

# Supervision of video and audio content in digital TV broadcasts

MICHAIL VLASENKO



**KTH Information and  
Communication Technology**

Master of Science Thesis  
Stockholm, Sweden 2007

COS/CCS 2007-30

# Supervision of video and audio content in digital TV broadcasts

Master thesis performed at Teracom AB

Michail Vlasenko  
Email: [michail@kth.se](mailto:michail@kth.se) or [michail.vlasenko@teracom.se](mailto:michail.vlasenko@teracom.se)

Examiner: Prof. Gerald Q. Maguire Jr, KTH  
Supervisor: Petri Hyvärinen, Teracom

## **Abstract**

An automatic system for supervision of the video and audio content in digital TV broadcasts was investigated in this master's thesis project. The main goal is to find the best and most cost effective solution for Teracom to verify that the broadcast TV content as received by remote receivers is the same as that incoming to Teracom from content providers. Different solutions to this problem will be presented.

The report begins with some background information about the Swedish terrestrial digital TV network and the MPEG-2 compression standard used to transport audio and video; including a description of the DVB Systems and Transport Stream protocol. It describes two current techniques for the supervision of the audio and video content, including an evaluation of these techniques.

The first solution is to monitor the video and audio content either by detecting common errors such as frozen picture, visible artifacts, or by comparing the content from two different sources, i.e. a comparison of the output and the input content. The later could be done using video fingerprinting. The second solution monitors the video and audio content indirectly by analyzing the Transport Stream. This could be done either by comparing two Transport Streams to verify that the broadcast signal is identical to the received signal or by detecting common errors in the streams.

Further two new potential solutions will be presented based on the research utilizing background knowledge of the MPEG-2 compression standard. The thesis ends with a summary with conclusions and evaluations of all four solutions and future work.

## **Sammanfattning**

Ett system för automatisk övervakning av ljud- och bildinnehåll i digitala TV sändningar var undersökt i detta exjobb. Målet är att hitta bästa och mest kostnadseffektiva lösningen för Teracom för verifiering av TV innehållet som tas emot av fjärrmottagare är densamma som Teracom får från sina tjänsteleverantörer. Olika lösningar till detta problem blir presenterade.

Presentationen startar med bakgrundsinformation om Sveriges marknät för digital TV och MPEG-2 komprimeringsstandarden som används för ljud- och bildsändningar. Den kommer att inkludera en kort beskrivning av DVB system och Transport ström protokoll. Två nuvarande tekniker för övervakningen av ljud- och bildinnehåll kommer att presenteras.

Första lösningen handlar om att övervaka TV innehåller antigen genom att detektera vanligast förekommande fel såsom fryst bild, tydliga artefakter eller genom en jämförelse av innehållet från två olika källor, dvs. en jämförelse av ingångs och utgångssignal. Den senare kan åstadkommas genom att använda en så kallad video fingeravtryck. Andra lösningen övervakar ljud- och bildinnehåll indirekt genom att analyser Transport strömmen. Detta görs genom en jämförelse av två Transport strömmar för verifiering om signalen är densamma samt genom detektering av vanligast förekommande fel i strömmarna.

Vidare, två nya potentiella lösningar kommer att presenteras med utgångspunkt från den bakgrundskunskap om MPEG-2 komprimerings standard som getts. Presentationen avslutas med en sammanfattning och utvärdering av alla fyra lösningar och framtida arbeten.

## Table of contents

<b>1. Introduction .....</b>	<b>1</b>
1.1 System description .....	1
1.2 The supervision system requested by Teracom.....	2
1.3 Teracom.....	2
<b>2. A description of Teracom's systems .....</b>	<b>3</b>
2.1.1 Architecture .....	3
2.1.2 Single Frequency Network.....	3
2.1.3 Net planning .....	3
2.2 Primary distribution system.....	4
2.3 Secondary distribution.....	6
<b>3. MPEG-2.....</b>	<b>7</b>
3.1 Introduction .....	7
3.2 Video compression methods – an overview .....	7
3.3 Video compression.....	8
3.3.1 Video basics.....	8
3.3.2 DCT coding .....	9
3.3.3 Quantization .....	10
3.3.4 Zigzag scanning.....	10
3.3.5 Run length code.....	10
3.3.6 VLC .....	10
3.3.7 Buffer occupancy control .....	11
3.3.8 Motion compensation techniques.....	11
3.3.9 Hierarchical structure of MPEG-2.....	12
3.4 Audio compression.....	13
3.4.1 Masking .....	13
3.4.2 Filter bank.....	14
3.4.3 Bit allocator .....	14
3.4.4 Scaler and quantizer .....	15
3.4.5 Multiplexer .....	15
3.4.6 MPEG Layer II characteristics .....	15
3.4.7 AC-3 .....	16
3.5 DVB Systems .....	18
3.5.1 Transport Stream .....	18
<b>4. Methods and analysis .....</b>	<b>21</b>
4.1 Introduction .....	21
4.2 IdeasUnlimited .....	21
4.2.1 Test bench.....	21
4.2.2 Single-ended mode .....	22
4.2.3 Double-ended mode.....	23
4.3 Agama .....	25
4.3.1 Test bench.....	25
4.3.2 Agama Analyzer .....	25
4.3.3 Agama Verifier.....	26
4.4 Investigation of DCT coefficients and Scale factors usability .....	27
4.4.1 Introduction .....	27
4.4.2 Test bench for video .....	27
4.4.3 Test cases for examination of video content .....	28
4.4.4 Conclusion of DCT coefficients usability .....	32

4.4.5 Detection of bit errors based on subsequent syntax errors .....	32
4.4.6 Test bench for audio .....	33
4.4.7 Conclusion of scale factors usability .....	35
4.5 Monitoring the digital data stream using signatures and syntax .....	36
<b>5. Conclusions .....</b>	<b>38</b>
5.1 Evaluation.....	38
5.2 Future work .....	38
<b>References .....</b>	<b>39</b>
<b>Appendix A .....</b>	<b>42</b>
<b>Appendix B.....</b>	<b>43</b>
<b>Appendix C .....</b>	<b>44</b>
<b>Appendix D .....</b>	<b>45</b>

## Abbreviations

ADC	Analog to Digital Converter
ATM	Asynchronous Transfer Mode
CAT	Conditional Access Table
DCT	Discrete Cosine Transform
DTT	Digital Terrestrial Television
DVB	Digital Video Broadcasting
GOP	Group Of Pictures
HDTV	High Definition TV
IDCT	Inverse Discrete Cosine Transform
MDCT	Modified Discrete Cosine Transform
MPEG	Moving Pictures Experts Group
NTP	Network Time Protocol
PAT	Program Association Table
PCR	Program Clock Reference
PES	Packetized Elementary Stream
PID	Packet Identifier
PMT	Program Map Table
PSI	Program Specific Information
PTS	Presentation Time Stamp
RLC	Run Length Code
SCFSI	Scale Factor Selector Information
SDH	Synchronous Digital Hierarchy (a transport protocol used by Teracom, primary for the analogue TV distribution)
SFN	Single Frequency Network
SI	Service Information
STC	System Time Clock
VLC	Variable Length Coding

# 1. Introduction

## 1.1 System description

In April 1999 on the behalf of the Swedish government, Teracom began broadcasting of the digital terrestrial TV using the Digital Video Broadcast - Terrestrial (DVB-T) standard [1]. At that time the network consisted of multiple transmitting towers each of which was fed with the modulated output of two multiplexers. This system initially provided coverage of 50% of the fixed households in Sweden. In DVB-T, a multiplexer is a collection of TV services carried on one single frequency allocation (it will be described in detail in section 3.5). Today there are six multiplexers (see figure 1); with four of them connected to transmission towers having 90% population coverage (the broadcast from multiplexers 1-4), the fifth multiplexer being connected to transmission towers offering 50% population coverage and the sixth multiplexer being connected to transmission towers only in Mälardalen region (Stockholm, Uppsala and Västerås).

There are in total 33 digital TV channels broadcasted with approximately 5-7 digital TV channels carried by each multiplexer. Most of these digital TV channels are scrambled using the Viaccess [8] encryption scheme, which allows only paying customers access. Other TV channels are free to view; these include for example public service channels from SVT, the privately owned TV4, TV6, and some other channels. Some of the TV channels have different regional content during certain times of the day. For instance, public service channel SVT2 is split up into 20 local news feeds several times each day. There are 54 transmission sites in Sweden, each with a large coverage area. For example, in Stockholm the site is located in Nacka.

Currently MPEG-2 video compression and MPEG-1 Layer II and AC-3 Dolby Digital audio compression are deployed for Standard Definition TV. For High Definition TV (HDTV) it is planned to use H.264 (MPEG-4) video compression. Audio compression is planned to be Dolby Digital AC3+ and MPEG-4 HE-AAC (High Efficiency AAC). Audio compression is further described in section 3.4.

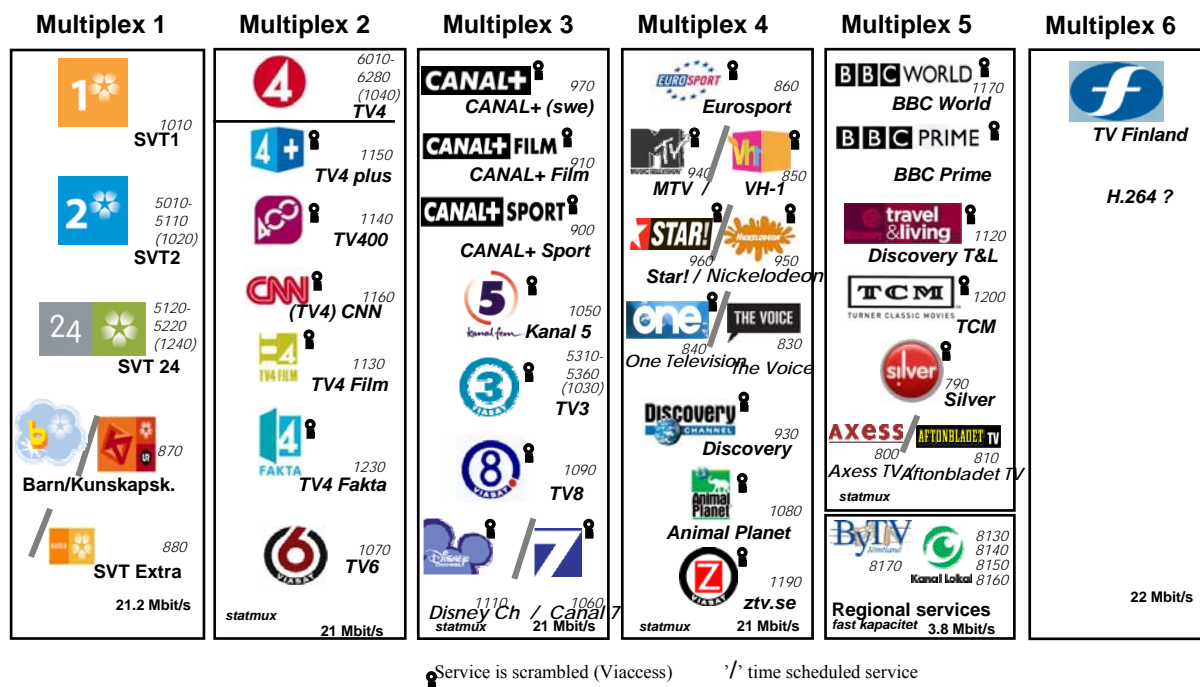


Figure 1: Programmes in Swedish digital terrestrial television broadcast as of January 2007 (Courtesy of Teracom). In each multiplexer there are certain number of TV channels, each one with a Packet Identifier (PID) number, further described in section 3.5.1. In addition, each multiplexer has a certain Bit Rate (typically 21 Mbit/s - as indicated at the bottom of each column).

## 1.2 The supervision system requested by Teracom

The supervision system is intended to monitor the content that is broadcasted by Teracom. Some of these contents are: video, audio, Program Specific Information / Service Information (PSI/SI), DVB subtitling, and teletext subtitling. This report will concentrate on the video and audio content.

If a failure is detected, then an alarm indicating the information about this failure should be routed to Teracom's central supervision system. A failure can occur for many different reasons, such as antenna collapse, a disconnection of the feed to the antenna, power failure at the antenna site, equipment failure, a link failure of one of the links in the backbone (fixed) transmission system, etc.

The monitoring of the video content involves the detection of a black or frozen picture, and visible bit errors in the signal. The monitoring of the audio content involves a detection of audio signal loss or audible failures in the audio signal.

The reason to utilize a supervision system is to ensure the quality of the broadcast services and provide a high QoS (Quality of Service) for Teracom's customers (the content providers). This supervision system should be able to quickly detect errors, and raise the appropriate alarm, then based upon these alarms Teracom should quickly resolve the problem - or at least take steps to alert their customers that there is a problem and that they are working to solve it.

## 1.3 Teracom

Teracom is a terrestrial network operator owned by the Swedish state. Before 1992, Teracom was a part of Televerket, but since then it has become an independent, state owned public service corporation. It is Sweden's largest TV and radio operator, and has broadcast radio and TV programs for almost 80 years [15]. The main customers are the Swedish public service television and radio broadcasting companies, Sveriges Television and Sveriges Radio, as well as the commercial television channel TV4. Another large customer is Boxer TV-Access AB, a company in which Teracom has a 70% ownership in. Boxer TV-Access AB offers individual households and entire buildings access to digital TV and interactive services [16]. The remaining 30% ownership is controlled by venture capital company 3i.

The term "content providers", for the purposes of this thesis, are the customers of Teracom. This means that the end receivers of the content are actually customers of these content providers and not customers of Teracom, since Teracom has no contractual relation to the end receivers of the content. These end receivers of content are typically homes which have a DVB-T receiver and one or more decoders (and perhaps decrypters) to view and listen to the program content from one or more content providers. Note that this definition of "content provider" differs from that usually used elsewhere, as much of the content is not actually provided by these entities, but rather these entities are what might have been called TV or radio stations, but for the fact that they do **not** actually do any broadcasting themselves. Strictly speaking Teracom's customers are "programming companies". (See pg. 7 of the Teracom Annual report for 2000 [39])



## 2. A description of Teracom's systems

### 2.1.1 Architecture

Digital Terrestrial Television (DTT), by which we mean digitally transmitted broadcast television in Sweden, utilizes a network built and maintained by Teracom. Their broadcast network consists of the following: Each transmission station is equipped with 1-6 multiplexers (depending on the site) connected to transmitters operated in the Ultra High Frequency (UHF) band. Each of the 54 large TV/FM transmitting stations is assigned 1-6 frequencies for DVB-T transmission. The net bit rate for each multiplexer is 22-24Mbit/s (i.e., this represents the aggregate rate of the content from all the programs for a given multiplexer). The coverage of each transmitting station is primarily planned assuming that each end receiver is connected to a roof-top antenna which is pointed at this transmission station's antenna.

### 2.1.2 Single Frequency Network

Teracom's network implements a Single Frequency Network (SFN). This means that several adjacent smaller transmission stations that have overlaying coverage areas simultaneously broadcast using the same frequency band. This requires a time synchronization of these transmitting stations. The time reference is provided by a GPS-receiver.

### 2.1.3 Net planning

Two channel encoding modes have been used for the deployed network. The first (and main alternative) uses the following channel encoding: FFT 8K, modulation 64-QAM, code rate 2/3, guard interval 112  $\mu$ s (1/8), and net bit rate 22.12 Mbit/s. The second encoding is deployed at the larger sites: FFT 8K, modulation 64-QAM, code rate 3/4, guard interval 224  $\mu$ s (1/4), and net bit rate 22.39 Mbit/s.

The size of the Fast Fourier Transform (FFT) and guard interval affects the characteristics of the SFN. Using an 8K FFT offers 6048 data carriers for each UHF channel (the other carriers are guard bands, system signaling, etc). The guard interval indicates the maximum time difference between signals from different transmission stations that can be managed at a reception point. The modulation method and code rate describes how data is modulated and error protected. The 64-QAM (Quadrature Amplitude Modulation) means that the transmitted data is coded into 64 different symbols that are modulated in phase **and** amplitude. That gives a symbol length of 6 bits, i.e. every carrier carries a 6 bit symbol. Code rate states the amount of payload out of the total number of symbols which are sent. A code rate of 2/3 means that 2/3 of all transmitted data is user data, while the rest is redundant information that enables error discovery and a limited amount of error correction.

Teracom has an agreement with the Swedish state to provide a public broadcast network covering 99.8% of the fixed households in Sweden. The Swedish state operates such a public network because in the event of war or weather disaster it provides a means to inform the population of the situation. At other times it provides a means for the public broadcasters (i.e., Swedish Television and Swedish Radio) to reach a large audience. This infrastructure is shared with commercial broadcasters for economic reasons. Teracom planned its network coverage using data about the location and number of households in Sweden provided by the Swedish Central Statistical Bureau (Statistiska Central Byrån - SCB). SCB is a central government authority for official statistics.

## 2.2 Primary distribution system

The primary distribution system consists of the complete chain from a content provider's submitting of program content to actual television transmission sites. This chain includes the compression of the content, re-multiplexing, creating or managing of service information, scrambling, and distribution to transmitters via a transport network, in this case Teracom's Asynchronous Transfer Mode (ATM) trunk network. The figure below shows a schematic of Teracom's primary distribution system.

