Perceptual Video Measurements

A tutorial by Dr John Emmett of Broadcast Project Research Ltd.

Summary

The separation of television production and display standards from the transmission standards has had the unusual effect of reviving interest in the Human visual system (HVS).

This Paper looks at recent advances in our understanding of this visual system, and the value of system independent HVS measurements are examined. Two examples of practical video measurements are described, both of which depend on perceptual elements. One example involves Colour Gamut and the other involves the detection of visual sequences that are liable to cause Photostimulated Epilepsy in some viewers.

INTRODUCTION

It may seem strange that any Paper in a measurement conference is concerned with factors that cannot be objectively or directly measured, but in any media form there is a final sensory link to the human subject. This final link is often ignored because of the lack of objective understanding, but equally it is becoming the most important link because it may offer the greatest opportunities for future development. The recent advances in low bit rate audio coding have come about simply as a result of only partial understanding of the hearing process.

One common human experience over the last 60 years, has been the exposure to cathode ray tubes (CRT's) showing interlaced television pictures, and this continues to influence most of us subconsciously when assessing new displays and processing techniques. Yet the unexpected experience of seeing MPEG coding artefacts on new and immature display systems, whilst these artefacts remain invisible on CRT based displays, has prompted a new interest in the perceptual and sensory elements of vision.

Out of this interest, and the rapid and almost parallel emergence of several new technologies for television and cinema display systems, there may come a new understanding of the measurements that have previously been accepted as representing visual display performance parameters. It is quite possible that the display measurement parameters that have come to be accepted at present, are simply comparing displays, and have little relevance to the image viewing experience.

MEASURING THE HUMAN VISUAL SYSTEM

Sensory and Perceptual

The Human *Sensory* system acquires a television image via an electrical transducer (display) in conjunction with the environment surrounding the viewer or listener.

Once we receive the sensory information via our neural networks, our brain processes this information. This is said to be the *perceptual* stage. Then and only then, can we be said to actually *experience* seeing.

Let us look first at what vision comprises. Under the brightest sunlight conditions, all of the visual information received by our brains comes from two wandering, roughly circular patches of intense data, no larger in angular size than a fist held at arm's length. At lesser light levels, down to those of a moonlit scene, surrounding monochrome information at reduced resolution augments these patches.

If we are to make any scientific value judgements out of what we "see", the first essential stage is to model the overall visual sense, by finding analogues that can be mathematically described.

Visual Modelling

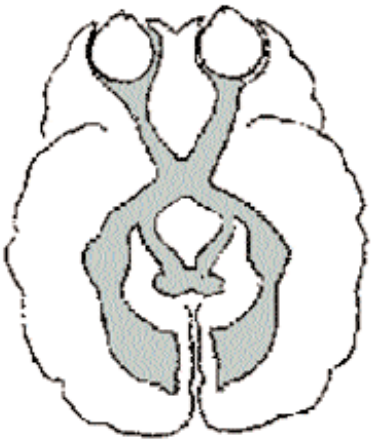
If a process can be divided into separate orthogonal (non-interfering) stages, then each stage should be easier to model, and the overall process can be understood as a cascade of individual stage models.

In the case of hearing, the physiological ear could be physically seen as a sensory mechanical transducer where the nerve impulses form a natural division from the Brain. Only the final stage of hearing involves the perceptual process, when the nerve impulses reach the brain and are analysed as sound. The nature of perceptual processes does not rule out Modelling (indeed, the science of Psychology depends on this!), but the results of the model will vary from individual to individual, and also with learning or experience. Of course, the vision experience also varies in the population, as well as with the age of the subject. Fortunately, unlike those in the hearing process, there are minimal "learned" responses to vision, and the realistic partition of vision models is not as hopeless a task as it may appears at first.

As a start on this process, we could test out lower level vision models, by dividing into the following virtual sections, and then test the individual concepts by comparing the predicted overall results with experimental figures:-

- Formation of the optical image on the retina
- Conversion of the image into nerve impulses
- Interaction between these pulses and their transmission to the brain
- Interpretation in the brain

Measurement and observation, not actually physical partition must establish these divisions, as a quick glance at the diagram of the visual cortex seen from below shows that physical division is difficult. Indeed there could be some physiological justification in considering the retina simply as part of the brain visible to the outside world.



