

The eyes, the ears and the brain – and how to cheat them

J.D. Tucker (John Drew Tucker Associates)

1. Early Days

In 1666, Sir Isaac Newton (1643 – 1727), the English mathematician, philosopher and astronomer, began his researches into the nature of light, and “modern theories” on light really date from this time. It was in 1675 that Newton conducted his famous experiment of passing a beam of light through a prism, causing the light to split into its constituent colours.

The three-colour method of printing was first demonstrated by the Franco-German engraver, Jacques Christophe Le Blon (1667 – 1741), but it was not until the beginning of the 19th century that the English physician and physicist, Thomas Young (1773 – 1829), explained how Le Blon’s method worked. It was due to the physiology of the eye. Young postulated that within the retina of the eye there were three distinct sets of colour-perceiving elements or *receptors*, each of them sensitive to wide ranges in the visible light spectrum but with maximum sensitivities in three different colour regions: towards the red, blue and green parts of the spectrum.

A further 50 years was to pass before the German physicist, physiologist and mathematician, Baron Herman von Helmholtz (1821 – 1894) and the Scottish physicist, James Clerk Maxwell (1831 –

Hearing and seeing are just two of our senses and we take them for granted, that is until we lose – or notice an impairment in – one or both of them.

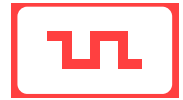
In this Article, brief mention is made of the physiology of the eye and the ear, and how the combination of this complex eye/ear/brain relationship can assist engineers in the design of compressed data equipment. Comments are then made to show that perhaps care should be taken not to push the exploitation of the so-called deficiencies of our seeing and hearing organs too far. Finally some comments are given on more work that needs to be done to try to determine just how much “cheating” can be tolerated by we humans before the results become unacceptable.

1879), were to apply Young’s theory to their own work. In 1861, Maxwell demonstrated that he could reproduce a colour scene by superimposing three lantern slides; one in red, another in green and the third in blue – the three additive primary colours.

By applying the Young-Helmholtz theory to his work, Maxwell was able to specify, quantitatively,

Original language: English
Manuscript received 7/1/97.

This article is adapted from a paper presented by the Author at the EBU/IAB Seminar “Made to Measure”, November 1996.



13 different colours and he illustrated them in the form of a triangular diagram. In the Maxwell colour diagram, the primary colours – red, green and blue – are located at the corners of an equilateral triangle. A mixture of the three primary colours, in equal amounts, produces white light near the centre of the triangle. The complementary colours – cyan, magenta and yellow – appear near the middle of the sides and are produced by mixtures of green and blue, blue and red, and red and green respectively. This is known as the *trichromatic* method of colour specification.

2. The eye/brain combination

To understand how we see, we need to be reminded of the various parts and functions of the human eye and, in particular, the basis of monochrome and colour vision.

However, we cannot consider the eye in isolation – the only difference between red and green light is the differing wavelengths of two extremely high frequency radio waves. The sensation of colour is produced in the brain and is not a physical property of light itself.

Surprisingly perhaps, the image produced on the retina of the eye is:

- upside down;
- left to right;
- curved and not flat;
- unsharp, except for a tiny area corresponding to the image of a small coin held at a distance of two metres.

The image moves rapidly across the retina every time we move our head or turn our eyes. Furthermore, our eyes do not allow distant and near objects to be made sharp at the same time.

Let us now consider the eye/brain combination (*Fig. 1*) by looking at what happens after the optical system of our eyes has formed its image on the retina.

Abbreviations	
DCT	Discrete cosine transform
HDTV	High-definition television
ISO	International Standards Organisation
ITU	International Telecommunication Union
MPEG	(ISO) Moving Picture Experts Group

