A Study of Enhanced PAL Television, Including a New Coding Technique, and New Methods for Analysing and Appraising Coding Techniques

Kevin Taylor

A thesis submitted in fulfilment of the requirements for the degree of PhD in Engineering in Electrical and Electronic Engineering at the University of Canterbury, Christchurch, New Zealand.

October 1997
Abstract

The introduction of colour television in the 1950's was an impressive achievement producing a full colour picture on new receivers whilst maintaining substantially complete compatibility with monochrome receivers. This was achieved by placing two colour difference signals on top of the existing monochrome (luminance) signal. The disadvantage of the technique was the lack of full separability of the signals in the receiver, resulting in signal cross-effects and some loss of picture quality. Although acceptable at the time of introduction improving source quality (due to improved cameras, digital sources, and improved recording) coupled with increasing quality expectations has caused these artefacts to become more noticeable and less acceptable. Consequently considerable work in the last twenty five years has concentrated on finding improved encoding and decoding techniques for creating compatible PAL signals. Unfortunately until very recently with the introduction of PALplus no satisfactory solution had been agreed upon. Work in the field has been made difficult through the complexities of the multi-dimensional PAL signal, the lack of common framework, and the difficulty of assessment and comparison.

This thesis provides new work in several areas in the field of enhanced PAL coding including: i) the development of a complete multi-dimensional representation for a television coding system, ii) the description of a new technique for determining the spatio-temporal characteristics of coding systems, iii) the description and application of a new approach to objective quality assessment, and iv) the detailed study of a new improved PAL coding technique.

This thesis also i) provides comprehensive background on the multi-dimensional interpretation of the PAL signal and television coding techniques, ii) develops a model for the human visual system chrominance threshold response, iii) places all current enhanced PAL coding techniques into a common theoretical framework, and iv) provides a detailed mathematical analysis of WC-PAL.

Work on all these areas was greatly assisted through the development by the author of a multi-dimensional graphical interface (PHIGSdraw) and image processing system (TVPROC).
The years covering the time spent on the work in this thesis have been challenging and at times very difficult. Several of the years were the most difficult of my life so far, and I hope they remain so. Early work in the field covered by this thesis led to the development of very bad OOS (RSI). During 1½ years on the sickness benefit I did not know if I would recover long term and to be able to continue my work, or indeed my career. I had many problems to face up to at this time, and now having overcome these and finally submitting this thesis I realise how much this time has changed me.

My greatest thanks must go to my parents for their unquestioning support during this time, especially when I returned to live with them for nearly two years. I received much support also from my original supervisor Dr. Chris Carey-Smith who was always encouraging and enthusiastic about my work.

I must acknowledge also Ian Goodwin of Broadcast Communications Ltd whose long term interest in my work has developed into a strong friendship. Ian's support and promotion of my work has continued long after BCL withdrew their official support for the research.

Others have worked with me in this field, including Antony Dean and Peter Casey. It was great to have these people to share ideas with. With the help of Peter Casey in particular I was able to continue for some time to be involved at an intellectual level in the research, despite being unable to actively participate to the degree I would have liked.

Lastly I must acknowledge my final supervisor Professor Des Taylor. He deserves credit for taking over the supervision of a Ph.D. in a field not his own, and for having the patience to wait as my original finish date kept shifting. Des Taylor's professional guidance has greatly helped me in producing a better thesis with a more rigorous mathematical basis.
# Contents

## CHAPTER 1 INTRODUCTION
- 1.1 The development of colour television: PAL, NTSC and SECAM ........................................ 1
- 1.2 The search for improvements to conventional PAL and NTSC ........................................... 2
- 1.3 PALplus and the current enhanced television situation ....................................................... 4
  - 1.3.1 The development of multi-dimensional television theory ......................................... 5
- 1.4 The assessment of enhanced coding techniques ............................................................... 5
- 1.5 Overview of Thesis .......................................................................................................... 6
  - 1.5.1 Theory of television and the human visual system .................................................. 6
  - 1.5.2 Methods for analysing and assessing television coding techniques ...................... 6
  - 1.5.3 Enhanced coding techniques .................................................................................... 7
  - 1.5.4 New simulation and display tools ............................................................................ 7
- 1.6 Publications based upon the work in this thesis ............................................................. 7

## CHAPTER 2 TELEVISION AND THE HUMAN VISUAL SYSTEM
- 2.1 The structure of the eye ................................................................................................. 9
- 2.2 Visual acuity (spatial) .................................................................................................. 10
  - 2.2.1 Spatial acuity for colour ......................................................................................... 11
  - 2.2.2 Spatial acuity and television .................................................................................. 11
  - 2.2.3 Acuity and motion .................................................................................................. 13
  - 2.2.4 Television's temporal bandwidth requirements for moving images ...................... 13
- 2.3 Temporal acuity .......................................................................................................... 14
  - 2.3.1 Temporal acuity for colour ................................................................................... 15
  - 2.3.2 Television's temporal resolution requirements ....................................................... 15
- 2.4 The response of the eye to illumination ........................................................................ 15
  - 2.4.1 Adaptation ............................................................................................................ 16
  - 2.4.2 Illumination and television ..................................................................................... 17
- 2.5 The response of the eye to colour ............................................................................... 17
  - 2.5.1 Colour and television ............................................................................................ 19
- 2.6 The achromatic spatio-temporal threshold of the eye .................................................... 19
  - 2.6.1 The threshold functions and television ................................................................... 21
  - 2.6.2 The supra-threshold response ............................................................................. 22
  - 2.6.3 The supra-threshold response and television ........................................................ 22
- 2.7 The chromatic spatio-temporal threshold of the eye ..................................................... 22
  - 2.7.1 Chromatic thresholds and television ..................................................................... 23
- 2.8 Modeling the eye's threshold characteristics ................................................................. 24
  - 2.8.1 Modification of model for non-isotropic spatial response .................................. 25
  - 2.8.2 A chromatic model ............................................................................................... 26
- 2.9 Picture quality evaluation ............................................................................................ 27
  - 2.9.1 Subjective assessment .......................................................................................... 28
  - 2.9.2 Objective assessment ........................................................................................... 29
- 2.10 Summary .................................................................................................................... 30

## CHAPTER 3 BACKGROUND
- 3.1 The Spectrum of the Scanned Signal ............................................................................ 31
  - 3.1.1 The spectrum of the image .................................................................................. 32
  - 3.1.2 Scanning the Image ............................................................................................. 33
  - 3.1.3 The spectrum of a scanned signal ....................................................................... 35
  - 3.1.4 The spectrum of a scanned signal in two dimensions ........................................ 36
  - 3.1.5 The spectrum of a scanned signal in three dimensions ....................................... 37
  - 3.1.6 The Nyquist limit ................................................................................................. 38
  - 3.1.7 The interpretation of movement in three dimensional spectral space .............. 39
5.1 Television as a Linear MIMO system................................................................. 93
5.1.1 Television as a periodically time varying system - a one-dimensional analysis 94
5.1.2 Identifying the system.................................................................................. 95
5.2 A simplified representation of the system characteristics......................... 97
5.2.1 Example of use of the defined system representation............................... 99
5.2.2 A complete representation of the television system................................. 101
5.3 Determining the characteristics for a given system....................................... 102
5.3.1 The frequency range of interest ................................................................. 103
5.3.2 Discrete representation of \( H(f, u) \) .............................................................. 103
5.3.3 Identification of \( H(f, u) \) with fewer input tests........................................ 103
5.3.4 Piecewise spectral analysis........................................................................ 104
5.3.5 Implementation of piecewise spectral analysis......................................... 107
5.3.6 The three-dimensional version ................................................................ 108
5.4 Results ........................................................................................................... 109
5.5 Summary ....................................................................................................... 116

CHAPTER 6 A NEW APPROACH TO OBJECTIVE ASSESSMENT..................... 117
6.1 Using the spatio-temporal characteristics to obtain a quality measure ....... 118
6.2 The effects of spatio-temporal energy distribution....................................... 118
6.3 The relative response of the HVS to spatio-temporal frequency............... 119
6.3.1 The response for the unwanted signals..................................................... 119
6.3.2 The response for the unwanted signals..................................................... 121
6.4 Calculating an objective quality measure.................................................... 121
6.4.1 Calculating the figure-of-merit for a wanted signal............................... 122
6.4.2 Calculating the figure-of-merit for the unwanted signal......................... 122
6.5 Normalising the figures-of-merit.................................................................. 123
6.6 Results using objective assessment.............................................................. 124
6.7 Comparison of objective and subjective assessment................................. 126
6.7.1 Choosing a data set................................................................................... 127
6.7.2 Performing objective assessment ............................................................. 127
6.7.3 Comparing the results ............................................................................. 127
6.7.4 A mathematical approach to determining the correlation....................... 129
6.8 Further development of this objective assessment method.......................... 131
6.9 Summary ..................................................................................................... 131

CHAPTER 7 K-PAL - A NEW ENHANCED PAL CODING TECHNIQUE............. 133
7.1 Placing luminance and chrominance in quadrature..................................... 133
7.1.1 Pre-filtering of U and V ........................................................................... 134
7.1.2 Modification of luminance ...................................................................... 134
7.1.3 Pre-filtering the luminance .................................................................... 136
7.1.4 A phase-segregated signal ...................................................................... 136
7.2 Separating Y and C in a decoder................................................................. 136
7.2.1 Decoding Chrominance .......................................................................... 136
7.2.2 Decoding Luminance ............................................................................. 137
7.3 Forming a PAL compatible signal ............................................................... 138
7.3.1 WC-PAL ................................................................................................. 138
7.3.2 K-PAL ..................................................................................................... 140
7.4 Compatible Reception .................................................................................. 141
7.4.1 Receiving the WC-PAL Signal ................................................................. 142
7.4.2 Receiving the K-PAL Signal .................................................................... 144
7.5 Enhanced Reception .................................................................................... 145
7.5.1 WC-PAL Decoding ................................................................................. 145
7.5.2 K-PAL Decoding .................................................................................... 147
7.6 Designing the filters..................................................................................... 149
7.6.1 The K-PAL band segregating filter ......................................................... 149
7.6.2 Designing the pre-filters \( R_{e1} F_{U} \) and \( R_{f} \) ........................................... 151
7.6.3 Effects of horizontal filters ...................................................................... 152
7.6.4 Rearranging the block diagrams ............................................................ 152
7.6.5 Minimising horizontal filtering ............................................................... 153
7.6.6 The effects of upper sideband loss .......................................................... 156
7.6.7 Designing $L_1$, $L_2$, and $H_1$ ................................................................. 157
7.6.8 Asymmetrical chrominance filtering ....................................................... 160
7.6.9 Enhanced horizontal resolution ............................................................... 162
7.7 Implementation details .................................................................................. 163
7.7.1 Over-range luminance ............................................................................ 163
7.8 Comparison of K-PAL and WC-PAL implementations ............................... 166
7.8.1 Choosing an implementation for $R_c$, $F_U$, and $R_T$ .......................... 167
7.8.2 The Luminance Response ........................................................................ 169
7.8.3 Simulation Results .................................................................................. 172
7.8.4 Luminance Aliasing ................................................................................. 174
7.8.5 Chrominance ......................................................................................... 176
7.8.6 Cross-chrominance ................................................................................. 178
7.8.7 Cross-luminance ..................................................................................... 184
7.8.8 Inter-chrominance ................................................................................. 185
7.8.9 Comparison of Results .......................................................................... 186
7.9 Summary .................................................................................................... 187

CHAPTER 8 : CONCLUSIONS ........................................................................ 189
8.1 The Human Visual System .......................................................................... 189
8.2 The Theory of PAL Signal and Coding Analysis ........................................ 189
8.3 Calculating the 3D spatio-temporal characteristics of a coding system - a new way of understanding PAL coding ................................................................. 190
8.4 A new approach to objective assessment .................................................. 190
8.5 A new improved PAL coding technique ...................................................... 191
8.6 Summary .................................................................................................... 191

APPENDIX A  THE COLOUR TELEVISION SIGNAL ..................................... 193
A.1 The gamma ............................................................................................... 193
A.2 The correction of phase errors .................................................................... 195

APPENDIX B  DESIGNING DISCRETE TIME FILTERS ................................ 197

APPENDIX C  COMB FILTERS USED FOR K-PAL AND WC-PAL .......... 201

APPENDIX D  AN EMPIRICAL CORRELATION MEASUREMENT ............ 205

APPENDIX E  TVPROC - A THREE-DIMENSIONAL IMAGE CODING SIMULATION ENVIRONMENT ............................................................. 207
E.1 Construction vs. simulation ........................................................................ 207
E.1.1 Computer simulation allows for ............................................................... 207
E.2 The development of TVPROC ................................................................. 208
E.2.1 Comparison with other systems ............................................................. 208
E.3 The structure of TVPROC ........................................................................ 209
E.4 The Hierarchical Structure ...................................................................... 210
E.4.1 Layer 1 - The image data structures .................................................... 210
E.4.2 As a summary this complex naming system has several advantages ... 213
E.4.3 Layer 2 - The matrix functions ............................................................ 213
E.4.4 Layer 3 - Image processing libraries ................................................... 213
E.4.5 Layer 4 - Simulation tools ..................................................................... 214
E.4.6 Layer 5 - User configurable shell scripts ............................................. 214
E.4.7 Layer 6 - The graphical interface ........................................................ 215
E.4.8 Packages Used By TVPROC ............................................................... 217
E.5 The functional Structure .......................................................................... 217
E.5.1 Test Image Generation ......................................................................... 218
E.5.2 Creating and Running a Simulation ..................................................... 219
CHAPTER 1
INTRODUCTION

This Chapter will provide some historical and a little theoretical background for the field of television and enhanced PAL coding.

1.1 The development of colour television: PAL, NTSC and SECAM

The current communications revolution began with radio in the beginning of the 20\textsuperscript{th} century. This form of mass communication had its golden era in the 1930s and 40s. At this time, television broadcasting began in several countries, although it was then seen as being only an adjunct to the popular radio broadcasting, a position that was soon to change. The early television systems transmitted only a black and white picture with the moving image being scanned inside a monochrome 'iconoscope' camera and converted into lines and interlaced fields (2 fields forming a picture frame). Thus the image could be transmitted as a radio signal, to be reconstructed into an image by the receiver.

As technology advanced improvements were made with better cameras, and larger, clearer receiver pictures. The most notable improvement was the introduction of colour. The first colour system appeared in America in 1954, using a standard decided upon by the National Television Services Committee (NTSC). By this time television already had an enormous market with monochrome receivers in many homes. The National Television Services Committee faced an important issue, one still faced by the enhanced television systems being developed now, the issue of compatibility. Replacing the monochrome television signal with a non-compatible colour one was not the answer, everyone was not about to throw away their old black and white sets overnight. Simultaneous transmission of a new colour signal was not an economic solution. A monochrome compatible colour signal was the answer. This meant using a system that allowed transmission of the extra colour information in such a way that a normal black and white picture would still appear on monochrome receivers.

The system finally chosen - named NTSC after the committee - used two quadrature modulated colour difference signals placed within the monochrome signal. The colour difference signals, I and Q, were collectively referred to as the chrominance, with the monochrome signal being the luminance, or Y. It was known that the amount of colour information in an image viewed by the human eye is far less important than the luminance information [Wasserman 1978], the reason why monochrome television had worked so well in the first place, NTSC used this property to justify the use of restricted bandwidths for the colour difference signals. The placement of the colour sub-carrier high in the spectrum of the luminance signal took advantage of another property of the eye - its reduced sensitivity to high frequency image detail. Also, as most of the luminance energy is concentrated in the low frequencies this placement minimised interference. In the colour receiver the colour difference signals were demodulated from within the luminance, and together with the existing luminance signal used to reconstruct a colour picture. On monochrome receivers the extra encoded colour signals would appear as a barely perceptible dot pattern on areas containing colour. As such NTSC essentially met most of the criteria as a compatible system.
Other countries were slower in following America's lead, with two different standards, Secam and PAL eventually being adopted. Secam was developed for France and the Soviet bloc with its first broadcast in the 1960s. Originally proposed by Henri de France, Secam avoided quadrature modulation by sending the two colour difference signals line alternately and reconstructing them by use of a line store in the receiver. PAL was a later development still, with first transmissions in Germany and the UK in 1967.

PAL was basically a modified form of NTSC. The changes involved slightly different resolutions and field rates, and the use of different colour difference signals, referred to as \( U \) and \( V \). The important difference was the inversion of one these on alternate lines, the phase-alternation line giving PAL its name. One of the problems with NTSC had been its sensitivity to phase errors in the sub-carrier causing the incorrect decoding of the quadrature colour signals, which produced picture hue changes. (NTSC was jokingly referred to as 'Never The Same Colour', the new PAL is 'Picture Always Lovely'). Due to the phase alternation hue errors on the alternate lines are complementary, and are cancelled by the eye to produce a sensation of the correct hue.

This visual cancelling of hue errors only works well for small phase errors, with an alternating colour pattern being visible at large phase errors. The errors in colour decoding also affect the luminance signal due to distortions introduced by gamma correction (correction for display characteristics). As the eye is more sensitive to changes in luminance this distortion is often more noticeable, and the resulting pattern of lighter and darker lines is known as Hanover bars [Patchett 1967]. This form of PAL is known as PAL-S (Simple PAL).

The introduction of a line store to the receiver, in a similar manner to Secam results in the version of PAL in use today, and is referred to as PAL-D (Delay line PAL). The addition of the line store reduces the effects of phase errors markedly. Instead of relying on the human eye to combine the lines visually, contributions from alternate lines are combined electronically.

Whilst NTSC, Secam and PAL are the three world colour television standards in current use there are many variations within these standards [CCIR Report 624-3 1990], mostly to do with the bandwidths of the luminance and chrominance signals.

1.2 The search for improvements to conventional PAL and NTSC

Whilst all three standards maintain compatibility with monochrome receivers, this is achieved at some cost to the colour picture. The chrominance \( C \) is placed on top of the luminance signal \( Y \), meaning there are 3 signals occupying the same area of spectrum and they are no longer uniquely separable. Their decoding consequently results in the cross-effects: cross-chrominance \( Y \rightarrow C \), inter-chrominance or \( U/V \) crosstalk \( U \leftrightarrow V \) and cross-luminance \( C \rightarrow Y \). Early attempts to eliminate these cross-effects through the use of notch and bandpass filters [Lent 1964] resulted in unacceptable reductions in picture resolution. Instead a partial reduction is obtained by use of luminance notch filters centred on the colour sub-carrier frequency. The trade-off is a reduction in horizontal luminance resolution in the receiver, lower even than that available on the old monochrome systems.

Other perceived defects in the standard television systems are the reduction of vertical resolution due to interlacing and sampling effects, described by the Kell factor [Kell et al 1940], and the presence of temporal aliasing effects such as inter-line and large area flicker [Townsend 1960].

From the 1970s onwards research began on finding ways for fixing these problems. Interlace to progressive scan conversion, and improved source filtering could reduce flicker and increase perceived vertical resolution [Roberts 1983]; special modulation techniques were proposed for regaining lost horizontal resolution [Fukinuki 1984]; and line or field based comb filters were used to improve luminance-chrominance separation [Auty 1977].

The name Clean PAL was coined for an enhanced version of PAL in which the cross-effects of conventional PAL would be eliminated. The methods available for developing such a system are heavily constrained by the requirements for compatibility. Table 1 shows the four paths that need to be considered.
1.2 THE SEARCH FOR IMPROVEMENTS TO CONVENTIONAL PAL AND NTSC

| 1. Standard PAL Coding          | Standard PAL Encoder ➞ Standard PAL Decoder |
| 2. Compatible Reception         | Enhanced Encoder ➞ Standard PAL Decoder    |
| 3. Compatible Transmission      | Standard PAL Encoder ➞ Enhanced Decoder    |
| 4. Full Enhancement             | Enhanced Encoder ➞ Enhanced Decoder        |

**Table 1.1 The Simulation Paths for Assessing an Enhanced PAL Technique**

To be acceptable path 4 must offer significantly improved picture quality compared to the standard path 1. As well paths 2 and 3 must provide picture quality at least as good as path 1, preferably better.

To begin with Clean PAL was tried using two dimensional filters, modifying the signal such that the luminance and chrominance information occupied differing parts of the spectrum - avoiding their interference. This approach is called *band segregation*. A number of methods of achieving this were developed using line based comb filters [Auty 1977]. Field and frame based filters were suggested as offering improved filtering characteristics [Drewery 1975], however the size of the large delays required (a frame delay is 625 times larger than a line delay) meant that such filters were considered too expensive and difficult to produce at the time.

In 1976 Weston proposed a two component system for the international exchange of digital Secam signals [Weston 1977]. This system, when applied to PAL, was later seen to provide a method for *clean PAL* coding [Oliphant 1980]. Later still, Weston Clean PAL, as the system had become known, was shown to achieve its effect through quadrature modulation of the luminance with the chrominance [Drewery 1984]. Thus a second technique for achieving chrominance/luminance separation was realised - *phase segregation*.

The obvious next technique to try was *time segregation*, with chrominance and high frequency luminance being sent alternately. A line based time segregation technique has been proposed [Holoch 1985] and named I-PAL. This technique and in particular a variant named I-PAL-M offers compatibility with standard PAL, whilst completely eliminating cross-effects when using an I-PAL decoder. A field based version of I-PAL has since been proposed (named *Colour Alternate Field PAL* or CAFFPAL) [Dean 1993], producing improved performance for stationary pictures. The ultimate use of time segregation is probably represented by the development of MAC (Multiplexed Analog Components [Annegarn et al, 1987]) (although this is not a compatible technique, of course).

During the 1980’s high speed mass storage was quickly becoming cheaper and more readily available. This meant that existing techniques could be improved by using more complex filters. Despite this their performance was still short of optimum. A major reason for this is the necessary trade-off between performance for stationary and for moving images. Line based techniques did not restrict image movement, but greatly limited vertical resolution. On the other hand field and frame based filters (temporal filters) offered increased vertical resolution, but produced motion artefacts. With field/frame filters being good for stationary images and line filters being good for moving ones the obvious answer was a *motion adaptive* coding technique [Teichner 1985].

Although motion adaptive coding appears the optimum solution reliable motion detection has proved difficult. The main reason for this is the enormous temporal aliasing present in television images. The temporal dimension is heavily under sampled, with pre-filtering rarely being added because of the unacceptable motion blurring that results. Added to this the complexity of the motion detector often outweighs that of the enhanced coding filters.

By the late 1980s the work on enhanced coding had led to the development of some improved commercial encoders, using line and field based filtering to reduce cross-effects in standard receivers. Also some top-of-the line receivers used line-based comb filters to improve their decoding. Significant improvements to the PAL (or NTSC) picture, however, require a matched modification of both encoder and decoder. Despite nearly 15 years of research no one technique stood out as a suitable upgrade standard, for either PAL or NTSC. With the impending HDTV revolution and the proposed introduction of the MAC system for satellite transmissions the incentives for significant improvement for terrestrial broadcasts were becoming great. To add to this the considerable improvements in camera technology were making the limitations of the current coding systems more and more obvious. Clearer,
sharper camera pictures meant more high frequency signal content, causing even worse cross-effects than before.

All these approaches to clean PAL coding are described in greater detail in Chapter 4. This Chapter places these techniques in a common framework, explains the theory behind them and compares their performance.

1.3 PALplus and the current enhanced television situation

In 1989 the PALplus project was initiated in Germany by a group of broadcasters and industry partners [Ellis 1994]. This group aimed to develop and finalise standards for an enhanced PAL system. The objectives for the system were:

- Full compatibility with conventional PAL.
- Full support of the 16:9 wide-screen format.
- Improved picture quality by reduction of cross-effects and increased luminance resolution.

For achieving these aims a number of possible algorithms were tested. One of these was a highly developed form of the BBC’s Weston Clean PAL [Croll 1992]. However, the system eventually chosen used motion adaptive time segregation on alternate fields. This system was named Motion Adaptive Colour Plus (MACP) [Vreeswijk 1993], and turns out to be a motion adaptive version of Dean’s CAFPAL. The PALplus system was finalised in early 1993, and broadcasts began in 1994 in several European countries.

The BBC, although losing out on PALplus have continued to push Weston Clean PAL as an enhanced studio standard. Rather amusingly their name for the system has continued to change every 9 months or so. Weston Clean PAL became Extended Studio PAL, then Single Wire Component, and at the moment is called Composite Compatible Component (COM3).

Despite efforts by the Japanese to develop a similar enhanced system for NTSC (called EDTV-II) [Fukinuki 1993], the Americans, the biggest players in the NTSC market have decided to circumvent enhanced NTSC completely. Instead they intend to jump directly into digital TV and HDTV [Rast 1993]. They are even implementing legislation to smooth the introduction path of the new technologies.

From recent events it might now appear that the need for further work on enhanced analogue television techniques has disappeared. However, this may not be entirely the case. For a start there will for a long time yet be a very large market of standard PAL receivers. Whilst offering significant improvements on enhanced receivers PALplus offers no improvement to existing receivers. Also, although PALplus is rapidly being embraced by the highly developed European PAL countries it is unlikely to attain great penetration of the world’s poorer PAL using countries. Considering the fact that this group includes both China and India there is likely to be a considerable and still growing market for conventional PAL in the medium-term future. Added to this is the newly realised set-top decoder market. Digital satellite transmission broadcasts of television are expected to increase enormously in the next few decades. In at least the short term this service will be provided to the consumer via set-top digital-to-PAL coders, which will connect to the consumers existing receiver. For receivers that have only composite video inputs (i.e. no SCART, S-video input) many of the benefits of digital transmission will be lost, unless some form of improved PAL coding is used in the set-top coders.

From these facts it may be suggested that there still exists now, and will exist for some time a market for enhanced PAL coding - especially for encoder only enhancements that can provide improvements on existing receivers [Olsen 1995].

The author has pursued this area, resulting in the development of a new enhanced PAL coding technique optimised for encoder only enhancement. Chapter 7 describes this technique (labelled K-PAL) in detail and compares it with WC-PAL.
1.3.1 The development of multi-dimensional television theory

The work on enhanced coding has been aided by new insights into the multi-dimensional nature of the television signal [Drewery 1975] and the effects of multi-dimensional sampling lattices on signal content [Tonge 1981]. Drewery looked at the concept of television's three dimensional spectrum based on early work [Mertz 1934] in studying the one dimensional television spectrum.

Whilst in one dimension the luminance and chrominance signals overlap completely, when viewed in terms of their three-dimensional spatio-temporal image content the chrominance signals actually occupy different areas of spatio-temporal frequency space. This suggested that the use of vertical and temporal filtering techniques would enable improved identification and separation of the Y (luminance), U and V (chrominance) components. Also, with knowledge of the eye's spatio-temporal frequency response [Robinson 1966] together with an understanding of multi-dimensional sampling theory the signal contents could be shaped using three dimensional filtering. This would allow the production of an optimum quality image, whilst minimising interference between the different signals.

Unfortunately despite these insights multi-dimensional analysis were not generally applied to the design, understanding or appraisal of enhanced coding techniques. There are perhaps two reasons for this.

Firstly the ideas were foreign to those already working in the field. The great majority are industry based researchers, and were more interested in building a solution than pursuing the theory any further. They preferred to remain with the familiar one dimensional approach. Thus vertical and temporal frequency components were shown as line frequency harmonics and vertical or temporal filters were described as line or field based comb filters, referring to their comb like appearance when viewed in one dimension. The operation of line and band segregation techniques were explained in terms of line-to-line signal differences, phase segregation in terms of line-to-line or field-to-field phase differences.

In later papers a combination of 1-D phase and time, and 1,2,3-D spectral descriptions were used to explain the techniques, creating a very confused approach. Often the same techniques were described completely differently by different authors. (Chapter 3 describes these many approaches to television signal theory, and explains the relationships between the 1, 2 and 3 dimensional views of the coded PAL signal).

The second reason for the lack of multi-dimensional theory is the difficulty of visualising or analysing the 3-D spectrum (graphing a 1-D function is a lot easier than displaying a 3-D one). To begin with all techniques were tested by implementing them in real-time hardware. 3-D signal analysis of the high bandwidth television signal was virtually impossible. It is only recently that the computing tools have become available for simulation of television coding techniques. Even with modern computers multi-dimensional signal analysis is still computationally intensive.

In Chapter 7 a new technique for obtaining accurate and complete characteristics for a television system is derived.

In the absence of true multi-dimensional analysis other approaches have been adopted. Pseudo spatio-temporal analysis can be carried out using spatial and parabolic zone plates [Weston 1982]. These are special test images that contain a range of spatio-temporal frequencies, varying like a frequency sweep across and down the image. The disadvantage of these is that it is difficult to obtain any quantitative information on a coding techniques characteristics - the results must be interpreted visually. Secondly these test images provide information on only a cross-section of spatio-temporal frequency space. The use of such test images for coding analysis is described in Chapter 4.

1.4 The assessment of enhanced coding techniques

The assessment of coding techniques is a very difficult area, and consequently one that has been somewhat neglected. A number of methods such as zone plates, multi-dimensional and mathematical analysis are available for examining the characteristics of a new coding technique. Once issues of complexity and implementation have been settled the image quality provided by the technique is one of the most important issues. However, for television, the performance of a technique in this respect is
CHAPTER 2 TELEVISION AND THE HUMAN VISUAL SYSTEM

completely dependent upon the perceptions of the viewer. Beauty (or image quality) is in the eye of the beholder. This means that standard error measures such as Signal-to-Noise Ratio (SNR) and Mean Square Error (MSE) are of limited use as they correlate poorly with subjective assessment.

The reason for this is that the human visual system (HVS) is both highly complex and non-linear. Also judgement of picture quality or degradation involves many intangibles - recognition of picture content, viewer attention, viewer experience, etc. The characteristics of the human visual system that impact on television are explained in Chapter 2. This Chapter also describes a number of mathematical models for the eye's response.

The complexity of picture quality evaluation has meant that the measurement of television coding picture quality has always been carried out to date through use of subjective assessment. A summary of the techniques used for this subjective assessment is given in Chapter 2, after the description of the HVS.

One of the problems with subjective assessment that is discussed is the ambiguity of the results. As an example, in a test carried out on 4 images with 10 observers for Weston Clean PAL [Oliphant 1980] the standard deviation of the results was far greater than the rated improvements in image quality.

Subjective testing has improved in recent times with the help of a set of standards issued by the CCIR [CCIR 1986], and use of more test images as well as image sequences. However, the results remain of limited usefulness, and the tests are highly labour intensive.

The benefits to be offered by objective testing have been long recognised. Although plenty of research has been done on developing techniques, based upon models of the human visual system [Budrikis 1973], no reliable alternatives to subjective testing appear to have emerged. Most of this work has been based upon analysis of the distortions present in the output image ie. the changes introduced by the coding [Lucas et al 1982]. This approach, aimed particularly at assessment of digital coding techniques, may not be the most appropriate for analogue television coding techniques.

In Chapter 6 a new technique for objective assessment will be described.

1.5 Overview of Thesis

This thesis can be divided into a number of separate subject areas.

1.5.1 Theory of television and the human visual system

Chapters 2 and 3 provide some background for the thesis on the implications of the human visual system for television, and the theory behind the television signal. These chapters provide a contribution to the field by bringing together from a wide range of sources information on the major building blocks in the understanding, and design of enhanced coding systems.

Chapter 2 describe in detail the characteristics of the human visual system (HVS) explaining its implications for television. The information in this chapter is sourced from a range of scientific papers on the study of the HVS, and represents a comprehensive compilation and interpretation not readily available elsewhere. A number of mathematical models for describing the HVS are presented, including a new model for the eye's spatio-temporal chromatic response developed by the author. This chapter also provides an introduction to picture quality assessment.

Chapter 3 presents the theory behind colour television. This includes scanning, sampling, PAL coding and an examination of the problems with conventional PAL. The relationships between 1, 2 and 3 dimension views of the PAL signal are explained in detail. The chapter provides a comprehensive explanation of the multi-dimensional nature of the television signal and applies this representation to the explanation of coding effects and comb filtering. There has been considerable confusion and inconsistency in the past when these things have been described. In contrast this chapter creates a concrete and consistent mathematical platform for understanding and describing the television signal and enhanced coding techniques. As such this chapter provides something not readily available elsewhere.
1.5.2 New methods for analysing and assessing television coding techniques

Chapter 5 presents an entirely new model for a colour television system as a Many Input Many Output (MIMO) linear periodically time-varying (LPTV) system. This provides a complete representation for the characteristics of the colour coding system. As such it is the first model of its kind proposed for television coding techniques and represents a very powerful tool for future development. This model can provide information on virtually every aspect of the system (e.g. cross effects, aliasing, spatio-temporal bandwidth, etc.). Chapter 5 also explains a numerical technique developed by the author for determining these characteristics based upon a computer simulation of the coding system. The results are three-dimensional spatio-temporal plots of the system’s characteristics (transfer functions) for every possible path in the system. These provide a view of many coding techniques never before seen and consequently can give enormous insight into how different coding techniques work and what effects they have on the television signal. Results are presented for standard PAL, I-PAL, and ColourPlus, with further results for Weston Clean PAL presented in Chapter 7. These multi-dimensional pictures of a system’s characteristics represent an enormous advance over conventional analysis techniques using zone plates and specialist images (see overview of conventional techniques in Chapter 4).

Chapter 6 combines the newly obtained spatio-temporal characteristic of a coding system with the HVS models described in Chapter 2. The result is a radically new approach to objective quality assessment that can be used for predicting the perceptual levels of resolution and artefacts in a coding system. Results obtained using this new method are compared against those obtained using normal subjective assessment methods (as described in Chapter 2) to show a medium-level correlation. The technique shows promise, but requires further work to fully develop it fully and improve the correlation.

1.5.3 Enhanced coding techniques and a new technique - K-PAL

Chapter 4 describes the current approaches to enhanced PAL, placing them in a common framework. The main approaches of band, time and phase segregation, and adaptive coding are described, with examples from current literature. Such a generalised framework for enhanced coding has not been described well previously, nor have the range of techniques covered in Chapter 4 been compared together. This framework helps to identify the many common elements amongst the many different techniques that have been proposed.

The highlight of this thesis is arguably the presentation of a new coding technique, named K-PAL. Chapter 7 describes K-PAL in detail and compares it with WC-PAL. The Chapter firstly derives a complete mathematical description of WC-PAL and K-PAL (such a description has not been presented elsewhere for WC-PAL). Implementation details and filter requirements are discussed and a full set of simulation results are presented using some of the new tools developed in Chapters 5 & 6 (spatio-temporal spectra and objective assessment). K-PAL is fundamentally different from WC-PAL and all other modern enhanced PAL techniques in its separate treatment of the U and V signal paths. Instead of combining U and V into one chrominance signal and then segregating Y and C, K-PAL separately segregates Y, U, V to produce an optimal signal for minimising cross-colour under compatible reception.

1.5.4 New simulation and display tools

The work described in this thesis required extensive simulation and computer assisted analysis. When the research was begun no software tools or image processing packages available were suitable for these tasks. Consequently during the course of this research a unique television image processing system has been developed by the author (named TVPROC), with a particular emphasis on multi-dimensional signal analysis. This system is described in Appendix E. Also contained in the Appendix (Appendix F) is a description of a specialised viewing package (named PHIGSdraw) created by the author for displaying 3D spatio-temporal characteristics.

In summary, the major new work presented in this thesis is (i) a new and complete representation of a television coding system as a 3-D spatio-temporal MIMO LPTV system, (ii) the development, implementation, and demonstration of an entirely new method for determining complete and accurate
spatio-temporal characteristics for any coding system, (iii) a fundamentally new approach to objective assessment, (iv) the discovery and analysis of all aspects of the new and optimal enhanced PAL coding technique named K-PAL.

1.6 Publications based upon the work in this thesis


CHAPTER 2

TELEVISION AND THE HUMAN VISUAL SYSTEM

This Chapter provides an overview of the mechanisms and characteristics of the human visual system, and compares this with the visual aspects of television. This is followed by the development of several models for the eye's spatio-temporal visual threshold, and a discussion of quality assessment methods.

2.1 The structure of the eye

Below is a diagram of the human eye showing the principal parts.

![Diagram of the human eye showing main features](fig:2-1, Wentworth 1955).

Several mechanisms within this structure are used to control the light entering the eye:

- The changing diameter of the *pupillary aperture*, which can vary from 2 - 8 mm.
- *Accommodation* - the adjustment of the position and curvature of the lens, changing the focus of the eye.
- eye movement - a number of small muscles around the eye improve vision by allowing tracking of moving objects, and by dithering movements called *saccades*.

The light entering the eye is thus adjusted for best effect on the retina. The optics of the eye are, however, imperfect, and it is impossible to bring the image properly into focus over the whole retina. Instead the eye aims for best focus in the *fovea* - the most central area of the retina. When light levels are adequate the *iris* is closed as far as possible to allow a sharper image and greater depth of field.

The retina, which covers almost the entire internal surface of the eye, is composed of over 100 million receptors. Relatively simple neural circuitry combines the signals from these receptors into the approximately 1 million nerves that pass through the optic fibre into the brain.

The light receptors in the retina are of two types, *rods* for night-time vision and *cones* for day-time vision (the names referring to their shapes). Although we see mostly by use of the cones under normal
daytime conditions there are far more rods than cones in the retina, the ratio being around 20 to 1 (approx. 6 million cones). The distribution and size of the receptors also varies greatly across the retina, with cones concentrated in the central region, rods in the peripheral areas and the size of both decreasing towards the centre. The different distributions of rods and cones is illustrated below.

![Figure 2.2 Relative density (and therefore acuity) for rods (scotopic) and cones (photopic) across the retina [fig. 36, Walsh 1958].](image)

The central area, containing almost no rods is called the macula lutea, and is around 2 - 2.5 mm. Smaller still within the macula lutea is the fovea, the area of peak cone density. This tiny portion of the retina, around 0.5 mm\(^2\), is estimated to contain over 34 000 cones, tightly packed in an extremely fine structure (1 - 3 μm in diameter). As would be expected, it is within the fovea that vision is the clearest. It is also interesting to note that there is an almost one-to-one connection between cones in the fovea and nerves in the optic fibre. In the outside areas of the retina very many rods or cones may be connected to each optic nerve.

Taken together this information points to the appearance of an image varying enormously depending upon which portion of the retina it is focused. Tiny involuntary eye movements (the saccades), and considerable higher neural processing in the brain help to produce the illusion of a consistent and homogenous image.

### 2.2 Visual acuity (spatial)

The visual acuity of the eye is determined by a number of factors.

1. The density of receptors in the retina.
2. The level of accommodation.
3. Aberrations in the lens.
4. Chromatic aberration - meaning different focal lengths are required for different light wavelengths.
5. The light level.

Factors (a) and (b) were discussed in the previous section. For (c) aberrations in the lens means that light passing through the outside portions of the lens is focused differently from light passing through the centre. This problem is minimised when, at high enough light levels, (above 10 cd/m\(^2\) field luminance) the pupil diameter is at a minimum. This restricts light to passing through the central portion of the lens.

The effect of (d), chromatic aberration, is reduced in the eye by (i) the lens material restricting light below 400 nm, and (ii) the existence of a yellow pigment covering the macula, corresponding to the central 3-4° of vision. This pigment attenuates the out of focus blue light.

The light level (e) affects the acuity by causing changes in the accommodation and pupil diameter, and by determining the relative contribution of rods and cones to the response. At very low light levels,
2.2 VISUAL ACUITY (SPATIAL)

When sight is predominantly by use of the rods their different distribution in the retina changes the visual acuity. This means that night-time vision is most acute 4-5° from the optical axis. For the viewer this means that the best image will be obtained by looking slightly to one side of an object. This can be noticed in viewing the stars on a dark night - they will appear brightest when looked at slightly to one side.

As visual acuity will be related to the receptor density, the distribution of receptors in the retina should provide an indication of the expected change in visual acuity across the retina. Maximum acuity lies in the foveal region, a tiny 1.7° out of a total field of view of approximately 140° horizontally and 100° vertically. As indicated earlier the cones within the foveal area are tightly packed, with an average diameter of 2 \( \mu m \), which corresponds to an angular diameter of 0.5 minutes of arc (1120 degree). According to the Nyquist sampling theorem the maximum theoretical spatial frequency the retina should be capable of resolving is therefore 1 cycle/min or 60 cycles/degree (cpd). Experimental results [Higgens et al 1948, 1950] with parallel line black and white gratings give a measured value of 0.5 - 0.7 cycles/min (30 - 42 cpd). This lower value is to be expected, due to the finite size of the receptors, and the many other imperfections described earlier. It is in fact quite amazing that the measured acuity comes so close to the theoretical limit.

The acuity of the eye, as well as depending upon the angle from the optical axis, also depends to some degree upon the orientation. Tests measuring acuity with black and white gratings [Higgens et al 1948, 1950] show that visual acuity is 10 to 20 % lower for diagonal detail, than for purely horizontal or vertical detail. It appears the human eye is designed to be more sensitive to horizontal and vertical edges. This may be due to the structure of the retina, or higher processing that is designed to detect such detail.

### 2.2.1 Spatial acuity for colour

In order to see colour the cones are of three types (usually called red, green, and blue), with each type being sensitive to a different portion of the visible spectrum (the eye's response to colour is described in more detail in section 2.5). Due to this fact it is to be expected that the eye's acuity for colour will be lower than for non-colour detail. Studies show that the different cone types are not present in equal amounts. Walraven and Bouman [Walraven 1966] measured population densities of red, green and blue receptors in the ratios of 40:20:1. This means that 65.6% of cones are red receptors, 32.8% are green, and only 1.6% are blue. It would be expected therefore, that the colour acuity of the eye be lower for those colours whose receptors are fewer. This is in fact the case, as shown by the experimental results below [Bedford 1950].

<table>
<thead>
<tr>
<th>Condition</th>
<th>Acuity in cpd</th>
<th>Acuity in % black &amp; white acuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>black-white</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>red-green</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>red-blue</td>
<td>8.5</td>
<td>23</td>
</tr>
<tr>
<td>green-blue</td>
<td>7</td>
<td>19</td>
</tr>
</tbody>
</table>

*Table 2.1: Acuity limits of human eye for different colour combinations [Bedford 1950].*

Other factors, such as chromatic aberration, and the attenuation of blue light are likely to contribute to the reduction of acuity towards the blue end of the spectrum.

### 2.2.2 Spatial acuity and television

The level of detail presented to the eye by a television screen is directly related to the distance a person sits from the screen. Studies on home television viewing conditions indicate that there is considerable variation in the viewing distance. Results from one study [Nathan 1985] showed that 90% of children's viewing was between 2.6H and 10.5H (where H = picture height), and for adults the 90% limits were 6.5H and 14H. In determining the level of detail required for maximum image quality it makes sense to design for the worst case. These statistics thus help explain the choice of 4H and 6H as the recommended distances for subjective testing of conventional television systems, with 6H being the
preferred viewing distance [CCIR Rec. 500-3 1986]. The subtended angles (Figure 2.3) for these viewing distances are given below in Table 2.2.

For comparison this table also includes the subtended angles for a computer monitor, which is typically viewed at a distance of around 2H.

Using the information in Table 2.1 the eye’s resolution limit in lines may be estimated for these different viewing distances.

![Figure 2.3 The viewing angle as related to viewing distance and picture size.](image)

<table>
<thead>
<tr>
<th>Signal</th>
<th>2H</th>
<th>4H</th>
<th>6H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>height</td>
<td>width</td>
<td>height</td>
</tr>
<tr>
<td>black-white</td>
<td>2100</td>
<td>2700</td>
<td>1000</td>
</tr>
<tr>
<td>red-green</td>
<td>840</td>
<td>1100</td>
<td>420</td>
</tr>
<tr>
<td>red-blue</td>
<td>480</td>
<td>640</td>
<td>240</td>
</tr>
<tr>
<td>green-blue</td>
<td>400</td>
<td>520</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 2.3 The estimated resolution limits of the eye in lines for several different viewing distances (values shown to 2 significant figures).

Obviously these resolution limits only apply to that portion of the picture in foveal vision. In practice a considerable amount of the picture will be in peripheral vision, and the resolution limit for these areas will be correspondingly lower. However it is difficult to predict at which areas of the picture the fovea (i.e. the viewer’s attention) will be directed, especially for moving images where the eye may be tracking motion across the display. Consequently displayed picture quality must be chosen to satisfy the requirements of foveal vision. The values shown in Table 2.3 thus indicate the required resolution in lines for a theoretically ideal display device.

These ideal values may be compared with those in Table 2.4 for the maximum resolutions of the standard television signals.
2.2 VISUAL ACUITY (SPATIAL)

<table>
<thead>
<tr>
<th>System</th>
<th>Signal</th>
<th>Current height</th>
<th>Current width</th>
<th>Best possible height</th>
<th>Best possible width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTSC</td>
<td>Y</td>
<td>288</td>
<td>320</td>
<td>480</td>
<td>433</td>
</tr>
<tr>
<td></td>
<td>I/Q</td>
<td>288</td>
<td>52/64</td>
<td>288</td>
<td>52/64</td>
</tr>
<tr>
<td>PAL 5 MHz</td>
<td>Y</td>
<td>346</td>
<td>384</td>
<td>576</td>
<td>520</td>
</tr>
<tr>
<td></td>
<td>U/V</td>
<td>144</td>
<td>63</td>
<td>288</td>
<td>63</td>
</tr>
<tr>
<td>PAL 5.5 MHz</td>
<td>Y</td>
<td>346</td>
<td>384</td>
<td>576</td>
<td>572</td>
</tr>
<tr>
<td></td>
<td>U/V</td>
<td>144</td>
<td>114</td>
<td>288</td>
<td>114</td>
</tr>
<tr>
<td>Computer Monitor</td>
<td>R/G/B</td>
<td>1024</td>
<td>1280</td>
<td>1024</td>
<td>1280</td>
</tr>
</tbody>
</table>

Table 2.4 Maximum resolution (in lines) possible for several different coding standards (Kell factor = 0.6).

<table>
<thead>
<tr>
<th>System</th>
<th>Signal</th>
<th>Current</th>
<th>Best possible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTSC (6H)</td>
<td>Y</td>
<td>40%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>I/Q</td>
<td>100%</td>
<td>15%</td>
</tr>
<tr>
<td>PAL 5 MHz</td>
<td>Y</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>U/V</td>
<td>50%</td>
<td>15%</td>
</tr>
<tr>
<td>PAL 5.5 MHz</td>
<td>Y</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>U/V</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Monitor</td>
<td>Y</td>
<td>50%</td>
<td>45%</td>
</tr>
<tr>
<td>(2H) Colour</td>
<td></td>
<td>120%</td>
<td>110%</td>
</tr>
</tbody>
</table>

Table 2.5 Resolution of television coding methods as a percentage of the eye’s resolution limits (rounded to nearest 5%). Percentages for colour are as compared to the Red-Green acuity of the eye.

Comparing actual resolutions with the eye’s resolution limits there is a considerable difference. Even at the greatest distance of 6H both PAL and NTSC have less than half the resolution required to match the acuity limits of the eye. In the horizontal dimension the situation is particularly bad, ranging down to less than a quarter (63 cycles/width) the ideal resolution (depending upon colour). Even a high resolution computer monitor does not reach close to the acuity limits of the eye (at 2H).

It appears that there is still considerable room for improvement in the spatial resolution of conventional television (and computer displays), particularly in the horizontal direction. The fact that the current standards are a long way from theoretical limits suggests that the moderate improvements obtainable with enhanced coding techniques (see end columns of Table 2.5) could offer worthwhile improvements in quality. It might also be possible to take advantage of the considerably lower acuity of the eye for blue colours, concentrating colour resolution improvements on red and green (note that NTSC already does this to some degree with the I and Q colour difference signals being different bandwidths).

2.2.3 Acuity and motion

The acuity of the eye when tracking motion (i.e. the dynamic acuity) reduces as the image velocity increases [Miller 1962]. This results from tracking errors in the eye, producing a reduced resolving ability. The eye must then have a maximum resolvable level of detail for a particular image velocity, which may be expressed as a temporal frequency (no. of cycles passing a fixed point in the image). This maximum has been estimated at 1000 Hz [Tonge 1985]. This corresponds to, for example, a resolution limit of 20 cpd for an image velocity of 50 degrees/second.

2.2.4 Television’s temporal bandwidth requirements for moving images

Applied to television the limit mentioned above implies a required temporal bandwidth of 1000 Hz, or 2000 fields/frames per second. A lower frame rate would cause visible aliasing. This conclusion begs
the question as to why 50/60 Hz systems appear to work so well with vastly lower frame rates. The answer is that the confusion of velocity caused by aliasing only occurs for images containing purely high spatial frequencies eg. such as a moving grating. For most practical picture content high spatial frequencies are almost always accompanied by low frequency 'boundary' information. This information is not aliased, and allows the eye to ignore alias effects present in the higher spatial frequencies.

It is likely due to this effect that an adequate appearance of motion can be achieved with frame rates as low as 8 per second. However, at such low rates aliasing effects (e.g. flicker, jerky movement) become very noticeable. The 50/60 Hz field rates chosen for television are, in fact, determined mostly by flicker considerations (discussed in next section). Because of the low field rates used some aliasing artefacts are still visible in the current systems, eg. the apparently reverse rotation of a wagon wheel or propeller, and the jerky motion of a quickly moving ball or vehicle in sports events. Considering the enormous frame rates required to eliminate these effects, 50/60 Hz systems are generally accepted as providing a reasonable compromise.

### 2.3 Temporal acuity

The ability of the stationary eye to detect temporal changes in brightness (flicker) is limited. Experimental studies [De Lange 1958] on a 2° central vision field give a critical flicker fusion (cff) frequency of around 40-50 Hz. At this frequency any flicker becomes undetectable and simply appears as the time average brightness.

Just as the spatial acuity changes across the retina, so does the temporal acuity. Unlike spatial acuity however, temporal acuity increases for non-central vision, up to a limit of 70-80 Hz. This is likely due to the higher percentage of rods in peripheral vision, which are more sensitive to flicker.

Consequently, for wider fields of vision the cff is higher.

The cff also changes with image brightness; the higher the brightness the more sensitive the eye is to flicker effects. This increase in cff with brightness is shown below, for a 15° field of view (~5H viewing distance) and a range of screen phosphor decay times. Note that most modern screens have short decay times of less than 2 ms. This relationship suggests that flicker can be minimised by viewing at relatively low light levels, such as in near darkness. This approach is used in picture theatres where films are usually displayed in almost complete darkness.

![Figure 2.4 Change of critical flicker fusion frequency (cff) with screen luminance, shown for several different display decay times [fig. 2-11 Wentworth 1955].](image)

---

1 It has also been postulated that peripheral vision has a higher temporal acuity to improve the detection of movement - to "better see danger approaching from behind".
2.4 THE RESPONSE OF THE EYE TO ILLUMINATION

2.3.1 Temporal acuity for colour

As with spatial acuity temporal acuity for colour is lower than for black and white. Measurements give a critical colour flicker fusion (ccff) frequency of between 7 Hz (5.3° field) [van der Horst 1969] and 12 Hz (1° field) [Wisowaty 1981]. Considering that colour vision is limited in the periphery of the eye it is likely that these areas are less sensitive to colour (chromatic) flicker, as opposed to being more sensitive as for achromatic acuity. There does not appear to be conclusive results as yet on how or whether the ccff varies for different colours.

2.3.2 Television's temporal resolution requirements

The temporal bandwidth requirements of television have already been discussed with regard to motion portrayal. In terms of flicker considerations Figure 2.4 appears to indicate that an 80 Hz picture rate is needed to completely eliminate flicker under all viewing conditions. In practice 60 Hz (NTSC) and 50 Hz (PAL) picture rates are used. This means that some flicker is still visible, especially for close viewing, and bright images. The difference between NTSC and PAL in this respect can be quite noticeable. Those used to NTSC can find the flicker present in a PAL receiver quite irritating. Flicker levels are made worse by the use of interlace, which causes an extra 25/30 Hz flicker component to be present. This is particularly noticeable in small picture areas, causing what is known as inter-line flicker (see section 3.3.1).

The table below compares the flicker requirements of the eye with the frame/field rates currently used for image display2.

<table>
<thead>
<tr>
<th>System</th>
<th>Black &amp; White</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human eye</td>
<td>60 - 80 Hz</td>
<td>7 - 12 Hz</td>
</tr>
<tr>
<td>NTSC</td>
<td>60 (30) Hz</td>
<td>60 (30) Hz</td>
</tr>
<tr>
<td>PAL</td>
<td>50 (25) Hz</td>
<td>50 (25) Hz</td>
</tr>
<tr>
<td>Computer monitor</td>
<td>70 / 76 Hz</td>
<td>70 / 76 Hz</td>
</tr>
</tbody>
</table>

Table 2.6 Temporal resolution limits of eye, television and a typical computer monitor.

It is noticeable that although the picture rates of television are too low to completely prevent achromatic flicker they are far higher than the colour cff. It appears that a reduction of chrominance temporal resolution by a factor of 2-4 times would likely be quite acceptable. Conversely, an improvement in the achromatic temporal resolution is needed.

2.4 The response of the eye to illumination

The human eye has an enormous dynamic range of nearly 1000 million over which it can function. This enormous range is helped by the different characteristics of the rods and cones. As already mentioned, rods are most sensitive at very low levels of illumination, where cones stop functioning. The diagram below shows this change from rod (scotopic) to cone (photopic) vision.

The changeover from rod to cone vision occurs at around 0.01 - 0.1 troland3. Figure 2.5 also shows how the perception of brightness follows a log relationship \( \Delta L \propto \log L \), known as Fechner's Law.

At any one time the eye cannot perceive this entire range of illumination. Instead it must adjust to the current average brightness level and perceive a range of illumination centred around this average. This range covers several orders of magnitude, but is perceivable in only around 100 contrast steps.

---

2 It is interesting to note that computer monitors, which are usually viewed at between 1-2 Hz offer much higher picture rates, typically 85 Hz interlaced, or 60-76 Hz non-interlaced. Due to the large field of view a computer screen occupies lower picture rates would otherwise cause eye strain due to considerable flicker.

3 The Troland is a measure of retinal illumination. One Troland is the retinal illumination produced by a surface having a luminance of 1 cd/m² when the pupil diameter is 1mm.
CHAPTER 2  TELEVISION AND THE HUMAN VISUAL SYSTEM

2.4.1 Adaptation

The process of brightness adjustment is called adaptation. The eye takes the average current value of the illuminance as a brightness mid-point, and black and white are then judged relative to this adaptation level. Under a constant state of adaptation the range of illumination perceivable is much reduced to 200-1000 (depends upon absolute illuminance). The adaptation process is relatively quick (several seconds) for changes from dark to light surroundings, but it is much slower (up to 30 mins [Wentworth 1955]) for changes in the other direction i.e. dark adaptation.

As well as general brightness levels the eye adapts itself to changes in the quality or spectral distribution of the illumination. This accounts for the fact that objects appear substantially the same colour, whether viewed under sunlight, through tinted glasses, or under a whole range of artificial illumination methods (incandescent, fluorescent, etc). As the graphs below show, the spectral distribution of these different light sources varies markedly.

Overall adaptation is a complex process, made up of the different processes of global, local, and lateral adaptation. These different luminance and colour adaptation mechanisms are responsible for the appearance of complementary colour after-images, and some optical illusions (e.g. the appearance of grey squares near the corners of a black square checker board pattern - a Hermann grid).
2.4.2 Illumination and television

The implications all this has for television are relatively straightforward. Firstly, for colour television the screen must be bright enough to operate in the photopic vision range. Testing recommendations [CCIR Rec. 500-3 1986] suggest a range of screen luminance up to 70 cd/m², which equates to an average pupillary illuminance of about 200 Td at 6 H (about 4 on the horizontal axis of Figure 2.5). Secondly the background illumination should be at approximately the average brightness of the screen, and at a standard white (CCIR Rec 500-3). It is also recommended that no brightly coloured objects be placed near the screen, as this would change the colour adaptation state of the eye, possibly changing the appearance of colours on screen.

2.5 The response of the eye to colour

As has already been mentioned the eye sees colour by use of three cone types, called the red, green, and blue (or R, G, B) cones. As well as different densities these each have different spectral sensitivity curves, shown below, with their combined response producing the photopic response.

![Figure 2.7 Relative sensitivities of red, green, and blue cones.](image)

It is generally accepted that the R, G, B cone outputs are converted inside the retina to a brightness (White-Black) and two colour difference channels (Red-Green and Yellow-Blue). This conversion process is shown in simplified form in Figure 2.8.

The characteristics of the three resulting channels are different, and are dealt with in more detail in later sections. The relative weighting for the R, G, B cones is not shown here, but will be different for each cone type, and will also likely change with colour adaptation. For the W-Bk channel it might be expected that the relative weightings of R, G, B will be in accordance with their relative sensitivities as shown in Figure 2.7.