The visual cortex.

It is therefore, remarkably hard to separate the sensory part of vision from the perceived part. One obvious example of this is the difficulty or impossibility of stopping yourself from seeing an obvious illusion. Nevertheless, there are a number of logical deductions and experiments that can be used to expose the sensory limits of vision.

Visual Acuity

Historically, the first objective measurements that were made on vision, concerned the acuity of vision of a "static" object[1]. Traditionally this was measured as the ability to read a letter character at a certain distance under ideal brightness conditions between 30 to 300 cd/m^2 . These levels are (not surprisingly) the preferred brightness levels for television displays, and acuity is normally the first (and often only) parameter to be considered for displays.

20/20 "Snellen" vision implies roughly 30 cycles per degree discrimination, which is the angular spacing of the arms of a letter "E" 8.75 mm high at

20 feet. Other distances are sometimes used, so that 6/6 vision is equivalent in resolving power to 20/20, and 6/60 will therefore be equivalent to 20/200 vision.

20/200 vision implies a 87.5mm high letter at 20 feet, and below this acuity of corrected vision, an observer is normally classed as visually impaired. Peripheral rod vision tends towards this acuity, so as the target brightness reduces to 0.1 cd/m^2, perceived acuity does reduce to this value.

Under ideal conditions, many people can resolve 50 cycles per degree, and this is finer than the physical spacing of the receptors on the retina, so we have not gone far into the field of sensory measurements before perceptual processing must come into play.

Hyperacuity is another important curiosity of visible acuity, where jagged edges, such as those produced on displays by scanning aliasing, can be detected to an accuracy of 0.025mRad. This is getting on for 10 times better than the raw visual acuity. This can be explained as phase sensitivity, but temporal integration could equally explain this phenomena. Other hyperacuitys that are sometimes invoked, such as the ability to discriminate the straightness of lines or the flatness of forms, are less straightforward in explanation and may not even require greater than normal acuity.

Perceived Resolution of Television Displays

From this simple acuity study, we can generate a model of the HVS response in the form of the modulation transfer function (3). This is a useful general measurement because it relates so well to display engineering measurements. However, as we shall see, it only forms part of the HVS model.

We may think of the possibilities of television display *temporal* effects (frame rate, interlace etc) affecting the visually *perceived* resolution, but an entirely different aspect of acuity perception came about as a result of micro probe analysis in the brains of primates such as monkeys (4). This study, and later studies based on real-time brain scanning in humans, has revealed a visual analysis taking place in the brain, where pattern frequencies were "seen" in octave wide bands, and in angular divisions of about 22.5 degrees.

The importance of this discovery is that it had been always assumed that visual aliasing was a severe impairment, (as indeed had been experienced in audio terms). However, as the HVS cannot distinguish the alias pattern "foldback" frequencies from the fundamental image frequencies, a little aliasing *can* be shown to be a good thing in displays! Notice however that this applies at a normal viewing distance, because unlike hearing, visual artefacts change as you get closer and further from the display.

If a display (and picture coding system) is optimised for performance at a given viewing distance, there is no value whatsoever in assessing the system close up.

Going deeper into the area of visual perception modelling, we must use the techniques of psychology. Here, the most valuable available assumption for our purposes is probably Webers' Law. This holds fairly well over a wide range of psychological stimuli, and it simply states that the smallest detectable change in a stimulus (called a just noticeable difference JND, in television terms) is proportional to the magnitude of that stimulus. In reality, it breaks down for vision at very low light levels, due to the presence of noise in the system, but the law holds well within the light levels of the television display regime.

