR) is controlled by the bit rate range (Minimum – Maximum) or the average bit rate (each pass is averaged, and then several averages are re-averaged) to achieve a close to uniform bit rate for the stream. Variable bit rate is used for a multi-pass encoding process.

7.2 Network Connectivity

7.2.1 IP address (Internet Protocol Address)

Every computer device that can be connected to the internet requires a unique address so that it can be found, rather like a telephone number. So to host a media stream that can be found on the internet, the media server requires a static IP address. Like telephone numbers in a directory, IP addresses are convertible to meaningful names by a Domain Name Server (DNS) process running on a web-hosting server. Consequently media servers and web-servers are closely coupled, and media is capable of being found using a web-based URL (Uniform Resource Locator) such as www.youtube.com.

The source device (the computer or IP enabled camera etc.) needs to be found by the media server. This means the device has to ‘join’ the media server’s network (a more permanent connection) or pass on its IP address for the session (a temporary connection lasting until the session ends).

Network Router

The link between the source device and the media server may not be a direct connection, but may run through other connecting devices (network servers and routers). A router is a device that redirects data to another connected device either on its own or another network. This is basically how the interconnectivity of the internet works, routing between the server acting as the source device’s internet service provider to other networks and finally being re-directed to the destination.

Multiplexer (MUX)

Obviously it would be inefficient if the stream of data being passed to a media server is solely dedicated to one source device. This is a poor use of bandwidth, (bandwidth being the range of frequencies available in the data stream), so multiple devices are streamed simultaneously. A multiplexer (MUX) is a device that is used to combine input streams into a single output stream which is then split back into individual streams using a de-multiplexer (DEMUX).

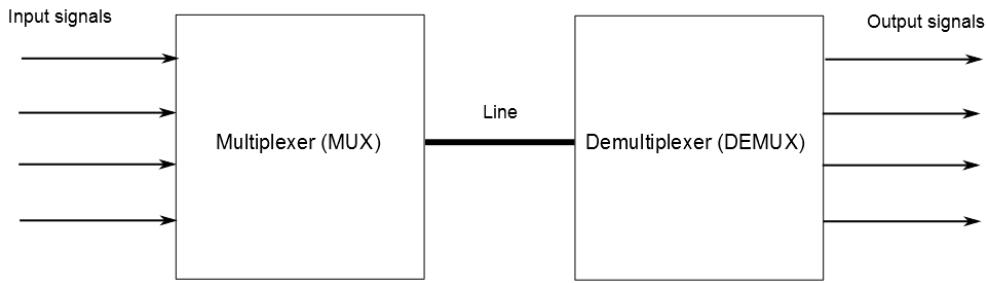


Figure 7-1: Multiplexer to Demultiplexer

7.2.2 Web-Host Servers

For a stream to be found it must be made available to web-browsers or stand-alone players (e.g. Windows Media Centre). The web-host server handles the web-site connectivity but need not be the same computer as the media server but both need a connection between them that isn't prone to interruption. This web site often has a web page (HTML –HyperText Markup Language) with a plug-in media player connected to the media stream being managed by the media server, and contains the web-site in which the page resides. Consequently it must link to the internet and Domain Name Server process to resolve the IP Address from the web-site name, so the routing can direct the end user's browser to the host.

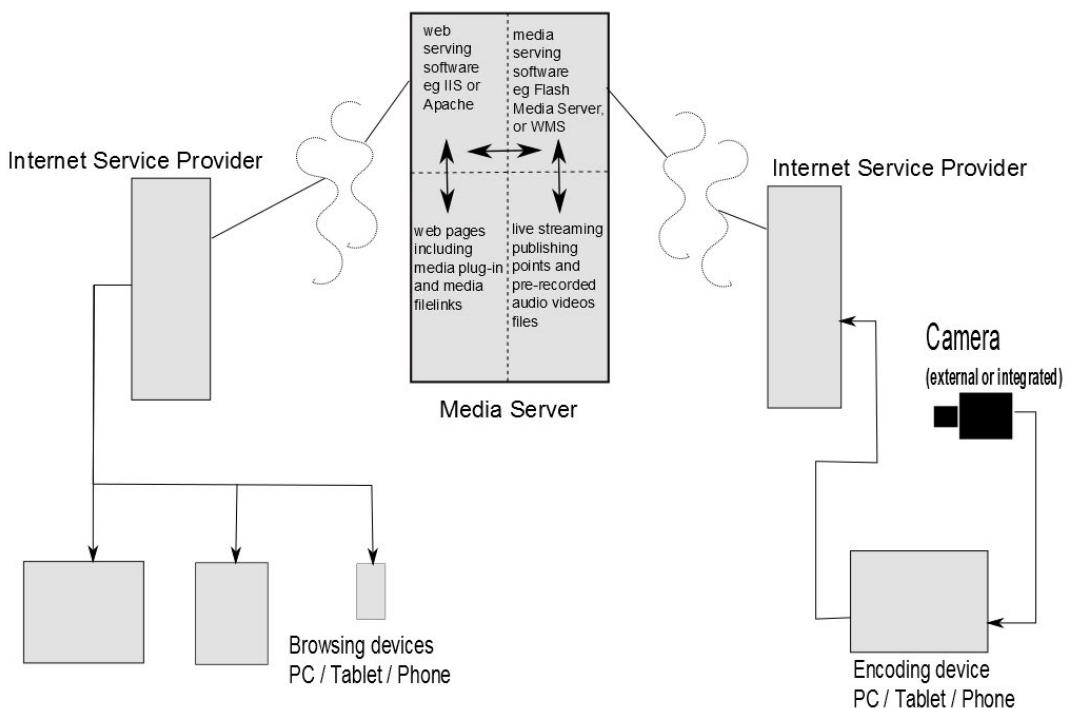


Figure 7-2: Connection diagram

The browser only needs to connect to the internet via its Internet Service provider to link to the media stream's web-host server (see figure 7-2 Connection diagram). Hence any internet device with media playing capability could view the stream including smart-phones, tablets and PCs providing it has the right decompressing codec and can manage the bit rate of the stream from the media server. This is why it is important to offer various streams of differing quality, bit rate and formats (e.g. a Windows Media Video .WMV format file may not play on an iPad without a conversion app).

A further file often created at the time of the stream hosting web page is an announcement file. These are particularly important to make potential viewers aware of the content and set links to the media stream.

Podcast and Vodcast

Podcasts are audio files that are available for download from a web hosting service (vodcasts are video inclusive podcasts) and differ from media streaming in so much as the content is downloaded then played on the user's device. Streamed media is viewable but not downloaded to the device. They are often announced by RSS feeds (a short web content file) to which your device has been connected.

7.3 Media Streaming Servers

A Media server is additional software that runs on a typical web server (or file server with web-host server connection). It requires a static IP – so its address does not alter on each session). The media server software needs to add additional protocols to those found on a simple web-server. In addition to Hypertext Transfer Protocol (HTTP) which is inbuilt with the web server, these additional protocols differ for proprietary server software. Real Time Messaging Protocol (RTMP) is for Adobe Flash systems along with HTTP Dynamic Streaming (HDS). Microsoft Media Services (MMS) is no longer supported for windows streaming and now uses HTTP and Real time Steaming Protocol (RTSP), and finally there is an HTTP Live Streaming (HLS) protocol which is for Apple iOS based systems.