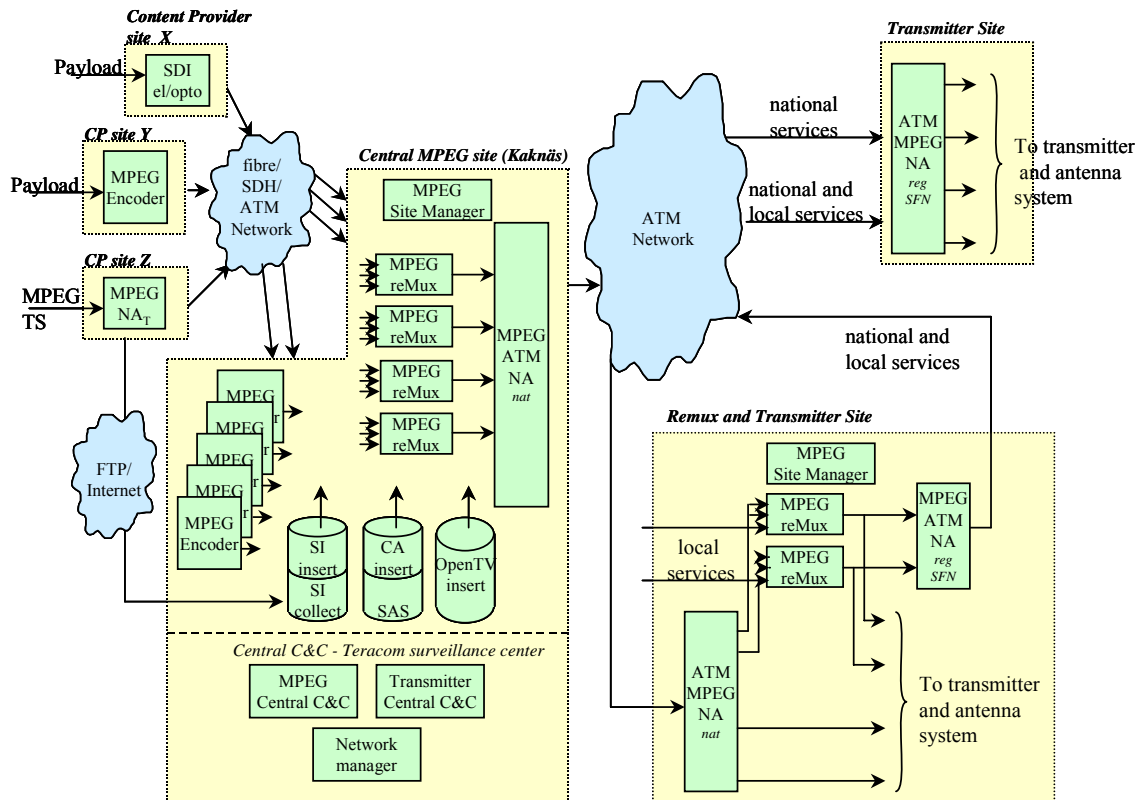


Figure 2: A schematic picture of primary distribution. (Courtesy of Teracom).

The first step in distribution of DTT services is the encoding and compression of the video and audio streams, these components are multiplexed together into one service and packetized into one MPEG-2 Transport Stream (TS). In figure 2, this process is labeled "MPEG Encoder". It is also possible to combine several services by re-multiplexing them into one TS (i.e., combining data from several Transport Streams).

There are several different possibilities for a encoding of the content provider services: encoding can be done by a content provider itself with its own equipment or using Teracom's provided equipment, or the encoding can be done at Teracom. In the last case a content provider delivers uncompressed signals to a site where the encoding is performed. This is done by sending the raw signals through a fiber network (generally an ATM or SDH network).

Joint Bit rate Regulation is used to provide more effective usage of a given transmitter's bandwidth. This requires that MPEG encoders assigned for different services have to cooperate and share the total channel capacity. Thus video components are allocated an instantaneous capacity or bit rate depending on the complexity of the current video content. This could be done to produce streams for a single multiplexer.

There are several possibilities for content providers to choose from when they determine how they want to send their services. Some content providers choose to send their content with a 16:9 aspect ratio. The image aspect ratio information is usually specified in the sequence header of the MPEG

video stream. Content providers may also add multi-channel audio (Dolby AC-3, and DTS) and DVB subtitles.

Re-multiplexing enables the downstream distributor of a TS to change the contents of this TS. There are several reasons that this might be done: the first is related to MPEG coding: for example adding DVB subtitles or scrambling. Next re-multiplexing occurs at the central site (Kaknäs) where all national services are re-multiplexed into one TS for each multiplexer. From the central site the TS is distributed further to regional re-multiplexing stations and transmission sites. The next re-multiplexing occurs in each region where the local content (such as local news or advertisement) is added to the national services. The TS is subsequently sent to the transmission stations where re-multiplexing may again take place.

Preserving a correct sense of time is very important in DTT because of the need for synchronization. Teracom is using an application based upon Network Time Protocol (NTP) [23] which broadcasts this time out to the MPEG equipment (encoders, re-multiplexers, etc.). The actual SFN transmitters use the timing signal from the Global Positioning System (GPS) as a time reference. As NTP and GPS are both derived from more accurate time sources at other strata, they are synchronized.

Program companies also deliver so called event information to Teracom. This contains information about the programs being sent, their start & duration times, category (film, news, etc.), and a description of the program. This information is sent as an Event Information Table (EIT) and it is part of the Service Information (SI) inside a Transport Stream.

Program Specific Information (PSI) and Service Information (SI) include different kinds of tables providing necessary system information. According to the MPEG standard this information includes such mandatory tables as the Program Association Table (PAT), Conditional Access Table (CAT), Program Map Table (PMT), and Network Information Table (NIT). Each of these will be described in more detail in section 3.5. The capacity for SI is around 1Mbit/s for each multiplexer.

The Conditional Access system utilizes a combination of scrambling and encryption to prevent unauthorized reception. The system for access control consists of a Subscriber Authorization System, Entitlement Management Message, and Entitlement Control Message. This system is connected to the customer database, Subscriber Management System, and MPEG re-multiplexers. The system manages the creation and encryption of keys and control messages (Entitlement Management Message and Entitlement Control Message) that are being sent. Encryption is done using the Viaccess algorithm [24]. The scrambling of TS is done in the MPEG re-multiplexers according to the DVB standard. The Subscriber Management System is managed by Boxer's<sup>1</sup> customer service organization and includes information about the subscription status of each customer. This information is sent to the Subscriber Authorization System that generates a unique Entitlement Management Message for each subscriber's smartcard. This Entitlement Management Message is encrypted such that only the intended smartcard may decrypt it. All components except teletext and PSI/SI are scrambled today (the reason for the former not being scrambled is that some receivers cannot descramble teletext).

The Interactive Data Platform consisted originally of two parts: the broadcast system and the return channel. The first connects applications to the receivers through DTT. The return channel is utilized to receive subscriber's replies from their receivers; but this system is not in use today. The broadcast system manages the compilation and viewing of OpenTV [9] applications. OpenTV has defined an Application Programming Interface. Applications are developed to utilize this standard interface. The Interactive Data Platform is also used for distribution of boot loaders (software updates for the receivers, it is used to initially load the operating system).

---

<sup>1</sup> Boxer TV-Access AB is the company that sells subscriptions for pay DTT channels in Sweden.

### 2.3 Secondary distribution

The secondary distribution system comprises the transmission system, infrastructure, coaxial, and antenna systems. Since this thesis is only concerned with the primary distribution system, the secondary distribution system will not be covered further (for details see [3]).

### 3. MPEG-2

#### 3.1 Introduction

We begin by introducing the encoding and decoding theory necessary to understand both the proposed solutions and the problems in monitoring the content of the received digital TV signal. We begin with the coding process as this will give us insight into the undesired effects on the audio and video content of not correctly receiving the intended signal.

Over the past 10 years digital communications have almost completely replaced analogue communication techniques. The main reasons are the robustness of the bit stream that contains digital information and the ability to transmit more TV channels (of the earlier resolution) using the same frequency allocation. The bit stream can be stored and recovered, transmitted and received, processed and manipulated virtually without errors [3]. In digital television this means that the picture reproduced on the home screen is **identical** to the picture in the studio. To fit all of this content into the assigned bandwidth, we must compress the digital data stream. In this compression, the main task is to reduce the bit-rate without loss of quality, this is based upon removing redundancy from the data stream - however, doing so means that failure to correctly receive and decode the received stream may not simply result in *small* errors, but may also result in **very large** errors. The compression technique which will be used exploits properties of the human visual and aural senses.

The core element of all DVB systems is the MPEG-2 coding standard. The MPEG-2 specification only defines the bit-stream syntax and decoding process [11]. The encoding process is not specified, which means that improvements in picture quality are possible. This means also that there are no requirements that encoders follow any particular model as long as the resulting data streams meet the specification, i.e. a freedom for the encoder developers allow an implementation of both low-cost and high-cost (high performance and high quality) encoders. This enables the improvement of an existing DVB system by upgrading the encoders without changing any of the receivers.

#### 3.2 Video compression methods – an overview

Digital video compression exploits the fact that successive frames of video often are similar to the previous and subsequent frames. A frame in this case can be seen as a still picture consisting of a set of color pixels. Pixels are also subject to compression since the changes in colors from pixel to pixel within a small area often are minimal. These two facts are due to temporal and spatial redundancy. We can think of a video sequence as a three dimensional array where two dimensions are the spatial (horizontal and vertical) directions of the picture, and the third dimension represents the time domain. Spatial redundancy is removed by first encoding regions of the image using the Discrete Cosine Transform (DCT), which allows us to remove some of the high spatial frequency image content, followed by quantization of the residual content. Temporal redundancy is exploited by using motion prediction techniques, thus we seek to track moving objects and simply include the fact that they have moved in a certain direction and orientation - rather than having to retransmit the object again - simply because its position within the image has shifted.

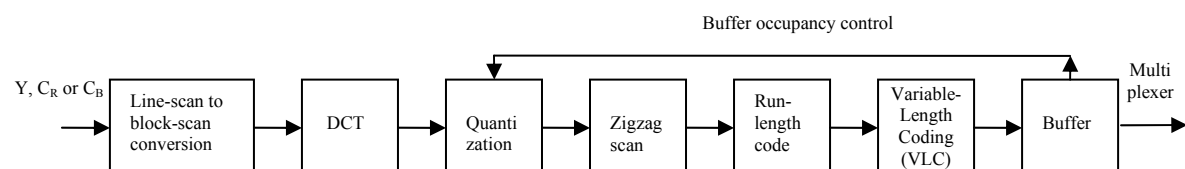


Figure 3: Basic DCT coder (adapted from [3]).

### 3.3 Video compression

#### 3.3.1 Video basics

There are three light properties related to color television that controls the human visual sensations when presented with this light. These properties are known as brightness, hue, and saturation. Red, blue, and green have been chosen as the primary colors for television. The proper combination of these three colors produces white. Luminance represents the brightness in the picture, i.e., the intensity of light in the picture. Chrominance represents the color information in the picture and is expressed by two of the three color signals minus the brightness component; these signals are known as the blue and the red color difference signals. In digital component systems the image signals are expressed as  $Y C_R C_B$  signals where  $Y$  represents the luminance component and  $C_R$  and  $C_B$  represents chrominance components.

To calculate  $Y C_R C_B$  values a translation has to be done. First, a compensation for the nonlinearity in the human visual system's perception of intensity is done by introducing a compensating nonlinearity, usually referred to as gamma correction. The conversion from gamma corrected RGB components generates a YUV color space. The translation to  $Y C_R C_B$  color space is obtained by scaling and offsetting the YUV color space. The result of the conversion from gamma corrected RGB components is represented as 8-bits per component (i.e., per  $Y$ ,  $C_R$ , and  $C_B$ ).

Since the human eye is less sensitive to color (chrominance) than luminance, bandwidth can be optimized by storing more luminance detail than color detail. A family of sampling rates, based on the reference frequency of 3.375 MHz has evolved. In figure 4, 4:2:2 sampling is shown. The sampling rate for the luminance component is 13.5 MHz ( $4 \times 3.375$  MHz); while the sampling rate for each of the chrominance components are 6.75 MHz ( $2 \times 3.375$  MHz). Using 8-bits per sample, the digital bandwidth of the uncompressed signal is 216 Mbit/s.

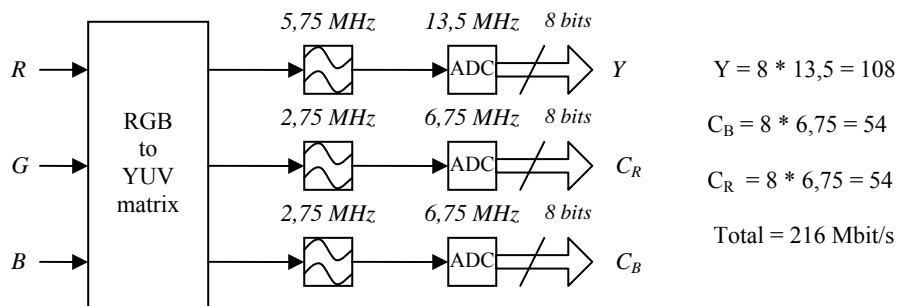


Figure 4: Video (4:2:2) sampling [11]; here 5.75 MHz and 2.75 MHz specify the bandwidth of luminance and chrominance signal. ADC is an abbreviation for Analog to Digital Converter.

Directly using a bit stream of 216Mbit/s for DTT is not possible (since this greatly exceeds the maximum per multiplexer bit rate), hence a method of reducing the bit rate is further needed. In most MPEG-2 coding applications, 4:2:0 sampling is used rather than 4:2:2. In 4:2:0, which is a relative relationship between chrominance and luminance, for every four samples of luminance forming a 2x2 array, two samples of chrominance exists, one a  $C_R$  sample and one a  $C_B$  sample. Note that the bit rate calculated for the 4:2:2 video sampling was based on the old CCIR-601 standard [10] which included methods of encoding 525-line 60 Hz and 625-line 50 Hz signals, both with 720 luminance samples and 360 chrominance samples per line. The new name of the standard is ITU-R BT.601 and it uses data bit rate of 270 Mbit/s for a 10-bit Serial Digital Interface. To reduce the data bit rate further a combination of various tools are used. Figure 3 shows a basic DCT encoder with the necessary steps to reduce the video data rate.

### 3.3.2 DCT coding

The DCT coding process transforms blocks of pixel data into blocks of frequency-domain coefficients. The purpose of using this transform is to assist the processing which removes spatial redundancy, by concentrating the signal energy into relatively few coefficients [11]. However, the DCT itself does not reduce the data rate and is totally reversible. The process of DCT, shown in figure 5, involves the transformation of an 8x8 array of luminance pixel amplitude values into an 8x8 array of DCT coefficient blocks where the resulting top left corner number is the DC coefficient representing the average DC level of luminance of the whole 8x8 array of pixels. The other coefficients indicate the size of the higher spatial frequency components of the original waveform and are called the AC coefficients. The mathematical definition of an NxN DCT is presented in appendix A.

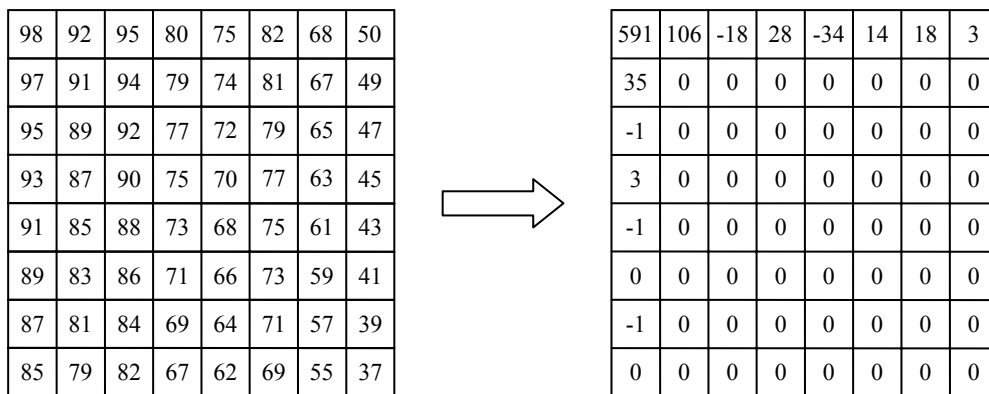


Figure 5: 8x8 blocks of pixel values transformed into 8x8 DCT transform coefficient values (adapted from [3]).

As we can see from the figure 5 most of the signal information following the transformation tends to be concentrated in a few low-frequency components of the DCT. The inverse DCT process (IDCT) reconstructs the exact original pixel values if and only if the DCT coefficients are kept unchanged. A combination of quantization and efficient coding techniques, such as Variable-Length Coding (VLC), makes a further data rate reduction possible. However, since quantization is performed after the transformation, the original signal can not be exactly reconstructed. Hence this is a lossy coding scheme. The choice of an 8x8 block size is the result of a compromise between an efficient energy compaction that requires a large screen area, and a reduced number of real-time DCT calculations that requires a small area [3].

Before compression, the original pictures are digitized by means of sampling structures chosen to achieve the required resolution. Luminance and chrominance are separated into 8x8 blocks of Y, C<sub>B</sub>, and C<sub>R</sub> values as described in section 3.3. Then, a macroblock is formed, as shown in figure 6. The ordering within a macroblock determines the sequence of blocks in which they are sent to the DCT coder.

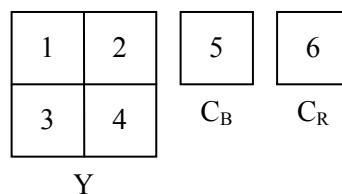


Figure 6: 4:2:0 Macroblock.

### 3.3.3 Quantization

The basic function of the quantization process is to divide each DCT coefficient by a number greater than one to generate numbers near or equal to zero. The point is that low-energy coefficients, representing small pixel-to-pixel variations, can be discarded without affecting the perceived resolution of the reconstructed picture. The main drawback of quantization is that it introduces artifacts. Two different weighting tables are used for luminance and chrominance quantization. The difference is due to the fact that chrominance information is less critical to human perception. Common for both quantization tables is that the dividing factor is small for DC and low-frequency components, and gradually increases for higher-frequency coefficients.

