

Codecs and Capability

Sean Cubitt
University of Melbourne

for Video Vortex, Institute of Network Cultures
<marijevanneck@networkcultures.org>

scubitt@unimelb.edu.au
Media and Communications Program
Faculty of Arts
Room 127 John Medley East
The University of Melbourne
Parkville VIC 3010
Australia

ABSTRACT : This paper addresses the specific technical qualities of Flash Video and its underlying codec. It suggests that the current divergence of high and low resolution media is indicative of a change in the nature of mediated social structures. It suggests that low-resolution media may have a greater utopian dimension, albeit still a very depleted one, than high-definition media.

BIO: Sean Cubitt is Director of the Program in Media and Communications at the University of Melbourne and Honorary Professor of the University of Dundee. His publications include *Timeshift: On Video Culture*, *Videography: Video Media as Art and Culture*, *Digital Aesthetics*, *Simulation and Social Theory*, *The Cinema Effect* and *EcoMedia*. He is the series editor for Leonardo Books at MIT Press. His current research is on public screens and the transformation of public space; and on genealogies of digital light.

Codecs and Capability

What makes a YouTube video good? Maybe it is the political tenor, or perhaps you like the ethics. Perhaps it looks nice. Or it's funny. Perhaps a YouTube video is good when it reaches a lot of people. But the great thing about the internet is that it allows every minor interest, every academic specialism, every rare and refined hobby a place, so the numbers really don't matter in the same way as the old media. Everyone has had that lovely serendipitous moment when you find exactly the right piece of data, exactly the right image, on the site dedicated to collecting photos of old street lights or the history of dye-transfer techniques. Popularity isn't in question. Looking nice, being funny, politics, even ethics are pretty much personal opinion in the globally connected, rapid and fragmentary culture of the internet post-2002. It may be better to ask what makes a YouTube video bad. Then we have some answers. Slow download. Too much fuzz in the image or the soundtrack. Stutter. Technical qualities are what make a bad video. Things that go wrong, like using a pine green title on a black background (<http://www.youtube.com/watch?v=inc76pe4Yf8>). There is always a workaround, an optimal way of using the tool that's available, but the tool has to be available, and a network tool has to be as nearly universally available as it can if it is to permit the serendipitous discovery of the lone like soul to yours among the billion pages.

As Adrian Mackenzie (2008: 54) notes, the way technical aspects of video formats operate 'pulls apart and reorganizes moving images goes further than simply transporting images ... Like so much software it institutes a relational ordering that articulates realities together that previously lay further apart'. In what follows, I hope to demonstrate how this occurs on YouTube. YouTube video formats run on the .flv format, a proprietary web format owned by Adobe, who in 2005 amalgamated with the original owners, Macromedia. Although users can upload in most of the popular low-resolution formats (Windows Media Player, QuickTime and Real Player), files are automatically converted to the .flv format (and the file extension automatically stripped from the file name). Audio format is mp3 and mono (although manual conversion of downloaded files can release latent stereo). The .flv format is supported by the H.263 codec (short for coding-decoding or compression-decompression; protocols used to squeeze audiovisual data for transmission and unsqueeze them for playback). Though H.263 was developed by the Sorenson Media company, it is largely seen as a tweak of the H.264 standard codec developed conjointly by the ITU and ISO in the suite of tools devised around MPEG-4, and is presented by the ITU as a 2005 development from H.261.