Adaptive streaming (HDS, HLS and Microsoft Smooth Streaming) requires the stream to be fragmented (which is how HTTP delivers content) and may utilise the MPEG-DASH codec. Adaptive streaming is where the content is streamed as fragments in a variety of bit rates with the computer automatically selecting the next most appropriate sized fragment based on its current playback state to minimise buffering. This differs from the older method of providing different dedicated streams at constant bit rates and the client selecting one most appropriate to their computers (or routers) connection capability.

7.3.1 Content Delivery

Unicast

In a Unicast scenario the client connects to the server on a one-to-one basis. The number of clients is limited by bandwidth considerations.

Multicast

In a multicast scenario the server streams to a multicast IP address (this is a special address on the client's network). This is a one-to-many basis and is an effective means of reaching many clients with less bandwidth overhead.

UDP v TCP

All content (including streams) is delivered across a network in packets. In User Datagram Protocol(UDP) the stream is sent without checking the connection and no acknowledgement of receipt is made. UDP is seen as unreliable but it is simpler and quicker. Transmission Control Protocol (TCP) is bi-directional so will check for receipt and retransmit missing packets. TCP is seen as reliable but slower. A good discussion on this is at http://en.wikipedia.org/wiki/User_Datagram_Protocol#Comparison_of_UDP_and_TCP

7.3.2 Live Streaming

A live stream needs to be seen at the time of broadcast. A live stream needs connection to a publishing point on the media server which connects to an encoding device. This may be another computer or a camera with IP addressable capability. The publishing point provides the connection between the content (live stream or pre-recorded) and the client's computer which links to it via a web-host request from the internet.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang

7.3.3 Video on Demand (VoD)

If the stream is recorded then it is treatable in the same way as any pre-recorded media file. Note; it is not advisable to record an encoded stream and then re-encode it as this would severely compromise its quality.

Streams and Playlists

Media streaming servers can have many pre-recorded files ready for streaming, often collected into separate playlists (one media file plays immediately after another). These playlist or file streams can be on a continuous loop, or awaiting selection by a viewer through the browser. This latter selection method is called video-on-demand, although it equally applies to audio files as well.

Bandwidth considerations

A media server can manage several streams and be linked to several web-hosts at the same time. This requires careful planning of the number of streams the media server can handle which is a function of its bandwidth connection. The more simultaneous streams being handled then the less the size of the bit rate is available for each of those streams. If a media server has a 1Gbit/second connection then it could only handle $1000 \times 1\text{Mbit}/\text{second}$ streams. However, full utilization of the bandwidth like this is not normally done; there are recommended bit rates for video streams based on destination image size and aspect ratio, e.g. a $1280^* 720$ HD video with stereo audio will require around a 2.5 Mbit/sec rate. A good source for this is:

http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_live/popup.html

Push – Pull

The relationship between the encoding source device and the media server is based on which device initiates the stream. i.e. it is PULLED from the source device by the media server (needed for video-on-demand), or it is PUSHED to the server by the source device to start the service (a broadcast need). The media server needs to know how the stream is to be initiated for it to start the service.

7.4 Storage

7.4.1 Read – Write speed

All devices that are used to store data (including audio and video) need to be able to write to the storage device faster than the transmitted data is fed to it, otherwise it must buffer (temporary memory store on a faster device) and then read from the buffer to maintain the data transmission sequence. Consequently if a device is used that cannot cope with the data transmission rate than it will fail or lose data e.g. using a low class SD card in a camcorder.

Data is written to storage devices and stored in a binary format but unlike data transmission the Kilo/Mega/Giga/Terra sizes are based on multiples of 1024 (2^{10}) not 1000. The speed of data being read from a storage device may be slower than required to play in real-time which would result in its being prone to stutter and freeze.

7.4.2 Simple Storage Devices Considerations

Tapes – early tapes (DVCAM, DV) required striping. This was to put a continuous time code on the tape before recording – however later tape devices (including mini-DV camcorders) made this unnecessary, though any discontinuities or repetitions in the timecode could cause problems when ingesting material to an editing workstation.

Cards – cards such as SDHC have a class rating which will denote the read/write speed of the card in Megabits per second, and storage size in Gigabytes. Always check to ensure the card will work with the device and check to see if there is a device firmware update, particularly if the device is more than a year old.

USB sticks – as with cards their read write speeds differ wildly – check using an on-line speed testing application, it is usually better to transfer video files to a hard disk before using the file for playback or editing.

CD and DVD disks have a read/write speed depending on the quality of the disk (recording speed), but another consideration is the data rate used in the writing process from such software as the video non-linear editor (NLE). An ‘average’ bit rate (based on Peak and minimum – Variable bit rate (VBR)) or ‘constant’ bit rate (CBR) needs to be selected, that will not only write to the disk but allow the disk to be played on the output device. Computers can write to disks comfortably at 9–11 Megabits per sec. But this needs to be slower (around 5 Mb per sec) if writing to a DVD that is to be played on an older standard-definition DVD player. You should consider the bit rate as part of your consideration of overall file size and the storage capacity on the disk. A good explanation of data rate calculation is given in:-
https://helpx.adobe.com/encore/using/project-planning.html#bit_budgeting

Hard Disks – Many older hard disks (often found in laptops) spin at 5400 rpm, which is too slow for video playback and a minimum 7200 rpm disk is needed. Hard disk read write at around 50–150 Mbytes per sec.

SSD – Solid State Disks are now finding favour due to faster read/write speeds than traditional hard disk technology. SSD read write speed is between 200–500 Megabytes per sec.

7.4.3 Network Storage

Connection speed

Any network storage has to pass data via the network card (NIC – Network Interface Controller), which needs to be as fast as possible (preferably a fibre connection of 1Gbits per sec – but a minimum of 10 Mbs for Ethernet). Unlike other storage considerations network traffic can be bandwidth throttled (i.e. the bit rate is reduced) and will affect speed. If the NIC card is under your administration always set it on maximum performance.

NAS, SAN, Cloud

NAS (Network Attached Storage) is what most people think of as network storage- an array of hard disks that allows for file storage remotely from your computer directly accessible via the network.

SAN (Storage Area Networks) are a separate network but pretty much do the same job as NAS differing in access protocol.

The Cloud is just another remote storage area (uses SAN technology) but accessed via the internet (typified by a URL connection – Uniform Resource Locator) not a local area network LAN connection (typified by a UNC connection – Universal Naming Convention).



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

7.4.4 Raid

RAID (Redundant Array of Independent (or inexpensive) Disks) allows the disk storage to have a measure of redundancy and/or striping to create a secure method of retrieving data should there be a disk failure. For media technology only a few RAID levels (configurations) are used (i.e. Levels 0 and 5 – see figure 7-3 Raid diagram).

Level 0 – usually this requires at least two disks and the data is striped across them. (Note: It can be put on one physical disk using two logical drives – but with little advantage). That essentially means data is split into blocks and distributed across the disks. Typically used in video storage applications as it is fast, there is no redundancy (no duplication) so recovery from a disk crash is almost impossible. Consequently if a disk fails then the file may not be fully recoverable. If you have RAID 0 on your disks, always be sure to keep an external copy of your original audio or video files.

Level 1 – disk mirroring, requires at least 2 disks but is slow as it writes the same data twice (once to each disk). Data is easy to recover as the system has full redundancy (disk duplicated). This level is good for general data and possibly audio only files. Many video editors feel this level is too slow for working with video files.