In general the nature of colour vision this means that to define a colour exactly requires the specification of three values. There is a number of different ways of doing this;

- R, G, B values.
- Hue, Saturation, Brightness (or Value).
- Luminance, and two colour difference (U/V or I/Q).
- The x, y, z CIE chromaticity co-ordinates.
The last of these is the most commonly used for colorimetry, and is conventionally plotted against the $x, y$ axes only, ignoring the $z$, or brightness value. The resulting diagram is shown in Figure 2.9.

![Diagram](image)

**Figure 2.8** Simplified diagram of retinal neural coding of responses from R, G, B cones [fig. 8-40, Uttal 1981].

![Diagram](image)

**Figure 2.9** The standardized $x, y$ CIE chromaticity diagram, showing the hue areas corresponding to the ten major Munsell hues [fig. 3-30, Wentworth 1955].
Along the spectrum locus shown are the colours produced by single wavelengths, and within this locus are all possible visible colours. This diagram also shows the subjective colour groups for different areas, with the central region being a set of different colours perceived as ‘white’.

### 2.5.1 Colour and television

The human eye does not perceive the spectral distribution of the light directly, only its intersection with the R, G, B cones response shown in Figure 2.7. This means that the sensation of one particular colour can be produced by many different spectral distributions. As an example, white (though not technically a colour), occurs when the response of the three cone types is the same. White would be produced by light having a uniform distribution over all frequencies, or by three distinct monochromatic colours, positioned so as to intersect with the individual cone responses. This property is made good use of in television, where by combining only three colours (the red, green and blue phosphors) most of the visual colours can be produced.

These three phosphors used in television each have their own chromaticity co-ordinates, and define a triangular area of colour, or colour gamut, that the screen is capable of displaying. This gamut is shown below on the chromaticity diagram.

![CIE chromaticity diagram showing gamuts for television and other colour processes](image)

*Figure 2.10 CIE chromaticity diagram showing gamuts for television and other colour processes [fig. 4-1, Wentworth 1955].*

As shown above the practical gamut differs from the ideal gamut, due to the fact that the screen phosphors are not ideally mono-chromatic and instead contain a distribution of frequencies. From Figure 2.10 it is obvious that television cannot reproduce the entire range of visible colours, even if different ideal phosphors were used (max. additive gamut on Figure 2.10). This has some significance for the use of colour screens in scientific visualisation and image processing but is quite acceptable for normal television entertainment viewing. It is interesting to note that the television gamut is still greater than that possible using other display media.

### 2.6 The achromatic spatio-temporal threshold of the eye

For a given adaptation state the detectability of a spatio-temporal stimulus depends upon its frequency. Generally speaking the closer the stimulus becomes to the spatial or temporal acuity limits of the eye
the higher the contrast detection threshold becomes (ie. the greater the attenuation characteristic of the eye).

Many experiments have been carried out to determine these characteristics; as a function of temporal frequency [Ives 1922, de Lange 1958], of spatial frequency [Schade 1956] and of combinations of spatial and temporal frequency [Robson 1960, van Nes 1967, Kelly 1972]. In most of these experiments the subject viewed the test area within a larger field at a constant luminance equal to the average value of the spatial/temporal test pattern. An example set of results from such an experiment is shown below [Robson 1966].

\[
L = L_0 \left(1 + m \cos 2 \pi \mu_0 x \cos 2 \pi f_0 t\right) \quad (2.1)
\]

where \(L_0\) is the average luminance, \(\mu_0\) the spatial frequency and \(f_0\) the temporal frequency. For the experiments the threshold is determined by adjusting \(m\) until the stimulus is just visible. The results obtained for a range of \(\mu_0\) and \(f_0\) define a spatial-temporal threshold surface for the eye, such as that shown below.
2.6 THE ACHROMATIC SPATIO-TEMPORAL THRESHOLD OF THE EYE

Studying Figure 2.11 and Figure 2.12 it is noticeable that in both dimensions the curves are low-pass at high frequencies and band-pass at low frequencies. Thus there is a set of spatial and temporal frequencies at which the threshold is a minimum i.e. at which the visual system is most sensitive. The null at zero frequency is to be expected, as this is an effectively DC component that will be removed by the eye's adaptation response.

The particular advantage of these experimental results is that, being obtained using a sine-wave test stimulus, linear theory and Fourier analysis can be used to predict the response for all other signals. This would make characterisation of the eye very simple, if it weren't for the fact that the visual system is neither linear nor time-invariant. Nevertheless, given a sufficiently constant adaptation state, and confining our study to threshold stimuli only linear theory is seen as being applicable [Budrikis 1973].

This generalisation of results using Fourier analysis is backed up by results from tests with square and triangle wave data [van der Horst 1969]. For these signals the detection threshold was found to be solely determined by the amplitude of the fundamental frequency. This resulted in the expected threshold ratios of 1:1.27:0.81 for the sine, square, and triangle waves respectively.

The data plotted in Figure 2.11 is for a set background luminance of \( L_0 = 6.3 \) mL. Considering the complicated nature of the eye's response the characteristic plotted might be expected to change at lower or higher levels of illuminance. In actual fact they remain almost identical in shape over a wide range of test illuminances [van Nes 1967, Kelly 1972]. Their amplitude, however, does change following the DeVries-Rose law [Lukas 1982] with \( m = L^{0.5} \). This relationship is illustrated by the experimental results plotted below. For very high illuminance levels (outside that encountered with television) the thresholds become constant, following Weber's Law [Lukas 1982].

![Figure 2.13 Change in threshold values M with mean retinal illuminance B_o, illustrating DeVries-Rose law [fig. 6, van Nes 1967]](image)

Note that the lines in this plot all have a slope of 0.5, as predicted.

An unknown factor in the measurement of these thresholds are the eye movements of the subject. It is quite likely that such movement would have an effect upon the thresholds, as it would cause spatial detail to vary temporally on the retina. More recent experiments [Kelly 1979, Burbeck 1980] using eye tracking and image stabilisation have verified this, measuring different threshold curves. Under normal viewing conditions, however, these small eye movements will always be present, and as far as television is concerned the non-stabilised response is more appropriate.

2.6.1 The threshold functions and television

The significance of these threshold characteristics for television is fairly obvious. When considering picture artefacts in the luminance (e.g. noise, distortion) that are close to visual threshold, the visibility of these artefacts should depend upon their spatio-temporal frequency components in the manner predicted by Figure 2.11 and Figure 2.12. According to these figures very low, or high spatial-temporal artefacts are likely to be far less noticeable than intermediate ones of equivalent amplitude. These threshold characteristics should provide a good indication of the visibility of artefacts, and could theoretically be used to predict perceived distortion levels in an image.
2.6.2 The supra-threshold response

The above results describe the contrast-sensitivity function for the eye at threshold. At levels above threshold (supra-threshold), the response changes. It has been found [Bowker 1983] that as the contrast is increased variations in the perceived contrast as a function of spatial-temporal frequency decrease. This is interpreted as being due to a compensation mechanism that compensates for high and low frequency rolloffs. At levels well above threshold this effect will cause the perceived contrast of a spatial-temporal frequency to correspond more closely to its physical contrast. Generally speaking this compensatory gain increases the further above threshold the stimulus becomes, meaning at high enough levels overcompensation can occur. An example of this is Mach Bands [Springer 1979], an optical illusion caused by visual overcompensation.

2.6.3 The supra-threshold response and television

The difference between the threshold and supra-threshold characteristics is important. Whilst the threshold characteristics are likely a good predictor of the visibility of low level image artefacts the same cannot be said of the main image information. This information should be well above threshold (supra-threshold) and thus it might be expected that the visual compensatory mechanisms will, on average, cause high frequencies to be just as visible as lower frequencies (within the acuity range of the eye). We must conclude therefore that the threshold characteristics, although a good weighting function for low-level distortions and artefacts at the threshold of perceptibility, cannot provide an appropriate weighting for the main image information.

It is of interest to note that supra-threshold effects are likely the reason that sharpness enhancement (boosting higher spatial frequencies) on video images is so successful. For low contrast image areas this enhancement will assist the compensatory mechanism of the eye. It also helps compensate for the loss of higher frequencies caused by image capture and display processes. As would be expected, this sharpness enhancement is usually inappropriate for high quality pictures.

2.7 The chromatic spatio-temporal threshold of the eye

Experiments have been carried out studying the chromatic threshold characteristics as well as the achromatic (luminance). These have variously determined the spatial [van der Horst 1967, Granger 1973], temporal [van der Horst 1969, Kelly 1977, Wisowaty 1981] and the combined spatial-temporal characteristics [Kelly 1974, van der Horst 1969]. Although there has been some confusion caused by the difficulty of generating purely chromatic stimuli the general differences found were

- Much lower cut-off of threshold curves for both spatial and temporal frequencies.
- A purely low-pass characteristic with no resonance peak at intermediate frequencies.

In other respects the chromaticity thresholds vary in a similar manner, following the deVries-Rose law for changing illumination levels.

Some chromatic threshold characteristics are shown below, taken from a paper by van der Horst.
Comparing these graphs with Figure 2.11 confirms the differences described (note that the above plots are threshold values, not 1/m as for Figure 2.11). The two sets of thresholds shown are for the Red-Green (R-G) and Yellow-Blue (Y-B) colour mechanisms (see Figure 2.8). As would be expected from earlier discussion the Y-B thresholds show a lower cutoff, at around 2-3 cpd compared to 2-3 cpd for R-G. The temporal cutoffs, however, are identical.

Supra-threshold measurements for chromatic thresholds do not appear to have been carried out to date (the author could find no information). It is possible that a similar gain mechanism to that for achromatic stimuli works in the chromatic paths. If it existed it would cause a flattening of the response characteristics, with the eye’s chromatic response cutoff more closely approaching its chromatic acuity limits (see sections 2.2.1, 2.3.1).

### 2.7.1 Chromatic thresholds and television

Just as for the achromatic thresholds the above results are a suitable weighting function for low-level chromatic artefacts (i.e. artefacts near threshold). It would appear that the human eye is likely to be much less sensitive to chromatic artefacts. However, once again, it is likely that these characteristics do not apply to stimuli well above threshold. We will therefore conclude, similarly to the achromatic result, that the chrominance thresholds are suitable only for judging chromatic artefacts near threshold, but not the main colour information.

This means they will not be reliable indication of the relative importance of spatial-temporal chrominance information in a television image. Figure 2.14 (a) and (b) are appropriate weighting functions for low-level chromatic artefacts and distortion only.
2.8 Modeling the eye’s threshold characteristics

In order to apply the previously described results to the appraisal of image distortion a mathematical model is required. A number of such models of varying complexity have been discussed and developed for the achromatic response [Budrikis 1972, 1973, Sarikson 1977, Granrath 1981, Lukas 1982]. The most common approach is to use a combination of independent spatial/temporal excitation and inhibition. Such a model is shown below.

![Combination spatial-temporal inhibition-excitation model used as a visual weighting mechanism](image-url)

where the functions in the model are

- \( L(x, y, t) \) - error signal (of spatio-temporal image)
- \( C(x, y, t) \) - weighted error signal
- \( V_e(t) \) - temporal excitation
- \( V_i(t) \) - temporal inhibition
- \( U_e(x, y) \) - spatial excitation
- \( U_i(x, y) \) - spatial inhibition

Referring to the threshold curves reproduced in Figure 2.11 inhibition accounts for the low-frequency decline in the characteristic, with excitation responsible for producing the overall low-pass response.

If the spatial response is considered isotropic then the two spatial variables \( x \) and \( y \) reduce to a single distance \( \rho \) and frequency to a direction independent spatial frequency \( u \), where

\[
\rho = \sqrt{x^2 + y^2}, \quad u = \sqrt{\frac{x^2}{\rho^2} + \frac{y^2}{\rho^2}}
\]

(2.2)

The transfer function of the system of Figure 2.15 thus becomes

\[
S(u, f) = G_e(u)H_e(f) - G_i(u)H_i(f)
\]

(2.3)

where

\[
G_e(u) = \int_0^\infty U_e(\rho)2\pi\rho J_0(2\pi\rho u)\,d\rho
\]

(2.4)

\[
H_e(f) = \int_0^\infty V_e(t)e^{-j2\pi\eta t}\,dt
\]

(2.5)

\( f \) is temporal frequency, and \( J_0 \) is the Bessel function of order zero.

Budrikis [Budrikis 1973] has studied a range of mathematical functions for use in the terms above and has compared the resulting models with spatio-temporal threshold data from Robson [1960] and Kelly [1972].

A best fit was found with a diffusion like time function for excitation, a Gaussian time function for inhibition, and Cauchy space functions. Placing these functions into (2.3) we obtain

\[
|S(u, f)| = A\left(e^{-(\tau_1^2+\tau_2^2)}e^{-\sigma_2^2\eta^2} - ke^{-2\pi^2\eta^2}\right)
\]

(2.6)

where \( A, \tau_1, \tau_2, \sigma_2, \sigma_1, \) and \( k \) are the model parameters. Matching to Robson's data gave parameter values of

\[
A = 853
\]
2.8 MODELING THE EYE'S THRESHOLD CHARACTERISTICS

\[
\begin{align*}
\tau_1 &= 496 \text{ ms} \\
\tau_2 &= 98 \text{ ms} \\
\sigma_e &= 8.59 \text{ min arc} \\
\sigma_i &= 32.4 \text{ min arc} \\
k &= 0.677
\end{align*}
\]

A summary of the assumptions made by Budrikis in developing this model are listed below:

- The response is linear and time invariant.
- The response is isotropic (the same in all directions).
- The eye is stationary (not tracking anything in the image).
- The response is separable into independent spatial and temporal terms.

For the general case all these assumptions are clearly incorrect. However, for well defined test conditions the model above is expected to provide a reasonable approximation.

2.8.1 Modification of model for non-isotropic spatial response

The model developed by Budrikis may be made slightly more accurate by removing the assumption of an isotropic response. As described in section 2.2 the eye’s acuity for diagonal frequencies is around 20% less than for vertical or horizontal frequencies. Eqn. (2.2) may be modified to more closely represent the astigmatic response.

\[
u = \sqrt{v_x^2 + v_y^2} \left(1 + 0.2\left(\frac{1}{2} - \frac{1}{2}\cos(4\arctan(v_x/v_y))\right)\right)
\]

(2.7)

This eqn. causes a smooth change in the achromatic threshold as shown below in Figure 2.16 (a). Previously circular spatial equi-threshold contours now become rectangular shaped, reflecting the increased acuity in the horizontal and vertical directions (Figure 2.16 (b)).

![Figure 2.16 The modified non-isotropic spatial response (a) change in acuity with angle, (b) equi-threshold spatial contours.](image)

The choice of a cosine characteristic is arbitrary, and was chosen as a simple function providing a smooth transition between horizontal/vertical and diagonal frequencies. The exact characteristic is not known, but the error in this approximation is likely to be small.

Using this modification to (2.6) the luminance model may be plotted as a function of spatio-temporal frequency. Figure 2.17 shows two views of a 3-D contour of the HVS luminance threshold. Note the similarity to Figure 2.16(b) in the spatial dimensions, and the tapering of the response with increasing temporal frequency - leading finally to cutoff at high temporal frequencies.
2.8.2 A chromatic model

No information could be found on models for the chromatic response. Consequently one has been developed by the author.

As discussed in section 2.7 and illustrated by Figure 2.14 the chromatic thresholds have a purely low-pass characteristic. The response appears to consist of excitation with no inhibition, meaning less terms should be required. Assuming once again that spatial and temporal terms are separable the chromatic model will consist of two terms, a spatial and a temporal excitation.

A number of models were tried, based upon the Gaussian, Cauchy, Diffusion and Poissonian functions used by Budrikis. However, none were found to provide a particularly good fit to the data in Figure 2.14. In the end a different model of the form shown below was used with parameters $A, \tau_1, \tau_2, n_1, n_2$.

$$S(u, f) = A \left(10^{(\log(\tau_1 + 1))^{n_1}} 10^{(\log(\tau_2 + 1))^{n_2}}\right)$$ (2.8)

This model was fitted to the chromatic threshold data in Figure 2.14, recorded by van der Horst [1969]. Two sets of parameters were determined for the R-G, and Y-B paths. The parameters were found using a simple gradient based search and error measures similar to those used by Budrikiš - a closeness-of-fit measure $P$, and an rms model deviation $D$. These measures are defined as

$$P = \sum_{i=1}^{N} \left\{ \log\left( \frac{S(u_i, f_i)}{m_i} \right) \right\}^2$$ (2.9)

$$D = \sqrt{P/(N - 5)}$$ (2.10)

where $m_i$ is the recorded threshold, $N$ the number of experimental data points, and $N-5$ is used due to the parameters providing five degrees of freedom.

Using this approach a very good fit was obtained for both R-G and Y-B data. The final parameter values found for best fit were:

- **R-G**
  - $A = 0.876$
  - $\tau_1 = 0.612$ s
  - $\tau_2 = 0.621$ s
  - $n_1 = 4.44$
  - $n_2 = 2.85$

- **Y-B**
  - $A = 2.02$
  - $\tau_1 = 0.589$ s
  - $\tau_2 = 0.647$ s
  - $n_1 = 4.53$
  - $n_2 = 1.67$

The threshold curves defined by these models are plotted below, along with the original data points from Figure 2.14.
2.9 Picture quality evaluation

For image coding and transmission systems the quality of the final image, as perceived by the intended audience, is of high priority. It might seem that considerable understanding gained of the visual system should make it possible to determine and compare the expected quality of coding schemes. Unfortunately, this is not yet the case. The eye is only the first level of processing in a perceptual process that involves a large portion of the human brain. Although the physiology of the eye is well understood, the psychology of perception is not [Armington 1978, Uttal 1981]. The effects the higher levels of visual perception will have on perceived image quality are difficult to predict. As an example, a viewer is likely to find distortions in the shape of printed text more annoying than similar distortions in a picture of trees and forest. Similarly, hue errors in the colour of a human face will be far more disturbing than similar errors in the colour of a car or house in the same picture.

For these reasons final decisions on picture quality have always been made on the basis of subjective testing results - the judgements of a group of human observers. Although such subjective testing is still the method of choice, it’s use has many problems and work continues on development of improved objective methods. Such methods will allow accurate computer prediction of subjective picture quality without use of human observers.
Currently used subjective and some proposed objective assessment methods will now be looked at in detail.

2.9.1 Subjective assessment

At its simplest subjective assessment involves presenting a set of pictures (or sequences) to a group of human observers, and asking them to rate the quality. These ratings are recorded as a single number (e.g. one to five), representing a range from bad to good. This is a statistical process and the results have a statistical uncertainty that must be taken into account when interpreting them.

There are two main problems with subjective assessment.

Firstly, there are a great many variables in these tests.

- number of observers.
- viewing distance.
- display brightness and contrast.
- background illumination colour and level.
- viewing experience of observer.
- test pictures chosen.
- order and timing of presentation of pictures.

A sizeable difference in any one of these variables could render comparison of two sets of results misleading.

To maintain the consistency and repeatability of results all assessments should be carried out under similar conditions, with these conditions simulating as closely as possible the expected viewing conditions for the final image. CCIR Recommendations 500-4 "Method for the subjective assessment of the quality of television pictures" [CCIR 1990], defines in detail such a set of conditions. These recommendations were first put together in 1974, and have been updated regularly since. They have led to many groups that work on video and television development setting up special CCIR approved viewing rooms for carrying out subjective tests.

The second problem with subjective testing is obtaining judgements from the observers in a consistent manner. To this end the CCIR recommendations provide several multi-point quality scales to be used by observers for recording their ratings. Table 2.7 - Table 2.9 show these scales for three different rating systems.

<table>
<thead>
<tr>
<th>Quality scale</th>
<th>Excellent</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 2.7 The CCIR picture Quality scale, used for non-categorical overall judgment of picture quality.*

<table>
<thead>
<tr>
<th>Impairment scale</th>
<th>Imperceptible</th>
<th>Perceptible, but not annoying</th>
<th>Slightly annoying</th>
<th>Annoying</th>
<th>Very annoying</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 2.8 The CCIR Impairment scale, used for categorical judgements of stated impairments eg. cross-colour, blockiness, noise.*
2.9 PICTURE QUALITY EVALUATION

Table 2.9 The CCIR Comparison scale, used to judge the subjective difference between a reference picture and a modified version of the picture (either improved or degraded).

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>Much worse</td>
</tr>
<tr>
<td>-2</td>
<td>Worse</td>
</tr>
<tr>
<td>-1</td>
<td>Slightly worse</td>
</tr>
<tr>
<td>0</td>
<td>The same</td>
</tr>
<tr>
<td>1</td>
<td>Slightly better</td>
</tr>
<tr>
<td>2</td>
<td>Better</td>
</tr>
<tr>
<td>3</td>
<td>Much Better</td>
</tr>
</tbody>
</table>

Below is an example of a set of quality gradings, obtained for a comparison of Weston Clean PAL and conventional System I PAL [Oliphant 1980].

These results use the five point Quality scale of Table 2.7, applied to a set of four test images for the four enhanced and compatible paths. Added to these plots by the author are error bars showing the standard deviation for each result (taken from Oliphant's tabulated results). It is interesting to note that the standard deviation is as great, or greater than any quality improvements shown on these graphs. It could be said that, from a statistical perspective, the results of this subjective assessment are highly inconclusive.

Overall the problems with subjective assessment are many:

- Results are non-deterministic
- Results from one test cannot be reliably compared with those from another.
- Assessment provides a very simple qualitative measure only.
- Results provide no information on why a particular quality measure was obtained.
- Meaningful results rely upon the use of a statistically representative set of images.
- The assessment tests are expensive to set up, and labour intensive to carry out.

2.9.2 Objective assessment

Objective assessment is intended to solve the problems described in the previous section by providing automated, repeatable, and well defined methods for measuring picture quality. Such assessment techniques aim to eliminate the highly variable human observer, whilst providing results that can be used as a reliable predictor of traditional subjective test results.

This obviously requires the use of a model of the visual system in order to simulate a human observer. A number of such models have been developed [Budrikis 1972, Sarikson 1977, Granrath 1981, Lukas 1982]. The complexity of these models vary, but their general form is shown below.
In operating upon the image error these methods should provide results equivalent to the CCIR impairment scale shown in Table 2.8, rating the perceptibility of any errors/impairments in the output image.

In Figure 2.20 this is achieved by first filtering the error signal by a weighting function representing the eye’s ability to detect different artefacts (see section 2.8). The detectability of a visual stimulus is also affected by the detail in its surroundings, referred to as the image activity. Thus a masking signal, derived from measurements of global and local activity is then used to adjust the filter output.

More advanced techniques use a feature extraction system to isolate particular types of errors [Miyahara 1988, Voran 1991]. This works better due to the fact that some types of errors eg. contour errors, blocking artefacts, and motion jerkiness, are more objectionable than others, eg. random noise, blurring, and colour distortions.

Overall these objective assessment methods have been shown to be better indicators of coded picture quality than traditional measures, such as mean-square-error (MSE) or the signal-to-noise ratio (SNR) [Lukas 1982, Miyahara 1988, Lowe 1993, Mishina 1995]. However, their use as a replacement for subjective assessment is not yet assured. Later, in Chapter 5, a new approach to objective assessment will be described, developed by the author and using the HVS threshold models described in this chapter.

2.10 Summary

The aim of television is to provide a high fidelity picture to the human viewer. This Chapter has shown how television is designed to achieve this through a detailed understanding of the HVS. The characteristics of the HVS were used to predict the potential room for picture improvement in television. It is the aim of the enhanced coding techniques described in later chapters to achieve this potential. The author has also presented, based upon current experimental results, mathematical models for the (i) eye’s threshold response. These include new work with (i) the modification of Budrikis achromatic model to account for non-isotropic spatial response, and (ii) the development of a close fitting achromatic threshold model.

Finally an overview of subjective quality assessment has been given, explaining the many problems inherent in the techniques currently used.
CHAPTER 3

BACKGROUND

The introduction outlined historical background leading to the current interest in Enhanced Television techniques. This chapter provides the theoretical background. A full treatment of this theory is difficult to find in any one place, and the author certainly found some of the concepts involved took time to understand. For this reason this chapter attempts to give a relatively complete coverage of the field, not just the theory necessary for the author's work alone. It is intended that this chapter be useful in providing an introduction to the field for a newcomer, with an emphasis on linking the requirements of image quality with the nature of multi-dimensional sampling and spectral analysis.

Many of the properties of and artefacts in a television image are determined by the scanning and sampling processes used to convert the image into a signal. Section 3.1 looks at this process, and the complex properties that result from multi-dimensional non-orthogonal sampling patterns.

Sections 3.2 and 3.3 explain PAL coding and its effects upon the coded signal, and the resulting picture. Use is made of the newly developed analysis tools (Chapter 5) to show these effects in spatio-temporal frequency space. This understanding of PAL paves the way for the development of enhanced PAL techniques to reduce the artefacts present in conventional PAL. Section 3.4 describes the various advanced filter techniques (e.g. comb filter, complementary coding), that are used as building blocks in enhanced PAL coders.

To provide a proper overview of the field a passing reference is made to some areas not directly studied by the author, such as scan upconversion and non-orthogonal sampling techniques.

Throughout an emphasis is placed upon linking the 1, 2 and 3 dimensional views of the PAL signal.

3.1 The Spectrum of the Scanned Signal

If we consider the world as being four dimensional, with time being the fourth dimension, then television represents a three dimensional slice of this world. A television image is a function of two spatial and one temporal dimension, with the spatial dimension of depth being lost. Its extent in the two spatial dimensions is finite, but may be considered as infinite in the temporal dimension - our television receiver will continue to display changing images as long as they are being sent to it. These extents are shown diagrammatically by Figure 3.1 where A and B are the physical dimensions.

![Figure 3.1 Extents of television image](image-url)
As far as the television camera is concerned there is only one finite sized image - the rest of image space is ignored. However, for the purposes of studying the image mathematically it is useful to consider the rest of image space as being filled with copies of our transmitted signal, i.e. that the image repeats in the vertical and horizontal dimensions. This is represented mathematically by , with the result of this effect shown by Figure 3.2.

\[ e(x, y, t) = e(x - pA, y - qB, t) \]  

(3.1)

Figure 3.2 Television image as an infinitely repeating set of identical pictures

This assumption results in no loss of generality, and is useful in understanding the effects of scanning.

3.1.1 The spectrum of the image

The image function along a horizontal line passing through Figure 3.2 will be periodic, with period A. It is a known property of Fourier theory that such a periodic signal is entirely represented by harmonics of the base period. If we consider the brightness of any point on our line to be represented as \( e(x) \) then this function may be expressed as a Fourier series

\[ e(x) = \sum_{m,n} E\left(\frac{m}{A}\right) \exp j2\pi \left(\frac{mx}{A}\right) \quad m,n \text{ integer} \]  

(3.2)

where \( E\left(\frac{m}{A}\right) \) defines the spectrum of this function and A is the width of the image. Note that this spectrum is discrete, only integer multiples of \( 1/A \) are allowed. Incorporating the vertical dimension \( y \) and the height of the image B, \( e(x) \) becomes

\[ e(x, y) = \sum_{m,n} \sum_{\nu_y,n_y} E\left(\frac{m}{A},\frac{n}{B}\right) \exp j2\pi \left(\frac{mx}{A} + \frac{ny}{B}\right) \]  

(3.3)

Thus the spatial spectrum is entirely discrete with the vertical and horizontal frequencies obeying the relationship \( \nu_y = \frac{n}{B} \) and \( \nu_x = \frac{m}{A} \). The temporal dimension, \( t \), may now be added also, to obtain

\[ e(x, y, t) = \sum_{m,n} \sum_{\nu_y} \sum_{\nu_t} E\left(\frac{m}{A},\frac{n}{B},\nu_t\right) \exp j2\pi \left(\frac{mx}{A} + \frac{ny}{B} + \nu_t t\right) dt \]  

(3.4)

The image in the temporal dimension is not considered as being periodic, and consequently the temporal spectrum is continuous. From (3.4) the spectral components of the image must lie on lines in spectral space as shown by Figure 3.3.
3.1 THE SPECTRUM OF THE SCANNED SIGNAL

As a summary the related properties of the television image and its spectrum in the three image dimensions are:

<table>
<thead>
<tr>
<th>dimension</th>
<th>Image space extent</th>
<th>Spectral space</th>
</tr>
</thead>
<tbody>
<tr>
<td>horizontal</td>
<td>finite/periodic</td>
<td>discrete</td>
</tr>
<tr>
<td>vertical</td>
<td>finite/periodic</td>
<td>discrete</td>
</tr>
<tr>
<td>temporal</td>
<td>infinite</td>
<td>continuous</td>
</tr>
</tbody>
</table>

Table 3.1 Related properties of image and spectral space for television.

3.1.2 Scanning the Image

For transmission, the three dimensional television image must be converted into a one dimensional signal. This is achieved by scanning. The scanning process, performed by the television camera, converts the three dimensional image into a sequence of frames, similar to the picture frames on a movie film. These frames are then split into a sequence of lines. The transmitted signal then consists of all these lines placed end to end to produce a continuous signal. In the receiver this signal is combined again into lines and fields which are displayed on screen, where they will reconstruct the appearance of the original image.

3.1.2.1 Practical scanning

Scanning is a sampling process, with each sample or scan theoretically representing an instantaneous point or line in spatio-temporal space. However, due to the nature of scanning equipment this is not completely the case. In a television camera each frame is obtained by integrating the image brightness over the frame period (or a portion thereof), and each scan line is obtained by integrating over an area of spatial image space determined by the size and shape of the scanning aperture.

This non-ideal scanning action can be viewed as an ideal scanning action that is performed on a transformed version of the image. The image transformation produced by the non-ideal scanning has been shown to be a filtering operation whose characteristics are determined by the aperture shape and size and the integration period [Mertz 1934]. In practice this filtering action has a low-pass effect, one which benefits the scanning process by reducing higher frequencies that would otherwise cause aliasing, but which also contributes to a reduction in picture resolution.

For the purposes of studying the scanned image the scanning operation will be considered as an ideal process, but one operating on a spatio-temporally pre-filtered image.
3.1.2.2 Progressive scan

The camera scanning spot moves in both the vertical and horizontal directions, tracing out a scanning path. The resulting scanning pattern for a progressive scanned image is shown by Figure 3.4(a).

![Figure 3.4 Image scanning path for progressive scan.](image)

Note that, unlike film, each frame of a television scanned image is not a 'snapshot' in time. The bottom corner of the frame is scanned some time after the top corner has been scanned. Similarly each line is slanted in the temporal dimension. This produces the skewed three dimensional scanning pattern shown in Figure 3.4(b).

3.1.2.3 Interlaced scan

Progressive scan is the simplest form of scanning, but most television systems use a slightly different method, called interlaced scan. For interlaced scan the image is scanned vertically at twice the rate, but half the number of lines per vertical scan. If the number of lines in each picture is equal to an integer plus a half then every two vertical scans will interlace. This effect is shown in Figure 3.5(a).

For a stationary picture - one unchanging in time - interlaced scan is almost equivalent to splitting each progressive scanned frame into two fields, one containing the odd lines, the other the even lines.

Interlaced scan has the advantage of allowing the vertical scan rate to be doubled, without increasing the line rate. This increases the resolution in the temporal dimension, but at the expense of some vertical resolution (in moving images).
3.1 THE SPECTRUM OF THE SCANNED SIGNAL

3.1.3 The spectrum of a scanned signal

In section 2.2.1 the image was described as repeating in the spatial dimensions. Using this assumption the movement of the scanning spot in the spatial dimensions can be described by the equations

\[ x = A f_h t \]  (3.5)
\[ y = -B f_v t \]  (3.6)

where \( f_h \) and \( f_v \) are the horizontal and vertical scanning frequencies and \( A \) and \( B \) are the image dimensions. The negative sign in (3.6) is due to the downward direction of the scanning. The path described by these functions in image space is shown in Figure 3.6.

![Figure 3.6 Representation of scanning path as a continuous line on the repeating image.](image)

By substituting equations (3.5) and (3.6) into (3.2) the spatial image is converted into a one dimensional temporal signal \( e_s(t) \).

\[ e_s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} E \left( \frac{m}{A}, \frac{n}{B} \right) \exp j2\pi (mf_h - nf_v) t \quad m, n \text{ integer} \]  (3.7)

From (3.7) a spatial frequency with \( m \) cycles/picture width and \( n \) cycles/picture height corresponds to a signal frequency of \( v_s \) Hz

\[ v_s = mf_h - nf_v \]  (3.8)

For an interlaced scan image with \( N \) lines/picture height (\( N \) must be odd) \( f_v \) and \( f_h \) are related by (3.9)

\[ f_v = \frac{2f_h}{N} \quad N \text{ integer} \]  (3.9)

Using (3.8) and (3.9) the signal spectrum for a scanned signal can be predicted. As the horizontal scanning frequency \( f_h \) is much higher than \( f_v \) the spectrum will consist of fundamental components at multiples of \( f_h \), with each of these surrounded by a series of components lying at multiples of \( f_v \). The one dimensional signal spectra for progressive and interlaced scan are thus shown in Figure 3.7.
Chapter 2: Background

3.1.4 The spectrum of a scanned signal in two dimensions

Because scanning is a sampling process it gives rise to periodically repeated spectra in Fourier space. Converting the signal spectrum to horizontal and vertical frequency axes, denoted $v_h$ and $v_v$, the repeat spectrum positions may be found for values of $m, n$ where equation (3.8) is zero.

Using (3.9) this requirement becomes

$$m - \frac{2n}{N} = 0$$

(3.10)

$$m - \frac{n}{N} = 0$$

(3.11)

for the interlaced and progressive scan systems respectively. Thus the repeat spectrum positions are shown by Figure 3.9.
3.1 THE SPECTRUM OF THE SCANNED SIGNAL

Figure 3.9 Repeat spectra positions (zero frequency) for (a) progressive scan, (b) interlaced scan.

Note that the horizontal scale is highly magnified and for large $N$, the repeat positions may be considered to be almost entirely in the vertical direction.

As with any sampled system there is a maximum frequency, the Nyquist frequency, that can be represented without aliasing. For the scanned image this limit exists only in the vertical direction, and is equal to $N/2$. The horizontal frequency is limited only by the bandwidth of the scanned signal.

3.1.5 The spectrum of a scanned signal in three dimensions

The two dimensional spectrum is only valid for still images with no temporal components. Equations (3.5) and (3.6) describe the scanning spot as also moving in the temporal direction. The spatio-temporal spectrum is thus fully described by

$$e_s(t) = \sum m \sum n \int E \left( \frac{m}{A}, \frac{n}{B}, v_i \right) \exp j2\pi \left( mf_h - nf_v + v_i \right) t \, dv_i$$

A spatio-temporal frequency of $m$ cycles/picture width, $n$ cycles/picture height and moving at $v_i$ corresponds to a signal frequency $v_s$.

$$v_s = mf_h - nf_v + v_i \tag{3.13}$$

The repeat spectra positions are again the solutions of $v_i = 0$ in (3.13). For $v_i = 0$ these positions are the same as the two dimensional case. Non-zero values of $v_i$ result in the two dimensional repeat spectra being repeated again, but with a temporal offset. Rearranging (3.13) and using (3.7) an equation predicting the repeat spectra temporal offsets is produced

$$v_s = \left( 2n - mN \right) \frac{f_v}{2} \tag{3.14}$$

As $m$, $n$ and $N$ are all integer the repeat spectra positions are at multiples of half the frame or picture frequency (the vertical scanning frequency). These spectrum centres all lie on a plane whose equation is given by (3.14), and illustrated by Figure 3.10.
3.1.6 The Nyquist limit

The repeat spectra positions lie very nearly in the vertical temporal plane. The scanning operation may thus be seen as sampling the image in the vertical and temporal dimensions. Consequently there will be a Nyquist limit for these dimensions. The Nyquist limit of the signal is a two dimensional function, however it is not unique, a fact discussed in detail in section 3.1.6. The standard interpretation of this limit for progressive and interlace scan is shown in Figure 3.11.

For an equivalent line rate ($f_h$ the same) an interlaced scan exchanges high vertical-temporal resolution for greater temporal resolution. The shape of the Nyquist limit for the interlaced scanned image also better matches the shape of spatio-temporal response of the eye (see 2.6), but only in the vertical-temporal dimensions.

A summary of the related properties of the television image and its spectrum in the three image dimensions are:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Image</th>
<th>Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>horizontal</td>
<td>continuous</td>
<td>not repeated</td>
</tr>
<tr>
<td>vertical</td>
<td>sampled</td>
<td>repeated</td>
</tr>
<tr>
<td>temporal</td>
<td>sampled</td>
<td>repeated</td>
</tr>
</tbody>
</table>

Table 3.2 Related properties of image and spectral space due to scanning
3.1 THE SPECTRUM OF THE SCANNED SIGNAL

The complete characteristics of the scanned image spectrum are found by combining the effects of scanning with those caused by the finite spatial image size (Table 3.2 and Table 3.1). We thus end up with different properties in each dimension. The horizontal spectrum is discrete and non-periodic, the vertical spectrum is discrete and periodic, and the temporal spectrum is continuous and periodic.

3.1.7 The interpretation of movement in three dimensional spectral space

Temporal frequencies in an image can be caused by the varying brightness of an object, e.g. a flashing light. Most, however, are the result of movement. To gain an understanding as to how movement effects the spectrum of the image consider the spatial frequency signal

\[ e(x, y, t) = \exp j2\pi \left( \frac{mx}{A} + \frac{ny}{B} + v_t t \right) \quad m, n \text{ integer} \]  

(3.15)

for \( v_t = 0 \). This frequency will thus lie entirely in the \( n, m \) plane in spectral space. If there is motion, described by

\[
\begin{align*}
x &= ut \\
y &= vt
\end{align*}
\]  

(3.16) \hspace{1cm} (3.17)

then (3.15) becomes

\[ e(x, y, t) = \exp j2\pi \left( \frac{mu}{A} + \frac{nv}{B} \right) t \]  

(3.18)

Thus a temporal frequency component of

\[ v_t = \left( \frac{mu}{A} + \frac{nv}{B} \right) \]  

(3.19)

has been introduced. Rearranging (3.19) we obtain an equation describing a plane passing though the origin.

\[
\begin{align*}
m \frac{u}{A} + n \frac{v}{B} - v_t &= 0 \\
m, n &\text{ integer}
\end{align*}
\]  

(3.20)

Thus when the image is stationary the image spectrum lies entirely in the \( m, n \) plane. When movement occurs the spectrum shifts to lie on a tilted plane whose angle of tilt is proportional to the velocity, and whose direction of tilt is in the direction of motion.

These properties are made use of in motion-adaptive coding, with the filtering being altered to track the shifting spatio-temporal spectrum of the image.

![Figure 3.12 Plane in which spatio-temporal frequency components lie for a (a) stationary image (b) moving image.](image)
3.1.8 Sampling the scanned signal.

To apply digital filtering or to simulate any scanning system sampling must be performed in the third dimension as well. This introduces spectral repetition to the scanning signal spectrum. If $f_s$ is the sampling frequency and $p$ an integer then the scanning signal spectrum will obey the relationship

$$E(v_s) = E(v_s - pf_s)$$

$p$ integer (3.21)

Spectral repetition has thus been introduced in the horizontal frequency dimension as well. Equation (3.13) now becomes

$$v_s - pf_s = mf_h - nf_s + v_r$$

$$v_r = mf_h - nf_s + v_s + pf_s$$

$m, n, p$ integer (3.22)

Just as equation (3.13) describes a plane upon which the repeat spectra lie, (3.22) describes a series of planes, obtained by varying $p$.

Depending on the choice of $f_s$ the repeat spectrum can become quite complex.

3.1.8.1 Orthogonal sampling and the primitive bandwidth

Equation (3.23) describes a relationship between the sampling frequency $f_s$ and the line frequency $f_h$

$$f_s = hf_h$$

If $h$ is an integer an orthogonal sampling pattern results, with an integer number of samples per line. The European digital standard [CCIR Rec 601-1 1986] uses a sampling frequency of $f_s = 13.5$ MHz. For standard PAL this is equivalent to exactly 864 samples per line (720 active), i.e. the sampling is orthogonal. For this example each set of repeat spectra described by (3.22) is offset in the horizontal direction by $m = 864$. One of the effects of this sampling is to introduce a Nyquist constraint to the horizontal dimension.

In three dimensions the Nyquist constraints define a surface around the repeat spectrum position. This surface contains those frequencies that will not cause aliasing and will be referred to as the Nyquist boundary. Parallels have been drawn between three dimensional sampling and the theory of X-ray crystallography [Tonge 1981], showing that the Nyquist boundary is the same as the first Brillouin zone of the sampling lattice. Tonge refers to the volume of spectral space enclosed by this surface as the primitive bandwidth of the resulting signal.

The Nyquist boundary containing the primitive bandwidth is not uniquely defined. For any boundary the only requirements are that the volume contained always remain same, and that the shape defined be capable of tessellating, without gaps or overlap, when placed at the repeat spectrum positions. This property also holds for two dimensional sampling, where the Nyquist boundary is also non-unique. As an illustration Figure 3.13 shows three different vertical-temporal boundaries that could be used for interlaced scanning.

Choosing the best shape to use will depend upon several factors.

(i) The required bandwidth for the signal in the different dimensions. For example, Figure 3.13(c) doubles the temporal resolution, at the expense of a loss in high vertical-temporal frequencies.

(ii) The filtering requirements for producing such a shape. Keeping the filtering as simple as possible, meaning a Nyquist boundary that is the product of several vertical/horizontal and temporal filters is often the best. Filtering for Figure 3.13(a) is likely to be more complex than that for (b) and (c).
3.1.8.2 Quincunx sampling

If \( h \) of equation ((3.23)) is non-integer the sampling pattern will no longer be orthogonal. Some of these non-orthogonal patterns are of special interest and are compared with orthogonal sampling in Figure 3.14.

![Figure 3.14](image)

Figure 3.14 The sampling patterns produced by (a) orthogonal sampling, (b) line quincunx, (c) field quincunx.

Line-quincunx (samples offset on alternate lines) uses a sampling frequency that produces a \( \frac{1}{2} \) sample offset every alternate line, i.e.

\[
f_s = \left( k + \frac{1}{2} \right) f_h \quad k \text{ integer}
\]

Field-quincunx (samples offset on alternate fields) has a \( \frac{1}{2} \) sample offset every alternate field, with a relationship

\[
f_s = kf_h + \left( l + \frac{1}{2} \right) f_v \quad k, l \text{ integer}
\]

These offsets mean line and field quincunx follow a four field sampling pattern (orthogonal interlaced sampling follows a two field pattern).

The different sampling patterns result in different repeat spectrum positions, and these in turn allow the definition of different alias boundaries. Field and line-quincunx sampling, for example, allow more of the primitive bandwidth of the signal to be concentrated in low spatial-temporal frequencies, as well as allowing higher frequencies along the vertical, horizontal and temporal axes. Figure 3.15 compares the alias boundaries that could be used for orthogonal and quincunx sampling. Note that line-quincunx sampling, in particular, allows the Nyquist boundary to become a shape similar to that expected for orthogonal sampling.
of the eye's spatio-temporal response (refer Figure 2.17). This means that better use is made of the available signal bandwidth. Similar to the two dimensional cases shown in Figure 3.13 these boundary surfaces must also tessellate, but in three dimensional space.

![Figure 3.15 The conventionally used spatio-temporal alias boundaries for (a) orthogonal sampling, (b) line quincunx](image)

### 3.1.9 Amplitude Modulation in three dimensions

Amplitude modulation is the process by which a linear frequency shift is applied to a signal. For a function \( g(t) \), which has the Fourier transform \( G(f) \), amplitude modulation may be represented as

\[
G_m(f) = \frac{1}{2} G(f) \left( \delta(f - f_m) + \delta(f + f_m) \right)
\]

where \( G_m(f) \) is the modulated signal and \( f_m \) is the modulating frequency. The result is to produce two copies of the baseband spectrum, centred on the positive and negative modulating frequencies. Generalising to more than one dimension scalar frequency is replaced with a vector \( \mathbf{v} = [v_x, v_y, v_z, ...] \). Thus (3.26) becomes

\[
G_m(\mathbf{v}) = \frac{1}{2} G(\mathbf{v}) \left( \delta(\mathbf{v} - \mathbf{v}_m) + \delta(\mathbf{v} + \mathbf{v}_m) \right)
\]

In multi-dimensional modulation the signal spectrum will be shifted to positions given by the vector \( \pm \mathbf{v}_m \).

For a scanned three dimensional image any modulation performed on the scanned signal can be mapped to a three-dimensional translation vector \( \mathbf{v}_m = [v_h, v_y, v_z] \) (or \( \mathbf{v}_m = [m, \hat{n}, \hat{v}] \) for the scanned signal). The components of this vector for an interface scanned image can be found by solving equation (3.13) for the signal modulating frequency \( v_s = f_m \). Consequently any operation performed upon the scanned television signal may be represented in terms of its spatio-temporal effects through use of this transformation vector.

### 3.2 PAL Television

The PAL television standard describes the requirements for the encoding and decoding of a scanned colour image and its consequential transmission and decoding. All PAL colour systems combine their colour signals in the same manner and use a standard field rate of \( f_v = 50 \) fields/s with \( N = 625 \) lines per frame. Transmission bandwidth does, however, vary with the UK standard being 5.5 MHz (System I PAL), whereas New Zealand uses a lower 5 MHz (System B, G). Other minor differences in PAL exist for different countries [CCIR Report 624-3].

#### 3.2.1.1 The PAL Coder

Figure 3.16 shows a block diagram of the PAL coder. As well as the colour picture signals field, line and sub-carrier synchronisation signals must be added. A sound signal must also be added. In studying the effects of the encoding process on the image the only section of real interest is the
3.2 PAL TELEVISION

colour encoding, shown within the dotted line. The synchronisation signals are required to enable the receiver to display the lines and fields in their correct positions on screen, but, under normal circumstances, have no effect on the picture. For detailed information on these synchronisation signals and the PAL standard refer to [TVNZ 1982, CCIR Report 624-3].

![Block diagram of PAL encoder.](image)

The colour encoding consists of quadrature modulating the $U$ & $V$ components with the chrominance sub-carrier and combining this signal with the luminance. Before combination some simple filtering is performed upon the $YUV$ components. A notch filter is optionally used to attenuate luminance information near the chrominance sub-carrier frequency - an attempt to reduce cross-colour (luminance to chrominance crosstalk). The chrominance is low-pass filtered to a horizontal bandwidth of approximately 1.5 MHz, compared to the 5-5.5 MHz bandwidth of the luminance. This filtering reduces the overlap of chrominance and luminance, and places the chrominance signal high in the luminance spectrum, which helps to limit the effects of cross-luminance (chrominance-to-luminance cross-talk). Synchronisation and sound signals are now added to produce a composite video signal, which is low-pass filtered to meet the channel bandwidth requirements, producing the baseband PAL signal spectrum, shown in Figure 3.17.

![Baseband spectrum of the PAL signal.](image)

Before modulation the $V$ component is multiplied by the PAL switch signal $s(t)$. This signal causes the $V$ component to be inverted on alternate lines.

The equation describing the final PAL signal (ignoring filtering) is thus

$$PAL = Y' + U' \sin(2\pi f_c t) + s(t)V' \cos(2\pi f_c t)$$ (3.28)

where $Y'$, $U'$ and $V'$ are the gamma corrected $YUV$ signals. Refer to Appendix A for the definition of these signals and their derivation from the $RGB$ components.
3.2.1.2 Choice of Sub-carrier

The sub-carrier frequency is calculated from the equation

\[ f_{sc} = \left(284 - \frac{1}{2}\right) f_h + \frac{f_y}{2} \]  

(3.29)

The choice of the sub-carrier frequency was governed mostly by monochrome compatibility requirements. To this end the sub-carrier is placed as high in the luminance spectrum as possible. This reduces the visibility of cross-luminance which will now appear only as high luminance frequencies - with these high frequencies being less noticeable to the eye. The quarter line offset causes a phase change from line to line, and the half field frequency offset causes a phase reversal every field. This results in the sub-carrier pattern due to cross-luminance being different on alternate lines and fields, with 8 fields required before the pattern repeats. These phase changes cause the interference to visually cancel out to a large degree.

If we map the sub-carrier signal frequency onto spatio-temporal frequency space using equation (3.13) for interlaced scanning we obtain the spatio-temporal vectors

\[ v_{sc} = \pm \left(284, 78, 18 \frac{1}{4}\right) \]

\[ \pm \left(283, -235, 6 \frac{1}{4}\right) \]  

(3.30)

It can thus be seen that from a spatio-temporal point of view \( f_{sc} \) consists of high vertical, horizontal and temporal offsets. In spatio-temporal frequency space \( f_{sc} \) is therefore placed at a position corresponding to the lowest response of the eye. This explains the apparent visual cancellation of the sub-carrier pattern.

3.2.2 The PAL Spectrum in one dimension

From equation (3.30) the spectrum of the PAL signal will consist of a baseband luminance signal plus a quadrature modulated chrominance spectrum, as shown in Figure 3.17. In close-up the luminance spectrum will appear as in Figure 3.7(b) with the \( U \) component spectrum being similar, but with its centre shifted to a horizontal position of \( f_{sc} \). The \( V \) component, however, is first convolved with \( S(f) \), the spectrum of the PAL switching signal \( s(t) \), before modulation by \( f_{sc} \). The modulation of the \( V \) component is equivalent to a modulation by a series of frequencies which are the sub-carrier \( f_{sc} \) convolved with the spectrum components of \( s(t) \). The \( V \) sub-carrier is thus:

\[ f_{scv} = \pm (2r + 1) \frac{f_h}{2} \]  

(3.31)

The main effect of this is to offset the \( V \) harmonics by \( \frac{1}{2} f_h \) from the \( U \) harmonics. It also repeats the \( V \) harmonics with an offset of \( f_h \), with the result that higher \( V \) harmonics are repeated on top of lower harmonics and vice-versa.

An alternative interpretation of the effect of \( s(t) \) is as a sampled sine wave whose period is 2 lines ( \( 2/f_h \) ) and where the samples occur at 90° and 270°. The result is the modulation of the \( V \) component, causing a frequency offset of \( \pm \frac{1}{2} f_h \).

Looking at the PAL spectrum near the sub-carrier the \( U \) & \( V \) line harmonic centres are seen to be offset from the \( Y \) line harmonic centres by \( \pm \frac{1}{4} f_h \). Looking closer still, the \( U \) & \( V \) picture harmonic centres are offset from the \( Y \) picture harmonic centres by \( \pm \frac{1}{4} f_p \) (\( f_p = f_s/2 \)).
Thus the U & V components interleave with the Y component for all horizontal and vertical frequencies. However, when temporal frequencies are taken into account the positions of the horizontal/vertical harmonics may be shifted by a temporal offset, the maximum being ±½fp. This range is indicated in Figure 3.18 for Y, U and V. It can be seen that there is an overlap in range between the different components. Consequently, it is possible, with certain temporal frequencies, for the luminance and chrominance harmonic components to coincide. For these spatio-temporal frequencies the luminance and chrominance become indistinguishable, and potentially inseparable.

3.2.3 The PAL spectrum in two dimensions

Looking at the PAL spectrum in the spatial frequency domain only, the luminance spectrum will consist of a series of repeat spectra, repeating in the vertical frequency dimension. In the horizontal dimension the maximum extent is determined by the maximum signal bandwidth, with the relationship m_{max} = v_{Smax}/f_h. For a 5 MHz signal with f_h = 15625 Hz this gives a maximum of m = 320 cycles per picture width. Ignoring the temporal offset of the sub-carrier, the first repeat positions of the U component will be at (284, 78) and (283, -235). Further repeat positions are found using the repeat vector (2, 625).

The repeat positions of the V spectrum are different, and may be found by substituting equation (3.31) into the scanning equation (3.13) that maps signal frequency to spatio-temporal frequency. Ignoring the temporal offsets again the first repeat positions will be at (284, 234) and (283, -79). Once again further repeat positions are found by applying the repeat vector (2, 625).
The $U$ and $V$ spectrum positions thus interleave, and the resulting PAL spatial frequency spectrum is shown in Figure 3.19.

Figure 3.19 Spatial-frequency spectrum of PAL signal for $r=0$. (NOTE: Shapes of spectrums shown are completely arbitrary and are intended to indicate relative position only).

3.2.4 The PAL spectrum in three dimensions

In three dimensions the luminance spectrum repeats according to equation (3.14), which describes the spectrum of an interlace scanned image. The $U$ & $V$ spectrum positions also repeat in a similar manner, but their basic repeat vectors now have temporal offsets as well.

$$U \begin{pmatrix} 284, 78, 18 \frac{1}{2} \\ 283, -234, -64 \frac{1}{2} \end{pmatrix} \quad \quad V \begin{pmatrix} 283, 79, -18 \frac{1}{4} \\ 284, 234, 6 \frac{1}{4} \end{pmatrix}$$

The resulting spectrum, viewed along each axis, and in three dimensional perspective, is shown by Figure 3.20(a)-(d). These illustrate the relative positions of the $Y$, $U$ and $V$ component spectra (with the shapes of the $U$ and $V$ spectra being chosen arbitrarily). The full spectrum is obtained when these spectra are repeated in the vertical-temporal dimensions, i.e. the luminance bound in Figure 3.20 is repeated at the points shown in Figure 3.10.
3.2 PAL TELEVISION

Figure 3.20 Different views of the three dimensional PAL spectrum, showing positional relationships for first spectrum repeat.

It should be noted that, contrary to the appearance of Figure 3.19, there are actually only two $U$ and two $V$ spectrum repeats for each $Y$ spectrum. The extra $U$ and $V$ spectra in Figure 3.19 overlap from neighbouring repeats of the baseband spectrum.

These figures confirm the comment made earlier that the $U$ & $V$ sub-carrier positions are placed on the very edge of the luminance alias boundary surface.

3.2.5 The PAL Decoder

Figure 3.21 shows a block diagram of the PAL decoder. As with the encoder the area of particular interest is the colour decoding section. In this section the luminance and chrominance are separated, by use of notch and bandpass filters, and the chrominance is demodulated to produce the $U$ and $V$ baseband signals. The technique used to decode the chrominance is of special interest. Before demodulation the $U$ and $V$ are passed through a delay whose length is not quite equal to one line. The output of this delay, called the PAL delay line, is added to the undelayed signal. Due to the inversion of the $V$ sub-carrier on alternate lines this action helps to cancel out sub-carrier phase errors, preventing them from causing picture hue errors. Exactly how this works is explained in Appendix A.
3.2.5.1 The response of the PAL delay line.

The PAL delay line is actually a simple filter, and consequently has an effect upon the frequency content of the chrominance signals. The PAL delay line belongs to a family of filters called comb filters, which are described in detail in section 3.4.2. The response of the filter is similar to that of a one line delay comb filter, but has a slant in the spatial dimensions - due to the filter delay being just less than one line. The spatial passband (defined as response > 0.707) for the filter, as applied to the $U$ and $V$ components is shown in Figure 3.22 (a) and (b).

From the positions of the stopbands shown in these figures it can be seen that a degree of $U$, $V$ separation is achieved. The delay line also attenuates the intermediate vertical frequencies of the $U$ and $V$ components - thus causing some picture resolution loss. This reduction in vertical colour resolution is, however, relatively small compared to the reduced horizontal resolution of the chrominance, which is less than 1/4 that of the luminance. Figure 3.23 shows a -3dB contour (half power) for the full spatio-temporal frequency response of Figure 3.22 (a), (b).
3.2.6 The response of the PAL system

The use of the channel filter combined with the luminance notch causes a reduction of the horizontal luminance bandwidth to well below the theoretical maximum of 320 cycles/width. Similarly the chrominance is restricted by the chrominance low-pass filters (and to some degree by the channel filter), meaning a maximum horizontal chrominance frequency of around 75 cycles/width. Unlike the luminance the chrominance is restricted vertically as well (to approx. 78 cycles/height), due to the action of the PAL delay-line.

The resulting response of the PAL system (encoder plus decoder) is shown in Figure 3.24 (a) - (c)

Figure 3.24 A -3dB spatio-temporal contour plot of the PAL delay line response for (a) U component, (b) V component

Figure 3.23 A -3dB spatio-temporal contour plot of the PAL delay line response for (a) U component, (b) V component
CHAPTER 2 BACKGROUND

These spectrum plots are based on a simulation of PAL colour encoding and decoding using TVPROC (see Appendix E), and were obtained using the spectral analysis techniques described in Chapter 4.

3.3 The Image Degradations Introduced by the PAL System.

There is a range of artefacts or image degradations present in the PAL system.

3.3.1 Scanning effects

From Nyquist theory a band-limited signal is fully represented by samples taken at a rate equal to or greater than the maximum frequency content. From these samples it is theoretically possible to reconstruct the original signal exactly by passing the samples through an ideal reconstruction filter (a low pass filter with infinite rolloff). In practice an ideal reconstruction is never possible, and the reconstructed signal will be different from the original i.e. artefacts will be introduced.

The transmitted television signal is a sampled version (in vertical and temporal dimensions) of the original video image. Ideally, this signal should be reconstructed at the receiver to produce the original picture, with the accuracy of the reconstruction filter determining the level of any artefacts introduced. In practice, the standard television receiver makes no attempt at reconstruction. Instead the sampled lines and fields are simply displayed directly on screen. The reason for this lack of reconstruction is that television relies upon the characteristics of the human eye to act as a reconstruction filter. As shown by 2.12 of section 2.6 the eye's response is effectively a low-pass characteristic, which is the basic requirement for a reconstruction filter.

The idea of using the eye to reconstruct the image is soundly based (as explained in Chapter 2), and greatly reduces the complexity of receiver that would otherwise be required. However the characteristic of the human eye is very complex - it is definitely not an ideal reconstruction filter. A further point is that the lowpass cut-off of the eye is actually too high to properly reconstruct the scanned image. This is illustrated by Figure 3.25, which shows a response contour for the eye, placed upon the spectrum of the scanned signal.

\[
\begin{align*}
\text{Figure 3.25} & \quad \text{The areas of the interlaced signal spectrum that are visible to the human eye.}
\end{align*}
\]

Section 2.8.1 showed that the shape of this contour in the vertical-temporal dimensions is diamond shaped, however the size shown is fairly arbitrary, and depends upon the level of response we are talking about. Suffice to say that Figure 3.25 shows a contour encompassing spatio-temporal frequencies that are quite noticeable to the eye.

The response contour not only encompasses the wanted baseband signal, but a number of the repeat spectra as well. This causes several artefacts to be perceived in the displayed image.
3.3 THE IMAGE DEGRADATIONS INTRODUCED BY THE PAL SYSTEM.

1. **Large Area Flicker**: The flicker fusion frequency for the eye is higher than television’s 25 Hz field rate, particularly in the eye’s peripheral vision. Thus the existence of spectra B is seen as a large area flicker of the screen.

2. **Inter-line flicker and line crawl**: Horizontal edges on objects and fine horizontal lines are often perceived as flashing, or moving, due to the effects of interlace. These artefacts are caused by spectra C.

3. **Line structure and Kell effect**: There is often a loss of sharpness in the vertical direction, with the subjective vertical resolution being far less than might be expected, an effect first explained in 1934 by Kell [Kell et al 1940]. This vertical resolution loss is consequently known as the Kell effect, and is due largely to the existence of spectra A (although B and C also contribute). For an interlaced display the Kell factor (ratio of perceived vertical resolution to actual resolution) varies from 0.5 to 0.7.

### 3.3.1.1 Aliasing

Aliasing is another scanning artefact present in television. Real world images contain a practically unlimited range of spatio-temporal frequencies, meaning that the limits of the spatio-temporal spectrum are not defined, and, in fact, far exceed that capable of being represented within the **primitive bandwidth** of the scanned signal. In practice these higher frequencies are statistically less likely to occur. Added to this the scanning characteristics, as explained in section 2.2, tend to have a low-pass characteristic attenuating frequencies outside the primitive bandwidth. However, any parts of the spectrum outside the **primitive bandwidth** will cause aliasing, which appears as image artefacts. Moving detail, for example, will cause temporal aliasing.

Aliasing is becoming more of a problem today as modern cameras offer improved resolution. This extra resolution is sometimes greater than that which can be contained within the primitive bandwidth of the interlaced signal. When this point is reached image quality becomes worse instead of better, due to the aliasing artefacts that are introduced.

### 3.3.1.2 Scan conversion

Many of these artefacts can be removed by introducing image pre-filtering before transmission and image reconstruction in the receiver. Pre-filtering removes frequencies that cause aliasing and can allow the use of sophisticated sampling patterns, such as line-quincunx sampling, to provide increased horizontal resolution. In the receiver conversion from interlace to progressive scan will reduce interline flicker, and vertical interpolation will remove the Kell effect. Such techniques are being proposed [Okada et al 1985, Biaslo et al 1990, Hentshel 1989] as methods for improving the resolution of conventional television to compete with and display HDTV images.

### 3.3.2 Cross-effects

The interference between the components of the PAL signal is one of the most objectionable defects of PAL television. As explained in section 2.3 the PAL encoding process is essentially non-reversible, which means that it is impossible to recover the original colour components exactly, the result being cross-effects.

There are two main types of cross-effect - **cross-luminance** (chrominance incorrectly decoded as luminance) and **cross-chrominance** (luminance incorrectly decoded as chrominance).

### 3.3.2.1 Cross luminance

Figure 3.26 shows how cross-luminance arises in the one dimensional spectrum of the PAL signal. The use of a low chrominance bandwidth plus a luminance decoder notch filter helps to reduce cross-luminance. However, these constraints are applied only in the horizontal dimension by simple filters operating on the scanned signal. Figure 3.27 shows the cross-luminance component from a simulated assessment of New Zealand (5 MHz) PAL. Figure 3.27 shows the cross-luminance caused independently by the U and V components, and tell us which spatio-temporal frequencies will appear in the luminance spectrum due to interference from the chrominance.
3.3.2.2 Cross chrominance

Cross-chrominance as it appears in the one dimensional spectrum is shown by Figure 3.28. Notch filtering the luminance signal before encoding is sometimes used to reduce the possible cross-chrominance by attenuating luminance that is close to the sub-carrier. Despite this, cross-chrominance still occurs, and the spatio-temporal cross-component frequencies, found by simulation of a standard New Zealand PAL system, are shown by Figure 3.29. Due to the different weights given to $U$ and $V$ in the PAL signal the $U$ cross-component is greater.
3.3 THE IMAGE DEGRADATIONS INTRODUCED BY THE PAL SYSTEM.

PAL signal spectrum

![Image of PAL signal spectrum]

**Figure 3.28** The occurrence of cross-chrominance in the PAL decoded chrominance signal.

**Figure 3.29** The cross-chrominance for New Zealand PAL caused by the luminance for (a) the $U$ component, (b) the $V$ component.

### 3.3.2.3 Inter-chrominance (or $U/V$ cross-talk)

For most television transmission vestigial sideband modulation is used. This can be used because only one of the luminance sidebands is required in order to recover the luminance signal. However, to recover a quadrature signal both sidebands are required. This means that for a chrominance bandwidth of 1 - 1.5 MHz the baseband signal bandwidth should be 5.4 - 5.9 MHz to avoid removing any part of the chrominance upper sideband. This requirement is met for the System I PAL used in Britain, which has a transmission bandwidth of 5.5 MHz. With New Zealand PAL, however, the transmission bandwidth is 5 MHz (System B, G), which means that the upper chrominance sideband is cut in half. Upon decoding the loss of part of one of the chrominance sidebands causes the confusion of the $U$ & $V$ components for any chrominance signal frequencies above 0.53 MHz.

For a simulation of standard New Zealand PAL this effect can be seen in the spatio-temporal spectrum of the $U$ and $V$ components.
CHAPTER 2  BACKGROUND

Due to the different weighting’s given to $U$ & $V$ in producing PAL this cross-effect appears greater in the $V \rightarrow U$ direction, with no cross-effect in the $U \rightarrow V$ direction being visible at -20 dB (although it is visible at -30 dB). Note that these thresholds are calculated relative to the maximum amplitudes of the $U$ and $V$ signals. In three-dimensional frequency space inter-chrominance occurs for any spatio-temporal frequency with a horizontal component of $m > 36$ c/pw (approx). As well as causing inter-chrominance this also causes an attenuation of the wanted chrominance. In the television picture these effects appear as hue errors in areas of colour detail.

3.3.3 Resolution Loss

Some aspects of the PAL system that affect resolution have already been discussed in section 2.3. The filters used in the PAL encoder and decoder also have an effect. For example, the luminance notch filter used to reduce cross-chrominance also reduces the effective horizontal luminance bandwidth to approximately 3.5 MHz or 224 c/pw, considerably lower than the original monochrome television system. Similarly inter-chrominance and the PAL line delay both contribute to a decrease in chrominance resolution.

3.3.4 Phase Errors

The PAL system is designed especially to be robust to phase errors. However in the PAL delay line decoder phase errors will still cause a slight saturation error, even if no hue errors. In the PAL-S receiver (no delay line) the hue errors on alternate lines cancel visually, instead of in the delay line. The saturation error however, which is also different on alternate lines, is more noticeable, producing an effect of vertical light and dark bars, referred to as Hanover Bars.

3.3.5 Enhanced PAL

The aim of Enhanced PAL is to reduce many of the artefacts in the current signal that have been described in this section. By use of different filtering and modulation techniques it is possible to reduce component cross-effects, whilst at the same time increase the perceived luminance and chrominance resolution by better matching the spectral characteristics of the signals to the eye’s spatio-temporal response. These improvements need be achieved without sacrificing PAL’s robustness to chrominance phase errors. Hue errors and saturation errors (Hanover bars) should be suppressed as well or better than in a conventional PAL delay line decoder. Chapter 4 describes a range of techniques that have been developed for achieving these aims, and a new technique developed by the author will be presented in Chapter 7.
3.4 Video Filters and Filtering Techniques

Many of the approaches to enhanced PAL achieve their signal improvements by use of filters applied to the PAL or individual YUV components. Conventionally, any filtering performed upon a video signal is done using standard analogue filter types, applied directly to the analogue video signal. Due to the nature of the scanned signal these filters can only act upon the horizontal dimension of the three dimensional video image. The vertical and temporal dimensions have been effectively sampled, meaning that a discrete filter must be used instead. Because the signal being processed is actually still analogue it is necessary to use a type of filter called a tapped delay line. This filter makes use of a line of fixed analogue delays to cause, at any one point in time, the signal to line up either vertically or temporally.

The tapped delay line filter is often referred to as a digital filter, as all digital filtering must use a filter of the form of equation B.1 (See Appendix B). However, this filter can be equally well applied to an analogue signal. The PAL delay line is an example of an analogue tapped delay line filter, with a block delay almost equal to one line. For digital input the delay must obviously be equal to an integer number of sampling periods ie $T_d = nT_s$; or if the delay is fixed then the sampling frequency must be chosen so as to meet this requirement. Tapped delay line filters can be divided into two types, IIR or recursive filters and non-recursive or FIR filters. A FIR filter differs from the IIR filter in having no feedback elements. Appendix B describes these filter types in more detail, and explains their differences in performance.

To date the use of FIR type filters has dominated in video processing applications. This is mainly due to their stability and linear phase characteristics. However, in video applications the order of the filter can be critical, with high speed delay storage, digital multipliers and adders being very expensive. IIR filters offer potentially the same filter performance, or better, for a much lower order filter (up to an order of magnitude less taps), and consequently a lower cost. For this reason their increased use in future applications is quite probable, and is an area for future investigation.

3.4.1 Designing FIR filters for television filtering

Filtering of the three dimensional television image may be performed in any one of the three image dimensions, or alternatively in several at once, using a multi-dimensional filter.

In most television systems horizontal filtering is performed using analogue filters. This can be done as the scanned video image is a continuous signal in the horizontal dimension.

For advanced digital systems, or for simulation of a system the scanned signal is usually sampled in the horizontal dimension as well. Any horizontal analogue filter must therefore be converted to an equivalent or similar IIR or FIR model. IIR filters are commonly used as digital equivalents to analogue filter types. However, due to the stability and relative simplicity of FIR filters all simulations performed by the author use FIR filters as approximations to the analogue filters. The design method use for FIR filters is usually the Fourier window technique. This technique is explained in Appendix B.

3.4.1.1 Vertical and temporal filters for an interlaced signal

Vertical and temporal filters may be designed using the Fourier window technique also. However, as the order of these filters in any practical application is critical (due to the large storage requirements for line and field delays), the window technique is seldom used. The response of these filters is also altered by the fact that the signal being filtered is interlaced.

In interlaced scanning the image is sampled in the vertical and temporal dimensions as shown by Figure 3.31.
Thus any vertical or temporal filter must have taps that lie on this sampling grid, with the filter delays for the different taps being multiples of the picture line time.

3.4.1.2 Modification of response due to interlace

The result of any operation performed upon the scanned signal is different from what might normally be expected, due to the fact that the sampling in the vertical-temporal direction is non-orthogonal. This difference amounts to a 'squashing' of the response in the vertical and temporal dimensions. i.e.

$$H'(v_v,v_t) = H\left(\frac{v_v}{2}, \frac{v_t}{2}\right)$$

This 'squashing' creates interesting effects due to the existence of repeat spectra in the spectrum of the sampled signal. Figure 3.32 illustrates this, showing the response of a low-pass filter applied to an interlaced signal.

The placement of the repeat spectra in an interlaced spectrum causes any vertical or temporal low-pass filter to confuse the low and high frequency components. This results in the filter passing high frequencies, as well as low frequencies. The filter function thus becomes a notch filter, instead of a low pass, with a response that is symmetrical around half the folding frequency.

3.4.1.3 Multi-dimensional filters

Designing a multi-dimensional filter can be more difficult - due, once again to the non-orthogonal nature of the scan sampling. If the scanned signal is sampled using a line or field quincunx sampling pattern it becomes even more complicated still, as the sample points do not line up horizontally either.

Two approaches to designing multi-dimensional filters are available.

1. The filter may be designed for an orthogonal sampling pattern that overlaps with the interlaced pattern. The coefficients for those points not shared by both patterns are then set to zero, producing a filter that can be implemented using the interlaced signal. This
action will, of course, alter the filter's response. The way in which the response is changed is similar to the way the filter response was altered by interlace in Figure 3.32.

2. A second method is to design the filter for an orthogonal set of sampling points, then rotate the coefficients by 45°, such that they then lie on the sampling grid of Figure 3.33 (b). By this technique a diamond filter for restricting the vertical and temporal frequencies of the interlaced signal can be designed by rotating a vertical/temporal low pass filter.

3.4.1.4 Variables Separable Filters

Multi-dimensional filters are far more complex than one dimensional ones. The number of taps for the same order filter becomes squared (or cubed), for example a 2nd order two dimensional filter requires 25 instead of 5 taps. An alternative approach to producing a required multi-dimensional characteristic is to combine several simpler filters. This method can be used if the required characteristic is variables separable, which means that it can be separated into the product of several independent characteristics. An example of this is a vertical/temporal low pass filter, which can be created by separately applying a vertical, then a temporal low pass filter. Another example is the filter characteristic shown in Figure 3.33 (b), which could be created by cascading two diagonal filters.

![Figure 3.33](image)

*Figure 3.33 The diamond filter for an interlaced signal in (b) may be obtained by a 45° rotation of the coefficients for the orthogonal lowpass filter in (a).*

3.4.2 Comb filters

The most popular type of filter used in advanced video applications is the comb filter. This filter is one of the simplest examples of a FIR filter, and is shown in its two classic forms in Figure 3.34 (a) and (b).

![Figure 3.34](image)

*Figure 3.34 Block diagram of comb filters - (a) 1 delay, (b) 2 delay.*

The amplitude response of each of these two filters is
\[ |H_1(v_c)| = \sqrt{\frac{1 + \cos(2\pi v_c T_d)}{2}} \]  
\[ |H_2(v_c)| = \frac{1 + \cos(2\pi v_c T_d)}{2} \]  

This response may be viewed in a number of different ways, (i) against one dimensional signal frequency \( v_d \) (Figure 3.35 with \( T_d = t_{\text{line}} \)), (ii) against vertical-temporal frequency (Figure 3.36), or (iii) against spatio-temporal (three dimensional) frequency, producing a picture similar to Figure 3.23. Compared to the 1 delay filter the 2 delay filter offers a slightly better characteristic, with a sharper response and more clearly defined passbands and stopbands. Another advantage of the 2 delay filter is its amplitude symmetry - meaning that the filter inverse has an identically shaped characteristic, and possesses skew-symmetry.

Figure 3.35 Amplitude response of (a) 1 delay comb filter, (b) 2 delay comb filter, plotted against signal frequency for \( T_d = t_{\text{line}} \).

Figure 3.36 Amplitude response of (a) 1 delay comb filter, (b) 2 delay comb filter, plotted against vertical/temporal frequency for \( T_d = t_{\text{line}} \).
The comb filter takes its name from the appearance of the characteristic when viewed against signal frequency. As shown by Figure 3.35 the filter appears to have many teeth, which are used to comb out harmonics in the spectrum that correspond to different vertical & temporal frequency groups (see Figure 3.18).

### 3.4.2.1 Basic comb filter Types

There are four main types of comb filters used in television processing, their difference being their delay periods and consequently their spatio-temporal response.

Figure 3.37 and Figure 3.38 show these filters - their tap positions and coefficients - and their corresponding vertical-temporal pass-bands. These are two dimensional contour plots, with the contour taken at -6dB ($H(v_y) = 0.5$).

For standard PAL with $N = 625$ the filter delays are of 1 line, 312 lines and 313 lines and 625 lines (1 frame) respectively. Doubling the delay times i.e. doubling the spacing of the filter taps, causes the filter characteristic to change. This change is equivalent to ‘squashing’ the response along the line in which the filter taps are applied.

As noted earlier the responses of purely vertical or temporal filters are modified by the effects of interlace.

Also shown in Figure 3.37 and Figure 3.38 are the positions of the $U$ and $V$ component sub-carriers. These serve to illustrate how well suited many of the comb filters are for use in the encoding and decoding of the PAL signal. Particularly suited are the 2, 313 and 1250 line filters. These three do a good job of selecting low frequency luminance, whilst rejecting the $U/V$ sub-carrier positions.

Most enhanced PAL techniques employ one or more of these filters, with some of the more advanced techniques using combinations, usually in a variables separable approach.
Figure 3.37 Frequency response and tap coefficients for 1 delay comb filter with delays of (a) 1 line, (b) 312 lines, (c) 313 lines, (d) 625 lines.
Figure 3.38 Frequency response and tap coefficients for 1 delay comb filter with delays of (a) 2 lines, (b) 624 lines, (c) 626 lines, (d) 1250 lines.
3.4.3 Complementary coding and decoding
The most common method for combining two signals without causing them to interfere, and in such a way that they can be individually recovered is complementary encoding and decoding. One filter is used in such a way that the characteristic applied to the first signal is the complement of that applied to the second signal. To achieve this the filter F must be designed to have clear stop-bands and passbands, with the result that wherever one signal exists the other will be excluded.

For PAL the two signals normally used in this technique are the luminance ($Y$) and chrominance ($C$) with the encoded result being represented by the equation

$$S = FC + (1-F)Y$$  \hspace{1cm} (3.35)

where $S$ is the PAL signal. Decoding is achieved by applying the same complementary filtering to the encoded signal, producing a separation of the two signals due to the fact that $(1-F) \times F \equiv 0$. In practice there will be residual cross-effects due to the transition band between stop-band and pass-band. The shorter this transition band, the lower the cross-effects (but the higher the filter order, and the more likely it is to produce ringing or motion artefacts).

The technique is illustrated by the block diagram in Figure 3.39 below.

![Figure 3.39 Block diagram of complementary encoding and decoding filters.](image)

3.5 Summary
This chapter has provided the necessary background for the understanding and development of enhanced PAL coding techniques. This has included the basic theory of PAL television and the historically used views of the PAL signal. The problems with the conventional delay line PAL system have been covered in detail, indicating the areas the enhanced coding techniques will try to improve. Throughout the author has placed an emphasis on viewing the image and PAL signal as a 3 dimensional spatio-temporal signal, with 3-dimensional spectral analysis being, in the author's opinion, the most appropriate approach to a complete understanding of PAL and image coding effects.

It has been noticeable in past treatment of the field that much confusion has existed in understanding the PAL signal. Past analysis has explained coding effects in terms of line-to-line phase differences, the positions of line and field harmonics, 'spectrum folding' and by use of horizontal or horizontal/vertical only spectra. It is the author's contention that these approaches have been inadequate and have hindered proper understanding. Many functions (e.g. a comb filter) are greatly simplified when viewed in 3 dimensions. Consequently this chapter has attempted to generalise all past views to the 3 dimensional spatio-temporal signal. The resultant representation of the PAL signal developed in this chapter will be used throughout the rest of this thesis as the basis for designing, describing and analysing television coding techniques.
CHAPTER 4

ENHANCED PAL CODING TECHNIQUES

Enhanced television has been an area of study for many years. Since very shortly after the introduction of colour, methods for improving picture quality further have been investigated. Unfortunately, in the work on enhanced compatible systems there has not been a co-ordinated approach. Many different techniques have been investigated by different researchers, (Lent 1974, Drewery 1976, Weston 1977, Holoch 1985, Teichner 1985) and explained in many different ways. This has made the field rather confusing, with the same techniques explained differently (e.g. in terms of line-to-line phase changes, one dimensional frequency interleaving, or three dimensional spectrums), and also implemented differently (e.g. sampling vs. modulation for Weston Clean PAL).

In view of these facts the author has in this Chapter attempted to put all known enhanced PAL techniques into a common framework. The operation of these techniques is explained in terms of the multi-dimensional spectral theory developed in the preceding Chapter. Many of the techniques have not been compared with each other elsewhere, and it is hoped that this summary of the field will cause the similarities and advantages of the different techniques to stand out.

4.1 The aim of Enhanced PAL

The aim of Enhanced PAL is to provide an improved quality picture through the use of better encoding and decoding methods, whilst still retaining standard PAL compatibility. These aims can be presented more specifically as:

- Reduction/elimination of component cross-effects.
- Maintenance or improvement of picture resolution.
- Compatibility of the enhanced PAL signal with a standard PAL decoder.
- Compatibility of enhanced PAL decoders with standard PAL signals.

The last two are very important. The greatest improvements to picture quality are likely to come from the use of a matched encoder and decoder, meaning modifications to both the encoding and decoding methods. Practically speaking, however, the transmission of any new encoded signal is likely to have a large market of standard PAL receivers to begin with. Similarly, any new decoder brought onto the market will likely still be receiving a standard PAL signal from many transmission sources. Consequently, in developing an enhanced PAL technique, there are four situations to be considered, for the four encoder-to-decoder paths. These four paths are shown below in Figure 4.1.

Referring to this diagram there is the obvious requirement that path (4), the enhanced path, should result in improved picture quality compared to path (1), the standard path. The compatible paths should offer quality not much worse than the standard path, with path (2) the Enhanced decoder path preferably offering an identical or improved quality picture. The whole issue of compatibility for enhanced television is very similar to the situation faced with the introduction of colour in the 1960's.
4.1 Compatibility - the requirements

The basic requirements for maintaining compatibility can be summarised as:

- The Y, U, and V signals must be used.
- Luminance must be a baseband signal.
- The colour difference signals should appear quadrature modulated by the sub-carrier frequency.
- The PAL line switching sense must be adhered to.

4.1.2 The approach for enhanced PAL

The basic problem with standard PAL is that there are three signals trying to share the same spectral space. This is obviously impossible to achieve in a reversible manner, using normal amplitude modulation techniques. The task of the enhanced PAL techniques is to find a way of segregating the Y and C (U+V) signals - by spectral band, phase or time, whilst allowing them to still 'appear' to be occupying the same spectral space in the manner of PAL. The area of conflict for these three signals within the PAL spectrum is only the portion close to the chrominance sub-carrier - the area where spectral overlap of Y and C occurs. For this reason the clever encoding and decoding techniques of the enhanced techniques are only ever applied to this region of the signal. This usually means the use of a band-pass or high-pass filter centred on the chrominance sub-carrier frequency. (As the chrominance overlaps the luminance right to the top of the luminance spectrum the band-pass and high-pass filters become effectively equivalent).

Figure 4.2 Simplified block diagrams of (a) standard PAL encoder, (b) standard PAL decoder.
4.1 THE AIM OF ENHANCED PAL

The standard PAL encoder and decoder and the general form for the enhanced PAL coder are shown in Figure 4.2 as simplified block diagrams (With the $U$, $V$ low pass filters translated to become a chrominance band pass).

These figures are much simplified, and are arranged to emphasise the main differences in the enhanced coder, compared with a standard coder. These differences are (i) the division of $Y$ into high and low frequency luminance, and (ii) the use of special coding for the chrominance and high frequency luminance.

![Figure 4.2 Simplified block diagrams for the general form of (a) enhanced PAL encoder, (b) enhanced PAL decoder.](image)

In practice the block diagrams of the systems usually appear quite different, despite producing similar signals. The differences are often due to attempts at minimising the number of processing blocks required. A simple example of this is shown below in Figure 4.4, where for Figure 4.3 the luminance notch filter may be obtained simply by subtracting the band pass filter output from the luminance (complementary filtering). A further reduction of blocks would be obtained by using the same bandpass filter for both $Y$ and $C$ paths.

![Figure 4.4 Use of complementary filtering to create a notch filter from a bandpass filter.](image)

4.1.3 Grouping the coding techniques

The special coding techniques applied to the luminance and chrominance in Figure 4.3 are designed to allow better separation of these in the decoder. The approaches used to achieve this may be broadly divided into three categories.

1. Band segregation.
2. Time segregation.
3. Phase segregation.
These can each be implemented in either of two forms:

1. Non-adaptive coding.

For adaptive coding the nature of the segregation is modified according to the content of the current signal (e.g. movement).

Many enhanced coding methods actually use a combination of these techniques. In fact, horizontal band segregation is used in all systems, as shown in Figure 4.3.

These different coding approaches will be examined in detail in this Chapter, with examples of their use from the current literature. Each has its own advantages and problems, and these will be discussed.

### 4.2 Band Segregation

Band segregation is one of the oldest enhanced PAL techniques [Drewery 1976, Auty 1977], and has been studied in the greatest detail. The segregation of luminance and chrominance is normally achieved by complementary filtering, a concept explained in section 3.4.3. This allows a single band segregating filter to be used for filtering both the luminance and the chrominance. The general form for band segregation is shown below.

The idea is that the portion of frequency space the chrominance occupies is restricted, and the luminance signal can then be placed everywhere the chrominance isn’t. This means that luminance and chrominance never occupy the same frequency space. Ideally the band segregating filter $F$ is equal to 1 at the frequencies we want chrominance (the pass-band), and 0 everywhere else (the stop-band). This means that

$$PAL = Y_L + [FC + (1-F)Y_H]$$

Where $Y_L$ is low frequency luminance and $Y_H$ is high frequency luminance, defined by the use of the chrominance bandpass filter $B$

$$Y_L = (1-B)Y, \quad Y_H = BY$$

Upon decoding
4.2 BAND SEGREGATION

\[ C' = F \cdot PAL = F^2 C + F(1-F) Y_H \]  
\[ Y_H' = (1-F) \cdot PAL = (1-F) F C + (1-F)^2 Y_H \]

As ideally \( F(1-F) = 0 \) and \( F^2 = F \) the luminance and chrominance will be recovered without interference or distortion. In practice, of course, \( F \) is not ideal, and will contain a finite sized transition band. This means that

(i) Residual cross-effects will still occur in the transition bands of \( F \).

(ii) \( Y_H \) and \( C \) are attenuated at frequencies near the edges of the pass-band. However, sharp filter cut-offs can cause objectionable ringing in images.

Better results can often be obtained using non-complementary filters. In this case separately optimised filters are used on \( Y_H \) and \( C \) independently. These filters are complementary in their pass-bands and stop-bands, but not in the transition regions. The improvements gained by this method are compared in Figure 4.6.

![Figure 4.6 Reduction in cross-effect obtained by using non-complementary filters (a) complementary \( F \), \((1-F)\) filters and cross-effects, (b) non-complementary filters \( F \), \( G \) and cross-effects.](image)

Non-complementary filtering can allow further reduction of cross-signals, without having to increase filter cutoff, but at the expense of a loss in the pass-band width for the filters.

The simplest form of band segregation is that used currently in conventional PAL. This is a very primitive form of non-complementary segregation, using the chrominance bandpass and luminance notch. As was shown in section 3.2.2 it does not work very effectively. This is the main reason for the search for better segregation methods. Early attempts at finding a better set of horizontal filters [Lent 1974] did not offer any significant improvements. The real breakthroughs came with the use of multi-dimensional filtering [Drewery 1976], and the application of comb filters, illustrated in Figure 4.7. Using filters that segregated the luminance and chrominance in terms of their vertical and temporal image frequencies provided a greatly increased frequency space in which segregation could be applied.
4.2.1 Two dimensional band segregation

Band segregation can be divided into two areas, for two and three-dimensional segregation. Two dimensional band segregation combines the use of the horizontal bandpass (F\textsubscript{1} in Figure 4.5) with a comb filter (F\textsubscript{2}).

\[ \begin{array}{c}
  \text{X} \\
  \downarrow \text{.5} \\
  \downarrow \text{.5} \\
  \downarrow \text{.25} \\
  \downarrow \text{.25}
\end{array} \]

(a) \hspace{1cm}
\[ \begin{array}{c}
  \text{X} \\
  \downarrow \text{.5} \\
  \downarrow \text{.5} \\
  \downarrow \text{.25} \\
  \downarrow \text{.25}
\end{array} \]

(b)

Figure 4.7 Block diagram of comb filters - (a) 1 delay, (b) 2 delay.

The comb filter operates in vertical/temporal frequency space, and has a one dimensional characteristic. Although the characteristic is one dimensional there is freedom in its orientation relative to the vertical and temporal frequency axes.

An important attribute of comb filters is their use of line and field delays to allow application of three-dimensional filtering to the one-dimensional PAL signal (as explained in detail in Chapter 3). Because of this the filters are expensive to implement, and there is a high incentive to keep them simple.

It is generally taken that the most important image information is contained in the lower spatial-temporal frequencies [Eckert, 1992, Cortelazzo et al, 1991]. Consequently the filter F should be chosen to maximise the chrominance spectral area around the U and V sub-carriers, whilst also preserving low vertical/temporal luminance. The four most commonly used comb filter orientations for achieving this are shown below [Clarke 1988], where the crosses indicate the U and V sub-carrier positions.

As illustrated in Figure 4.8 (a), the purely vertical comb filter is the most economical in terms of delays, and has been studied extensively [Auty 1977, Drewery 1976, Clarke 1982 (June, July), Teichner 1986, Poz 1990], and has found some application in professional coders. Very recently vertical comb filters have even been used in top-of-the-line domestic decoders.

The disadvantage of the vertical comb is the considerable restriction of vertical luminance and chrominance frequencies, to less than 12% of their full resolution. (Although, in comparison with a normal delay-line decoder the reduction is only around 50% (2-tap comb) or 36% (3-tap comb)).

Diagonal vertical-temporal filtering is offered by 312- and 313-line field based combs (Figure 4.8 (b), (c)). Of these two the 312-line comb has be the better characteristic, and arguably offers the best preservation of wanted luminance and chrominance of all four types of comb filter. Because of this, and despite its need for 312 times as many delay elements the 312-line comb filter has also been studied in detail [Drewery 1976, Clarke 1982, 1988, Teichner 1986, Poz 1990]. This type of field based filtering has also been used in some top-of-the-line professional coders.

The fourth purely temporal characteristic involves the greatest number of delays (four times as much as the 312-line comb). This filter can offer perfect separation, and full vertical resolution for stationary images. However, the temporal resolution is restricted to only 2-3 Hz (instead of 25 Hz). This filter has been studied [Clarke 1982 (June, July), Teichner 1985, Poz 1990], but has generally been found to have unacceptable motion performance. In particular it causes colour smearing with movement.

For each of these filter types, the order of the filter may be increased to improve separation and resolution by sharpening the transition region. Up to a point this provides improved performance. As an example Poz [1990] found that a 5-tap vertical comb performed better than a 3-tap comb. Filters of much higher order, however, offer diminishing returns, and can cause visible ringing in the picture. Consequently higher order filters have not been investigated in detail in available literature.
4.2 BAND SEGREGATION

4.2.2 Three-dimensional band segregation

Using a two-dimensional vertical-temporal filter in combination with the horizontal bandpass could theoretically provide the greatest freedom for optimisation of chrominance and luminance responses. The simplest approach is simply to combine two one-dimensional comb filters. Teichner [1988] has studied the two most obvious such filters.

The first is a combination of the 1-line and 625-line combs (Figure 4.8 (a) & (d)), the second a combination of 312-line and 313-line combs.
These filters have been studied by others in various forms [Drewery 1976, Clarke 1988, Cortelazzo 1990, Poz 1990]. Looking at Figure 4.9 it can be seen that, once again, the field based filter appears best matched to the PAL spectrum and the positions of the U and V sub-carriers. (Note that these diagrams, unlike Figure 4.8, show the upper right quadrant only.) The line and frame based filter of Figure 4.9 (a) has four passband areas with the important chrominance only being centred in two of these. As a result spectral space is wasted, and the chrominance frequencies are restricted more than for the field based filter.

Instead of placing two filters in series we can use a single two-dimensional filter, with the tap values found by convolving the two one-dimensional filters. The tap values for the two-dimensional versions of the filters above are thus

![Diagrams showing tap values for two-dimensional filters](image)

These diagrams show another advantage of the 312/313-line delay filter - its use of a fewer overall delays. The total delay is 1876 lines, compared to 2504 lines for the 1/625-line filter.

Modified, or higher order versions of these filters are possible, although, as before, a point of diminishing returns is quickly reached. High order two-dimensional filters require enormous delays, and can introduce ringing and motion artefacts. Some have been studied [Cortelazzo 1990, Clarke 1988], but have generally not provided great enough improvements to justify the extra complexity.

Due to the relative simplicity of most of the filters non-complementary filtering has been little used. In one trial of non-complementary filtering [Clarke 1988] it did not prove to offer noticeable advantages. Non-complementary filtering is, however, used in the BBC's highly complex enhanced PAL proposal [Croll 1992] which combines band and phase segregation (see section 4.4).
4.2.3 Compatibility of band segregation techniques
As all the band segregation techniques are basically just extensions of the filtering mechanism used in standard PAL they all generate compatible PAL signals. In most cases the E-PAL→PAL path results in a slightly improved picture compared to the PAL→PAL path, at the cost of a more restricted chrominance bandwidth.

For the PAL→E-PAL path the picture will also be slightly improved compared to the standard path. Reasonable levels of cross-effects will still be present, having been introduced by the PAL encoder in a process that is non-reversible. However, the band segregation E-PAL decoder is much better matched to the PAL spectrum, allowing Y and C to be better separated, with a reduced loss of luminance resolution.

The E-PAL→E-PAL path will obviously provide the greatest improvements, as long as the same type of segregating filters are used in encoder and decoder.

The fact that all three enhanced paths can offer improvements is the reason why band-segregating techniques have already been adopted for use in conventional systems, despite no common type of band segregation being agreed on.

4.2.4 Problems with band segregation
A summary of the problems with band segregation as an enhanced technique is:

- It restricts the allowable chrominance frequencies to a narrower range than PAL.
- The quality of Y, C separation is dependant upon the filter cutoff. If the band segregation filter characteristics are not sufficiently sharp cross-effects will still exist for the filter transition region.
- Using higher order filters to improve filter sharpness may cause ringing in the picture.
- The best filters require use of considerable storage (from 1-3 frames).

4.3 Time segregation
Time segregation allows separation of Y_H and C by transmitting them at different times. If they never exist in the PAL signal concurrently they cannot interfere with each other. This principle is illustrated below for a segregation interval of T seconds.

The decoder must then store Y_H and C and use this stored information to replace the missing signals. This process is shown below, and generalised block diagrams of a time segregating encoder and decoder are shown in Figure 4.13.

*Figure 4.11 The principle of time segregation with luminance and chrominance transmitted in separate time intervals.*
For an enhanced system this technique can achieve perfect separation of luminance and chrominance. The disadvantage is that time segregation requires that half the $Y_H$ and $C$ information be discarded, and in the decoder repetition of these signals will cause picture errors. These problems can be minimised by choosing $T$ equal to a line, field or frame period. Within any television image the information in neighbouring areas (either vertically or temporally) is highly correlated. The changes between two lines or fields are normally quite small and thus repetition of alternate lines, fields or frames should result in minimal error. In practice disturbing errors can still be introduced, when picture content is changing quickly along the vertical or temporal dimensions.

4.3.1 I-PAL - Line based segregation

A time segregated signal allows an enhanced decoder to achieve perfect separation. However, the resulting PAL signal is not entirely compatible with a conventional PAL decoder. This decoder expects chrominance and high frequency luminance on all lines/fields, not just some. The PAL decoder does, however, have a simple form of chrominance storage with the chrominance delay line. An enhanced technique labelled I-PAL [Holoch 1985] takes advantage of this by using line based segregation with $T = t_{line}$. The PAL delay line will act so as to store and repeat the chrominance, producing a correctly decoded PAL signal. Although this means that I-PAL is acceptable as a compatible enhanced system, it still has a number of problems.

(i) The PAL delay line is slightly less than one line.

(ii) The amplitude of the chrominance signal will be halved, implying a loss in colour saturation.
(iii) Lack of chrominance on alternate lines means the delay line can no longer be used to correct phase errors.
(iv) High frequency luminance is still missing on alternate lines

Luckily there are a number of mitigating factors.
(i) The delay line is close enough to a line to make no difference.
(ii) Loss of colour saturation may be compensated by colour adjustment on the receiver, at the expense of a 6 dB reduction in signal-to-noise.
(iii) Modern television systems suffer much less from phase errors than when PAL was first introduced.
(iv) Most high frequency luminance is removed by the notch filter anyway.

4.3.2 Improving compatibility
A suggestion made by Holoch [1985] for improving compatibility was to add repetition of the chrominance to the encoder. The encoding and decoding sequence shown in Figure 4.12 then becomes

<table>
<thead>
<tr>
<th>Line No</th>
<th>Encoder</th>
<th>Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$C_1$</td>
<td>$C_1$</td>
</tr>
<tr>
<td></td>
<td>store</td>
<td>store</td>
</tr>
<tr>
<td>2</td>
<td>$C_1$</td>
<td>$C_1$</td>
</tr>
<tr>
<td></td>
<td>$Y_{1H2}$</td>
<td>$Y_{1H2}$</td>
</tr>
<tr>
<td></td>
<td>store</td>
<td>store</td>
</tr>
<tr>
<td>3</td>
<td>$C_3$</td>
<td>$C_3$</td>
</tr>
<tr>
<td></td>
<td>store</td>
<td>store</td>
</tr>
<tr>
<td>4</td>
<td>$C_3$</td>
<td>$C_3$</td>
</tr>
<tr>
<td></td>
<td>$Y_{1H4}$</td>
<td>$Y_{1H4}$</td>
</tr>
<tr>
<td></td>
<td>store</td>
<td>store</td>
</tr>
<tr>
<td>5</td>
<td>$C_4$</td>
<td>$C_5$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Y_{1H4}$</td>
</tr>
</tbody>
</table>

*Figure 4.14 Improved time segregation coding technique, transmitting chrominance on all lines.*

This modification requires that a line store be added to the encoder as well. In the decoder the identical chrominance on alternate lines can be subtracted to recover $Y_H$. As chrominance is transmitted on all lines now the problems stated as (ii) and (iii) previously disappear. This modified version of I-PAL is labelled I-PAL-M by Holoch.

4.3.3 Field based segregation
I-PAL-M still leaves $Y_H$ missing on alternate lines. Ideally we need to include both $C$ and $Y_H$ on all lines, but still be able to separate them. One way to achieve this would be by inversion of one of the signals in alternate intervals. As shown below this would allow full separation of $Y_H$ and $C$, even though both are always present.

The only problem with this idea is that inverting one of the signals on alternate lines is unlikely to be compatible with conventional decoders. Luckily, however, there is one case for which compatibility can be maintained perfectly. A field delay of 312 lines is almost exactly equal to an odd number of sub-carrier half cycles ($312*\text{tine} = 88530.4992$ cycles). This means that in standard PAL modulated chrominance will normally appear inverted when viewed across a 312 line delay. Consequently, for $T = 312*\text{tine}$ the encoder sequence in Figure 4.15 will appear completely compatible to a standard PAL decoder. Antony Dean [Dean 1992], working with the author at the University of Canterbury recognised this fact, and designed the field based system shown below.

This technique was labelled CAFPAL (Colour Alternate Field PAL) [Dean 1993], as it is based upon the same principal as I-PAL. Strictly speaking colour information exists on all fields, however, the coding principle works on alternate fields.
4.3.4 Reducing sub-sampling and repetition errors

As mentioned previously the loss of information and signal repetition in time segregation can introduce picture errors. These errors are of two types.

(i) Aliasing, due to discarding of information on alternate time intervals $T_i$ i.e. subjective-sampling.

(ii) Attenuation of high frequencies due to repetition of decoded $Y_H$ and $C$ signals.

Looking first at (i), the sub-sampling process is illustrated Figure 4.17, along with its effects on the spectrum of the signals.
4.3 TIME SEGREGATION

The sub-sampling is effectively performed along a line through the sampled vertical-temporal \((m, n)\) space. Thus the aliasing shown above will occur along this line. Where

\[ T = t_{\text{line}} \]  
Aliasing of high vertical frequencies to low vertical frequencies.

\[ T = t_{\text{field}} \]  
Aliasing of high vertical-temporal frequencies to low vertical-temporal frequencies.

The obvious way to prevent this aliasing is by pre-filtering the signal prior to the sub-sampling process, limiting the maximum frequency to \(1/(4T)\).

The pre-filtering must be applied along the same vertical-temporal axis as the sub-sampling and, as shown, prevents aliasing by stopping overlap of the spectra. CAFPAL and ColourPlus both use such pre-filtering, and a CAFPAL encoder using a 2 tap pre-filter is shown below.

This 2-tap pre-filter has the vertical-temporal characteristic shown earlier in Figure 4.8 (b). Along its line of operation in vertical-temporal frequency space it has the response shown in Figure 4.20.
As can be seen this pre-filter still allows considerable residual aliasing. Higher order filters can be used to reduce this aliasing further.

Ideally a similar post-filter should be used in the decoder to reconstruct the original signal. This post-filter would remove the aliased spectra placed at odd multiples of \(1/(2T)\), and interpolate the signal to provide \( Y_H \) and \( C \) for the missing intervals. Once again higher order filters would provide a better reconstruction, with less residual aliasing. In practice CAFPAL and ColourPlus just use repetition of the signal. This simple reconstruction is actually equivalent to applying a 2 tap comb filter to the decoded sub-sample signal. The resulting output characteristics for \( Y \) and \( Y_H \) and \( C \) are below.

Thus the fully decoded \( Y_H \) and \( C \) are restricted to low frequencies along the line of the segregation (vertical or vertical-temporal), and contain some residual aliasing.

### 4.3.5 Compatibility of time segregation

The more advanced versions of time segregation are all quite compatible with conventional PAL decoders. For the E-PAL→PAL path the picture quality will be slightly lower than normal, due to the introduction of aliasing and loss of some frequencies.

The PAL→E-PAL path, compatible transmission, on the other hand, is likely to offer poorer performance than for a conventional decoder. The decoder will expect certain relationships to exist between the \( Y_H \) and \( C \) signals. The lack of these relationships will cause re-appearance of cross-effects, probably in a worse manner than for a conventional decoder. The lack of any pre-filtering will also mean the introduction of aliasing into \( Y_H \) and \( C \).

The E-PAL→E-PAL enhanced path will offer very good quality, notably the complete removal of cross-effects. Extra horizontal luminance resolution will be gained, at the expense of a loss in vertical or vertical-temporal detail, and the introduction of a small amount of aliasing.

### 4.3.6 Problems with Time Segregation

As with other approaches time segregation techniques have some disadvantages. These include

1. Poor performance for compatible paths
2. Reduction of aliasing dependant upon use of high order pre- and post-filters.
3. I-PAL does not allow chrominance phase error correction
4. Loss of higher vertical or vertical-temporal frequencies in \( Y_H \) and \( C \).

Of these the poorer performance for compatible paths has perhaps been the major stumbling block of time segregation.
### 4.4 Phase Segregation

Phase segregation is the most difficult of the three methods to explain, and is covered in greater detail, with supporting mathematics in Chapter 7. Implementing phase segregation is complicated by the existing quadrature modulation of the $U/V$ signals around the chrominance sub-carrier. Somehow this must still be preserved in order to maintain compatibility.

Simple phase segregation using *quadrature amplitude modulation* (QAM) can only be performed properly with two signals, whereas in PAL there are three. The solution is to find some way of combining the three signals such that at any one frequency only two signals need to exist in quadrature. Ideally signals coded thus can then be perfectly separated by some form of synchronous demodulation.

To achieve this in PAL requires three steps:

1. Band segregation of $U$ and $V$ chrominance signals so that they occupy different areas of frequency space.
2. Modification of the luminance signal so as to cause $Y_H$ to appear as a *double side-band* (DSB) signal modulated at $f_{se}$.
3. Combination of $Y, U, V$ signals in such a way that all three remain in phase quadrature, but are still PAL compatible.

There are, so far, two methods of achieving this - the BBC's Weston Clean PAL (WC-PAL) [Weston 1977, Oliphant 1980], and K-PAL a new technique developed by the author. These two techniques have similar requirements for steps 1 and 2 above, but differ in the 3rd step. The requirements for each step will be dealt with separately.

#### 4.4.1 Pre-filtering of $U$ & $V$

The chrominance signals must be pre-filtered to ensure that when modulated by the chrominance $U$ and $V$ sub-carriers they do not overlap. If $R_C(f)$ is the filter then this requirement may be represented as

$$R_C(f) R_C\left(f \pm \frac{f_{line}}{2}\right) = 0 \quad (4.5)$$

where $f_{line}$ is the line frequency. In practice for real finite cutoff filters this will not be achievable. However, we will require that (4.5) be satisfied as closely as possible. $R_C(f)$ must also be chosen so as to preserve as much of the important chrominance frequencies as possible.

Three suitable one-dimensional characteristics are shown in Figure 4.22 for $R_C(f)$, for comb filters based on line, field, and frame delays.

(a) Delay = 1 line
CHAPTER 4 ENHANCED PAL CODING TECHNIQUES

Figure 4.22 The three possible comb filter responses for chrominance pre-filtering, showing how the resulting chrominance signals fit together in the PAL spectrum.

As for the band segregation techniques the 312-line comb filter appears to preserve the greatest amount of chrominance information. Two dimensional pre-filtering is also possible, and is proposed for the BBC system [Croll 1992]. Such filters would be similar to those shown earlier in Figure 4.9. However, in the author’s opinion the use of two-dimensional filtering is unnecessary. No two dimensional filter will preserve more chrominance than the 312 line filter of Figure 4.22 (c). As the luminance is pre-filtered separately two dimensional filtering can offer no extra benefits.

4.4.2 Luminance modification

The luminance is a baseband signal, and as such cannot have another signal placed in quadrature to it. It was found, however, with the original Weston Clean PAL that by sampling the signal at the sub-Nyquist rate of $2f_{sc}$, and deliberately aliasing the luminance an appearance of modulation around $f_{sc}$ could be obtained. This process is shown below in Figure 4.23, with its analogue equivalent.

Figure 4.23 Creation of a DSB signal from the baseband luminance by use of (a) sampling, (b) PAL modification.

The cutoff for the low-pass filter after D/A conversion should be either 5 or 5.5 MHz, depending on whether System I or B/G PAL is being used.
The resulting aliasing around \( f_{sc} \) causes the high frequency luminance to appear as a double side-band (DSB) signal modulated at \( f_{sc} \). The use of \( 2f_{sc} \) modulation and a band-pass filter centred on \( f_{sc} \) in Figure 4.23 (b) is commonly known as PAL modification [Bruch 1965].

The one problem with this method is the aliasing introduced. This problem can be fixed by pre-filtering the luminance, such that \( Y \) and \( Y_{alias} \) occupy different areas of spectral space. This will allow the decoder to separate out \( Y \) from \( Y_{alias} \) again. The requirement for such a filter is

\[
R_Y(f) \left( R_Y(f \pm 2f_{sc}) \right) = 0
\]

For one-dimensional comb-filters it happens that the same set of filters as for the chrominance band-segregation are suitable for \( R_Y(f) \) (see Figure 4.22). For two-dimensional filters, however, a different horizontal-vertical filter is also possible [Drewery 1984]. This filter, shown Figure 4.25, attempts to preserve horizontal and vertical detail, by removing a portion of the diagonal frequencies. It has the advantage of being a more complex filter without requiring temporal filtering and the use of field or frame delays. On the other hand, it can only provide a maximum horizontal resolution of 5.2 MHz (332 cpw), compared to 5.5 MHz (352 cpw) for other filters under system I PAL.

### 4.4.3 Luminance chrominance combination

WC-PAL and the author’s new technique differ in this area.

#### 4.4.3.1 Weston Clean PAL

For this technique the \( U \) and \( \pm V \) signals are added together and modulated by a single carrier at \( f_{sc} \). This and the modified luminance are then passed into a PAL combiner. The entire encoding process, including pre-filtering and luminance modification is shown Figure 4.26.

The PAL combiner is quite ingenious in that it maintains the \( U\pm V \) and \( Y' \) signals always in quadrature whilst also making \( U \) and \( V \) appear at the correct sub-carrier phases for normal PAL demodulation. The graphs of luminance and chrominance phase angle in Figure 4.27 show how this is achieved.
The PAL combiner acts as a vertical filter that adds opposing phase shifts to the luminance and chrominance signals. Relative to the sub-carrier the absolute phase of $Y'$ and $U\pm V$ varies continuously with vertical frequency, however, two important requirements are met:

- Luminance and chrominance are, at any one vertical frequency, always $90^\circ$ apart i.e. in quadrature.
- Compared at the $U$ and $V$ sub-carrier positions the chrominance signals are $90^\circ$ out of phase, as is required for PAL.

The PAL combining is a reversible process [Weston 1977] and consequently a WC-PAL decoder can perfectly separate luminance and chrominance. At the same time low frequency chrominance will be correctly decoded by a conventional PAL decoder. The technique has the extra advantage of completely eliminating cross-chrominance near the chrominance sub-carrier positions, due to the luminance being $90^\circ$ out of phase with the decoder sub-carrier at these points.

The one obvious disadvantage of the technique is that as the distance from the sub-carrier frequency increases the chrominance has an increasing phase error. This results in significant chrominance decoding errors in the conventional PAL decoder (attenuation of chrominance and $U/V$ cross-talk).

Considerable work has been done, particularly by the BBC, on the further development of WC-PAL. Since the first description of the technique in 1977 [Weston 1977] this has included testing and the discovery of the principle of phase segregation underlying WC-PAL [Oliphant 1980, 1981, Drewery 1984, 1986]. This has culminated in the proposal of the technique as the basis for a complete enhanced PAL system [Sandbank 1990, Croll 1992, Jones 1992, Storey 1992]. Other alternative uses for the technique have been investigated, particularly as an enhanced studio and recording format, variously known as Extended Studio PAL [Thomas 1992], Single Wire Component [Jones 1993, EBU 1994], and most recently as Composite Compatible Component (COM³) [Itu 1994, Snell 1995, Jones 1995]. This proposed format has actually been adopted by some manufacturers of professional television equipment (e.g. Snell & Wilcox).

### 4.4.3.2 K-PAL

The new K-PAL technique developed by the author is intended to produce a more compatible signal by ensuring that all $U$ and $V$ frequencies are correctly in phase with their respective sub-carriers. The K-PAL encoder appears as in Figure 4.28.
The idea is that $U$ and $V$ are separately phase-segregated with the luminance close to them. We thus have two signals containing $U + Y'$ and $V + Y'$ in phase quadrature. These two signals are then restricted to occupy separate areas of frequency space by the two filters $F_U$ and $F_V$, where

$$F_U(f)F_V(f) = 0$$

As for the chrominance pre-filter $F_U$ and $F_V$ may be of several forms. However, if they are vertical filters, such as that shown in Figure 4.22 (a) then their phase diagrams will appear as

![Figure 4.29 Luminance and chrominance phase in the K-PAL signal, where $R_C$ is a vertical filter.](image)

which may be compared with Figure 4.27 for WC-PAL. It can be seen that phase quadrature is always maintained, whilst also keeping $U$ and $V$ correctly in phase with their respective sub-carriers. Consequently chrominance demodulation errors in the conventional PAL decoder are avoided. Also, rejection of cross-chrominance is improved. The problems with this method result from $F_U$ and $F_V$ being in practice non-ideal filters, with finite cutoff. This means that for enhanced reception the combination of $Y, U, V$ is, unlike WC-PAL, not completely reversible. Phase segregation is not maintained completely in the transition regions between $F_U$ and $F_V$. As $F_U$ and $F_V$ automatically restrict the $U$ and $V$ frequencies to different areas of the spectrum the pre-filtering of $U$ and $V$ becomes unnecessary. $F_U$ and $F_V$ perform the phase combination and chrominance pre-filtering concurrently, and may be represented as shifted versions of $R_C$. The K-PAL coder is further simplified by use of complementary filtering, with

$$F_V(f) = 1 - F_U(f)$$

In conclusion, K-PAL can offer a better compatible picture than WC-PAL, at the expense of a loss of perfect separability for luminance and chrominance in an enhanced decoder. For a more detailed description of K-PAL and WC-PAL see Chapter 7.
4.4.4 Compatibility of phase segregation techniques.

The compatibility of WC-PAL and K-PAL has already been discussed to some degree. The main points are that both offer some improved performance for compatible reception, with K-PAL being slightly better. For compatible transmission a WC-PAL decoder apparently has reduced cross-effects compared to the conventional PAL decoder [Clarke 1982, 1988]. K-PAL, however, is not expected to perform so well for compatible transmission. On the other hand WC-PAL has the disadvantage in both cases of introducing U/V cross-talk, resulting in hue errors on colour transitions. Both methods introduce luminance aliasing, although its visibility is reduced by the high vertical-temporal offset produced by the sub-carrier frequency in the luminance modifier. In a conventional PAL decoder the luminance notch will remove most aliasing.

4.4.5 Problems with phase segregation

The problems differ slightly for K-PAL and WC-PAL as follows:

1. Residual luminance aliasing will always be present.
2. Chrominance resolution is restricted much more than conventional PAL.
3. WC-PAL causes U/V cross-talk.
4. K-PAL has residual cross-effects at the edges of the band segregating filters.
5. For both phase segregation can only be properly maintained when both chrominance sidebands are present. Thus for system B/G PAL phase segregation will only work across a region of $2(5-f_{sc}) = 1.2$ MHz.

One of the notable advantages of the phase segregation techniques, however, is the ability to use different and completely independent pre-filtering for luminance and chrominance.

4.5 Adaptive coding

Adaptive coding attempts to improve the luminance and chrominance signals by dynamically altering the coding parameters to suit the current signal content. All of the techniques described so far work very well for certain picture content (e.g., a vertical comb works perfectly when there are no line-to-line changes). However, they can fail badly under certain different picture content (e.g., the vertical comb fails for sharp vertical transitions). An adaptive coder will detect these changes in picture content and, when the change is significant enough, alter the coding method to suit. The types of adaptive coding used for television are relatively simple, with virtually all techniques being of the form shown below.

![Adaptive coding diagram](image)

Figure 4.30 The general form for an adaptive television coder.

An image filter is used to detect changes in the signal’s spectral content, and this is passed to a decision unit that selects between two types of coding. The filter used may be either linear or non-linear (e.g., gradient based [Kohne 1991]) with non-linear filters often offering improved characteristics for the
same degree of complexity. These filters, and consequently the type of adaptation can be divided into two classes; namely,

- vertical adaptation.
- motion adaptation.

### 4.5.1 Vertical adaptation

Vertical line-based comb filters work well for image content that has little vertical frequency content. When vertical frequencies are present, however, a line-based comb filter will incorrectly decode chrominance and luminance (leaving 'hanging dots' on vertical colour transitions), and greatly restrict the vertical resolution. Under this circumstance it can be better to revert to simple horizontal band segregation. These two situations are shown in Figure 4.31.

![Figure 4.31](image)

**Figure 4.31** The PAL spectrum for adaptive band segregation using (a) vertical comb filter for horizontal detail, and (b) horizontal low and high pass filters for vertical detail.

Thus the adaptive coder must detect the presence of vertical detail and switch between vertical and horizontal band segregation when needed. The simplest way of detecting vertical detail would simply be to subtract one line from the next to detect any line-to-line changes. For an encoder this would be quite satisfactory. For the decoder, however, operating upon the PAL-composite signal this would not work. The chrominance sub-carrier follows an approximately four cycle period, and would be interpreted as vertical luminance detail. The answer is to use the difference obtained over a four line delay, where the chrominance sub-carrier will be in phase. This filter is, however, less effective at detecting fine vertical detail (eg. a fine horizontal line).

Band segregation with vertical adaptation will perform better than the standard methods described in section 4.2, but still fails for some cases. If both vertical and horizontal frequencies are present (ie. diagonal detail) cross-effects will return, and luminance resolution will be lost.

Vertical adaptation has the advantage of requiring few storage delays (no temporal filtering). However, it has to date only been studied for improved PAL decoding [Teichner 1986, Perlman 1990, Markhauser 1990, Kohne 1991]. It has been implemented in some top-of-the-line decoders (eg. Mitsubishi HS-E70 S-VHS VCR).

### 4.5.2 Motion adaptation

Vertical adaptation operates in the horizontal-vertical plane. Motion adaptation operates in the vertical-temporal plane, and requires a motion detecting filter. As for vertical adaptation a simple frame difference filter is sufficient for a coder. For a decoder, however, a four frame filter is needed [Teichner 1985], in order to cancel the sub-carrier. Such a large delay (2500 lines or 8 fields) just for the motion detector (let alone the band segregating filters) may be difficult to justify.
The most obvious band segregating filters for a motion adaptive coder are the vertical and temporal combs (Figure 4.8), which operate as shown in Figure 4.32. Such a coder has been extensively studied by Teichner [1985, 1986, 1988].

Similar to vertical adaptation, failure of the technique will still occur. For motion adaptation this will be when there is significant moving vertical detail, in which case the temporal comb will allow cross-effects to reappear, as well as removing much of the vertical chrominance detail in the moving area.

Types of segregation different to that above may be used, such as field or line combs [Ito 1994], and field / line based time segregation as in ColourPlus [Vreeswijk 1993].

**4.5.2.1 ColourPlus**

This is the technique used in PALplus, the European proposed EDTV system [Croll 1992, Jensen 1993, Matzel 1993, Dreier 1994, Brockhurst 1994, Ellis 1994, PALplus group 1994, Riemann 1994]. It is a technique designed for enhanced encoding and decoding and the basic system is shown below, where IFA stands for Intra-Frame Averaging.

![Simplified block diagram of ColourPlus motion adaptive encoder.](image)

When no motion is present the coding uses field based time segregation, the same principle as in Dean's CAFPAL [Dean 1992]. When motion is detected the system reverts to simple lowpass segregation. A modified form [Westerkamp 1994], reverts to line based band segregation. An addition to the ColourPlus coder is the film/camera modes of operation. This takes advantage of the fact that for picture information from a film source every two fields are from the same film frame. Consequently
these fields represent the same instant in time and there will be no movement between them, meaning that motion adaptation is unnecessary.

The information on which mode is currently being used is transmitted to the receiver. When in camera mode, however, the encoder and decoder motion detectors operate independently. This means that there is a likelihood of encoder and decoder being in different motion adaptive modes occasionally. While this will not cause the coding to fail completely it will reduce picture quality.

4.5.3 Adaptive coding in more detail

In practice adaptive coding does not use a simple switch to change between coding techniques. The sharp change that would result could cause visible picture artefacts. Instead a fading mechanism is used, as shown below, with an adaptive co-efficient $K_m$ varying from 0 - 1.

![Diagram of adaptive detector](image)

*Figure 4.34 Detail of an adaptive detector for two mode adaptive coding.*

This adaptive detector consists of a rectifier, low pass filter, and non-linear characteristic or threshold function. The shape of the non-linear characteristic will determine the smoothness of the change between coding techniques. The low-pass filter will choose the adaptation signal (eg. movement, or vertical detail) and determine how quickly changes between states can be made.

Adaptive phase segregation has not been described in the literature to date. Such a technique would, however, be quite feasible, with the adaptive parameter being used to choose appropriate chrominance and luminance pre-filtering.

4.5.4 Compatibility of adaptive techniques

For compatible transmission (PAL→E-PAL) adaptive decoders have been found to perform very well [Teichner 1985, 1986]. Considering that for conventional PAL the $Y$ and $C$ signals are irretrievably confused adaptive decoding is very good at making the best of a bad situation.

For compatible reception (E-PAL→PAL) some improvements are also possible. Some cross-effects will still be present, but the adaptive coding is likely to allow better decoded resolution for a similar reduction in cross-effects.

4.5.5 Problems with adaptive coding

Adaptive coding is more complex than normal coding. The adaptive detector by itself is likely more complex than any of the single non-adaptive coding techniques. Overall the problems can be summarised as

1. Considerable extra complexity.
2. Need for a vertical or motion detection filter.
3. Need for two types of coding to be present in coder.
4. Changeover artefacts can be introduced.
5. Adaptive encoders and decoders may not perform identically.

These problems aside, however, adaptive coding should be ultimately capable of delivering the best quality enhanced picture of all the enhanced coding techniques presented.
4.6 Present methods for analysing television coding systems

With such a range of possible coding techniques, and each technique having many possible implementations understanding and comparing their characteristics becomes very important. There are two aspects to the appraisal of enhanced coding techniques; namely,

1. Understanding how well they work and what effects they have on the signal.
2. Determining the resulting improvements or reductions in picture quality.

Of these two aspects the second, deciding the overall quality of the coding technique, is perhaps the most difficult. To date this has been achieved almost exclusively through the use of subjective testing (see section 2.9) although in Chapter 6 the author presents a new objective assessment technique.

The first point, the understanding and analysis of a coding technique is made difficult by the many variables involved. There are three components to be coded, six cross-component paths, and three dimensions be considered. Many techniques also involve the use of time-varying functions (eg. modulation, sampling). Methods have been developed to help study these.

To date these methods have concentrated upon the use of special television test images, with the characteristics of the coding system being inferred from changes visible in the decoded image. General test images, the standard television test patterns, have been in use since the beginning of television. In recent times more specific test images, such as spatial zones, temporal zones and colour patches have been used. These provide much more information. Mathematical analysis is another available method. This section provides an overview of these analysis methods, and shows some of the results that are obtained using them to analyse a PAL system. These results will be useful for comparison with those obtained in Chapter 5, using a new three dimensional spectral analysis method.

4.6.1 Mathematical analysis

Enhanced television coding techniques are relatively simple in comparison with digital coding techniques (eg MPEG). Using knowledge of the relationship between the television signal and its spatio-temporal spectrum (section 3.1) it is possible to represent mathematically the characteristics of a coding system. To date such an approach has been used mostly to gain conceptual insight [Fukinuki, 1991, Isnardi, 1988], and occasionally to describe sub-sections of a coder [Teichner, 1988, Clarke, 1988]. However, it has not been used to provide an overall description of a coder's characteristics.

4.6.2 Image domain analysis

Using special test images is one way of gaining information on a system characteristic. Television test patterns have long been used to provide information on coding performance, two such patterns are shown in Figure 4.35

![Figure 4.35](image)

Figure 4.35 Two television test patterns, (a) colour and freq. burst pattern, (b) RMA resolution chart.
These test patterns use a combination of colour bars, vertical and horizontal lines to provide some information on the cross-effects, step, phase and frequency response of the system. However, the information that they are capable of providing is very limited.

A more useful test picture is the zone plate. Weston, 1980, , Fukinuki, 1986,). This special test image can provide comprehensive information on the spectral characteristic of a system while working entirely within the image domain ie. no Fourier analysis is required. It also has the advantages of being easy to use (just sit the picture in front of the camera), and simple to evaluate.

There is a range of different zone plates, with each one containing a two-dimensional range of spatio-temporal frequencies, equivalent to a planar slice through spatio-temporal frequency space. Figure 4.36 illustrates this for the commonly used circular zone plate [Drewery, 1978, Weston, 1982] and parabolic zone plate [Alvarez, 1990].

4.6.2.1 Stationary zone plates
Stationary zone plates provide information on a coding systems spatial response. Normally vertical image frequency varies in one direction and horizontal frequency in the other direction. Thus every point in the image represents a different combination of vertical and horizontal frequency. Figure 4.37 shows two such zone plates, the circular zone plate and the hyperbolic zone plate [Drewery, 1978] (the difference being the directions in which horizontal and vertical frequency vary).
4.6.2.2 Moving zone plate

The spatial zone plate provides little information on temporal characteristics. To achieve this a moving image is required which means the use of an image sequence. A number of such images have been described [Weston, 1980, Fukinuki, 1986, Alvarez, 1990]. Figure 4.38 shows two of these, a moving circular zone, and a parabolic zone. Using these images can provide information on different sections of frequency space (see Figure 4.36). The moving circular zone contains a full range of spatial frequencies, all with the same temporal offset (equivalent to shifting the circular zone plate along the $f_t$ axis as shown in Figure 4.36). In the parabolic zone temporal frequency varies horizontally and vertical frequency vertically. Both these images can be changed to study different areas of frequency space by varying the temporal offset for the circular zone and the horizontal frequency offset for the parabolic zone.

It may be noted that all the zone plates shown here consist of alternating black and white bands. Although this is traditionally the form zone plates have been used in it is not ideal. As noted by Drewery [1982], the sharp transitions introduce unwanted harmonics, preventing the zone plate from working accurately. The ideal is a smoothly changing sine wave. All the zone plates used by the author were consequentlly of this form.

As may be appreciated there is a considerable range of different zone plates and many are needed to provide a full picture of a system's spatio-temporal characteristics. In using these test images the author has generalised them into two categories: the spatial zone plate and the temporal zone plate.

4.6.3 The spatial zone plate

The circular zone plate is a useful example of a spatial frequency test image. This image has the advantage of horizontal frequency varying horizontally, and vice versa for vertical frequency. Using smooth sine wave transitions this spatial zone may be defined

$$
e(x, y, t) = \cos \left( 2\pi \left( \frac{M_{\text{max}}}{2A} x + \frac{N_{\text{max}}}{2B} y + v_t \right) \right)$$

(4.9)

where $M_{\text{max}}$, $N_{\text{max}}$ are the maximum horizontal and vertical frequencies contained in the image, $v_t$ is the temporal frequency offset, and $A$, $B$ are the image dimensions. For this image horizontal frequency varies with $x$ and vertical frequency with $y$. For $v_t = 0$ the image can be represented by a single frame, however, for $v_t \neq 0$ the image will change with time (the rings will appear to converge, or diverge) and the image must be represented as a sequence, with a repeating period of $1/v_t$. 

---

Figure 4.37 Two stationary zone plates, (a) circular zone plate, (b) hyperbolic zone plate.

Figure 4.38 Two moving zone plates (arrows indicate direction and speed of movement), (a) moving circular zone, (b) parabolic zone.
Below is an example of the use of this test image on a conventional PAL coding system. The simulation was run three times, with the image being used as a luminance, or chrominance (U or V) only test signal and the other inputs being left blank. In practice a chrominance only PAL signal is impossible, however under simulation this approach allows us to isolate the effects of each of the nine possible paths through the PAL system (three intra-component and six inter- or cross-component). Figure 4.39 shows the resulting output images for the three simulations. The output images have been displayed as separate Y, U, and V signals, and have been 'rectified' (bottom amplitude range of signal removed) to help reveal the magnitude changes in the zone plates' sine wave envelope. If this 'rectification' is not performed it becomes difficult for the human eye to judge attenuation of the envelope correctly.

The first row of images shows the luminance response and cross-luminance effects. The restricted horizontal bandwidth of PAL can be seen, as also the diagonal frequencies that cause cross-chrominance. The position of this cross-chrominance corresponds quite well with the theoretical positions of the chrominance shown in figure 3.20 in Chapter 3.

Figure 4.39 The output component signals obtained when a spatial zone image is applied separately to the Y, U and V inputs of a simulation of System B/G PAL.

The second and third rows show U and V signals, cross-luminance and U/V cross-talk. Both U and V have greatly restricted horizontal bandwidth, and show a loss of intermediate vertical frequencies due to the PAL delay line. Cross-luminance (V→Y, U→Y) is caused mostly by medium horizontal chrominance frequencies, with the effects of low frequencies being removed by the luminance notch.
Ideally for PAL the quadrature modulation of $U$ and $V$ signals should mean that they are perfectly separable. In system B/G PAL, however, the upper chrominance sideband is restricted to only $5f_{sc} = 0.6$ MHz. This causes confusion of any chrominance signals above 0.6 MHz, resulting in a low-level of $U/V$ cross-talk for intermediate horizontal chrominance frequencies.

Although the test zone plate used here contains no temporal offsets this is not true for all the output images. To fully appreciate these results they must be viewed as sequences, whereupon the temporal offsets of the cross-luminance and cross-chrominance become evident. For $Y \rightarrow U$ and $Y \rightarrow V$ the circles will appear to converge and diverge at varying speeds, and for $U \rightarrow Y$ and $V \rightarrow Y$ the dots appear to 'crawl' slowly diagonally across the image.

The use of the spatial zone as a comparison and analysis tool is further illustrated by comparing the first row of Figure 4.39 with the images below, obtained from a simulation of I-PAL [Holoch 1985] (see section 4.3.1).

![Figure 4.40](image)

*Figure 4.40 The output component signals obtained when a spatial zone image is applied to the luminance input of an I-PAL simulation.*

Most obvious is the almost complete elimination of cross-chrominance, with $Y \rightarrow U$ and $Y \rightarrow V$ being almost blank. Figure 4.39 $Y \rightarrow Y$ shows an increased horizontal luminance bandwidth compared to conventional PAL in Figure 4.40 $Y \rightarrow Y$. However, this is at the expense of some aliasing of diagonal frequencies. If viewed as a sequence this aliasing is seen to have temporal offsets as well.

### 4.6.4 The temporal (parabolic) zone

The spatial zone plate provides no information on the temporal characteristics of the system paths, and little information for the cross-effects paths. To better analyse the temporal characteristics a cross-section through temporal frequency space is needed. This is provided by the temporal zone image, described by the equation

$$e(x, y, t) = \cos \left(2\pi \left(M_x x + \frac{N_{\text{max},y} y + \frac{v_{\text{max}} x}{2A} t}{2B}\right)\right)$$  \hspace{1cm} (4.10)

where $N_{\text{max},y}$, $v_{\text{max}}$ are the maximum vertical and temporal frequencies contained in the image, $M_x$ is the horizontal frequency offset and $A$, $B$ are the image dimensions. For the resulting image vertical frequency varies along $y$ (as for spatial zone), and temporal frequency varies with $x$.

The spatial and temporal zones are excellent test images for providing a qualitative impression of a system's spatio-temporal characteristic. However, they have two main disadvantages. The first is the fact that it is very difficult to obtain any quantitative information about the system characteristic from the images - they must be displayed and interpreted visually by an observer. The second disadvantage is that both image types provide information on only a cross-section of spatio-temporal frequency space. To obtain more complete information a whole series of image sequences with differing frequency offsets would have to be generated to cover the whole of spatio-temporal frequency space. These disadvantages are completely absent from a new method of analysis developed by the author.
This method uses direct spectral analysis to gain a complete description of a system’s spatio-temporal characteristics and is described in detail in Chapter 5.

4.7 Summary

This Chapter has provided an overview of enhanced PAL coding, describing all the main approaches and techniques that have been developed in current literature. At present no direct comparison has been made of all these techniques and as has been mentioned previously the common basis of many techniques has been somewhat obscured by the different approaches used in explaining them. This Chapter places all these techniques in a common framework as being forms of either band, time, or phase segregation used to achieve separation of luminance and chrominance signals. All the techniques can be described in terms of their spatio-temporal characteristics. This has allowed a more direct comparison of the comparative advantages and disadvantages of the techniques. Unfortunately space has not permitted a detailed quantitative analysis and comparison of the many techniques. Such an undertaking would be worthwhile but massive in scale considering the many permutations of techniques that are possible. Instead Chapter 7 will undertake a more detailed study and performance evaluation of the phase segregation techniques.

This Chapter has also presented a brief overview of the currently used methods for spectral analysis of coding techniques. This is important in appreciating the further advancements in spectral analysis achieved by the author and described in Chapter 5.
CHAPTER 5

NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

When studying a particular enhanced television technique the combination of encoder and decoder may be regarded as the system under study. The function of this system is to take an input image and produce an output image whose quality is as close as is possible to that of the original image. The ability of a system to achieve this can be judged to a great degree from the system transfer function. The transfer function provides information on the relationship between the input and resulting output of the system. It can thus be said to characterise the system. Being able to characterise a system allows us to go a long way towards answering the questions posed in the introduction - how good is the system? and what, exactly, is wrong with it?

There are several ways of characterising a system. Some of those important for a television system are:

- step response
- phase response
- frequency response

The last of these, the frequency response, arguably provides the greatest amount of information about the system. For this reason the author's work to date has concentrated on finding methods for evaluating this response, although it must be pointed out that to fully characterise a system the step and phase response must also be considered.

5.1 Television as a Linear MIMO system

An analogue colour television system (encoder plus decoder) may be represented as a MIMO system $S$, as shown below,

$$
\begin{array}{c}
\text{x}_i(r) \\
\text{x}_o(r) \\
\text{x}_t(r)
\end{array}
\rightarrow
\begin{array}{c}
\text{S} \\
\text{PAL system} \\
\text{encoder & decoder}
\end{array}
\rightarrow
\begin{array}{c}
\text{y}_i(r) \\
\text{y}_o(r) \\
\text{y}_t(r)
\end{array}
$$

*Figure 5.1 The colour television system*

where $x_i(r)$, $x_o(r)$, $x_t(r)$ are the gamma corrected $Y, U, V$ picture signals obtained from the image source (eg. camera), $y_i(r)$, $y_o(r)$, $y_t(r)$ are the decoded picture signals displayed at the receiver, and the vector $r$ may be either time $r = [t]$ or a three-dimensional vector $r = [x \ y \ t]$. These signals will be regarded as being independent (though in practice this is not strictly the case). If the system is linear then the Fourier transform of the outputs will be related to the Fourier transform of the inputs by,
CHAPTER 5 NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE
SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

\[
X(v) = T(v) Y(v)
\]

\[
\begin{bmatrix}
Y_x(v) \\
Y_y(v) \\
Y_z(v)
\end{bmatrix} =
\begin{bmatrix}
T_{yy}(v) & T_{yu}(v) & T_{yz}(v) \\
T_{uy}(v) & T_{uu}(v) & T_{uz}(v) \\
T_{vy}(v) & T_{vu}(v) & T_{vv}(v)
\end{bmatrix}
\begin{bmatrix}
X_x(v) \\
X_y(v) \\
X_z(v)
\end{bmatrix}
\]

(5.1)

where the vector \(v\) is either one dimensional signal frequency \(v = [f]\) or a spatial-temporal frequency vector \(v = [v_x, v_y, v_z]\), and \(T(v)\) is the system transfer function.

The nine functions in (5.1) may be divided into two groups; the main diagonal representing the \textit{intra-component} paths, and the other functions representing the \textit{inter} or \textit{cross-component} paths. For an ideal system all intra-component paths would be unity, and all cross-component paths zero, i.e.

\[
T(v) = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

(5.2)

In practice this is not the case, and the unwanted effects that arise are commonly referred to as:

- \(T_{yy}\), \(T_{yu}\), \(T_{yz}\) - cross-chrominance (or cross-colour)
- \(T_{uy}\), \(T_{uu}\), \(T_{uz}\) - cross-luminance
- \(T_{vy}\), \(T_{vu}\), \(T_{vv}\) - inter-chrominance (or U-V crosstalk)

The remaining \(T_{yy}, T_{yu}, T_{uv}\) define the luminance and chrominance responses for the system \(S\).

5.1.1 Television as a periodically time varying system - a one-dimensional analysis

Many of the paths of a conventional or enhanced analogue television system contain time-varying functions - modulation and switching (PAL switch) in conventional PAL, and sampling, sub-sampling, and phase-segregation in different enhanced techniques. These functions are linear (superposition holds true), but vary in time in a periodic manner. Thus \(S\) is more correctly represented as a linear periodically time-varying (LPTV) system, with each path in \(S\) being of the general form shown in Figure 5.2 where \(r\) has been replaced with \(t\), and \(v\) with \(f\) for the present one dimensional analysis. In Figure 5.2 \(m_i(t)\) are sinusoidally time-varying parameters and \(H_i(f)\) are the lumped transfer functions associated with these parameters. In practice there may be any number of functions existing before or after \(m_i(t)\). For the purposes of this analysis all of these will be viewed together in terms of an equivalent transfer function \(H_i(f)\) placed after \(m_i(t)\).

At the end of this analysis \(t\) and \(f\) will be replaced again with their three-dimensional vectors. The results for the one and three-dimensional cases are very similar, and thus most of the derivation can be shown for the one dimensional case only, and later generalised to three dimensions.

For input \(x(t)\) and output \(y(t)\) any path in the PAL system can be defined by the transfer functions \(H_0, H_1, \ldots, H_N\) and the sinusoidal modulation functions \(m_1, m_2, \ldots, m_N\) where

\[
m_i(t) = A_i \cos(2\pi f_i t + \phi_i)
\]

(5.3)

It must be noted that the representation in Figure 5.2 is not valid for systems containing adaptive parameters (e.g. motion adaptive filtering). Such systems are no longer fully linear. However, if the adaptive controls are simple (e.g. 2 state adaptation) it may be possible to model the system as two (or more) LPTV systems. For example, for a motion adaptive coder there would be two systems, one for stationary image content, and a second for motion.

Omitting adaptive systems from our analysis for the present, the output \(y(t)\) of any path through \(S\) may be described by

\[
y(t) = \int_{-\infty}^{\infty} h_0(t - \tau) x(\tau) d\tau + \sum_{i=1}^{N} \int_{-\infty}^{\infty} h_i(t - \tau) x(\tau) m_i(\tau) d\tau
\]

(5.4)
5.1 TELEVISION AS A LINEAR MIMO SYSTEM

where \( h_i(t) = \text{IFT}(H_i(f)) \).

\[ x(t) \rightarrow H(f) \rightarrow y(t) \]

\[ \begin{align*}
M_i(f) &= \frac{1}{2} A_i \left( \delta(f + f_i) e^{-j\phi} + \delta(f - f_i) e^{j\phi} \right) \\
Y(f) &= H_0(f)X(f) + \sum_{i=1}^{N} H_i(f) \left( X(f) * M_i(f) \right)
\end{align*} \]  

(5.5)  

(5.6)

where * indicates convolution. Substituting for \( M_i(f) \) (5.5) becomes

\[ Y(f) = H_0(f)X(f) + \frac{1}{2} \sum_{i=1}^{N} A_i H_i(f) \left( X(f + f_i) e^{-j\phi} + X(f - f_i) e^{j\phi} \right) \]  

(5.7)

Putting \( G_i(f) = A_i H_i(f) \) and \( G_0(f) = H_0(f) \) (5.7) becomes

\[ Y(f) = G_0(f)X(f) + \frac{1}{2} \sum_{i=1}^{N} G_i(f) \left( X(f + f_i) e^{-j\phi} + X(f - f_i) e^{j\phi} \right) \]  

(5.8)

\( G_i(f) \) will be referred to as the modulation transfer function.

Each path of the system \( S \) may now be fully described by a baseband transfer function, \( G_0(f) \) and \( N \) modulation transfer functions \( G_i(f) \).

5.1.2 Identifying the system

If no prior knowledge of \( S \) is assumed (except that it is LPTV), then \( G_i(f) \) must be determined solely by comparing output \( x(t) \) with input \( y(t) \). One method of identification for linear systems is to 'sample' the system characteristics, using single frequency input signals. This is equivalent to 'sampling' the Fourier characteristics with a delta function. If \( u \) is the input frequency sample point chosen, then a suitable input signal is

\[ x(t) = \begin{cases} 
1 & u = 0 \\
2 \cos(2\pi ur) & u > 0
\end{cases} \]  

(5.9)

Note that as we are considering real signals only it is necessary only to test the system for positive \( u \).
If \( x \) is expressed as a function of \( u \) as well as \( t \), then the Fourier transform of \( x(t) \) (w.r.t \( t \)) is

\[
X(f, u) = \delta(|f| - u)
\]  

(5.10)

Substituting for \( X(f) \) in (5.8) and replacing \( Y(f) \) with \( H(f, u) \) we obtain

\[
H(f, u) = G_0(f) \delta(|f| - u) + \frac{1}{2} \sum_{i=1}^{N} G_i(f) \left( \delta(|f + f_i| - u)e^{j\phi} + \delta(|f - f_i| - u)e^{-j\phi} \right)
\]  

(5.11)

The resulting function \( H(f, u) \) contains a complete description of the system’s characteristics as a set of output spectra for every input frequency \( u \), and will be referred to as the input-output characteristic.

This relationship between \( H(f, u) \) and \( G_i(f) \) can be represented graphically as shown below, with \( G_i(f) \) all existing along the lines \( |f \pm f_i| = u \) in the \( f, u \) plane of \( H(f, u) \). For three frequencies \( f_0, f_a, f_b \) the paths on which \( G_0(f), G_a(f), \) and \( G_b(f) \) lie on are shown.

![Figure 5.3 Representation of \( G(f) \) on \( f, u \) plane of \( H(f, u) \).](image)

If input and output frequencies are specified separately as \( w \) and \( f \), then for any input signal with a spectrum of \( X(w) \) the output spectrum may be determined from the input-output characteristic as

\[
Y(f) = \int_{-\infty}^{\infty} H(f, w) X(w) \, dw
\]  

(5.12)

If \( X(w) \) is defined by (5.10) then the input-output characteristic may be easily determined from the output of the system

\[
H(f, u) = Y(f) \bigg|_{X(w) = \delta(|w| - u)}
\]  

(5.13)

Complete identification of the system thus requires (i) determination of \( H(f, u) \), and (ii) solving of (5.11) for \( G(f) \).

Unfortunately achieving this is, in practice, very difficult. There are several reasons for this:
5.2 A SIMPLIFIED REPRESENTATION OF THE SYSTEM CHARACTERISTICS

- Determination of $G_i(f)$ requires the solution of a potentially very large set of linear equations, particularly for the case where $f$ and $u$ are three-dimensional.
- All points in $H(f, u)$ must be known concurrently (i.e., solutions cannot be built up a portion at a time).
- As $N$ and $f_i$ are unknown $G_i(f)$ must be represented for all possible $f_i$.

To illustrate, if an area of spatio-temporal frequency space is sampled on a $32 \times 32 \times 32$ grid, and the system under study is tested for $8 \times 8 \times 8$ different spatio-temporal input frequencies then:

- Determination of $G_i(f)$ requires the solution of $32^3 \times 8^3 \equiv 1.7e7$ equations.
- $H(f, u)$ contains $32^3/2 \times 8^3 \equiv 8e6$ points (ignoring complex conjugate portion of $H$).
- There are $8^3 = 512$ possible $G_i(f)$.

Although use of matrix techniques and mathematical symmetries can reduce the computation considerably, the problem is still computationally impractical in its present form. Consequently the information on the system must be left in the form of $H(f, u)$. The disadvantage of this is that (i) $H(f, u)$ is a very sparse function for small $N$, and (ii) it has twice as many dimensions as $G_i(f)$. For analysis purposes it is useful to extract the most important information from the input-output characteristic and represent this in a simplified form.

### 5.2 A simplified representation of the system characteristics

When appraising an analogue television coding technique we are most interested in

1) The response of the intra-component (wanted) signal paths.
2) The causes and effects of artefacts in these paths (e.g., aliasing).
3) The causes and effects of any inter-component (cross-component) artefacts.

where an artefact is any unwanted output signal not originally present in the input signal. This can be stated more specifically as separate criteria for the intra and inter component paths.

For the intra-component paths we desire to know

- The baseband (time invariant) transfer function.
- The output artefacts that arise due to the effects of time-varying parameters, i.e., the sum of all the modulation transfer functions.
- The input frequencies that are affected by time-varying parameters, i.e., the modulation transfer functions as they appear viewed from the input.

These will be referred to as the baseband transfer function $T(f)$, input modulation characteristic $I(u)$, and output modulation characteristic $O(f)$. $O(f)$ represents the lumped contributions of all $H_i(f)$, with $I(u)$ representing the sum of all $H_i(f)$ as they would appear if translated into front of the time-varying parameters $m_i(t)$ (see Figure 5.2).

For the cross-component paths any output signal is an artefact; there is no reason to distinguish between time-varying and time-invariant parts. Thus we will define only two functions for these paths.

- The level of output cross-effects existing at each output frequency $f$, equivalent to the sum of all transfer functions $H_0$ to $H_N$.
- The contribution of each input frequency $u$ to these cross-effects, defined as the sum of all transfer functions $H_0$ to $H_N$ as translated to in front of $m_i(t)$.

These will be referred to as the output and input cross-effect characteristics, $O^c(f)$ and $I^c(u)$.

As far as cross-effects and artefacts are concerned the phase relationships between input and output signals are relatively unimportant. This is because (i) these signals are of different frequency, and (ii) phase shifts in unwanted signals will make little difference to their visibility. We will therefore ignore phase information in these functions, and consider only the relative levels, or magnitude.
CHAPTER 5 NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

For the baseband transfer function, representing the wanted signal, phase effects are definitely important; phase distortion can have a considerable effect on picture quality. However, these effects are better studied in the image domain, and for the present analysis all the functions described will be defined as containing magnitude only.

The properties of the system, as represented by the functions that have just been defined, may now be represented in terms of \( H(f, u) \).

\[
T(f) = \int_0^1 |H(f, u)| \delta(|f| - u) \, du \tag{5.14}
\]

\[
O(f) = \int_0^1 |H(f, u)| \left[ 1 - \delta(|f| - u) \right] \, du \tag{5.15}
\]

\[
I(u) = \int_0^1 |H(f, u)| \left[ 1 - \delta(|f| - u) \right] \, df \tag{5.16}
\]

\[
O^c(f) = \int_0^1 |H(f, u)| \, du \tag{5.17}
\]

\[
I^c(u) = \int_0^1 |H(f, u)| \, df \tag{5.18}
\]

Referring back to Figure 5.3 it can be seen that these functions are projections through \( H(f, u) \) onto the \( u \) and \( f \) axes. This is almost equivalent to summing the magnitudes of the modulation transfer functions.

\[
T(f) = |G_0(f)| \tag{5.19}
\]

\[
O(f) = \sum_{i=1}^N |G_i(f)| \tag{5.20}
\]

\[
I(u) = \frac{1}{2} \sum_{i=1}^N |G_i(u + f_i)| + |G_i(u - f_i)| \tag{5.21}
\]

\[
O^c(f) = |G_0(f)| + \sum_{i=1}^N |G_i(f)| \tag{5.22}
\]

\[
I^c(u) = |G_0(u)| + \frac{1}{2} \sum_{i=1}^N |G_i(u + f_i)| + |G_i(u - f_i)| \tag{5.23}
\]

This representation, although helpful conceptually fails at the points of intersection of the different \( G_i(f) \). For these input frequencies the fact that two different modulation transfer functions, or the conjugates of one (at \( f=0 \)) produce the same output frequency means that relative phase differences in \( m_i(t) \) and \( G_i(f) \) can result in the reinforcing, or cancellation of the output signal. Defined in terms of \( H(f, u) \) these effects are automatically accounted for.

The intersections within the input-output characteristic do, however, cause problems for extracting the baseband transfer function \( G_0(f) \). Due to the overlap of \( G_0(f) \) and \( G_i(f) \) at points \((u, f) = (\pm f_i, \pm f_i)\) the functions become inseparable at these points. A portion of the characteristic that should be part of \( G_i(f) \) is consequently defined by (5.14) as being part of \( G_0(f) \).

Strictly speaking this is not an error. A frequency \( f_0 \) modulated by \( 2f_0 \) ends up back at \( f_0 \). Thus any input frequency that is exactly half the frequency of a time-varying parameter undergoes no frequency shift. In a purely theoretical sense this result is therefore part of the time-invariant response of the system, and is truly part of \( T(f) \). In practice, however, anything passing through \( G_i(f) \) represents an unwanted signal, and ought to be considered as part of \( O(f) \) and \( R(u) \), not \( T(f) \).
In any case, for real world systems $N$ is small (i.e. $N < 10^4$) and any errors in the result will be small, amounting to $2N$ points in error. For the continuous case these points occupy an infinitely small portion of the response (due to the delta functions), however, in the discrete case, depending upon the size of the steps chosen, they can have significance (see later in section 5.3.2).

To illustrate these effects and the appearance of the various functions that have been described an example will be given for a portion of a PAL coding system.

### 5.2.1 Example of use of the defined system representation

Figure 5.4 shows a portion of an enhanced PAL coding system that incorporates a PAL modifier in the encoder.

![Block diagram for a section of an example enhanced PAL system, showing Y/U encoding and Y decoding.](image)

The input-output characteristics for the intra-component $Y$->$Y$ path, and the cross-component $U$->$Y$ path are shown below, for $f$ and $u$ being one-dimensional signal frequency. The characteristics for these two paths are represented as $H_{YY}(f, u)$ and $H_{UY}(f, u)$ respectively.

![Input-output characteristics for the Y->Y and U->Y paths of the system shown by Figure 5.4](image)

---

4 Normally $N$ will equal no more than 2 for Enhanced television systems with sinusoidally time-varying parameters. However if the time-varying parameter is non-sinusoidal then it must be represented by an infinite series of carriers. In this case, for a practical analysis, it is only really necessary to only consider the first few carriers, which will carry most of the energy; omitting higher harmonics of the function will introduce little error.
Applying equations (5.14)-(5.18) to the function plotted above results in the five functions shown below, for $T_{Y}(f)$, $O_{Y}(f)$, $I_{Y}(u)$, $O'_{U}(f)$, and $I'_{U}(u)$.

![Luminance transfer function](image)

![Luminance alias input](image)

![Luminance alias output](image)

![Chrominance input](image)

![Chrominance output](image)

Figure 5.6  The characteristic functions for the (a) $Y \rightarrow Y$, and (b) $U \rightarrow Y$ paths of Figure 5.4

Figure 5.6 (a) shows the characteristics for the $Y \rightarrow Y$ path. Evident is the effect of the notch filters and the 0.5 scaling of the area around $f_{sc}$. The time-varying functions show the effect of the PAL modifier, which causes an artefact to be introduced at frequencies near $f_{sc}$. Note also the impulse at $f_{sc}$ in the time-invariant transfer function, and the zeros at $f_{sc}$ for the time-varying portions. These arise due to the errors in separating $G_{y}(f)$ and $G_{f}(f)$ discussed earlier, and occur at half the frequency of the time-varying parameter, which in this case was $2f_{sc}$.

In Figure 5.6 (b), the $U \rightarrow Y$ path, it is low frequency chrominance that causes the introduction of high-frequency artefacts into the luminance path. As the $U \rightarrow Y$ path is a cross-component path no time-invariant transfer function is given.

For the $U \rightarrow Y$ path the frequency shift between input and output is fairly obvious. However, for the $Y \rightarrow Y$ path there does not appear to have been any shift at all. Unfortunately information on the value of $f_{i}$ has been lost in equations (5.14) - (5.18). Considering the fact that for most practical systems $N$ will be small we may preserve most of the information about $f_{i}$ by use of the equation

$$W(f) = \sum_{i=0}^{N} \delta(f \pm f_{i}) \int_{f_{sc}}^{+\infty} G_{i}(w) \, dw$$

(5.24)

$W(f)$ will consist of $N$ delta functions each weighted by the sum of the corresponding modulation transfer function $G_{i}(f)$. $W(f)$ thus provides us with a picture of the different $f_{i}$ present in the system, with the magnitude indicating the contribution of each time-varying function to the output. The actual magnitude of $W(f)$ is unimportant, of more interest is the relative contributions of the different modulation transfer functions. Consequently we will only deal with a normalised version $W'(f)$ in practice.
With no knowledge of the exact $G_i(f)$, $W(f)$ must be determined using $H(f, u)$. This can be achieved by summing along each of the lines $f = u + f_i$ or $f = u - f_i$ in Figure 5.3

$$W(f) = \int_0^\infty |H(u + f, u)| + |H(u - f, u)| + |H(-u + f, u)| + |H(-u - f, u)| \, du$$

For $f = f_i$, $W(f)$ will provide us with an appropriate weighting for this time-varying parameter.

As for the inter-component functions the intersections of the modulation transfer functions will result in errors in $W(f)$. This occurs due to the ambiguity existing at these points. For an input frequency $u = f_c$ that results in an output frequency $f = f_d$ the time-varying parameter could be either $f_i = f_d - f_c$ or $f_i = f_d + f_c$. Once again, however, for practical systems with small $N$ the errors will be negligible, amounting to a low-level noise in $W(f)$.

Using (5.26) on $|H_{yy}(f, u)|$ and $|H_{uv}(f, u)|$ as they are plotted in Figure 5.5 we can determine $W_{yy}(f)$ and $W_{uv}(f)$. These are shown below, after normalisation.

$$W_{yy}(f) = \begin{cases} 1 & \text{for } f = 0 \\ \frac{1}{2} & \text{for } f = \pm f_c \end{cases}$$

$$W_{uv}(f) = \begin{cases} 1 & \text{for } f = f_c \\ \frac{1}{2} & \text{for } f = 0 \\ \frac{1}{4} & \text{for } f = \pm f_c \end{cases}$$

Figure 5.7 $W(f)$ for the Y->Y and U->Y paths, showing the value and relative contributions of all time-varying parameters in Figure 5.4.

From this graph we can see that the luminance path Y->Y has most of its energy travelling through $G_y(f)$, with a small portion undergoing a shift of $f_i = 2f_c$. This means that, contrary to the appearance of Figure 5.6 (a) the signals centred on $f_c$ are not unchanged. The $2f_c$ shift is equivalent to a reflection about $f_c$ frequencies above $f_c$ in $I_{yy}(u)$ end up below $f_c$ in $O_{yy}(f)$, and vice-versa. For the U->Y path all signals undergo a shift of $f_i = f_c$, just as Figure 5.6 (b) shows.

### 5.2.2 A complete representation of the television system

It has been shown that $T(v)$ (as defined by (5.1)) is insufficient to properly represent an LPTV system. Instead this will be replaced with the range of functions that has been described. This results in the 30 functions grouped below

$$T(v) = \begin{bmatrix} T_{yy}(v) \\ T_{uv}(v) \\ T_{vy}(v) \end{bmatrix}$$

(5.27)
where \( f \) and \( u \) have now been replaced with vectors \( v \) and \( w \). As stated before these vectors may be either one-dimensional \((v = [f], w = [u])\), or three-dimensional \((v = [v_x, v_y, v_t], w = [w_x, w_y, w_t])\).

The significance of these functions is summarised below.

- **\( T(v) \)** - The time-invariant transfer function for the Y, U, V components, i.e. the intra-component paths
- **\( O(v) \)** - Shows the output artefacts (i.e. frequencies) that arise in the intra-component paths due to time-varying parameters (e.g. aliasing).
- **\( I(w) \)** - Shows the input frequencies that result in the output artefacts in \( O(v) \)
- **\( O^c(v) \)** - Shows the output frequencies that arise due to cross-component effects.
- **\( I^c(w) \)** - Shows the input frequencies that result in cross-component effects.
- **\( W(v) \)** - Identifies the period (i.e. frequency) of any existing time-varying parameters and shows their relative contributions.

### 5.3 Determining the characteristics for a given system

Completely and accurately determining the spatio-temporal characteristics for a television coding system is a task that has never been performed previously. This is in large part due to the limited techniques that have been used.

The presently used methods for analysis and appraisal were described in detail in section 4.6. These analysis methods provided very limited, and usually only qualitative information on the television system characteristics.

This section will deal with the development on a practical method for numerically determining the complete characteristics.

As stated earlier the ideal way to identify an LPTV system is by recording its output response for every possible input frequency.

\[
H(f, u) = Y(f)
\]

where \( f \) is output frequency, \( w \) input frequency, and \( u \) the test frequency.
5.3 DETERMINING THE CHARACTERISTICS FOR A GIVEN SYSTEM

Achieving this would be impractical if it were not for a number of mitigating factors.

- Only a certain range of frequencies is of interest.
- For computer based calculation $H(f, u)$ must be represented on a finite number of points i.e. $H(f, u)$ will be discrete.
- Only a limited number of input frequencies need to be tested to obtain a useful representation of $H(f, u)$.

These constraints will be dealt with separately.

5.3.1 The frequency range of interest

As explained in Chapter 3 the television signal is sampled in the vertical and temporal dimensions, and band-limited in the horizontal. This gives spatio-temporal frequency limits for the system of

\[
\begin{align*}
  v_{x_{\text{max}}} &= \frac{B}{f_h} \quad \text{cycles per picture width (cpw)} \quad (5.34) \\
  v_{y_{\text{max}}} &= \frac{N}{2} \quad \text{cycles per picture height (cph)} \quad (5.35) \\
  v_{t_{\text{max}}} &= \frac{f_v}{2} \quad \text{Hz} \quad (5.36)
\end{align*}
\]

where $B$ is the nominal transmission bandwidth, $f_h$ the line rate, $f_v$ the field rate, and $N$ the number of lines per frame. For B/G PAL with $B = 5 \text{ MHz}$, $f_h = 15625 \text{ Hz}$, $f_v = 50 \text{ Hz}$, and $N = 625$ these limits are

\[
\begin{align*}
  v_{x_{\text{max}}} &= 320 \quad \text{cpw} \quad (5.37) \\
  v_{y_{\text{max}}} &= 312.5 \quad \text{cph} \quad (5.38) \\
  v_{t_{\text{max}}} &= 25 \quad \text{Hz} \quad (5.39)
\end{align*}
\]

In studying a B/G PAL coding system we therefore need test only for frequencies up to these limits.

5.3.2 Discrete representation of $H(f, u)$

For the purposes of computer based calculation $H(f, u)$ must be represented on a finite array of points. Ignoring for the present the possible three-dimensional nature of $f$ and $u$ the discrete version $H^D(j, k)$ may be represented as

\[
H^D(j, k) = H(f_j, u_k) = Y(f)_{X(u) = \delta(|u|-u_k)} \quad (5.40)
\]

where

\[
\begin{align*}
  j &= j \frac{f_{\text{max}}}{L} \quad j = -L \ldots L - 1 \quad (5.41) \\
  k &= k \frac{f_{\text{max}}}{L} \quad k = 0 \ldots L - 1 \quad (5.42)
\end{align*}
\]

and $L$ is the number of points chosen for representation of $H(f, u)$ within the frequency range of interest. Note that $H^D(j, k)$ is defined for both positive and negative $j$, and has one more negative frequency than positive. This asymmetry is needed to allow for later use of the DFT.

5.3.3 Identification of $H(f, u)$ with fewer input tests

Considering that computation of $H(f, u)$ increases with $L^4$ (for the three dimensional case), there is a considerable incentive to find ways of reducing the number of calculations required. One way of achieving this is to reduce the number of input frequencies $u_k$ for which the system is tested. The resulting characteristic may be interpolated to occupy all $L$ points. The disadvantage of this is a significant reduction in the accuracy of the result.
An alternative approach is to test the system with blocks of input frequencies. This will be referred to as piecewise spectral analysis. The advantage of this method is that for relatively simple systems (few time-varying parameters) the accuracy of the result is largely preserved.

5.3.4 Piecewise spectral analysis

For this method spectral space is divided into $M$ blocks of $K = L/M$ points each. $H(f, u)$ is then calculated from the output obtained for each block of input frequencies, instead of a single input frequency. Equation (5.40) becomes

$$H^P(j, k) = Y(f_j)^X(w) = B(w)$$

(5.43)

where

$$B(w) = \sum_{l=0}^{K-1} \delta(|w| - (nK + l))$$

(5.44)

and

$$n = \text{int}\left(\frac{k}{K}\right) \quad k = 0 \ldots L - 1$$

(5.45)

The start frequency for each block will be $u_n = n \frac{f_{\text{max}}}{M}$, and the offset frequency within the block will be $u_l = l \frac{f_{\text{max}}}{L}$. Due to the use of these frequency blocks $H^P(j, k)$ no longer needs to be defined for all $k$. Instead $k$ may be replaced with the block number $n$.

$$H^P(j, n) = H^P(j, k) \quad nK \leq k < (n+1)K$$

(5.46)

This reduces the number of points required to represent the system characteristic by a factor of $K$.

The use of this piecewise spectral analysis changes the results in a manner that is best illustrated graphically. Below is a modified version of Figure 5.3 (upper quadrant only) showing the form $H^P(j, k)$ takes when plotted against $j, k, n$.

![Figure 5.8](image)

*Figure 5.8 Piecewise version of Figure 5.3 (upper quadrant only), showing overlap of $G_0(f)$ and $G_1(f)$ and intersection of $G_d(f)$ and $G_d(f)$.*

This diagram shows the division of $u$ into blocks of frequencies. Within each of these blocks $H(f_j, u_k)$ is the same for all $v_k = v_n + v_l$ where $n$ is the block number and $l = 0 \ldots K-1$. Each time-varying parameter $f_i$ shifts these input frequency blocks and changes the magnitude and phase of each frequency according to the appropriate function $G(f_i)$. The result is that $H(f_j, u_k)$ is determined accurately for all $f_i$, but suffers a loss of accuracy (by a factor of $K$), along the $u_k$ axis.

Compared to testing all input frequencies separately the use of a piecewise block method means two main differences.
• The error at the intersection of two characteristics now involves a whole block instead of just one point.
• If two characteristics are too close to each other the blocks overlap, with a resulting error.

An alternative way of viewing these effects is by comparing input and output spectra. Below, shown for two dimensional frequency are possible outputs for the functions $G_0(f)$, $G_a(f)$, and $G_b(f)$ as plotted in Figure 5.8.

For the output spectrum any frequencies that are still within the input block boundary are deemed to belong to the time-invariant characteristic $G_0(f)$. All frequencies outside the boundary are seen as having been affected by time-varying parameters. For the first and third blocks shown above ($G_0(f)$, $G_b(f)$) this approach works. However, for $G_a(f)$ some frequencies appear inside the boundary and some outside. Consequently there will be a confusion of $G_a(f)$ with $G_0(f)$. This is the same effect that caused the overlap of blocks in Figure 5.8.

This confusion is only of importance where it occurs with $G_0(f)$, as we are not concerned with separately determining $G_a(f)$, $G_b(f)$, ... these all being lumped together as the time-varying characteristics. The problem may be avoided entirely if all time-varying parameters have a frequency greater than the block size.

$$f_i > \frac{f_{\text{max}}}{M} \quad i = 1 \ldots N$$

(5.47)

where $f_i$ is the frequency offset for $G_i(f)$, $f_{\text{max}}$ the frequency testing limit and $M$ the number of blocks. For the systems under study this requirement is normally easily satisfied.

The system characteristic functions defined by (5.14)-(5.18) and (5.26) may now be converted to a discrete form and expressed in terms of $H^m(j, n)$, as
106  CHAPTER 5  NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

\[ T(j) = \left| \mathcal{H}^R(j, \text{int}\left(\frac{j}{K}\right)) \right| \]

(5.48)

\[ O^a(j) = \sum_{n=0}^{M-1} \mathcal{H}^R(j, n) - T(j) \]

(5.49)

\[ I^a(n) = \sum_{j=-L}^{L-1} \mathcal{H}^R(j, n) - \sum_{l=1}^{K-1} \left| \mathcal{H}^R(nK+l, n) \right| \]

(5.50)

\[ O^c(j) = \sum_{n=0}^{M-1} \mathcal{H}^R(j, n) \]

(5.51)

\[ I^c(n) = \sum_{j=-L}^{L-1} \mathcal{H}^R(j, n) \]

(5.52)

\[ W(j) = \sum_{n=0}^{M-1} \mathcal{H}^R((nK-j)\%L, n) \]

(5.53)

where \( \% \) represents the modulus function.

The real advantage of the piecewise block analysis is that although the resolution of \( I(n) \) and \( f(n) \) is reduced by a factor of \( K \), the resolution of the other functions is preserved. To illustrate this the example shown earlier in Figure 5.5 is reproduced here in the modified form defined by (5.43).

\[ H(f_j, u_n) \]

(5.48) - (5.53) are shown below for the system in Figure 5.4. (Note that although these have been plotted as continuous they are actually discrete).

Comparing Figure 5.11 with Figure 5.6 and Figure 5.7 a number of observations can be made:

- The intersection error in the Y->Y path (Figure 5.11 (a)) has increased markedly.
- The input functions \( I_{YY}, I_{YU} \) are of lower resolution (blocky).
- \( W^a(f_j) \) shows blocks instead of single frequencies (delta functions). Some positional accuracy in determining \( f_j \) has obviously been lost.

Most noticeable is the intersection error. This error is large and for this example probably unacceptable. With \( M \) blocks this error will occupy a \( 1/M \) portion of the characteristic. If, for example, \( M = 8 \) an eighth of the characteristic (12%) will be in error. For three dimensional analysis however, the error becomes much less significant, being equal to a \( 1/(4M^3) \) portion. The factor of 4...
arises due to the existence of 8 instead of 2 quadrants in three dimensions. For \( M = 8 \) this means that \( \frac{1}{2048} \) or .05 % will be in error. Consequently, for full three-dimensional analysis the intersection error is generally negligible.

![Image of luminance transfer function](image)

![Image of characteristic functions for (a) \( Y \rightarrow Y \), and (b) \( U \rightarrow Y \) paths of Figure 5.4, as calculated using piecewise spectral analysis.](image)

5.3.5 Implementation of piecewise spectral analysis

This method must be used in conjunction with a computer simulation of the system under test. Such a simulation will normally operate in the image domain. Use of the Fourier transform is thus required to produce the input for the simulation, and transform the output back into a spectrum. For spatio-temporal analysis with \( \mathbf{r} = [x \ y \ t] \) and \( \mathbf{v} = [v_x \ v_y \ v_t] \) this process is shown below.

![Diagram of 3D FFT process](image)

Figure 5.12 Use of spectral analysis with an image domain simulation.
Instead of determining \( H(f, u_n) \) the characteristic functions may be calculated directly from the output spectrum. If this calculation begins by testing the system with block \( n = 0 \) and continues until block \( n = M-1 \) is reached then the functions may be determined in an iterative manner, as follows

\[
T_x(j) = T_{x,-1}(j) + |Y(j, n)|X(j, n) \tag{5.54}
\]

\[
O_x(j) = O_{x,-1}(j) + |Y(j, n)|(1 - X(j, n)) \tag{5.55}
\]

\[
I(n) = \frac{1}{k} \sum_{j=-L}^{L-1} |Y(j, n)|(1 - X(j, n)) \tag{5.56}
\]

\[
O_x'(j) = O_{x,-1}'(j) + Y(j, n) \tag{5.57}
\]

\[
I'(n) = \frac{1}{k} \sum_{j=-L}^{L-1} |Y(j, n)| \tag{5.58}
\]

\[
W_x(j) = W_{x,-1}(j) + |Y((nK - j)\%L, n)| \tag{5.59}
\]

where \( X(j, n) = B(j, n), \quad j = -L \ldots L-1, \quad n = 0 \ldots M-1 \)

and \( B(j, n) \) is defined by

\[
B(j, n) = \sum_{l=-L}^{L-1} \delta[j - (nK + l)] \tag{5.60}
\]

### 5.3.6 The three-dimensional version

To determine the spatio-temporal characteristics all the variables must be replaced with three-dimensional vectors.

\[
\mathbf{j} = [j_x \quad j_y \quad j_z]
\]

\[
\mathbf{n} = [n_x \quad n_y \quad n_z]
\]

Instead of calculation along one dimension the functions must now be evaluated in all three. Thus the input frequency block described by (5.44) becomes a three dimensional version \( B_3(j, n) \)

\[
B_3(j, n) = \sum_{l_x=-L_x}^{L_x-1} \delta[j_x - (n_xK_x + l_x)] \sum_{l_y=-L_y}^{L_y-1} \delta[j_y - (n_yK_y + l_y)] \sum_{l_z=-L_z}^{L_z-1} \delta[j_z - (n_zK_z + l_z)] \tag{5.61}
\]

where

\[
\mathbf{j} = [j_x \quad j_y \quad j_z] = [-L_x \quad -L_y \quad -L_z] \ldots [L_x-1 \quad L_x-1 \quad L_z-1]
\]

\[
\mathbf{n} = [n_x \quad n_y \quad n_z] = [0 \quad 0 \quad 0] \ldots [M_x-1 \quad M_y-1 \quad M_z-1]
\]

It can be seen why only the one-dimensional notation has been used so far.

In order to create a real valued image it is necessary for the input spectrum to be symmetric about zero. This means that the frequency blocks must be produced in pairs, for negative and positive frequencies. For the one dimensional case \( B(j, n) \) automatically achieves this. For the three-dimensional case \( B_3(j, n) \), as defined above, does more than this, producing a frequency block in all 8 quadrants. This would be acceptable if the characteristics were identical in all 8 quadrants. However they are not (eg. the position of \( U/V \) chrominance sub-carriers). Thus in order to correctly distinguish the areas of \( H(f, v) \) in different quadrants each \( B_3(j, n) \) must be split into four separate sets of two blocks. Each set of two blocks will be in diagonally opposite quadrants. This new function will be denoted by \( B_3^q(j, n) \).

\[
B_3^q(j, n) = \sum_{l_x=-L_x}^{L_x-1} \delta(s_x^q j_x - (n_xK_x + l_x)) \sum_{l_y=-L_y}^{L_y-1} \delta(s_y^q j_y - (n_yK_y + l_y)) \sum_{l_z=-L_z}^{L_z-1} \delta(s_z^q j_z - (n_zK_z + l_z)) \tag{5.63}
\]
5.4 RESULTS

where \( s^q = [s_{sq}^q, s_{sq}^q, s_{sq}^q] \quad q = 1, 2, 3, 4 \)

and \( s^q = [1 1 1], s^2 = [-1 1 1], s^3 = [1 -1 1], s^4 = [1 1 -1] \)

The full process for determining the three-dimensional spatio-temporal characteristics of a colour television system can now be outlined below in Figure 5.13.

As may be seen considerable computation is required for producing the full characteristics of the chosen system.

Choosing values for some of the parameters provides an indication of the level of computation required. In order to take advantage of the Fast Fourier Transform all parameters are chosen as powers of two. For \( L = [32 16 16], M = [8 8 8], K = [2 2 2] \) the FFT will need to be of size 64 x 32 x 32 and the characteristic functions will each contain 65 536 points. The number of simulations will be \( No. \) components \( x M_x x M_y x M_z x No. \) quadrants/2 = 3 x 8 x 8 x 8 x 4 = 6144.

Despite the enormous computation obviously required the task is feasible. Most of the spectral results presented in this document were produced using the above parameter values. For a simulated television system of average complexity calculation of the complete characteristics takes around 20 hours on three SUN SparcStation 2s (one for each component).

Using a new multi-dimensional simulation system developed by the author (see Appendix E) it has been possible to implement the algorithm of Figure 5.13 and determine the spatio-temporal characteristics for several PAL coding techniques. Presented here are some of the results obtained.
from an evaluation of standard B,G PAL, Holoch's I-PAL [Holoch 1985], and ColourPlus in film mode (no motion adaption) [Vreeswijk 1993].

Figure 5.14-Figure 5.19 represent a subset of the functions described by equations (5.27) - (5.32) with each being displayed as a three-dimensional contour plot showing the areas of the spatio-temporal frequency characteristic that are above the stated threshold level (e.g. -3 dB). This display was done using PHIGSdraw, another tool developed by the author (described in Appendix F). The characteristics have been normalised relative to their maximum amplitude in the standard PAL signal. With this display method it is normally necessary to view a range of three-dimensional contour plots to gain a full appreciation of the characteristic. Space does not permit this here, and instead only one contour is shown for each function. Despite this it is possible to interpret much from what is shown.

Figure 5.14 shows the characteristics for the Y, U, V paths in a simulation of conventional PAL [8]. For PAL time-varying effects are negligible in these paths. From these plots the considerable difference between luminance and chrominance horizontal bandwidth is obvious, as also is the effects of the PAL delay line - removing intermediate vertical chrominance frequencies.

Figure 5.14 Spatio-temporal linear characteristics of Y, U & V paths in B,G PAL.
Figure 5.15 Spatio-temporal cross-chrominance characteristics for Y \(\rightarrow U, V\) paths in B.G PAL.

Figure 5.15 and Figure 5.16 show PAL's cross-effect characteristics. On the second line are the image frequencies introduced by cross-effects, and on the top line the main input frequencies that cause them. From Figure 5.15(a), (b) it is evident that cross-chrominance is caused by different areas of horizontal-vertical frequency for the different components of $U$ and $V$. These different areas will correspond to the positions of the $U$, $V$ sub-carriers. Due to the smaller relative amplitude of the $U$ component (compared to $V$) the cross-effects are greater in Figure 5.15(c).

In Figure 5.16 it can be seen that the high horizontal frequencies of cross-luminance (Figure 5.16(c), (d)) are caused by low horizontal chrominance frequencies (Figure 5.16(a), (b)). Very low horizontal frequency causes no cross-luminance (at the threshold shown), likely because of the decoder luminance notch. At first glance these results appear to show simply a horizontal shift of frequencies between input and output. However the actual values of the shifts, as shown by Figure 5.16(e) and (f) are not just horizontal. These two plots, as we would expect, identify the $U$ and $V$ sub-carrier positions in spatio-temporal frequency space - cross-luminance consists of chrominance frequencies shifted by the spatio-temporal offsets of $[284 \pm 78 \pm 18\%]$ and $[284 \pm 234 \pm 6\%]$. 
These results for PAL can be compared with those for I-PAL and ColourPlus. I-PAL uses line-based time-segregation of luminance and chrominance. Consequently most of its effects appear in the vertical direction. ColourPlus, on the other hand, uses field-based time-segregation, with coding effects occurring in a vertical-temporal direction. Unlike I-PAL ColourPlus also performs pre and post-filtering of the signal. Note that Figure 5.17(c), (d) show evidence of the finite block size used in the piecewise analysis.

Comparing Figure 5.17(a), (b) with Figure 5.14(a) the enhanced techniques both show improved horizontal bandwidth, though at the expense of a reduction in bandwidth elsewhere.
5.4 RESULTS

Figure 5.17 Spatio-temporal characteristics of the luminance path in I-PAL and ColourPlus (no motion adaption), showing both linear and non-linear effects.
CHAPTER 5 NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

(a) $T_{uu}(v)$, -3 dB

(b) $T_{uv}(w)$, -3 dB

(c) $I_{uu}(w)$, -3 dB

(d) $I_{uv}(w)$, -12 dB

(e) $O_{uu}(v)$, -3 dB

(f) $O_{uv}(v)$, -12 dB
5.4 RESULTS

Figure 5.18 Spatio-temporal characteristics of the chrominance V path in I-PAL and ColourPlus (no motion adaption), showing both linear and non-linear effects.

Figure 5.19 Spatio-temporal characteristics of cross-effect paths in I-PAL and ColourPlus (film mode) (Y \rightarrow U & U \rightarrow Y).

Unlike standard PAL both techniques contain time-varying parameters in the wanted signal (intra-component) paths which are responsible for the appearance of aliasing. For I-PAL intermediate vertical frequencies (Figure 5.17(c)) alias to low and high vertical frequencies (Figure 5.17(o)). In
CHAPTER 5 NEW METHODS FOR REPRESENTING AND CALCULATING THE COMPLETE SPATIO-TEMPORAL CHARACTERISTICS OF A CODING SYSTEM

ColourPlus it is intermediate vertical-temporal frequencies that cause aliasing (Figure 5.17(d)). Due to the pre-filtering used the level of aliasing is much lower in ColourPlus, requiring a threshold of -12 dB before it becomes visible in the contour plots.

Figure 5.18 shows a similar set of plots for the U component. Similar shaping of the characteristics is evident, and once again the level of aliasing is lower for ColourPlus. The nature of the aliasing is indicated by the values of $f$, shown in Figure 5.18(e), (f). For both coding techniques there is a value at zero frequency, corresponding to the unaliased, wanted signal. The time-varying (alias) parameters exist at $[0 \ 156\% \ 0\ 25\%]$ for I-PAL and $[0 \ 312\% \ 0\ 0\ 0\ 25]$ for ColourPlus.

Lastly Figure 5.19 shows some of the output cross-effect characteristics for the enhanced techniques. Comparing Figure 5.19(a), (b) with and Figure 5.19(c), (d) with Figure 5.16 it is obvious that cross-effects have been significantly reduced. Once again ColourPlus has better performance - in Figure 5.19(b) no cross-chrominance can be seen above -3 dB.

Apart from a few points in some plots most of these characteristics appear to have been little affected by the errors discussed previously. This verifies the prediction that, for low $N$ and three-dimensional functions errors should not be significant.

5.5 Summary

This Chapter has attempted to develop a solid mathematical platform for the analysis of enhanced television coding techniques. A television system is represented as a Linear-Periodically-Time-Varying Many-Input-Many-Output system. For analysis purposes the characteristics of this system can be simplified into 30 three-dimensional spatio-temporal functions that describe the wanted signals, main signal artefacts, and cross-component artefacts. This representation is intended to replace the incomplete and confusing approaches that have been used to date with one that is self-consistent and accurate. Also described is a numerical method for determining the complete spatio-temporal characteristics of a coding system. This method has been proved by its use in the analysis of many coding techniques. Results obtained for several techniques, presented as spectral contours, illustrate the considerable insight this representation can offer. So useful has this representation been found that it has been used throughout this thesis for explaining coding concepts (see Chapters 3 & 6). The results obtained are reliable and quantitative, and lend themselves to further use in the automated appraisal of coding quality (see next Chapter).

Finally the author would like to propose that the representation used in this Chapter be used as a standard model for future analysis and comparison of enhanced coding techniques.
CHAPTER 6
A NEW APPROACH TO OBJECTIVE ASSESSMENT

As discussed in Chapter 2 the problem of objective assessment is one that has not been solved satisfactorily. In studying different television coding techniques and the many variations possible in each, the number of performance variables is large (e.g. cross-effects, resolution, aliasing, etc). As the performance of the coding technique is ultimately determined by viewer perceptions accurate rating of different implementations is very difficult without comparative subjective tests.

As discussed in section 2.9 these subjective tests have many problems, not least being the fact that they are labour intensive and limited in the number of comparisons that can be made. The author’s motivation for the development of a new objective technique came from two directions.

First the facilities and resources for performing CCIR subjective tests were not available. The large number of techniques and implementation being studied (Chapters 4) would likely have been impractical to evaluate anyway.

Second, the author was looking for a way of summarising the enormous amount of information provided by the newly developed spectral analysis (Chapter 5). For specific detailed analysis of coding techniques the detailed spatio-temporal information provided is ideal. However, for making general overall comparisons of performance the complete spatio-temporal characteristics represent a level of information overload.

The result has been work on the development of an entirely new approach to objective assessment. In common with previous approaches this new method uses models of the eye’s response to estimate the visibility of different spatio-temporal frequencies. The difference with the new approach is what these response models are applied to. Conventional objective assessment compares differences or ‘distortions’ between input and output test pictures. The new method uses no test pictures, but instead provides a prediction of expected image quality based upon a HVS response weighted measure of the spatio-temporal characteristics of the coding technique. How this achieved will now be explained and results from using this new approach will be presented. Also presented will be a comparison with results obtained using conventional subjective assessment. It is also perhaps worth noting here that the objective assessment technique described provides results for a full colour image, unlike most previous attempts at objective assessment [Budrikis 1972, Sarikson 1977, Granrath 1981, Lukas 1982, Miyahara 1988, Lowe 1993, Mishina 1995] which have been applied to only greyscale images.

This new method is by no means fully developed and the following sections discuss in detail many of the issues that remain unresolved. The method’s promise is, however, hopefully demonstrated. Results from the method as it is described here have been used here and in other chapters (Chapter 7) as an aid in comparison and appraisal of coding techniques.
6.1 Using the spatio-temporal characteristics to obtain a quality measure

It stands to reason that the coding quality of a system will be related to its spatio-temporal characteristics. The larger the volume of the spatio-temporal characteristics for the chrominance and luminance paths the more information or bandwidth the system can carry and thus the higher the signal fidelity will be. Similarly the smaller the volume of the cross-component and aliasing characteristics the lower the level of introduced picture artefacts should be. In practice there are two further factors to be considered.

1. The energy in and statistical occurrence of the different spatio-temporal frequencies.
2. The relative visibility and importance of different spatio-temporal frequencies to the human observer.

If the coding system loses or distorts frequencies that are either unimportant to the HVS, or very rare in occurrence then the overall impact on image quality will be small. Similarly a small loss of very common or visually significant frequencies may have a large impact on picture quality.

In studying these factors further it is helpful to consider the wanted signals (luminance and chrominance) and unwanted signals (cross-effects and signal aliasing) separately. We will now look at the first factor mentioned above in detail.

6.2 The effects of spatio-temporal energy distribution

For the wanted signals of luminance and chrominance if certain high spatio-temporal frequencies are seldom present, or only present at very low levels then the loss of these frequencies in a coding system is likely to have little effect on overall image quality. This would suggest that any quality measure using the spatio-temporal characteristics should be weighted according to the statistical likelihood of each luminance chrominance frequency in the source image.

Some measurements have been made of spatio-temporal energy in television sequences [Eckert, 1992, Cortelazzo et al, 1991]. These showed a fairly consistent inverse relationship between both temporal and spatial frequencies and the average image energy. The higher the spatio-temporal frequency the less energy it has. What is surprising is the steepness of this inverse relationship, with the energy of the maximum temporal and spatial frequencies in a 50Hz 762x384 sequence being around $10^{-5}$–$10^{-6}$ times less than for the lowest signal frequencies [Eckert, 1992]. If used as a weighting function for the wanted signals this distribution would cause high spatio-temporal frequencies to "have an incredibly tiny weighting (or significance), suggesting their effects on overall image quality must be close to zero. Similarly, for the unwanted signals the distribution suggests that the relative energy of high spatio-temporal frequencies is so low that any artefacts they cause can be practically ignored.

Actual experience contradicts such conclusions. As an example, the loss of horizontal luminance frequencies above 224cph (3.5MHz) causes a very noticeable softening in picture detail. Higher frequencies are certainly less important that lower frequencies, but, the author suggests, not to the degree suggested by their statistical distribution. It is likely that the information extracted by the Human Visual System (HVS) from different spatio-temporal frequencies is out of proportion with the statistical energy represented by these frequencies. In other words high spatio-temporal frequencies, even with very low relative image energy convey considerable important information to the HVS. The exact relationship is unknown to the author and needs further investigation.

For the unwanted signals the statistical distribution might seem more relevant, as the artefacts carry no real image information to the HVS i.e. they have no perceptual meaning. However, once again, the statistical distribution does not appear to match subjective perception of artefacts properly. For example the moving high diagonal luminance frequencies that cause the most visible coarse cross-colour have an extremely low statistical energy. Despite this, such cross-colour still remains a significant visual artefact. The spatio-temporal distribution of image frequencies [Eckert, 1992, Cortelazzo et al, 1991] suggests that the tiny energy levels of such high luminance frequencies would more than compensate for the higher visibility of the artefacts they cause. In practice, although these
artefacts occur in only small picture areas and only on occasional picture material they can be quite disturbing to the viewer. It is likely that the human observer does not average out these artefacts over the rest of the image, or over the time between their occurrence.

All this discussion illustrates the difficulty of applying statistical weighting to a quality measure based upon spatio-temporal characteristics. Obviously there must be some correlation, artefacts that occur often are worse than ones that occur rarely. However, the best way to use information on the statistical energy is unclear. More information is needed on the importance of different spatio-temporal frequencies to the HVS and on how the HVS responds to artefacts. It is likely that such information does not properly exist as yet within the body of knowledge on the HVS. In light of this uncertainty on how to apply a statistical weighting the author has decided to omit the use of statistical weighting until an appropriate model can be found. Instead the objective assessment will be divided into two categories: one for stationary (no motion) image content, the second for moving (motion) image content. This simple partition circumvents to some degree the need to decide the relative importance of high versus low temporal frequencies. Instead the relative importance of results for stationary and moving image content will be left for later determination.

The second factor in determining our quality measure will now be looked at.

6.3 The relative response of the HVS to spatio-temporal frequency

One of the most important factors for any objective assessment technique is taking into account the importance and visibility of different picture content to a human observer. Once again this needs to be considered separately for the wanted and unwanted signals.

6.3.1 The response for the unwanted signals

For the wanted signal there are two points to consider:

1. Can the HVS perceive a particular spatio-temporal frequency.
2. How important is each spatio-temporal frequency.

Whether the HVS can perceive a particular spatio-temporal frequency is dependent upon whether it is above the visual threshold and whether it is within the acuity limits of the eye.

In Chapter 2 (Table) it was shown that, at normal viewing distances, the spatio-temporal limits of the television system are well within the acuity limits of the eye. However, if a particular frequency has low amplitude it may still be below the visual threshold. Some advanced coding techniques (e.g. MPEG, JPEG compression) take advantage of this fact.

Adaptive coding based on the magnitude of different signal frequencies is, in the field of PAL coding, quite primitive, and as stated in Chapter 5 the spatio-temporal characteristics can be calculated for non-adaptive systems only. Consequently as the amplitude of each frequency is unknown we must assume for now that all spatio-temporal frequencies can be perceived (i.e. they are within HVS acuity limits and above the visual threshold).

The second question is far more difficult and is related to the knowledge required in the previous section. The threshold characteristics described in Chapter 2 showed an HVS response curve of decreasing response with increasing spatio-temporal frequency. The wanted signals however, will normally be significantly above the visual threshold. Limited supra-threshold testing [Bowker, D.O., 1983] has showed a flattening of the response with frequency above visual threshold. Active mechanisms in the eye compensate for the high frequency rolloff. Unfortunately a general form for the supra-threshold response is not known. Determining such a response is complicated by two further factors.

First, despite a poor visual response for high spatio-temporal frequencies perceptually they are quite important. These frequencies carry the edge information that is critical to image understanding and object recognition [Uttal 1981]. Second, although there is a reduction in response for high temporal frequencies as well, this can be completely negated if the viewer's eye tracks the movement. Such
tracking effectively eliminates or 'zeros' any temporal offsets by causing the picture to become stationary on the viewer's retina.

Considering all these factors the author was unable to decide upon a consistent approach for determining relative importance of different frequencies and the wanted signals. The best we can say is that, in general, higher frequencies are less important than lower frequencies i.e. average image quality suffers more if we remove lower spatio-temporal frequencies than if we remove higher ones. In the author's opinion further research is needed to accurately determine this relationship.

Due to the lack of information in this area the author has omitted any spatio-temporal frequency weighting for the wanted signals. Instead the quality measure will be for the present be simply a spectral bandwidth measure. This measures the volume of a -3dB boundary volume for the spatio-temporal characteristics and is somewhat equivalent to stating the -3dB bandwidth for a 1-D signal system. Theoretically the larger the -3dB volume (i.e. spatio-temporal bandwidth) the fewer signal frequencies will be lost and the greater the image fidelity will be. This volume measure has been modified slightly to try and account for the differing importance of high and low frequencies.

![Figure 6.1 Illustration of modified bandwidth measure](image)

As Figure 6.1 illustrates, two different characteristics can have the spectral volume (or in 2-D area) yet contain frequencies of differing importance. For this reason a modified volume measure only measures 'base' volume i.e. volume contained in this characteristic up to the first -3dB boundary along each axis. Using this approach the higher horizontal vertical frequencies in Figure 6.1(a) are not included in the volume measure causing Figure 6.1(b) to give a higher quality measure than (a). This helps to indicate that (b) loses more of the important low diagonal frequencies than does (a). The resulting bandwidth measure still suffers from the basic lack of a HVS importance weighting function, and as such may not be a particularly good objective quality measure. The bandwidth measure also suffers from an assumption of the equivalence of temporal and spatial frequencies. Doubling the vertical bandwidth and halving the temporal bandwidth will likely result in the same bandwidth measure. Whether the resulting picture quality is the same is unlikely. The partitioning of the quality measures into motion and no-motion measures should help to reduce this problem. For
no-motion the bandwidth measure becomes effectively a spatial area measure (no temporal frequencies are included) and the problem of spatial/temporal equivalence does not arise.

6.3.2 The response for the unwanted signals

Deciding upon satisfactory HVS weighting for the unwanted signals (i.e. cross-effects, aliasing, etc) is much easier than for the wanted signals. First, for most television coding techniques the unwanted signals introduced are at or near visual threshold most of the time. Second, these artefacts contain no perceptual information. They represent no recognisable picture content and in many ways are quite random in both form and movement. As such they are less likely to be affected by the higher perceptual areas of the HVS (e.g. object recognition, edge detection). Third, largely because of the above facts the eye is unlikely to track any movement (i.e. temporal frequency offsets) in the unwanted signals. As their motion is generally fast and unpredictable it will be very difficult for the viewer’s eye to track the artefacts even if they so wished. All these factors suggest that the visual threshold response models described in Chapter 2 should provide a satisfactory weighting function for the overall subjective visibility of unwanted signals in both the luminance and chrominance.

6.4 Calculating an objective quality measure

The implementation of the objective assessment method will now be described. The results for the method will be presented as two sets of numbers for each of the wanted or unwanted signals, one set for motion and the second for no-motion. These numbers will be known as figures-of-merit (FOM). The intention is that the higher the figure-of-merit for the wanted signals and the lower the figures-of-merit for the unwanted signals the better the predicted subjective image quality for that coding technique.

Figure 6.2 No-motion spectral characteristics for conventional PAL coding system.
In calculating the figures-of-merit the first step is to determine the coding system’s full spatio-temporal characteristics. This is achieved using the methods described in Chapter 5 with one difference. Instead of determining the complete characteristics in one go two sets are calculated, one for input frequencies with zero temporal offset (no motion), the second for input frequencies with non-zero temporal frequencies (motion).

Note that although the input frequencies may have zero temporal offset the output spectra can contain temporal frequencies. Figure 6.2(b) shows the cross-luminance caused by stationary spatial frequencies in conventional PAL coding. In practice, as Figure 6.2 indicates the no motion input spectra will contain some low temporal frequencies due to the discrete size of the blocks used in the spectral analysis. With the parameters used for all simulation results presented in this chapter the spectra are divided into eight blocks in each dimension, meaning the stationary spectra will actually contain temporal frequencies within a range of ±3.125Hz. As has been indicated the objective assessment is performed differently for the wanted and unwanted signals.

6.4.1 Calculating the figure-of-merit for a wanted signal

The wanted signals are the Y, U, V signals without aliasing or distortion. These are represented according to the analysis of Chapter 5 and equation 5.27 as $T_{YY}$, $T_{UU}$, $T_{VV}$. The figures-of-merit are found by determining the proportion of spatio-temporal frequencies that are above a given threshold for the motion ($m$) and no motion ($nm$) spectra. This provides a measure of the spatio-temporal bandwidth of the signals. This calculation is represented by equations (6.1) and (6.2).

$$F_{nm}^{rr} = \frac{\sum_{v_t} \sum_{v_i} \left( T_{rr}(v_t, v_i, v_r) \geq T \right)}{k_N N_r} v_t = 0$$

$$F_{nm}^{rr} = \frac{\sum_{v_t} \sum_{v_i} \sum_{v_r} \left( T_{rr}(v_t, v_i, v_r) \geq T \right)}{(N_r - K) N_N N_r} v_t \neq 0$$

where $T$ is the threshold (typically set at the $\frac{1}{2}$ poser point i.e. -3dB), $N_N$, $N_r$, $N_r$ are the number of points in the characteristic in each dimension, and $K$ is the number of points in each block (see section 5.33). For implementation these equations are modified slightly to ensure only the ‘base’ bandwidth as defined in section 6.3.1 is being measured.

6.4.2 Calculating the figure-of-merit for the unwanted signal

The unwanted signals are the cross-effects $O^c(v)$ (eqn 4.65) and intra-component artefacts $O(v)$ (eqn 4.63). The figure-of-merit for each of these functions is calculated by summation after weighting by the appropriate visual response characteristic. For example the calculation of the $Y \rightarrow U$ cross-colour figure-of-merit is defined by

$$F_{nm}^{nu} = \sum_{v_t} \sum_{v_i} S_{uy}(v_t, v_i, v_r) \frac{O^c_{nu}(v_t, v_i, v_r)}{v_t = 0}$$

$$F_{nm}^{nu} = \sum_{v_t} \sum_{v_i} \sum_{v_r} S_{uy}(v_t, v_i, v_r) \frac{O^c_{nu}(v_t, v_i, v_r)}{v_t \neq 0}$$

where $S_{uy}(v_t, v_i, v_r)$ is the $U$ chrominance threshold model described in section 2.82. Thus the figure-of-merit represents the sum of the relative contributions of each spatio-temporal frequency in the characteristic - where a frequency’s relative contribution is determined by (i) the magnitude of the characteristic and (ii) the visual threshold for that frequency. If a characteristic has a high value (i.e. close to 1) over a large volume of spatio-temporal frequency space and these same frequencies have a low visual threshold (i.e. high visibility) then the resulting figure-of-merit will have a high value,
meaning that this particular unwanted signal is likely to be present in high subjectively perceivable levels in decoded images.

The result of all these calculations is a set of 24 values, as listed in Table 6.1. These values are intended to provide an indication of the subjective performance of the coding system in each of 12 categories, for both motion and no motion.

<table>
<thead>
<tr>
<th></th>
<th>Motion</th>
<th>No Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>resolution/bandwidth</td>
<td>$F_{yy}^m$</td>
<td>$F_{yy}^m$</td>
</tr>
<tr>
<td>aliasing</td>
<td>$F_{ya}^m$</td>
<td>$F_{ya}^m$</td>
</tr>
<tr>
<td>cross-luminance</td>
<td>$F_{YU}^m$</td>
<td>$F_{YU}^m$</td>
</tr>
<tr>
<td>cross-colour</td>
<td>$F_{YV}^m$</td>
<td>$F_{YV}^m$</td>
</tr>
<tr>
<td>U/V cross-talk</td>
<td>$F_{UV}^m$</td>
<td>$F_{UV}^m$</td>
</tr>
</tbody>
</table>

Table 6.1 The full set of motion and no-motion figures-of-merit calculated by the author's objective assessment method.

6.5 Normalising the figures-of-merit

Taken in isolation one set of figures-of-merit means little, similar to the way subjective testing results for one picture and one coding technique will mean little. The figure-of-merit's true value is as a comparison tool. It is thus useful to normalise the values to a standard set.

In this case we will choose a conventional PAL coding system. The objective assessment results obtained for PAL will be used as normalisation values for all other systems, with the figures-of-merit being expressed as a percentage of those for conventional PAL. For example a value of 110 for $F_{yy}$ would mean a coding system has 10% greater spatio-temporal luminance bandwidth than conventional PAL.

Producing these normalised values is not simply a matter of dividing each figure-of-merit by a reference figure-of-merit. Some of the normalisation values such as $F_{ya}$ for aliasing are zero for conventional PAL. Using these values for normalising other figures-of-merit is thus meaningless. Also, normalising each figure-of-merit individually loses information on the relative significance of different figures-of-merit when compared to each other, for example the relative significance of chrominance artefacts such as chrominance aliasing, cross-colour and U/V cross-talk.

As discussed earlier the relative significance of wanted versus unwanted signals is unknown. Also unknown is the relative importance of luminance and chrominance information and artefacts. Once again these are areas needing further research. It is, however, reasonable to assume that different unwanted signals which appear in the same component (e.g. luminance aliasing, cross-luminance) are being affected by the same HVS characteristic, and may thus be directly compared with each other. The figures-of-merit can thus be divided into four groups. Within each group the figure-of-merit values will all be normalised to the same reference value. The reference value used will be the most significant artefact in the reference system. The resulting reference values and signal groups are shown in Table 6.2.

It is occasionally useful to study chrominance effects separately for $U$ and $V$ (e.g. K-PAL coding can treat $U$ and $V$ differently - see section 7.6.8). Generally speaking however $U$ and $V$ are treated identically in all coders. It is thus helpful to simplify the objective assessment results further by combining $U$ and $V$ into a single measure of chrominance. Objective assessment results can then be summarised with two sets of seven values as shown in Table 6.3.
124 CHAPTER 2 A NEW APPROACH TO OBJECTIVE ASSESSMENT

Signal Group | Relevant FOMs | Reference FOM
---|---|---
Luminance bandwidth | \( F_{yy} \) | \( F_{yy} \)
Chrominance bandwidth | \( F_{uu} \), \( F_{vv} \) | \( F_{vv} \)
Luminance artefacts | \( F_{ys}, F_{ys} \) | \( F_{ys} \)
Chrominance artefacts | \( F_{uu}, F_{us}, F_{vu}, F_{uv} \) | \( F_{vu} \)

Table 6.2 Grouping of figures-of-merit for normalisation and relative comparison.

<table>
<thead>
<tr>
<th></th>
<th>Motion</th>
<th>No Motion</th>
</tr>
</thead>
</table>
luminance bandwidth | \( F_{yy}^{m} \) | \( F_{yy}^{n} \)
chrominance bandwidth | \( F_{cc}^{m} \) | \( F_{cc}^{n} \)
luminance aliasing | \( F_{ys}^{m} \) | \( F_{ys}^{n} \)
chrominance aliasing | \( F_{us}^{m} \) | \( F_{us}^{n} \)
cross-luminance | \( F_{ct}^{m} \) | \( F_{ct}^{n} \)
cross-colour | \( F_{uc}^{m} \) | \( F_{uc}^{n} \)
U/V cross-talk | \( F_{cc}^{m} \) | \( F_{cc}^{n} \)

Table 6.3 Set of combined figures-of-merit from table Table 6.2, reducing their number from 24 to 14.

Once again these will be normalised w.r.t. the largest PAL figure-of-merit in each category.

### 6.6 Results using objective assessment

The objective assessment technique described has been applied to a wide range of coding techniques by the author. Space does not permit presentation and discussion of all these results here. Instead a small subset consisting of conventional PAL, WC-PAL coding, and CAFPAL (basically non-adaptive ColourPlus) will be looked at (See Chapter 4 for more on these techniques). Some further examples and results obtained using the author’s new technique are shown in Chapter 7 where these objective assessment figures-of-merit are used to compare the performance of different implementations of K-PAL and WC-PAL.

The separate figures-of-merit for both **motion** and **no-motion** will be plotted under the headings of luminance bandwidth, chrominance bandwidth, luminance artefacts, and chrominance artefacts, as grouped in Table 6.2. Note in each case for PAL the figure-of-merit is automatically 100% with the other techniques having figures-of-merit relative to the PAL result.

Looking first at the bandwidth measures in Figure 6.3 and Figure 6.4 we see that both WC-PAL and ColourPlus (i.e. CAFPAL) show increases in luminance and chrominance bandwidth for most cases. For luminance the increase is relatively small (basically it is just due to the extension of horizontal resolution); for chrominance it is much larger, although the enormous increase in stationary chrominance bandwidth for ColourPlus is somewhat offset by a lower than normal bandwidth for moving chrominance. Comparing luminance and chrominance artefacts we see the real advantages of both enhanced techniques. All artefacts are enormously reduced compared to conventional PAL. For ColourPlus a counterbalance to the improvement in cross-luminance is the introduction of significant levels of chrominance and luminance aliasing. For stationary luminance the luminance aliasing is on a par with cross-luminance in conventional PAL. In practice, however, as explained earlier, these results do not take into account the statistical likelihood of the signals that cause these artefacts. Intermediate chrominance frequencies that cause cross-luminance in PAL will be far more common than the high diagonal luminance frequencies that cause luminance aliasing in ColourPlus.
Overall these objective assessment results provide a good indication of the important performance attributes of the coding techniques. Both enhanced techniques are seen to provide increased resolution and reduced cross-effects. As a summary of the information in the spectral characteristics they are invaluable. Using the figure-of-merit values it is possible, at a glance, to compare the performance of several techniques. A figure-of-merit can provide us with an indication of whether a technique has improved resolution, decreased cross-effects, and other introduced artefacts (e.g. aliasing). However whether the trends shown in Figure 6.3 - Figure 6.6 will accurately reflect perceived image quality is an unanswered question.
To answer this question we must directly compare results from objective and subjective assessment and find a correlation between them.

6.7 Comparison of objective and subjective assessment

Comparing objective and subjective assessment is not a straightforward task for several reasons.

1. The two methods give different types of results. Subjective assessment provides one overall quality measure, whereas the author’s objective assessment gives 7 different quality measures.

2. The two methods operate on different data sets. Subjective assessment measures the quality of a set of output images/sequences. Objective assessment gives quality parameters for the actual coding system.

3. As explained in section 2 subjective results are highly variable and inaccurate in themselves. The results of subjective assessment are not a perfect data set to compare against.

Because of these factors it might appear that we are, colloquially speaking, trying to compare chalk and cheese. However these differences can be bridged to some degree if further knowledge about the test sequences is available. To compare objective and subjective assessment results we need to judge:

1. Whether the sequence contains predominantly stationary or moving content.

2. If the sequence has a critical level of important high frequency luminance or chrominance detail.

3. What artefacts the sequence is most likely to produce.

In performing subjective assessment there are no quantitative measures made of these factors. Generally speaking the sequences are chosen to be critical in one of the above areas, so as to noticeably show up performance differences between coding techniques. However, the degree of stationary vs. moving or high vs. low frequency content is not measured. Also the assessment results consist of a single quality judgement with no information collected on what image aspects determined their judgement. For example, there may be both a loss in luminance frequencies and an increase in U/V cross-talk. For the viewer, however, it may be the hue errors in someone’s face (caused by U/V cross-talk) that causes them to give a lower rating. Added to this different effects might cancel each other. A viewer might give two differently coded sequences the same rating because an improvement in cross-luminance in the second was offset by a loss in chrominance resolution. Whilst acknowledging the inherent ambiguities in any judgements it is, nevertheless, possible to make an informed guess as to the nature of the image content and what aspects affect its quality most.
6.7 COMPARISON OF OBJECTIVE AND SUBJECTIVE ASSESSMENT

6.7.1 Choosing a data set

The next step is to choose a set of coding techniques and test sequences and perform a set of comparative subjective and objective tests. Unfortunately the resources required for performing subjective assessment were beyond those available to the author. Instead it was decided to choose an existing set of published subjective results. Objective assessment results for the same coding techniques could then be produced for comparison. For this the BBC research department report *PAL decoding: multi-dimensional filter design for chrominance -luminance separation* by C. K. P. Clarke [Nov 1988] was chosen. This document compares six PAL decoding techniques (including a conventional PAL decoder) and presents subjective test results for these as tested on 12 image sequences (or stills). The results were tabulated using the CCIR 5 point scale of table 2.7, rating the images from 1-bad to 5-excellent. The decoders tested are listed in Table 6.4.

<table>
<thead>
<tr>
<th>Test Coder [Clarke Nov 1988]</th>
</tr>
</thead>
<tbody>
<tr>
<td>conventional delay-line</td>
</tr>
<tr>
<td>2-line Weston</td>
</tr>
<tr>
<td>2-field comb</td>
</tr>
<tr>
<td>4-field filter</td>
</tr>
<tr>
<td>2-picture Weston</td>
</tr>
<tr>
<td>8-field filter</td>
</tr>
</tbody>
</table>

Table 6.4 The six PAL decoders on which subjective assessment was performed.

An advantage of the results given in this particular document is that a detailed description is provided of the test sequences, describing which artefacts they produce most, and what aspects of the sequence are most testing (e.g. moving chrominance effects, luminance detail, etc). This is invaluable in deciding which objective parameter should be used for comparing with subjective results.

6.7.2 Performing objective assessment

Clarke's paper provides a detailed description for each of the 6 decoders allowing the author to replicate them for simulation using TVPROC (see Appendix E). Through numerical spectral analysis (as described in Chapter 5) each decoder was analysed for a conventionally coded 5.5 MHz (System I) PAL signal. Full spatio-temporal characteristics for each system (consisting of a conventional PAL coder and one of the enhanced PAL decoders) were calculated. Space does not permit presentation of these characteristics here. However by themselves the spectra provide considerable insight into the coding techniques and are very useful performance comparison tools (see Chapter 7).

6.7.3 Comparing the results

Out of the 12 test sequences 3 were stationary pictures and the other 9 contained motion (though these often contain stationary content as well). The names given to these sequences are listed in Table 6.5 below.

Table 6.5 also shows the parameters judged to be the most critical for each sequence - taken from Clarke's description of the sequences (see Appendix G).

For comparison purposes two of these sequences will be omitted. First the Teletext sequence was, according to Clarke "seriously affected by differential phase distortion". It was intended to test for immunity to phase distortion. As the spectral characteristics provide no information on phase response then there can be no expected correlation with this sequence. The second sequence omitted will be Pendulum, for the simple reason that this sequence showed no rating differences for the subjective assessment, and is consequently no use as a comparative sequence.
The task now becomes one of showing a correlation between the subjective ratings and an increase or decrease in the relevant figures-of-merit. It should be expected, for example, that an increase in subjective rating would correlate with an increase in luminance bandwidth or a decrease in cross-colour and cross-luminance. Thus there should be a positive correlation with the wanted signals (luminance bandwidth and chrominance bandwidth) and a negative correlation with the unwanted signal impairments / artefacts. It is unlikely, however, that this correlation is linear. A doubling in system bandwidth will improve picture quality, but not by a factor of 2. Empirically by inspection of the data the author has chosen a log relationship, comparing the subjective ratings against the log of the objective figures-of-merit. Figure 6.7 shows correlation graphs for the Young Couple test sequence, using the no-motion figures-of-merit.

Table 6.5 The 12 test image sequences/stills used by Clarke for subjective assessment.

<table>
<thead>
<tr>
<th>Test Image or Sequence</th>
<th>Motion</th>
<th>Y bandwidth</th>
<th>C bandwidth</th>
<th>Cross-colour</th>
<th>Cross-luminance</th>
<th>U/V cross-talk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young Couple</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orbit</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teletext Graphics</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pendulum</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Captions</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train Set 1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train Set 2</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotating Disk</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast Disk</td>
<td>✔</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show-Jumping</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snooker</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.5 suggests that the critical parameters for the young Couple are luminance bandwidth and cross-colour. Looking at the graphs for these in Figure 6.7 there does indeed appear to be a positive correlation for luminance bandwidth and a negative correlation for cross-colour. The correlation is nowhere near perfect and even with the figure-of-merit being plotted on a log scale the correlation is still not linear, especially for cross-colour. However, in general an increase in subjective rating is matched by an increase in luminance bandwidth and a decrease in cross-colour.
In Figure 6.8 the motion figures-of-merit for the Fast Disk show some positive correlation for chrominance bandwidth, but no obvious trend for the other parameters.

Unfortunately we come here to one of the problems mentioned earlier. In reality several parameters may have contributed to the subjective rating. 2-picture WC-PAL, for example, is responsible for ruining the correlation for chrominance. It has a relatively high chrominance figure-of-merit, but the lowest subjective rating. However it also has the worst U/V cross-talk. Thus if we were to combine the figures-of-merit for chrominance bandwidth and U/V cross-talk we might get a better match.

### 6.7.4 A mathematical approach to determining the correlation

Measuring the degree of correlation between objective and subjective assessment methods for these results is very difficult. For a start the correlation can in no way be expected to be linear, and consequently linear correlation measures cannot be used. Instead we must find an alternative. Ideally the results from objective assessment should act as a predictor of subjective assessment results. This, in its most basic form requires two things:

1. An increase in subjective rating should be matched by an increase in the objective rating.
2. The larger the difference in subjective rating is the larger the difference in objective rating should be.

Thus we may compare the rating for each sequence with those for every other sequence and evaluate their differences in objective and subjective ratings. If, in every case, the rules above hold we may define the subjective and objective ratings as having a high correlation.

Using these rules the author has developed an empirical correlation index (defined in Appendix D) which has a value between -1 and 1. For good agreement between objective and subjective assessment the wanted signals (luminance and chrominance bandwidth) should have an index value of close to 1. For the unwanted signals (e.g. cross-effects) a strong correlation means an index of close to -1, i.e. the lower the cross-effects the higher the subjective rating.

Using this approach indices have been calculated for all the sequences listed in Table 6.5. Figure 6.9 and Figure 6.10 show the results obtained for each image. In these figures only the indices for the relevant performance parameters (as listed in Table 6.5) are shown. The indices used are either the motion or no-motion figures-of-merit depending on whether the sequence is predominantly moving or stationary (e.g. the C=>C bar for Fast Disk shown in Figure 6.9 is for the motion figure-of-merit).

In analysing these results it was found that two of the moving sequences in fact showed far stronger correlation for the no-motion figures-of-merit than for the motion. These two sequences were Captions and Interview. This suggested that the subjective judgement of these sequences was
actually determined mainly by non-moving effects. The nature of these sequences backs this up. Captions consists of ‘...white letters scrolling vertically ...in the manner of programme credits’, the sequence Interview is described as ‘an animated conversation...(where the) camera gradually zooms in for a close-up’. In both these cases there is no fast movement. As stated earlier the no-motion figures-of-merit actually include some low temporal frequencies. It is therefore likely that the frequency content of these two sequences is more suited to use of the no-motion figures-of-merit. Consequently in Figure 6.9 and Figure 6.10 the no-motion figures-of-merit have been used for Captions and Interview.

Looking at Figure 6.9 and Figure 6.10 the results appear quite good - there is definitely a correlation between the subjective and objective results. For the wanted signals the correlation is always positive, as expected, and similarly for the unwanted signals the correlation is mostly negative. The correlations are not, however, high, and are close to 1 (or -1) in only a few cases. Considering all the
assumptions that have been made, the incompleteness of the objective assessment method, and the
difficulties in comparing subjective and objective results these results are, in the author’s opinion,
quite favourable. Notable is the fact that the correlations are strongest for stationary images (i.e.
young couple, and orbit). This may suggest that the motion figures-of-merit are less accurate
predictors of subjective results, perhaps because of the spatial/temporal equivalence problem
discussed earlier in section 6.4.1.

6.8 Further development of this objective assessment method
As has already been stated this method is still very much undeveloped. There are many questions
still to be answered, such as:

1. A full measurement of the supra-threshold response of the HVS.
2. Determination of the relative importance of different spatio-temporal frequencies to a
   viewer’s judgement of image quality.
3. Methods for reconciling objective and subjective results.
4. The development of a useful framework for presenting objective assessment results.
5. A comparison of the performance of this objective method against other methods (see
   section 2.9.2).

6.9 Summary
This Chapter has outlined a new approach to objective assessment, based upon the use of models of
the HVS response applied to calculated spatio-temporal characteristics for a coding system. The
result is a set of quality measures, or figures-of-merit, which predict the expected performance of the
coding system in any one of a number of performance categories (e.g. luminance
resolution/bandwidth, cross-colour, etc).

This approach is fundamentally different to existing approaches which calculate performance
measures based upon HVS weighting of the ‘errors’ between input and output images. This new
approach relies on the author’s new method for calculating the complete spatio-temporal
characteristics for a television coding system. Its feasibility has been demonstrated by its application
to six coding techniques in this chapter, and to K-PAL and WC-PAL in Chapter 7.

As yet the method is not fully developed, and has many areas needing further work. Its potential has,
however, been demonstrated by comparison with an existing set of subjective results [Clarke 1987].
A correlation was shown between these two sets of results. Although the correlation was admittedly
not strong it is expected that this might improve greatly once the method is more fully developed.

Even if not used as an objective assessment tool the figure-of-merit results provide a very useful
tool in comparing many different aspects of performance in television coding techniques. This fact
has been illustrated by examples in both this Chapter and Chapter 7. In these examples the figures-
of-merit do an excellent job of showing bandwidth improvements, cross-effect reduction and the
introduction of artefacts such as aliasing and U/V cross-talk.
CHAPTER 7
K-PAL - A NEW ENHANCED PAL CODING TECHNIQUE

This Chapter describes in detail a new enhanced PAL coding technique developed and studied by the author. The technique, named K-PAL, was introduced in Chapter 4 and uses phase segregation to achieve separation of luminance and chrominance signals. K-PAL is similar to WC-PAL in this respect, and together these two techniques represent different approaches to compatible phase segregation. The theory and performance of WC-PAL has been studied some in detail in the literature [Oliphant 1980, 1981, Drewery 1984, 1986], and the possibilities for various implementations of WC-PAL have also been studied [Sandbank 1990, Croll 1992]. WC-PAL has consequently been chosen as a basis for comparison in studying K-PAL.

This Chapter begins with a detailed mathematical analysis of K-PAL and WC-PAL, proving the basis of phase segregation as a compatible clean coding technique, and providing insight into how the techniques work. This leads to the derivation of equations that describe the complete response for luminance and chrominance in both K-PAL and WC-PAL, under enhanced and compatible reception. The resulting complete analysis of WC-PAL has not been presented in this form elsewhere.

The mathematical analysis is based upon ideal filters. Later sections deal with design considerations for non-ideal vertical-temporal filters and horizontal filters, and with implementation considerations such as reducing coder component count, and preventing over-voltage outputs.

The final section provides a detailed comparison of two simulations each of K-PAL and WC-PAL. All the relevant performance parameters are compared (e.g. luminance resolution, luminance aliasing, cross-chrominance, ...). All available tools are used in this comparison - mathematical modelling and test images, and the newly developed spatio-temporal spectra and objective assessment results.

7.1 Placing luminance and chrominance in quadrature

As discussed in previous Chapters cross-effects in PAL occur because high frequency luminance $Y(f)$, and chrominance $C(f)$ are not combined in a reversible manner. One well known method for maintaining separation of two signals is quadrature amplitude modulation (QAM). A QAM signal consists of two double side-band (DSB) signals modulated on the same carrier, but at 90° to each other. It is conventionally represented as

$$q(t) = s_i(t)\cos(2\pi f_c t) + s_q(t)\sin(2\pi f_c t)$$  \hspace{1cm} (7.1)

where $f_c$ is the carrier frequency, and $s_i(t), s_q(t)$ are the two baseband signals to be modulated on the in phase and quadrature carriers. The two signals may be described as being in phase segregation and are separable because of this.

With PAL phase segregation might appear, at first, to be inapplicable as (i) there are three, not two signals to be combined, and (ii) the luminance is a baseband signal. However, by special modification of the $Y$ and $C$ signals a form of quadrature phase segregation is possible. The modification process requires four steps.
1. Pre-filter $U$ and $V$ to ensure that at any one frequency in $C(f)$ only one of the $U$ or $V$ signals is present.
2. Modify the luminance so as to appear a DSB signal modulated at $f_{ce}$.
3. Pre-filter the luminance to allow its recovery and 'de'-modification in the decoder.
4. Combine the $C(f)$ and modified $Y(f)$ signals at $90^\circ$ to each other in a manner that maintains PAL compatibility.

The way in which these processes are carried out is greatly restricted by the requirements of the fourth step, the maintenance of PAL compatibility. Despite all the modifications the output signal must appear completely normal to a conventional PAL decoder. This is perhaps the most complex requirement, and can be met in two different ways represented by the enhanced techniques of Weston Clean PAL (WC-PAL) [Weston 1977, Oliphant 1980] and the author’s K-PAL. As discussed briefly in Chapter 4 both techniques are nearly identical as far as the first three steps are concerned. It is in the fourth step that they differ markedly.

The requirements for the first three steps will now be presented in detail, along with proof of the separability of the resulting signals. K-PAL and WC-PAL will then be analysed separately to provide insight into their different approaches to achieving PAL compatibility. The mathematical description of each system will be used to show that, under ideal conditions (and ideal filters), both techniques allow complete separation of luminance and chrominance for the enhanced path, as well as correct decoding in a conventional PAL decoder.

### 7.1.1 Pre-filtering of $U$ and $V$

The first step is to modify the chrominance signals by use of a pre-filter. The chrominance normally consists of two signals, $U$ and $V$, placed in quadrature and overlapping each other in vertical-temporal frequency space. In order to apply phase segregation we need to ensure that only one of the chrominance signals exists with the luminance at any one point in frequency space. This means making sure that $U$ and $V$ occupy different areas of spatio-temporal frequency space. To achieve this we may make use of the fact that the $U$ and $V$ sub-carriers are in different positions in spatio-temporal frequency space (see figures 3.19 and 3.20 in Chapter 3), with their positions being

\begin{align}
U_{sc} & = \pm f_{sc} \\
V_{sc} & = \pm f_{sc} \pm \frac{1}{2} f_{las}
\end{align}

To prevent overlap of $U$ and $V$ we may thus use a pre-filter that satisfies the requirement

\begin{align}
R_c(f \pm f_{sc})R_c(f \pm f_{sc} \pm \frac{1}{2} f_{las}) & = 0
\end{align}

This is equivalent to

\begin{align}
R_c(f)R_c(f \pm \frac{1}{2} f_{las}) & = 0
\end{align}

In order to preserve the most important chrominance frequencies $R_c(f)$ should be substantially low-pass in nature. Figures 3.37 and 3.38 in Chapter 3 show some of the comb filter types that would meet these requirements.

### 7.1.2 Modification of luminance

The next step after pre-filtering of the chrominance is to modify the luminance into a DSB signal.

As was discussed in Chapter 4 all enhanced PAL techniques only operate on the high frequency luminance close to the chrominance sub-carrier. For these purposes $Y$ is normally separated into $Y_L$ and $Y_H$, by use of a bandpass filter (or highpass) as below
7.1 PLACING LUMINANCE AND CHROMINANCE IN QUADRATURE

Figure 7.1 Complementary separation of luminance into high and low frequency signals.

Any modification of luminance will only be applied to \( Y_H \), with \( Y_L \) remaining unchanged throughout. Considering this fact the distinction between \( Y_L \) and \( Y_H \) will be dropped and all references to \( Y \) will be assumed to refer to \( Y_H \) only.

The modified version of \( Y \) (i.e. \( Y_H \)) needs to be of the form

\[
y_m(t) = s_i(t)\cos(2\pi f_s t) + 2y(t)\sin(2\pi f_s t)
\]

and yet still contain a baseband version of \( Y \) for compatible decoding.

This can be achieved by setting \( s_i(t) \) to

\[
s_i(t) = 2y(t)\cos(2\pi f_s t) + 2\bar{y}(t)\sin(2\pi f_s t)
\]

where \( ^\wedge \) indicates the Hilbert transform. Taking the Fourier transform of (7.5) and (7.6) we obtain

\[
Y_m(f) = Y(f) + \frac{1}{2}\left[ Y(f + 2f_s)\left( 1 \pm \text{sgn}(f + 2f_s) \right) \right] + \frac{1}{2}\left[ Y(f - 2f_s)\left( 1 - \text{sgn}(f - 2f_s) \right) \right]
\]

where \( Y(f + 2f_s)\left( 1 \pm \text{sgn}(f + 2f_s) \right) \) is a compacted form of the expression

\[
Y(f + 2f_s)\left( 1 + \text{sgn}(f + 2f_s) \right) + Y(f - 2f_s)\left( 1 - \text{sgn}(f - 2f_s) \right).
\]

This luminance modification process can be implemented using the block diagram shown below

Figure 7.2 Block diagram for modification of high frequency luminance into a DSB signal at \( f_s \).

Thus a \( 2f_s \) modulated version of \( Y \) is added to the luminance, creating a signal that appears to be modulated in phase with the carrier. The process of modulation by two times the sub-carrier, then bandpass filtering is a well known technique referred to as PAL modification, the entire device being called a PAL modifier [Bruch 1966]. If we instead wish the output luminance signal to be in quadrature with the carrier the PAL modified signal is instead subtracted from the baseband signal.

For further analysis a PAL modified signal will be indicated by the symbol \( ^\sim \) above the function, such that

\[
\bar{Y}(f) = \frac{1}{2}B(f)\left[ Y(f + 2f_s)\left( 1 \pm \text{sgn}(f + 2f_s) \right) \right]
\]

Thus the equations for modified in phase and quadrature luminance become.

\[
Y_{m}(f) = Y(f) + \bar{Y}(f)
\]

\[
Y_{mq}(f) = Y(f) - \bar{Y}(f)
\]
7.1.3 Pre-filtering the luminance

The modification of the luminance has added an aliased version of $Y$ to the signal. To enable this to be separated from the original luminance in an enhanced decoder a pre-filter must be applied to $Y$. If this pre-filter is represented by $R_y(f)$ the requirements for the filter may be represented as

$$R_y(f)R_y(f \pm 2f_w) = 0 \quad (7.11)$$

As was discussed in Chapter 6 most filters suitable for $R_c$ will also satisfy the above requirement. Thus the comb filters of Figure 6.22 are also suitable for $R_y$. The other possibility for $R_y$ is a vertical-horizontal filter of the form shown in figure 6.25 (Chapter 6).

Using the pre-filter the modified PAL signal becomes

$$Y_y(f) = R_y(f)Y(f) + \tilde{R}_y(f)\tilde{Y}(f) \quad (7.12)$$

Setting $Y_y(f) = R_y(f)Y(f)$ (7.12) may be simplified to

$$Y_y(f) = Y_p(f) + \tilde{Y}_p(f) \quad (7.13)$$

7.1.4 A phase-segregated signal

The generalised form of a phase-segregated PAL signal may now be presented as

$$P(t) = y_p(t) + \tilde{y}_p(t) + c_p(t)\sin(2\pi f_st) \quad (7.14)$$

F.T. $\Rightarrow P(f) = Y_p(f) + \tilde{Y}_p(f) + \frac{4}{T}C_p(f \pm f_w)$

where $C_p(f)$ is the pre-filtered chrominance signal containing either $R_c(f)U(f)$ or $R'_c(f)V(f)$. The superscript $s$ indicates modulation by the PAL switch.

7.2 Separating $Y$ and $C$ in a decoder

The luminance and chrominance signals have now been combined into a phase segregated signal. We must now prove that they can be recovered, without cross-effects, in order to validate phase segregation as an enhanced coding method.

7.2.1 Decoding Chrominance

The block diagram below shows the conventional PAL method for decoding one of the quadrature chrominance signals.

```
\[ \text{Figure 7.3 Block diagram of conventional PAL chrominance decoding.} \]
```

where $B$ is a chrominance bandpass filter, $L_c$ a chrominance lowpass filter, and $C_d$ is the decoded chrominance (either $U$ or $V$), defined by

$$C_d(f) = L_c(f) \left[ \frac{\pi}{2} B(f \pm f_w)P(f \pm f_w) \right] \quad (7.15)$$

The use of the bandpass filter is actually equivalent to the lowpass, with $B(f \pm f_w) = L_c(f)$ for $|f| < f_w$. Consequently the bandpass filter can be left out of the present analysis.

Expanding (7.15) using the definition for $P(f)$ in (7.14) results in
7.2 SEPARATING Y AND C IN A DECODER

\[ C_d(f) = \frac{1}{2} L_c(f) \left( \mp Y_p(f \pm f_{sc}) \mp \bar{Y}_p(f \pm f_{sc}) + \frac{1}{2} C_p(f \pm 2f_{sc}) - \frac{1}{2} C_p(f) \right) \]  \hspace{1cm} (7.16)

If all chrominance frequencies are below \( f_{sc} \) and the chrominance lowpass filter \( L_c \) has a cutoff below \( f_{sc} \) then (7.16) reduces to

\[ C_d(f) = C_p(f) + \frac{1}{2} L_c(f) \left( \mp Y_p(f \pm f_{sc}) \mp \bar{Y}_p(f \pm f_{sc}) \right) \]  \hspace{1cm} (7.17)

Using the definition for \( \bar{Y} \) made in (7.8) it can be shown that

\[ \mp \bar{Y}_p(f \pm f_{sc}) L_c(f) = \pm Y_p(f \pm f_{sc}) \]  \hspace{1cm} (7.18)

with the result that the two luminance terms in (7.17) cancel, leaving only the chrominance term.

\[ C_d(f) = C_p(f) \]  \hspace{1cm} (7.19)

The pre-filtered chrominance signal has now been successfully recovered, with no cross-chrominance.

### 7.2.2 Decoding Luminance

The luminance signal must now be recovered. An obvious method for achieving this is by subtracting already decoded chrominance from the original PAL signal. To do this \( C_d(f) \) must first be decoded, and then remodulated again.

This operation, shown in Figure 7.4 can be represented as

\[ Y_d(f) = P(f) - \left( \mp \frac{1}{2} P(f \pm f_{sc}) L_c(f) \right) \ast \frac{1}{2} \delta(f \pm f_{sc}) \]  \hspace{1cm} (7.20)

\[ \Rightarrow Y_d(f) = P(f) + \frac{1}{4} L_c(f \pm f_{sc}) \left( P(f \pm 2f_{sc}) + P(f) \right) \]  \hspace{1cm} (7.21)

where \( \ast \) indicates convolution.

If \( P(f + 2f_{sc}) = 0 \) \( f > 0 \)

and \( P(f - 2f_{sc}) = 0 \) \( f < 0 \)

then we may replace the shifted lowpass chrominance filter with a chrominance bandpass, ie. \( B(f) = L_c(f \pm f_{sc}) \).

Considering only the bandpass region of the PAL signal (7.20) becomes

\[ Y_d(f) = \frac{1}{2} \left( P(f) + B(f) P(f \pm 2f_{sc}) \right) \]  \hspace{1cm} (7.23)

which, using (7.8) again, becomes

\[ Y_d(f) = \frac{1}{2} \left( P(f) + \bar{P}(f) \right) \]  \hspace{1cm} (7.24)

The result suggests that the block diagram in Figure 7.4 may be replaced with that shown below in Figure 7.5.
This circuit is ideally sufficient to completely separate the luminance from the PAL signal without cross-luminance.

### 7.3 Forming a PAL compatible signal

The previous sections have illustrated the principles behind phase segregation of chrominance and luminance. Now a way must be found to apply these principles whilst still maintaining the $U$ and $V$ signals in their correct phase relationships, thus allowing decoding in a conventional PAL decoder.

The steps described so far apply equally well to K-PAL and WC-PAL. In their approaches to achieving PAL compatibility, however, the techniques differ markedly. The differences in these approaches will now be explained, and the resulting PAL signals analysed.

#### 7.3.1 WC-PAL

The block diagram below shows how a WC-PAL signal may be formed

![Generalised block diagram for a WC-PAL decoder.](image)

Note that this diagram is a different, but equivalent, arrangement of that commonly presented in the literature for WC-PAL [Oliphant 1980]. The arrangement chosen here is intended to help with deriving a description of the signal, as well as allowing comparison with K-PAL.

If we set

$$Y_s(f) = R_s(f)Y_u(f)$$

and

$$C_{sc}(f) = \frac{1}{2s}
\left(
L_c(f \pm f_{sc})R_c(f \pm f_{sc})U(f \pm f_{sc}) + 
\right.
$$

then the full WC-PAL signal is described by

$$P(f) = Y_s(f) + \frac{1}{2}(1+e^{-2\pi f_{line}})\left(Y_u(f) + \hat{Y}_p(f)\right) + \frac{1}{2}(1-e^{-2\pi f_{line}})C_{sc}(f)$$

(7.27)

where $T_{line} = 1/f_{line}$. Making the definitions
FORMING A PAL COMPATIBLE SIGNAL

\[ S_y(f) = \frac{1}{2} \left(1 + e^{-j2\pi n_{lin}}\right) \]  
\[ S_c(f) = \frac{1}{2} \left(1 - e^{-j2\pi n_{lin}}\right) \]  

Equation (7.27) becomes

\[ P(f) = Y_s(f) + S_y(f)(Y_p(f) + \bar{Y}_p(f)) + S_c(f)C_c(f) \]  

\( S_y(f) \) and \( S_c(f) \) are vertical comb filters and together form the WC-PAL assembler [Drewery 1984]. To view their response against vertical spatial frequency we may use equation 3.13 and set \( f = nf_r \). Taking the magnitude and phase of (7.28) and (7.29) gives the functions defined below, (7.31) and (7.32).

\[ S_y(f) = \sqrt{\frac{1 + \cos \left(\frac{2\pi n}{N}\right)}{2}} \arg \left\{ -\frac{n}{N} \right\} \]  
\[ S_c(f) = \sqrt{\frac{1 + \cos \left(\frac{2\pi n}{N}\right)}{2}} \arg \left\{ \frac{\pi}{2} - \frac{n}{N} \right\} \]  

where \( N = 625 \) is the number of lines per frame. These equations describe the responses of the WC-PAL assembly filters as a function of vertical spatial frequency \( n \) in cycles/picture height.

7.3.1.1 WC-PAL amplitude and phase

In order to better understand how WC-PAL achieves compatibility we can plot the amplitude and phase characteristics of the assembler filters. This directly relates to the phase of the luminance and chrominance signals in WC-PAL.

The functions in (7.31) and (7.32) are plotted in Figure 7.7.

From these plots it can be seen that in WC-PAL the chrominance \( U \) and \( V \) sub-carriers are multiplied by \( \frac{1}{\sqrt{2}} \), and phase shifted by \( \pm 45^\circ \). When the \( \sqrt{2} \) chrominance scaling factor in Figure 7.6 is taken into account the result is a correct amplitude of unity at the \( U \) and \( V \) sub-carrier positions. The phase shifts mean the \( U \) and \( V \) sub-carriers also appear \( 90^\circ \) apart, whilst still maintaining the modified luminance signal at \( 90^\circ \) to all chrominance. As stated in Chapter 4 two important requirements are met

1. The \( Y \) and \( C \) signals are \( 90^\circ \) apart at all frequencies.
2. At the PAL U and V sub-carrier positions the chrominance signals have the correct amplitude and are 90° out of phase.

Obviously, for higher chrominance frequencies that are not close to the sub-carrier this PAL compatibility will begin to break down; there will be a phase error in the modulated chrominance and an amplitude error. The effects of this will be discussed in more detail in section 6.4.1.

### 7.3.2 K-PAL

In distinction to WC-PAL K-PAL uses two different filters \( F_u(f) \) and \( F_v(f) \) to select areas around the U and V sub-carriers. Phase segregation is then applied separately to these two areas. This is done in such a way that the chrominance sub-carriers in the \( F_u \) and \( F_v \) areas are 90° apart, meeting the requirements for PAL compatibility.

Figure 7.8 shows the block diagram for the K-PAL encoder. Ideally the two areas of frequency space selected by \( F_u \) and \( F_v \) should not overlap, i.e.

\[
F_u(f)F_v(f) = 0 \tag{7.33}
\]

As only \( U \) passes through \( F_u \) and \( V \) through \( F_v \) these filters should be chosen to select the most important chrominance frequencies. These requirements end up being identical to those for the chrominance pre-filters meaning we may define \( F_u \) and \( F_v \) as

\[
F_u(f) = R_c(f \pm f_w) \tag{7.34}
\]

\[
F_v(f) = R_c(f \pm (f_w + f_{line})) \tag{7.35}
\]

![Figure 7.8 General form for K-PAL encoder.](image)

Chrominance pre-filtering using \( R_c \) is therefore be replaced by the use of \( F_u \) and \( F_v \). In order to reduce the component count complementary filters can be used for \( F_u \) and \( F_v \) with

\[
F_v(f) = 1 - F_u(f) \tag{7.36}
\]

Using this definition for \( F_v \) and equation (7.25) defining \( Y_p \) the signal produced by Figure 7.8 may be described by

\[
P(f) = F_u(f)[Y_p(f) - \bar{Y}_p(f) + U_w(f)]
\]

\[
+ (1 - F_u(f))[Y_p(f) + \bar{Y}_p(f) + V_w(f)] \tag{7.37}
\]
where

\[
U_{sc}(f) = \frac{1}{2} L_C(f \pm f_{sc}) U(f \pm f_{sc}) \tag{7.38}
\]
\[
V_{sc}(f) = \frac{1}{2} L_C(f \pm (f_{sc} + \frac{1}{2} f_{sim})) V(f \pm (f_{sc} + \frac{1}{2} f_{sim})) \tag{7.39}
\]

Equation (7.37) can be simplified to

\[
P(f) = Y_p(f) + (1 - 2F_U(f)) Y_p(f) + F_U(f) U_{sc}(f) + (1 - F_U(f)) V_{sc}(f) \tag{7.40}
\]

This equation defines the general form for the K-PAL compatible PAL signal.

### 7.3.2.1 K-PAL amplitude and phase

As with WC-PAL a better understanding of the mechanism behind K-PAL can be gained by looking at the phase diagrams. For comparison with WC-PAL we will assume \(F_U\) is a vertical filter (although in practice it may not be). The luminance and chrominance amplitude and phase diagrams for K-PAL thus appear below.

![Magnitude and phase plots against vertical frequency for K-PAL chrominance and luminance signals.](image)

As can be seen K-PAL also meets the basic requirements for PAL compatibility, with the \(U\) and \(V\) signals being in the correct places with their sub-carriers 90° out of phase. K-PAL has the added advantage of maintaining the \(U\) and \(V\) signals completely in phase with their sub-carriers, unlike WC-PAL. However, as will be shown later, this is only the case for an ideal \(F_U\) i.e. a spatio-temporal low-pass filter with a zero width transition region.

### 7.4 Compatible Reception

It has been shown that the K-PAL and WC-PAL signals meet the basic requirements for PAL compatibility. They are, however, still different from a standard PAL signal. These differences inevitably lead to changes in the received signal on a conventional PAL receiver. In designing the enhanced encoders for maximum compatibility we need to know what these changes are. Below is a basic PAL decoder.
In this PAL decoder $B$ is the $f_{sc}$ centred bandpass filter, $N$ is a sub-carrier notch, $L_C$ is the chrominance low-pass filter, and $D_U$ and $D_V$ produce the $U$ and $V$ delay line outputs. $Y_d$, $U_d$, and $V_d$ are the decoded signal components.

As in section 6.2.1 $B$ is subsumed by $L_C$ and for the purposes of a simple mathematical analysis may be ignored. The phase shift $s$ is included to allow for correct synchronisation to the enhanced signal. We will now look at the performance of this decoder on WC-PAL and K-PAL encoded signals.

### 7.4.1 Receiving the WC-PAL Signal

The compatible reception of a WC-PAL signal will now be analysed. This looks at the effect of passing the WC-PAL signal of (7.30) into the conventional PAL decoder shown in Figure 7.10.

#### 7.4.1.1 Decoded Chrominance

For a WC-PAL encoded signal we must add a compensatory phase shift of $s = 45^\circ$ to the PAL decoder. In an actual receiving situation this phase adjustment will occur automatically, with the decoder synchronising to the received sub-carrier.

Considering only the $U$ path

$$U_d(f) = 2L_C(f)[P(f)D_U(f)] \ast \frac{1}{\sqrt{2}}[(1 + j)\delta(f + f_{sc}) + (1 - j)\delta(f - f_{sc})]$$

(7.41)

where $\ast$ indicates convolution.

The PAL delay-line has the response

$$D_U(f) = \frac{1}{2}(1 + e^{-j2\pi T_D})$$

(7.42)

where $T_D = 284/f_{sc}$.

It can be shown that $D_U(f \pm f_{sc}) = D_U(f)$, meaning that (7.41) becomes

$$U_d(f) = \frac{1}{\sqrt{2}}L_C(f)D_U(f)[(1 + j)P(f + f_{sc}) + (1 - j)P(f - f_{sc})]$$

(7.43)

Denoting $\pm f_{sc}$ frequency shifts by the use of $\ast$ and $\dot{\ast}$ superscripts (7.43) can be represented as

$$U_d = \frac{1}{\sqrt{2}}L_C D_U[(1 + j)P^* + (1 - j)P^*]$$

(7.44)

where the variable $f$ has been omitted from all functions temporarily. This form of representation with $f$ omitted will be used considerably from now on, as it allows a simpler presentation of the functions being studied.
Using the identities

\[ S^+ = S^+ = \frac{1}{2}(1 + jE) \]  
\[ S^- = S^- = \frac{1}{2}(1 - jE) \]  

where \( E = e^{j2\pi f_{11}} \), and substituting for \( P \) using the WC-PAL PAL signal defined earlier in (7.30), equation (7.44) becomes

\[
U_d = \frac{1}{\sqrt{2}} L_c D_0 \left[ \left( 1 + j \right) \left( S_r^+ Y_p^+ + S_c^+ C_w^+ \right) + \left( 1 - j \right) \left( S_r^+ Y_p^- + S_c^+ C_w^+ \right) \right]
\]  

(7.47)

It can be proven that

\[ L_c Y_p^+ = L_c Y_p^+ \]  
\[ L_c C_w^+ = L_c C_w^+ = \frac{1}{\sqrt{2}} L_c \left( L_c R_c U + L_c R_c V \right) \]  

(7.48)

(7.49)

where the \( s \) superscript indicates modulation by the PAL switch. Using these equation (7.47) becomes

\[
U_d = \frac{1}{\sqrt{2}} L_c D_0 \left[ \left( 1 + j \right) S_r^+ \left( Y_p^+ + Y_p^- \right) + \left( 1 - j \right) S_r^+ \left( Y_p^+ + Y_p^- \right) \right] + \frac{1}{\sqrt{2}} L_c \left( L_c R_c U + L_c R_c V \right) \]  

(7.50)

Using the identities for \( S_r^\pm \) and \( S_c^\pm \) in (7.45), (7.46) the following further identities may be calculated.

\[ S_r^+ + S_r^- = S_c^+ + S_c^- = 1 \]  
\[ jS_r^+ - jS_r^- = jS_c^+ - jS_c^- = E \]  

(7.51)

(7.52)

Consequently (7.50) simplifies to

\[
U_d = L_c D_0 S_r R_c U + L_c D_0 S_r R_c V + \sqrt{2} L_c D_0 S_c Y_p^\pm \]

(7.53)

where \( Y_p^\pm = R_p^\pm Y^\pm \). With \( L_c \) being a horizontal filter the PAL switched version \( L_c^* \) is identical to \( L_c \).

For conventional PAL the decoded chrominance signal for the \( U \) component is

\[
U_d = L_c^2 D_0 U + L_c D_0 Y^\pm \]

(7.54)

Comparing (7.53) and (7.54) there are several differences in the decoded chrominance signals.

1. Reduction of cross-chrominance by \( \frac{1}{\sqrt{2}} S_c R_r^\pm \)
2. Attenuation of chrominance by \( S_r R_c \)
3. Introduction of \( V \to U \) cross-talk

These changes are both good (1) and bad (2, 3), and will be discussed in more detail later.

### 7.4.1.2 Decoded Luminance

For conventional PAL luminance decoding simply involves the application of the notch filter \( N \). Thus from equation (7.30) the received luminance signal for a conventional PAL decoder operating with a WC-PAL encoded signal is

\[
Y_d = NS_l R_l Y + NS_p R_p \bar{Y} + NS_c C_w
\]  

(7.55)

This may be compared with the normal decoded PAL luminance signal.
\[ Y_d = NY + N(U_v + V_v) \]  \hspace{1cm} (7.56)

where \(U_v\) and \(V_v\) are the conventionally encoded chrominance signals (no pre-filtering). Comparing (7.55) and (7.56) the main differences are.

1. Attenuation of luminance by \(S_y R_y\).
2. Introduction of an aliased luminance signal \(\bar{Y}\).
3. Change in the nature of the cross-luminance.

The change in the cross-luminance will not be discussed in detail here, but it is likely that the WC-PAL signal produces a reduction in cross-luminance. Mostly this will be due to the effects of the chrominance pre-filters \(R_C\).

7.4.2 Receiving the K-PAL Signal

A similar analysis to that just performed for WC-PAL will now be done for compatible reception of a K-PAL signal.

7.4.2.1 Decoded Chrominance

For K-PAL the phase shift \(s\) may be set to 0°. Considering once again the \(U\) path only and using the shortened notation that has been developed the decoded chrominance is

\[ U_d = 2L_C D_u^+ P^+ \hspace{1cm} (7.57) \]

Using the fact that \(D_u^+ = D_u\) (ie. \(D_u(f \pm f_{sc}) = D_u(f)\)) and substituting for \(P\) using (7.40) this equation expands to

\[ U_d = L_C D_u \left( \begin{array}{c} Y_p^+ + (1 - 2F_u^-) \bar{Y}_p^+ + F_u^+ U_v^+ + (1 - F_u^-) V_v^+ \\ Y_p^- + (1 - 2F_u^-) \bar{Y}_p^- + F_u^- U_v^- + (1 - F_u^-) V_v^- \end{array} \right) \hspace{1cm} (7.58) \]

Using the identity in (7.48), and assuming that \(F_u^+ = F_u^-\) (see section 6.8.2, equation (7.85)) this simplifies to

\[ U_d = L_C D_u \left( 2Y_p^+ (1 - F_u^-) + F_u^+ U_v^+ + (1 - F_u^-) V_v^+ \right) \hspace{1cm} (7.59) \]

The chrominance lowpass \(L_C\) will remove all frequencies above \(f_{sc}/2\). After filtering \(U_v\) will reinforce and \(V_v\) will cancel, leaving the decoded chrominance signal as

\[ U_d = L_C^2 D_u F_u^- U_v + 2L_C D_u (1 - F_u^-) R_y^+ Y_p^\hspace{1cm} (7.60) \]

Comparing this result with that for conventional PAL in (7.54) we have two differences:

1. Reduction in cross-chrominance by a factor of \((1 - F_u^-) R_y^+\).
2. Attenuation of chrominance by \(F_u^-\).

Unlike WC-PAL, however, there is no inter-chrominance (UV cross-talk).

7.4.2.2 Decoded Luminance

Applying the PAL luminance notch filter \(N\) to (7.40) we obtain

\[ Y_d = NR_y Y + N(1 - 2F_u^-) \bar{R}_y Y + NF_u U_v + N(1 - F_u^-) V_v \hspace{1cm} (7.61) \]
7.5 ENHANCED RECEPTION

The resulting decoder luminance has similar properties to that for WC-PAL, with an attenuation of \( Y \) by \( R_y \), the introduction of luminance aliasing, and a change in the cross-luminance.

7.5 Enhanced Reception

Enhanced decoding of the K-PAL and WC-PAL signals will now be analysed. This should produce improved signals compared to compatible decoding. The most importance difference should be the elimination of cross-effects.

7.5.1 WC-PAL Decoding

A WC-PAL decoder may be represented as in Figure 7.11

![Figure 7.11 General form of a Weston Clean PAL decoder.](image)

The combination of \( S_y \) and \( S_c \) in WC-PAL coding is known as PAL assembly and disassembly and forms a transparent network when cascaded in the encoder and decoder. Theoretically the \( Y \) and \( C \) signals will emerge unmodified and perfectly separated from this network. A compensating delay has been added to the above decoder to correctly match its phase to that of the encoder. In practice the phase of the modulators would be adjusted instead (this occurs automatically through sub-carrier locking).

7.5.1.1 WC-PAL - Decoding the Chrominance

Decoding the chrominance involves passing the PAL signal through the PAL disassembler, demodulating, then separating out \( U \) and \( V \). From Figure 7.11 the decoded chrominance signal in the \( U \) path may be described by

\[
U_d = \sqrt{2} L_c S_c^\frac{z}{\sqrt{E^z}} \frac{1}{E^z} P^k
\]  

(7.62)

where the delay of 'tline is represented by \( 1/E (E = e^{-2\sigma_{\text{line}}} \).

Substituting for \( P \) using the WC-PAL signal defined in (7.30) this equation becomes

\[
U_d = \sqrt{2} L_c \left( \frac{1}{E^z} S_c^2 S_c^k \left( Y_p^+ + \tilde{Y}_p^+ \right) + \frac{1}{E^z} S_c^2 S_c^k \right)
\]  

(7.63)

Using the identity for \( \tilde{Y}_p^2 \) in (7.48) this expression can be expanded and rearranged to become

\[
U_d = \sqrt{2} L_c \left( Y_p^+ \left( \frac{1}{E^z} S_c^2 S_c^k + \frac{1}{E^z} S_c^2 S_c^k \right) + Y_p^- \left( \frac{1}{E^z} S_c^2 S_c^k + \frac{1}{E^z} S_c^2 S_c^k \right) \right) + \frac{1}{E^z} S_c^2 S_c^k
\]  

(7.64)

Using identities for \( S_y^2 \) in (7.45) and (7.46), and the further identity
\[ E^\pm = \pm jE \]  
(7.65)

equation (7.64) becomes

\[ U_d = \sqrt{2} L_c \left( \frac{1}{jE} S_c^* S_c - \frac{1}{jE} S_c S_c^* \right) + \frac{\pm 1}{jE} S_c^2 C_w \]  
(7.66)

Consequently in (7.66) the luminance terms cancel, leaving

\[ U_d = \sqrt{2} L_c \left( \frac{1}{jE} S_c^2 C_w^* - \frac{1}{jE} S_c S_c^2 C_w \right) \]  
(7.67)

Evaluating \( S_c^2 \) using the identities in (7.45) and (7.46) provides the further identity

\[ S_c^2 = \frac{1}{4} \left( 1 \mp 2 jE - E^2 \right) \]  
(7.68)

If this identity is used in (7.67) most of the terms cancel, leaving only \( C_w^* \). Using the identity for \( C_w^* \) in (7.49) and then substituting for \( C_w \) using (7.26) we obtain a final result for the decoded chrominance.

\[ U_d = L_c^2 R_c^2 U + L_c^2 R_c R_c^* Y \]  
(7.69)

where the superscript \( s \) indicates modulation by the PAL switch. Ideally \( R_c R_c^* \) will equal zero (equation (7.4)) and the \( V \) term will disappear. In practice, however, \( R_c R_c^* \) will not quite equal zero meaning some \( V \) chrominance will remain (U/V cross-talk).

### 7.5.1.2 WC-PAL - Decoding the Luminance

Once again, if we consider only high frequency luminance the bandpass filter \( B \) can be ignored. With \( E \) defined as \( e^{-j2\pi f_{sc}} \) the decoded luminance signal \( Y_d \) may be found

\[ Y_d = R_r \left( \frac{1}{E} S_r P + \frac{1}{E} \bar{S}_r \bar{P} \right) \]  
(7.70)

Substituting for \( P \) using (7.30) this becomes

\[ Y_d = R_r \left( \frac{1}{E} S_r^2 (Y_r + \bar{Y}_r) + \frac{1}{E} S_r S_c C_w + \frac{1}{E} \bar{S}_r^2 (\bar{Y}_r + \bar{Y}_r) + \frac{1}{E} \bar{S}_r \bar{S}_r C_w \right) \]  
(7.71)

It can be shown that the following identities hold

\[ \bar{S}_r = S_c \]  
(7.72)

\[ \bar{S}_c = S_r \]  
(7.73)

\[ \bar{Y}_r = Y_r \]  
(7.74)

\[ \bar{C}_w = C_w \]  
(7.75)

\[ \bar{E} = -E \]  
(7.76)

Using these (7.71) becomes

\[ Y_d = \frac{1}{E} R_r \left( (S_r^2 - S_c^2)(Y_r + \bar{Y}_r) + (S_r S_c - S_c S_r)C_w \right) \]  
(7.77)

(7.78)
As $S_e^2 - S_c^2 = e^{-\beta g_{\text{ave}}}$ therefore $\frac{1}{E}(S_e^2 - S_c^2) = 1$ meaning the luminance terms will reinforce, whilst the chrominance terms cancel. Replacing $Y_p$ the final result is

$$Y_d = R_y^2 Y + R_y (1 - R_y) \vec{P}$$ (7.79)

For an ideal luminance pre-filter $R_y(1 - R_y)$ is equal to zero and thus the aliased luminance should disappear, leaving only the original pre-filtered luminance.

### 7.5.2 K-PAL Decoding

The K-PAL decoder is almost identically the reverse of the encoder. First the band segregating filters separate out the $U$ and $V$ areas of the PAL signal, then chrominance demodulation and luminance demodulation are applied separately to each channel.

![Figure 7.12 Block diagram of K-PAL enhanced decoder.](image)

The final chrominance and luminance characteristics of a decoded K-PAL signal can now be analysed. As before, only the bandpass region ($Y = Y^p$) will be considered. This means that the bandpass filter $B$ will be ignored for the following analysis.

#### 7.5.2.1 K-PAL - Decoding the Luminance

The decoder luminance path may be represented mathematically as

$$Y_d = \frac{1}{2} R_y \left( P - \vec{F}_u \vec{P} + (1 - F_u) P + (1 - \vec{F}_u) \vec{P} \right)$$ (7.80)

which simplifies to

$$Y_d = \frac{1}{2} R_y \left( P + (1 - 2 \vec{F}_u) \vec{P} \right)$$ (7.81)

Using equation (7.40) defining the K-PAL PAL signal, (7.81) expands to

$$Y_d = \frac{1}{2} R_y \left( Y_p + \left( 1 - 2 F_u \right) \vec{P} + F_u U_x + \left( 1 - F_u \right) V_x + \left( 1 - 2 F_u \right) \vec{P} \right)$$ (7.82)

Using the identity for $\vec{Y}_p$ in (7.74) and the further identities
\[ \tilde{U}_{sc} = U_{sc} \quad (7.83) \]
\[ \tilde{V}_{sc} = -V_{sc} \quad (7.84) \]

Equation (7.82) becomes
\[ Y_d = \frac{1}{2} R_f \left[ 2\left(1 - 2\tilde{F}_u + 2\tilde{F}_u^2\right)Y_u + 2\left(1 - F_u - \tilde{F}_u\right)\tilde{Y}_u + \right. \]
\[ \left. (F_u + \tilde{F}_u - 2\tilde{F}_u^2)U_{sc} + \left(3\tilde{F}_u - F_u - 2\tilde{F}_u^2\right)V_{sc} \right] \quad (7.85) \]

For perfect separation of the luminance we need the terms in front of \( U_{sc} \) and \( V_{sc} \) to become zero. This requires that
\[ \tilde{F}_u = F_u \quad (7.86) \]
\[ F_u^2 = F_u^2 \quad (7.87) \]

If \( F_u \) is designed to meet these requirements (as closely as possible) then the result in (7.65) will reduce to
\[ Y_d = R_f^2 Y + R_f (1 - R_f) (1 - 2F_u) \bar{Y} \quad (7.88) \]

where \( Y_p \) has been replaced with \( R_f Y \).

Once again, for an ideal luminance pre-filter \( R_f(1 - R_f) \) should equal zero, thus removing the luminance alias and leaving only the original pre-filtered luminance.

In practice the assumption made in (7.87) will not hold completely meaning \( F_u^2 \) will not quite equal \( F_u \) and some residual cross-effects will remain (see section 7.8.6).

The requirement that \( \tilde{F}_u = F_u \) (\( F_u(f \mp f_w) = F_u(f) \)) is an extra restriction on \( F_u \) that does not exist for \( R_c \) in WC-PAL. How \( F_u \) is modified so as to meet this requirement is discussed later in section 6.6.1.

### 7.5.2.2 K-PAL - Decoding the Chrominance

Considering only the \( U \) path in the K-PAL decoder of Figure 7.12
\[ U_d = 2L_C \left( F_u U_u P^* + \frac{1}{2} \delta(f \pm f_w) \right) \quad (7.89) \]
\[ \Leftrightarrow U_d = L_C F_u^* P^* \quad (7.90) \]

Substituting for \( P \) using the K-PAL signal of (7.40) this equation expands to
\[ U_d = \frac{1}{2} L_C F_u^* \left( Y_u^* + (1 - 2F_u^*) \tilde{Y}_u^* + F_u^2 U_{sc}^* + (1 - F_u^*) V_{sc}^* \right) \quad (7.91) \]

Using the identity for \( \tilde{Y}_u^* \) in (7.48) and the fact that (7.86) implies that \( F_u^* = F_u^* \) the expression above simplifies to
\[ U_d = L_C \left( 2\left(1 - F_u^2\right)Y_u^* + F_u^2 U_{sc}^* + F_u^* (1 - F_u^*) V_{sc}^* \right) \quad (7.92) \]

Using the further identities
\[ U_{sc}^* = 2U \quad (7.92) \]
\[ V_{sc}^* + V_{sc}^* = 0 \quad (7.93) \]
and substituting for $Y_p$ equation (7.91) becomes

$$U_d = P_c F_u^{-2} U + 2 \left( F_u^{-} - F_u^{+} \right) R_u^+ Y^+$$

(7.94)

Once again if $F_u^{+} = F_u^{-}$ exactly then the cross-luminance portion of this result will cancel, leaving only the $U$ channel chrominance.

### 7.6 Designing the filters

The previous section has shown the form of the K-PAL and WC-PAL signals for idealised filters. In practice neither comb filters nor horizontal filters are ideal (in terms of their having non-zero width transition bands). As a result, the way the filters are designed can have considerable effects on how close the actual performance comes to the ideal. This section will look at the practical nature of these filters and how they can be designed to maximise overall performance.

Firstly we will look at the design of the K-PAL band-segregating filter.

#### 7.6.1 The K-PAL band segregating filter

In section 7.5.2 it was shown that for K-PAL to decode correctly the band segregating filter must meet the requirement $F_u^{+} = F_u^{-}$ (equation (7.86)). This is in addition to the previously stated requirement of $F_u f (f)F_u (f \pm 1 f_{sc}) = 0$ in (7.33). The standard set of comb filters described in Chapter 3 do not meet this new requirement. Indeed, for most of them

$$F_u^{+} = 1 - F_u$$

(7.95)

The requirement in (7.86) is needed to ensure that $F_u$ selects the areas near the sub-carrier. To do this the sub-carrier must be placed in phase across the filter. Across a line delay, however, the sub-carrier is $90^\circ$ out of phase, which means we must either use delays slightly different from one line, or introduce a compensating phase shift. This gives us three main ways of constructing the band-segregating filter.

1. Add a diagonal slope to the filter by using delays that are slightly more (or less) than one line ie. $T_D = 283.5 f_{sc}$ or $T_D = 284 f_{sc}$ (remembering that $T_{line} = 283.7516 f_{sc}$).
2. Add a $90^\circ$ phase shift to alternate lines in the filter by use of a phase delay. This can be achieved using a phase-shift network or a small time delay, eg. $T_p = .25 f_{sc}$.
3. Add a $90^\circ$ phase shift by use of a double PAL modifier.

To illustrate these modifications and the effects they have on the characteristics the diagrams below show four different versions of a three tap vertical comb filter. The frequency characteristic is plotted as a contour plot against vertical and horizontal frequency.

---

**Figure 7.13 Original three tap vertical comb filter.**
Looking at the filter responses against horizontal and vertical frequency method 3 would appear to provide the best characteristics, selecting a large and symmetrical portion of the U chrominance signal. Method 3 also allows the use of full line delays. Its disadvantage is its complexity, requiring the cascading of two modulators and two bandpass filters.

Method 1 (Figure 7.14) is the simplest, but produces a skewed response, and has the disadvantage of not using full line delays.

When comparing all three, method 2 (Figure 7.15) represents a compromise between the others. This filter uses full line delays, and a small 0.25/fsc time delay to introduce the 90° phase shift. This phase shift is only accurate close to fsc, and consequently the response of the filter is only good within a bounded region around fsc. In practice, however, the region of interest for chrominance/luminance
7.6 DESIGNING THE FILTERS

separation is only within 1.5 MHz either side of $f_{sc}$. Because of this the filter response shown in Figure 7.15 is quite acceptable, with a relatively sharp response over the main region of interest.

When it comes to digital implementation using non-sub-carrier locked sampling (eg. Rec. 501 @ 13.5 MHz) methods 1 and 2 may both be unsuitable, due to their requirement for delays of a non-integer multiple of the sampling frequency. Instead method 3 may be more suitable. More likely, however, is the use of method 2 with the delay replaced by a digital phase shift filter. For the results presented in this thesis the problem of non-sub-carrier locked sampling did not apply. All simulations were run using $8f_{sc}$ sampling which produces an almost exactly orthogonal sampling lattice. Consequently method 2 has been used in all simulations of K-PAL presented in this thesis.

7.6.2 Designing the pre-filters $R_C/F_U$ and $R_Y$

The requirements for the filters $R_C/F_U$ and $R_Y$ have been derived in earlier sections, and some of the possibilities for these filters were described in Chapter 4. In general there are two variables to be considered:

- **filter order** - Number of delays and taps.
- **filter type** - 1D vertical, temporal, or vertical-temporal diagonal.
  
  2D vertical-temporal, or vertical-horizontal.

Complicating the choice of these filters is the fact that, for K-PAL, some parameters are affected by the response of both filters. Table 7.1 lists which filters are responsible for affecting the response of each parameter in both K-PAL and WC-PAL.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>K-PAL</th>
<th>WC-PAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compatible</td>
<td>Enhanced</td>
</tr>
<tr>
<td>Luminance</td>
<td>---</td>
<td>$R_Y$ &amp; $F_U$</td>
</tr>
<tr>
<td>Luminance aliasing</td>
<td>---</td>
<td>$R_Y$ &amp; $F_U$</td>
</tr>
<tr>
<td>Chrominance</td>
<td>$F_U$</td>
<td>$F_U$</td>
</tr>
<tr>
<td>Cross-chrominance</td>
<td>$R_Y$ &amp; $F_U$</td>
<td>$R_Y$ &amp; $F_U$</td>
</tr>
<tr>
<td>Cross-luminance</td>
<td>---</td>
<td>$R_Y$ &amp; $F_U$</td>
</tr>
<tr>
<td>Inter-chrominance</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

*Table 7.1 Filters affecting each performance parameter.*

In terms of performance improvement the importance of the different parameters could be rated as shown below.

<table>
<thead>
<tr>
<th></th>
<th>Compatible</th>
<th>Enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>most important</strong></td>
<td>Cross-chrominance</td>
<td>Luminance aliasing</td>
</tr>
<tr>
<td></td>
<td>Chrominance resolution</td>
<td>Chrominance resolution</td>
</tr>
<tr>
<td></td>
<td>$U/V$ crosstalk</td>
<td>$U/V$ crosstalk</td>
</tr>
<tr>
<td><strong>least important</strong></td>
<td>Cross-chrominance (K-PAL only)</td>
<td>$U/V$ crosstalk (WC-PAL only)</td>
</tr>
</tbody>
</table>

*Table 7.2 A possible rating for the importance of performance parameters in considering each path.*

In order to minimise compatible cross-chrominance $R_C/F_U$ must be kept as a vertical filter so as to best match the vertical characteristic of the PAL delay line ($D_y$). Increasing the order of $R_C/F_U$ will increase chrominance resolution and reduce $U/V$ crosstalk for WC-PAL. It may be remembered that in Chapter 4 the best characteristic was concluded as being a one-dimensional vertical filter, with two-dimensional filters offering no extra benefits.
For $R_f$ higher order filters will increase luminance resolution and reduce luminance aliasing. As temporal delays are expensive very high order filters must be avoided. There is, however, some freedom to choose different filter orientations (eg. vertical vs. temporal filtering). Section 7.8.1 will describe the exact choice of the filters used in simulation.

### 7.6.3 Effects of horizontal filters

Just as the comb filter responses are non-ideal the horizontal filters are also not ideal, containing ripple in the passband, and having a finite roll off in the stopband. The fact that there are several filters in series (eg a bandpass filter in encoder and then decoder) serves to accentuate these effects. The result is that:

- Enhanced coding techniques will not work properly (ie cross-effects cannot be fully suppressed) within filter transition regions.
- For phase segregation any asymmetry in the sidebands about $f_{se}$ will reduce the effectiveness of the technique.

These effects are the reason why WC-PAL, although theoretically offering zero cross-effects, in practice has residual cross-effects present. This is shown in the objective assessment results of Figure 7.49 and Figure 7.54 where a residual level of cross-effects is shown in the enhanced WC-PAL path, despite the theory predicting zero cross-effects (equations (7.69) and (7.79)).

The problems caused by non-ideal filters can be minimised by

1. Designing the pass-band to be as flat as possible
2. Choosing a sharp cutoff
3. Avoiding the use of unnecessary filtering
4. Maintaining the best symmetry possible around $f_{se}$

The first two points are a matter of filter complexity and the trade-offs required for implementation. The third point provides the impetus for a further redesign of the coders. The fourth point will be dealt with later.

Avoiding unnecessary filtering can be achieved by two steps. First by rearranging the block diagrams into a minimal realisation, and second, by changing the way the horizontal filtering is performed.

### 7.6.4 Rearranging the block diagrams

There are a number of reasons for simplifying the implementation diagrams of the coders. As well as improving the signal for hardware implementation purposes we need to minimise the number of components required to construct the coder. In particular we must minimise the use of expensive storage delays (line / field stores), and the overall signal delay introduced by the coder. These aims can be achieved by a considerable rearrangement of the block diagrams presented previously for K-PAL and WC-PAL. We will also add here the equalising delays needed to compensate for the delays inherent in comb filters.

#### 7.6.4.1 K-PAL

A number of changes can be made to the K-PAL coder that reduce component count.

- $F_V$ is replaced with $1-F_V$.
- The same luminance modifier can be used for both the $U$ and $V$ paths.

Making these changes the encoder can be rearranged to obtain the new block diagram below
7.6 DESIGNING THE FILTERS

Figure 7.17 Simplified and rearranged block diagram for K-PAL encoder.

This diagram also includes the significant equalising delays that are needed. As mentioned, equalising delays are needed for all horizontal filters as well. However, these delays are very small compared to the comb filter delays, and will for the present be ignored. The dashed line around $R_y$ and $T_{eq}[R_y]$ is intended to indicate that this equalising delay can be obtained from within $R_y$, thus requiring no extra components. This is not the case for $T_{eq}[F_u]$ and the other two $T_{eq}[R_y]$ delays.

For the decoder similar simplifications can be made. The simplified K-PAL decoder, including compensating delays for the comb filters, is shown below.

Figure 7.18 Simplified and rearranged block diagram for K-PAL decoder

Once again some of the compensating delays can be obtained from within $F_u$ and $R_y$. The resulting decoder delay is identical to the encoder, with the storage requirements being slightly less.

7.6.4.2 WC-PAL

The most commonly used simplified form for WC-PAL is based upon the first proposed and most basic implementation of the technique [Oliphant 1980]. For this case the luminance and chrominance pre-filters are only 2 tap filters (as opposed to the 3 tap filters used in WC-PAL (1)). The use of these filters allows use of a very simplified form for the coder, requiring only 3 delays for the entire encoder, and 2 for the decoder.

For the more general implementations this level of reduction may not be possible. Some simplification can be achieved by combining $S_y$ and $S_C$ into one filter. Also, in the decoder the use of two chrominance post-filters can be avoided by subtracting a post-filtered $U$ signal from the chrominance to obtain the $V$ signal (this is based on the assumption that $R_C$ is a complementary filter).

7.6.5 Minimising horizontal filtering

We will look first at the K-PAL signal. Equation (7.40) for the encoded signal may be modified to include low frequency luminance and the effects of all the horizontal filters. Considering only the $V$ path (i.e. where $F_U = 0$) the response of the encoder can be described as
\[ P = \frac{L_p (1 - B)Y}{Y_L} + \frac{L_p R_p BY}{Y_H} + \frac{L_p B \tilde{R} \tilde{B} \tilde{Y}}{\tilde{Y}_H} + L_p V_w \]  

(7.96)

where \( L_p \) is the channel lowpass, and \( B \) the bandpass. As indicated the luminance portion of the PAL signal may be divided into three parts, the low frequency luminance \( Y_L \), the high luminance \( Y_H \), and the modified high frequency luminance \( \tilde{Y}_H \). The diagram below shows the horizontal filters that affect each of these parts.

\[ \text{Figure 7.19 The general shape of the horizontal filters affecting the low, high, and modified high frequency luminance.} \]

These diagrams verify the assertion made earlier that a chrominance bandpass and highpass are effectively equivalent. The bandpass filtering applied at the channel limit is subsumed by the channel lowpass. The bandpass may thus be replaced with a highpass, or a complementary filter \( 1 - L_i \) where \( L_i \) is a lowpass with a cutoff frequency similar to that of the lower end of the original bandpass. For the \( \tilde{Y} \) path further modifications can be made.

As the bandpass is symmetrical about \( f_w \), and PAL modification causes the flipping (or mirroring) of the sidebands around \( f_w \) then \( \tilde{B} = B \), meaning that the first bandpass is superfluous. The second bandpass may now be, once again, replaced with a highpass, or, if placed before the PAL modifier, a PAL modified lowpass, to be designated \( \tilde{L}_2 \). The PAL signal described by (7.96) now becomes

\[ P = \frac{L_p L Y}{Y_L} + \frac{L_p R_p (1 - L_2) Y}{Y_H} + \frac{L_p \tilde{L}_2 \tilde{R} \tilde{Y}}{\tilde{Y}_H} + L_p V_w \]  

(7.97)

and Figure 7.19 becomes
7.6 DESIGNING THE FILTERS

Unnecessary filtering is now avoided. Adding these changes to the simplified block diagram in Figure 7.17 the encoder now becomes

![Diagram of encoder with modified filtering]

Figure 7.21 Simplified K-PAL encoder with minimum horizontal filtering.

Similar modifications can be made to the decoder, replacing the bandpass filter with a highpass $H_1$, or a complementary highpass using $L_1$.

![Diagram of decoder with modified filtering]

Figure 7.22 Simplified K-PAL decoder with minimum horizontal filtering.
The new decoder response is thus

\[ Y_d = L_P P + \frac{1}{2} R_R (1 - L_P) P + \frac{1}{2} R_R \tilde{H}_1 \left(1 - 2 \tilde{F}_U \right) \tilde{P} \]  

(7.98)

and for chrominance (in \( U \) path)

\[ U_d = L_C \left( \delta(f \pm f_{\omega}) \ast \left( H_1 L_P U_{\omega} \right) \right) \]  

(7.99)

7.6.6 The effects of upper sideband loss

Looking at the chrominance signal in the decoder just before demodulation (as described by (7.99)) the response of the horizontal filters affecting the chrominance will appear similar to that in Figure 7.23 below (where \( L_C \) has been transposed to \( f_{\omega} \)).

![Figure 7.23](image)

Figure 7.23 The filters affecting the chrominance signal around \( f_{\omega} \).

As can be seen there is only a limited amount of space between the sub-carrier and the channel limit set by \( L_P \). This causes problems for demodulation of the chrominance signal. The chrominance is a QAM signal, and its correct demodulation requires that both the sidebands be present. The cutoff of \( L_P \), however, restricts the width of the upper sideband to 0.57 MHz for systems B, G PAL (5 MHz), and 1.07 MHz for system I PAL (5.5 MHz). This limit will be labelled as \( f_L \) where \( f_L = f_{\text{channel}} - f_{\text{sc}} \). The cutoff of \( L_C \) is traditionally set to around 1.2-1.6 MHz in the encoder. As this is well above \( f_L \) some of the upper chrominance sideband will inevitably be lost. This effect will obviously be worse for system B, G PAL.

For conventional PAL decoding any chrominance frequencies above \( f_L \) will be attenuated and incorrectly decoded, causing \( U/V \) cross-talk. The attenuation means that the effective received chrominance bandwidth is reduced to \( f_L \), and the \( U/V \) cross-talk causes hue errors on sharp horizontal colour transitions.

For K-PAL and WC-PAL the upper sideband loss causes not only problems with chrominance decoding, but also ruins cross-effect suppression. \( Y \) and \( C \) can only be properly segregated across a region of 1.14 MHz for system B, G and 2.14 MHz for system I. The failure of the enhanced techniques outside this region is illustrated in Figure 7.24 using a luminance spatial zone and WC-PAL coding.

---

5 In fact, the author wonders why, especially for B, G PAL such a high chrominance bandwidth is still used in the encoder. The extra chrominance bandwidth results in virtually no increase in received chrominance resolution, but instead causes \( U/V \) cross-talk and increased cross-luminance (from the wider lower sideband). It would seem to make more sense to use a restricted chrominance bandwidth, to perhaps 0.6-0.7 MHz in the case of B, G PAL.
Although WC-PAL theoretically produces clean coding with zero cross-effects in practice this is only achieved across a narrow region (<f₁) near the sub-carrier, meaning that in Figure 7.24 cross-chrominance is still present outside this region. As this region is almost twice as narrow in B, G PAL than in I PAL it might seem that system B, G PAL would be particularly unsuited to phase segregation techniques. In practice, however, it is in this narrowest region that the most significant cross-effects occur. Removing cross-effects in an area inside 0.5 MHz either side of fₛ is sufficient to greatly improve picture quality. The images shown in Figure 7.24(b) and (c) bear out this assertion. Comparing them with the same picture under conventional PAL reception Figure 7.24(a) both show greatly reduced cross-chrominance, with the increased improvement under system I being small.

The failure of phase segregation outside the region described above has further implications for WC-PAL. The WC-PAL assembler / disassembler ideally form a transparent network. The response of the combined assembler / disassembler is a sine wave. When the chrominance is demodulated, or the luminance passed through a PAL modifier the positive and negative frequencies of the signal around fₛ are superimposed on each other. The two sine wave responses are now at 180⁰ to each other, and add to give a uniform response. If, however, one of the sidebands is attenuated a uniform response is no longer obtained. The result is a sine wave shaped attenuation of the chrominance or high frequency luminance. The greatest attenuation occurs for intermediate vertical frequencies.

With all the above points in mind the PAL system being used should be taken into account when implementing a phase segregation technique. All horizontal filters (band-pass, low-pass & high-pass) should be designed to restrict the phase segregation only to the region that can properly support it (ie 3.86 - 5 MHz under system B, G and 3.36 - 5.5 MHz under system I). As well the chrominance lowpass Lc should be designed with a cutoff that matches closely to f₁.

### 7.6.7 Designing L₁, L₂, and H₁

The design of these filters is determined by three requirements

1. Maximising cancellation of cross-chrominance for compatible reception.
2. Maximising suppression of cross-effects in an enhanced decoder.
3. Minimising distortion (attenuation) of luminance in passband region.

These will be looked at one at a time.

#### 7.6.7.1 Minimising cross-chrominance - compatible decoding

The elimination of cross-chrominance in a conventional decoder is dependant upon the cancellation of the luminance sidebands around fₛ. To ensure that this occurs the modified luminance from the K-PAL encoder should be as symmetrical as possible.
Equation (7.97) describes the PAL signal in the stopband of $F_u$ (ie the $V$ passband). Choosing also the passband of $R_y$ (setting $R_y = 1$) and considering only the luminance (7.97) becomes

$$P = L_p Y + L_r \tilde{I}_2 \tilde{Y}$$

(7.100)

The decoded $V$ chrominance signal will thus be (similar to (7.99))

$$V_d = L_c \left( \mp \frac{1}{2} \delta(f \pm f_w) * P \right)$$

(7.101)

Using the fact that for any arbitrary function $F$

$$L_c \left( \mp \frac{1}{2} \delta(f \pm f_w) * \tilde{F} \right) \equiv \pm \frac{1}{2} F(f \pm f_w)$$

(7.102)

equation (7.101) becomes

$$V_d = \frac{1}{2} \left( Y^+ \left( L_r L_z^+ - L_r^+ \right) - Y^- \left( L_r^+ L_z^- - L_r^- \right) \right)$$

(7.103)

From this elimination of cross-chrominance requires that

$$L_r L_z^+ - L_r^+ = L_r^+ L_z^- - L_r^- = 0$$

(7.104)

Thus for a given channel filter $L_2$ should be chosen such that

$$L_2(f) = \frac{L_r(f)}{L_r(f - 2f_w)} \quad f > 0$$

(7.105)

For negative frequencies the filter response will simply be a mirror image. An example of a filter defined in this way is shown below.

![Figure 7.25 'Ideal' response of filter $L_2$ as defined by equation (7.105).](image)

The filter is identical to $L_p$ above approximately $2f_{sc} - f_{channel}$ (4 MHz). Unfortunately the lower part of the characteristic of $L_2$ is impractical. We must set a ceiling on the gain of $L_2$, and thus some cross-chrominance will return for lower frequency luminance below $2f_{sc} - f_{channel}$. This is to be expected, as it is outside the bandpass region. However, a certain level of gain, perhaps a factor of 2, may be left, and will help to compensate for attenuation of the upper sideband by the channel lowpass.

7.6.7.2 Minimising cross-chrominance - enhanced decoding

For compatible reception we can only control the symmetry of the upper sideband of $Y$. This means that cross-chrominance returns for luminance frequencies below $f_{sc}$. In the enhanced decoder the...
filter \( H_i \) can be used to shape the lower sideband of \( Y \), ensuring proper cancellation will occur with the upper sideband of \( \bar{Y} \).

For the K-PAL decoder equation (7.101) is instead

\[
V_d = L_C \left( \pm \frac{1}{2} \delta(f \pm f_c) * H_i P \right)
\]  
(7.106)

Considering the passband of \( R_y \) we can substitute for \( P \) using (7.100). Ignoring the chrominance signal (7.106) becomes

\[
V_d = L_C \left( \pm \frac{1}{2} \delta(f \pm f_c) \ast (H_i L_p Y + H_i L_p \bar{L}_p \bar{Y}) \right)
\]  
(7.107)

Using the definition made in (7.102) this equation provides a requirement for elimination of cross-chrominance. For cancellation of a particular luminance frequency to occur requires that

\[
H_i L_p = \bar{H}_i \bar{L}_p \bar{L}_2
\]  
(7.108)

Considering first luminance frequencies below \( f_c \) then

\[
L_p = 1, \quad L_2 = 1 \quad |f| < f_c
\]  
(7.109)

If we assume that \( H_i \) is a highpass with a cutoff below \( f_c \) then \( \bar{H}_i = 1 \) for \( |f| < f_c \) and

\[
H_i = \bar{L}_p \bar{L}_2 \quad |f| < f_c
\]  
(7.110)

For luminance frequencies above \( f_c \) \( H_i = 1, \bar{L}_p = 1 \) and (7.108) becomes

\[
\Rightarrow H_i = \frac{\bar{L}_p \bar{L}_2}{L_2} \quad |f| > f_c
\]  
(7.111)

Equations (7.110) and (7.111) appear to provide conflicting requirements for \( H_i \), with the first defining the requirements for cancellation of low frequency luminance (below \( f_c \)), and the second defining the requirements for luminance above \( f_c \). However, if we take into account the fact that \( \bar{L}_p = 0 \) for \( |f| < f_c \), then both equations define \( H_i = 0 \) for \( |f| < f_c \). The statistical energy of the luminance signal decreases with increasing frequency [Ref 1], thus cancellation of lower frequencies is more important than for higher frequencies. Consequently we will choose (7.114) to define \( H_i \) near \( f_c \).

7.6.7.3 Minimising attenuation of luminance in passband

The overall response of the luminance path may be found by substituting (7.97) into (7.98). Considering only the passband of the luminance pre-filter where \( R_y = 1 \), the decoded luminance may be described by

\[
Y_d = \left( L_2 L_p + L_p L_1 (1 - L_1) + \frac{1}{2} L_3 L_p (1 - L_1) + \frac{1}{2} L_4 L_p (1 - L_1)^2 + \frac{1}{2} \bar{H}_i \bar{L}_p \bar{L}_2 \right) Y + \left( L_1 L_p \bar{L}_2 + \frac{1}{2} \bar{H}_i \bar{L}_p \bar{L}_1 \right) \bar{Y}
\]  
(7.112)

The lowpass filter has a much lower cutoff than the channel filter \( L_p \), therefore we may assume that

\[
L_1 L_p \equiv L_1
\]  
(7.113)

Using this identity and simplifying (7.112) gives
Ignoring the modified luminance signal (which is responsible for residual aliasing), we can achieve a flat response up to the channel limit if
\[ \frac{1}{3} \left( L_1 + L_p + \bar{H}_1 \bar{L}_p \bar{L}_2 \right) = 1 \quad f < f_{\text{channel}} \quad (7.115) \]

If considering only frequencies where \( f < f_{\text{channel}} \) then we may make a further assumption that
\[ L_2 = \bar{H}_1 = 1 \quad f < f_{\text{channel}} \quad (7.116) \]

Equation (7.115) thus becomes
\[ L_1 + L_p + \bar{L}_p = \frac{1}{3} \quad f < f_{\text{channel}} \quad (7.117) \]

Rearranging this we can obtain an expression defining the lowpass filter \( L_1 \).
\[ L_1 = 2 \left( L_p + \bar{L}_p \right) \quad f < f_{\text{channel}} \quad (7.118) \]

Given a channel response \( L_p \) if \( L_1 \) is designed to meet the above requirement then a flat response for high frequency luminance can be assured.

### 7.6.7.4 Minimising residual aliasing

Taking equation (7.114) and ignoring the wanted luminance signal \( Y \) the luminance alias, in the \( R_y \) passband is
\[ Y_d = \left( L_1 \bar{L}_2 + \frac{1}{2} \bar{H}_1 \bar{L}_1 \right) \bar{Y} \quad (7.119) \]

The response of these filters may be assumed positive only, and thus minimal aliasing can only be achieved if
\[ L_1 \bar{L}_2 = 0 \quad \text{and} \quad H_1 L_1 = 0 \quad (7.120) \]

In practice, as these filters have a finite cutoff this cannot be achieved perfectly. Instead we may choose an acceptable level of residual alias. We will arbitrarily set this at \( \frac{1}{4} \) amplitude, or -12 dB. This requires that
\[ \max(L_1 \bar{L}_2) = 0.25, \quad \max\left( \frac{1}{2} H_1 L_1 \right) = 0.25 \quad (7.121) \]

This can be achieved by choosing \( \bar{L}_2 \) and \( L_1 \) as complementary filters where
\[ \bar{L}_2 = (1 - L_1) \quad (7.122) \]

\( L_1 \) can be chosen as a 'semi-complementary' filter where the -3 dB point of \( L_1 \) coincides with the -3 dB point of \( H_1 \). These results thus provide further design constraints on \( L_1, L_2, H_1 \).

### 7.6.8 Asymmetrical chrominance filtering

As indicated earlier the implementations studied in this Chapter (section 6.8) are only two of many possible variations. One variation worth looking at is the use of differing bandwidths for the \( U \) and \( V \) signals. If the visual response weighted figures-of-merit for cross-chrominance (see Figure 7.49 later) are separated into individual figures-of-merit for \( U \) and \( V \) as in Figure 7.26 we see that the visibility of \( Y \leftrightarrow V \) cross-chrominance is predicted to be considerably greater than that for \( Y \leftrightarrow U \) (note that the higher the figure-of-merit the worse the cross-colour). This is due to the fact that, as
explained in section 2.2 of Chapter 2, the eye's resolving ability is lower for the colour frequencies contained in $U$ (predominantly Yellow-Blue) than for those in $V$ (predominantly Red-Green).

![Figure 7.26 Comparison of separate figures-of-merit cross-chrominance into the U and V paths.](image)

This suggests that image quality may be improved by placing a differing emphasis on the $U$ and $V$ signals. Extra $V$ bandwidth and minimising $Y \rightarrow V$ cross-chrominance is more important than bandwidth and cross-chrominance for the $U$ signal. This reasoning is used in NTSC, where the $I$ and $Q$ colour difference signals have quite different horizontal bandwidths of 1.5 MHz and 0.5 MHz respectively.

A redesign of the band-segregating filter $F_U$ can allow an exchange of increased $V$ bandwidth for decreased $U$ bandwidth. This also increases the area over which luminance is out of phase with the $V$ signal, meaning less $V$ cross-chrominance (at the expense of increased $U$ cross-chrominance). Figure 7.27 below shows one such filter. The

![Figure 7.27 Asymmetrical chrominance pre-filter $F_U/F_V$ in (b), compared with chrominance pre-filter used in K-PAL2 - (a).](image)

This filter increases the vertical bandwidth of the $V$ signal from just under 70 cph in K-PAL2 to around 85 cph, with a corresponding decrease of the $U$ bandwidth to 50 cph. This filter has been used in a simulation labelled K-PAL3, and the results using this new filter are compared below in Figure 7.28 with K-PAL1 and K-PAL2.
Figure 7.28 Comparison of results obtained using asymmetrical pre-filter (K-PAL3) with those from Figure 7.26.

Figure 7.28 shows that the asymmetrical pre-filter has indeed succeeded in reducing overall cross-chrominance, by decreasing \( V \) cross-chrominance at the expense of a small increase in \( U \) cross-chrominance. For moving picture content cross-chrominance is now almost perfectly balanced between \( U \) and \( V \). The advantage of this technique is also, as noted, a likely increase in perceived vertical chrominance resolution, due to the increase in bandwidth of the more important \( V \) signal.

### 7.6.9 Enhanced horizontal resolution

As well as reducing cross-effects most enhanced PAL techniques considerably improve horizontal resolution. Instead of a limit of around 3.5 MHz (224 c/pw) the received picture can be restored to a horizontal resolution matching the channel limit. This equates to 320 c/pw for system B, G PAL, and 352 c/pw for system I. Phase segregation techniques, however, are actually capable of providing further horizontal resolution extension, to beyond the channel limit. This is achieved by using the aliased spectrum produced by the luminance modifier to store the extra resolution, in a technique known as spectrum folding. How this works is illustrated below.

Figure 7.29 shows the pre-filtered luminance for implementation 1, which used a vertical filter for \( R_y \). From the spectra there is the appearance of the high frequency luminance being folded (or rotated) about \( f_{sc} \), with the aliased spectra slotting into the ‘holes’ in the baseband luminance. In Figure 7.29 (b) although all frequencies beyond the channel limit are lost from the baseband luminance, a copy of what was lost is retained in the aliased spectra. This can then be unfolded in the decoder to regain the lost resolution.
The disadvantage of the method is that, in order to accommodate the extra luminance in the aliased spectra more luminance must be removed from the baseband signal. Increased horizontal resolution thus becomes a trade-off against reduced diagonal luminance resolution.

Another disadvantage is that decoding the extended resolution luminance signal is more complicated, due to the fact that the high frequency luminance exists in only one sideband around \( f_c \). Because of this, after decoding the extra luminance will have only half the amplitude of the normal luminance. Thus a filter must be used to compensate the higher frequencies.

### 7.7 Implementation details

#### 7.7.1 Over-range luminance

The luminance modification process adds an aliased copy of the luminance to the original baseband luminance. This causes a doubling of high frequency luminance energy, and could theoretically cause over-range luminance values to result. This might mean that the PAL signal amplitude exceeds its allowed limits. In a worst case scenario the encoded luminance signal would be twice its normal amplitude. Such a situation could potentially cause problems for transmitters, recording equipment, and A/D conversion.

The situation is not, however, as bad as it at first appears. The worst case example can only occur when high frequency luminance is of high amplitude. For such a signal any low frequency luminance present will be of low amplitude. As chrominance is of restricted bandwidth it only accompanies low frequency luminance signals, and will thus, in this situation, also be of low amplitude, leaving extra room for an increase in luminance amplitude. To determine whether there will be enough room we must study the PAL signal limits and the enhanced signals.

The equations for the WC-PAL and K-PAL signals were derived earlier and are given in (7.30) and (7.40). To consider the worst case scenario we can take an input luminance signal

\[
y(t) = A \cos 2\pi f_s t
\]

where \( f_s \) is chosen as a high horizontal frequency. Taking the Fourier transform a normal PAL signal for this input would simply be

\[
P(f) = \frac{A}{2} \delta(f \pm f_s)
\]

#### 7.7.1.1 WC-PAL luminance signal

Considering first WC-PAL, and substituting for \( Y \) in (7.30) the PAL signal for this enhanced technique is

\[
P_{wc}(f) = S_r(f)R_r(f) \frac{1}{2} A \delta(f \pm f_s) + S_r(f)R_r(f \pm 2f_c) \frac{1}{2} A \delta(f \pm 2f_c \pm f_s)
\]

Evaluating this expression and using the identities for \( S_c(f \pm f_w) \) in (7.45) and (7.46) we obtain

\[
P_{wc}(f) = \frac{A}{2} \left( S_r(\pm f_s)R_r(\pm f_s) + S_c(\pm f_s)R_r(\pm f_s) \right)
\]

Using the further definitions for the WC-PAL assembler filters in (7.28) and (7.29) equation (7.126) becomes

\[
P_{wc}(f) = \frac{A}{2} R_r(\pm f_s)
\]
This result means that the problem with over-range values will never occur for WC-PAL, as the response of $R_Y$ is never greater than one, and thus the amplitude of the expression in (7.127) will never exceed that for a normal PAL signal in (7.124).

### 7.7.1.2 K-PAL luminance signal

For K-PAL the situation is different. Substituting (7.123) into (7.40) we obtain, after simplification, the expression

$$P_{kipal}(f) = AR_Y(\pm f_s)\left[1 - F_U(\pm f_s)\right]$$

Even if $F_U$ and $R_Y$ never exceed one this expression could still be up to twice the normal PAL amplitude.

Comparing this function with normal PAL we can define a relative signal level function $G_Y$

$$G_Y = \frac{P_{kipal}(f)}{P(f)}$$

Plotting this for K-PAL implementations 1 and 2 we obtain the graphs below

![Graphs](image)

*Figure 7.30 Relative luminance signal levels for K-PAL as compared with normal PAL, (a) Implementation 1, (b) Implementation 2.*

At their peaks the signal levels in these implementations are 1.45 and 2 times greater than normal PAL.

The normal limits of the PAL signal are shown in Table 7.3 [CCIR Report 624-3].

<table>
<thead>
<tr>
<th><strong>Table 7.3 Normal limits for PAL signal.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>synchronising level</td>
</tr>
<tr>
<td>luminance black</td>
</tr>
<tr>
<td>luminance white</td>
</tr>
<tr>
<td>maximum signal level</td>
</tr>
</tbody>
</table>

The ultimate upper and lower limits of the synchronising and maximum signal levels represent a range of $\pm 186\%$ for the luminance. It might thus appear that although implementation 2 is outside this range, implementation 1 may have enough room (only requiring $\pm 145\%$). In practice the situation is more complicated, as chrominance and high frequency luminance may be present at the same time.
In order to determine what colours and signal levels will cause an over range output we can consider a signal where each gamma corrected $R'$, $G'$, $B'$ component contains both a low (in this case DC) and high frequency. As with the luminance the $R'$, $G'$, $B'$ components are constrained to within the range 0 - 700 mV. Input signals with maximum worst case amplitude may thus be defined as

$$R' = A_r + K_r \cos 2\pi f_r t$$  \hspace{1cm} (7.130)

$$G' = A_g + K_g \cos 2\pi f_g t$$  \hspace{1cm} (7.131)

$$B' = A_b + K_b \cos 2\pi f_b t$$  \hspace{1cm} (7.132)

where $0 \text{ mV} \leq A_r \leq 700 \text{ mV}$ \hspace{1cm} (7.133)

and

$$K_r = \begin{cases} A_r & A_r \leq 350 \text{ mV} \\ 700 \text{ mV} - A_r & A_r > 350 \text{ mV} \end{cases}$$  \hspace{1cm} (7.134)

The frequencies $f_r$, $f_g$ and $f_b$ must be above the cutoff of the chrominance lowpass $L_C$. Using these definitions the maximum and minimum excursions for the combined PAL signal will be.

$$P_{\text{max}} = Y_{\text{max}} + |C|_{\text{max}}$$  \hspace{1cm} (7.135)

$$P_{\text{min}} = Y_{\text{min}} - |C|_{\text{max}}$$  \hspace{1cm} (7.136)

where

$$Y_{\text{max}} = 0.299 (A_r + G_r K_r) + 0.587 (A_g + G_r K_g) + 0.114 (A_b + G_r K_b)$$  \hspace{1cm} (7.137)

$$Y_{\text{min}} = 0.299 (A_r - G_r K_r) + 0.587 (A_g - G_r K_g) + 0.114 (A_b - G_r K_b)$$  \hspace{1cm} (7.138)

$$|C|_{\text{max}} = \sqrt{U^2 + V^2}$$  \hspace{1cm} (7.139)

and

$$U = 0.493 (A_b - Y_L)$$  \hspace{1cm} (7.140)

$$V = 0.877 (A_r - Y_L)$$  \hspace{1cm} (7.141)

$$Y_L = 0.299 A_r + 0.587 A_g + 0.114 A_b$$  \hspace{1cm} (7.142)

These equations use the definitions for the PAL signal made in Appendix A.

Setting $G_r$ to the worst case values of 1.45 and 2, as obtained from Figure 7.30 we can now consider a full range of RGB signals and determine the maximum and minimum excursions of the encoded signals. Such an analysis has been performed numerically using the equations listed above. The results are shown below in Table 7.4, compared with normal PAL.

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Min. mV</th>
<th>Max. mV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal PAL</td>
<td>-273</td>
<td>1000</td>
</tr>
<tr>
<td>K-PAL (Implementation 1)</td>
<td>-317</td>
<td>1075</td>
</tr>
<tr>
<td>K-PAL (Implementation 2)</td>
<td>-440</td>
<td>1217</td>
</tr>
</tbody>
</table>

Table 7.4 Comparison of maximum and minimum excursions of PAL signal for different encoders.

We see that both implementations can cause over-range signals, both at the top and bottom limits. K-PAL 1, however, is only just outside the range, whereas K-PAL 2 is well outside. A better feeling for the significance of these results can be obtained by plotting the over-range values as a function of RGB colour space. Figure 7.32 and Figure 7.33 show the areas of RGB colour that cause over and under-range signals.
For K-PAL 1 under-range values can only occur for a tiny 0.07% of possible colours (Figure 7.31 (a)). Over-range values are more likely, but still only caused by a tiny range of colours (1.5% in Figure 7.31 (b)). For K-PAL 2 the situation is quite different, with 4.6% of colours capable of causing under-range values, and a huge 35% causing over-range signals.

In practice the likelihood of out-of-range values occurring is much smaller than just these results suggest. For any out-of-range signal to occur all of the following conditions must be present

1. A colour signal within the colour space areas shown in Figure 7.31 and Figure 7.32.
2. Maximum (or near maximum) amplitude signals on all three RGB inputs.
3. Presence of both low ($< L_C$) and high ($> L_C$) frequencies in same picture area.
4. A vertical frequency offset placing the signal at or near the peaks of the luminance 'envelope' response shown in Figure 7.30

An accurate estimation of the statistical likelihood of these conditions occurring is beyond the scope of this analysis, but it is likely to be very uncommon. Even when an over-range signal does occur it will likely be present only in a very small area of the picture.

7.8 Comparison of K-PAL and WC-PAL implementations

We will now perform a detailed comparison of the mathematical and simulated characteristics of K-PAL and WC-PAL, using a range of the tools available for analysis.

In comparing the performance of these television coding techniques there are a great many areas to be considered. The trade-offs between these all are difficult to judge, as many of the parameters
cannot be easily compared. For example, is an increase in luminance resolution worth a resulting increase in aliasing, or a decrease in cross-luminance worth loosing extra chrominance resolution? As a step towards making these judgements this section will look separately at each of the following performance parameters:

- luminance resolution
- luminance aliasing
- chrominance resolution
- cross-chrominance
- cross-luminance
- U/V cross-talk (inter-chrominance)

In studying these we are interested in finding out:

1. The performance differences between K-PAL and WC-PAL
2. The differences between enhanced and compatible reception.
3. The changes in performance with different implementations of the technique ie. choice of $R_C$, $F_U$, and $R_T$.

The possible variations in implementation are too many to consider in detail. Here only two implementations of WC-PAL and K-PAL will be looked at. For each implementation the WC-PAL and K-PAL versions will be chosen as similar as possible, to allow meaningful comparison of the performance of the two coding techniques.

### 7.8.1 Choosing an implementation for $R_C / F_U$ and $R_T$

Taking all the factors described earlier into account a choice of filters has been made for the two implementations of K-PAL and WC-PAL.

<table>
<thead>
<tr>
<th>Implementation</th>
<th>$R_C / F_U$</th>
<th>$R_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st order vertical</td>
<td>1st order vertical</td>
</tr>
<tr>
<td>2</td>
<td>5th order vertical</td>
<td>3rd order vertical-temporal</td>
</tr>
</tbody>
</table>

*Table 7.5 Choice of filters for two implementations of K-PAL and WC-PAL.*

Block diagrams of these filters are shown in Appendix C. The response of these filters is shown below as a cross-sectional amplitude characteristic and as a vertical-temporal contour.

For implementation 1 the filters for $R_C / F_U$ and $R_T$ are both the same, and their response is shown below in Figure 7.33. These are the simplest filters possible for use in $R_C / F_U$. 

![Luminance / chrominance filter - Ver 1](a)

![Luminance / chrominance filter - Ver 1](b)
For implementation 2 the chrominance and luminance filters are different. For chrominance $R_c$ has a much sharper characteristic, for luminance $R_y$ has a wider aperture, due to it being vertical-temporal instead of just vertical.

Note in these plots that the amplitude response is plotted only for positive frequencies, and is equivalent to a cross-section through the contour plots along the line $f_r = 0$. Note also that for $F_0$ in K-PAL these filters will need to be modified in accordance with the approach described in section 7.6.1.

For each implementation, and each technique four methods will be used to analyse the six performance parameters listed earlier.

1. Examination of the derived mathematical function describing the response.
2. Plots in 1 (vertical), 2 (vertical-temporal), or 3 (spatio-temporal) dimensions of the frequency response characteristic, taken from the describing equations or actual simulation results.
3. Decoded test images, e.g. spatial zone plate, or BBC young couple.
4. Objective assessment results obtained using the new assessment technique described in Chapter 6.

At the end of this section a summary will be presented of the most important results.

7.8.2 The Luminance Response

The luminance responses for the four paths of interest may be taken from equations (7.85), (7.79), (7.61), and (7.55) and are listed below

<table>
<thead>
<tr>
<th>Path</th>
<th>K-PAL</th>
<th>WC-PAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced</td>
<td>( R_Y^2 \left( 1 - 2F_U + 2F_U^2 \right) )</td>
<td>( R_Y^2 )</td>
</tr>
<tr>
<td>Compatible</td>
<td>( NR_Y )</td>
<td>( NS_Y R_Y )</td>
</tr>
<tr>
<td>Conventional</td>
<td>( N )</td>
<td></td>
</tr>
</tbody>
</table>

Also listed for comparison is the luminance response for conventional PAL.

These equations are intended to describe the vertical-temporal response and (except for the PAL notch) ignore the effects of horizontal filters.

The functions that make up these responses are

- \( R_Y \) - Luminance pre-filter
- \( F_U \) - K-PAL chrominance pre-filter / band segregating filter
- \( S_Y \) - WC-PAL luminance assembler
- \( N \) - PAL decoder luminance notch

For the compatible paths the notch filter is likely to remove most high frequency luminance, meaning the effects of \( R_Y \) and \( S_Y \) will be small.

For WC-PAL enhanced reception the response is as expected in (7.144), being the combination of the encoder and decoder luminance pre-filters. For enhanced K-PAL decoding, however, the response has been modified by the band segregating filter \( F_U \) as well. Figure 7.36 shows equation (7.144) and (7.145) for the enhanced luminance response of WC-PAL and K-PAL, under both implementations. The effect of \( F_U \) on the luminance response can be seen in producing a marked difference between K-PAL and WC-PAL in both implementations. This difference includes a factor of two attenuation of low / high vertical frequencies. The considerable distortion of the luminance response for K-PAL is clearly unacceptable and must be compensated for.
Figure 7.36 Comparison of K-PAL and WC-PAL enhanced luminance responses for, (a), (b) vertical luminance pre-filter in implementation (Ver) 1, and (c), (d) vertical-temporal pre-filter in implementation (Ver) 2.

7.8.2.1 Luminance ripple compensation

The function $1 - 2F_u + 2F_u^2$ that is causing the luminance distortion in K-PAL will be referred to as the luminance ripple function. This function is plotted below in Figure 7.37 for $F_u$ in implementations 1 and 2.
This function causes the unwanted distortion of the luminance response that is evident in Figure 7.36(b) and (d). These distortions result in, at worst, an attenuation of the luminance by -6 dB. This can be compensated for in the decoder by using a filter with the ideal characteristic

\[ F_c(\text{ideal}) = \frac{1}{1 - 2F_u + 2F_u^2} \]  

(7.148)

This ideal filter is impossible to realise in practice, but to a first approximation we may use

\[ F_c = 1 + 4F_u - 4F_u^2 \]  

(7.149)

Using this compensating filter the ripple characteristics in Figure 7.36 are reduced significantly from a maximum of -6 dB to around +1 dB, as shown in Figure 7.38. This level of distortion is far more acceptable.

When the compensating filter is added to the K-PAL decoder the luminance responses shown in Figure 7.36 (b) and (d) become those depicted in Figure 7.39.
The K-PAL luminance responses now matches that of WC-PAL much more closely. The disadvantage of the compensating filter $F_C$ is that it requires twice as many delays as $F_U$, and results in increased decoder complexity. As an alternative $F_C$ may be placed in the encoder (i.e. pre-compensation). Doing this, however, means a distortion of the luminance response for compatible reception - although such distortion would likely have little effect on the picture in a conventional PAL decoder as most high frequency luminance is removed by the notch filter anyway.

7.8.2.2 Luminance distortion in Implementation 1

For K-PAL implementation 1 the luminance distortion can actually be taken advantage of and used to improve the luminance response. Instead of applying a compensating filter we can simply scale the luminance by a factor of two ($F_C = 2$). Figure 7.36 (b) then becomes

![Graph of K-PAL Enhanced - Ver 1 - Luminance (Y)](image)

Comparing Figure 7.40 with the luminance response for WC-PAL plotted in Figure 7.36(a) we see that the luminance ripple function and the luminance pre-filter characteristic actually complement each other to some degree, resulting in a wider luminance response than otherwise. At the cost of some ripple (~1 dB) the vertical luminance bandwidth has been greatly increased, from around 40 cph to nearly 70 cph. K-PAL implementation 1 is thus a particularly useful version of K-PAL, providing unusually good luminance bandwidth with minimum coding complexity and no decoder luminance compensation required.

7.8.3 Simulation Results

The simulations used to produce the results presented for K-PAL in this Chapter incorporate the decoder changes that have just been described. The compensating filter used is different for two implementations, being (7.149) for K-PAL2 and simply a scaling factor of 2 for K-PAL1.

The new methods for spectral analysis and objective assessment described in Chapters 5 and 6 have been applied to these simulations, producing a set of results for each performance parameter. Figure 7.41 presents the objective assessment results for the luminance signal.
7.8 COMPARISON OF K-PAL AND WC-PAL IMPLEMENTATIONS

In Figure 7.41 the figure-of-merit bandwidth measures for the luminance bandwidth have all been normalised w.r.t. the bandwidth of conventional PAL. Looking first at compatible reception it may seem surprising that all techniques produce a lower bandwidth than conventional PAL. The reason for this is revealed by studying the spatio-temporal contour plots of the luminance response. Comparing the plots for compatible reception of PAL and K-PAL in Figure 7.42 we can see an extra loss of intermediate vertical frequencies for the K-PAL encoded signal. This is due to the wider response of the encoder bandpass filter, causing the luminance pre-filter to remove more intermediate luminance frequencies than the PAL decoder notch by itself.

Under enhanced reception (Figure 7.43) the advantages of the pre-filter become apparent, producing an extended horizontal bandwidth in all cases, and a better figure-of-merit than PAL for most. For the several cases where the figures-of-merit are still lower than PAL the extra horizontal bandwidth has not been enough to offset the loss of luminance bandwidth due to the encoder bandpass filter. This brings out one of the problems with the bandwidth measure. As discussed in Chapter 6 it does not take into account the statistical likelihood, or visual importance of different luminance frequencies. In practice the extra horizontal bandwidth is likely to be more important for picture quality than the lost diagonal frequencies. This is somewhat verified by the young couple test images shown later in Figure 7.53(d) and (e), where the luminance detail in the striped shirt is considerably clearer.
Figure 7.43 also shows the differences between implementations of K-PAL and WC-PAL. Comparing (a) and (c) we can see that K-PAL1’s higher figure-of-merit in Figure 7.41 is due to its extra vertical bandwidth. This is a result of the effects discussed earlier and the extended luminance response shown in Figure 7.40. For implementation 2 the theoretical luminance responses (Figure 7.36(c) and Figure 7.39(b)) show only small differences, and as a result the spatio-temporal contours for K-PAL and WC-PAL are very similar. The objective assessment results indicate that the small differences present give K-PAL a very slight bandwidth advantage.

The differences between the vertical and vertical-temporal luminance pre-filters are particularly obvious. The higher vertical bandwidth of Figure 7.43(b) and (d) is responsible for implementation 2’s high figure-of-merit for stationary image content. Conversely the extra temporal bandwidth of Figure 7.43(a) and (c) produces improved performance for moving image content. For these cases the figures-of-merit are better than normal PAL.

7.8.4 Luminance Aliasing
The modification applied to the luminance in both K-PAL and WC-PAL introduces aliasing that cannot be completely removed again. The functions describing the residual aliasing, under both compatible and enhanced reception are (from (7.85), (7.79), (7.61), and (7.55))

Enhanced K-PAL

\[ F_C R_T (1 - R_T) (1 - F_U) \]  \hspace{1cm} (7.150)

\[ R_T (1 - R_T) \]  \hspace{1cm} (7.151)
7.8 COMPARISON OF K-PAL AND WC-PAL IMPLEMENTATIONS

WC-PAL

Compatible
K-PAL
\[ N(1 - R_y)(1 - 2F_v) \]  
(7.152)

WC-PAL
\[ N(1 - R_y)S_r \]  
(7.153)

Conventional
PAL
No aliasing  
(7.154)

where the K-PAL luminance is now modified by the ripple compensating filter \( F_c \).

For the compatible paths most luminance aliasing will be eliminated by the sub-carrier notch filter. The visibility of any residual aliasing will be minimised by the high spatio-temporal offset of the sub-carrier. This hypothesis is supported by the objective assessment results in Figure 7.44 where residual aliasing in compatible reception is always lower than under enhanced reception.

For enhanced reception the alias levels are residual amounts of the PAL modified luminance \( \tilde{Y} \) that have been left in the transition regions of the luminance pre-filter \( R_y \). These residual amounts are graphed below Figure 7.45 for the functions described by (7.130) - (7.133).

\[ N \]  
(7.154)
Figure 7.45 Residual aliasing levels for enhanced coding in WC-PAL (a), (b), and K-PAL (c), (d).

Figure 7.45(a), (c) for implementation 1 shows that increased luminance aliasing is one price paid by K-PAL1 for its increased luminance bandwidth. Aside from this the generally lower level of aliasing in WC-PAL is likely due to the fact that K-PAL places more energy in the high frequency luminance (see section 7.7.1).

Comparing results for stationary and moving picture content the most visible aliasing occurs for moving luminance. This is to be expected, as the 2f_c shift that is applied by the PAL modifier causes stationary frequencies to alias to frequencies with a high spatio-temporal offset, and consequent low visibility.

Overall Figure 7.44 shows that all implementations tested produce some luminance aliasing. As normal PAL has no aliasing the figures-of-merit shown above have been normalised w.r.t. PAL cross-luminance. These results therefore show that luminance artefacts produced by luminance aliasing are likely to be much less noticeable than those caused by cross-luminance. Figure 7.44 may be compared directly with cross-luminance results presented later in Figure 7.54. In comparing these we see that even with enhanced coding cross-luminance is still a far more visually perceivable artefact than luminance aliasing. As discussed in Chapter 6 the visibility of artefacts as indicated by the objective figures-of-merit does not take into account the likelihood of their occurrence. In this case the high frequency luminance that produces luminance aliasing is likely to be much rarer than the low frequency chrominance that causes cross-luminance, reducing the significance of the luminance aliasing artefact even further. Results using the young couple test image in Figure 7.53 in section 7.8.6 appear to back this up, there being no noticeable distortion of the fine luminance detail in the striped shirt.

7.8.5 Chrominance

From equations (7.94), (7.69), (7.60), and (7.53) the chrominance responses for K-PAL, WC-PAL, and conventional PAL are

<table>
<thead>
<tr>
<th>Implementation</th>
<th>K-PAL</th>
<th>WC-PAL</th>
<th>K-PAL</th>
<th>WC-PAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced</td>
<td>$F_v^{+2}$</td>
<td>$R_c^2$</td>
<td>$D_v F_v^3$</td>
<td>$D_v R_c S_r$</td>
</tr>
<tr>
<td>Compatible</td>
<td>$D_v$</td>
<td>$D_v$</td>
<td>$D_v$</td>
<td>$D_v$</td>
</tr>
<tr>
<td>Conventional</td>
<td>$D_v$</td>
<td>$D_v$</td>
<td>$D_v$</td>
<td>$D_v$</td>
</tr>
</tbody>
</table>
where $R_c$ is the WC-PAL chrominance pre-filter, and $D_u$ the PAL delay line. The chrominance horizontal lowpass filter $L_c$ has been omitted from these functions; it has a virtually identical effect on all paths.

Obviously, from these functions the chrominance resolution will in most cases be reduced compared to conventional PAL. However the reduction is mostly due to a loss of vertical chrominance frequencies. Considering the normal imbalance between chrominance horizontal and vertical resolution this is perhaps not unreasonable.

The objective assessment results for chrominance bandwidth are shown below in Figure 7.46. Under compatible reception the chrominance bandwidth is, as indicated, always lower than conventional PAL. The extra factor in $S_Y$ in (7.158) means that WC-PAL has a lower chrominance bandwidth than K-PAL for both implementations.

For enhanced reception chrominance bandwidth is lower than PAL for implementation 1, but higher than PAL for implementation 2. Looking at plots of $R_c / F_u$ in Figure 7.47 we can see why.

In implementation 1 the response of the cascaded pre- and post-filters has a greater attenuation of low chrominance frequencies than the simple delay-line. In implementation 2, however, the higher order chrominance pre-filter produces an improved low frequency response compared to the delay-line, resulting in a chrominance bandwidth higher than PAL and almost twice as high as implementation 1. Note that the plots in Figure 7.47 will be identical for both WC-PAL and K-PAL.
A puzzling result in Figure 7.46 is the differences between K-PAL and WC-PAL for enhanced reception, implementation 2. Equations (7.56) and (7.57) predict that the enhanced responses should be identical. For implementation 1 this is indeed the case; for implementation 2, however, WC-PAL has significantly lower bandwidth than K-PAL. The spatio-temporal contours for the enhanced chrominance responses, shown in Figure 7.48, back up this difference. Figure 7.48(a) and (c) representing implementation 1 are identical, but Figure 7.48(b) and (d) for implementation 2 are different. The difference is a loss of diagonal chrominance frequencies in WC-PAL2.

The reason for this lies in the failure of the WC-PAL assembler / disassembler when the sidebands around \( f_c \) are not perfectly symmetrical. As discussed in section 7.6.6 the resulting attenuation of chrominance occurs mostly in the intermediate vertical frequencies. In implementation 1 the lower vertical chrominance bandwidth seen Figure 7.48(a), (c) means that these frequencies have already been removed by the pre-filter. In implementation 2, however, the higher order pre-filter preserves these frequencies, meaning that their loss in WC-PAL can be noticed, causing a lower figure-of-merit compared to K-PAL.

### 7.8.6 Cross-chrominance

This is the parameter that shows up some of the most dramatic differences and improvements for K-PAL and WC-PAL. The cross-chrominance has been expressed in the functions derived earlier (e.g. (7.53) and (7.60)), in the form
Comparison of K-PAL and WC-PAL Implementations

\[ C = G Y^4 \]  \hspace{1cm} (7.160)

Where \( G \) is to be defined. In order to know the 'envelope' or maximum level of cross-chrominance for each luminance input frequency we will set \( Y = 1 \). The luminance pre-filter \( R_Y \) is a complementary filter, and the definition of this in equation (7.11) implies that

\[ R_Y^* = 1 - R_Y \]  \hspace{1cm} (7.161)

Using these definitions (7.160) becomes simply

\[ C = G \]  \hspace{1cm} (7.162)

Using this identity, and equations (7.94), (7.69), (7.60), (7.53) and (7.54) derived earlier the formulas describing the cross-chrominance envelope are

<table>
<thead>
<tr>
<th>Enhanced</th>
<th>K-PAL</th>
<th>WC-PAL</th>
<th>(7.163)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-PAL</td>
<td>((1 - F_Y^<em>)F_Y^</em>)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>WC-PAL</td>
<td>0</td>
<td>((1 - F_Y^*)D_u)</td>
<td>(7.164)</td>
</tr>
<tr>
<td>Compatible</td>
<td>K-PAL</td>
<td>((-\frac{1}{2}S_cD_u)</td>
<td>(7.165)</td>
</tr>
<tr>
<td>WC-PAL</td>
<td>((-\frac{1}{2}S_cD_u)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>PAL</td>
<td>(D_{u})</td>
<td>(7.166)</td>
</tr>
</tbody>
</table>

It must be noted here that these results are using the assumption of equal energy across the luminance spectrum \((Y = 1)\). Under this assumption for the enhanced techniques the formulae are the result of two cancelling effects. For a particular luminance frequency \( Y \) the presence of the luminance alias \( \tilde{Y} \) in the spectrum effectively doubles the energy available for causing cross-chrominance. However, because half of the frequencies that cause cross-chrominance are removed by the luminance pre-filter the average cross-chrominance remains the same.

In practice, however, this reasoning does not hold completely, as the luminance frequencies removed by \( R_Y \) are high diagonal (or high vertical-temporal) frequencies that occur less often. Because of this the actual levels of cross-chrominance for the enhanced paths will be lower than the formulae above might suggest. For the worst case, where the picture content contains no high diagonal / diagonal-temporal frequencies there is a doubling of the high frequency luminance energy in the enhanced signal. In this case there will be a factor of 2 increase in the expressions in (7.163)-(7.166).

Looking now at the objective assessment results in Figure 7.49 we see that although WC-PAL has theoretically zero cross-chrominance, in practice due to non-ideal horizontal filters this is not achieved. WC-PAL enhanced reception is, however, still significantly better than K-PAL, particularly for moving picture content.
Although the enhanced K-PAL path does contain higher levels of cross-chrominance compared to WC-PAL it is still considerably lower than normal PAL. Figure 7.50 shows plots of equation (7.163) for K-PAL implementations 1 and 2, compared against normal PAL (eqn. (7.167)). Coarse cross-colour occurring at low vertical frequencies has been almost completely removed. The result is a figure-of-merit for K-PAL enhanced reception that is 60-80% lower than normal PAL (with WC-PAL being 85-90% lower).

Under compatible reception the differences in WC-PAL and K-PAL are reversed, with K-PAL producing less cross-chrominance than WC-PAL. The difference is not as significant as for enhanced reception, with a maximum difference of around 8%. Plotting (7.66) and (7.67) in Figure 7.51 we can see how these improvements are gained. Most important is the lower level of coarse cross-colour in K-PAL for both implementations, although this comes at the expense of an increase in fine cross-colour. This trade-off between decreased coarse cross-colour and increased fine cross-colour is likely the reason why K-PAL2 does not show particularly significant improvements in its figure-of-merit compared to K-PAL1. In Figure 7.51 WC-PAL represents a step below K-PAL1, with slightly more coarse cross-colour, and less fine cross-colour. Although the level of compatible cross-chrominance is theoretically the same for all implementations of WC-PAL (equation 7.166) is independent of $R_C$ and $R_Y$). Figure 7.49 does show some differences, once again this is likely due to horizontal filter effects.

The theoretical and objective assessment results may be compared with some test results in a more conventional form. Figure 7.52 and Figure 7.53 show a selection of pictures for WC-PAL2 and K-
PAL2 under both enhanced and compatible reception. The two test images chosen are the commonly used luminance spatial zone, and a section of the BBC young couple test image. A number of observations may be made, backing up the previous results.

1. Under compatible reception both K-PAL and WC-PAL reduce coarse cross-colour (compare (a) with (b) and (c) in both figures).

1. Under compatible reception K-PAL reduces cross-colour more than WC-PAL (compare (b) and (c)).

1. Under enhanced reception WC-PAL reduces cross-colour more than K-PAL (compare (d) and (e)), with K-PAL having residual find cross-colour.

**Figure 7.51** Comparison of compatible cross-chrominance responses for K-PAL and WC-PAL with normal PAL, (a) K-PAL1, (b) K-PAL2, (c) WC-PAL (both implementations).
Figure 7.52 Luminance spatial zone upon compatible reception with encoder (a) Normal PAL, (b) K-PAL2, (c) WC-PAL2; enhanced reception for (d) K-PAL2 coding, (e) WC-PAL2 coding.
Figure 7.53 Section of young couple test image upon compatible reception with encoder (a) Normal PAL, (b) K-PAL2, (c) WC-PAL2; enhanced reception for (d) K-PAL2 coding, (e) WC-PAL2 coding.
7.8.7 Cross-luminance

Cross-luminance is generally a less noticeable artefact than cross-chrominance, but is still important. Below are the equations describing the level of cross-luminance in WC-PAL, K-PAL and conventional PAL. These are taken from equations (7.85), (7.79), (7.61), and (7.55) as derived earlier, and for K-PAL now includes the ripple compensating filter $F_c$.

\[
\text{Enhanced K-PAL:} \quad \frac{1}{2} F_c \left(1 - F_u \right) F_u R_y \quad (7.168)
\]

\[
\text{WC-PAL:} \quad 0 \quad (7.169)
\]

\[
\text{Compatible K-PAL:} \quad \frac{1}{2} N \quad (7.170)
\]

\[
\text{WC-PAL:} \quad \frac{1}{2} S_c N \quad (7.171)
\]

\[
\text{Conventional PAL:} \quad N \quad (7.172)
\]

Once again these results are based on the assumption that the chrominance pre-filter $R_c$ IF$_u$ is discarding half the chrominance energy. In practice, as for the cross-luminance case, the frequencies discarded are statistically much less common. For the average picture content most of the chrominance energy will reside in the passband of $R_c$ IF$_u$. In the extreme case none of the chrominance frequencies will end up being discarded, meaning that the level of cross-luminance will be twice the level equations (7.168)-(7.171) suggest. To allow comparison of this worst case scenario the plots of the cross-luminance characteristics in Figure 7.55 have been scaled by a factor of two.

Looking first at the enhanced path WC-PAL once again shows the best result with WC-PAL figures-of-merit for cross-luminance in Figure 7.54 being half those of K-PAL. However, these results fall well short of the zero cross-effects predicted by (7.169). As was explained in a previous section this is due to the failure of phase segregation in the horizontal filter transition regions, a factor (7.169) does not take into account.

Figure 7.54 Objective assessment results for K-PAL and WC-PAL cross-luminance.

Figure 7.55 plots equation (7.68) for K-PAL implementations 1 and 2. Both implementations produce their greatest cross-luminance at low vertical luminance frequencies. At these low vertical frequencies, which will be the most visible ones, cross-luminance is only around a half that of normal PAL and has a similar characteristic in both implementations. This correlates well with the results shown in Figure 7.54 where the figures-of-merit for K-PAL enhanced reception are just under 50% of PAL and are almost identical for implementations 1 and 2.
7.8 COMPARISON OF K-PAL AND WC-PAL IMPLEMENTATIONS

Under compatible reception the difference between WC-PAL and K-PAL is slightly less, though WC-PAL is still significantly better than K-PAL. Looking at equations (7.71) and (7.72) (ignoring effects of the notch filter) plotted in Figure 7.56 we see why. WC-PAL has a far higher attenuation of the most visible low luminance frequencies than does K-PAL.

Overall the results in Figure 7.54 suggest that both techniques offer smaller improvements for cross-luminance than they do for cross-chrominance.

7.8.8 Inter-chrominance

Inter-chrominance or U/V crosstalk does not normally occur much in PAL coding due to the synchronous demodulation of the quadrature modulated signals $U$ and $V$ (although phase distortion during transmission can cause quite bad U/V crosstalk). In K-PAL the chrominance signals are also modulated on separate carriers, meaning they can be fully separated again. WC-PAL, however, combines $U$ and $V$ into a single signal preventing them from being completely separated in a decoder. The result is inter-chrominance for both enhanced and compatible reception. The ‘envelope’ of this inter-chrominance is taken from equations (7.69) and (7.53)

\[
R_c(1 - R_c)
\]  

(7.173)
If horizontal filter effects are ignored the equivalent inter-chrominance equations for K-PAL and conventional PAL will be zero.

In practice, as Figure 7.57 shows, the levels of inter-chrominance are not quite zero for PAL and K-PAL, this being mostly due to the partial loss of the chrominance upper sideband with the resulting failure of the chrominance QAM (see section 7.6.6). These inter-chrominance values are, however, still considerably lower than WC-PAL.

These results have been normalised w.r.t. the level of cross-chrominance present in normal PAL. This allows us to compare inter-chrominance with the other colour artefact cross-chrominance which was shown in Figure 7.49. Compared to the cross-chrominance normally present in PAL Figure 7.57 suggests that inter-chrominance in WC-PAL is insignificant, particularly for stationary picture content where it is almost non-existent. Under enhanced reception and moving image content however, inter-chrominance is actually on a par with cross-chrominance (Figure 7.49) as a chrominance artefact. Any further improvements in cross-chrominance for WC-PAL would therefore have to be accompanied by a reduction in inter-chrominance in order to gain an overall increase in picture quality.

### 7.8.9 Comparison of Results

Table 7.6 provides a summary of the 'winners' of each performance category, based upon the objective assessment results that have been presented. Under compatible reception there was virtually no difference between the techniques for luminance bandwidth and this slot is therefore blank.

The important differences summarised by Table 7.6 are:

1. K-PAL performs better for compatible reception.
2. WC-PAL performs better for enhanced reception.
3. Overall K-PAL preserves more bandwidth.
4. Overall WC-PAL has less chrominance and luminance artefacts.

K-PAL does appear to achieve its main aim of reducing cross-chrominance in a conventional PAL decoder, although WC-PAL is not far behind. Section 7.6.8 has explained how modification of the K-PAL coding by use of an asymmetrical chrominance filter can further improve the reduction of cross-chrominance. One K-PAL failing is its relatively poor performance for cross-effect suppression under enhanced reception. This is partly the result of the coder being optimised for compatible reception. However, even under compatible reception WC-PAL performs better for...
cross-luminance. This is an area for improvement of K-PAL, and some ideas are described in the Conclusions in Chapter 8.

<table>
<thead>
<tr>
<th>Performance category</th>
<th>Compatible</th>
<th>Enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminance bandwidth</td>
<td>K-PAL</td>
<td>K-PAL</td>
</tr>
<tr>
<td>Chrominance bandwidth</td>
<td>K-PAL</td>
<td>WC-PAL</td>
</tr>
<tr>
<td>Cross-chrominance</td>
<td>K-PAL</td>
<td>WC-PAL</td>
</tr>
<tr>
<td>Cross-luminance</td>
<td>WC-PAL</td>
<td>WC-PAL</td>
</tr>
<tr>
<td>Luminance aliasing</td>
<td>WC-PAL</td>
<td>WC-PAL</td>
</tr>
<tr>
<td>UV Cross-talk</td>
<td>K-PAL</td>
<td>K-PAL</td>
</tr>
</tbody>
</table>

*Table 7.6 The 'winners' of each performance category.*

The results presented have covered two implementations of both techniques, one using low order comb filters and the second more complex higher order (and larger delay) filters. The use of more complex filters helps to shorten filter transition regions and allows the coder’s performance to better approach the theoretical ideal. As expected, the more complex filtering in implementation 2 does produce improvements in virtually all performance categories, increasing bandwidth and reducing artefacts. However, with the exception of chrominance bandwidth, the improvement is quite small, almost indistinguishable in some cases. It is therefore debatable if this improved performance is worth the considerable increase in comb filter storage requirements, from 7 lines in implementation 1, to over 12 fields (3765 lines) in implementation 2. Large coder signal delays can also cause synchronisation problems in a studio environment.

**7.9 Summary**

This Chapter has described in considerable detail two coding techniques, WC-PAL and the author’s K-PAL, that utilise phase segregation to improve reception of the PAL signal. Overall the treatment has included, for both WC-PAL and K-PAL:

1. Complete mathematical description of phase segregation principles in enhanced PAL coding.
2. Derivation of formulae describing received signals for WC-PAL and K-PAL under both enhanced and compatible reception.
3. Design considerations of luminance and chrominance pre-filters - including the particular requirements of the K-PAL band-segregating filter.
4. Discussion and derivation of optimal forms for all horizontal filters.

In particular many of the design issues surrounding the design of a K-PAL coder have been identified and discussed:

1. Creation of a minimal block implementation suitable for hardware construction.
2. The use of asymmetrical luminance filtering to increase chrominance and cross-colour performance.
3. A study of the signal content that can cause over-range PAL signal values in a K-PAL encoder.

The Chapter ends with a separate study of each performance parameter (bandwidth, aliasing, etc) for two implementations of K-PAL and WC-PAL. Using objective assessment, response plots, spatio-temporal spectra, and standard test images the performance of the two techniques and two implementations is compared.

K-PAL is shown to be a viable alternative to other enhanced PAL coding techniques, with no significant drawbacks or performance problems. The overall conclusion is that K-PAL offers
improved performance under compatible reception (significantly reduced cross-colour), but gives slightly inferior performance for enhanced reception. Performance for both techniques improves with the use of higher order pre-filters. However, the improvement is not great and may arguably not be worth the extra complexity.

This Chapter has covered many aspects of K-PAL and WC-PAL, but has by no means been exhaustive. Space constraints have meant that some areas, such as enhanced reception of a conventional PAL signal, have been omitted.
CHAPTER 8

CONCLUSIONS

This Chapter will present an overall summary of the areas covered in this thesis, and the new work contained therein. Detailed discussion and summary of the work has already been presented on a Chapter by Chapter basis. Consequently this Chapter to some degree repeats these summaries, as well as tying together all the author’s new work.

8.1 The Human Visual System

The display of television images is ultimately aimed at pleasing the human viewer. A good understanding of the human visual system (HVS) and its properties is therefore crucial to improving television coding techniques. Chapter 2 presents an overview of this field and consequently derives a number of limits and requirements for television, as well as some models for the eye’s response. This information has been used later in Chapter 6 as part of the author’s new objective assessment method. Chapter 2 served mostly as background, but contains some new work with the modification and development of mathematical models for the eye’s chromatic and achromatic threshold response. This Chapter does in itself remain a contribution to the field in bringing together in one place information on the HVS that is not widely available or known to those working in the field of enhanced television.

8.2 The Theory of PAL Signal and Coding Analysis

Chapters 3 and 4 dealt with the background to PAL signal analysis and provided an overview of enhanced coding techniques. A theme of these Chapters has been the development of a consistent theoretical framework for coding analysis.

As explained in Chapter 3 the PAL signal has previously been presented in various combinations of 1, 2, or 3 dimensions, and has been explained in terms of line and field harmonics, as well as vertical and temporal frequencies. Similarly the operation of coding systems has been variously described in terms of line-to-line or field-to-field phase differences, spectrum folding, and comb filtering. It has been the author’s contention that these various approaches are inadequate and have hindered proper understanding. Similar coding techniques have often been described in completely different ways, as well as being presented in very different forms. Because of this, one aim of Chapters 3 & 4 has been an attempt to standardise this theory and to produce an all encompassing framework for the study of enhanced television.

To this end Chapter 3 has extended the early work of Drewery [Drewery 1977] and has generalised all previous approaches to the three dimensional view of the television signal as a sampled three dimensional spatio-temporal signal. It is hoped that Chapter 3 may provide a useful introduction to the theory for a newcomer to the field and help illustrate the relationships between the different existing views of PAL coding. It is the author’s belief that PAL should be viewed as a three dimensional spatio-temporal signal first and foremost, with simpler analysis used for specific cases only.

Moving on from PAL signal analysis to PAL coding Chapter 4 has summarised previously published PAL coding techniques. In keeping with the development of a common framework for enhanced PAL
analysis the similarities between the different techniques are explained, and they are all placed in categories of either band, phase, or time segregation.

Once again it is intended that Chapter 4 may prove useful to a newcomer to the field. When presented in a common framework the similarities and differences between the many enhanced coding techniques become more apparent, and are far more easily understood.

8.3 Calculating the 3D spatio-temporal characteristics of a coding system - a new way of understanding PAL coding

The mathematical representation of the PAL signal derived in Chapter 3 has been further extended in Chapter 5 with a generalised mathematical representation for a scanned moving colour image coding system. This coding system is characterised as a linear periodically time varying (LPTV) many-input many-output (MIMO) system. Chapter 5 develops a solid mathematical platform for the analysis of enhanced television coding techniques with a representation that is complete, self consistent, and accurate. The coding system is defined by a set of three dimensional 'transfer functions' or 'characteristics' that describe both the main (intra) signal paths for Y, U, V and the cross (inter) component paths. With three main signals and six possible cross-signal paths there is a total of nine characteristics required to describe the coding system. This representation has been further extended to distinguish artefacts (time varying components such as aliasing) within the main signal paths. After modification to reduce the number of dimensions a set of 30 three dimensional spatio-temporal functions are shown to provide a complete representation of a television coding system.

A possible problem with this representation is its complexity. However, it is the only one known to the author that can describe a system's characteristics completely. When it comes to analysing and comparing systems not all of this information may be needed in which case the results can be simplified. However, in the author's opinion it is important to always start from the complete representation before simplifying the analysis for a specific example.

A mathematical representation is little use without being able to apply it to actual coding systems. Computer simulation is a common method for testing coding techniques before implementation in hardware. Chapter 5 consequently describes in detail a numerical method for analysing an existing simulated coding system and calculating its characteristics using the representation defined earlier. This method is computationally intensive but definitely practical. This has been proven by use of the method for determining the characteristics of a wide range of coding techniques with many of these (e.g. WC-PAL, K-PAL, ...) being included throughout this thesis. To the author's knowledge these are the only known plots of actual coding system characteristics, as opposed to the conceptual three-dimensional diagrams used in all other literature.

The author's new representation has proven to provide significant insight into the understanding of coding techniques studied. It shows up even small differences in the characteristics that would be difficult to analyse using other approaches.

The disadvantage of the method, besides its complexity is that it cannot be used for coding techniques that employ non-linear operations. Simple examples, such as two mode adaptive coding may perhaps be modelled as two separate linear systems. However, more modern digital techniques such as MPEG are well outside the scope of the method. Further work on expanding the range of coding techniques that can be represented may be useful. However, it is the author's belief that the representation is not well suited to non-linear coding techniques - it having been developed first and foremost as a tool for enhanced analogue coding techniques. In view of the increased use of digital / non-linear techniques and the predicted demise of conventional analogue based television coding (e.g. the rise of HDTV) any future work might be better concentrated on finding alternative fields where this representation might also apply and could be used to provide insight.
8.4 A new approach to objective assessment

The availability of complete information on the spatio-temporal characteristics of a coding system, as provided by the new methods described in Chapter 5, has allowed the author to trial a new approach to objective assessment. As Chapters 2 and 6 explained the existing assessment methods have many inherent problems. Unfortunately, satisfactory alternative objective techniques have not yet been proven.

Chapter 6 has described a new objective assessment method based upon application of models of the HVS response to the spatio-temporal characteristics of the coding system. This method was applied to a set of six enhanced PAL coding techniques and compared with an existing set of subjective assessment results for these techniques. The results, whilst not providing extremely high correlation, do show some promise. As explained in Chapter 6 the method is at a very early stage of development. Considerable further work is needed and many questions and issues need to be looked at further.

The issues of a replacement quality assessment method aside, the figures-of-merit still provide a very useful set of performance parameters that can be used in comparing coding systems. They are made good use of in Chapter 7 for comparing the performance of K-PAL and WC-PAL.

8.5 A new improved PAL coding technique

Arguably the highlight of the author's new work is the discovery and development of a new PAL coding technique. Designated K-PAL this technique is similar to WC-PAL in its use of phase segregation, but different from WC-PAL and all other known techniques in its independent treatment of \( U \) and \( V \) signals. Normally \( U \) and \( V \) are coded into a single chrominance signal which is then combined with the luminance, in K-PAL they are treated separately.

The most important advantage of the technique is its very good suppression of cross-colour in conventional PAL decoders. Unlike most previous techniques for reducing cross-colour this is achieved without sacrificing luminance and chrominance vertical bandwidth. Chapter 7 has described K-PAL in detail, providing a complete mathematical derivation of its characteristics for both compatible and enhanced reception. Many other aspects were also studied including:

1. Choice of and design of luminance pre-filter and chrominance band segregation filter.
2. Optimal design of horizontal filters.
3. Obtaining a minimal block implementation.
4. Problems with over-range luminance signals.
5. Asymmetrical chrominance filtering (using differing vertical bandwidths for the \( U \) and \( V \) signals).

All aspects of the performance of K-PAL including bandwidth, cross-effects, and aliasing have been studied, using both conventional methods and the new methods of spatio-temporal spectra and objective assessment developed by the author. As a reference a parallel study of WC-PAL was undertaken. This in itself represents a more complete description of Weston Clean PAL than anything that is available in current literature.

The conclusion reached is that K-PAL performs better than any existing technique for compatible suppression of cross-colour whilst maintaining maximum chrominance and luminance bandwidth. Under enhanced reception K-PAL performs on a par with other enhanced techniques, but not quite as well as WC-PAL. As a result of this work and the conclusions made the author is actively pursuing commercial involvement in the further development of the K-PAL technique.

The author's treatment of K-PAL has been quite comprehensive. Any further work should concentrate on:

1. Construction of a hardware prototype.
2. Conducting CCIR subjective testing to confirm performance improvements.
3. Study of compatibility issues for studio equipment, transmission equipment, and existing coding techniques.

8.6 Summary

Overall this thesis has contributed considerable new work to the field of PAL television coding. The major new work in this thesis has been (i) a new and complete representation of a television coding system as a 3-D spatio-temporal MIMO LPTV system, (ii) the development, implementation, and demonstration of an entirely new method for determining complete and accurate spatio-temporal characteristics for any coding system, (iii) a fundamentally new approach to objective assessment, (iv) the discovery and analysis of all aspects of the new and optimal enhanced PAL coding technique named K-PAL. A number of journal and conference papers have been or are soon to be published.

It is the author's hope that this work may have some impact on the very confusing and dated approaches used currently in television analysis. It is also hoped that the new coding technique K-PAL will find application in commercial PAL broadcast equipment.

The work in this thesis also resulted in the development by the author of a sophisticated multi-dimensional image processing and spectra display system (TVPROC and PHIGSDRAW), without which much of the work could not have been achieved.
APPENDIX A

THE COLOUR TELEVISION SIGNAL

Colour in the television image is treated in a similar manner to the way the human eye treats colour. A colour image is considered as consisting of red (R), green (G) and blue (B) components, which become three separate picture signals. In the receiver these three signals are used to illuminate red, green and blue phosphors to reconstruct colour images.

For transmission purposes these three colour signals are combined to form a luminance (Y) and two colour difference signals (U & V). These are produced according to (A.1)-(A.3), where the weightings used to produce the luminance signal are in accordance with the relative response of the red, green and blue cones of the eye (see figure 2.7)

\[
Y = 0.299R + 0.587G + 0.114B \\
U = 0.493(B - Y) \\
V = 0.877(R - Y)
\]

The use of these different signals for transmission has three main benefits:

1. The luminance signal is identical to the black and white signal used by monochrome receivers. This allows the backward compatibility of colour transmission.
2. For transmission of monochrome pictures (i.e. as for black and white movies, or in picture area containing no colour), the colour difference signals are zero.
3. The eye’s response to colour is lower than its response to luminance. Thus the use of U and V to carry all the colour information allows a reduced bandwidth to be used for these signals, without affecting the luminance signal.

The eye’s response to colour is lower than its response to luminance. Thus the use of U and V to carry all the colour information allows a reduced bandwidth to be used for these signals, without affecting the luminance signal.

A.1 The gamma

Regarded overall a television system ought to be linear. However, a receiver CRT does not emit light in direct proportion to the voltage applied to it, due to the non-linearity of the beam current against grid voltage characteristic. This relationship is represented by equation (A.4)

\[ L \propto V_g^\gamma \]

where \( V_g \) is the crt voltage and \( \gamma \) is referred to as the receiver gamma\(^1\). In the UK and New Zealand a value of \( \gamma = 2.2 \) is assumed for the receiver. To compensate for this receiver characteristic the

\(^1\)The gamma of a picture tube is defined as the slope of the curve giving the logarithm of the luminance reproduced as a function of the logarithm of the video signal voltage when the brightness control of the receiver is set so as to make this curver as straight as possible in a luminance range corresponding to a contrast of at least 1/40.
Signals are *gamma corrected* before transmission. These gamma corrected signals are referred to as $R'$, $G'$, $B'$ where

$$R' = R'^{\gamma}$$
$$G' = G'^{\gamma}$$
$$B' = B'^{\gamma}$$

Thus (A.1) - (A.3) become

\[
Y = (0.299R + 0.587G + 0.114B)^{\gamma} \tag{A.8}
\]
\[
U = (0.493(B - Y))^{\gamma} \tag{A.9}
\]
\[
V = (0.877(R - Y))^{\gamma} \tag{A.10}
\]

However, the equations in fact used for PAL are [CCIR Report 624-3]

\[
Y' = 0.299R^{\gamma} + 0.587G^{\gamma} + 0.114B^{\gamma} \tag{A.11}
\]
\[
U' = 0.493(B^{\gamma} - Y') \tag{A.12}
\]
\[
V' = 0.877(R^{\gamma} - Y') \tag{A.13}
\]

These are clearly incorrect for any colour signal except when the red, green and blue values are identical ($R = G = B$). Under normal circumstances the errors created are actually cancelled upon recombination of the $YUV$ signals to produce $R'$, $G'$ and $B'$. However, for high luminance frequencies where the colour information is not transmitted; or when there is an error in the colour difference signals (caused by sub-carrier phase errors), the incorrect encoding of $Y'$, $U'$, $V'$ will cause picture saturation (luminance) errors.

To illustrate this effect consider a colour consisting of $R = x$, $G = 0$, $B = 0$. The true gamma corrected luminance will be

$$Y' = (0.3x)^{\gamma} \tag{A.14}$$

But from (A.11) it is actually

$$Y' = 0.3x^{\gamma} \tag{A.15}$$

If the colour difference signals, with their compensating errors, are not transmitted then the received luminance for (A.14) and (A.15) is

- True - 0.3$x$
- Actual - 0.09x

The reproduced luminance is thus less than it should be. This is known as the 'failure of constant luminance', and causes a loss of luminance detail in coloured picture areas. It also gives rise to an artefact known as Hanover Bars, which occurs in PAL-S receivers (no delay line) when phase errors are present.
A.2 The correction of phase errors

Before demodulation the chrominance is first added to a delayed version of itself. The delay used is called the PAL delay line, its delay period being nearly equal to one television line. The importance of the delay line is in removing chrominance phase errors that would produce hue errors in the receiver picture. How this is achieved is illustrated by $V_i \equiv V_i' + V_i''$ and $U_i \equiv U_i'' + U_i'''$.

Figure A.1.
The chrominance phasor can be represented as the resultant of the two chrominance signals $U$ and $V$. Due to the PAL line switch this resultant will lie in a different quadrant on alternate lines. When a phase error occurs the chrominance phasor is rotated in the direction of the phase error. In the delay line the $U$ component of the delayed signal is inverted before summation causing the phase errors for alternate lines to lie in opposite directions. Upon summation these opposing errors cancel to produce the correct chrominance phasor. Thus chrominance phase errors are eliminated.

This form of phase error correction will not work accurately for very large phase errors, or when the colour information varies between the two lines (i.e. when the chrominance has a vertical frequency component). Even under ideal conditions a small saturation loss will still occur, the gamma correction of the $U$ and $V$ components causing this error.

As the delay line is applied before demodulation it is impossible to invert the $V$ component without also inverting the $U$ component as well. To solve this problem the output of the delay line is both added and subtracted from the original signal, producing the two outputs $2U'$ and $\pm2V'$. The length of the delay must also be chosen so as to be an integer number of half wavelengths of the sub-carrier frequency $f_{sc}$. The delay that is closest to one line is

$$T_d = \frac{2835}{f_{sc}} = 0.9991t_{line} \quad (A.16)$$

The delay line thus corrects phase errors, and at the same time separates the $U$ and $V$ components. This separation, however, only works perfectly when there are no line to line changes in the image. This is perhaps unimportant, as synchronous demodulation is still used in most television decoders.
APPENDIX B
DISCRETE TIME FILTERS AND THE WINDOW DESIGN METHOD

A tapped delay line filter may be represented by
\[
e_j(t) = \sum_{k=0}^{N} c_k e(t-kT_d) + \sum_{k=0}^{N} d_k e_j(t-kT_d)
\]

where \(T_d\) is the block delay time. Figure B.1 shows a generalised diagram of this type of filter.

\[
H(f) = \frac{\sum_{k=0}^{M} c_k e^{-j2\pi k f d}}{1 + \sum_{k=0}^{M} d_k e^{-j2\pi k f d}}
\]

Tapped delay line filters can be divided into two types, IIR or recursive filters and non-recursive or FIR filters. A FIR filter differs from the IIR filter in having no feedback elements, as Figure B.2 shows.

---

Figure B.1 Block diagram of general digital filter (IIR filter).

Figure B.2 Block diagram of non-recursive digital filter (FIR filter).
For the FIR filter (B.1) and (B.2) reduce to

\[ e_f(t) = \sum_{k=0}^{N} c_k e(t - kT_d) \]  

\[ H(f) = \sum_{k=0}^{M} c_k e^{-j2\pi f k} \]  

The important differences between these two filter types are summarised below.

**IIR**
- Potentially unstable
- Non-linear phase response
- Normally designed using analog filter prototypes (e.g., chebychev)

**FIR**
- Always stable
- Linear phase characteristics
- Normally designed using Fourier domain window techniques
- Much higher order filters are needed to achieve a required cutoff rate.
- Has no analog counterpart

The design technique used to create the FIR filters is the *Fourier window technique*. This technique is illustrated by Figure B.3

![Diagram](image.png)

*Figure B.3 The window technique for designing digital FIR filters. The *s indicate sampling points, which become the filter coefficients after windowing.*

This technique uses the Inverse Discrete Fourier Transform (sampling points indicated by *) to find the impulse response - (b) for the desired frequency response - (a). This impulse response is then
windowed with an appropriate window (e.g. Hamming, Kaiser) to produce a set of filter tap coefficients - (c). The window function chosen depends upon the desired cut-off v. sidelobe level requirements. The actual response of the FIR filter - (d) may be found by performing a Fourier Transform upon the windowed impulse response. This response will be the convolution of the desired response with the Fourier Transform of the window function.

To approximate the desired characteristics closely FIR filters must often be of a relatively high order (large number of taps). This high filter order can sometimes cause problems with ringing - due to a large overshoot in the filter’s step response.
APPENDIX C
COMB FILTERS USED FOR K-PAL AND WC-PAL

Figure C.1 V3_COMB - Vertical luminance pre-filter used in K-PAL1 and WC-PAL1 coders.

Figure C.1 VT5_COMB - Vertical-temporal luminance pre-filter used in K-PAL2 and WC-PAL2 coders.

Figure C.1 KP3_COMB - Low order band-segregation comb filter used in K-PAL1 coder.

Figure C.1 KP7_COMB - High order band-segregation comb filter used in K-PAL2 coder.

Figure C.1 YP7_COMP - Luminance ripple compensating filter used in K-PAL2 decoder.
Figure C.1 ASS_C and ASS_Y - Assembly filters used in WC-PAL.
APPENDIX D

AN EMPIRICAL CORRELATION MEASUREMENT

As discussed in Chapter 6 any relationship between the author's new objective assessment results and subjective assessment results is definitely not linear. This means we cannot use normal linear correlation techniques. Considering the variability of the results obtained it is difficult to fit any function to the relationship (e.g. log, square root, ...) in order to produce a linear correlation. Instead the author has defined an empirical measure. This measure is based on the premise that a good correlation should mean that, as stated in Section 6.7.4,

1. For wanted signals an increase in subjective rating should be matched by an increase in objective rating.
2. For unwanted signals (artefacts) an increase in subjective rating should be matched by a decrease in objective rating (i.e. a decrease in the visually perceptible levels of the artefact).

If we define \( S(i) \) and \( S(j) \) as the subjective ratings for two different coding techniques and \( O(i) \) and \( O(j) \) as the equivalent objective ratings for these coding techniques then for the wanted signals we require that a positive difference between \( S(i) \) and \( S(j) \) be matched by a positive difference between \( O(i) \) and \( O(j) \). This positive correlation may be stated mathematically as:

\[
(S(i) - S(j))(O(i) - O(j)) > 0
\]

For unwanted signals we require a negative correlation, i.e.

\[
(S(i) - S(j))(O(i) - O(j)) < 0
\]

If a range of techniques is being compared we may use this approach to compare each result with the others. The total number of comparisons is thus \( \binom{N}{2} \). Obviously the more comparisons obey the requirements of equation (A.5) and (A.6) the stronger the overall correlation must be. Thus we may define a simple correlation index:

\[
C = \frac{\sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \text{sgn}(d)}{\binom{N}{2}}
\]

where \( d = (S(i) - S(j))(O(i) - O(j)) \)

where \( N \) is the number of techniques tested. This formula yields a number between -1 and 1. If \( C = 1 \) then all comparisons showed a positive correlation; if \( C = -1 \) then all comparisons showed a negative correlation.

Equation (A.7) is useful, but doesn’t take into account the significance of the differences measured. Section 6.7.4 stated a further requirement.

- The larger the difference in subjective rating is the larger the difference in objective rating should be.
In order to measure this we must therefore include a weighting function in Equation (A.7). The obvious choice for a weighting function is $d$ as defined in Equation (A.8). If both the subjective and objective differences are large then $d$ will be large. If only one of the subjective or objective differences are large then $d$ will be relatively small. If both differences are small then $d$ will be very small. To illustrate consider the following examples.

<table>
<thead>
<tr>
<th>$S(i) - S(j)$</th>
<th>$O(i) - O(j)$</th>
<th>$d$</th>
<th>$\sqrt{d}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>100</td>
<td>400</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

Table D.1 Example of subjective and objective rating differences and the resulting values of $d$.

Thus when both the subjective and objective differences are large then $d$ is the largest. When either of the differences is smaller (and therefore less significant) then $d$ is smaller. The difference between the largest and smallest values of $d$ is, however, quite large. For this reason the square root of $d$ is used instead.

To ensure the resulting correlation values of $C$ remain between -1 and 1 the result must be normalised to the sum of the new weighting function.

Consequently the final empirical correlation measure is defined as:

$$C = \frac{\sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \text{sgn}(d)\sqrt{d}}{\sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \sqrt{d}}$$  \hspace{1cm} (D.9)$$

where $d = (S(i) - S(j))(O(i) - O(j))$  \hspace{1cm} (D.10)
APPENDIX E

TVPROC - A THREE-DIMENSIONAL IMAGE CODING SIMULATION ENVIRONMENT

TVPROC is a signal/image processing system that can be used for the testing of image processing algorithms. It is a general system, but one with special features particularly suited for the simulation of television and video processing techniques.

TVPROC is different from many similar systems (eg VIPS, MATLAB, IMPROC, Khoros) in the fact that no interactive user interface is provided. This means that most image processing commands cannot be executed interactively. Consequently TVPROC is far better suited to algorithm assessment, rather than algorithm development, with assessment being the task TVPROC is particularly designed for. The system enables the user to implement an algorithm, then make use of a large variety of tools to test and analyse this algorithm.

As TVPROC does not provide interactive processing there is no main program or package that must be run in order to use the system. The basic system consists of several C function libraries containing a variety of simple image processing functions, together with a collection of tools and utilities for use in creating and running simulations, generating test images, and viewing and analysing results. Extensive use is made of existing SUN packages to provide extra facilities for TVPROC (for example an image viewer by John Bradly called 'xv' is used for image display).

For some of the most commonly used operations (eg image display) a simple windows user interface is provided through the medium of the OpenWindows File Manager Icon bindings and Custom Command Menu.

This chapter starts with a discussion of the motivation for the development of TVPROC. It continues with an overview of the system which includes its historical development, basic structure and operation, plus a description of its important features, with examples.

E.1 Construction vs. simulation

Television and video have enormous bandwidths, several orders of magnitude greater than audio signals. It is largely this factor that makes the implementation of any video coding technique in hardware both complex and expensive. The hardware used will also likely consist of dedicated circuitry, with any flexibility of operation being limited in scope.

For these reasons computer simulation of the television coding techniques being studied appears as a sensible alternative to a real-time hardware implementation.

E.1.1 Computer simulation allows for

- greater flexibility
- reduced complexity - synchronisation problems no longer exists
- and reduced overheads - both in hardware costs (assuming computer facilities already exist) and in the time taken to implement the technique (i.e. build and test the hardware).
Computer simulation has other advantages, allowing the use of computer generated test images, and computer analysis. These are more difficult to achieve when using hardware.

Simulation does have its disadvantages however, the most obvious being the loss of real-time processing. This is an important loss, considering that a television picture is intended to be viewed at a particular picture rate. There is also a limit on the amount of data that can practically be processed, due to computer storage limitations, and processing requirements. With television’s very high bandwidth (CCIR Rec. 601-1 sets a requirement of 216Mbits/s) this is a very important constraint.

For simulation to provide a useful alternative to hardware implementation ways must be found to work around these constraints.