The organisations involved are key to the functioning of the global telecommunications system. ITU is the International Telecommunications Union, established in 1865 as the International Telegraph Union, and widely regarded as the oldest intergovernmental treaty body still functioning today. Among its tasks are ensuring the interoperability of telecom systems, pricing régimes for international calling, and infrastructures for audiovisual services, including moving image encoding, in this instance for low bit-rate communications. The ISO is the International Organisation for Standardisation, who have responsibility for ensuring, where possible, that the nomenclature and technical operation of engineering occurs in compatible ways across the world. ISO numbers are attached to products from screws to photographic emulsions,. ISO is a partner with the IEC, the International Electrotechnical Organisation which has similar jurisdiction over electrical and electronic activities, through the Joint Technical Committee which covers the whole range of digital media, from hardware and software to networks and interfaces. Unlike the ITU, both IEC and ISO are non-governmental organisations. But in recent decades, the ITU has increasingly welcomed corporations as influential participants in discussions, especially as a realist reaction to the privatisation of national telecommunications systems under neo-liberal pressure since the 1980s (MacLean 2003). The NGOs, meanwhile, have a long history of cooperation with governments and, for similar reasons and in a similar timeframe to the ITU, with corporations as well. The Motion Picture Experts Group is a sub-committee of the IEC/ISO Joint Technical Committee, but publication of the H.263 codec specification is undertaken by the ITU.

There is no internet without the standardisation of internet protocols; and there is no exchange of moving pictures without standardisation of the codecs on which the various proprietary players can function. Optimising for low bit-rate is in one perspective an entirely proper and democratising principle. It is all too easy to envisage codecs restricted to high bandwidth clients, excluding the majority of the world from access to content. The way in which such standards are arrived at should bring to mind the Internet Engineering Task Force slogan, 'We don't believe in presidents, kings, or voting; we believe in rough consensus and running code'. While undoubtedly commercial and political pressures are brought to bear, the central activity of these organisations is a rational debate among rational individuals, setting themselves goals, debating means in a shared language, and producing common norms (Russell 2006). It is in this sense a model version of the Habermasian public sphere (Froomkin 2003). I am less concerned here with the critique of this consensus as an exclusive and elite mode of operation which excludes the proper influence of civil society or elected representatives, and more with the resulting capabilities of the codec and the software it supports. In the Habermasian version, public sphere discussion should lead to an optimal result, in much the same way that post-Hayek economics asserts that the rationality of the

market will always provide the best possible products. But the .flv format and the H.263 codec do not provide the best possible image. Like VHS, and so many other victors of previous format wars, they are only good-enough.

The H.263 codec combines a restricted colour palette with a vector-based predictive function to minimise the need to repeat information concerning pixels whose colour doesn't change. The colour space used is YCbCr, which codes for luminance (Y) and two chroma channels (C), blue and red, on the principle that the panchromatic Y channel captures the necessary detail, while the absence of green (as used in almost all colour film, television and high-end storage media like DVD and Blu-Ray) minimises redundancy because the green channel overlaps with both red and blue, especially in the yellow segment of the spectrum. This is a variant of 8-bit colour graphics, which allows a range of 256 colours (actually slightly reduced from that because some of the hues are indistinguishable to the human eye; it is unclear whether .flv has compensated for the old problem produced when 8-bit colour for web applications tried to install its own palette on the local desktop, giving rise to various colour conflicts).

The layout of the image is compiled according to 'a hierarchical structure with four primary layers. From top to bottom the layers are: Picture; Group of Blocks, or slice, or video picture segment; Macroblock; Block' (ITU 2005:13). In default mode, a block is a section of 4x4 pixels. Using the hierarchy of macroblock (16x16 lines of Y and 8x8 lines of Cr and Cb) allows the codec to treat pixels not only as units but as averaged areas of colour and luminance, and the group of blocks, comprising 16 lines, also allows for such commonly observed features as the positioning of sky or grass at top and bottom of images respectively (a slice is the same size as a Group of Blocks but its shape can be varied). The picture comprises 16x16 Groups of Blocks. These Groups of Blocks are also the units in which the vector predictions take place: 'Motion vectors are restricted such that all pixels referenced by them are within the coded picture area', (ITU 2005: 12) (though virtual pixels off-screen are allowable in enhanced, more bandwidth-hungry versions than the format used on YouTube). Vectors predict movement based on sequence from an initial image. Encoding artefacts are increasingly likely in hand-held sequences when the prediction system is more likely to predict wrongly or as the Flash Video white paper has it, 'If your camera is not steady, most of the image moves, causing a high percentage of pixels in the video to change from frame to frame. A steady camera reduces the number of pixels that change from frame to frame, giving you better quality at higher compression rates (lower data rates' (Macromedia 2004: 11). Other artefacts are likely as a result of mismatches between this codec and others feeding into it, such as the PAL 625-line television standard in use in most of the world and the 525-line NTSC standard used in the

USA. Typical would be the mismatch of sound and image due to the different clock-cycles in incoming data streams and the use of the H.263 and mp3 codecs for video and audio respectively. The .flv format includes metadata for synchronising the two processes, but shifting between already lossy codecs invariably results in long-term degradation of the datastream.