### 3.3.4 Zigzag scanning

The next step for the two-dimensional quantized DCT blocks is to undergo a zigzag scanning pattern to facilitate the subsequent encoding and transmission using a one-dimensional channel. Different scanning patterns are available depending on the pixel-to-pixel variations in the picture. The type of pattern chosen must be defined in the encoded bit stream in order to control the decoder.

### 3.3.5 Run length code

In run length coding (RLC) each nonzero coefficient after the DC value is coded with a two-parameter (run, level) code word, the number of zeroes preceding a particular nonzero coefficient and its level after quantization.

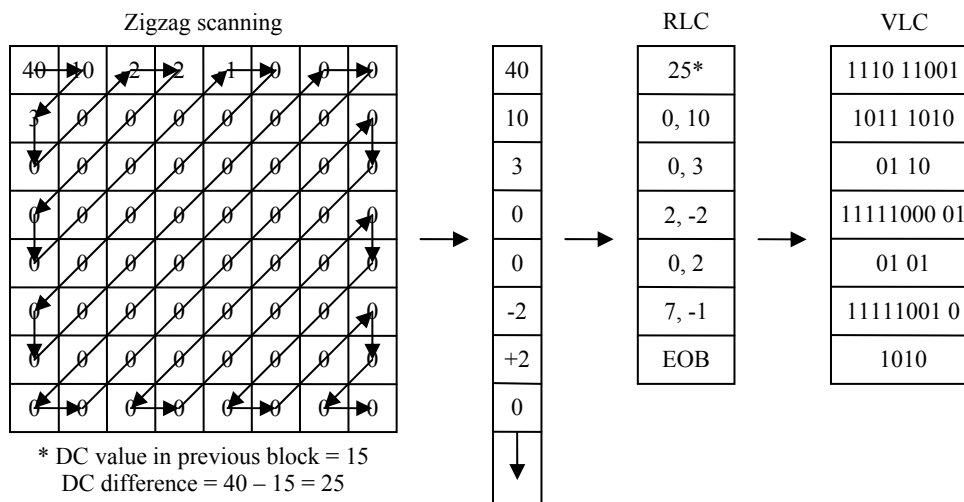


Figure 7: Zigzag scanning followed by RLC and VLC.

### 3.3.6 VLC

The RLC code words are further allocated short code words to frequently occurring levels and long code words to infrequently occurring levels. There are special tables for such code words. A short code word signals the end of block (EOB), which means that all following coefficients in the block are zeroes. Variable Length Coding (VLC), also called Huffman coding or entropy coding, is based on the probability of identical amplitude values in the picture. In the example in figure 7, the data corresponding to the original DCT coefficient block with  $8 \times 8 \times 8 = 512$  bits, is reduced to 48 bits after VLC encoding.



### 3.3.7 Buffer occupancy control

A buffer occupancy control mechanism ensures that no buffer underflow or overflow occurs. This is necessary since VLC code words can be produced at variable bit rates depending on the picture complexity. These values are written to a buffer memory. The reading from this memory is done at a fixed bit rate - in order to generate a fixed output bit rate. If the buffer becomes full, the quantization can be made coarser, by increasing the scaling factor of the quantizer. Note that in the case of re-multiplexing of VLC encoded data, one could employ cross input channel coding schemes so that the buffer limit is related to the aggregated bit rates and not simply to the bit rate of a single source, thus potentially allowing slightly higher quality (as there would be lower quantization error). However, this is not considered further in this thesis.

### 3.3.8 Motion compensation techniques

Motion compensation is based upon inter frame prediction – this is based upon detecting the displacement of picture details between two successive frames and emitting a motion vector to indicate the new position of these details in the current frame. Motion estimation is performed in macroblocks only on the luminance signals. A displacement vector is estimated for each macroblock, which corresponds to a 16x16 pixel block size.

The method of determining the displacement vector is called block matching. The reference block in the current frame is moved around its position within a search area in the previous frame until the best offset is selected on the basis of a measurement of the minimum error between the block being coded and the prediction. The measurement is accomplished with the DCT block values. Hierarchical block matching is an attempt to increase the size of the search area and at the same time keeping the necessary processing at the reasonable level.

There are three types of frames in the motion compensation prediction: An intra-coded I-frame has no reference to other frames and consists of intrablocks only. I-frames reduce spatial redundancy only and achieve a moderate compression. Predictive coded P-frames allow a higher data compression compared to I-frames. P-frames are coded with reference to a previous I- or P-frame. Coding errors can propagate between P-frames. The third type of prediction frames is called B-frame (Bidirectional predictive). These frames are coded both with reference to previous I- or P-frames and future I- or P-frames. They provide the most data compression, but do not propagate errors because they are not used as reference. However, in order to reconstruct a B-frame two (P- and/or I-frame) frames must first be decoded within a frame sequence.

A frame sequence is usually called a group of pictures (GOP) and allows the encoder to choose the right combination of frame types. The encoding order of frames is different from the display order, see figure 8. There is only one I-frame in each GOP. The first coded frame in a GOP must be an I-frame.

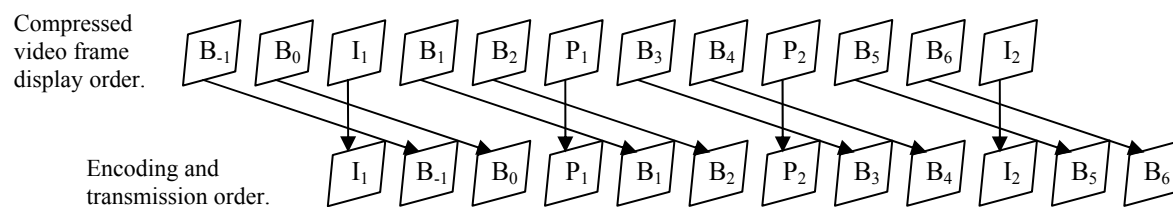


Figure 8: Frame order in video data (adapted from [3]).

### 3.3.9 Hierarchical structure of MPEG-2

The hierarchy of MPEG-2 coded video data consists of following six layers:

- DCT block layer consisting of 8x8 luminance and chrominance pixels transformed into DCT coefficients.
- Macroblock layer, consisting of a group of DCT blocks which correspond to a 16x16 coefficients. The macroblock header contains the information of its type and corresponding motion vectors.
- Slice layer is formed of one or several macroblocks. It can be the whole picture of a single macroblock. The header informs about the slice position within the picture and the quantizer factor.
- Frame or picture layer tells the decoder about what kind of frame is sent, i.e. I-, P-, or B-frame. The header indicates the frame transmission order allowing the decoder to display frames in the right order. There is also information about resolution, synchronization, and the range of motion vectors.
- GOP layer describes the size of the GOP and the number of B-frames between two P-frames. The header contains the timing information.
- Sequence layer includes information about the size of each picture, the aspect ratio, the bit rate for the picture in the sequence, and the buffer size requirements

All the features that can be described by these hierarchical subsets have been defined in different levels and profiles to make the decoding process easier and faster. DTT normally uses the Main Profile @ Main Level (MP@ML) for standard definition television which is associated with following parameters:

- Frame types: I, P, and B.
- Chroma sampling: 4:2:0
- Samples/line: 720
- Lines/frame: 576
- Frames/second: 25
- Maximum bit rate (Mbps): 15
- MPEG-2 MP@ML has no restriction to the number of consecutively coded B frames. In DVD, it is limited to no more than two B frames.

The MPEG-2 MP@HL profile was originally intended for HDTV applications, but nowadays many operators are using MPEG-4 AVC HP@L4 as their HDTV broadcast standard in order to save considerable bandwidth compared to MPEG-2 systems [13]. However, this thesis will not cover MPEG-4 coding.

### 3.4 Audio compression

Using the same approach for audio signals, a 16-bit-resolution stereo audio signal sampled linearly at 48 kHz will produce an audio data bit rate of 1.54 Mbps, while a multi-channel surround system (e.g. Dolby 5.1 surround) will produce a data rate of about 4.5 Mbps [3]. In a similar manner to the video signal, the audio signal redundancy is removed using source coding techniques. Psychoacoustic masking techniques are used to identify and remove irrelevant content. The MPEG-1 audio coding specification [25] contains three layers with increasing compression and increasing implementation complexity. MPEG-2 audio has a similar division into layers I, II, and III and uses the same coding algorithm. The difference is the extension in MPEG-2 to support multi-channel audio coding and surround sound with up to five full bandwidth channels. As mentioned in section 1.1, Teracom is utilizing MPEG-1 layer II as the primary audio channel in their DTT network. MPEG-1 layer II has proven to perform better than MPEG-1 layer III at high bitrates (192 to 384 kbit/s) and is generally more error resilient than MPEG-1 layer III, due to its lower complexity, so MPEG-1 layer II is considered optimal, and is the de facto standard, for broadcast applications. The typical bit rate for MPEG-1 layer II audio broadcasts in DTT network is 256kbit/s (128kbit/s per channel).

#### 3.4.1 Masking

Audio encoding exploits a property of the Human Aural Sensation (HAS) called **masking**. This means, if a tone of a certain frequency and amplitude is present, then other tones or noise of similar frequency, but of much lower amplitude, cannot be heard by the human ear. In this way, the louder tone masks the softer tone, and there is no need to encode the softer tone, thus reducing the data rate. This encoding property is a form of perceptual encoding, meaning that the perceptual quality of the reproduced sound is not affected. To illustrate this, in figure 9, Seppo Kalli in [3] considers a 1-kHz tone at a sound pressure level of 45 dB, which will raise the hearing threshold to 27 dB, meaning that sounds below 27 dB are inaudible. If we use the 6 dB-per-bit rule, we will only need 3 bits to encode this tone ( $45-27=18$  dB;  $18/6=3$  bits). The masking effects exist both in frequency (called spectral masking) and in time (called temporal masking). Temporal masking means that a loud tone of finite duration will mask a softer tone that quickly follows it. [2] Even if the masker tone suddenly disappears, the masking threshold does not disappear simultaneously, it takes some time before the masked tone will be audible. These effects are called pre- and postmasking. Usually postmasking lasts longer than premasking. The bandwidth, around a masking tone, over which spectral masking occurs is called the critical bandwidth.

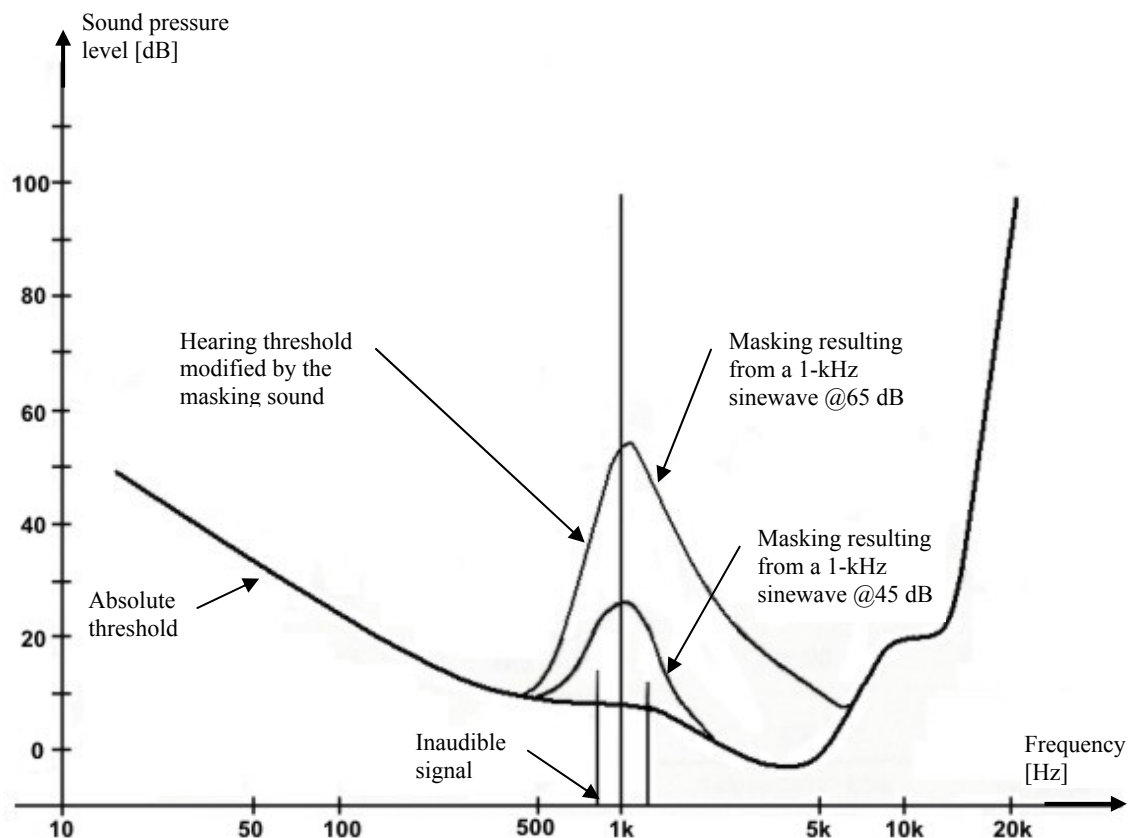


Figure 9: Absolute hearing and frequency masking thresholds (adapted from [3]). The frequency range of sound perception is between 20Hz and 20 kHz. Signals below the absolute threshold in sound pressure level are inaudible.

The basic structure of a perceptual encoder consists of a filter bank, a bit allocator, a scaler, a quantizer processor, and a data multiplexer.

### 3.4.2 Filter bank

The aim with a filter bank is to try to simulate a psychoacoustic model of HAS and decompose the signal spectrum into subbands. There are three types of filter banks:

1. The subband bank divides the signal spectrum into equal-width frequency subbands, similar to the HAS process of dividing the audio spectrum into critical bandwidths. There are 32 subbands in MPEG layers I and II. A polyphase quadrature mirror filter (PQMF) is one example of a subband filter.
2. The transform bank uses a modified DCT (MDCT) algorithm to convert the time domain audio signal into a large number of subbands.
3. The hybrid filter bank combines subband filters with MDCT, thus providing a finer frequency resolution, such as the one used in MPEG layer III (MP3).

### 3.4.3 Bit allocator

The bit allocation is calculated from the difference between the computed spectral signal envelope and computed masking curve. This difference determines the maximum number of bits necessary to encode all spectral components of the audio signal. See, the example in section 3.4.1. MPEG encoders use a forward adaptive bit allocation process, meaning that bit allocation calculation is made based upon the input signal in the encoder only. The masking threshold is calculated in order to determine the level of noise which each band in the filterbank is allowed to contain. This information is further used in the bit allocation.

### 3.4.4 Scaler and quantizer

Scaling is carried out by the block floating-point system which normalizes the highest value in a block of data to the full scale. All block data values are then quantized with a quantizing step size determined by the bit allocator. A block of data is made of 12 consecutive samples and an audio time frame consists of  $12 \times 32 = 384$  samples, which corresponds to 8ms of audio at a 48 kHz sampling rate in layer I and 24ms in layer II, the later consisting of  $12 \times 3 \times 32 = 1152$  samples. In MPEG layer I, a 512-sample FFT is used to accurately analyze the frequency and energy content of the incoming audio signal. In MPEG layers II and III, a 1024-sample FFT is used.

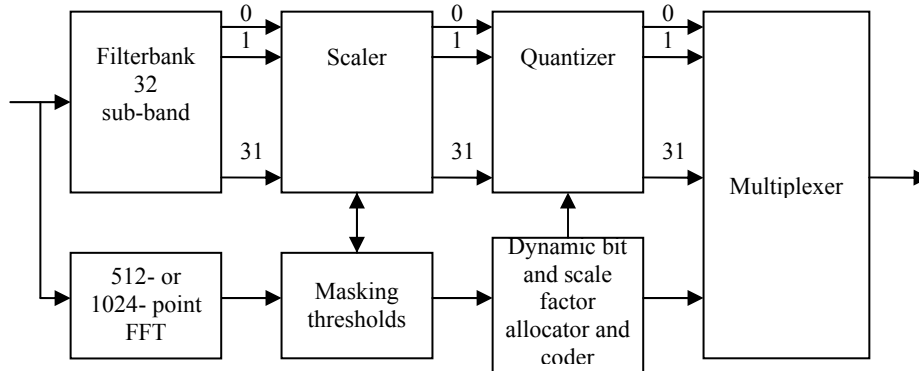


Figure 10: Block diagram of an MPEG audio encoder. (adapted from [3])

### 3.4.5 Multiplexer

Blocks of 12 data samples are multiplied by the corresponding scale factor and input into the bit allocator to form audio frames in the encoded bit stream.

### 3.4.6 MPEG Layer II characteristics

The audio MPEG Layer II frame structure is shown in figure 11. It starts with a 32 bit header containing a synchronization code word, information about actual sampling frequency, data rate, type of emphasis, and type of MPEG layer. It is optionally allowed by a cyclic redundancy check field providing protection of the header information. After this are a bit allocation field, Scale Factor Selector Information (SCFSI), scale factors, and the subband samples. The audio frame ends with an ancillary data field. This ancillary data field may contain program associated data or other messages. The size and structure of this final field can be defined by the user. The length of the frame in bytes is calculated as follows:

$$\text{Length} = 1152 * \text{bitrate} / \text{sampling rate} / 8$$

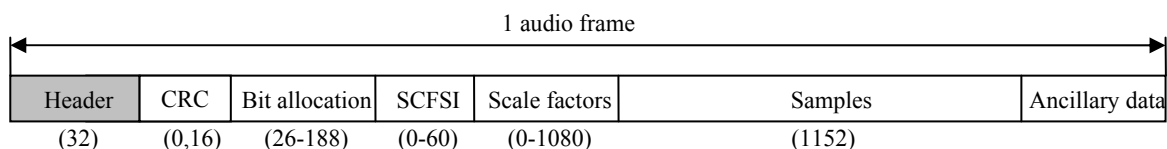


Figure 11: Audio MPEG layer II frame structure (adapted from [3]).

In MPEG-2 multichannel audio, in addition to left and right loudspeaker channels there are also two surround loudspeakers channels (Ls and Rs) and a center loudspeaker channel C. Instead of one multi-channel program, a second independent stereo pair may be transmitted. This could be deployed in services that require bilingual programmes or multilingual dialogues or commentaries in addition to the main multichannel service. The standard supports the transmission of up to 7 multilingual/commentary channels. The transmission of the MPEG-2 audio multichannel is realized by exploiting the ancillary data field in MPEG-1 audio frame. One of the features of MPEG-2 audio is its

backward compatibility with MPEG-1 coded mono, stereo, or dual channel audio programmes meaning that an MPEG-1 audio decoder is able to properly decode the basic stereo information of a multichannel program. This feature is achieved by an appropriate downmix of the audio information in all five channels, thus creating two channels (Lo and Ro). In later developments of the multichannel coding standard, it was decided to include a non-backward compatible CODEC in order to provide a significant quality improvement over backwards-compatible CODECs. One such CODEC is Dolby Digital AC-3.

### 3.4.7 AC-3

AC-3 or Dolby Digital is an audio compression standard containing up to six discrete channels of sound, with five channels for normal-range speakers (20 Hz – 20 kHz) (Right front, Left front, Center, Right rear, and Left rear) and one channel (20 Hz – 120 Hz) for the subwoofer which provides low frequency effects [17]. Hence AC-3 is very similar to the MPEG-2 layer II standard. The AC-3 audio coding scheme has been selected as the default audio standard for Advanced Television Systems Committee (ATSC) broadcasting (an alternative standard to DVB-T, used in North America among others). However, AC-3 is one of the audio standards which can be used with DVB-T.

The AC-3 encoding process is somewhat similar to MPEG, but with some differences. The block diagram is shown in figure 12. First, there is a transformation of audio samples to the frequency domain, using a 512-point MDCT filter bank. Next, a block floating-point system converts each transform into an exponent and mantissa pair. The mantissa is a part of the floating-point number that contains its significant digits. For example, the number 123.45 can be represented as a decimal floating-point number with integer significand 12345 and exponent  $-2$ . The mantissas are quantized with a variable number of bits, based on a parametric bit allocation model which uses a psychoacoustic masking to determine the number of bits for each mantissa in a given frequency band.

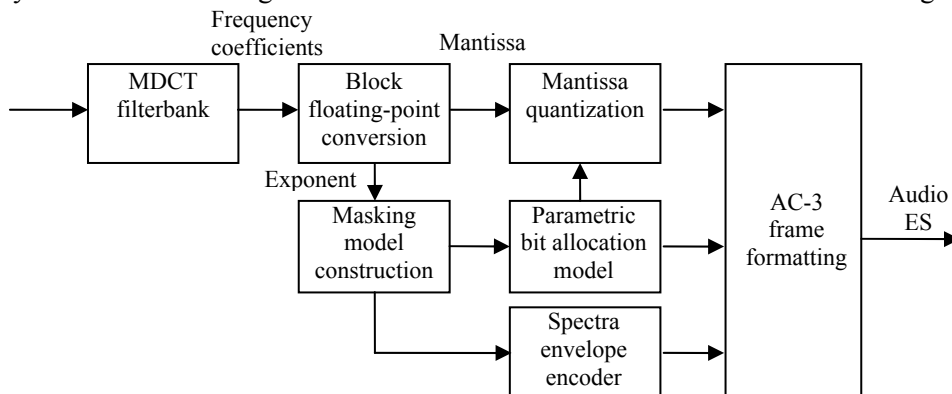


Figure 12: AC-3 encoder block diagram (adapted from [3]).

The spectra envelope act as a scale factor for each mantissa, based on the exponent value. Both the encoded spectral envelope data and quantized mantissa data are formatted into an AC-3 sync frame consisting of six audio blocks. Figure 13 shows an AC-3 sync frame. Each frame consists of  $256 \times 6 = 1535$  audio samples (i.e. with blocks of 256 samples each). Auxiliary block at the end of the frame is reserved for control or status information of system transmission. In each audio block there are different kinds of flags (block switch flags etc.), the data of exponents, bit allocation, and mantissas. In AC-3, there is like in MPEG a capability of down-mixing the signal to stereo or mono only.

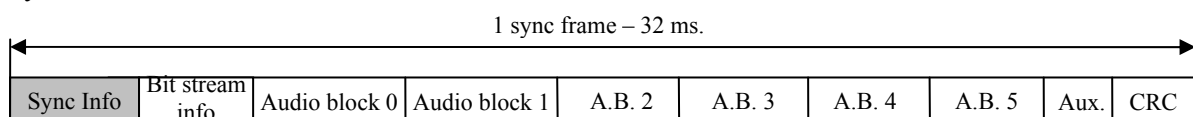


Figure 13: AC-3 sync frame [3].

Comparison between MPEG layer II and AC-3:

<u>Audio schemes</u>	<u>Total bit rate (kbit/s)</u>	<u>Filter bank</u>	<u>Frame length @48kHz (ms)</u>	<u>Bit rate target, (kbit/s per channel)</u>
MPEG layer II	32-448	PQMF	24	128
AC-3	32-640	MDCT	32	64

### 3.5 DVB Systems

The main enhancement of the MPEG-2 standard from the MPEG-1 standard is the introduction of a system layer specification, in this way forming a hierarchy of different data streams. Independent audio, video, or data sequences form independent data streams called Elementary Streams (ES). The system layer defines the combination of separate audio and video streams into a single stream for storage (Program Stream) or transmission (Transmission Stream). It also includes the timing and other information needed to demultiplex the audio and video streams and to synchronize the audio and video after decoding.

In a Packetized Elementary Stream (PES) packetizer, elementary streams are separated into packets of variable lengths. Each PES packet contains data from one ES (see figure 14).

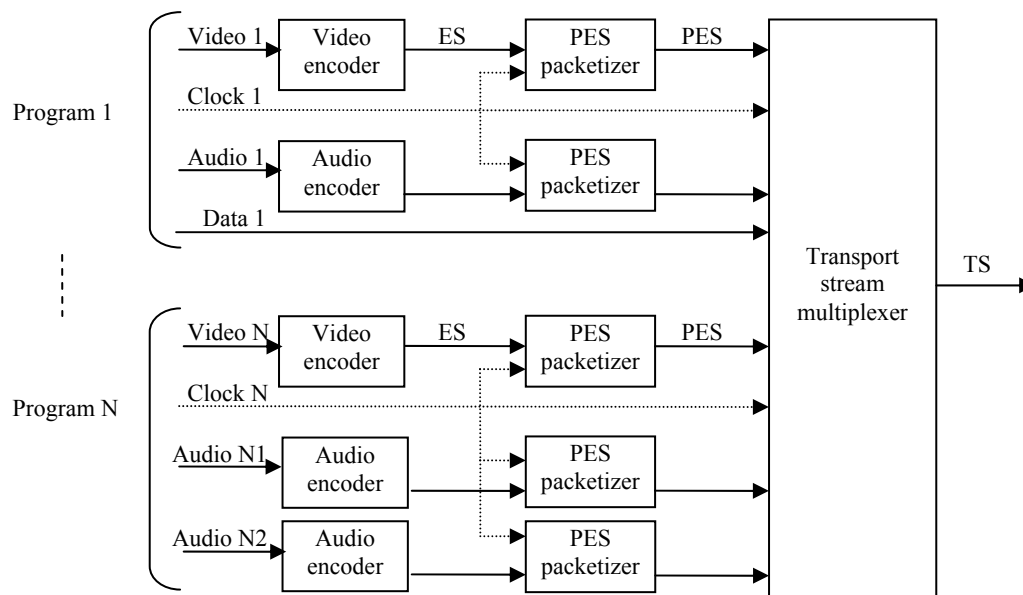


Figure 14: MPEG-2 TS multiplexer system. [4]

The synchronization of audio and video is solved by using presentation time stamps (PTS) and decoding time stamps (DTS). These time stamps define when a presentation unit should be decoded and displayed. For video a presentation unit is a picture and for audio it is a set of subband samples sent in one audio frame. In audio, the presentation and decoding is done simultaneously. In video, depending on whether it is an I- or P-frame, the presentation and decoding time may be different. I- and P-frames are decoded before a B-frame.

#### 3.5.1 Transport Stream

A Transport Stream is defined for transmission networks that may suffer from occasional transmission errors, such networks includes: DVB-T or DVB-S (DVB-Satellite). PES packets from various elementary streams are combined to form a Program. A Transport Stream may include several Programs, each with its own time base. In general, relatively long variable-length PES packets are packetized into shorter TS packets with a fixed size of 188 bytes. The reason is that a fixed packet size makes error recovery easier and faster, but it has a higher cost in terms of per packet overhead (which leads to a higher overall overhead).

Each packet starts with a TS header followed by an Adaptation field (see figure 15), followed by data from one PES packet. In the TS header there is information consisting of synchronization, flags, error detection, timing, etc. The Packet Identifier (PID) is used to distinguish between different elementary streams and different Program Specific Information (PSI), see section 3.5.2.



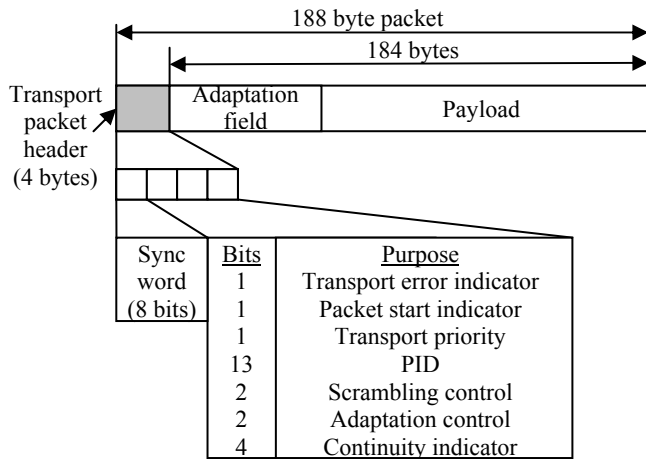


Figure 15: Transport packet structure in MPEG-2. [3]

In the Adaptation field the Program Clock References (PCR) are transmitted, which are the samples of the system clock in the encoder. These samples are used to synchronize the system time bases of the encoder and the decoder. The Adaptation field is optional and has a variable length. Note that if the adaptation field is longer, then the Payload field must be shorter - as the overall packet length is fixed. The adaptation control field in the TS header indicates the presence of an adaptation field or payload.

### 3.5.2 Program Specific Information

The program descriptions and the assignments of PESs and PIDs are contained in specialized TS streams called Program Specific Information (PSI). PSI is structured into four tables: The first one is called Program Association Table (PAT) and it has always PID 0. The PAT lists the number of all programs contained in the TS together with the PID value of the TS packets which includes the Program Map Table (PMT) section of each program.

PMT is the second table in this hierarchy. This table (see figure 16) includes PID values of each elementary stream packet. In Appendix B, an example is presented of a whole TS representing all PIDs, their stream type, etc.

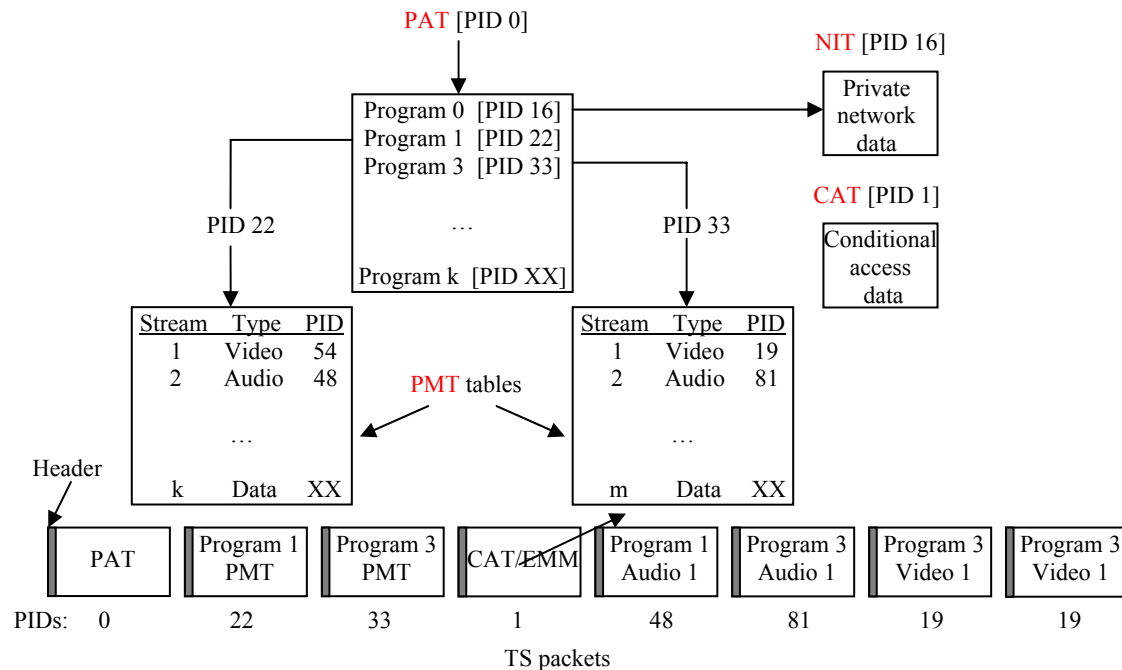


Figure 16: Audio, Video, and Data packets in a MPEG-2 TS stream. [14]

The third table is called the Conditional Access Table (CAT) and it provides information about the scrambling systems and their PIDs, this is called an Entitlement Management Message.

The fourth table is called the Network Information Table (NIT) and it is a private table not specified in MPEG-2. In general, this table contains physical network parameters, such as channel frequencies, modulation characteristics, etc.

### 3.5.3 Timing and synchronization

The system layer takes care of the synchronization of the encoding and decoding process. The delay between these processes is assumed to be constant, even though the delay through each of the encoder and decoder buffers is variable. Both the encoders and multiplexers use the same timing reference called the System Time Clock (STC). In TS, the STC samples are transmitted as Program Clock Reference (PCR) values in the adaptation field. The frequency of the STC is 27 MHz. Reconstruction of STC in the decoder is done with the help of time stamps transmitted in the system stream (see figure 17). Each program in a TS may have its own time base and consequently its own PCR values, but the programs may also share the same time base.

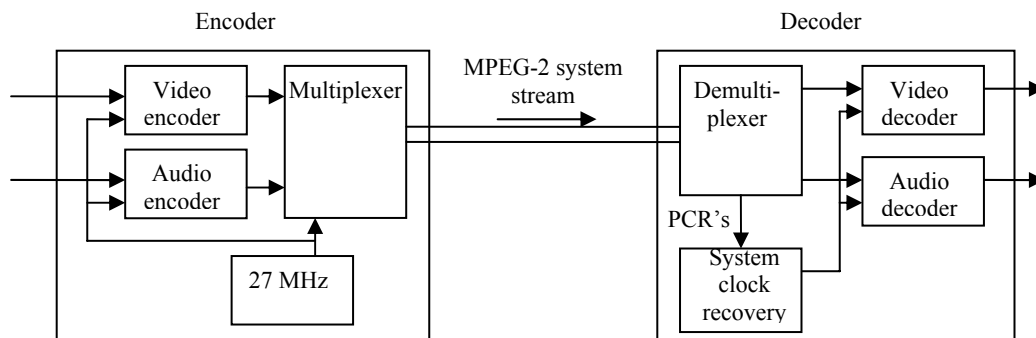


Figure 17: System time clock recovery. [4]

The transmission delay of the STC may vary, but the distance between two consecutive PCR values should not exceed 100ms according to ISO standard IEC 13818-1:2007 [26] and 40ms according to DVB standard [27]. This transmission delay variation is also called jitter. That is why System Clock Recovery needs to be performed in the decoding process. A Phase Locked Loop (PLL) is usually used to smooth the jitter in the received PCR. In this PLL, locally generated PCR values are compared to the received PCR values from the TS.

## 4. Methods and analysis

### 4.1 Introduction

This chapter will describe existing products which Teracom is considering as potential equipment for their supervision system, along with some new ideas of how to solve this problem. As mentioned in the introduction there are two approaches for monitoring that will be initially considered. Both approaches (see sections 4.2 and 4.3) have their advantages and disadvantages, but the key metrics for selecting one over the other will be their ability to correctly detect failure and their cost.

In order to evaluate these products, there are some questions that must be asked: How reliable is this technique? Is it really worth the cost if there has to be a monitoring of 33 TV channels at 54 different sites across Sweden? In the next part we will try to give answers to these questions.

### 4.2 IdeasUnlimited

IdeasUnlimited is a British company that has developed a product family named ContentProbe [7]. Their products use a technique called Media FingerPrinting which allows the system to compare video and audio signals in real-time. The hardware consists of a box with a web enabled broadcast network device. The input signal is analogue composite video or Serial Digital Interface. The software runs under an embedded Microsoft Windows XP operating system.

Three different units are available:

1. The Fault Tracker (FTE1000) unit monitors: video is present, the video froze, there is audio silence, and presence of an audio tone. These parameters have a limited ability to indicate if the TV content is correct or not.
2. The ContentProbe Verification (FTV1000) unit makes fingerprints of any audio and video signal, which it then monitors, and compares with other signals in real time.
3. The Compliance Recording (FTS1000) unit stores the audio and video input into a Windows Media 9 format and allows clients to view the stream over a LAN or WAN.

The client software is based on Omnibus System's G3 desktop [28] and can be used to configure every unit.