In the development of TVPROC solutions were found to some of these problems by using real-time video capture and display on the computer, processing reduced size images (representing only a portion of the screen area), and using reduced precision data formats.

### E.2 The development of TVPROC

The development of TVPROC arose from the need for a system capable of processing colour video (television) sequences, to be used in the simulation of enhanced PAL techniques. At the time there was no system available that was capable of handling three dimensional multiple-component (colour) data. VIPS, IMPROC and MATLAB, the packages available, all used two dimensional data structures. The large bandwidth of television image sequences meant there was also a need for a system capable of handling large datasets quickly and efficiently. The available packages were far better suited to complex analysis of small data sets, rather than the simpler processing of large data sets.

Consequently TVPROC was created to meet these needs, with a special three dimensional, multiple-component data structure, a multiple precision data format, and the use of compiled simulations to ensure quick processing. As the author’s emphasis was on functionality rather than user friendliness the system is command line based, and contains separately executable commands to help in the creation and running of simulations, and the creation of test images.

To reduce development time and enhance the capabilities of the system many facilities were provided by making use of other packages, rather than creating built-in facilities. Three dimensional signal analysis and spectrum display however, was not available within any existing packages, and so facilities for this were built in, and have become an integral part of TVPROC (See Appendix F).

With the usefulness of the system becoming evident refinement was continued, with the addition of user configurability, user extensible function libraries and some graphical interfaces.

The result, named TVPROC, may be used as a general purpose image processing system, but has special features which make it particularly suited for the study of television or video coding technique. The features which distinguish TVPROC are:

- three dimensional image processing
- support for colour (multiple component) images
- speed and memory (disk storage) efficiency
- graphical entry of algorithms
- three dimensional spectral analysis and spectrum display

### E.2.1 Comparison with other systems

Table 3.1 shows a comparison of other image processing system against TVPROC on some of the most critical points. It can be seen that TVPROC is, as yet, limited as a general purpose system. Also, the newly developed KHOROS system appears to offer many of the capabilities of TVPROC in a far more self-consistent package. It is doubtful that TVPROC would have been developed if
KHOROS had been available when work was begun. However, the KHOROS method of running simulations - calling separately compiled commands and passing data by use of temporary disk files means the system has a definite speed disadvantage compared to TVPROC. Its facilities for three dimensional analysis and display are also not as highly developed as those for TVPROC.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>TVPROC</th>
<th>MATLAB</th>
<th>IMPROC</th>
<th>KHOROS</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Interface</td>
<td>graphical interface compiled</td>
<td>command line interpreted</td>
<td>command line interpreted</td>
<td>graphical interface interpreted</td>
</tr>
<tr>
<td>Function libraries</td>
<td>limited</td>
<td>extensive</td>
<td>medium</td>
<td>extensive</td>
</tr>
<tr>
<td>Data structures</td>
<td>3-D real/complex multiple precision (max 32 bit)</td>
<td>2-D real/complex double precision (64 bit)</td>
<td>2-D real/complex double precision (64 bit)</td>
<td>3-D real/complex multiple precision (max 64 bit)</td>
</tr>
<tr>
<td>Spectral Analysis</td>
<td>extensive tools</td>
<td>limited</td>
<td>limited</td>
<td>some tools</td>
</tr>
<tr>
<td>Support for colour</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Processing speed</td>
<td>high</td>
<td>medium / low</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>Memory/disk efficiency</td>
<td>high</td>
<td>medium</td>
<td>medium</td>
<td>low</td>
</tr>
</tbody>
</table>

Table E.1 Comparison of TVPROC image processing system with similar existing packages.

E.3 The structure of TVPROC

TVPROC is not a standalone self-contained image processing package. There is no main program or interface which must be run and through which all use of the system is made. TVPROC is more a collection of function libraries, image utilities and simulation tools which operate on a common data format. This collection links together a number of different packages and facilities in order to provide all the capabilities required for a complete image processing system. This approach has its advantages and disadvantages. Firstly there is no one self-consistent user interface through which all facilities are accessed. All of TVPROC's facilities can be accessed via a command line interface, with a limited graphical interface (File Manager Icon Bindings and Custom Command Menu) being provided for some operations. However, the many other packages made use of by TVPROC all have their own user interfaces, which are in some cases quite different from each other. This means a steep learning curve for the completely novice user. If all packages and facilities are new the use of TVPROC may at first appear messy and complicated. For the experienced user, however, the approach offers power and flexibility. Instead of being restricted to limited inbuilt facilities for, as an example, the animation of images, the user can configure TVPROC to use the package of their choice. One of the problems with many systems is that they are good for some operations, but limited in other respects. With TVPROC this problem is circumvented by making use of the best package for the job. As an example, a full set of MATLAB interface functions are provided, to allow users to take advantage of MATLAB's excellent one and two dimensional graphing and signal analysis capabilities. Thus far all design of filters for use in TVPROC simulations has been carried out in MATLAB, using MATLAB's signal library. Specialised functions are provided to allow the filters so designed to be then saved as TVPROC data files, ready for use in simulation.

In this way TVPROC is sometimes useful simply because it can provide an interface path between two existing applications. For example, using TVPROC it is possible to take the contents of a MATLAB matrix and display them as an image using the X Viewer display package. This facility does now exist in the latest version of MATLAB, however the display facilities are very primitive compared to X Viewer's powerful capabilities.
The relationship between the many different sections that together form TVPROC can be represented as a hierarchically layered structure in Figure E.1, with the user choosing their access to the system through any one of these layers.

![Hierarchically layered structure of TVPROC interface](image)

When discussing the use of TVPROC, however, the system is better represented by a functional structure, basically a data-flow diagram, one that illustrates the relationship between different parts of the system when performing specific operations.

### E.4 The Hierarchical Structure

To some degree the hierarchical structure shown in Figure E.1 reflects the development of TVPROC, with the data structure and basic image processing functions at the centre, and the more recently added command line and graphical interface in the outer layers. The structure also reflects the user's access to TVPROC facilities. At the most basic level TVPROC can be used via the outermost layer only. For more sophisticated use, and greater flexibility, the commands and C functions of the inner layers can be used directly. User access should not, however, go beyond layer 2 - the software data structure and disk data format should remain hidden from the user.

#### E.4.1 Layer 1 - The image data structures

All TVPROC data and images are stored using a special matrix and disk format. This format is capable of storing three dimensional real and complex data with 32 bit precision in memory and variable precision (8,16,32 bits per component) on disk. A maximum precision of 32 bits is used (i.e. no double precision - 64 bits) in order to reduce memory requirements. Video images have a limited dynamic range (8-bits is normally regarded as sufficient) and hardware implementation of video processing algorithms is normally restricted in precision. Consequently 32 bit precision was regarded as being entirely adequate for most video processing. A variable precision disk format also allows storage requirements to be minimised. All image sources and image displays use a maximum of 8-bits per component and thus the 8-bit disk format is used predominantly, although 32-bit precision is required for things such as spectral analysis.

Associated with each image (or data matrix) is a set of attributes and parameters, which provide important information about the stored data.
The software data structure

Within all TVPROC C functions data is stored on the C TMatrix structure, of the form:

```c
struct TMatrix {
    int ncols, nrows, nslices;
    int complexflag;
    int saveformat;
    float *dataP;
    char *dataname;
    char *comments;
} TMatrix;
```

The parts of this structure are explained below.

- **ncols, nrows, nslices**: The dimensions of the data. If the data is only two dimensional then nslices = 1.
- **complexflag**: Indicates real or complex data.
- **saveformat**: A flag that indicates which of TVPROC's three variable-precision data formats is used to store the matrix when on disk. This flag is set either by the matrix loading function, to indicate what format the loaded data was in, or by the user/software, before saving the data, to indicate the desired save format.
- **dataP**: The data array pointer. All data is stored as a one dimensional array. Multi-dimensional data is stored row by row, slice by slice. Complex data is stored as real and complex pairs, with the array being twice as long.
- **dataname**: The filename associated with this data. When saving, the data will be saved in a file with this name. When loading, the dataname is the name of the file from which the data was loaded. The dataname normally conforms to a special syntax, which encodes limited information about the data (see later).
- **comments**: Any extra information can be placed in this string. TVPROC makes special use of this part of the structure as a store for parameter definitions of the form

  parameter = value

For example, the sampling frequency used for the data might be stored

  sampling_frequency = 13.5e6

Any number of comments/definitions can be placed in the comments, each separated by a ' \n' (newline) character.

The disk data format

When the data in the TVPROC matrix is saved it is placed in a TVPROC data file. The format for this file is slightly different depending upon which precision is chosen, but basically consists of header information followed by the data, stored in binary (not ascii) form. For the 8 and 16 bit formats the data must be converted from the 32 bit precision used in the TMatrix structure. In order to preserve accuracy this conversion is done by scaling the matrix data to the 8 or 16 bit range, and then placing an offset and multiplier within the file header. These can then be used to recover the data upon reloading.

Data labelling and identification

A special naming system is used to distinguish different data files and the type of data that they contain. Each data file or TVPROC matrix has a dataname, which becomes the name of the data file when stored on disk. This dataname is divided into a number of fields, separated by the underscore character '_', and containing the attributes of the data file. The full format for the dataname is

  name_simname_stage_pictype#datatype_component.dat
Not all fields must be filled at once, with an order of preference being used. Thus if an image does not have certain attributes the relevant fields will be omitted. The first field is the identifying name of the image or data. If all other fields are omitted the most basic name for a data file is of the form name.dat e.g.

asterix.dat

**Storing multiple component data**

Multiple component data (eg a colour image) is stored as separate data files, one for each component. The files will have identical names, except for the component field, which will denote the component name of the data. For example a colour image called asterix could be stored in YUV format as:

asterix_y.dat
asterix_u.dat
asterix_v.dat

**Storing data/image sequences**

TVPROC is capable of handling three dimensional images. Most other packages, however, are not. Consequently for the purposes of image creation/capture or image display/animation data must be stored as a sequence of two dimensional data files on disk. A number field is used to indicate the position of a data file in such a sequence. For example a 4 frame animated sequence of asterix might be stored as

asterix_1_r.dat asterix_1_g.dat asterix_1_b.dat
asterix_2_r.dat asterix_2_g.dat asterix_2_b.dat
asterix_3_r.dat asterix_3_g.dat asterix_3_b.dat
asterix_4_r.dat asterix_4_g.dat asterix_4_b.dat

Such sequences could be created by use of a package such as MATLAB. Using the TVPROC combine command these data files could then be combined into three 3-dimensional data files, ready for use by TVPROC. The result would be

asterix_1_r.dat asterix_1_g.dat asterix_1_b.dat

Note that the sequence number still remains. Most TVPROC data files will still contain a sequence number, even if they are three-dimensional. If they are not part of a sequence the number will always be 1.

**Naming a processed image**

After processing of data by a simulation the data file will be given a simulation name (the simname field will be filled). The name of consequent simulations will also be appended to this field. For example, after being processed by a PAL encoder (pal-E) and decoder (pal-D) an image might be stored as

asterix_pal-E:pal-D_1_y.dat
asterix_pal-E:pal-D_1_u.dat
asterix_pal-E:pal-D_1_v.dat

The other fields in the dataname indicate other information about the image data type and how the data is arranged. For example, whether the data is stored as frames or interlaced fields (pictype = f), and whether the data file contains magnitude or phase information (pictype = m or p).

Many image processing functions will look at the matrix attributes and modify their behaviour accordingly. (matrix_yuv will issue a warning and perform no operation if the data is already in Y,U,V format). Similarly these functions also alter the attributes to reflect changes made. For example, the fft function will change the attributes to indicate that the result matrix contains spectral data.
E.4.2 As a summary this complex naming system has several advantages

- Allows unique naming and storage of large sets of results.
- Provides the user with information on data contents simply by reading filenames.
- Allows software to operate on data in an intelligent manner ie modify their actions according to the data passed.

See the TVPROC Reference Manual for more information on the TVPROC data structures, formats and naming system.

E.4.3 Layer 2 - The matrix functions

Under normal circumstances all access to image data is made through a library of C matrix functions. This library provides functions for the creation and destruction of image matrices, their loading and saving on disk, and the setting and inquiry of matrix parameters and attributes.

Example E.1 Loading, saving and destroying a matrix.

The data in the TVPROC file testfile_y.dat is loaded, then saved again with a new dataname (and filename) testfile_v.dat.

```c
strcpy(Matrix.dataname, "testfile_y.dat")
load_matrix(&Matrix)
change_attr(&Matrix, S_COMPONENT , "v")
save_matrix(&Matrix)
destroy_tvmatrix(&Matrix)
```

This process is thus independent of the matrix contents.

Also provided is a variety of simple matrix manipulation functions. These can perform operations such as the initialisation, copying, limiting, adding, and multiplication of matrices. All functions take separate input and output arguments, with matrices being passed by reference, denoted in C by the use of the & character.

Example E.2 Format and usage of TVPROC matrix functions

The operation

\[ Y = X + 0.5 \]

is achieved by use of the scalarop function.

```c
scalarop (&Y, &X, '+', 0.5);
```

If the same matrix is specified as both the input and output matrix e.g.

```c
scalarop (&Y, &Y, '+', 0.5)
```

then the matrices contents are replaced directly with the result. This point appears trivial but is very important in minimising memory usage and memory paging when handling large datasets.

E.4.4 Layer 3 - Image processing libraries

The next layer contains several image processing libraries. The functions they contain are also written in C, and use a similar syntax to the matrix functions of layer 2. However, all (or most) access by these functions to the matrix contents is made via the matrix functions of the second layer. The user may add their own image processing functions accessing the TVPROC matrix in a similar manner.
The functions already in the TVPROC library include filters, Fourier transform, trigonometric functions as well as a range of functions designed particularly for television processing (e.g., RGB to YUV conversion, interlaced to progressive scan conversion and modulation).

E.4.5 Layer 4 - Simulation tools

All TVPROC commands and tools lie in the fourth layer. These are all compiled C programs which make use of functions from the Matrix and image processing libraries. The commands can be divided into several groups:

- Image creation
- Image display
- Image file manipulation and format conversion
- Simulation creation
- Simulation running

The use of these commands will be dealt with in detail in section E.5. (For a full list of all TVPROC commands refer to the TVPROC Reference Manual).

All commands use a standard command line syntax of the form:

```
  command [ -option1 <arguments> ] [ -option2 ] free_argument1 free_argument2 ... .
```

This syntax allows for any number of options, with or without arguments, and any number of free arguments (so called because they are not attached to any options). As an example the image2spectrum command has the following command line usage:

```
  image2spectrum [ -name ] [ -mag ] [ -phase ] [ -limit <max min> ] [ -quadshift ] [ -ralias ] [ -norm <component> ] [ -db ] imagefile
```

The use of this command is illustrated by Example E.3.

<table>
<thead>
<tr>
<th>Example E.3 Usage of TVPROC commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input file: testimage_1-y.dat</td>
</tr>
<tr>
<td>Command: image2spectrum -quadshift -mag -norm y testimage_1-y.dat</td>
</tr>
<tr>
<td>Output file: testimage:(y)_{1ms}y.dat</td>
</tr>
</tbody>
</table>

The resulting output file thus contains the quadrants shifted magnitude spectrum of the input file, (the number of dimensions of the spectrum will depend upon the dimensions of the input file). The output file will also be normalised to the defined limits of the y component. Note that all these changes are reflected in the output filename.

E.4.6 Layer 5 - User configurable shell scripts

In order to provide for a high degree of flexibility many TVPROC commands have a large number of options. To simplify the use of these commands and allow new users to personalise them a collection of Unix shell script files is provided for each user. Each command has its own script file, which can be used in place of the command, and will automatically provide option defaults. For example the image_spectrum script provides the options shown in Example E.3, and automatically chooses the correct normalising component.

Some of these script files can be quite sophisticated, providing different defaults for different types of files (as the image_spectrum script does selecting a correct normalising component). Another example is the data_image script, which will convert a data file to either an animated image sequence (command data2anim), or a three-dimensional PHIGSdraw contour metafile (using find_contour) depending upon whether the data file contains image or spectral data.
In this manner the user script files simplify the use of the TVPROC commands, provide an extra level of software intelligence, and also allow user customisability.

E.4.7 Layer 6 - The graphical interface

The final layer is the graphical interface. This can be divided into two parts:

- Graphical command interface
- Graphical simulation interface

**Graphical Command Interface**

Time has not permitted the development of a sophisticated user interface for TVPROC. Instead, as an intermediary measure use is made of existing interfaces - in particular the *OpenWindows File Manager*. Using the File Manager icon file bindings and custom command menu a windows interface is provided for some of the more commonly used TVPROC commands. This interface appears as shown in Figure E.2

![Figure E.2 Example of TVPROC user interface using OpenWindows.](Figure E.2 Example of TVPROC user interface using OpenWindows.)

**File Manager Icon Bindings**

The *OpenWindows File Manager* displays files individually as icons, with the icon used being determined by filename syntax. Thus, as seen in Figure E.2 files with the extension `.sim` are displayed with the `SIM` icon, and files with the `.net` extension are displayed as `~`. The icon bindings link each of these icon types to a particular command that is activated upon a mouse double-click on the icon. Example E.4 illustrates the use of this facility to run the commands.
Example E.4 Use of icons to create a simulation.

Double-clicking on the *pal_d.net* icon will cause the TVPROC command `create_function` to be run on this file, thus converting the netlist file to a C simulation function.

```
netlist file  TVPROC command script  simulation function
pal_d.net  create_function  pal_d.c
```

*Figure E.3 Conversion of net file to C file.*

Similarly double-clicking on the *pal-D.sim* icon will cause TVPROC to create a simulation, according to the simulation definition information within this file, resulting in the appearance of the executable file *simtv_pal-D*.

```
simulation definition  TVPROC command script  executable simulation
pal-D.sim  create_sim  simtv_pal-D
```

*Figure E.4 Creation of a simulation from a definition file.*

Example E.5 Use of icons to convert and display an image.

Viewing an image may be done by double-clicking on any data file, which will then be converted to a viewable image format. Double-clicking on this file will then cause the image to be displayed on screen.

```
TVPROC datafile  TVPROC command script  viewable image format

testpic.dat  data_image  testpic.tif

TVPROC command script  Displayed on screen
testpic.dat  display_image
```

*Figure E.5 Displaying the contents of a TVPROC data file.*

**Custom Menu**

The custom menu contains commands which can be applied to any files within the File Manager window. A file or group of files is first selected with the mouse pointer then the appropriate command chosen from the menu. This allows the same command to be applied to any number of data files at once.

Both the File Manager icon bindings and Custom menu make use of the user's shell script files instead of calling the commands directly. This allows the user to customise the actions performed via the File Manager interfaces.

**Graphical Simulation Creation**

Once again instead of developing a special graphical interface an existing package is used. In this case the package is the PROTEL schematic editor and post-processor. This package is normally used for creating electrical schematics, which can then be converted to a netlist (a list of components
and all pin connections between components). This resulting netlist is normally used for the creation of a printed circuit board layout. Despite this application appearing quite different PROTEL was found to be an ideal tool for creating algorithm flow diagrams. The netlist from these simulation schematics, could then be converted by TVPROC into a C simulation function (see example E.4).

In order for PROTEL to be used in this manner a special image processing ‘component’ library had to be added (see the TVPROC Reference Manual for a list of components). The components in this library each represent an image processing function from one of TVPROC (or the user’s) libraries. Connections between components in the simulation schematic indicate the flow of image data between functions. Figure E.6 shows an example of the type of simulation schematic that can be created (see Figure E.12 for another example).

![Figure E.6 Simulation schematic for a PAL decoder.](image)

The use of PROTEL in this manner eliminates the need for any C programming knowledge in creating a simulation using TVPROC. The schematic thus created also provides the user with an excellent diagram of their algorithm (one that is far more visually recognisable than those produced by KHOROS). All the block diagrams presented in this thesis, of Enhanced PAL techniques, are the actual simulation schematics used in TVPROC for studying the technique.

The disadvantages of the method are the lack of (as yet) any conditional branching or control functions, and the lack of loops and feedback. Any algorithm requiring iterative loops, or feedback paths cannot be entered as a simulation schematic, although it can still be implemented as a C function using the image processing libraries. These facilities will likely be added in the future.

E.4.8 Packages Used By TVPROC

TVPROC makes use of a number of existing packages on the SUN system, in order to provide facilities that do not exist within TVPROC.

**XViewer:** Used to display colour or black & white images

**Movietool:** Displays colour or black & white image sequences.

**PHIGSdraw:** Displays three dimensional spectral contours (was created by the author, see Appendix F).

**PPM Utilities:** Used for image format conversions.

**MATLAB:** A range of MATLAB mex functions are provided by TVPROC to allow direct loading, saving and manipulation of TVPROC data files from within MATLAB. MATLAB is a good tool for filter design and 1 or 2-D graphing.

**PROTEL:** Creates simulation schematics and generates netlists that can be used by TVPROC to create simulation functions.

**OpenWindows File Manager:** Used to provide a graphical interface for TVPROC, via the File Manager Custom Command menu and icon bindings.

It should be noted that TVPROC is not tied directly to these packages. Instead interfaces are provided that can be tied to particular packages for particular functions. The user is free to make the choice of what package is used by re-configuring TVPROC.

### E.5 The functional Structure

The previous section has explained the hierarchical structure of TVPROC. This section explains the relationship between the different parts of TVPROC and how they are used for creating a simulation,
creating test images and viewing results. With the use of examples this section should provide an appreciation of TVPROC's capabilities.

Figure E.7 sums up the operation of TVPROC.

The user creates a simulation, then tests the simulation on a variety of test images (or non-image data) producing results that can then be viewed or analysed.

A simulation cannot be altered without recreating (re-compiling) it (although this is relatively simple task). However, a range of parameters can be stored independently on disk (e.g., filter coefficients) and can be altered to change the way the simulation runs without requiring re-compilation. Figure E.8 shows a far more detailed functional block diagram of TVPROC with those sections not directly part of TVPROC being placed outside the dotted line.

Each part of this diagram will now be discussed in greater detail, under the headings of test image generation, creating and running simulations, and displaying results.

E.5.1 Test Image Generation

TVPROC provides commands for the creation of several types of test image. These include the spatial and temporal (parabolic) zones, colour bars and special spectrally constrained images (for use in spectral analysis - see Chapter 4). These commands create TVPROC matrix data files automatically. However, test images taken from sources external to TVPROC will have to be
converted to the right format. In TVPROC such conversions are performed via the Portable PixMap (PPM) image format, using the PPM utilities library. This process is illustrated by Figure E.9, and carried out by the `image2data` command.

![Diagram](image)

**Figure E.9 Converting other image formats to TVPROC.**

**Example E.6 Conversion of an image to TVPROC format.**

If an image exists in the GIF image format it may be converted to TVPROC format.

<table>
<thead>
<tr>
<th>Input file:</th>
<th>asterix.gif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command:</td>
<td><code>image2data -pipe giftoppm -yuv -format 1 asterix.gif</code></td>
</tr>
<tr>
<td>Output:</td>
<td><code>asterix_l-y.dat</code></td>
</tr>
<tr>
<td></td>
<td><code>asterix_l_u.dat</code></td>
</tr>
<tr>
<td></td>
<td><code>asterix_l_v.dat</code></td>
</tr>
</tbody>
</table>

The resulting output is the 3 component files, in YUV format, of the colour image `asterix`, in TVPROC 8-bit disk format.

Image sequences can be input to TVPROC as separately numbered images, one for each frame. Upon conversion they will be placed into one three dimensional TVPROC image file (or one for each component).

**E.5.2 Creating and Running a Simulation**

Creating a simulation can be divided into two steps

(i) Creating simulation functions

(ii) Compiling these functions to create a simulation.

**Creating a Simulation Function**

A simulation function can be created directly by actually writing a C function using the TVPROC image processing libraries, or it can be generated from a simulation schematic. This second method is illustrated in Figure E.10.
The simulation schematic, such as that shown in Figure E.6 for the PAL decoder, is created by the placement and linking of image processing 'components' within TVPROC's special component library, using the PROTEL schematic editor. Each of these components is linked via a component definitions file to one of TVPROC's C image processing functions.

Example E.7 Use of a graphical component to represent a TVPROC function.

Figure E.11 shows the modulate component from the graphical component library.

```
MODULATE
```

**Figure E.11** The modulate component from the TVPROC schematic library

This component is linked to the C tvmodulate function in TVPROC's image processing library by the definition

```
MODULATE = tvmodulate($2, $1, freq, phase).
```

The input (1) and output (2) pins of the component are linked to the image matrices $1$ and $2$ respectively, with the non-matrix arguments of freq and phase being provided by the component 'value'.

Pre-defined constants may also be used in the schematic, and are defined in a schematic definitions file.
Example E.8 Use of schematic definitions to define constants.

\[\text{tline} = 6.4 \times 10^{-5}\]
\[fsc = 4.43361875 \times 10^6\]

Figure E.12 PAL chrominance decoder (chrom_d) showing use of definitions tline and fsc

The filter names are defined slightly differently. The schematic filter ‘value’ (c-lpf) is combined with the function name (chrom_d) to produce a filename `chrom_d_c-lpf`. This is the name of a TVPROC data file which will be expected to contain the filter coefficients for the chrominance low-pass.

To convert the simulation schematic to a simulation function it must first be converted to a netlist (this process also generates an error report showing unconnected or incorrectly connected components, see Figure E.10). The netlist is then converted to a C function that uses the image processing functions in the TVPROC libraries. The resulting function is referred to as a *simulation function*.

Just as the image processing libraries are user extendible so is the schematic component library, allowing the user to create new components to represent new functions.

A useful feature is that a simulation may itself be used as a component in the building of a larger simulation. In this case the simulation is referred to as a *simulation subfunction*. Thus the chrominance decoder of Figure E.12 will become a component in the PAL decoder of Figure E.6. This new component representing the chrominance decoder is shown in Figure E.13.

Figure E.13 Schematic component representing a TVPROC subfunction.

Any existing function may, of course, be redefined as a subfunction, for use in another schematic.

This facility allows large or complex algorithms to be created from smaller building blocks. It also allows the user to create new image processing functions graphically, instead of writing a C program.

**Creating a Simulation**

A simulation is created by compiling any number of simulation functions and linking them with a simulation run shell that provides a user interface, error checking, and loads the data to be processed. This process is illustrated by Figure E.14.
For each simulation to be created the user must provide a simulation definition file with information about the simulation. Example E.9 shows the contents of one such file.

Example E.9 The TVPROC simulation definitions file.

A definition file for a simulation of the PAL system might appear as below

```
sampling_frequency = 17734475
simulation_name = nzpal
components = (y, u, v)
fuction = pal_e
function = pal_d
picture_min_in = 1
picture_min_out = 4
```

The resulting simulation will be created using the `pal_e.c` and `pal_d.c` functions and will be called `simtv_nzpal`. It will expect three input image files in YUV format and will produce an output with a minimum of 4 frames. (This allows temporal effects to be seen). The input image, however, is allowed to be as little as 1 frame. (In this case the image would be repeated temporally, so as to simulate a stationary picture, and allowing temporal effects on stationary pictures to be analysed).

**Running a Simulation**

The result of the above process is an executable program and must be run with arguments that provide the name of the images/data to be processed. The usage syntax for any simulation is the same, and is shown below for the simulation created for the simulation definitions file in Example E.9.

```
Usage: simtv_nzpal [ -s ] [ -v ] [ -out <path> ] [ -x ] [ -spectrum <opt1 opt2 opt3 opt4 opt5> ] [ -empty <components> ] [ -outformat <outformat> ] imagefile ...
```

More information on the usage of this compiled simulation and any other TVPROC function can be obtained by using the `tvhelp` facility (see section E.6).
Example E.10 Running a compiled simulation

The PAL simulation created according to the definition file shown in Example E.9 is run below on a test image called spatialzone. The verbose option is selected.

```
command_line> simtv_nzpal spatialzone:Y -v

Loading test image 'spatialzone:Y_i_y.dat'

***** SIMULATION FUNCTION 'pal_e' *****

Block L1 - FIR_FILTER (c-lpf)
Block L2 - FIR_FILTER (c-lpf)
Block N1 - FIR_FILTER (y-notch)
Block M1 - MODULATE (fsc(0))
Block M2 - MODULATE (fsc(90))
Block P1 - PAL-SWITCH
Block A1 - ADD
Block A2 - ADD
Block L3 - FIR_FILTER (pal-lpf)

***** SIMULATION FUNCTION 'pal_d' *****

Block B1 - FIR_FILTER (c-bpf)
Block N1 - FIR_FILTER (y-notch)
Block D1 - DELAY (283.5/fsc)
Linear delay is 0 fields, 1 lines, -1 cols
Block I2 - INVERT
Block A1 - ADD
Block S1 - SUBTRACT
Block M1 - MODULATE (fsc(90))
Block M2 - MODULATE (fsc(0))
Block L1 - FIR_FILTER (c-lpf)
Block P1 - PAL-SWITCH
Block L2 - FIR_FILTER (c-lpf)

Saving processed YUV / RGB components

********** FINISHED **********
```

Simulations will normally only produce results for those signals that are indicated as outputs in the schematic. However, if the stage save option -s is used results will be produced for any part of the simulation that is specially labelled in the simulation schematic. Using this option intermediary results from any stage of the algorithm can be viewed.

E.5.3 Displaying Results

The results produced by the simulation are placed on disk as a number of TVPROC data files. Depending upon the test image and the simulation parameters used these results may be single images, image sequences or spectrums.

Displaying Images

TVPROC does not have inbuilt facilities for image display. Instead TVPROC data must be converted into the image formats required by other image display packages. This process is basically the reverse of Figure E.9. In the default configuration for TVPROC an X windows shareware package called ‘xv’ is used for single image display. TVPROC uses this package to display images in colour, as separate components, and in progressive or interlaced scan format.

If a TVPROC data file is three dimensional then it will be converted to an image sequence.
Displaying an Image Sequence

An image sequence is displayed in TVPROC at present using the SUNView Movietool package. Figure E.15 shows this package as it is used to display the results from the PAL simulation.

![Animation tool used to display image sequences from TVPROC](image)

Once again the sequence may be displayed in colour, or as separate component images. A TVPROC command `data2anim` converts TVPROC data files to any special formats required by the animation package used. For display of television image sequences this conversion is done so as to simulate the effects of display on an interlaced television screen.

Displaying Spectrums

Spectrums are displayed using the three dimensional viewing package PHIGSdraw. Before being displayed they must first be converted to PHIGSdraw graphics metafiles, using the `find-contour` command. The spectrums are then displayed as a series of two or three dimensional contours, as shown in many places in Chapters 3, 4 and 6. Appendix F explains this process in greater detail.

E.6 Getting help

TVPROC has inbuilt help for all commands and functions. For command line utilities and programs the incorrect use of the command, or execution with no arguments will cause the provision of usage information. For example, typing `combine` with no options will produce the on screen output of

```
Usage: combine [-n <nfiles>] [-o <outfile>] [-delete] imagefile
```

For fuller information on commands, and help information on functions in any of the TVPROC libraries the `tvhelp` command is used. This basically just prints out the contents of the first comment in the C source code for the command or function. For example:

```
command_line> tvhelp combine
```

```
combine.c:

NAME
combine - Combines a number of 2D TVPROC data files into 1 3D data file.

SYNOPSIS
combine [-n <nfiles>] [-o <outfile>] [-delete] imagefile

DESCRIPTION
```
The 2D data files are expected on disk in the format

\[ \text{imagename} \_\# \_\text{component}.dat \]

**OUTPUT**

One 3D data file of the form

\[ \text{imagename} \_1 \_\text{component}.dat \]

**RETURN VALUES**

- Succeed 0
- Failure 1

### E.7 Error handling

All TVPROC functions and programs perform argument checking and error handling. If an error is detected then the function will issue an error message and quit. This function will then pass an error value to its calling function, which will also issue an error message, and so on. Using this error handling system the occurrence of an error will cause a hierarchical list of error messages to be displayed, as the error is passed up through the calling functions.

The error messages are all of the form:

\[ \text{ERROR}\{\text{function\_name}\} : \text{Error message} \]

**Example E.11** The production of TVPROC error messages. illustrates the way error messages occur.

```
Example E.11 The production of TVPROC error messages.
If, for a simulation of PAL one of the filter data files does not exist the on screen messages and errors would appear as:

command_line> simtv_pal asterix

| Loading test image 'asterix_1_y.dat'
| ***** SIMULATION FUNCTION 'pal_e' *****
| ERROR(get_matrixinfo): Error opening data file
| '/home/users/image/taylork/tvproc/sim_def/pal_e-y-notch.dat'
| ERROR(load_data): File 'pal_e-y-notch.dat' does not exist, or is not
| TVPROC data format
| ERROR(tv_fir_filter): Error loading Fir filter matrix
| ERROR(simtv_pal): An error occurred in the running of function
| pal_e
```

This approach causes error messages to sometimes appear a little convoluted, but is useful in indicating exactly when, why and in which function an error occurred. For a system that is still undergoing development this is a very useful feature.
E.8 Summary

TVPROC has proved itself an excellent tool in the simulation and study of enhanced television techniques. Its suitability for other applications has also been demonstrated. The system has been recently used in the simulation of non-linear noise filters for video, the testing of an MPEG (Motion adaptive video coding) algorithm, and the deblurring of a newspaper photograph taken of the alleged sighting of a New Zealand Moa! The ability to handle large datasets and colour images was particularly important for the latter two. The digitised image of the Moa photograph was over 6 million pixels per component (3000 x 2000).

One disadvantage of TVPROC is its present reliance upon a large number of external packages for the provision of many facilities. Although much of the power of the system has come from this linking together of the best available packages it does pose problems for the portability of the system.

Further work is needed on perhaps replacing these packages with TVPROC’s own facilities, as well as developing a better user interface, and extending the schematic and image processing libraries.
APPENDIX F

PHIGSDRAW - A THREE-DIMENSIONAL VIEWING PACKAGE

F.1 Overview

![Diagram of PHIGSDraw with labeled panels: view panel, command panel, status panel, mouse panel.]

*Figure F.1 PHIGSDraw being used to display a spatio-temporal spectrum.*

As the package stands *PHIGSDraw* is an interactive three dimensional viewing package, written in C using the SUN PHIGS graphics libraries and the Sunview windows interface. The important features of the package are:

- The display of three dimensional objects
- Use of three dimensional shading and hidden line/surface effects
- Interactive manipulation of displayed objects
- Comprehensive graphics libraries provided for both C and MATLAB.
- Compatibility with TVPROC.

*PHIGSDraw* is different from some data visualisation packages in not operating directly upon the raw data. The user does not present the data to *PHIGSDraw* and then select the graphing or display methods to be used. As its name suggests *PHIGSDraw* is closer to a three dimensional CAD drawing package, providing a range of geometrical drawing commands that can be used to place lines, points and surfaces in three dimensional space. Libraries of drawing and graphing commands are provided for both C and MATLAB. Using the commands in these libraries the user must write a program to...
display their data. The resulting display and graphing commands are written to a graphics metafile, which can be loaded and displayed using PHIGSdraw.

A one-dimensional graph, for example, can be created by drawing axes (command `axis(...)`) and a multi-point line (command `p_polyline(x, y, z)`). The graph data will then be used to provide the point coordinates for the line, with `x`, `y` and `z` thus being data vectors for the graph (for the case of a one-dimensional graph `z` would be held constant).

The graphics metafile that is created by use of the C and MATLAB commands, or by the TVPROC contouring program `find_contour`, is said to define a display object. This object can then be viewed using PHIGSdraw.

**F.2 The development of PHIGSdraw**

PHIGSdraw was developed initially for display of the three dimensional spectrums generated by TVPROC, as well as an aid in visualising the three dimensional concepts involved in studying television processing systems. The ability to view three-dimensional structures in an interactive manner (rotate, zoom etc) greatly aids their understanding and interpretation.

At the time the author began the only package available within the University Department that could offer such three dimensional viewing capabilities on the SUN platform was Ape. Ape however, is a complex and relatively slow visualisation package. The difficulty of using this package, and its relative unsuitability to the task convinced the author to look for an alternative. This alternative was found in PHIGS (Programmers Hierarchical Interactive Graphics System), a library of graphics functions designed for use in generating, manipulating and displaying three dimensional objects. Some preliminary work had been done using PHIGS by Niles Oien [1992] in writing a program to display bone structures. This program was used as a starting point for the development of a package that would meet the author's more complex requirements. During this process the potential of the package was realised, and, with the existence of only one other similarly capable package within the Department it was decided to continue developing PHIGSdraw to produce a fully general three dimensional viewing package.

To achieve this an extensive range of graphics drawing commands were incorporated, useable from both MATLAB and C. The package was extended to allow the use of different display modes and three dimensional effects, the saving of display views, and the user customisation of the package operation. The user interface was also improved.

To enhance the display of volume data (three dimensional spectrums are of this type) a separate contouring program was created. This program was designed to operate on the TVPROC data format and produce three dimensional threshold contours or surfaces for viewing in PHIGSdraw.

**F.3 PHIGSdraw - the display capabilities**

**F.3.1 The Display Model**

Figure F.2 shows the display model used by PHIGSdraw.

The view box defines the portion of three dimensional space that is mapped onto the display window. Any part of the display object outside of the view box will not be seen. This means that objects intersecting the boundaries of the view box will appear to have those parts outside of the view box 'cut off'. The size and dimensions of the box are determined by the viewing extents, which are a cube defined by the two points: 

\[(x_{\text{min}}, y_{\text{min}}, z_{\text{min}}) \& (x_{\text{max}}, y_{\text{max}}, z_{\text{max}})\]

The extents must be specified for each display object. Note that the extents do not have to be of an equal size in all dimensions, although they will appear as having equal size when viewed. As an example the extents of a television spectrum might be

\[
x = \pm 352 \text{ (m in c/ph)}
\]

\[
y = \pm 312.5 \text{ (n in c/pw)}
\]
The view of the currently displayed object can be changed by the user. This is done by shifting, or rotating the viewpoint position, relative to the view box. From the users point of view it is easier to think of moving the object, rather than moving the viewpoint. For this reason PHIGSdraw represents any changes of viewpoint as an object transformation. The three possible transformations that can be applied are then

- rotation (around X, Y or Z axes)
- shifting (in X, Y or Z directions)
- scaling (zooming closer in, or further away).

These actions will alter the position and orientation of the displayed object within the view box, thus changing the picture shown in the display window.

F.3.2 The Display Object

The display object is defined as being the set of three dimensional drawing primitives, with attributes, that are created by displaying the contents of a particular graphics metafile. The drawing primitives used within PHIGSdraw are:

- points (markers)
- lines
- surfaces
- text

Each of these drawing primitives has associated attributes

- colour
- size (or line width)
- style

The display object that these drawing primitives create may, optionally, be multi-layered. This means, as illustrated by Figure F.3 that the object may consist of a continuously viewable base layer
plus any number of separately viewable floating layers. The base layer is always visible, but at any one time only one of the floating layers will be visible, with the user being able to choose which floating layer to display.

This multi-layered option is designed specifically for the display of multiple threshold contours from volume data. In this case the base layer would contain axes and labelling information, and each of the floating layers would display three dimensional surface contours for different threshold values. Such an object is shown in Figure F.4, with three views, one for each of three floating layers which are threshold contours at -3, -6 and -12 dB. The axes and labels, being part of the base layer, are always visible.

This facility could also be used to display four dimensional data, with the parameter for the fourth dimension being varied for each floating layer. Displaying each floating layer one at a time would then provide an indication of how the three dimensional function changed with variation of the fourth parameter.

**F.3.3 Display modes and three dimensional effects**

Different display modes can be used when viewing a display object. These modes determine which three dimensional rendering effects are used to enhance the displayed view of the object that is projected onto the display window. The currently available effects are:

- perspective
- surface shading (ambient / direct / speculative)
The display capabilities

- hidden line / surface removal
- depth shading.

These effects are used in order to better provide the illusion of a three dimensional object. However, they can also slow down the display process. In some cases they may make interactive viewing (i.e. animated zooming or rotation) of the object difficult, each changed view taking several seconds to display. For this reason, when viewing simple objects that contain no surfaces different display modes can be selected, disabling some of the slower three dimensional rendering effects.

F.3.4 The Graphics Commands

A PHIGSdraw graphics metafile is created by using special drawing commands within C or MATLAB. Each of these commands takes an input of one or more three dimensional coordinates, display attributes and the name of a graphics metafile. The command then places its data into the metafile, appending it to any data already contained. Depending upon how the user wishes to display their data the graphics commands can be used much like commands in other packages. For example the command

```c
p_mesh (*demo_pic*, xp, yp, zp)
```

places a mesh, using the data in the two dimensional arrays xp, yp and zp into the metafile “demo_pic”. This usage is very similar to MATLAB’s own mesh command (although MATLAB’s can take only the x and y arguments). The main difference is in the way the command is executed. Nothing is actually drawn until the metafile is finally loaded into PHIGSdraw however, all commands operate as if they were drawing directly into PHIGSdraw’s view box.

The C and MATLAB graphics libraries provide a comprehensive range of commands from the drawing primitives (line, mesh, surface, text, etc) to the creation of geometrical objects (spheres, cubes, dodecahedrons, etc). (For a list of all graphics commands available in the C and MATLAB PHIGSdraw libraries see the TVPROC Reference Manual). As well as drawing commands there is also a number of control commands, used to specify view box extents, the number of floating layers, graphics modes, etc.

Many of the PHIGSdraw commands have long argument lists, needed to allow the specification of the many drawing primitive attributes. In order to simplify their usage the commands may be called with varying numbers of arguments specified. (This list must be null terminated in C). Default values are provided for the missing arguments. As an example the d_text command has the following syntax:

```c
d_text(text, ref_ptP, < colour, size, orientation, leadline, text_font, offsetP, text_quality>, 0)
```

but may be used with only the first three arguments, the other arguments being given internally defined defaults. eg

```c
d_text("demo_pic", "hello world", text_point)
```

Defaults:  
<table>
<thead>
<tr>
<th>Argument</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>colour</td>
<td>white</td>
</tr>
<tr>
<td>size</td>
<td>20</td>
</tr>
<tr>
<td>orientation</td>
<td>right</td>
</tr>
<tr>
<td>leadline</td>
<td>off</td>
</tr>
<tr>
<td>text_font</td>
<td>mono</td>
</tr>
<tr>
<td>offset</td>
<td>0</td>
</tr>
<tr>
<td>text_quality</td>
<td>high</td>
</tr>
</tbody>
</table>

The arguments are normally arranged in order of importance, meaning that most of the time only the first few arguments need to be specified by the user.
Figure F.5 Block diagram of PHIGSdraw showing external data and internal links with the PHIGS system.

Figure F.5 shows a block diagram of the PHIGSdraw software. This software provides an interface to the PHIGSdraw graphics commands, as well as an interactive viewing interface for the PHIGSdraw graphics transformations. The software also takes care of the loading of graphics metafiles, saving of object views and setting of the display configuration. Four of the sections shown in Figure F.5 are labelled in Figure F.1 as

1. View Panel
2. Control panel
3. Status panel
4. Mouse panel

The function of these will be explained within this chapter.

F.4.1 Setting the configuration

Before any display objects can be loaded from a metafile PHIGSdraw must be configured by loading either one of the default configurations, or a user created configuration file. Below is the contents of an example user configuration file.

```
win_height = 600
win_width = 600
proj_type = PERSPECTIVE
perspective_dist = 24
double_buffered = OFF
hidden_line_removal = ON
z_buffer = ON
dothcue = ON
```
PHIGSDRAW - THE SOFTWARE PACKAGE

nrotate = 12
shift_step = 0.05
zoom_step = 1.05
rotate_speed = 1
shift_speed = 1
zoom_speed = 1

The configuration file defines the size of the display window, the use of three dimensional rendering effects and the setting of parameters that control the interactive viewing interfaces. (The control panel and mouse panel). It is often useful to use different configuration files depending upon the type of display object being viewed. Simple line drawn objects, as discussed earlier, do not require many of the three dimensional rendering effects to be enabled, and can be rotated and moved in small increments as their redraw time is fast. For such objects the (MODE 1) configuration is best. For other objects that are more complex and contain surfaces (MODE 2) is provided. The user’s own custom configuration may be loaded with the (USER CONFIG) option.

F.4.2 Loading the display object from a graphics metafile

Pressing the (LOAD) button will cause PHIGSdraw to load and display the objects defined in one or more graphics metafiles. The name(s) of the graphics metafiles to be loaded must be placed in a special text file called phigsdraw.load. This may be done by editing the file, or by double-clicking on a PHIGSdraw metafile icon. (If TVPROG File Manager bindings have been set up PHIGSdraw graphics metafiles will all be displayed as At;). If more than one metafile is specified then several objects will be displayed concurrently. This can be a useful method of comparing several objects (eg. two spectrums).

As explained in section F.3.4 the PHIGSdraw metafile actually contains a list of commands, plus their associated data. When the metafile is loaded these commands are interpreted and translated into a sequence of appropriate PHIGS graphics commands. As an example the p_axis command will consist of a list of text and line drawing commands. With this modular approach new commands can be easily added to PHIGSdraw by creating a new translation function for the command.

F.4.3 Viewing the display object

The object(s) that have been loaded are automatically displayed with the current rendering effects applied. The sizes of the objects are determined by the extents specified in the graphics metafile.

Once loaded the object(s) may be viewed from any angle by applying the transformations of rotation, translation and scaling. There are three different interfaces through which these transformations can be applied: the command panel, the mouse panel, and the macro interpreter.

The command panel

The command panel provides buttons to select the transformation options shown in Table F.1.

<table>
<thead>
<tr>
<th>transformation</th>
<th>X axis</th>
<th>Y axis</th>
<th>Z axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>rotation (step)</td>
<td>(R-X)</td>
<td>(R-Y)</td>
<td>(R-Z)</td>
</tr>
<tr>
<td>rotation (180° animated)</td>
<td>(Flip-X)</td>
<td>(Flip-Y)</td>
<td>(Flip-Z)</td>
</tr>
<tr>
<td>translation</td>
<td>(=&gt;)</td>
<td>(==&gt;)</td>
<td>(UP)</td>
</tr>
<tr>
<td>scaling</td>
<td>(ZOOM IN)</td>
<td>(ZOOM OUT)</td>
<td></td>
</tr>
</tbody>
</table>

Table F.1 Transformations applied with buttons on PHIGSdraw command panel.

Except for the (Flip-) option all of these result in the transformation of the object by a discrete amount. For example, each press of the (ZOOM IN) button might cause the display object to be enlarged by 10% (ie. scale factor = 1.1). The increments used for the rotation, translation (shift) and zoom are defined in the configuration file (see earlier).
Mouse panel

The control panel options only allow transformations to be applied in pre-defined increments. An alternative to this is to use the movement of the mouse pointer on screen to provide a more flexible and intuitive interface. This is achieved by mapping the horizontal and vertical movements of the mouse, within the mouse panel, to two of the possible object transformations. The mappings available are listed below.

<table>
<thead>
<tr>
<th>Mapping</th>
<th>Horizontal mouse movement</th>
<th>Vertical mouse movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rotation around Y axis</td>
<td>Rotation around X axis</td>
</tr>
<tr>
<td>2</td>
<td>Translation along X axis</td>
<td>Translation along Y axis</td>
</tr>
<tr>
<td>3</td>
<td>Zoom in/out</td>
<td>Translation along Z axis</td>
</tr>
</tbody>
</table>

Table F.2 Image transformation mappings for mouse.

The mapping used may be changed by pressing the middle mouse button. The mapping is applied by pressing the left mouse button and 'dragging' the object, with the transformation applied being determined by the current mapping. If, for example, mapping 3 was selected then the 4 directions of mouse movement would produce the effects listed below:

- drag left - zoom in
- drag right - zoom out
- drag up - move object further into screen.
- drag down - move object further out of screen.

Note that the transformations are applied relative to the view panel, not the object.

Using the mouse in this manner gives the user the illusion of being able to 'pick-up' and manipulate the object on screen.

Macro interpreter

To allow more complex display operations to be performed a simple transformation macro interpreter is built into PHIGSdraw. This interpreter takes as its input an ASCII text file which contains a list of loading and viewing commands. All of these commands are of the format

`command parameter <repeat_No>`

Execution of the commands operates similarly to many of the control panel options, with the `parameter` providing the transformation increment amount (eg. degrees rotation) and the `repeat_No` being equivalent to the number of times the option is repeated.

For example, the command

`rotate-x 6 60`

will cause the displayed object(s) to be rotated around the X-axis by 360°, in 60 steps of 6°. The object will thus appear to 'spin' around the X-axis. Macro and lists are used to create some of the more complex commands available on the control panel. The (Demo) option, for example, has the macro command list shown below:

```
zoom-out 1.05 30
show 1
zoom-in 1.05 30
wait 1
tilt 3 10
wait 2
rotate-z 6 60
wait 2
rotate-x 6 60
wait 2
zoom-out 1.05 60
```

This list of commands makes the object zoom onto the screen, tilt, rotate round the Z-axis, rotate round the X-axis, then zoom out till it disappears.
The status panel

The top area of the mouse panel also serves as the status display panel, containing some basic information on the current status of the package.

- Names of currently loaded metafiles
- Configuration mode
- Mouse interface mapping mode.

F.4.4 Saving a view

For documentation purposes a particular view of an object can be saved as a bitmap image. This is done simply by pressing one of the save buttons in the command panel. The result is a colour or black and white bitmap saved on disk. This can then be printed, or included in a document.

F.4.5 Creating a PHIGSdraw graphics metafile

A PHIGSdraw graphics metafile can be created by using the commands provided in the C and MATLAB command libraries, or by use of the TVPROC find_contour command (designed for displaying 3D data).

Using MATLAB

A library of MATLAB functions is provided to allow the plotting of data contained in MATLAB matrices. Each command takes the metafile file name as an argument and writes data into that file. In creating the metafile the first command used is normally the d_extents command, specifying the view box extents to be used when displaying the metafile contents.

Below is an example of MATLAB program used to create a three dimensional picture of a suggested 'ideal' PAL spectrum [Drewery 1977, Fig. 29]. Figure F.6 shows a view of the resulting displayed object in PHIGSdraw.

```matlab
% SETUP EXTENTS OF SPECTRUM
ftmax = 25;  % Hz
fvmax = 312.5;
fmmax = 352;

% SIZES OF Y, U, V COMPONENTS
Yheight = 25;  % Hz
Uheight = 12.5;
Vheight = 12.5;

Yvradius = 312.5;  % cycles / height
Uvradius = 78*sqrt(2);
Vvradius = 78*sqrt(2);

Yhradius = 320;  % cycles / width
Uvradius = 80*sqrt(2);
Vvradius = 80*sqrt(2);

% POSITIONS OF YUV COMPONENTS
Ypos = [0; 0; 0];
Upos(1, :) = [78; 284; 18.75];
Upos(2, :) = [-234; 284; -6.25];
Upos(3, :) = [234; -284; 6.25];
Upos(4, :) = [-78; -284; -18.75];

Vpos(1, :) = [78; -284; 18.75];
Vpos(2, :) = [-234; -284; -6.25];
Vpos(3, :) = [234; 284; 6.25];
Vpos(4, :) = [-78; 284; -18.75];

% CREATE GRAPHICS METAFILE
```
nsteps = 20; mode = 2;
file = 'idealpall'
clear_dfile(file);
write_dfile(file);

extents(file, -fvmax, fvmax, -fhmax, fhmax, -ftmax, ftmax);

axes(file, 'v', 'h', 't', 7, 2, 1, 20);

box(file, 8);

cone(file, [Ypos, Ypos], [Yvradius, Yvradius], [Yheight, -Yheight], nsteps, mode, white);

for i = 1:4,
    cone(file, [Upos(i,:), Upos(i,:)], [Uvradius, uvradius], [Uheight, -Uheight], nsteps, mode, red);
    cone(file, [Vpos(i,:), Vpos(i,:)], [Vvradius, Vvradius], [Vheight, -Vheight], nsteps, mode, blue);
end,

Photograph of Ideal PAL in colour

Figure F.6 Rendering of a 3D image of an idealised PAL spectrum [Drewery 1976] using PHIGSdraw.

Using C

Using C to create PHIGSdraw metafile is very similar to using MATLAB, with the same commands being available, and very similar arguments. The C commands, like MATLAB, also take variable argument lists, however, these must be null terminated for C.

Writing a Conversion Program

When wishing to display data from a source external to TVPROC, C or MATLAB it is usually necessary to write a conversion program, in either C or MATLAB. This program must load the data, from whatever format it is stored in, and use PHIGSdraw drawing commands to graph and display the data according to the users requirements. An example of this being done is a program written by the author to convert zero sheet data (generated as part of a techniques for deblurring images) into a PHIGSdraw metafile, to enable the zero-sheets to be viewed using PHIGSdraw.

F.4.6 Volume data contouring

Volume data, such as CT data, or a television spectrum is impossible to display in its entirety. For three dimensional display such data is normally viewed as three dimensional contour surfaces (similar to a two dimensional contour plot) with the contour surfaces denoting a particular threshold boundary, or data value within the volume (e.g. the skin, or bone in a CT scan).

The find_contour command is a command line program which can produce such contours, taking as its input a three dimensional TVPROC data file and a set of threshold values. The output is in the form of a PHIGSdraw graphics metafile, containing a multi-layered object for which each layer contains the contour surfaces of a different threshold.

Depending upon how the user wishes to view the results in PHIGSdraw several output modes are available, to suit different PHIGSdraw display modes. These different modes produce objects that are constructed from line contours, hidden line meshes, or shaded surfaces. Line contours display quickly in PHIGSdraw, and can be animated interactively to provide a good three dimensional perception of the object. However, such objects can be confusing when displayed without
movement. Mesh and surface contours are slow to display, but more useful for documentation purposes. All of the spectrum plots shown within this thesis were created using the mesh option.

A configuration file is used to define the operation of these modes, as well as a number of other parameters.

**Volume data contouring - how it works**

The `find_contour` command produces threshold surface contours from volume data (three dimensional data). These contours indicate areas in the data where the specified threshold is crossed. This concept is illustrated in two dimensions below.

![Figure F.7 Contouring operation performed in 2D, (a) original 2D data set, (b) a set of multiple threshold contours.](image)

The process used by the `find_contour` command to convert the data to line and surface contours is shown below. The original data set is first thresholded at the chosen value. The resulting data is passed through a vertical and horizontal edge detector, and then a contour follower to produce a set of multi-point lines that can be plotted in PHIGSdraw.

![Figure F.8 Process used by find_contour to convert a data set to set of contour lines.](image)

To produce a three dimensional contour the same method is used. The three dimensional data is divided into a series of two dimensional slices which are each separately converted into line contours. Depending upon the complexity of the contour plot required this process can be repeated along each of the three axis, resulting in three intersecting sets of contours. This is illustrated by below.
When combined these form an object with a mesh appearance.

Surface contours are at present created by displaying each two dimensional contour as a solid object of a thickness equal to the distance between contours. This has a similar effect to building the object from a series of layered wooden cutouts.

Note that the line contours provide no information on which areas of the data are above or below the threshold they only indicate the threshold boundaries.

**F.5 Summary**

**PHIGSdraw** has been developed as a general tool to aid in the display and visualisation of three dimensional structures. The package created allows the use of C, MATLAB or TVPROC to plot mathematical structures in three dimensions, or display multi-layered contours produced from volume data. The interactive nature of the package in viewing the object and the provision of three dimensional rendering effects is important in providing the user with a good appreciation of the structure of the object(s) displayed.

**PHIGSdraw** has, thus far, proved invaluable in the study of enhanced PAL techniques, as well as finding applications in other areas. These areas include the display of X-ray diffraction patterns, CT volume data, and image zero-sheets (multi-dimensional structures used in image restoration techniques).
REFERENCES


BRUCH, W. (1966), 'A New Approach to Centrally Correcting the Phase of a PAL Colour TV Signal (PAL-PAL Transcoder), Electronic Engineering, April


CLARKE, C.K.P. (1982), 'High quality decoding for PAL inputs to digital YUV studios', BBC Research Department, BBC RD 1982/12, July.


CLARKE, C.K.P. (1986), Colour encoding and decoding techniques for line-locked sampled PAL and NTSC television signals, BBC Research Department, BBC RD 1986/2, March.


REFERENCES


SANDBANK, C.P., DREWEY, J.O. AND WESTON, M., (1990), 'Phase Segregation, Motion Compensation and DATV for Enhanced PAL and High Definition Television', BBC Research Department.


STOREY, R., (1992), 'BBC investigations into improved PAL', IEE Colloquium on 'Enhanced and Widescreen PAL', Digest No. 006, January, 4/1-3.


REFERENCES


TVNZ, (1963), New Zealand Standard Specifications for Television Transmission and Reception (1963), New Zealand Standards Institute, Wellington, April.


WATKINSON, J. (1994), The Engineer’s Guide to Decoding and Encoding, Snell and Wilcox Ltd.

WATKINSON, J. (1994), The Engineer’s Guide to Motion Compensation, Snell and Wilcox Ltd.


