Automation, Benchmarking and Analysis of Digital Video Recorder and Media Player Solution

By

Hitesh R.Viradiya

08MCE020

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AHMEDABAD-382481

May 2010

Automation, Benchmarking and Analysis of Digital Video Recorder and Media Player Solution

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

Master of Technology in Computer Science and Engineering

By

Hitesh R.Viradiya

08MCE020

NIRMA UNIVERSITY
INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AHMEDABAD-382481

May 2010

Declaration

I hereby declare that

i) The work presented in the thesis comprises my original work towards the degree of Master of Technology in Computer Science and Engineering at Nirma University and has not been submitted elsewhere for a degree.

ii) Due acknowledgement has been made in the text to all other material used.

Hitesh R. Viradiya

Certificate

This is to certify that the Major Project entitled ”Automation, Benchmarking and Analysis of Digital Video Recorder and Media Player Solution” submitted by Hitesh R. Viradiya (08MCE020), towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University of Science and Technology, Ahmedabad is the record of work carried out by him in Multimedia team at ST Microelectronics, Greater Noida, under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. S.N. Pradhan
M.Tech-Coordinator,
Department of Computer Engineering,
Institute of Technology,
Nirma University, Ahmedabad

Mr. Amit kumar Srivastava
Project Manager,
Multimedia team,
ST Microelectronics,
Greater Noida

Dr. Ketan Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad,

Prof. D. J. Patel
Professor and Head,
Department of Computer Engineering,
Institute of Technology,
Nirma University, Ahmedabad

Abstract

Electronic products have long become an integral part of the modern lifestyle. Demand and supply of such products with smaller size, lower power consumption and better performance are ever increasing. What is the future of home entertainment and computing and how consumers cope with rapid changes of digital life?

The PC is an amazing product; every year prices fall, while processing capabilities and storage capacity continue to escalate. However, the personal computer is not an ideal platform for home entertainment. The system hardware and software of PC are not designed for real-time audio video stream processing. The Digital Platform is a technical proposal for a new architecture for home entertainment and computing. The Digital Video Recorder (DVR) - a recorder that records on a computer hard disk has evolved as a low-cost, high performance consumer product. DVR is a popular security device that provides real time monitoring, video capturing and video playback from storages.

STMicroelectronics continues to lead the market for set-top box decoders with the various chip family and delivers its own solution of Digital Video Recorder abbreviated as DVR & Media Player as MP and that has already presence across the globe. The scope of this project thesis is to do automation, benchmarking and analysis of the ST's DVR & MP solution and hence this thesis contains overview of Set Top Box basics, different video compression standards, DVB standards and DVR & MP features. Depending upon DVR and MP features, different test suits are developed to test all those features in automated way. DivX Certification is achieved as one of the benchmark of STMP. The key feature targeted for Analysis is Audio/Video Synchronization, so as covered various methods of AV Sync and how they differ one another with pros and cons.

Acknowledgements

Any fruitful effort in a new work needs a direction and guiding hands that shows the way. I take this opportunity to express my sincere gratitude & it gives me enormous pleasure in expressing my thankfulness towards Dr S. N Pradhan, Professor, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for his precious guidance, support and inspiration to pursue my thesis. I am thankful to him for the valuable time he spent with me to carry on my thesis and shaped the project.

My proud privilege and bliss to bring my indebtedness and warm gratitude to Mr. Amit kumar Srivastava, Project Manager ST Microelectronics, Greater Noida, for his implication to improve quality of work and providing constant motivation and supports to be persistent the project thesis work. I am also thankful to Mr. Ankit Singhal, Sr. Software Engineer, for his brand support in all respects to prolong my project work.

I am thankful to all the team members of Multimedia team (Neerajkumar Singh, Rohit Sharma, Gurpreetsingh Dhami, Jyotirmoy roy, Deepak Pandey and Neeraj aggarwal) of ST Microelectronics, Greater Noida, for their special awareness and suggestions towards the project work.

- Hitesh R. Viradiya
08MCE020

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter contains the introduction of set-top box, Digital Video Recorder, and Media Player.