Level 5 – block striping and parity. This requires 3 disks minimum, data is striped across all the disks (except one) and the block parity is put on the excepted disk. This is done repeatedly using a different disk for parity on each block. One disk can fail and be rebuilt from the others by using the parity blocks on remaining disks. Raid 5 is also popular with video editors as long as the raid controller is fast enough. It is slower than level 0 but faster than Level 1 and has enough redundancy for disk recovery.

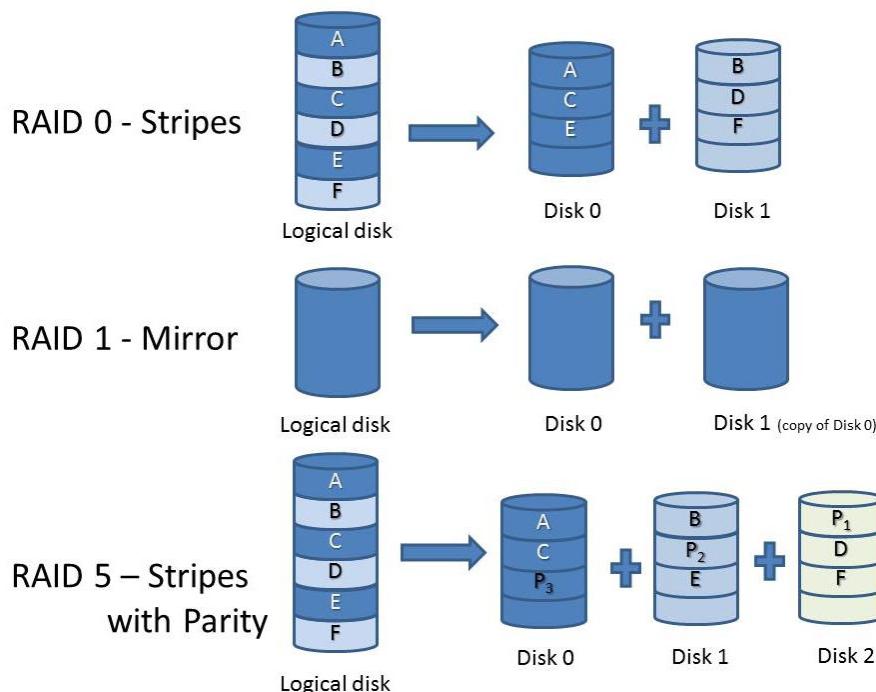


Figure 7-3: Raid Diagram

8 TV Transmission

In this chapter, we learn about television transmission and typical devices used within broadcasting.

Broadcasting is the process of conveying the programme content in a ‘one-to-many’ emission. Unlike streaming, where the bandwidth employed (and, hence, cost) is directly proportional to the number of viewers utilising it, broadcast emission is a highly cost-effective method of providing content where there is a very large number of viewers. The cost is fixed and a single emission will feed as many viewers as want to watch, through a process using either terrestrial transmission at UHF (Ultra High Frequency) radio waves or by Direct Broadcast Satellite (DBS) using microwave frequencies.

Learning Outcomes

- Identify appropriate signal routing for various TV scenarios
- Recognise the importance (and know the location) of published technical standards

8.1 Overview of Broadcast Signal Chain

From the camera to your receiver....

Within a studio environment, a multi-camera set-up will often be used. This allows decisions to be made ‘live’ during the program about the viewing angles, framing, etc., with the director instructing the vision mixer (person) who controls the vision mixer (device). (In the US, the device is often referred to as the vision switcher to avoid confusion). Other vision signals may be generated entirely electronically (e.g. computer-based graphics, captions, animation etc. – see section 8-3). In addition, the production may call for colour keying techniques to be used (e.g. weather map). The electronic signals from all the cameras and other devices have to be mixed together in such a way that they are synchronised to each other to ensure that the top left hand corner of each raster scan on each device starts at precisely the same moment.

Alongside the section of the ‘gallery’ (control room) housing the programme production staff (director, vision mixer, VT operator, autocue operator, producer, production assistant etc.) is a section for vision engineering or vision control. The vision engineers are responsible for setting up the cameras and other signals to be fed around the studio and into the programme output. They constantly monitor the instrumentation to ensure that the video signals remain within allowable limits for transmission. In a studio vision control gallery, they would share the vision control gallery with a lighting vision supervisor who operates the lighting console, under the direction of the Lighting Director.



Figure 8-1: Typical small studio gallery

Allied to the vision is the sound signal under the control of a sound supervisor. At any given moment, the audio heard by the viewer could be created naturally and captured using microphones or it could be an entirely electronically-simulated signal (e.g. synthesised music). In a similar signal path to those of the video signals, the audio signals are controlled at a mixing desk. Unlike the vision signals which require so many parameters to be controlled simultaneously that it is impossible to do so while switching the signals to the output, the sound supervisor monitors the levels and mixes the appropriate inputs to create the final output at the mixing desk in a combined operation.

The output of the live sound and vision mixers may be fed out directly to the onward transmission system or it may be recorded 'as live'. In effect, the editing decisions are being made as a live process during the recording (thus vastly reducing the time needed both to make the show and for any post-production if recorded 'as live') although, sometimes, multiple recordings are made from some of the cameras as well as the output of the vision mixer, allowing for possible further editing, depending on the programme's content. This is generally the only effective way of capturing content that is largely spontaneous and unrehearsed, such as game shows or televised panel shows such as the BBC's 'Question Time' where a panel of politicians responds to questions posed by the audience.

8.2 Signal Routing & Compatibility

At all stages in a TV system, the signals for video and audio must be presented in formats that are consistent and compatible with the infrastructure system and the viewer's receiver. Consider a simplified diagram for an Outside Broadcast (OB) programme such as an important political by-election being covered by a nightly news programme. It might look something like this:

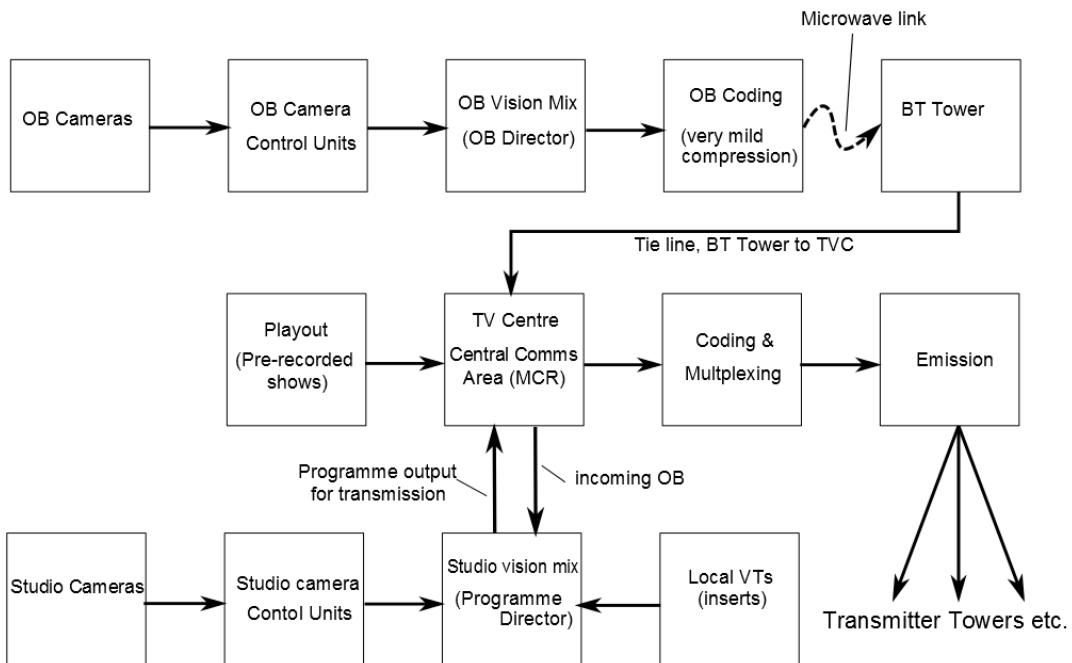


Figure 8-2: Outside Broadcast overview

A photograph of the Apollo Hotel building at night. A large red circular logo with a white lightbulb icon is on the left. To its right, the text reads "CISO Conference" in large white letters, with "Produced by Inspired" below it. On the right side of the image, there is a white rectangular overlay containing the text "Apollo Hotel 1, Groenlandsekade Vinkeveen, Amsterdam, NL Dec 5th 2019".

Listen, learn & build relationships with our Network of CISOs & Cyber Security Leaders

Inspired

Click on the ad to read more

It should be clear that there are many stages during which changing the standard used for the signal could cause havoc. The audio will follow a similar path but remains separate from the video until final encoding and transmission. It is all too easy for the audio and video to arrive ‘out of sync’ with each other caused by unequal processing delays at one stage or another. The aim is always to maintain compatibility and synchronisation at each stage by adhering to the agreed technical standards and to use the highest quality signal for as long as possible.

The situation is further compounded by the fact that whilst 20 years ago, every stage in that block diagram would have been handled by a single organisation (namely the BBC) co-ordinating connections via its central communications area (CCA, known elsewhere in the industry as a ‘master control room’ (MCR)), in the sadly now closed Television Centre in London, the privatisation of most of the ‘behind the scenes’ functions that has happened over the last 20 years means that very little is now handled by what the public still imagines to be “The BBC” and, in fact, at least half a dozen separate companies now provide the functions shown above and the signals are passed through a plethora of different companies’ equipment. (The principles of the diagram, above, still apply, though CCA has split its satellite operations away from London whilst the traditional MCR function has been moved a few hundred metres up the road to the ‘Broadcast Centre’.)

Any breakdown in understanding between all these companies that used to be ‘under single management’ (the BBC) during a live transmission can (and has) resulted in technical breakdown and the viewer losing the signal in some way. Thus – STANDARDS ARE IMPORTANT! Choosing the wrong one can result in failure to get ‘on air’. That applies right from the moment the first content is captured so agreement on standards to be used and adherence to them by every element of the production must be undertaken during early planning as it will affect the equipment inventory required for the show, some of which may need to be booked and hired in advance. The major broadcasters all publish technical requirements for programme makers and their engineering staff. Content must pass technical checks on arrival and will be rejected if it fails. Costs may be incurred in such circumstances.

Studio & OB Equipment Operation

When working in a studio or OB truck, equipment can be more easily controlled than when working on location with a portable single camera. In order to do so effectively, however, it is still necessary to understand how the equipment works and, therefore, how it can be operated most effectively.

8.2.1 Vision Control

The vision control gallery contains the remote control panels for the cameras which can be adjusted so that the video signal occupies an optimum range of luma values without too many overshoots being unintentionally clipped (leading to peaking (or ‘white crushing’)) or undershoots being unintentionally clipped (leading to ‘black crushing’). With white crushing, light (but not white) subjects are overexposed and appear to be as white as the genuinely white subjects. With black crushing, dark (but not black) subjects are underexposed and appear to be as black as the genuinely black subjects. If either of these effects occurs, detail in the picture is lost and the overall lightness of the subject can appear wrong as well.

Thinking about the vision control gallery in an outside broadcast truck at a live transmission of a football match (where one end of the pitch might be in sunlight and the other end in shadow) should give an idea of the necessity of doing this. The staff must continuously monitor and adjust the cameras. Altering the illumination in a studio will have similar, if more controllable, effects.



Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

**Unlock your life.
Bookboon Premium is your key.**

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com

Click on the ad to read more

The control is exercised (in ‘manual’ mode) by the vision control operator(s) adjusting the iris & gain (white level) and sit (black level) of the camera’s signal levels at the remote control panel which feeds signals back to the camera control unit and, ultimately, back to the camera itself. The same remote control panels (RCPs) for the cameras also adjust the gains of the amplifiers built into the camera heads that provide amplification to the red, green and blue signals coming from the sensors in the cameras back to the camera control units (CCUs) in the studio apparatus room. (As usual, providing gain at the ‘front end’ means that the signals to be sent onwards are at a higher level, less susceptible to low-level interfering signals and, therefore, maintain a better ‘signal to noise ratio’, resulting in higher picture quality). By adjusting the gains, the overall colour of the image from each camera can be balanced so that colours are correctly reproduced and so that there is no variation between cameras. The latter is particularly important in a multi-camera shoot since, while the eye is not sensitive to (spatially) finely-detailed chroma, it is particularly sensitive to even small *changes* in hue between shots, especially on flesh tones. When correctly adjusted, switching from one camera to another should not cause a participant’s face to change hue.

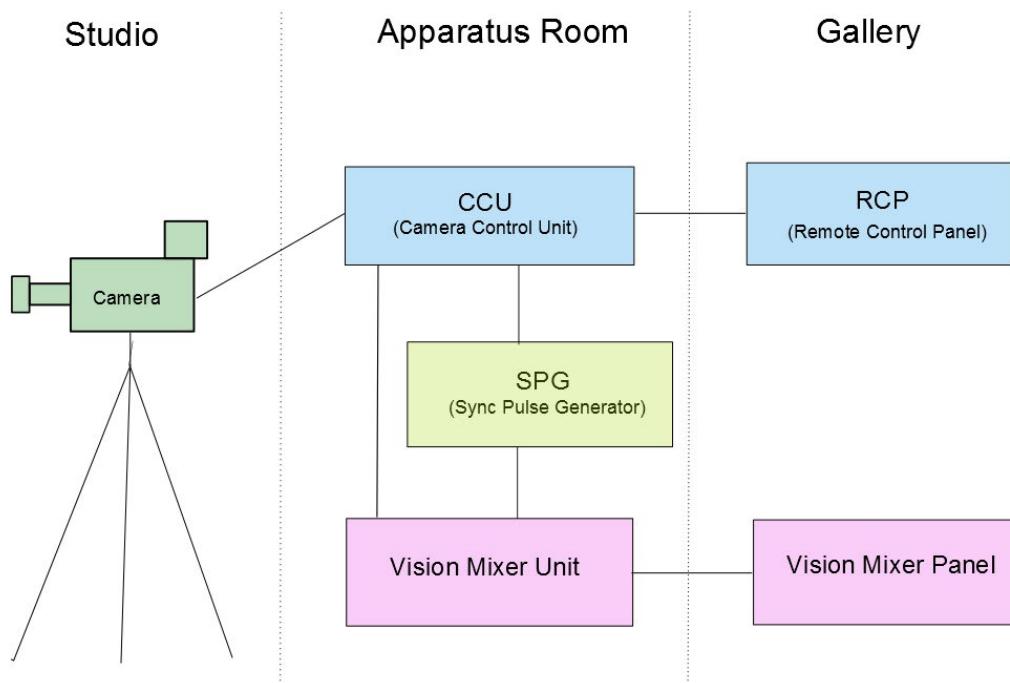


Figure 8-3: Vision Control

As explained in a previous chapter, the use of a black/white/grey multi-level test card (where available) and the vectorscope enables a reasonable setting of the colour balance for each camera by homing in on the central spot which is where both the U and V components are zero (i.e. no chroma) ensuring that the neutral black, white and greys are reproduced correctly without any chroma bias. However, it is usually necessary to ensure that other settings on the cameras’ remote control panels match each other otherwise it may prove difficult to get a match between cameras. A direct split-screen comparison between different cameras will usually enable fine-tuning to minimise any problems.