#### 4.2.1 Test bench

Several tests were performed in order to evaluate if products from IdeasUnlimited could be used in Teracom's supervision monitoring system. The test bench examined their operation in two types of modes: Single-ended mode, where no comparison of video content in the Device under Test (DUT) was performed and double-ended mode, where the DUT was comparing reference and test input.

#### 4.2.2 Single-ended mode

In single-ended mode failures such as black screen detection, frozen frame detection, hard compressed video content (i.e., re-encoded to a video bit rate of 0.5-1.0 Mbit/s), detection of video decoding errors in the content, and audio silence detection were tested and evaluated using the IdeasUnlimited FaultTracker unit. This product is the least expensive of the three and provides only basic monitoring of the content. In figure 18 the test bench is illustrated. The monitoring of content is done within the FaultTracker. On a client PC the monitoring status can be viewed; along with the possibility of viewing screenshots and live video streaming. The time server is an NTP time server and it is necessary for the synchronization of system time and dates of all the equipment used in this test.

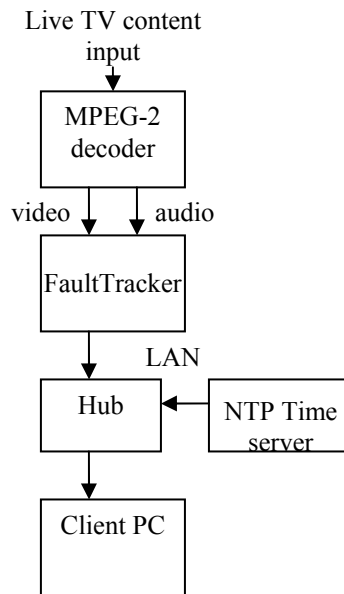


Figure 18: The test bench for single-ended mode test.

Frozen frame detection of the video signal is measured in percent. Thus 100% percent is a totally frozen frame. Slow movement content could generate a false alarm, but there is a possibility to configure the duration before an alarm should be generated. Black screen detection acts in the same way as frozen screen detection.

In order to create video bit errors, equipment that could attenuate the RF input signal level was used. A noise generator could be used instead, but the result would be equivalent since it is Carrier to Noise ratio that we are interested in, and which when reduced generates the desired amount of bit errors. When the content on the screen was frozen, video bit errors could result in apparent movements on the screen, thus the content was classified as a moving picture, which is of course completely incorrect. Hard compressed content was created with an encoder by changing of the video bit rate value. The result was that the FaultTracker could not detect this type of content as a failure until it became frozen.

Audio silence was possible to detect, but since most of the TV content during the daytime is silent (e.g., SVT channels are often broadcasting only a TV schedule with no sound, during the day time), the system had no ability to detect whether the silence was intended in the content or it was an effect of signal distortion. In other words, a lot of false alarms were generated. For several reasons Teracom is not able to get information from content providers about when the content actually is silent, hence the audio silence detection is unused. One of the reasons is that the monitoring should be independent from the information of content providers in case the information is wrong.

The conclusion was that single-ended mode failure detection is not sufficient for Teracom's supervision monitoring system.

### 4.2.3 Double-ended mode

In the double-ended mode two live content inputs were used. The reference input was from the TV tower feed (Kaknäs) through a fiber network. The second one, the test input was from a DTT antenna (in this case the one in Nacka). Verification was performed on the same TS multiplex (Multiplex 1 containing SVT channels) and the TV channel carrying SVT2 was used for comparison.

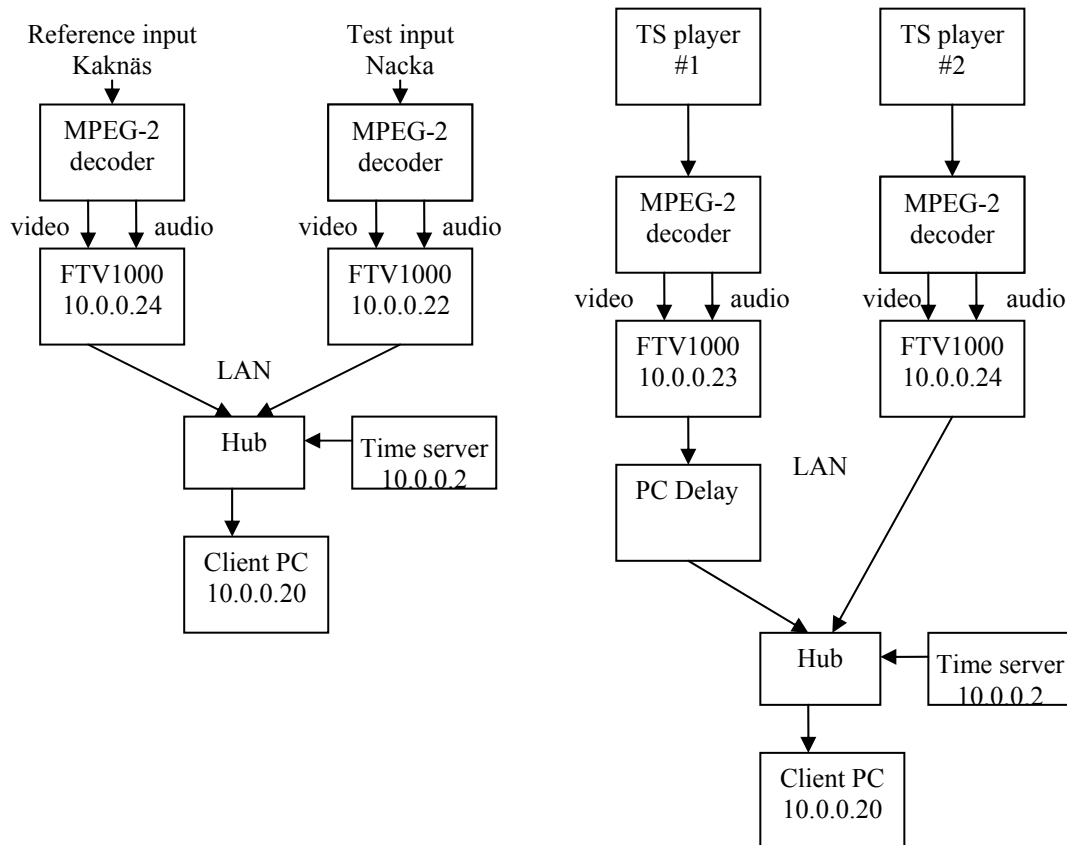


Figure 19 (a) and (b): Two test benches for double-ended comparison tests. In (a) two live streams are compared. In (b) two recorded TS streams are compared.

First, the test bench described in figure 19 (a) was used for testing (in principle) the same conditions that were tested in single-ended mode as described in previous section. The comparison was performed in IdeasUnlimited ContentProbe Verification (FTV1000) unit with the IP address 10.0.0.24. A content FingerPrinting technique was used in order to detect unmatched content. It took approximately 12 seconds to detect a difference in content and 5 seconds to detect that the content is the same. In the test bench described in 19 (b) it was possible to delay content for comparison, by using two pre-recorded TS streams and different delay times.

A test with down-scaled video content was performed in the following way: 720x576 (5Mbit/s SVT2 ABC) and 352x288 (4Mbit/s SVT2 ABC) were compared with each other. The system is able to match the contents even if the resolutions of the video are different. The verification of the hard compressed video with standard video content was performed as follows: the original bit rate (4.9-6.0 Mbit/s) for a service (SVT2 ABC) was compared with the same service compressed to 1.5 Mbit/s and 1.0 Mbit/s. The later bit rate comparison (at 1Mbit/s) resulted in an alarm, while the previous one (at 1.5 Mbit/s) did not. The explanation is that at 1.0 Mbit/s the content had too many blocking artifacts, and consequently was difficult to recognize as being the same as the original content. The FTV1000 was also able to detect different aspect ratios (4:3 or 16:9). The detection of audio mismatch worked as well.

The test bench in 19 (b) could also detect different failures on the Ethernet level (e.g. the loss of packets or a decrease of available bandwidth); since the verification is performed at one of the FTV1000 units, it is very important that the internet connection between these two does not incorrectly affect the monitoring properties.

A conclusion of testing the double-ended mode is that it is much more effective than the single-ended mode, but still far from acceptable. The system may generate alarms which are incorrect and at the same time miss errors such as visible bit errors or audible errors. The price of each unit is high and for monitoring of all broadcast TV content Teracom would require very substantial investments since the monitoring has to be performed at sites spread over the whole country.



Figure 20: Client PC software displaying content with visible bit errors (Nacka (streaming)), but classified by the equipment as matched video content; shown in the top left corner of the figure. The green colored boxes in the figure indicate a correct detection, while the blue colored icons indicate events that need attention from the user. The four boxes in the top right corner of the screen shot show the different inputs to the system.

### 4.3 Agama

Agama [29] is a Swedish company based in Linköping that has developed a monitoring solution for IPTV. Instead of using analog input for analyzing the audio and video content, Agama's solution monitors the services within a TS containing IPTV broadcast contents. The main advantage compared with IdeasUnlimited solution is that the monitoring occurs on the packet level, thus there is no need to decode all the services in a Transport Stream.

#### 4.3.1 Test bench

The equipment from Agama consists of a PC that runs Fedora Linux OS [30]. There are two PCI cards installed: DVB-T (a DVB-T receiver) and DVB-ASI (an asynchronous serial interface - generally used for coaxial cable connection to/from satellite transmission systems, interfacility links, or telecommunication networks). There is no particular reason to have different kind of cards, this was done simply to test different kinds of inputs and to compare them. The software consists of Agama Analyzer and Verification where the first is able to analyze the TS up to the macroblock layer (see section 3.3.9) and the latter verifies that two TS's match.

#### 4.3.2 Agama Analyzer

The Agama real time analyzer (shown in figure 21) is able to monitor TS packets and detect errors in the stream. Errors are described in the message table and a color graph presents the content error status over time. Different colors indicate how seriously the errors affect the subjective video quality. The analyzer is also able to detect Frozen Frame (in the figure marked as FF), when the video content is frozen.

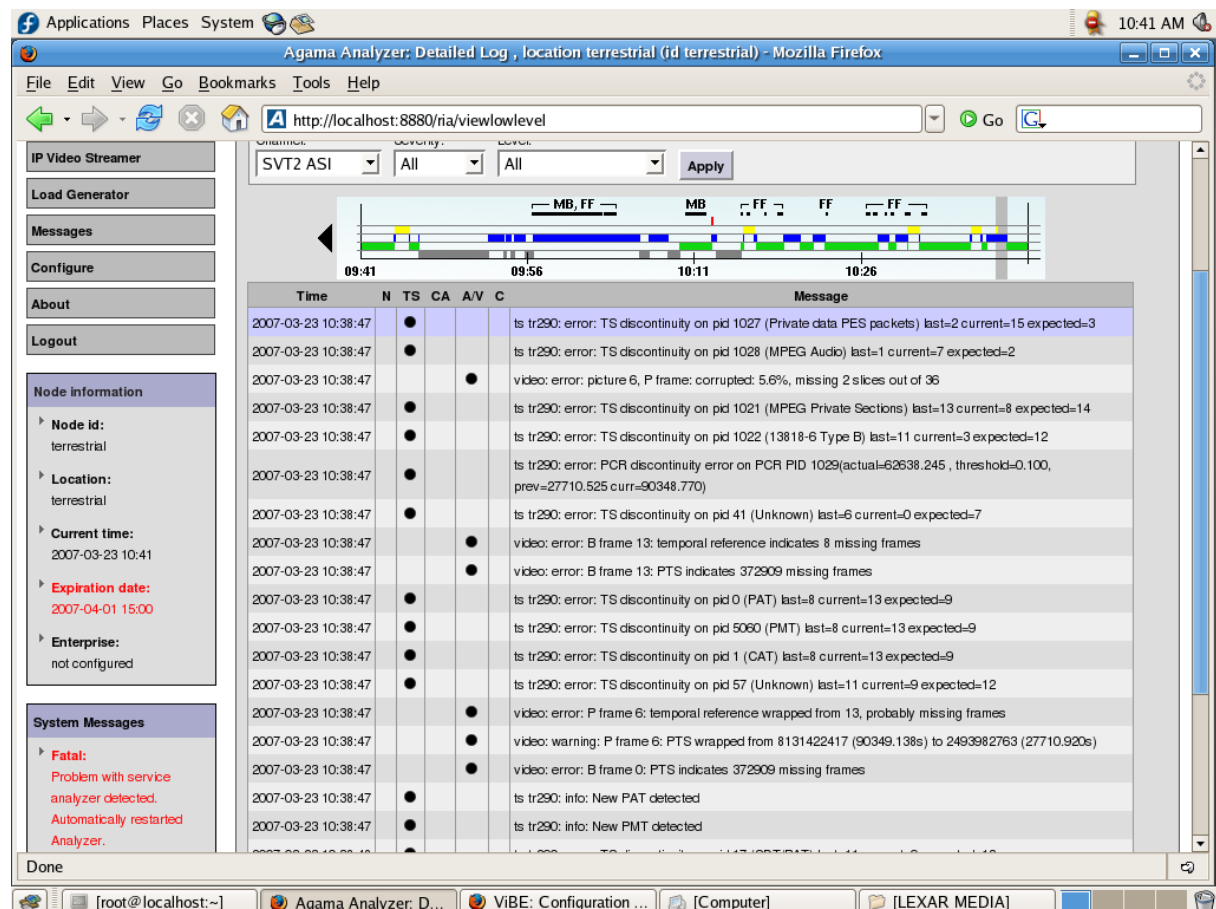
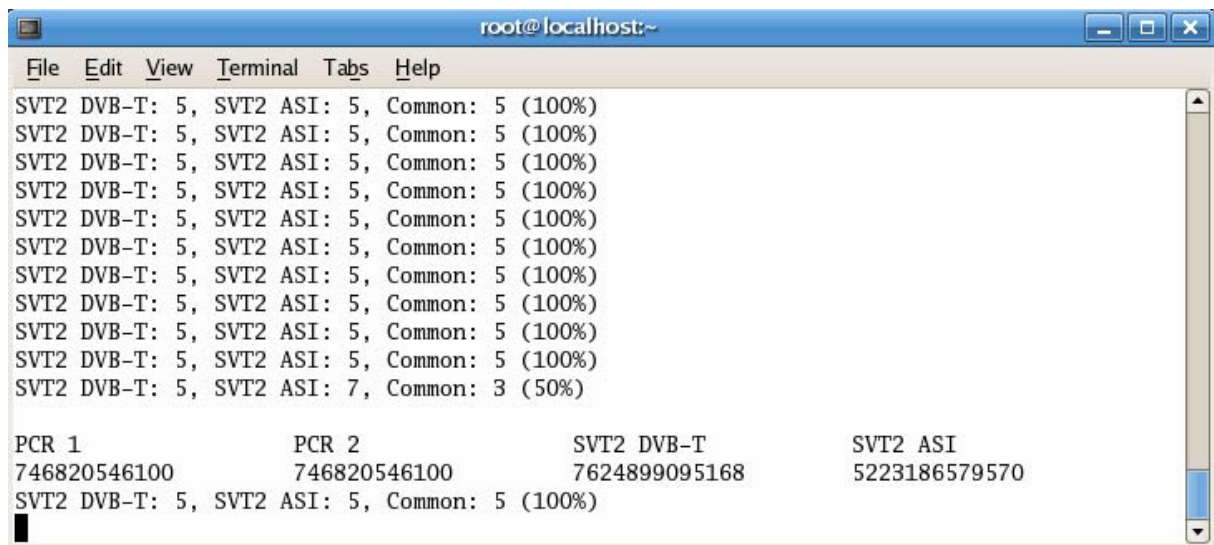


Figure 21: A screenshot from Agama's Real Time Analyzer.

### 4.3.3 Agama Verifier

Although this program is still under development, we had a chance to test the first version of it. The basic idea is to compare two services carried within TS's: one coming through the DVB-T card and the second one coming through the ASI card.

The verification of these two services is done with help of the PCR values (described in section 3.5.3) and the checksum values (Agama calls them "finger print values") from the packet's header. The verification program receives two finger prints values from both inputs and removes the finger print values that are exactly the same, i.e. with the same PCR and finger print value for both input 1 and input 2. Then, it checks for the finger prints that have the same PCR value, but different finger print values and displays them. An example is shown in figure 22. The percentage displayed indicates how equal the two packet streams are. 100% means that all the packets completely matched each other.



```
root@localhost:~
File Edit View Terminal Tabs Help
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
SVT2 DVB-T: 5, SVT2 ASI: 7, Common: 3 (50%)

PCR 1          PCR 2          SVT2 DVB-T      SVT2 ASI
746820546100   746820546100   7624899095168   5223186579570
SVT2 DVB-T: 5, SVT2 ASI: 5, Common: 5 (100%)
```

Figure 22: A screen-shot from the basic Agama Verifier program.



## 4.4 Investigation of DCT coefficients and Scale factors usability

### 4.4.1 Introduction

This section will present my own investigation utilizing the background knowledge that has been presented earlier in this report. The basic idea is to use DCT coefficients (described in section 3.3.2) in order to detect different kinds of video content. Detecting of the audio content is done by examining the Scale factors.

### 4.4.2 Test bench for video

In order to extract DCT coefficients from a transport stream, a modified program from the open source project called “mpeg2dec version 0.4.1” [22] was used. Basically, the program decodes an MPEG-2 video stream and extracts the DC coefficients for luminance blocks (Y:B0, Y:B1, Y:B2, and Y:B3) for all three type of frames (I, P and B-frames). The code for this program is written in C.

The original transport stream of video content was recorded from the live network by using Acterna (now JDSU) Transport Stream Recorder (v1) [31] and demultiplexed by using Interra Systems Vega H264 Transport stream analyzer (v6.1) [32].

Since bit errors can propagate from I-frames to both P- and B-frames (see section 3.3.8), we limit our investigation to study only the DC coefficients from I-frames. The information from I-frames should be enough to get an overview of what DC coefficients can tell us about the video content.