Flash video (.flv) marks what the H.263 codec refers to as 'initial frames' by the animation term 'keyframes', uncompressed frames which are used as a reference for filling in compressed frames that come between them. In Flash vector animation, a similar process is used to do the 'in-betweening', extrapolating from first and last frames the action needed to move from one to the other. This information is encoded not as full-frame animation but as an instruction set, which requires far fewer lines of code. As the Flash Video white paper notes, 'A lower keyframe rate (such as one keyframe every six seconds) will result in a softer or blurrier image but reduces the bandwidth demand' (Macromedia 2004: 13). The format uses progressive scanning of the image rather than interlacing, the process common to television and the 1080i High Definition television (HDTV) codecs which repeats each image in two separate fields, the first scanning odd, the second even lines, which removes the flicker perceived in bright cathode ray tube (CRT) monitors. Progressive scanning is native to most computer screens and digital light programming (DLP) projectors. Flash also permits manipulation of the frame rate, set at 25 frames per second for terrestrial TV transmission, as low as 10 fps. Bitrates can also be selected with an eye to the end-user, with options ranging from 56kbps to 2 Mbps, and can be encoded using a variable bit-rate which requires more time to encode but speeds decoding. Such tools lower the quality of the image in exchange for reductions in file size and processing power required to decompress them at the receiving end. Delivery is speeded up by streaming, in which the whole file is never delivered or cached by the end user, and only the viewed parts of the video arrive at the end-user screen. Remaining bandwidth can then be used for interaction, and for the additional services which streaming server software offers for traffic management and security.

In what sense is this better or worse than any other codec? The emerging standard for HDTV, Blu-Ray and high-definition DVD are 1080i and 1080p (interlaced and progressive variants of the 1080 line standard) with frame rates limited by bandwidth to between 24-30 fps, although 60 fps is available (for example on Apple's QuickTime HD Trailers site). They use full RGB colour rather than the restricted palette of YCbCr. As a result, they display at near-35mm film quality on screens which can be over a metre (measured on the diagonal) for domestic use, and 20 metres or more for external public projections and datascreens. Take up on HDTV has been slow due to pricing, worries over reliability and longevity of screens, and format wars, but is picking up rapidly in the

late 2000s. According to ongoing research by Jeff Bird at the University of Melbourne, owners cite anxiety about street-crime and terrorism, the expense of cinema-going, and the increased control over time as important reasons for moving to HDTV. There is some evidence that owners gain status from their screens, but that factor is bound to diminish as prices drop to bring them in range of working-class consumers. At the same time, the popularity of YouTube and the rise of third-screen audiovisual consumption on iPods, mobiles, PDAs and handheld game consoles like PSP suggest that resolution, colour accuracy, refresh rates and size are not the sole drivers of demand in emergent video markets.