When adjusting the camera settings in a studio or OB production, it is obviously necessary to work on a video signal which is currently on preview rather than being transmitted, so that any changes cannot be seen by the viewer. However, after initial adjustment, which may be by use of a neutral test card, the operators also have to make subjective assessment based on image quality as seen on the monitors. This, in turn, requires that the monitors are all set up to the same standard so a periodic job of a vision supervisor is to align the monitors. (As with the cameras, equipment should be allowed to ‘warm up’ for an hour or two as the operating temperature of the equipment must be stable before making the adjustments, otherwise they will drift while the temperature changes.)

A suitable test signal exists to enable the monitors to be adjusted, called PLUGE (see figure 8-4). This consists of a zero chroma signal having four blocks of neutral grey up to peak white (110mV, 200mV, 450mV and 700mV), plus two black bars at +/- 2% of the peak video signal (+/- 14mV) centred about the black level.

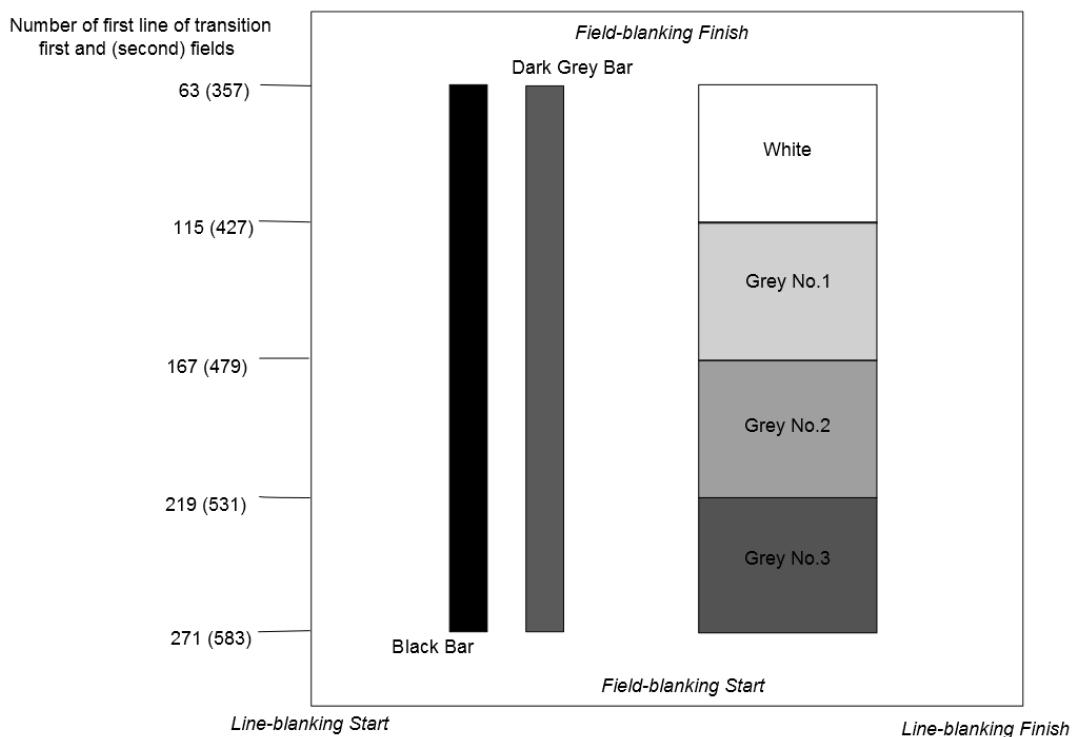


Figure 8-4: PLUGE components

The contrast control on most monitors actually adjusts the peak white level while that marked brightness actually controls the black sit (i.e. black level). When correctly set up, the -2% black bar should be invisible while the +2% bar should be only just apparent. The peak white should be set to $80\text{cd}/\text{m}^2$, where the monitors are assumed to be used in a subdued and correct colour-temperature-lit gallery (6500K). [For those who may be interested, the complete procedure can be found in EBU Technical Recommendation R23 <https://tech.ebu.ch/docs/r/r023.pdf>].

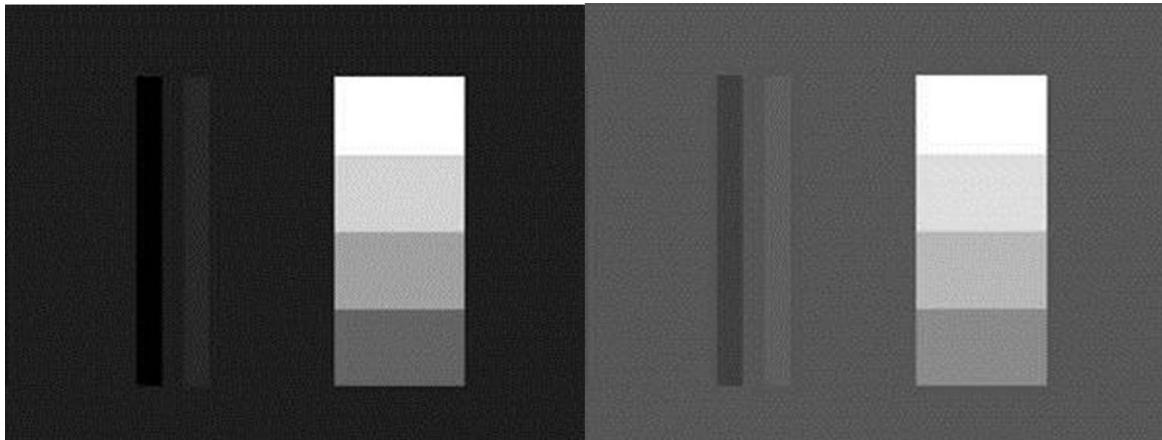


Figure 8-5: PLUGE 'Brightness' comparison

In figure 8-5, the left hand picture is where reducing the brightness control is just about to make the left hand black bar invisible relative to the background, which would be correct when taken a little further than shown. The right hand picture is with the brightness control too high, at the start of the test.

An advertisement for MT Højgaard. On the left, a woman with long blonde hair, wearing a blue tank top, is smiling and looking upwards while wearing a black VR headset. On the right, there is a red rectangular graphic containing text and a logo. The logo consists of a stylized 'M' icon followed by the text 'MT Højgaard'. Below the logo, the words 'BEDRE LØSNINGER' are written in large, bold, white capital letters. Underneath that, a paragraph of Danish text reads: 'I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?' At the bottom of the red section, the website 'mth.dk/vorestilgang' is listed in white. A green button at the bottom right contains a white hand cursor icon pointing to the text 'Click on the ad to read more'.