From the hierarchical structure of MPEG-2 (see section 3.3.9) we know that each frame contains a number of slices, namely 36. Further, each slice is built up of 45 macroblocks and each macroblock has 4 luminance and 2 chrominance blocks. In each block we have 8x8 DCT coefficients, 64 values (see section 3.3.2), where the top left corner value is the DC coefficient and all the other values are the AC coefficients. We also know that most of the signal information after the transformation tends to be concentrated in the DC value; thus we are only interested to study this particular value of each block.

The output from “mpeg2dec” is the DC coefficient and its differential code for each luminance block. A DC coefficient’s differential code describes the difference between the DC coefficient value of current decoded block and that of the previous block. It can be useful in case we need to compare a current value with the previous one.

Looking at the DC coefficients over time (represented by I-frames) will give us a view of how these coefficients are changing in different video streams.

#### 4.4.3 Test cases for examination of video content

Similarly to the test benches of IdeasUnlimited and Agama, we are interested in detection of different kinds of errors by examining the DC coefficients.

Four different video contents were chosen for examination. The first one was static content (frozen frame) where the content was only changing every 10<sup>th</sup> second. This video content was recorded from Swedish TV channel SVT1. This video content is called “TV-tablå” and is usually broadcasted during the daytime. Below, there are two screen shots from it.

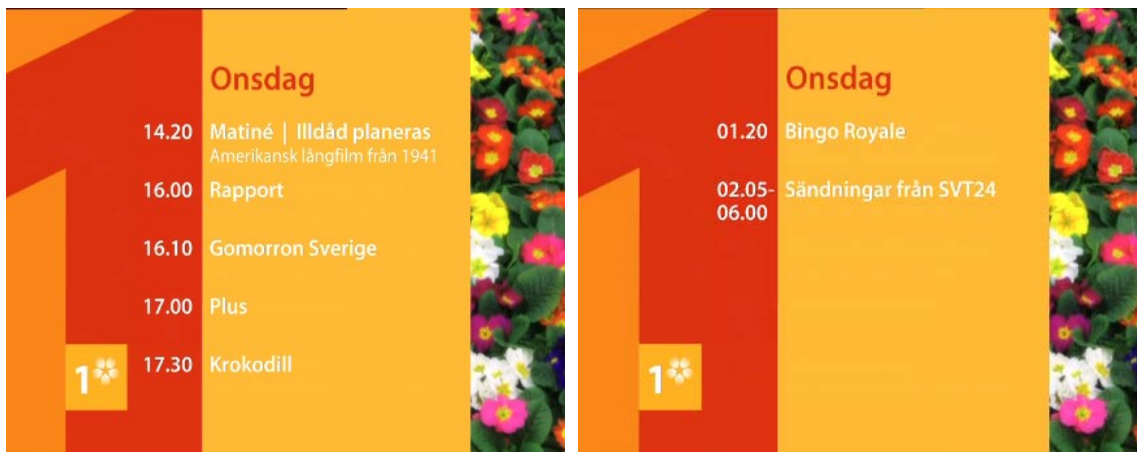


Figure 23: Screen shots from the “TV-tablå” (SVT1).

Examining the DC values of this content readily shows that both were frozen picture for 10 second periods, with some change of a certain part of the content every 10 seconds. By looking at the mean values of the DC coefficients from each I-frames (see figure 24), we can see that DC coefficients quickly tells us that frozen video content generates the same mean value for each I-frame while the content is static, which seems logical. When the video content is changed (see I-frames 12 and 30), the mean value changes and during this time the content is not frozen anymore. From a single-ended point of view this could be a simple way of detecting frozen video content, which does occur from time to time in digital TV broadcasts.

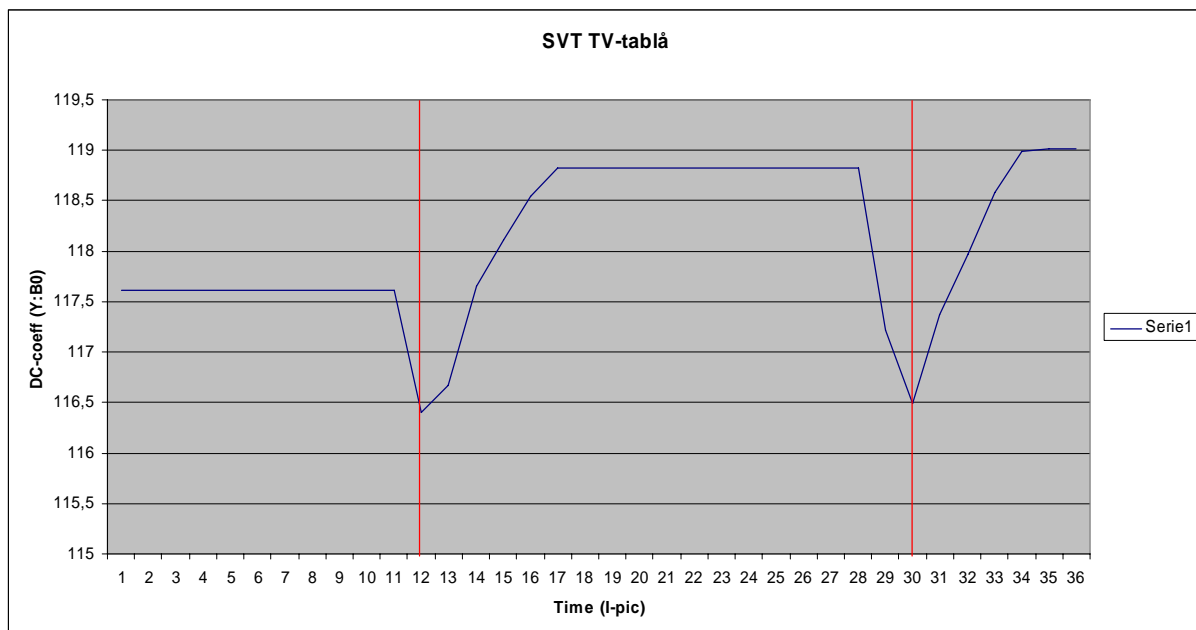


Figure 24: A diagram showing the mean value of DC-coefficients over time (based upon I-frames).

The next case is similar video content to previous case. In this video approximately 98% of the picture is frozen while the rest of the image has only small changes. Figure 25 shows a test picture from TV channel SVT1 in 16:9 aspect ratio. This video content has only two movements: the clock, with numbers changing every second, and the moving stripe, down to the left of the circle. This line is moving from left to right, indicating for the viewer that the picture is not frozen, since it is a test picture.

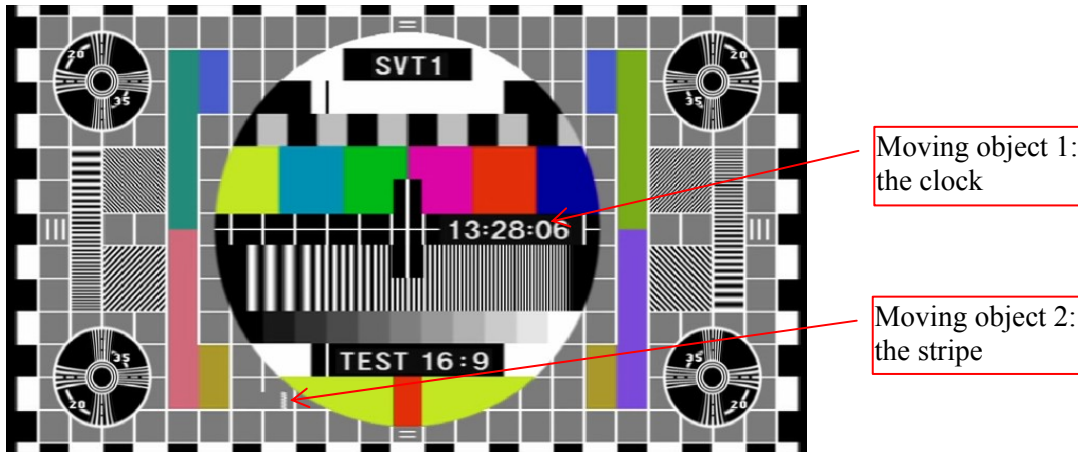


Figure 25: Screen shot from a video test picture (SVT1).

We examine this frame in the same way as the previous case. By looking at a mean value, we can see that despite most of the picture is not moving, small movements are being detected, see figure 26.

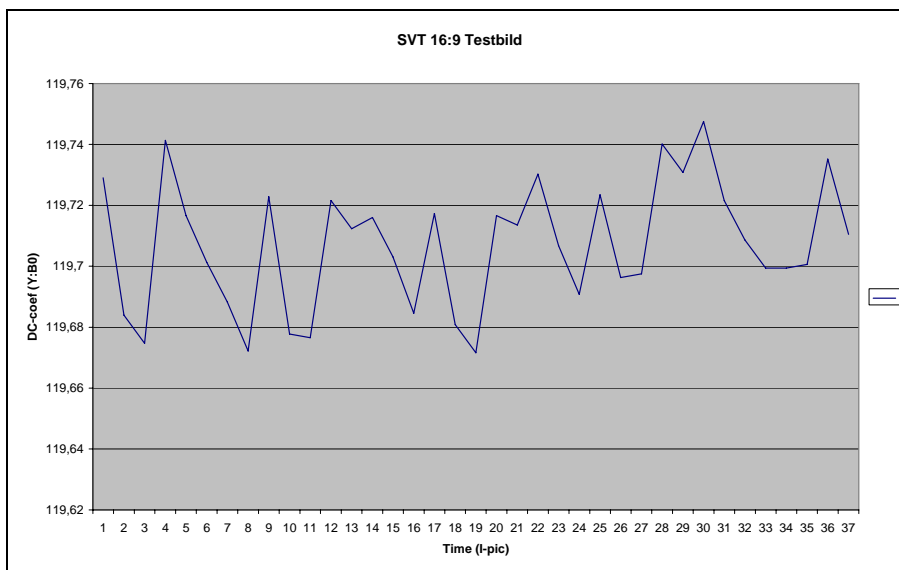


Figure 26: A diagram showing the mean value of DC-coefficients over time (based upon I-frames).

The test pattern above is just an example of a correct video with limited visual changes. A typical screen image of a news reader (also know as “a talking head”) with minimal movement (often just the lips) will result in a similar DC coefficient change as a function of time.

The next two cases concern video content with bit errors. Bit errors were generated by attenuating the RF signal to the receiver. The attenuation was chosen such that the video would still be possible to view when visible artifacts appear.



Figure 27: Screen shot from a video with bit error, represented as green blocks. (TV4).

The video in figure 27 has small movements, with just the TV host moving, while the background is still. This content is similar to the previous two cases; thus we use same method, i.e. examine how mean value of the DC component varies.

Bit errors that we are trying to detect are also called syntax errors. They can be divided into different groups [19]. One such group of syntax errors occurs when the decoded data is beyond the range of allowed values in the MPEG-2 standard, e.g. the value is negative. The video shown in figure 27 has such values, thus instead of calculating the mean value, we calculate the variance in value. This video sequence contains bit errors in following I-frames: 2, 7, 28, 31, 34. These are detected in figure 28. For example figure 27 is I-frame 2 below. Still, this method proves to be insufficient when the video content contains lot of movement and/or scene changes.

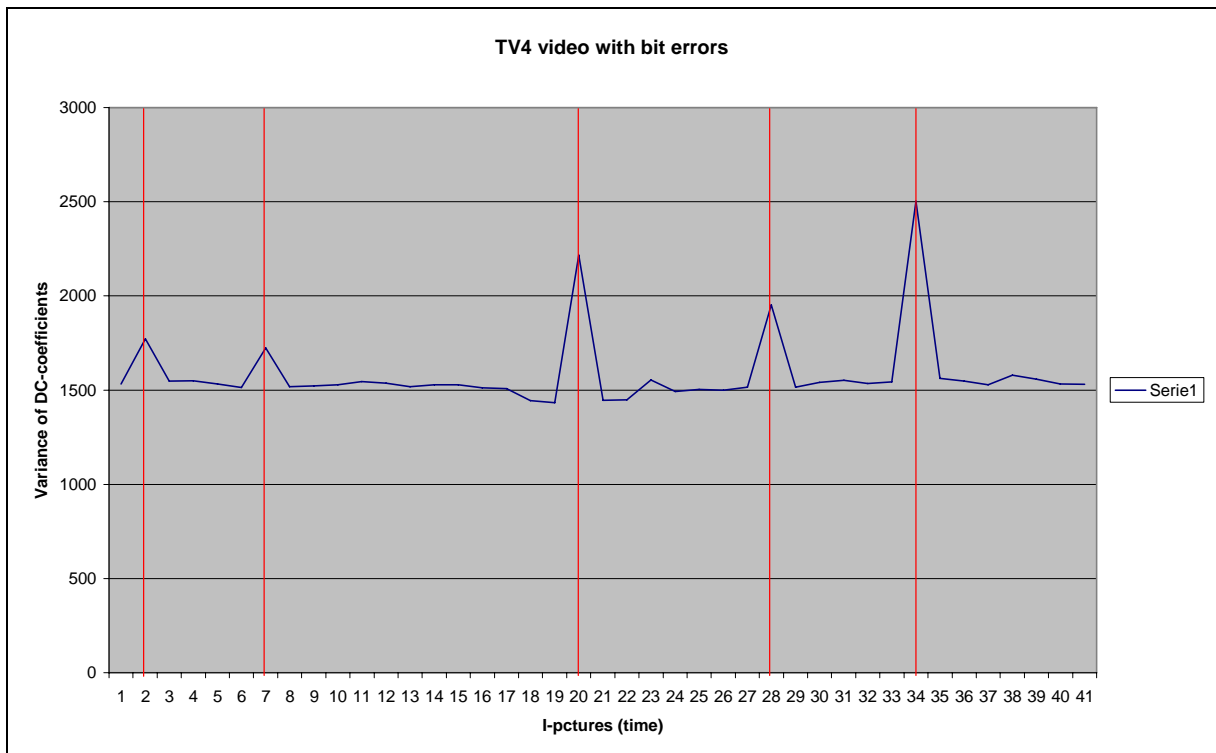


Figure 28: The figure shows variance values of DC-coefficients over time (based upon I-frames).

The fourth case contains a music video sequence (recorded from “The Voice TV”) with lot of movements and bit errors. Figure 30 shows the variance of the DC coefficients as a function of the I-frame number. We see that we are not able to detect the presence of bit errors using this method in this video content.

The screen shot in figure 29 is split up into macroblocks (vertical) and slices (horizontal). There are totally 45 macroblocks and 36 slices, as mentioned previously. Figure 29 shows bit errors that appear in slice 20 starting from 9<sup>th</sup> macroblock (coloured green in fig.29) and continuing to the last macroblock on this slice.

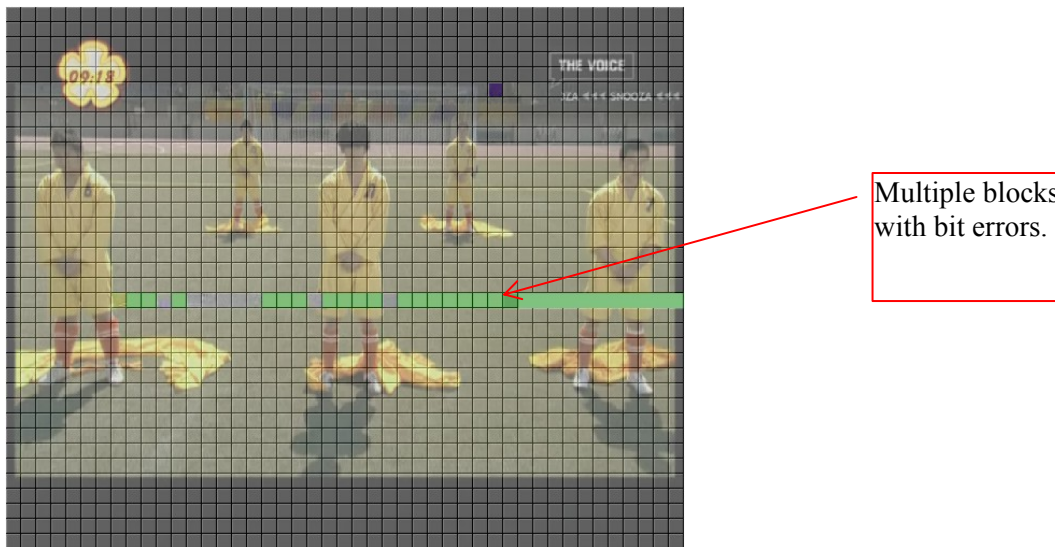


Figure 29: Screen shot from a music video with bit errors (from The Voice TV).

Figure 30 shows the variance of the DC coefficients. Bit errors appear in I-frames: 2, 12, 32, 33, 34, 35. We can see that variance of the average DC value gives us no information about where the bit errors are in this specific video content.