A vector is a contradictory creature. It is an algebraic statement of the curvature of a line, a static description of movement. In its use in H.263, it is a statistically normative description of the succession of frames, frames that are no longer composed of fields, as in the old CRT interlace, but which are constrained by the exclusion, in default mode, of any offscreen determinant. The predictive vector, then, is analogous to both the political management of populations and the economic efficiency of distribution, at the expense of truth either to the illumination of the recorded scene or the impression of the perceiving eye. De-realising and disembodiment, the low-res predictive vector codec conforms vision (and such communication tools as Skype which also use it) to current social conditions, while at the same time enabling them. Relying on the human ability to extrapolate likelihood from what is given to sight, H.263 demands a labour of interpretation from its end-users, an active engagement working on the ripples and blocks. Here too it is contradictory: on the one hand undermining the supposed passivity of audiences, on the other increasing the unpaid labour of attention through which audiences provide the bases for the advertising economy, and now the unpaid production of content. In H.263, the unpredictable movement of the ideal vector as a trajectory towards the unknowable future (Cubitt 2004) is tamed, brought into line with the practice of actuarial risk-management and business and investment planning. Crucial to fast-delivery .flv files is the economical direction of action: the less change there is, the lower the bandwidth demands. YouTube's technical specification militates against change. To the extent that it limits us to the unchanging network of normal communication, and excludes us from what is beyond the frame and from change, it is an art without hope.

The March 2008 edition of *American Cinematographer* carries a report on Bill Viola's HD shoot of a video installation for the 2007 Venice Biennale (Stasukevich 2008). Video art practice is no longer condemned to the U-matic reel-to-reel aesthetic of Wegman's *Selected Works*, nor bound to accommodate itself to the affordances of low-resolution kit. High-definition lends itself to the ritual

mysticism of Viola's practice, and his reach back to Baroque roots for a language of spiritual experience. It is proper to the rare, the site-specific, what opposes itself to the ordinary running of the world. Viola opposes the normativity of low-resolution with the exceptionalism of high-definition, which removes the image stream from the banality of the quotidian, at its best seeking to force open a chink in the familiar to let the light of something absolutely other flicker in the interstices. We might call it the sublime.

In a 1936 letter to Walter Benjamin, Adorno described high and low culture as 'the two torn halves of an integral freedom to which, however, they do not add up' (Adorno 1977). We might today speak of the gap between high and low resolution in similar terms, with the startling exception that the high and low no longer have class correlates. In the monstrous reorganisation of society as population management and of knowledge as data flow, we face an impossible choice. We may succumb to slack-jawed immersion in spectacle, from Vegas to the Sydney Olympics., or accede to the fragmented and ephemeral world of connectivity. The oscillation between hi-res and lo-res paths is the impossible choice between the sublime and despair. Despair and the sublime are the two torn halves of a single oppression: the removal of the object of contemplation from the realm of what can be communicated. In the spectacle of the immersive sublime, what we gaze upon is other, wordless and worldless, beyond history or debate. In the connective despair, what we seek to communicate, the very content of cellular networks, is ourselves, but that is the one thing that cannot be communicated in a world of hyperindividuation. Choose, we seem to be told, between the cinematic spectacle of 9/11 or the connective mobile images from Abu Ghraib. Choose between the unspeakable, sublime icon created by Islamists whose faith does not allow icons. Or pick the degraded and degrading mobile images from Abu Ghraib, where connectivity becomes an extension of the humiliation which is the goal of torturers. The binarism of hi-res and lo-res takes us to the sick heart of the contemporary world.

Is this all there is? The challenge once posed by barred access to high-resolution video has migrated to the challenge of bandwidth management, and as with Wegman's talking torso, the new contenders are using low-end solutions to make high-concept works. And there are other opportunities, migrating between the two worlds. In a paper presented in Seoul in 2007, Ross Harley (2007) argues that YouTube is potentially a model for video art archives. Low-resolution versions online produce demand for high-resolution distribution copies, just as teaser trailers do for the latest Hollywood soaps and sitcoms. And although there are problems to overcome in copyrights, control and governance, the principle of distributed production and exhibition are in

tune with the origins of video art. Several clips from Jean-Luc Godard's *Histoire(s) du cinéma* are available on YouTube. Otherwise, since their sporadic transmission in some European territories in the 1990s, and apart from some gallery and cinema showings, the only consumer format release was in the form of a CD set of the soundtrack, although recently Gaumont have managed to release a 4 DVD region 0 set of the programmes. The long delay arose because of the problems of rights clearances. Godard is an old man, and frankly uninterested in spending his remaining years chasing copyright owners. Thankfully Gaumont have done the work at last. In the interim, the memory of the series has been kept alive with lo-res transfers to .flv. The results are more than usually murky. Godard's palette is rich, his compositions densely layered, the soundtrack is complex enough to have been released as a free-standing CD set. The transfers often include edge coding and other artefacts, and the blocky pixellation is as far removed from high definition as one might expect. As Alexander Horwath observed, however, 'exist in diverse image, sound and text variants, but Godard's own way of working, in the *Histoire(s)* and elsewhere, is wholly contrary to totality and completeness.. Thus it seems quite logical that the material results of his work on the history and histories of cinema can never really be completely "boxed", distributed or owned - except in the rare, fleeting moment of projection (in which the constant sensuous overload makes it impossible to "capture" the whole anyway)' (Horwath 1998). YouTube is just another efflorescence of a perpetually (and perceptually) incomplete project.

Yet something else does shine through the murk: a desire, a sentiment of what is missing, that is not available in the high density images of spectacular high definition. YouTube lives on the precarious labour of its contributors. It offers them the old lie – at least five years old – that you are never alone in the network, and the network is where you can express and be a self. But Harley is right in this: that it is exactly because of their failings that YouTube and other lo-res media are less frightening than the alternative. They lie, but they are permeable. The arguments of the generation of '68 return: the dominant cinema constantly produces films which, despite their ideological project, cannot help addressing the contradictions in the dominant (see Comolli and Narboni 1977). Today, that role is being undertaken by software. The despair of the networked soul is still capable of a grainy, lossy, lo-res vision of hope. What makes a YouTube video good is the same as what makes it bad.

REFERENCES

Adorno, Theodor W (1977), 'Letters to Walter Benjamin', 18 March 1936 in Ernst Bloch et al, Aesthetics and Politics, New Left Books, London.

- Comolli, Jean-Louis and Pierre Narboni (1977), 'Cinema/ Ideology/ Criticism' (1) and (2), trans Susan Bennett, in John Ellis (ed), Screen Reader 1: Cinema/Ideology/Politics, SEFT, London, 2-11 and 36-46.
- Cubitt, Sean (2004), The Cinema Effect, MIT Press, Cambridge MA.
- Froomkin, A Michael (2003), 'Habermas@Discourse.net: Toward a Critical Theory of Cyberspace' in Harvard Law Review, Volume 116, January, 751-873.
- Harley, Ross (2007), 'Totally Busted: Do We Need A YouTube For Video Art?', paper presented at the Video Art Archive Network Forum, Gallery Loop, Yonsei University, Seoul, 8-9 November.
- Horwath, Alexander (1998), 'The Man With The Magnétoscope - Jean-Luc Godard's monumental *Histoire(s) du cinéma* as SoundImageTextBook', Senses of Cinema, trans Aileen Dierig, http://www.sensesofcinema.com/contents/01/15/godard_horwath.html, retrieved 22 March 2008.
- ITU (2005), ITU-T Recommendation H.263 Video coding for low bit rate communication, ITU-T: SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS - Infrastructure of audiovisual services – Coding of moving video, International Telecommunication Union, Geneva, <http://www.itu.int/rec/T-REC-H.263/en>, retrieved 4 March 2008.
- Mackenzi, Adrian (2008), 'Codecs' in Matthew Fuller, Software Studies: A Lexicon, MIT Press, Cambridge MA, 48-55.
- MacLean, Don (2003), 'The Quest for Inclusive Governance of Global ICTs: Lessons from the ITU in the Limits of National Sovereignty' in Information Technologies and International Development, v.1 n.1, 1-18.
- Macromedia (2004), Flash Video Primer, Adobe, www.adobe.com/devnet/flash/articles/flash_flv.pdf, retrieved 11 March 2008.
- Russell, Andrew L (2006), "'Rough Consensus and Running Code and the Internet-OSI Standards War' in IEEE Annals of the History of Computing, July-September 2006, 48-61.
- Stasukevich, Iain (2008), 'Short Takes: Crossing Over in Bill Viola's *Ocean Without a Shore*', American Cinematographer, 89(3), March, 10-14.