There is a problem these days when using non-CRT monitors (i.e. LCD or LED) in that many of them tend to ‘clamp’ at black level any incoming signal that is below blanking level, making it difficult to adjust the ‘sit’ correctly. For this reason, some operators are forced to set the left black bar at blanking with the surrounding black a couple of percent above blanking to compensate and the right hand bar a couple of percent above that. This is a compromise as it can result in the whole programme having an artificially raised black level. It is, of course, better to use the correct broadcast-designed equipment in the gallery and generate the correct level signals to the viewer!

It is normal for the studio lighting to be controlled by the vision control gallery too, since altering the illumination also alters all the luma and chroma levels, especially when working with conventional halogen bulbs, which run at a lower colour temperature (more red) as they are dimmed. The operators can ensure that as the lighting is set up for each scene, the camera controls can be adjusted if necessary.

8.2.2 Vision Mixing

In any realistic live (or recorded ‘as live’) scenario, more than one camera is used. A vision mixer is used in the production gallery to select between different sources, and it is possible to wipe, or mix and fade (dissolve), between different inputs. The inputs may include a selection of cameras, internally-generated full-screen colours, video tape recorders (or hard disk recorders), caption generators, graphics and animation generators, test signal generators, incoming live sources from outside the studio complex etc.

Vision mixers can also carry out operations such as colour separation overlay (chroma keying) and other special effects so that one signal can be replaced with another on just part of a screen, having a particular shape (e.g. circle, diamond etc). A studio may also have a graphic generator for captions. This will be a computer-based device but the output from the computer must be in the form of a valid broadcast video signal, not just the output that normally goes to the computer operator’s operational display, as the signal must be fed into the standard broadcast video chain.

8.2.2.1 Synchronisation, Genlock

In general, many video signals are being worked on at the same time in a studio. This can only be undertaken if all these signals are synchronised. This means that the scanning spot reaches any given point on the picture (e.g. the top left) on all devices simultaneously. (If this were not the case then a cut or a mix would try to have picture one with the spot at the top of the screen while mixing or cutting with picture two whose spot may be in the middle of the screen – lack of sync on cuts and mixes doesn’t work!)

How is this achieved? Basically, each studio would have a synchronising master signal (usually generated in a device called a sync pulse generator – SPG) based in the studio apparatus room. In practice, a large studio complex might have a central master SPG which then feeds sync to each studio SPG which acts as a buffer slave. The master SPG, along with master TimeCode generator, may be tied to a remote atomic time clock provided (in the UK) by organisations allied to the National Physical Laboratories, so that the TimeCode and sync pulses are extremely accurate. In broadcast studios, the tolerance on timing is such that line syncs on signals to be switched are expected to be within 12ns (nano seconds) of each other – and that's just for standard definition signals!

These sync signals have traditionally comprised a PAL signal where all the visual content is black so there are sync pulses and the PAL colour burst but nothing else. Consequently they are sometimes referred to as 'black & burst' or occasionally 'colour, black'. On some equipment, the signal is referred to as a Timing Reference Signal or TRS, (not to be confused with Tip Ring Sleeve on audio jack plugs!). HD signals do not, of course, carry a PAL colour burst and the sync pulse is more complicated, being a 'tri-level' sync pulse which swings both below and above the blanking level of 0v.

Next, each video device in the system (cameras, VTRs, test pattern generators, caption generators etc.) is 'slaved' to the studio SPG signal. This means that these devices must have an input called (variously) sync in, timing reference in, or genlock (GL) in. **Genlock** is the mechanism by which each device slaves itself to the master timing reference.

In non-professional equipment (and increasingly nowadays in some that is used in professional studios too), synchronisation is sometimes achieved using 'framestores'. In this technique, unsynchronised inputs are presented to the vision mixer. Inside, digital memory holds (at any given moment) a buffer of at least a frame of each input's video signal but each frame is then 'played out' of memory synchronously with all the other signals that are being treated similarly. (In other words, it delays them all so that they can be played out 'in sync' with each other and the master sync pulse generator.) In theory, the technique works and, for a time, all is well. The snag is that if the inputs are all 'free running' at slightly different rates, there will come a time when the input and the output of the memory have drifted so much that the buffer is full and it is no longer possible to maintain sync so an input will 'slip a cog' backwards or forwards by a whole frame. This is noticeable as a momentary judder or a jump on moving objects. However, modern crystal-controlled oscillators are so stable that this is very rarely seen these days unless a show has been running for many hours.

However, this technique means that the video signal is always delayed by at least one frame behind ‘real time’ so the sound must now be delayed similarly. Thus, the audio must now have delay lines added and be fed with the same sync signals, so the whole setup becomes somewhat more complex again. If equipment is concatenated (i.e. positioned so that signals are ‘daisy chained’) then the re-synchronisers each impose a delay of at least one frame (40ms). Several in series create a significant delay. In general, the practice is best avoided wherever possible, with SPGs continuing to be used but amateur OBs where SPG is not available will tend to use this type of equipment and it is certainly better than having no sync at all!

8.2.2.2 Signal distribution

It is frequently necessary to send the same signals to a multiplicity of devices. For example, the component camera signals often have to be sent to the mixer and a number of monitors. Each device takes some current, so it is not a good idea to hang multiple devices off the same source, although it is sometimes possible to ‘daisy chain’ a couple of devices via their ‘loop through’ connections. (There comes a point where not enough current can be supplied by the source.) Ideally, a separate copy of each signal should be sent. Multiple copies are generated in a device called a distribution amplifier. (The number of outputs is often between 4 and 10). Thus, one signal comes into the DA and many replicated copies of it leave.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.



Click on the ad to read more

In an analogue studio, since the signals have been designed so that R,G,B, Y,U,V, Pal and black & burst all operate within the same range of voltages, it is possible to use the same DAs for any and all such signals, thus saving on complexity and cost. (In digital studios, only SDI DAs are generally required anyway.)

In all cases, at the high frequencies involved, cable doesn't behave like it does at d.c. and starts to exhibit strange properties. A cable at high frequencies must be treated as a 'transmission line' (cf. tying one end of a rope to a wall then giving a pulse to the rope with a flick of the wrist – the wave travels along the rope then bounces back) and the end of the cable run must be terminated with a resistor having the same resistance as the 'characteristic impedance' of the cable. In all normal video systems, this is 75 ohms. However, some equipment has a terminating resistor built in to it (usually switchable but unfortunately fixed where domestic equipment has been used rather than professional). This can lead to cases where a line believed to be terminated correctly is not terminated at all or a line believed to be unterminated already is, yet is terminated externally again (double termination). In either case, for an analogue signal, the luma and chroma values arriving at the equipment will be wrong. A quick way to find out is to hang an analogue oscilloscope on the 'looped through' luma socket of the equipment. If the peak-to-peak voltage is 1V (-300mV sync + 700mV picture) then the line is correctly terminated. If it is 2V then it is unterminated, if it is $\frac{1}{2}$ V then it is double terminated. (A corresponding test on the chroma sockets should also be performed, remembering that the range of the Pb and Pr signals would be -350mV to +350mV when correctly terminated.) A test pattern of electronically-generated colour bars will facilitate these tests.