1.1 Set-top box (STB)

1.1.1 Definition

A small box sits on the standard Television which internally receives input signal coming from cable, satellite or terrestrial, and converts its content into specific form and display it onto the TV screen.[3]

To have more specific, it is a receiver which receives digital transmissions via the conduit it was designed to interface to, and then process the data in order that both the video and audio information can be sent to an external display and amplifier for the viewer to experience.

1.1.2 Architecture

As shown in Figure 1.1 set-top box is made up of mainly two parts:

a. **Front End:** Tuner, FEC and Digital Demodulator

b. **Back End:** TS Demux, AV Decoder, Digital Video Encoder, Audio DAC & CPU with on site memory

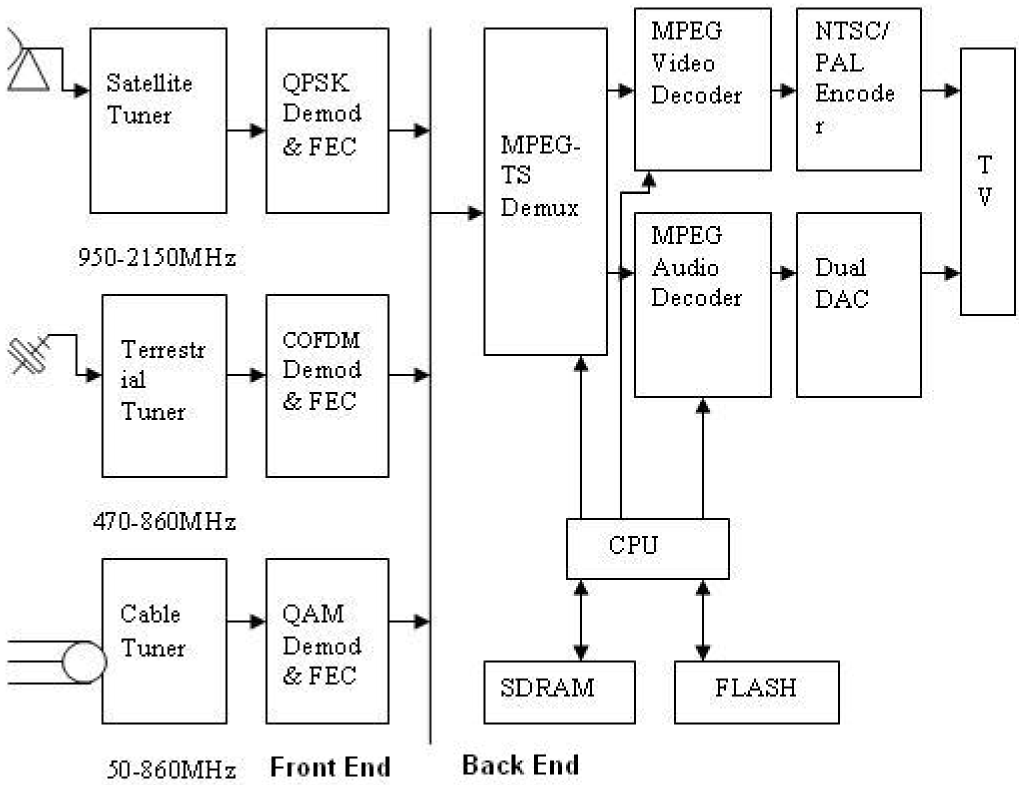


Figure 1.1: Block Diagram of STB

Front End components:

a. **Tuner:** A module or device which converts low-amplitude radio-frequency signals into a form suitable for processing by further modules or equipment. Here suitable format is error corrected TS (Transport Stream) format.

b. **FEC:** Forward Error Correction is a system of error control for data transmission, whereby the sender adds redundant data to its messages, also known as an error-correction code. This allows the receiver to detect and correct errors (within some bound) without the need to ask the sender for additional data.

c. **Digital Demodulator:**

QPSK demodulator: demodulates the signal coming from satellite.

QAM demodulator: demodulates the signal coming from cable.

OFDM demodulator: demodulates the signal coming from terrestrial.

Back End components:

a. **TS Demux:** Demultiplexer (demux) extracts all the useful information from the TS, since MPEG-2 data streams consist of a number of unique data packets and uses packet ID to identify each packet that contains data, A/V and interactive services. The demux examines every packet ID, selects packets, decrypts them, and then forwards them to their specific decoder.

b. **AV Decoder:** Decoders are required to convert the digital bit stream back into the format accessible by the subscriber. The video decoder converts video packets into a sequence of pictures. Next, the audio bit stream is sent to the audio decoder for decompression so it can be sent to the speakers. Table formats are decompressed using data decoders. Then, the decoded data is presented to the set-top processor. Sometimes, JPEG and MP3 decoder are also embedded to process digital still picture and compressed music.

c. **Digital Video Encoder:** Encode the elementary stream data, extract actual video information and convert into proper display format like NTSC/PAL format.

d. **Audio DAC:** Same as Digital Video encoder this converts digital audio (came from digital medium) into analog and passes it to speaker into TV.

e. **CPU with on site memory:** CPU handles all the tasks, like initialize various STB hardware components, interactive TV applications, manage hardware interruptions, pull data from memory and run programs with the use of SDRAM as primary memory and FLASH as secondary memory. CPU in STB is typically a 32bit processor with speed ranges from 50MHz to 300MHz. It always contains an arithmetic logic unit, a control unit and a clock.

1.1.3 Data-flow in set-top box

In Front-End,

- Tuner receives a digital signal from anyone of three source and tunes to a particular channel in the frequency range mentioned in diagram respectively for each one. For example, in case of satellite it tunes 950-2250 MHz.
- Then signal is digitally demodulated using QPSK demodulator (in case of satellite) to obtain digital data.
- Forward Error Correction (FEC) decoder will check error and send error corrected TS to Back-end.

In Back-End,

- TS demultiplexer extracts audio, video and clock and sends to AV decoder and smart card.
- Smart card determines access rights to various digital services.
- AV decoder decompresses audio video data which are to be displayed on TV after DAC and Video encoder processing respectively.

1.1.4 STB Softwares

There are different software layers in a STB. The structural diagram of general STB software is shown in Figure 1.2. An operating system (OS) is the most important piece of software in a STB. An OS is a suitable of programmes used to manage the resources in a STB. In particular it is the OS, which talks to the STB hardware and manage their functions such as scheduling real time tasks, managing limited memory resources[4].

At the heart of any STB OS is the "Kernel" layer, which is stored in ROM. Once the

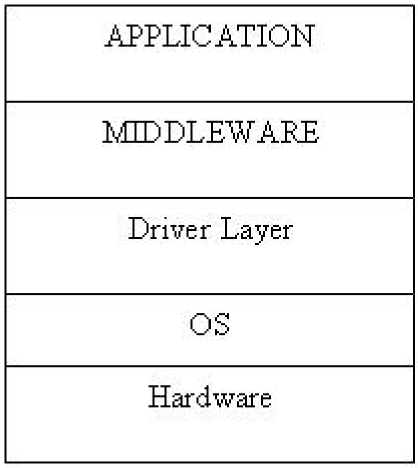


Figure 1.2: Block Diagram of STB

STB is powered up, the kernel will be loaded first and remains in memory until the STB is powered down again. Typically the kernel is responsible for managing memory resources, real time applications and high-speed data transmission. The kernel supports multi threading and multi tasking which allows a STB to execute different sections of a program and different programmes simultaneously.

The STB also requires 'drivers' to control the various hardware devices. Every hardware component in the STB must have a driver. A driver is a program that translates commands from the TV viewer to a format that is recognizable by the hardware device.

Finally a STB OS needs to incorporate a set of Application Programme Interfaces which are used by the programmers to write high-level applications for a specific API. An API is basically a set of building blocks used by software developers to write programs that are specific to a STB OS environment. Central to the new software architecture of a STB is a connection layer that acts as communications bridge between the OS and the 'subscriber applications' called 'Middleware'.

Middleware is a relatively new term in the set top business. It represents the logical abstraction of the middle and upper layers of the communication software stack used in set top software and communication system. Middleware is used to isolate set top application programs from the details of the underlying hardware and network components. Thus set top applications can operate transparently across a network without having to be concerned with the underlying network protocols. This considerably reduces the complexity of content development because applications can be written to take advantage of a common API.