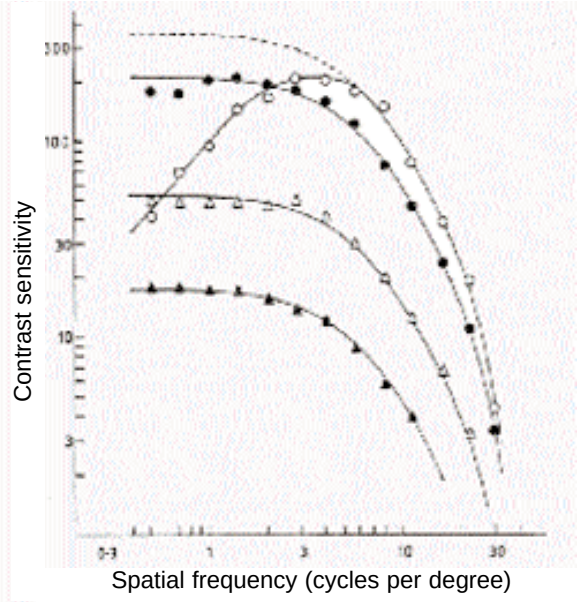
Looking into the significance of just noticeable differences, Robson (1) in the 1960's performed a thorough study of visible contrast ratios, both of patterns and flashing against visual acuity (sharpness of vision). He also investigated the cross effects of modulating the patterns with flicker, and the flicker with patterns. Both the curves produced by him (shown in the figures below), are at first quite surprising, in that they show a significant similarity and cross linking between spatial and temporal (flicker) responses:-

The fall off in responses at low frequencies may not appear surprising to an electronic engineer, as it can be viewed simply as a result of a kind of "AC coupling" in the visual system.

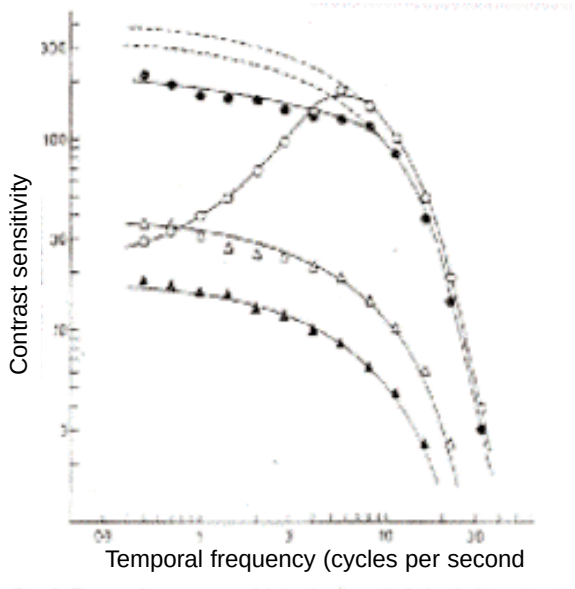
Out of this work, Budrikis (2) in the 1970's, formed the combined three dimensional filter response shown below, with the responses mirrored about the axis. This form of filter response, can be used as our vision "model" not only in order to assess display systems, but also in the field of display standards conversion, aspect ratio conversion, and a host of other two dimensional picture processing.

CONCLUSIONS

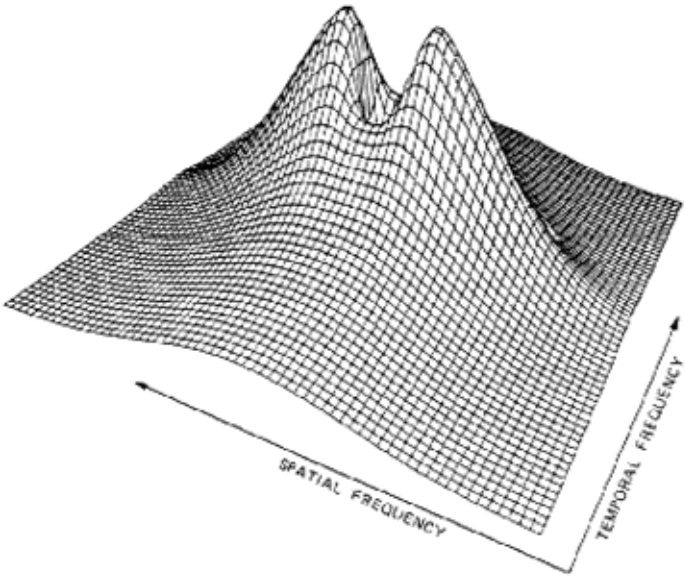
At this stage we probably have a sufficiently good model set to be useful in analysis of new display systems, especially as no thorough HVS work has been used to date for television dis-



Spatial contrast sensitivity



Temporal contrast sensitivity



Perspective view of spatio-temporal frequency response

play analysis. There was no need to do this analysis before, given only one primary form of display.