While the same principles apply to SDI (and especially HD-SDI), an ordinary oscilloscope will prove more problematic to read as the signal consists of a series of extremely rapid changes of voltage. As with attempts to use a digital oscilloscope that has not been designed to read an analogue video signal, a digital oscilloscope attempting to display an SDI signal will appear to be displaying random noise as it struggles and fails to find a sync with which to provide stability. The sensible approach is to use a portable waveform monitor designed to take in SDI signals and convert them into the same human-readable format as the traditional analogue signal waveform monitor.

BNC cables and sockets are normally used in professional environments (see figure 8-6). 'Prosumer' kit often has RCA/phono connectors.

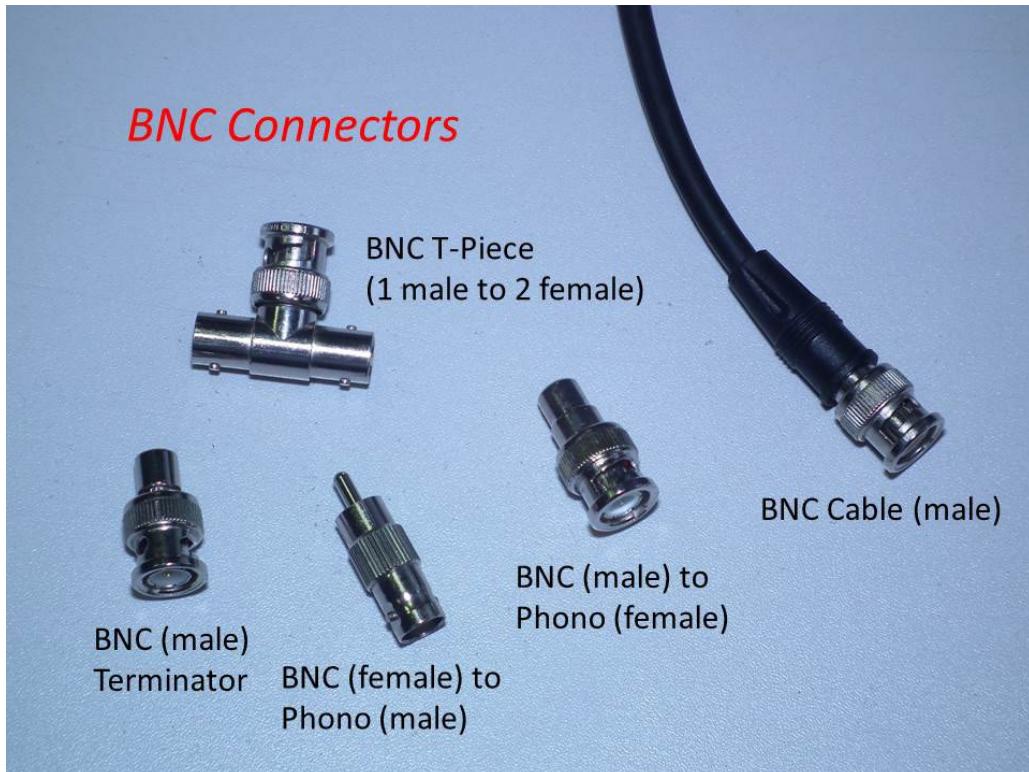


Figure 8-6: BNC Connection fittings

8.3 Effects: Electronic Graphics, Colour Keying

8.3.1 Graphics

Electronic graphics on live video signals, including sub-titling, is carried out using special equipment. The form of the graphics is set up using a PC (or similar) with a broadcast-standard video output.

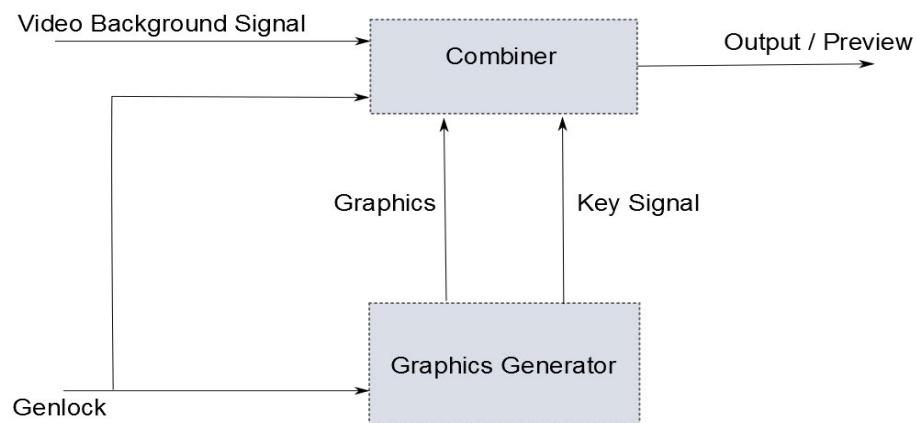


Figure 8-7: Graphics overview

The combiner (see figure 8-7) can be part of another piece of equipment (such as part of a vision mixer). On the other hand, the generator and combiner could be a single unit (so the signals shown as graphics and key signal would be internal). However, that would mean that the output of the vision mixer would then have to be passed through the caption generator, delaying it by at least a frame and introducing the need for audio delays again.

The purpose of the key signal is to indicate how transparent the graphic is. No transparency is sometimes referred to as 'punch through'. Its significance is that when you combine the video background signal with a light or white graphic, the video level must be cut to avoid the combination (the addition) of video and graphic going out of range (i.e. above 700mV). Of course, impressive graphics can be added to recorded material using the power of modern video editors and associated software. The diagram above (figure 8-6) is for live studio work.

8.3.2 Colour Keying

Colour keying (or chroma keying), originally called Colour Separation Overlay (CSO), is the technique whereby those parts of a picture within a certain colour range (generally a combination of hue and saturation) are replaced by another picture or video sequence. This enables a contributor in a local TV studio to appear against an appropriate background of the locality, such as a well-known landmark. It is also popularly used for providing weather maps behind a presenter.

CISO Conference
Produced by **Inspired**

**Apollo Hotel 1, Groenlandsekade
Vinkeveen, Amsterdam, NL**
Dec 5th 2019

**Listen, learn & build relationships with our
Network of CISOs & Cyber Security Leaders**

Inspired



Figure 8-8: Green Screen Studio

Blue was originally used as the keying colour since blue is the opposite hue of skin tone. Generally green is now a more popular background to indicate keying (hence called 'Green Screen' – see figure 8-8), as it is easier to get a lower noise signal from the higher-level green sensor output (enabling crisper keyed edges) even when the lighting is sub-optimal. However, both colours may be used at the same time when using a lenticular reflective background screen. This enables two cameras (one fitted with blue LEDs and the other with green LEDs) to illuminate the reflective screen. The nature of the tiny lenticular beads is to reflect just the LED light back to its source (cf. car number plates) so each camera sees only its own chroma key colour in the background and, in a sophisticated vision mixer, the background perspective can be tailored to each camera's angle of view to allow both cameras to be used during the shoot. (Conventional chroma key backcloth can normally only be used with one camera at a time.)