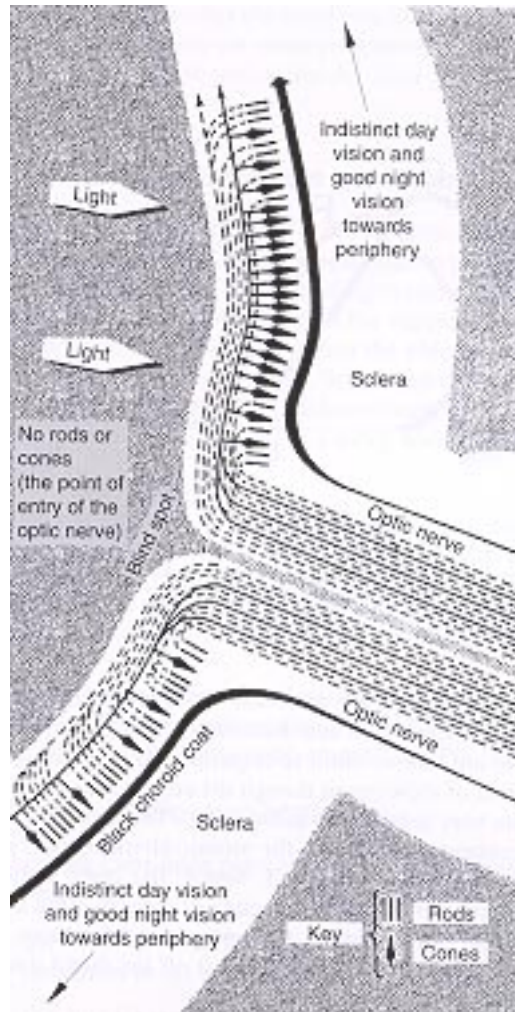


Figure 1
Cross-section of the human eye.

2.1. Monochrome vision

The light-sensitive layer of the human eye consists of millions of extremely small receptors known as *rods* and *cones*. Broadly speaking, the rods are dominant under conditions of low illumination (e.g. moonlight, starlight) and give us monochromatic vision.

2.2. Colour vision

The cones in the light-sensitive layer of the eye enable us to see colour. Without going into too much detail, it is possible to quantify the spectral absorption of light at the cones by measuring the amount of light reflected by the retina back through the pupil of the eye and noting how this amount alters at various wavelengths. The resulting curves (*Fig. 2*) indicate the presence of a red-absorbing, a green-absorbing and a blue-absorbing pigment in the cones.

The presence of cone pigments having different spectral absorptions clearly provides a means of distinguishing between changes in the spectral

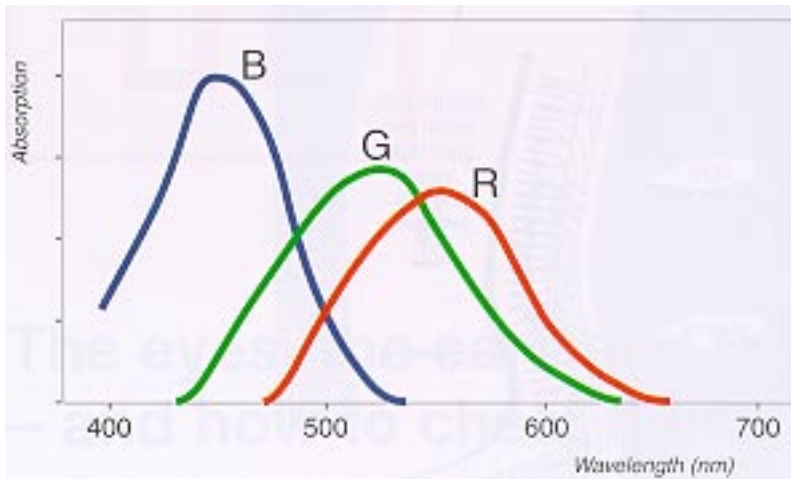


Figure 2
Typical spectral
absorption curves for
the human eye.

composition of a light source, as distinct from changes in its overall intensity. Thus colour vision becomes possible.

The colour-sensitive mosaic elements of the retina are concentrated near its centre which means that we are colour blind over peripheral areas of our field of view (even though the edges of the retina are very sensitive to movement). The nerve cross-connections between the mosaic elements pass in front of the retina and, hence, the optic nerve proper has to pierce the retina in order to reach the brain. In so doing, the optic nerve produces a “blind spot” in our vision, just off the direct line-of-sight.