Subjective differences in new types of viewed displays can only partly be explained in terms of temporal/spatial interdependence (an interdependence that Kell of the US RCA laboratories first detected in the 1930's). Other display impairments can probably be minimised once it is realised that the HVS is far from perfect, and has its own trade-offs. Among the ranges of other parameters that require more than this brief paper to consider are:-

Brightness, contrast ratio and acuity trade-off.
Colorimetry, gamma and white reference.
Interaction of the viewing environment with the display.
The masking effect of visible gratings (colour stripes, scan raster).

The ultimate test of HVS modelling could eventually come in the concatenation of display technologies, (i.e. in telecine transfer, within our area of expertise). If we could predict more about this complex process, an objective concept of the "film look" might lie within our grasp at long last.

BIBLIOGRAPHY

1) Robson, J. G. (1966). "Spatial and Temporal Contrast sensitivity Functions of the Visual System". *J. Opt. Soc Am., 56, 1143.*

2) Budrikis Z. L. "Model Approximations to Visual Spatio-temporal Sine-Wave Threshold Data", *Bell system tech Journal Vol. 52, No. 9, November 1973.*

3) P.F van Kessel et al. "A Comparison of Alternative High-Definition Display Technologies to CRT" *SMPTE Journal Aug 2000.*

4) William E.Glenn "Interlace and Progressive Scan Comparisons Based on Visual Perception Data", *SMPTE Journal Feb 2000.*

5) Chronicle, E.P. and Wilkins, A.J. (1996). Gratings that induce distortions mask superimposed targets. *Perception, 25, 661-668.*

And an excellent general "read": -

Richard L. Gregory, "Eye and Brain", Oxford University Press, 1998.

A.P Ginsberg, "Visual Information Processing Based on Spatial Filters Constrained by Biological Data", Air Force Aerospace Medical Research Lab Tech Report 1978: AMRL-TR-78-129.

R.L. and K.K. Valois, "Spatial Vision", Oxford University Press, 1988.

A.J.Wilkins, "Visual Stress", Oxford Univ. Press, 1995.

Appendix 1

PRACTICAL MEASURING EQUIPMENT

Photoepilepsy Monitoring

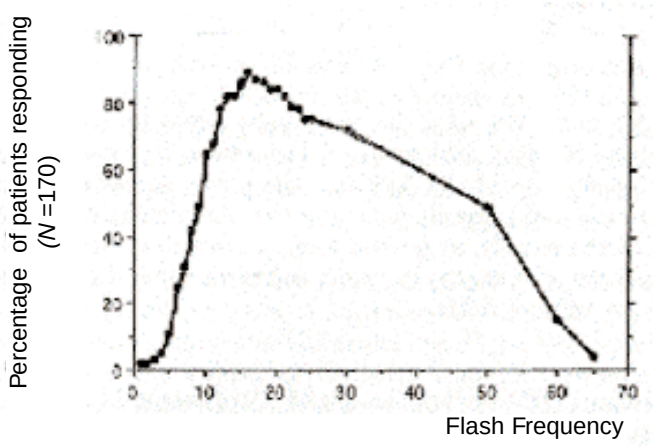
Photosensitive epilepsy can be defined as recurrent convulsions precipitated by visual stimuli[1]. In the UK, it occurs in approximately 1 in 4000 of the population, with an incidence of 1 in 100,000 per annum. In addition there in an unknown number of photosensitive persons who have as yet not had a convulsion.

Seventy-six per cent of patients have their first convulsion between the ages of 8 and 20, whilst only 11 per cent have their first photosensitive convulsion above the age of 20. The condition is more prevalent amongst girls, to a ratio of nearly 2 to 1.

The photosensitive population can be divided into three groups:

Persons who only ever have convulsions in the presence of a flickering light source or visual pattern;

Persons who have convulsions both with flickering light source (or pattern) and without any such stimuli being present;



Persons who are sensitive to intermittent photic stimulation (IPS) in the EEG laboratory, but have not had a convulsion with flickering light or patterns in the outside world.

Persons in group 1 and group 2 are probably aware that television viewing is a contributory factor to their seizures, and if they watch television, they will be aware of the viewing distance, background lighting and other patient specific conditions which may result in those seizures. This still leaves the members of these groups vulnerable to visual material beyond their control, i.e. the programme material itself!

Translating Patient Statistics into the Video Domain.

Patient statistics can be tied quite closely to the visual stimulus values that were presented to them in medical studies. Approaching from the other direction, the recommended television viewing conditions can be related via the display Gamma law, to give the visual stimulus values presented in home viewing conditions.

The overall result allows quantitative figures to be put onto the technical television parameters that are likely to provoke either PSE, or the pattern sensitive form.

Wilkins produced the statistics shown in the following graph from the patient statistics obtained by Harding:-

In order to produce the filter response curve that we require for the video waveform detector, we first need to

apply a “brick-wall” alias filter at 25 Hz to this diagram. This is a necessity because of the sampled nature of the video waveform, and would be modified up to 30Hz for the US and Japanese television systems. At the same time we can transform the percentage of patients responding, directly into the inverse of the light modulation depth. This transform accepts that a modulation depth of just 10% (plus and minus 10cd) will just induce activity at the most sensitive frequency, whilst 100% modulation (plus and minus 100cd) will just not induce activity statistically at 3 Hz. The resulting filter characteristic is used in the “Gordon” equipment (see Ref (5)), developed in a Research project initiated by the UK broadcaster Channel Four Television. The equipment can also incorporate Red flash detectors, as well as extreme pattern detection, representing two lesser, but still significant possible initiators of PSE.

Colour Gamut

Video signals generally originate in the form of the three colour representations Red, Green and Blue. Each of these signals is quite independent, and exists in a full system bandwidth form[2].

As a first stage of data reduction, and for ease of processing, Component Video systems generally work on just one luminance (brightness) signal Y, and two colour difference signals. The colour difference signals are often referred to by their mathematical relationship as “R-Y” and “B-Y”, although in digital systems where there is a code offset but no other difference, the shorthand terms “Cr” and “Cb” is often used.

In the case of Digital component systems to the SMPTE 601 standard, the sampling rate for the colour difference signals is a half of that used for the Luminance signal, and the colour difference bandwidth is consequently reduced. In this form, the signal is often referred to as “4.2.2”.[3]

Composite television systems, such as PAL and NTSC, use a second stage of data reduction, whereby both the Luminance and the colour difference signals were bandwidth filtered further, and sometimes have their peak excursions limited. In the basic form, the colour difference signals are then known as “U” and “V” signals. In the case of the NTSC standard, the U and V signals are further matrix encoded and bandwidth reduced into “I” and “Q” signals[4].

Digital emission standards (MPEG for instance), generally accept the core 601 signals, although adaptive data reduction may well result in transient or static reduction of the effective component bandwidths.

Whatever the transparency of the emission standard, the final display will use some form of conversion back into the three separate Red, Green and Blue signals. As a result of conversions between the colour representation forms back along the signal path, some *relative* combinations of R, G and B signals can be produced that result in colours that do not exist in the original palate. These are the so-called **illegal** colours. However, the major problem with conversion and processing in different signal domains (RGB and Y, R-Y, B-Y), is the relative ease of producing signals that end up outside the **Gamut** (range) of reproducible signal levels.

The visible result for the viewer of out-of-Gamut errors is often unpleasant Hue changes (colour tints) in the high or low light areas. This can then lead to outright rejection of otherwise broadcast worthy material.