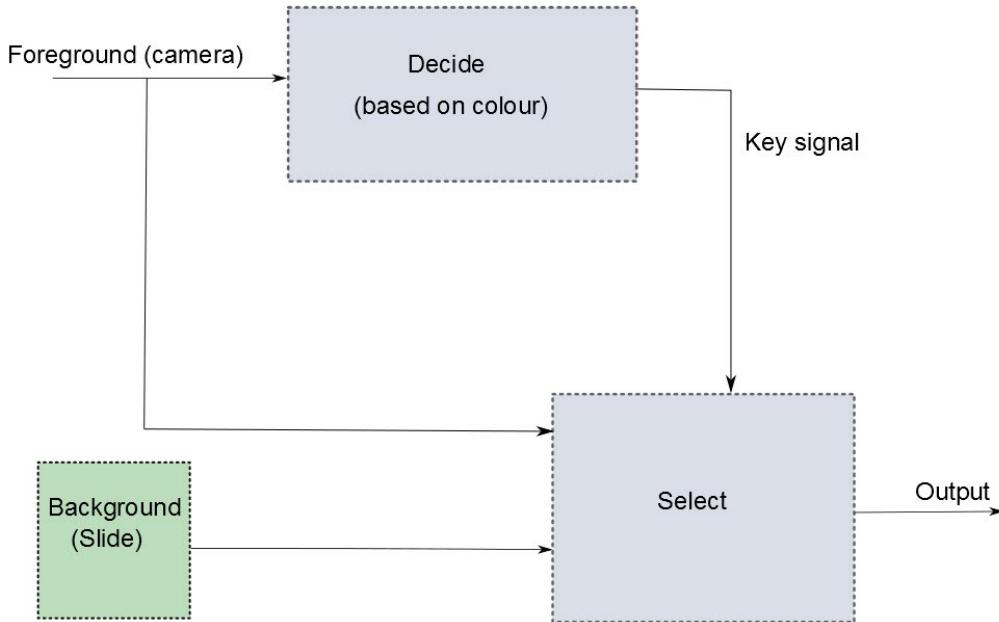


Figure 8-9: Colour Keying

Figure 8-9 is a block diagram of a colour keyer. Colour keying can generally be carried out within video mixers. It is usually possible to select the range of colours on which the system will key: this is generally necessary as the actor/presenter (referred to as the 'talent') can turn up in their own clothes, and the settings will depend on the background colour and lighting in use. The key signal becomes 'true' when the foreground camera sees the colour determined by the combination of luma and chroma values set by the chroma key controls on the vision mixer. For any other combination of luma and chroma values, the key signal is 'false'. Whenever the signal is true, it allows through the background picture to the output; where it is false, the foreground signal goes through.

Experience shows that the result can depend on the nature of the lighting and the fabric of the background material. Shadows can be a particular problem.

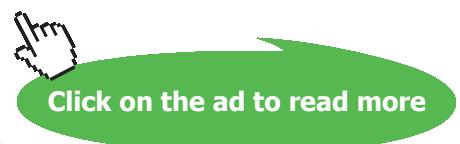
8.4 Distribution

8.4.1 Transmission

Transmission (or, as the EBU refers to it, emission) can take a variety of forms. Conventionally, the signal is relayed from the Master Control Room to transmitter towers around the country which transmit radio-frequency (RF) signals. These are picked up on consumers' roof-mounted aerials and fed to their TV receivers. However, there is now, of course, a plethora of alternatives, including direct-to-home cable TV (where the cable could be copper coax or optical fibre), direct broadcast by satellite with a roof-mounted satellite dish, a web-streaming service or even video over 3G/4G mobile telephony. IPTV is also expected to grow, where the signal is delivered over the internet infrastructure but is not in a web-stream format (which is usually of low bandwidth and, hence, poor quality) but a higher-bandwidth signal that is received by a dedicated 'thin client' (i.e. a computer disguised as a set-top box that is dedicated to the task). In each case, the image seen will differ, depending on the bandwidth available to the type of intended receiver. Image and sound quality are also dependent on the receiver itself. A 3 inch display on a mobile device cannot easily make full use of the same resolution available to the owner of a 50 inch wall-hung screen and, therefore, requires less bandwidth, so less is transmitted.



The advertisement features a photograph of a man with dark hair and a beard, wearing a white striped shirt, sitting at a desk and holding his head in his hands in a gesture of stress or despair. In the background, there's a window with white horizontal blinds and a laptop on the desk. On the left side of the ad, there's a small thumbnail image of an eBook cover titled "Your Boss: Sorted!" by Patrick Forsyth, described as having 55 pages. The main text on the right side reads: "Max's next Bookboon eBook Your Boss: Sorted! By Patrick Forsyth - 55 pages". Below this, a large blue banner contains the text "Unlock your life. Bookboon Premium is your key." in white and orange. At the bottom of the banner, it says "2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business." and provides the website "bookboon.com".



With ‘digital’ TVs (usually flat screen LCD, plasma, OLED, etc), it is not obvious what is going on inside. Generally an interlace to sequential conversion operation takes place inside the set but this can result in quite poor quality images depending on the algorithm chosen by the set manufacturer. Even with a good de-interlacing algorithm, the image quality of an interlaced signal viewed on a flat screen can never be as good as a CRT since a de-interlaced signal will show, simultaneously, both fields of an object moving across the screen which would have been captured at two different instants in time. (Interlacing is covered in another chapter, see earlier.) Also, most flat panel displays will scale the image to fit the screen which results in imprecise alignment between the original sampled pixels and those being displayed, leading to visual distortions and aberrations. However, the generation of HD material coincided with the end of manufacture of CRTs and, other than a few hideously expensive broadcast quality monitors, none exists in practice for the domestic viewer.

Analogue & Digital – Pros & cons

For any given country, in the days of standard definition and analogue transmission, there was one standard. This allowed reliable transmission of the common standard and, within the limitations of analogue systems (usually mainly centred on the problems of noise), there was good reliability and ease of use. Digital transmission now gives us a plethora of different standards depending on the application. Although there are advantages with digital transmission (less susceptible to noise, up to the ‘digital cliff’, flexibility of transmission methods, ability to modify signals using computer-based mathematical techniques etc) and the ability to cram more channels into a given bandwidth (giving commercial advantages), there are disadvantages too. These include the problem of compatibility; with a range of different standards to choose from, converting from one to another is not (as is commonly assumed) a lossless process. In particular, the concatenation of bit-rate reduced (compressed) signals is a major problem for broadcasters, which is why the minimum amount of compression of the material supplied to them should always be practiced.

8.4.2 Supplying content to broadcasters

It is imperative that, long before any production starts, the technical standards required by the broadcasters should be known and all material should be captured in a way that is compatible with those standards. Failure to supply the material to the specified standard will result in automatic rejection. A wise student would, therefore ensure at least some knowledge of the required technical standards for submission of content.

In the UK, the major broadcasters have agreed harmonised technical standards which take account of much which has been touched upon in this book and a lot more besides! Each broadcaster issues some specific modifications to the standard but these are primarily concerned with the details of delivery methods, contact phone numbers, robustness rules for live OB working etc. The generic document on technical standards for delivery is held by the Digital Production Partnership at the following location and will provide the basis on which to build content ready to meet the needs of the broadcaster (i.e. your client!): <http://www.digitalproductionpartnership.co.uk/what-we-do/technical-standards/delivery-standards/>

A typical ‘media’ student, whether interested in TV or Radio, producing, directing, camera or sound operations etc., should, by now, have appreciated the importance of understanding at least some of the basic technical aspects that go into making TV and radio programmes. A read through the technical delivery requirements for the broadcasters linked from the website above will leave nobody in any doubt as to the need to ‘get it right in pre-production’ rather than trying to ‘fix it in post’!



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang