



In previous Hot Topics (The Circuit Switch to Session Switch Transition June 2002) we have shown how adding in additional media components into a session should intuitively increase the length of a session. Adding video to voice in a conversational exchange should in theory give two people more to talk about than if they were just having a voice conversation. This assumption is the basis for 'giving away' video on top of existing voice services ie setting a tariff for voice and video that is at parity with existing voice services. The cost of delivery is higher in terms of occupied radio and network bandwidth but this is more than offset by an increase in average session length, which translates into an increase in billable minutes.

There are however two provisos – the first and most obvious is the power drain and battery capacity in the handset, the second proviso is the quality of the exchange. Through the 1990's, call lengths increased year on year by small but significant amounts. Partly this was due to tariff reductions but also it was due to reductions in handset power drain, better voice quality and a lower dropped call rate. The reductions in handset power drain were partly due to improvements in processor efficiency but also to an increase in network density (having a base station in reasonably close proximity) and to a steady improvement in handover and power control. Whatever the reasons, better voice quality, lower dropped call rates, lower power drain and more battery 'bandwidth' helped to increase billable minutes of use.

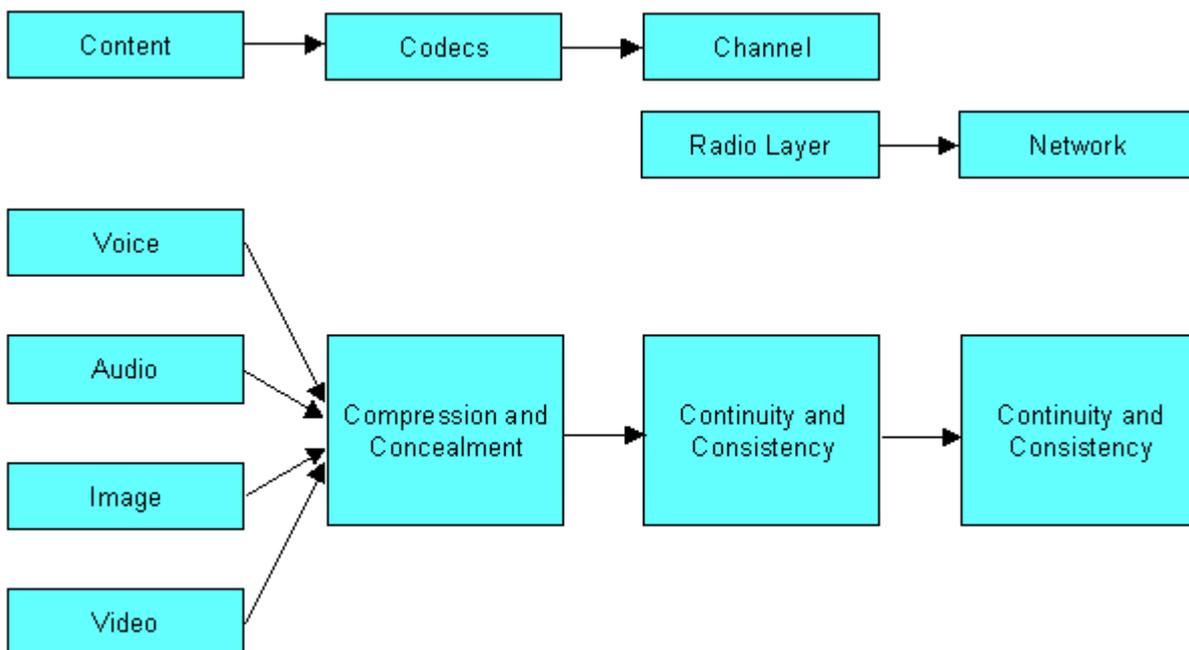


Figure 1: The Impact of Quality on Session Length

Consider the additional components in a multi-media session. Figure 1 identifies four content streams, voice, audio, image and video. The quality metrics in voice are

reasonably well understood and can be measured using a mean opinion score. The speech synthesis voice codecs used in cellular to-day work by doing a time domain to frequency domain transform to identify redundancy and then exploit sample to sample similarities to reduce the source coded bit rate. All present voice codecs also have error concealment – for example frame substitution when channel error rates are high.

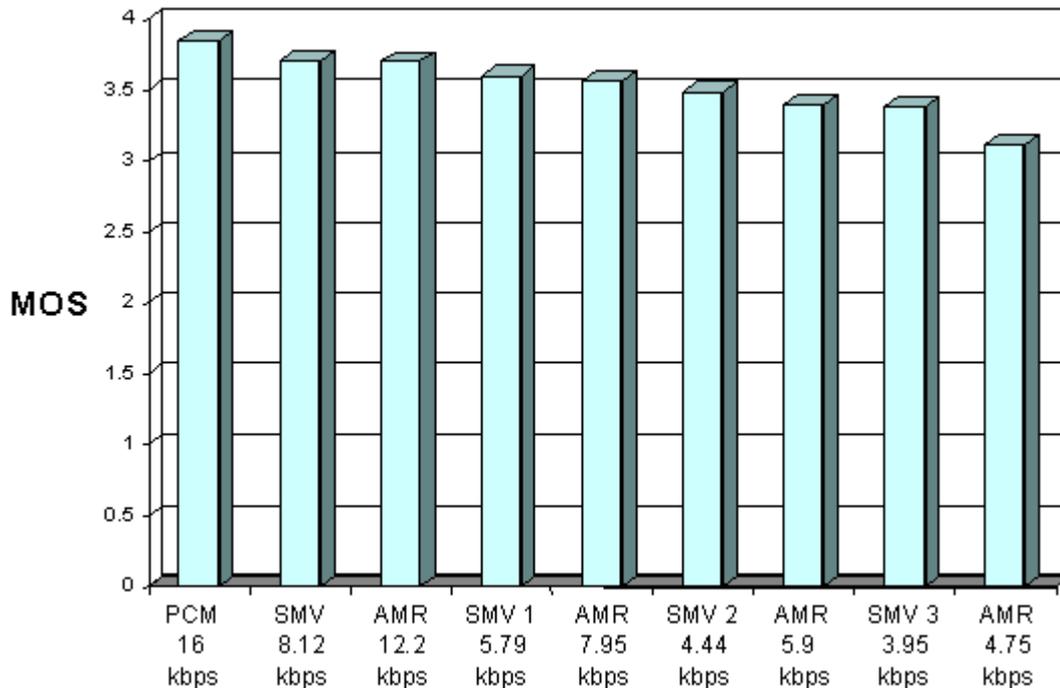


Figure 2: SMV Modes vs AMR Rates (Listening Tests at Dynastat Labs 10/00)
 (Source: www.cdg.org)

Figure 2 shows some typical bandwidth quality trade offs comparing the selectable mode vocoders originally specified by 3GPP2 and the adaptive multi rate vocoders specified by 3GPP1. The SMV coders provide a better quality/ bandwidth trade off. The cost is some additional processor overhead and some additional processor delay.

Similarly audio quality is reasonably well defined (courtesy of the audio industry) and can be described quantitatively in terms of frequency response, dynamic range and signal to noise ratio.

Image quality is trickier. In J-PEG encoding, the usual quality metric is to use the Q rating used in digital cameras. As with voice codecs, a J-PEG encoder does a time domain to frequency domain transform and then exploits block to block similarities to reduce the source encoder bit rate. Ignoring block to block differences reduces the Q. Fine camera mode has a Q factor of 90, standard camera mode has a Q factor of 70. To take an extreme example, a Q90 image of 172820 bytes reduces to a Q5 image of 12,095 bytes but only by trading a significant reduction in captured image quality. Note that these compression techniques are primarily intended to improve storage bandwidth efficiency rather than delivery bandwidth efficiency. In the above example, the 172,820 byte file would take just over 40 seconds to send over a 33kbps uncoded channel. The 12095 kilobyte file would take less than three seconds to send but the original image quality would be lower **and** the file would be less resilient to channel

errors introduced in the radio layer and network.

M-PEG is similar TO J-PEG but does frame to frame comparisons to reduce the source encoder bit rate (sometimes described as differential coding). Differentially encoded bit streams are generally intolerant to discontinuities introduced either at the radio layer or in the network.

To summarise, voice, image and video encoders produce a representation of the original content in which some of the original quality will have been lost. The encoded bit streams are then sent across a relatively high error rate radio channel and into a network that will introduce delay and, if buffering is allowed, will introduce delay variability. Delay and delay variability may further reduce voice, audio, image or video quality. The error concealment used in voice and video encoding will help mask some but not all of these 'channel effects'.

Finally the quality as perceived by the user will also be determined by the quality of the audio and video components used in the receiver. There is no point in sending high bandwidth audio to a device with a poor quality audio output. There is no point in sending a 15 frame per second video stream to a device with a display with a 50 millisecond refresh rate (which cannot manage anything faster than 12 frames per second).

Picture quality is particularly hard to pin down. One solution is to use 'subjective quality factor' which is more or less equivalent to the mean opinion scoring used in voice. A more objective method is to use a method known as the Modulation Transfer Function Area.

The quality of a video stream is partly a function of the frame rate and colour depth, but also a function of the contrast ratio, resolution and brightness. Contrast ratio defines the dynamic range of the display (the ratio of the brightest white the system can generate divided by the darkest black), resolution defines the display's ability to resolve fine detail expressed as the number of horizontal and vertical pixels. Brightness is, well, brightness, measured in foot Lamberts or candelas per square meter.

Brightness captures attention, contrast conveys information.

To be meaningful, brightness and contrast need to be characterised across the range of spatial frequencies being displayed. Spatial frequency is the ratio of large features to small features – the smaller the features in an image, the higher the spatial frequency. The overall number of pixels in the display determines the limiting resolution. The modulation transfer function is a way of comparing contrast to spatial frequency. As features get smaller, the contrast ratio will reduce as shown in Figure 3.

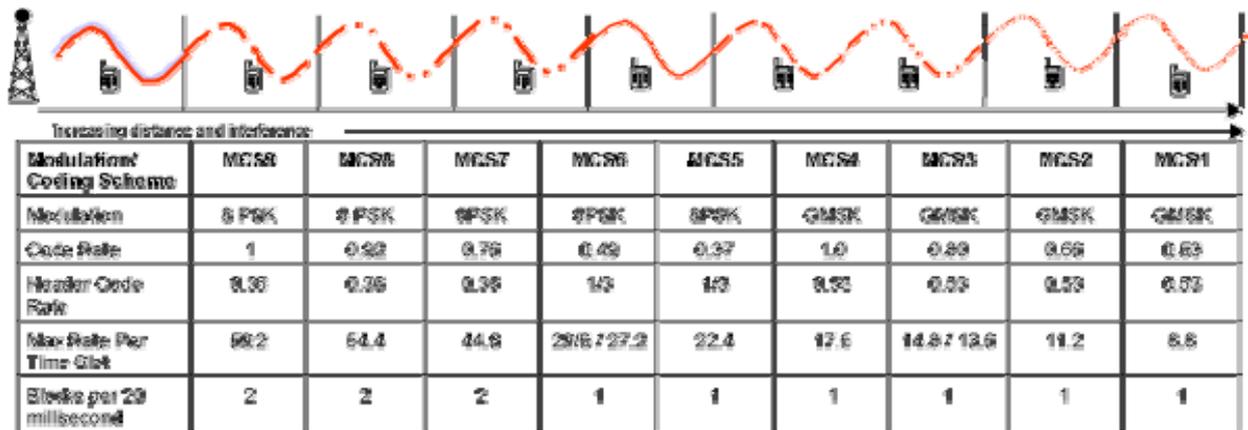


Figure 3: Modulation Transfer Function

However this measure does not take into account the limitations of human vision. We need a minimum contrast for an image to become distinguishable and this is measured using a Contrast Threshold Function.

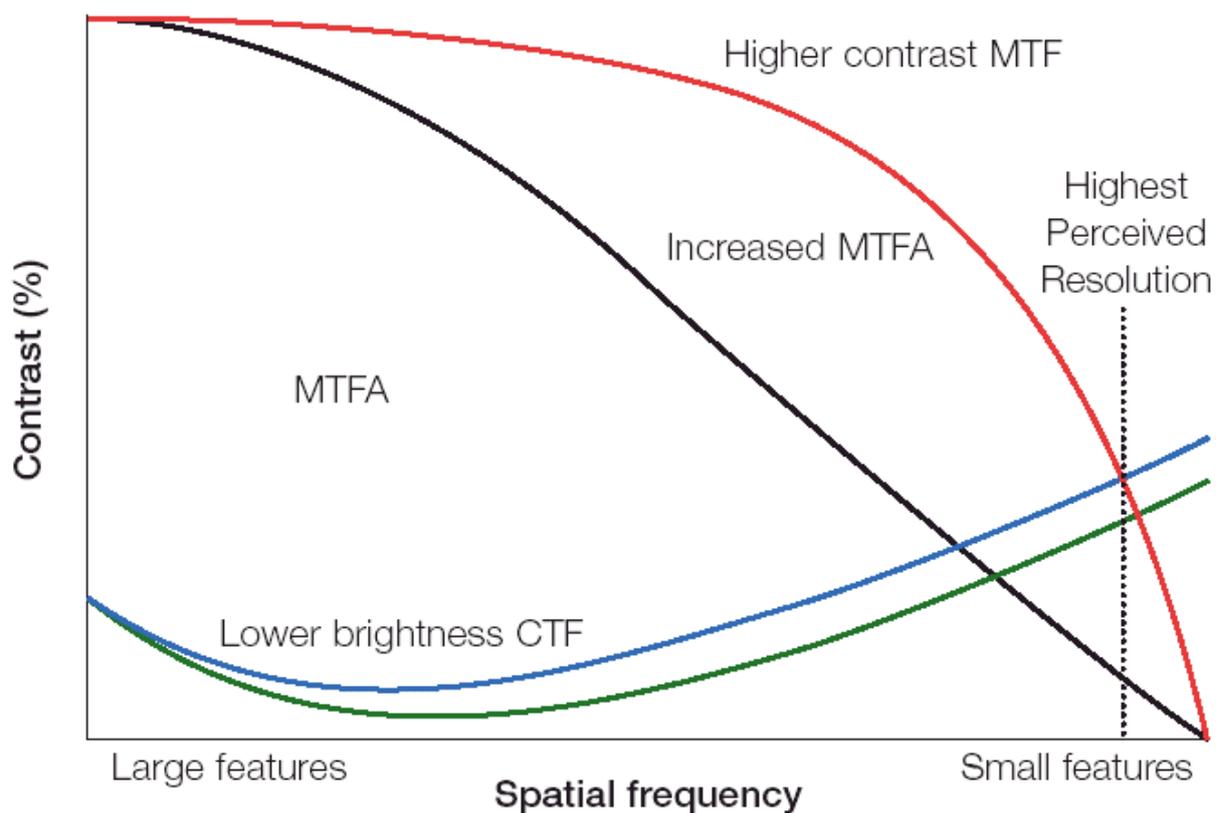


Figure 4: Contrast Threshold Function and MTFA

Adding together the modulation transfer function of the imaging system and the contrast threshold function of human vision yields a crossover point which determines the highest perceptible resolution (shown in Figure 4).

Increasing the brightness does not make much difference to the maximum perceived

resolution whereas increasing the contrast ratio significantly increases the amount of image content conveyed to the viewer. This is shown in Figure 5.

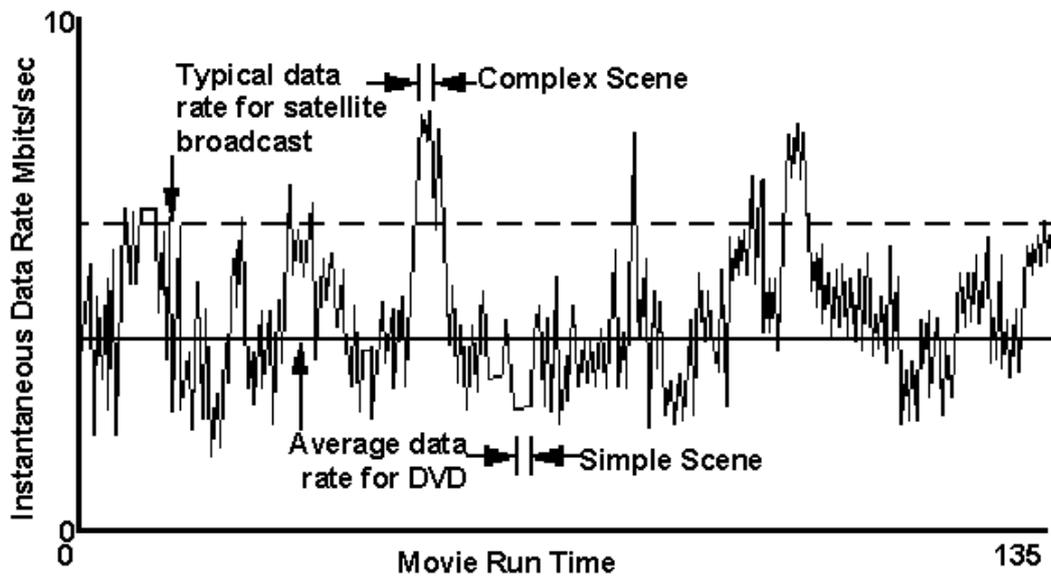


Figure 5: Effect of Contrast on MTFA

The above examples are from Clarity Visual Systems www.clarityvisual.com and are applied to benchmarking quality in larger display systems but the same rule sets apply to smaller micro-displays. These performance metrics are important because they define the real world experience of the user – a better quality image makes for a more immersive experience, the more immersive the experience, the longer the session will last.

So having sorted out display quality, let's just revisit the source encoder.

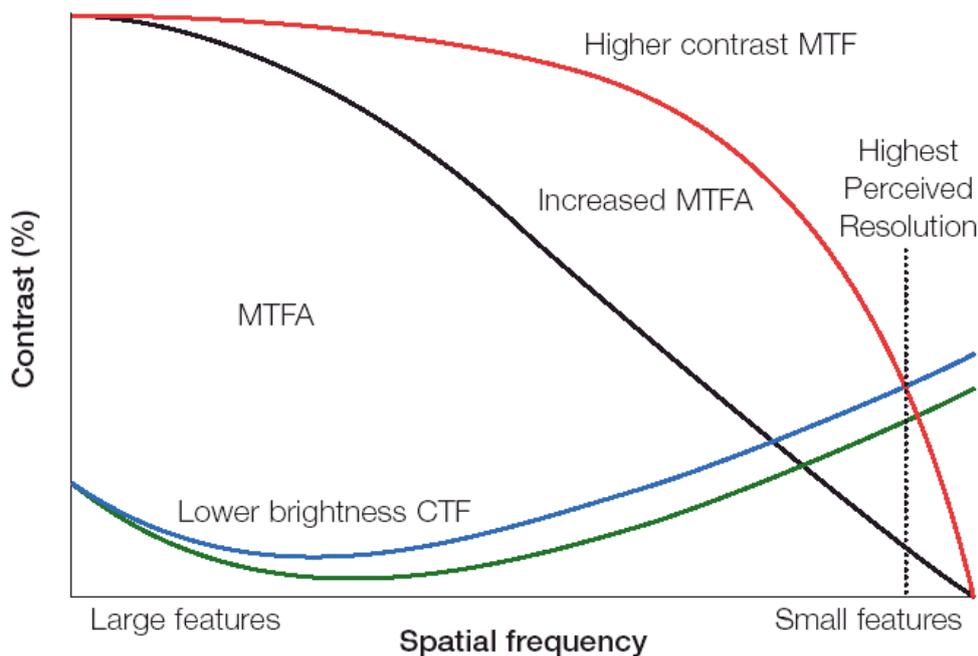


Figure 5: Use of M-PEG7 Descriptor in Scene Classification

Figure 6 shows a traffic trace from an M-PEG2 video file with minimum and maximum data rates ranging from 2 to 7 Mbps. Note the 'complex scene' bit rate excursion as the entropy of the image increases – flashbulbs going off at a Press Conference or a crowd standing up at a football match could be possible causes.

There are three choices. Either to band limit the excursion which will result in variable quality for the user, or to buffer the bit stream (which will result in variable delay) or to track the excursion by dedicating some additional instantaneous bandwidth to the event. Of these, option three provides the best user experience but requires a flexible radio layer (flexible layer one) and flexible access to available network bandwidth.

Which illustrates our overall point that multi-media quality is a function of getting the source coding right and matching source coding efficiently to a radio layer and network that can deliver continuity and consistency. A lack of continuity and consistency will conspire to shorten session length. Quality has to be qualified in terms of the end to end channel which includes the source encoder, channel coder, the radio layer, the network and (assuming this is a mobile to mobile session) the decoder and audio and display capabilities in the receiver. If buffering is needed, the memory at various points in the channel also needs to be considered.

In voice and video, particularly conversational voice and video, dropped call rates or more specifically, dropped **session** rates need to be considered. Here, protocols such as 3G 324M (specifically the addition of the H.245 signalling channel) should help.

To summarise – an important objective in 3G networks is to create the conditions whereby session lengths can and will increase over time. Quality and consistency are important pre-conditions that need to be met before this happens. Quality is a composite of power efficient source coding, channel coding, a flexible layer one, a deterministic network and good audio and video capabilities in the receiver. Consistency is dependent on trying to avoid too much band limiting of the channel and careful management of session maintenance protocols (the need to avoid dropped sessions).

About RTT Technology Topics

RTT Technology Topics reflect areas of research that we are presently working on.

We aim to introduce new terminology and new ideas to clarify present and future technology and business issues.

Do pass these Technology Topics on to your colleagues, encourage them to join our [Push List](#) and respond with comments.

Contact RTT

[RTT](#), the [Shosteck Group](#) and [The Mobile World](#) are presently working on a number of research and forecasting projects in the cellular, two way radio, satellite and broadcasting industry.

If you would like more information on this work then please contact

geoff@rttonline.com

00 44 208 744 3163