Out-of-Gamut signal levels.
The Out-of-Gamut signal level problem is often illustrated by a three dimensional colour space drawing, which shows a diamond shaped Y, R-Y, B-Y colour space, squeezed into the larger RGB colour space that is cubic in shape.

This complexity is not needed to show how the problem exists, as I hope the following thought experiment will show: -

The original Red Green and Blue signals *each* occupy voltage levels of 0 to 700mV peak.

When converted into Y, R-Y, and B-Y forms, the new signals must also exist within the same voltage ranges, the colour difference signals occupying ranges of + or- 350mV and the luminance signal being obtained from the original RGB components using a the typical matrix:-

$Y' = 0.3R' + 0.59G' + 0.11B'$ [5]

Now, if processing is applied to any of the Y, R-Y, and B-Y signals, the resultant voltage limits may well remain within the allowed range for each of the individual Y, R-Y, and B-Y channels. However when these are converted to RGB form, the resultant voltage levels may be outside the allowed limits.

For instance, adding a full excursion (peak white) luminance signal (+700mV) to an R-Y colour difference signal at any negative level from zero to -350mV, will produce Red signal levels of +350mV to +700mV, within allowed levels. However, should the R-Y colour difference signal have a positive voltage level, the resultant Red signal level will reach above 700mV, and thus be *Out-of-Gamut*.

Although both analogue and digital component systems allow for some signal over and under excursions (of the order of + or- 7%), clipping or limiting of any one colour component will imbalance the Hue of the resultant displayed colour.

Under and Overshoots.
There are several mechanisms that can produce transient under or overshoot excursions on video signals. In general, none of the excursions produced in any of these ways will seriously affect the final picture quality. This is because the under and overshoots are visually masked by the picture area edges that create them. This masking has an analogy to the way in which low bandwidth chroma information blends within higher definition luminance pictures in composite transmission standards.

Curiously, the cause of most under or overshoots is probably the direct result of filtering or various forms of bandwidth *reduction.* This typically occurs in the process of standards conversion or resizing of pictures (such as in Aspect Ratio Conversion), and it can occur in either the vertical or the horizontal planes. Thinking in classical terms of the video waveform, the idea of over and undershoots on the horizontal waveform will appear somewhat as in the diagram below:-

In practice, the same phenomena can occur in the vertical direction, especially as a result of Aspect Ratio Conversions. In the video waveform, such over and undershoots may manifest themselves as complete lines out-of-gamut.

Since none of these out-of-gamut excursions will seriously affect picture quality, The EBU Recommendation R103 (4), specifies that an out-of-gamut failure will only occur if more than 1% of the picture area is affected by out-of-gamut excursions. For instance, this will effectively “let through” two complete affected picture lines which may occur above and below Aspect Ratio conversions,

Short electrical domain over

or undershoots (<1 microsecond), can be filtered out, along with some high frequency noise, by means of an "IRE" luminance measurement filter, or a more recent equivalent.

The overall result is that for Production purposes, picture areas that may result in visible impairments should trigger some kind of out-of-Gamut alarm.

For a typical example of such an Alarm, see the specification of the "Hugh" equipment in Reference (5).

Appendix References

4) *www.ebu.ch*

5) *www.bpr.org.uk*

[1] Notice that only the object is static, the eyes of the observer are free to move, so the image certainly is not static on the Retina.
[1] (Harding & Jeavons, 1994)

[2] A Synchronising signal is sometimes combined with one or more of these basic signals, usually Green if on only on one channel, or it may be distributed on a separate fourth channel.
[3] Historically reflecting the approximate sampling rates compared to the NTSC subcarrier frequency.
[4] I and Q refer to the components used for In-phase and Quadrature modulation of the NTSC colour subcarrier.
[5] Y', R' G' B' strictly refer to the gamma corrected colour components, although this is only a non linearity of the transfer function, and does not significantly affect the over or under shoot area where out-of-gamut signals occur.

Dr John R. Emmett

Broadcast Project Research Ltd,
Teddington Studios
Teddington, Middlesex
TW11 9NT, GB