Our pupils expand and contract to control the intensity of the image but can offer only a 20:1 control ratio. The average eye has 10% flare, which would be intolerable in any camera. Nevertheless, despite these drawbacks, we miss little that is going on around us; our surroundings appear stable despite the somewhat strange images formed on the retina. We have a remarkable ability to judge whether or not a line is straight, we can see well enough over an intensity range of a thousand-million-to-one and most people will insist that their whole field of view is coloured. We are not conscious of our blind spots and it is to be noted that blindness is different from seeing black; in a dark room we see mostly grey and have to put on the light to experience the sensation of seeing a deep black colour.

At first sight, it is attractive to think that the spectral absorption of light by the cone pigments results in separate signals being transmitted from each cone along separate nerve fibres to the brain. However, there are estimated to be about 6 million cones in a human retina and only about 1 million nerve fibres. More-

over, these nerve fibres are also required to transmit signals from the rods which are estimated to number about 100 million. Quite clearly the concept of one-to-one connections between the eye and the brain is false and some “sharing” or “coding” must take place.

The brightness or *luminosity* of an image is also important and some people believe that the retina has a fourth light-sensitive element just to look after this aspect of the image. It is likely that the luminosity signal is composed of signals from separate red, green and blue cones and that this combination takes place before transmission down the nerve fibres to the brain.

All broadcast engineers know about the low luminosity of blue signals. This can be demonstrated by adding red, green and blue lights to produce white. It is found that the luminosity of the red and green beams are similar, whilst that of the blue is very much lower. Thus in colour television, for this reason, the contribution of the blue signal to the luminance signal amounts to about 10%.

If allowance is made for the fact that all blue lights affect not only the blue-absorbing pigments but also the green- and red-absorbing pigments, it is found that the contribution to luminosity of the blue-absorbing pigment on its own is only about 1%. This may seem strange, but it is a beneficial arrangement because the eye is not corrected for chromatic aberrations, so when the red and green images are in focus the blue images are out of focus.

Thus a luminosity signal, based mainly on the red- and green-absorbing pigments, can provide better resolution of fine detail than one based equally on all three. This effect is accentuated by the fact that the blue pigment is well separated along the wavelength scale from the green pigment, whilst the red and green pigments are positioned quite close to one another.

It may be assumed from the above that a luminosity signal is formed in the retina at an early stage in the visual chain, after absorption of light by the three cone pigments. If electrodes are placed close to the output path of the cones, a luminosity type of signal can be detected. If the electrodes are placed a little further away, signals are picked up which, relative to a resting potential, are either electrically negative for green wavelengths and positive for red, or negative for blue and positive for yellow. There is therefore some evidence for thinking that, as in the transmission of colour television signals, so in humans the data is coded into a luminosity and two sum-and-difference signals.



■ 2.3. Adaptation

A further important property of the eye/brain combination is called *adaptation*.

The brain is usually regarded as the place where the visual information is interpreted and it is in the brain where “instantaneous adaptation” occurs. If we pass suddenly from a tungsten-lit area into daylight or vice versa, white objects in the field look white immediately. Yet it has been shown experimentally – by adapting an observer’s two eyes separately and comparing their responses to colour quality – that adjustments in sensitivity are not complete until several minutes after such changes in the illuminant colour have occurred.

Adaptation introduces considerable complexity into the visual chain, but its main purpose would appear to be to minimize the effects that changes in illuminants or viewing conditions have on the appearance of scenes.

Trichromacy, luminosity, stability of colour matches, differential bandwidth requirements and adaptation seem to come from the nature of the retina. The complexity of the retina is so great, with interconnections between the neighbouring rods and cones, that it seems likely that a great deal of coding of information about the image is carried out prior to the passage of the signal along the optic nerve.

There is much more to study about how we see, but perhaps enough has been said to illustrate the importance of the wonderful eye/brain combination that enables us to see, to perceive colour, and to give an insight into some of the complexities of the human viewing system.

Perhaps the above may be best summarized with the following sentence taken from [1]: *Percep-*

tions depend on what the mind can bring to meet what the eye sees, on what concept can fit the percept, on what sense can be read from the appearances.

■ 3. The ear/brain combination

Within the last twenty years or so, we have learned of the importance of taking precautions to prevent or minimize damage to our hearing system. Also, we have become more aware of the various hearing problems that not only affect the elderly, but persons of all ages, and broadcasters have introduced aids such as Teletext/subtitling to allow these persons to enjoy a more normal life-style.

With the advent of audio digital techniques, our hearing system is being called upon to interpret occasional artefacts that are unfamiliar to its analogue nature and to try to make sense of them; more of this later.

To have an appreciation of how we hear, it is necessary to learn something about the various parts and functions of our hearing system (see Fig. 3).

■ 3.1. Structure of the ear

There are three main parts of the ear that enable us to hear:

- the *outer ear* which collects sounds and channels them to the eardrum;
- the *middle ear* which transforms the movements of the eardrum into vibrations of cochlea fluid;
- the *inner ear* which converts the vibrations of the cochlea fluid into electrical impulses which are passed to the brain.

When the electrical impulses generated by a sound leave the cochlea, they are conveyed along

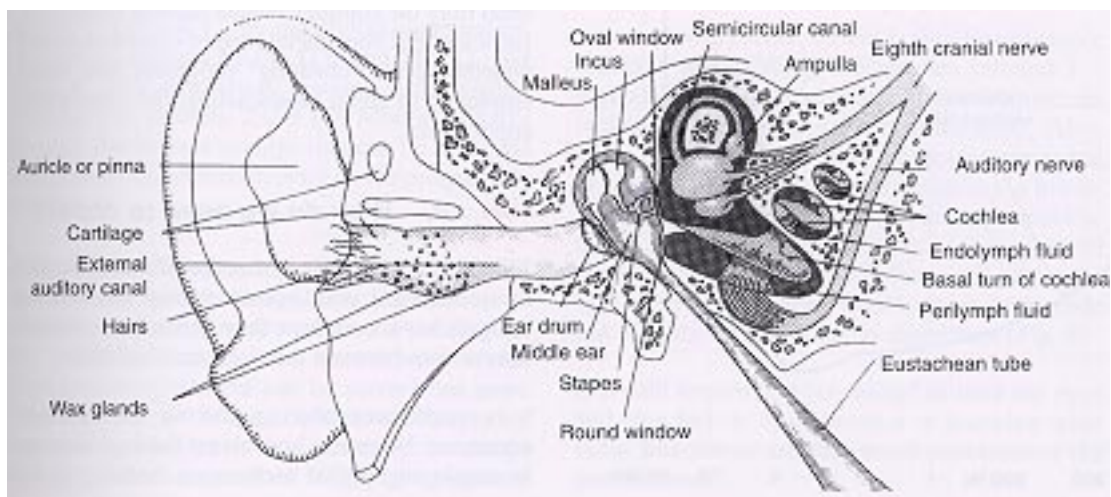


Figure 3
Cross-section of the human ear.

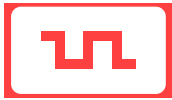


Figure 4 (upper)
Neural pathways showing how sound is transmitted from the cochlea to the brain.

Figure 5 (lower)
Effective hearing acuity of the human ear.

the auditory nerve towards the cortex of the brain (see Fig. 4). It is here that the final decoding of the signal takes place so that we experience the sensation of sound. On the way to the cortex, signals pass through lower levels of the central nervous system, including the brainstem. Arrival at each level is marked by a burst of electrical activity which can be detected through electrodes attached to the head.

There is an important neural connection between the ears which enables the brain to compare the

sounds received by both ears. This enhances our hearing when there is a background of noise (such as at a cocktail party) and assists us in locating the direction of sounds.

Although much-sophisticated processing of the sound signal takes place in the inner ear, it is the brain which performs the function of raising the signals to our consciousness and it is in the brain that we actually hear the sound.

While we can close our eyes, our ears never sleep, although the brain does have the ability to attend to some sounds whilst ignoring others.

3.2. The hearing range

Our hearing range has evolved so that it provides us with a range which is best adapted to our communication needs. The ear of a healthy young person can detect sounds across a range 20 – 16,000 Hz. Of course it does not respond equally to all sounds in this range, being more sensitive to sounds lying in the 1,000 – 4,000 Hz range. This sensitivity range has evolved over thousands of years – from the era when man needed to hunt for food. In those times, his aural (and visual) senses needed to be much more acute than our present-day needs, which are mainly those of communication.

The hearing acuity of the human ear is most sensitive in the range 1 – 4 kHz and becomes progressively less sensitive towards the lower and, particularly, the higher frequencies (Fig. 5).

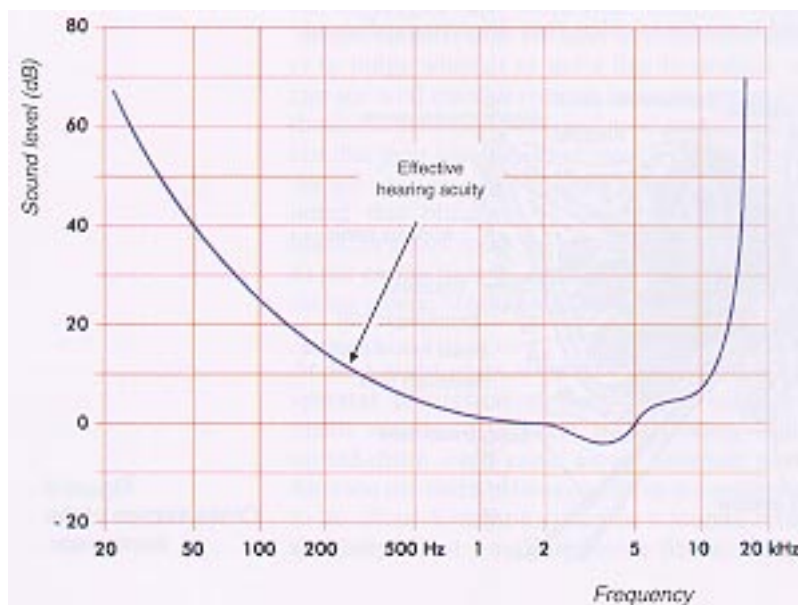
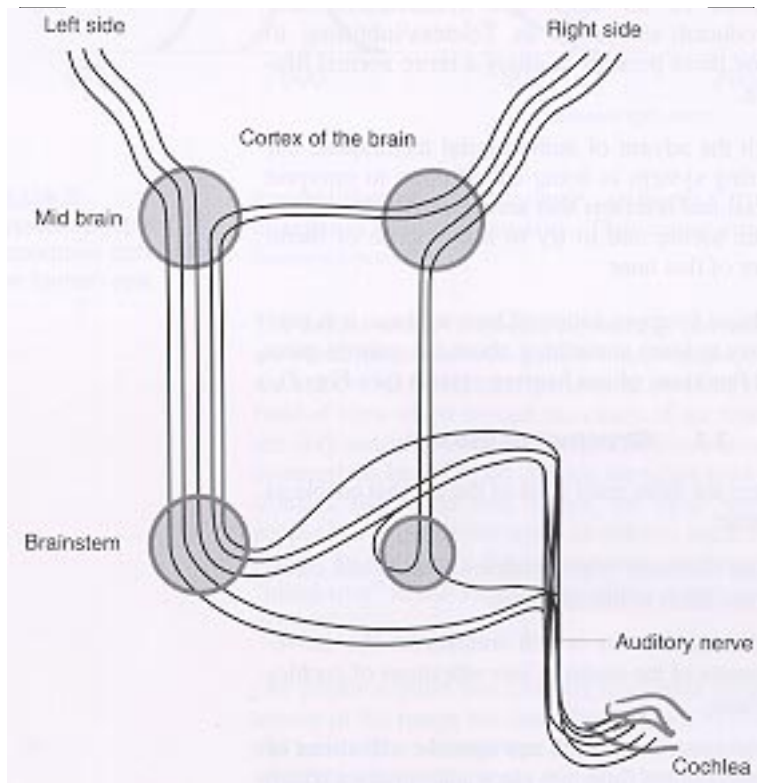
Another characteristic of the human ear is its inability to discriminate between a high volume sound at a particular frequency and a lower volume sound close to it in frequency. Both of these characteristics of the ear are made use of in audio digital coding techniques.

There is much more study necessary to comprehend fully the complex human hearing system, but sufficient has been mentioned to comment on and understand the “cheating” processes now being employed in audio broadcasting and other audio applications.

4. Why do we need to cheat?

This brief examination of some of the basic factors concerned with human seeing and hearing systems has shown how they depend on complex interactions between the eyes, ears and brain.

It is worth remembering that we are analogue creatures. No matter how clever the engineers are in employing digital techniques, nothing in the





foreseeable future (as far as the Author knows) can avoid the necessity of converting digitally-processed video and audio signals back into analogue form for us to be able to see, hear and understand the information being conveyed.

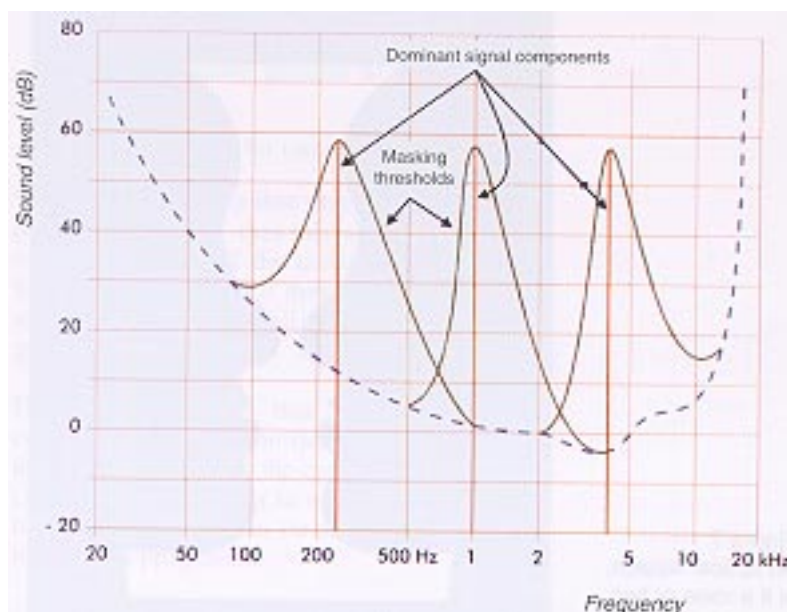
The main reason for needing to exploit the deficiencies of (i.e. to “cheat”) the human visual and aural senses is due to the recent advances in widescreen and HDTV services, and the employment of bit-rate-reduction (compression) techniques to minimize the bandwidth penalties that would otherwise be imposed. Bandwidth is an important commodity which is related directly to cost. Thus, a means of significantly reducing the bit-rate, without loss of quality, is of great importance for both the transmission and the storage of data: the search for such a means has provided a rare example of how the life and physical sciences have come together to help solve problems in the broadcast field. But in doing so – dare I say it – it has raised many more problems that will also have to be solved! The Author wonders if the old expressions “*it fell on deaf ears*” and “*what the eye does not see . . .*” were ready-made phrases for the present time!

To be more serious, how may we make use of the psychophysiology of the eye and the ear to assist in this task of bandwidth reduction?

The human eye is unable to perceive colours in fine detail. This fact is exploited by the analogue television systems in use today, which restrict the bandwidth of the transmitted colour information. Digital compression algorithms also make use of this fact.

Each frame of a typical picture is very often similar to the previous and subsequent frames. Also, the changes from pixel to pixel within a small area of the picture are normally minimal. These two facts are exploited – using techniques known respectively as *temporal redundancy* and *spatial redundancy* – by the digital compression algorithms defined by the ITU and MPEG, and used by other proprietary systems, for transmission rates of between 2 and 140 Mbit/s. Clearly though, the bit-rate savings that can be achieved using these algorithms depend on the programme content.

Spatial redundancy, using DCT transformation followed by quantization, allows a reduction of the high-frequency picture content, depending on the available bit-rate. The quantization of the high-frequency content can be carried out more coarsely than is the case with lower-frequency content, as the eye is less sensitive to abrupt changes (which correspond to high frequencies)



throughout the picture area. Temporal redundancy, on the other hand, is exploited by using motion compensation techniques.

Figure 6
The effect of audio frequency masking.

There are problems, as yet unresolved, in certain applications which use bit-rate compression – especially in digital video effects and chroma-keying (colour separation overlay) where an expert can detect visual artefacts that are unacceptable.

Let us now consider digital audio compression. The same arguments apply here as to video bit-rate reduction, namely the aim is to reduce the bandwidth and storage capacity of data to a minimum, without introducing unacceptable artefacts. The so-called *perceptual coding* technique used here takes advantage of the dynamic characteristics of the human ear to reduce the number of bits required to reproduce the audio material correctly.

As mentioned earlier, the ear is significantly more sensitive to mid-range frequencies (around 1 – 4 kHz) than to the lower and higher frequencies (called the *dynamic sensitivity* of the ears). Also, when the ear is subjected to a high-volume sound at what is termed a *dominant frequency*, a lower-volume sound that is close in frequency cannot be heard. This is called *temporal (or frequency) masking*. These two aural “defects” (dynamic sensitivity and temporal masking) are both made use of in the design of audio algorithms (Fig. 6).

It is well known that “cheating” of both our eyes and ears has its limitations; it is knowing what those limitations are that poses problems at the present time.

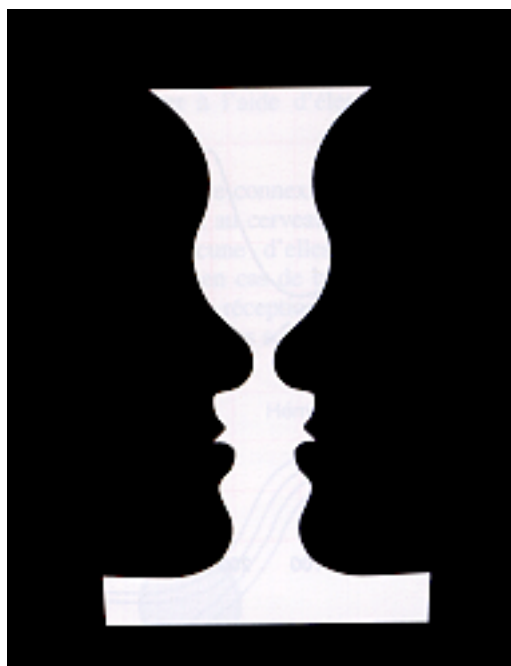


Figure 7
An optical illusion.
Is it a vase or two
people facing each
other?

5. *How far should the cheating of our visual and aural senses go?*

Data compression is a destructive process and a balance has to be struck between the degree of compression being offered and the level of distortion that can be tolerated. Like so many other things, compression is a trade-off and “you get what you pay for”.

The Author’s concern is that, at the present time, digits are the “flavour of the month” and we seem to be rushing headlong into more and more compression without due regard to the consequences.

It can be argued that when a compressed video signal is destined for viewing only, then the features of the eye can be exploited and more compression can be used than if the signal is to be subjected to more digital processing. In other words, it is not just the picture content that is important when considering the compression to be used but, equally, we must know if the programme is “ready” for distribution to the viewer, or if it will be further “processed” for contribution.

As mentioned earlier, the performance of compression systems is dependent on the picture content, and digital impairments vary with time and within the picture area. The viewing experience with digital compression is quite different from analogue and the limits of acceptability are often,

but not always, defined by those critical, mostly unpredictable, picture sequences in live programmes.

Another problem which occurs is that of cascading several compression systems, particularly of different types, in a broadcast chain. Applications that include a commentary feed or which involve a live interactive game show – both of which can introduce too much coding delay on a distribution link – produce problems that the eye and ear cannot tolerate. Of course one solution to many of these problems is to increase the bit-rate, but that begins to defeat the object.

Encoding equipment cannot readily determine what is actual movement in a scene, and is easily confused. Noise also presents a problem to encoding equipment in that it cannot easily distinguish between random noise and movement. Because noise appears as moving dots on a picture, encoders interpret this as requiring motion vectors to be assigned to each dot. Since encoders generally operate to compress the signal bit-rate to a fixed value, this results in the encoder deciding it can decrease the number of moving objects because our eyes are not able to resolve much detail in moving objects. This action results in increasing the size of the dots and making them appear as large moving “blocks”, which is unacceptable to our eyes. This effect is just one of the many complex problems posed by bit-rate-reduction methods on programme material, both visual and aural.

As previously mentioned, the brain is where the visual (and aural) interpretation takes place. We can only appreciate and recognize images and sounds that are familiar to us – things that we have learned during our lifetime.

The brain makes valiant efforts to perceive familiar patterns in the jumble of signals which it is receiving. It does this by comparing the signals with its long- and short-term memory of standard patterns in which allowance has been made for certain standard distortions such as the effect of perspective on apparent size.

Unfamiliar patterns cause the brain to search for meaning before deciding what it is. Sometimes it decides on more than one meaning, because the data presented is confusing. The optical illusion in *Fig. 7* illustrates this; both objects are familiar, but not usually presented in this fashion.

This confusion can also happen with bit-rate reduction when the “cheating” introduces artefacts that baffle our understanding of what we are asking our eyes/ears/brain combination to interpret.



The subjective assessment of bit-rate-reduction systems is time-consuming: it requires the introduction of standard sequences of moving material – covering a range of programme types which include some content of high-colour saturation, noise, lag, etc. – to ensure the tests are as realistic as possible.

So let me be deliberately provocative and say that subjective assessment of new coding or transmission systems, evaluated by a panel of observers, is still the only satisfactory method of assessing the quality of bit-rate-compressed systems – in spite of it being time-consuming both in its operation and in the analysis of results.

Being able to *objectively measure* the quality of compressed systems appears to the Author to be a difficult problem to solve because all compression systems introduce new kinds of impairments on the video and audio data, so that measurement methods used in analogue television are no longer useful in the digital domain.

However, there is a lot of work currently underway in this area (by the MOSAIC Consortium, TAPESTRIES Consortium and others) and it will be interesting to monitor the results over the coming months.

What we are searching for are *objective measurements* to satisfy the user-perceived quality obtained by *subjective assessment* means. This is easily said, but it is a complex subject with answers having to be found, not just for the distribution of data, but for all the many areas of data manipulation (storage, editing, monitoring, etc.) that programme-providers are involved with.

Perhaps we will have succeeded in this task when we can predict – by objective measurements – both the upper bit-rate level(s) and the lower bit-rates for all areas of programme provision, storage, contribution and distribution, which:

- avoid unacceptable artefacts in the compressed data;
- provide acceptable pictures *to the viewer at home*;
- satisfy the criteria for transmission costs.

There is an argument that says: if there are only occasional artefacts, then this is acceptable to the viewer, especially if the transmission costs are low. Possibly so, but if those artefacts have the effect of making the ball invisible just as it is going into the net . . .

Programme providers that allow transmission costs alone to dictate bit-rate reduction do so at their peril because it is the customer who will decide in the end whether to watch or switch off if the “cheating” goes too far and mars the enjoyment of the programme.

6. Conclusions

Timing is always a critical factor when implementing new technologies such as digitally-compressed audio and video, particularly when getting new products to the marketplace. Whilst it is perfectly reasonable and sensible to exploit the deficiencies of the human eye and ear in the march of progress, let us be careful not to go too far, too quickly, without understanding the full consequences of our actions.

Acknowledgements

The author would like to thank RE UK for permission to reproduce *Figs. 5 and 6*.

Bibliography

- [1] Wilson, M.H. and Brocklebank, R.W.: Contemporary Physics, 3, 91 (1961).

Mr John D. Tucker has worked over the years for the BBC, Associated-Rediffusion and EMI Broadcast Division. In 1976, he was awarded a UK Independent Broadcasting Authority (IBA) Research Fellowship, concerned with a study to improve the effectiveness of television programmes to certain handicapped children: the blind and partially sighted, the deaf and those with partial hearing. A report on these studies was published under the IBA Fellowship Scheme in April 1979.

Mr Tucker formed his own consultancy practice, specializing in broadcasting and allied fields, in 1978. Apart from this work, he takes a great interest in “bridging the gap” between the physical and life sciences and, for several years, was a Fellow of the Royal Society of Medicine. He also served on the Audio/Visual Committee of the Royal College of Surgeons.

John D. Tucker is a Fellow of the Institute of Electrical Engineers (UK) and a Member of the Institute of Electrical and Electronics Engineers (USA). He is a member of the Advisory Committee and Faculty of the International Academy of Broadcasting, Montreux, Switzerland and lectures there annually on the psychophysical aspects of hearing and seeing.