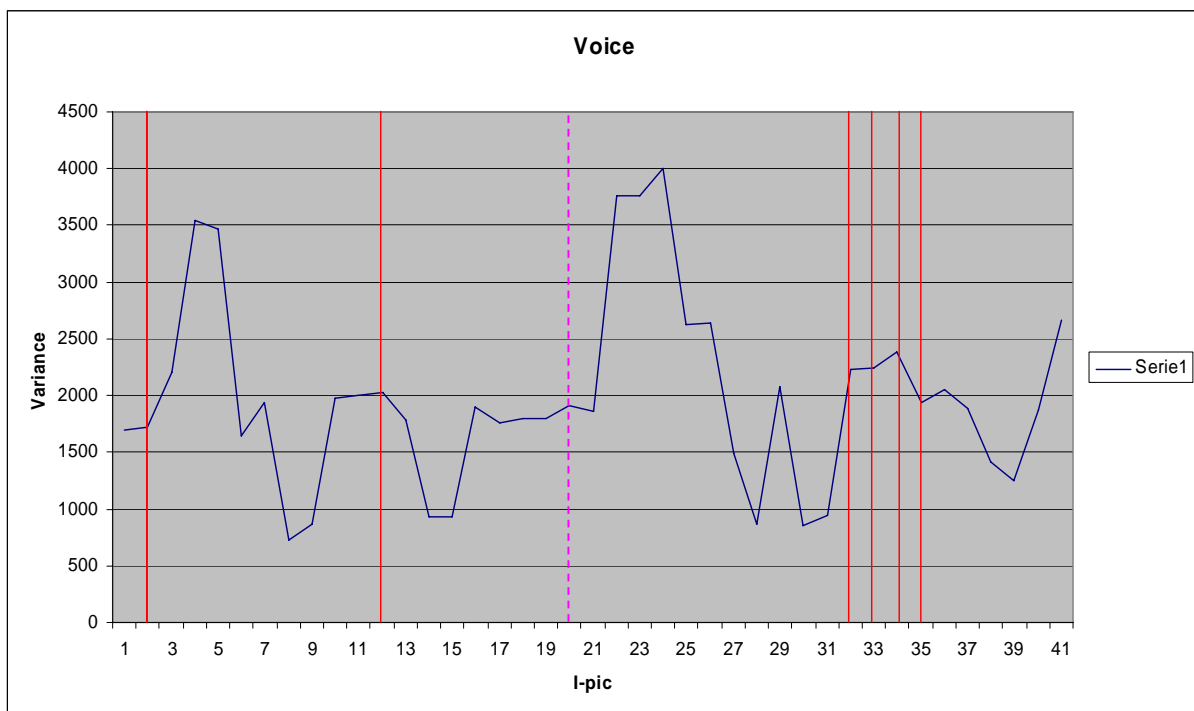


Figure 30: A diagram showing the variance value of DC-coefficients in time (based upon I-frames).

What we see in slice 20 in figure 29 is called desynchronization, which is caused by the transmission errors. VLCs (see section 3.3.6) used in MPEG-2 coding produces random bit sequences, if we loose synchronization, then the coded information in all the subsequent bits are undecodable until the next synchronization codeword appears [19]. In this case until the next slice. Detecting such errors is not possible by looking at the variance of the DC coefficient, but is detectable based on the detection of desynchronization! Desynchronization can be detected based upon the syntax errors which will occur after desynchronization.

#### 4.4.4 Conclusion of DCT coefficients usability

In single ended detection the use of mean or variance values of DC coefficients is useful only when the goal is to detect frozen or black video frames. For detection of bit errors other approaches have to be applied. These alternative approaches need to consider different kinds of syntax errors and based upon them infer the existence of bit errors.

For a double ended system, where it is possible to compare two identical video contents coming from two different sources, here DC coefficients can be used to detect frozen and black screen. For video containing bit errors the same method as for single ended detection based upon syntax errors has to be applied. Comparing DC coefficients from two different sources is thus possible in a double ended system. However, this may be both a complicated and an expensive procedure since we have to decode the complete TS for each content source. A complete monitoring system includes a large number of sources. In section 4.5 another potential solution of monitoring transport stream will be proposed. This solution's basic idea will be to add information into the transport stream, thus offering benefits of a double ended monitoring system, but using a single ended monitor.

#### 4.4.5 Detection of bit errors based on subsequent syntax errors

There are several commercial and open source systems which provide detection of bit errors based upon the detection of syntax errors. One example is DekTec's DTC-320 Streamexpert [33] or The Code Project's MPEG-2 analyzer [34]. Another one is Tektronix MTS4EA Video ES Analyzer [35]. The latter is a powerful MPEG analyzing tool (I had a chance to use a demo version of it), but its main drawback is its price. Such analyzer can cost more than 10000 €.

#### 4.4.6 Test bench for audio

For audio analysis two different programs were used. The first one is called “DecMPA v0.4.1” [36] and it is as “mpeg2dec” developed as an open source project. It extracts scale factor values from an MPEG-1 layer II audio file. The program decodes up to 3 scale factors for each subband and channel. See more information about MPEG audio and scale factors see section 3.4. The second program is the “Thales Mercury” (now Thomson Grass Valley) TNM2000 Series Audio stream analyzer [37] and it works basically in the same way. The second program had a better graphical overview, so the screen shots from this program will be used to illustrate this section of the thesis.

Why extract the scale factors? Starting with a similar approach as for video, we are trying to find a unique coefficient from which we can easily detect different kinds of content. Since the scale factor is a multiplier that is use to scale the samples in order to maximize the resolution of the quantizer, thus we would expect the scale factor to be unique for each audio frame. Scale factors have a range from 0 to 62. Appendix C [21] gives a list of values for each of the scale factors. Whereas Layer I encodes data in single groups of 12 samples for each subband, Layer II encodes data in 3 groups of 12 samples for each subband, such that the encoder forms frames of 3 by 12 by 32 = 1152 samples per audio channel [20]. The encoder is coding with a unique scale factor for each group of 12 samples only if necessary to avoid audible distortion. In figure 31, an overview of scale factors from an audio file is presented showing that some subbands have only one scale factor, some have two or three, while some have no scale factors at all. Figure 31 visualizes an audio frame from a transport stream from an SVT channel, where a person is speaking. The stream has a bitrate of 256 kbps and a frequency of 48 kHz. The shaded box does not represent any particular value; it is just a selected value, since the figure is just a screen shot. The value in each box is a scale factors (or several) for each subband and channel.

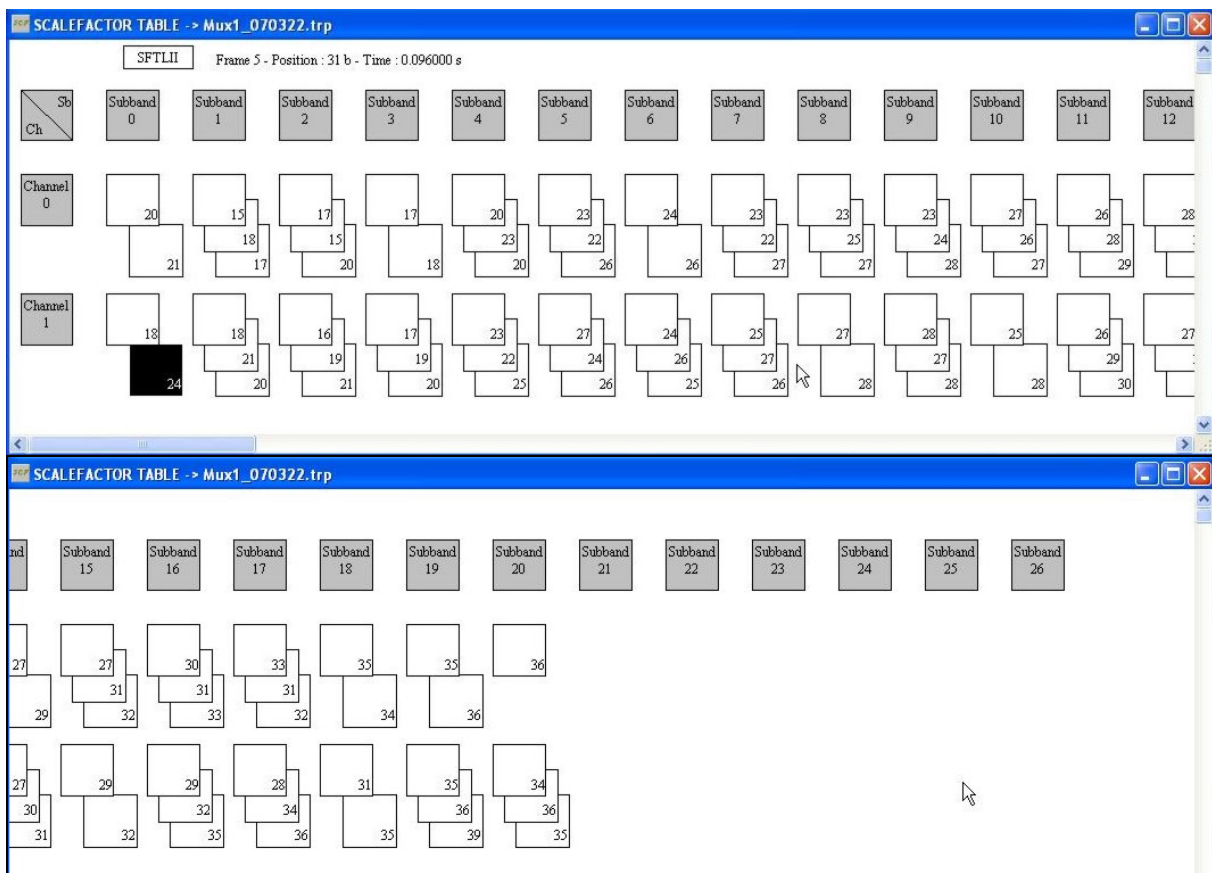


Figure 31: A screen shot from Thales Audio stream analyzer showing scale factors for a single audio frame.

The question now is what does the scale factors tell us about the quality of the content or the content itself. To answer these questions, we must first examine some other examples of audio content. The next case we study is a single tone. Though it is not often such audio content is broadcasted, still it may appear, e.g. in conjunction with a test pattern (such as shown in figure 25). The scale factors of a tone are visualized in figure 32. We see that there is almost all scale factors are zero outside of the lower frequency subbands. Appendix D [20] gives a list of critical bandwidths and which frequency bounds they represented.

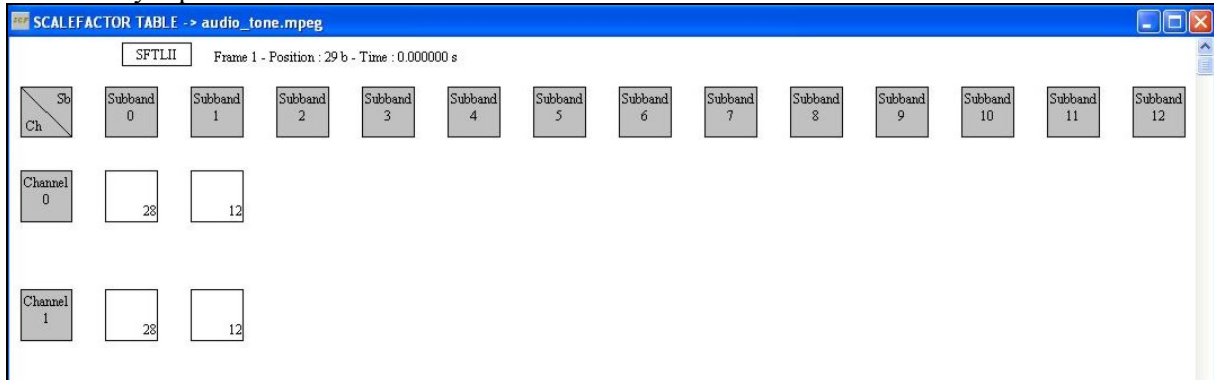


Figure 32: A screen shot showing scale factors for a single audio tone frame.

Next case we look at silence. Figure 33 shows that scale factors are present only in subband 0 with a value of 42 or 43. Given these two examples, we are able to detect both audio silence and a single tone in a similar way as detecting frozen pictures in video. Detecting silence in audio could be seen as analogous to detecting a frozen picture in video.

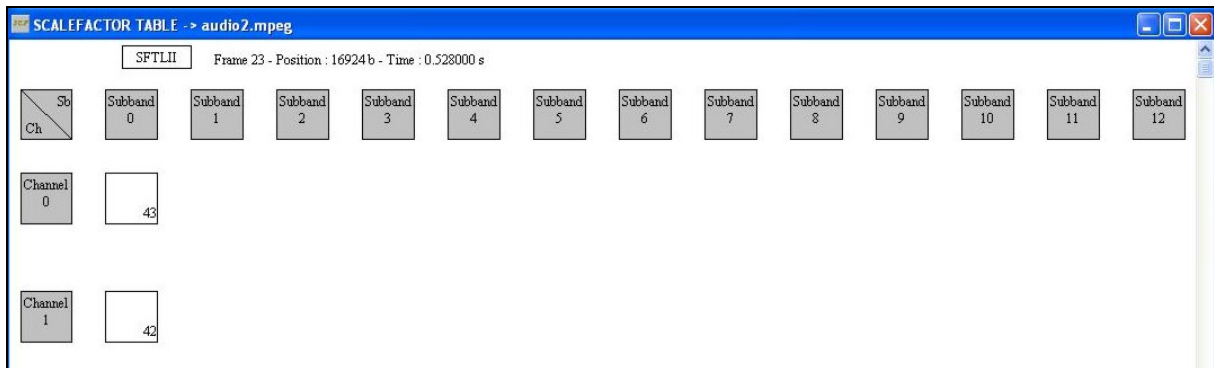


Figure 33: A screen shot showing scale factors for a single audio silence frame.



The final and the most interesting case concerning the audio analysis is an audio stream with bit errors. In the earlier video analysis we saw that usually bit errors results in syntax errors causing DC coefficients to be corrupted or even lost in a certain macroblock. An audio stream containing sufficient bit errors will result in loss of sound, since the CODEC will not be able to synchronize with the encode.

In figure 34 an audio frame with bit errors is visualized. The audio stream contains one or more bit errors. In this figure, subband 3 has one syntax error, namely a scalefactor with the value 63. This is impossible since the largest scalefactor number is 62. Therefore, the audio analyzer marks this as a wrong index. This is very similar to corrupted DC coefficients, which could also have a number beyond the range of allowed values.

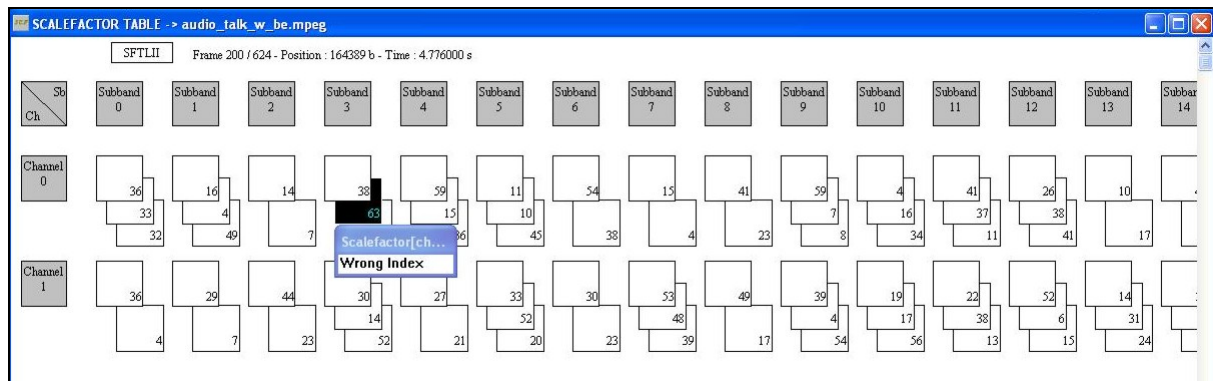


Figure 34: A print screen showing Scale factors for an audio frame with bit errors.

#### 4.4.7 Conclusion of scale factors usability

Scale factors proved to be useful in cases such as detection silence or a tone. In both cases a simple method of detection could be applied. In the case of bit errors the same approach as for video was considered, i.e. syntax errors and impossible scale factor values.

#### 4.5 Monitoring the digital data stream using signatures and syntax

The idea of adding information to the transport stream thus enabling a double ended monitoring system was suggested by Prof. Gerald Q. Maguire Jr. Since using DC coefficients and Scale Factors is a fairly complicated way of monitoring the audio and video content and also proved to be insufficient, we want to know whether we could detect bit errors *without* decoding the transport stream and without even being concerned with what this content consists of (i.e. the content could be encrypted, as many TV channels are in Teracom's DTT network). The remainder of the section is taken almost verbatim from Prof. Maguire's email to me on 2007.11.27.

Transport Stream  $A_1$  coming from a TV channel studio (e.g. SVT1) is received by Teracom; Teracom's contract concerning this channel is to deliver this out XXX antennas on multiplexer Y (where XXX is the number of antennas and Y is the particular multiplexer to be used). Thus it should be possible to compute a signature over features of this Transport Stream which can be examined at later points in the path from the content provider to a DVB-T receiver listening to multiplexer Y located near each of the XXX antennas.

Thus what we want to do is compute a hash over some parts of  $A_1$  from point  $t_1$  to  $t_2$  and compare this to a hash computed over these same parts of  $A_1$  from point  $t_1$  to  $t_2$  at each receiver  $R_i$  listening to multiplexer Y - for all  $i$  in the set XXX. If the hash matches there is nothing to do (except occasionally send an "I'm well" message to the monitoring system - so as to confirm that the monitoring station is still functioning). If the hash does not match at receiver  $R_j$  then the computer attached to receiver  $R_j$  should send a message to Teracom's monitoring system.

Teracom's monitoring system collects the reports of mismatches and based upon these reports generates alarms.

The advantage of this method is that:

1. The distributed system requires a number of monitoring sites proportional to the numbers of antennas (~54).
2. At each monitoring site a computer with ~6 receivers is needed - depending upon the computational load, processor used, and bus bandwidth this may require from 1 to 6 PCs.
3. Each monitoring site needs an uplink to the Teracom monitoring system (but this can be a rather low bandwidth link - as it only needs to report when it has a mismatch - perhaps time stamped and signed).

Note that these receivers could be physically located at the 54 antenna sites or could be distributed around the country.

In addition, Teracom needs to have ~33 TS processing units at the head end (Kaknäs) to compute the hash over the incoming TSs from the content providers. The resulting hashing and intervals can be encoded as part of the NIT information and hence distributed to all of the downstream systems.

So the key questions are:

- Q1. How to determine a suitable interval  $t_1$  to  $t_2$  to work over?
- Q2. What parts of the stream(s) in this interval should be excluded/included in the computations?
- Q3. What sort of hash should be performed?

By saying "hash" the computation could actually produce a vector of sub-values to help indicate where the error is, but in the interest of keeping things simple for a first round - we could simply focus on matching versus not matching a set of bytes in an ordered subset of a stream.

For Q1, we could either use STC times or PCR values - thus dividing the stream into time epochs; or between NITs with specific values in them (for example these could be a sequence number and a hash for the hashes of all the hashes over the desired fields since the last such special NIT).

Hashes could be separately computed per: Transport priority, PID, Scrambling control, Adaptation control, and/or Continuity indicator transport packets.

The idea is that in a digital distribution network, we do not care what the particular effects on the payloads are, but rather are concerned about:

1. Correctly delivering the transport packets which are to be transmitted.
2. Understanding how many of these (perhaps by type) are not correctly received.
3. Spotting complete connectivity failures (i.e., not being able to synchronize the receiver with the transmitter for some period of time).

The clear advantage of this solution is that it focuses on measuring the correct delivery and erroneous delivery of TS packets and not on the properties of the video or audio content - which may even be encrypted and for which you should **not** have to have the key.

Thinking about the problem in terms of verification of what is received from the content providers with respect to what is transmitted, it is also clear that each DVB-T monitoring receiver only needs about 15GBytes/day of storage to store all of the 6 multiplexers worth of data. Thus one can simply store the received data, then if there is an error, Teracom's central monitoring site could even access the raw received data to see what the error was in detail if they wanted - for example using one of the programs from the Dektec.com web site [38].

Two open questions remain:

1. How much time can be allowed to elapse from when the error occurs to when it is indicated?
2. How much delay does there need to be from the end of a frame of a TS until you should (or could) output the hash in the NIT? (For example, delaying this information by 200ms would allow time for computation via pipelining could be easily supported. This would add very little to the time before an error occurs and when it was detected.)

## 5. Conclusions

### 5.1 Evaluation

After describing four different systems, some evaluations can be made. The first two existing solutions from IdeasUnlimited and Agama Technologies have been tested in Teracom's laboratory and evaluated by Teracom. Both solutions have their advantages and disadvantages. Agama's solution turned out to be more suitable given the demands for the monitoring system needed by Teracom. It provides a quick and accurate detection of frozen frames, black screen, and the most important - a double ended comparison of the content, thus detecting whether it is correct or not.

Monitoring the content and detection of bit errors by using DC coefficients and Scale factors is sometimes useful, but proved to be a complicated approach due to the fact that each TV program within the TS need to be decoded and DC coefficients and Scale factors needs to be carefully examined. This requires powerful and expensive amounts of processing and substantial software investments. At the same time, the advantage is that we can use an already existing (in the TS) signature over some specific details about the content, thus telling us whether it is corrupted or correct.

Monitoring the content with help of signatures and syntax (as proposed by Prof. Maguire) could be a cheaper alternative and probably even more effective than the use of DC coefficient and Scale factors. However, this requires altering the TS by the Teracom DTT network (i.e. specifically changing of the NIT) which is currently impossible due to several technical and bureaucratic reasons. However, since the DTT network will be changed (MPEG-2 will be replaced sooner or later by MPEG-4 and in future maybe by MPEG-7) there will be greater demands on monitoring the broadcasted content. An integration of the complete DTT system, including digital set-top boxes at the end users would allow a greater supervision of entire DTT network, thus providing a quick discovery of different kinds of errors. The clear advantage of this monitoring procedure compared to the method using DC coefficients and Scale factors is that it does not really care about the coding method, and/or whether the TS traffic is encrypted or not. Hence the current technical and bureaucratic problems need to be overcome.

This master's thesis gave me an enormous amount of new knowledge in a subject that I was interested in, but did not know much about. Since this master's thesis has been done at Teracom, it also gave me insight into how the Swedish digital terrestrial network is built up and maintained. I learned how MPEG-2 compression is done, how the DVB-T system works, what DC coefficients are and what they mean, as well as Scale factors and what desynchronisation means in terms of a TS.

### 5.2 Future work

Next thing to be done could be developing a program based on the knowledge gained in the course of this thesis. A person who would follow up on this thesis could benefit from the research performed on different monitoring methods, perhaps by combining some parts of several methods. However, the technical development is occurring very fast, in couple of years new coding techniques will replace the older ones and new supervision procedures will be necessary. Hopefully, they will be based upon earlier techniques, so that the knowledge gained in the course of this thesis project will not go to waste.

## References

- [1] ETSI Standard: EN 300 744 V1.5.1 (2004-11), *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*.
- [2] Barry G. Haskell, Atul Puri and Arun N. Netravali. 1997. *Digital Video: An introduction to MPEG-2*. ISBN 0-412-08411-2, New York: Chapman & Hall.
- [3] Seppo Kalli. 1998. *80528 Digital Media: Course material*. Tampere University of Technology.
- [4] Vesa Lundén. 1996. *MPEG-2 Standard and Hierarchical Coding of Video Signals: A Thesis for the degree of Licentiate of Technology*. Tampere University of Technology.
- [5] Mats Røjne. 2003. *Digital-TV via mark, satellit och kabel*. ISBN 91-44-03054-1, Lund: Studentlitteratur.
- [6] Agama technologies, *Why use the Agama IPTV monitoring solution?* (2004) [www] [http://www.agama.tv/pdf/whitepaper\\_why\\_agama.pdf](http://www.agama.tv/pdf/whitepaper_why_agama.pdf) Last access on 2007-02-19
- [7] IdeasUnlimited, *ContentProbe Product Summary*, (June 2005) [www] <http://www.ideasunlimited.tv/download/ContentProbeProductSummary.pdf> Last access on 2007-02-19
- [8] Wikipedia. Viaccess [www] <http://en.wikipedia.org/wiki/Viaccess> . Last access on 2007-03-01
- [9] Open TV [www] <http://www.opentv.com/products/servicemgmt.htm> Last access on 2007-12-14
- [10] Wikipedia. CCIR 601 [www] [http://en.wikipedia.org/wiki/CCIR\\_601](http://en.wikipedia.org/wiki/CCIR_601) . Last access on 2007-03-06
- [11] Dr. S.R.Ely. 1995. *MPEG video coding: A simple introduction*. EBU Technical Review. [www] [http://www.ebu.ch/en/technical/trev/trev\\_266-ely.pdf](http://www.ebu.ch/en/technical/trev/trev_266-ely.pdf) Last access on 2007-03-07
- [12] Katsunao Takahashi, Maki Sugiura, Masato Murai, Akihisa Kodate, Hideyoshi Tominaga. 1998. *A Proposal of "Video Fingerprints" for MPEG-7 Visual Descriptor*. MPEG-7 AHG meeting in Lancaster
- [13] Wikipedia. HDTV [www] <http://en.wikipedia.org/wiki/Hdvtv> . Last access on 2007-03-19
- [14] Jeffery O. Noah. May 1997. *A rational approach to testing MPEG-2*. IEEE Spectrum, Volume 34, Number 5.
- [15] Boxer. [www] <http://www.boxer.se/?page=248> Last access on 2007-05-02
- [16] E24. [www] [http://www.e24.se/dynamiskt/tjansteforetag/did\\_14788692.asp](http://www.e24.se/dynamiskt/tjansteforetag/did_14788692.asp) Last access on 2007-05-02
- [17] Dolby Digital. AC-3 [www] [http://www.dolby.com/assets/pdf/tech\\_library/37\\_ac3-flex.pdf](http://www.dolby.com/assets/pdf/tech_library/37_ac3-flex.pdf)
- [18] O. Lehtoranta, T.D. Hamalainen, and V. Lappalainen, "Detecting corrupted intra macroblocks in H.263 video", *Multimedia Signal Processing*, 2002 IEEE Workshop on Volume, Issue , 9-11 Dec. 2002 Page(s): 33 – 36

- [19] Jihua Cao and Zhaohua Wang, "Analysis and detection of transmission errors for MPEG-2 videosegnal" Circuits and Systems, 2000. IEEE APCCAS 2000. The 2000 IEEE Asia-Pacific Conference on Volume, Issue , 2000 Page(s):105 – 108 Digital Object Identifier 10.1109/APCCAS.2000.913417
- [20] Davis Yen Pan, "Digital Audio Compression", Digital Technical Journal Vol.5 No.2, Spring 1993
- [21] ISO/IEC 11 172-3: 12393 (E) Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio Annex B, page 45
- [22] libmpeg2 downloadable files [www] <http://libmpeg2.sourceforge.net/downloads.html> Last access on 2007-11-20
- [23] Wikipedia, Network Time Protocol [www] [http://en.wikipedia.org/wiki/Network\\_Time\\_Protocol](http://en.wikipedia.org/wiki/Network_Time_Protocol) Last access on 2007-12-11
- [24] Viaccess [www] <http://www.viaccess.com/en/> Last access on 2007-12-11
- [25] ISO/IEC 11 172-3, Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio [www] <http://www.iso.ch/cate/d22412.html>
- [26] ISO/IEC 13818-1:2007, Information technology -- Generic coding of moving pictures and associated audio information: Systems [www] [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=44169](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44169) Last access on 2007-12-14
- [27] ETSI TR 101 290 V1.2.1, DVB; Measurement guidelines for DVB systems, [www] ETSI TR 101 290, <http://broadcasting.ru/pdf-standard-specifications/measurement/tr101290.v1.2.1.pdf> Last access on 2007-12-14
- [28] Omnibus [www] [http://www.omnibus.tv/products/itx\\_features.html](http://www.omnibus.tv/products/itx_features.html) Last access on 2007-12-14
- [29] Agama Technologies [www] [www.agama.tv](http://www.agama.tv) Last access on 2007-12-14
- [30] Fedora Project [www] <http://fedoraproject.org/> Last access on 2007-12-14
- [31] JDSU DTS-330 Digital Broadcast Test Platform, [www] [http://www.jdsu.com/test\\_and\\_measurement/products/descriptions/dts-330/index.html?cat=product\\_categories&subcat=digital\\_video\\_test](http://www.jdsu.com/test_and_measurement/products/descriptions/dts-330/index.html?cat=product_categories&subcat=digital_video_test) Last access on 2007-12-14
- [32] Interra systems Vega analyzer, [www] [http://www.interrasystems.com/dms/dms\\_vega.php](http://www.interrasystems.com/dms/dms_vega.php) Last access on 2007-12-14
- [33] DekTec DTC-320 Streamexpert, [www] <http://www.dektec.com/Products/DTC-320/index.asp> Last access on 2007-12-14
- [34] The Code Project, [www] [http://www.codeproject.com/KB/audio-video/program\\_stream\\_analyzer.aspx](http://www.codeproject.com/KB/audio-video/program_stream_analyzer.aspx) Last access on 2007-12-14

- [35] Tektronix MTS4EA, [www] [http://www.tek.com/products/video\\_test/mts4ea/index.html](http://www.tek.com/products/video_test/mts4ea/index.html)  
Last access on 2007-12-14
- [36] DecMPA v0.4.1 - Simple MPEG Audio Decoder, [www] <http://decmpa.sourceforge.net/>  
Last access on 2007-12-14
- [37] Thales Mercury Analyzer, [www] <http://www.grassvalley.com/products/tbm/mercury/>  
Last access on 2007-12-14
- [38] DekTec Applications, [www] <http://www.dektec.com/Downloads/Applications.asp#DTC-350>  
Last access on 2007-12-14
- [39] Teracom annual report 2000, [www] [http://www.teracom.se/pub/4801/Årsrapport\\_2000\\_Eng.pdf](http://www.teracom.se/pub/4801/Årsrapport_2000_Eng.pdf)  
Last access on 2007-12-19

## Appendix A

Definition of NxN DCT:

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{j=0}^7 \sum_{k=0}^7 f(j,k) \cos\left(\frac{(2j+1)u\pi}{16}\right) \cos\left(\frac{(2k+1)v\pi}{16}\right)$$

Where:

$f(j,k)$  = the original samples in the 8x8 pixel block.

$F(u,v)$  = coefficients of the 8x8 DCT block.

$u$  = the normalized horizontal frequency ( $0 < u < 7$ ).

$v$  = the normalized vertical frequency ( $0 < v < 7$ ).

$$\begin{aligned} C(u), C(v) &= 1/\sqrt{2} && \text{if } u, v = 0 \\ C(u), C(v) &= 1 && \text{if } u, v = 1, 2, \dots, 7 \end{aligned}$$



## Appendix B

A screen-shot has been taken from a DVB TS Interra systems Vega [32] analyzer program. It shows different kind of information about SVT multiplexer 1 in Teracom DTT. The most interesting part is information about PIDs connected to each Stream Type.

PID	Program	Pkt. Count	Rate (Mb/s)	% Bandwidth	TblErr	CCErrors	Stream Type
0		2695576	0.016	0.072	280	116	PAT
1		2694990	0.016	0.072	290	116	CAT
16		519882	0.003	0.014	48	7	NIT
17		808488	0.005	0.024	15	29	SDT
18		24227498	0.147	0.665	12898	210	EIT
20		51205	0.000	0.000	2	4	TDT TOT
21		466936	0.003	0.012	0	1	MIP
870		2694955	0.016	0.072	116	111	PMT
871	Barn/Kunskapsk.	340485	0.002	0.008	0	1	Private Sections
872	Barn/Kunskapsk.	8512065	0.050	0.227	0	81	DSM-CC
878	Barn/Kunskapsk.	44807666	0.263	1.190	0	331	Audio
879	Barn/Kunskapsk.	705390708	5.591	25.277	0	4681	Video
880		2694968	0.016	0.072	306	100	PMT
1002	Shared	127657498	0.750	3.392	0	983	DSM-CC
1004	Shared	38406474	0.226	1.020	0	310	Teletext
1010		2694952	0.016	0.072	94	105	PMT
1011	SVT1	340485	0.002	0.008	0	1	Private Sections
1012	SVT1	8512070	0.050	0.226	0	85	DSM-CC
1016	SVT1 Tal txt	11853793	0.069	0.314	0	106	Audio
1017	SVT1	80008717	0.470	2.126	0	558	AC-3 Audio
1018	SVT1	44807402	0.263	1.190	0	378	Audio
1019	SVT1	889403293	1.800	8.140	0	5740	Video
1021	Shared	340484	0.002	0.010	0	3	Private Sections
1022	Shared	8512083	0.050	0.226	0	70	DSM-CC
1026	SVT2 Tal txt	11853814	0.069	0.314	0	100	Audio
1027	Shared	77678552	0.470	2.126	0	217	AC-3 Audio
1028	Shared	44484692	0.263	1.190	0	1105	Audio
1029	Shared	670255943	5.492	24.829	0	16387	Video
1241	SVT24 ABC	340485	0.002	0.010	0	0	Private Sections
1242	SVT24 ABC	8512084	0.050	0.226	0	69	DSM-CC
1248	SVT24 ABC	44807602	0.263	1.189	0	371	Audio
1249	SVT24 ABC	713787563	4.724	21.360	0	5876	Video
1280		2694960	0.016	0.072	319	108	PMT
1290		2694980	0.016	0.072	316	102	PMT
5060		2694965	0.016	0.072	136	104	PMT
5070		2694966	0.016	0.072	147	104	PMT
5170		2694970	0.016	0.072	148	106	PMT
8191		150822886	0.878	3.968	0	0	NULL PID
		11884	0.000	0.000	0	0	User Private

## Appendix C

Take from ISO/IEC 11 172-3, Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio. [25]

### Tables

Table B.1 -- Layer I,II scalefactors

index	scalefactor	index	scalefactor
0	2,00000000000000	32	0,00123039165029
1	1,58740105196820	33	0,00097656250000
2	1,25992104989487	34	0,00077509816991
3	1,00000000000000	35	0,00061519582514
4	0,79370052598410	36	0,00048828125000
5	0,62996052494744	37	0,00038754908495
6	0,50000000000000	38	0,00030759791257
7	0,39685026299205	39	0,00024414062500
8	0,31498026247372	40	0,00019377454248
9	0,25000000000000	41	0,00015379895629
10	0,19842513149602	42	0,00012207031250
11	0,15749013123686	43	0,00009688727124
12	0,12500000000000	44	0,00007689947814
13	0,09921256574801	45	0,00006103515625
14	0,07874506561843	46	0,00004844363562
15	0,06250000000000	47	0,00003844973907
16	0,04960628287401	48	0,00003051757813
17	0,03937253280921	49	0,00002422181781
18	0,03125000000000	50	0,00001922486954
19	0,02480314143700	51	0,00001525878906
20	0,01968626640461	52	0,00001211090890
21	0,01562500000000	53	0,00000961243477
22	0,01240157071850	54	0,00000762939453
23	0,00984313320230	55	0,00000605545445
24	0,00781250000000	56	0,00000480621738
25	0,00620078535925	57	0,00000381469727
26	0,00492156660115	58	0,00000302772723
27	0,00390625000000	59	0,00000240310869
28	0,00310039267963	60	0,00000190734863
29	0,00246078330058	61	0,00000151386361
30	0,00195312500000	62	0,00000120155435
31	0,00155019633981		

## Appendix D

### Approximate Critical Band Boundaries

---

Band Number	Frequency (Hz) <sup>1</sup>	Band Number	Frequency (Hz) <sup>1</sup>
0	50	14	1,970
1	95	15	2,340
2	140	16	2,720
3	235	17	3,280
4	330	18	3,840
5	420	19	4,690
6	560	20	5,440
7	660	21	6,375
8	800	22	7,690
9	940	23	9,375
10	1,125	24	11,625
11	1,265	25	15,375
12	1,500	26	20,250
13	1,735		

---

<sup>1</sup>Frequencies are at the upper end of the band.

---