1.2 Digital Video Recorder (DVR)

1.2.1 Definition

A digital video recorder (DVR) or personal video recorder (PVR) is a de- vice that records video in a digital format to a disk drive or other memory medium within a device.

1.2.2 Architecture

As shown in Figure 1.3, DVR is having two implementation.

a. **Non-loop backing DVR:** Simple way of digital video recording.

b. **Loop backing DVR:** used to prevent loosing of data while recording as it takes sometime in decoding and the live stream data might get changed by new data.

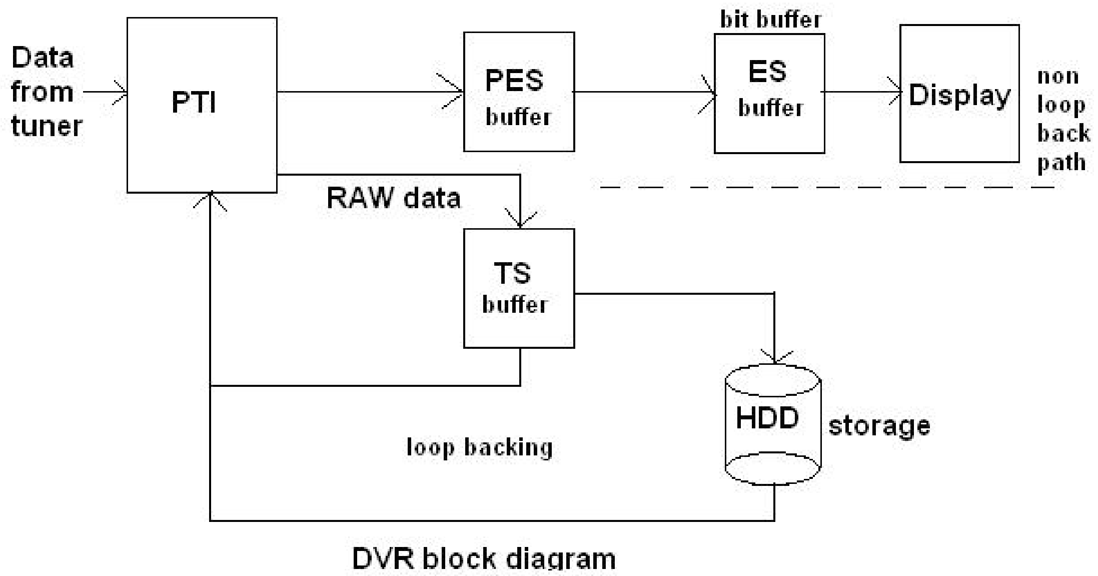


Figure 1.3: Block Diagram of DVR

1.2.3 DVR Operations

There are two independent operations in the DVR: Recording and Playback which are explained as follows[5]:

a. **Recording:** There are 3 steps in the recording operation:

(1) *Video capture & compression:* Analog video data is first digitized to Y/Cb/Cr format using an analog video decoder before passing to Video Processing Front End (VPFE). Captured video data is time-stamped and stored in memory subsystem. It is then passed to DSP for processing. The DSP controls the operation and perform real-time video compression and decompression. Compressed data is stored in system memory in elementary stream (ES) format.

(2) *Audio capture & compression:* Audio data is digitized by an audio codec. The data is also time-stamped and passed to DSP for processing. Compressed data is also stored in system memory in ES format.

(3) *A/V data multiplexing and storage:* The audio and video ES data is first

packetized into PES data and then multiplexed and packetized into fixed size transport stream packets before being stored in the hard disk drive.

b. **Playback:** There are also 3 steps in the playback operation, applied in the reversed order of recording:

(1) *A/V data de-multiplexing:* Transport stream data are de-multiplexed and de-packetized to separate A/V elementary streams.

(2) *Audio decoding and playback:* Audio bit-stream is decoded and the reconstructed audio data is played back audio codec.

(3) *Video decoding and playback:* Video bit-stream is decoded and the reconstructed video frame is displayed Video Processing Back End (VPBE). The timing of the display is controlled by the audio/video synchronization mechanism.

1.2.4 DVR Features

Digital video recorders are starting to make there way into the living rooms of millions of consumers. There are a number of features that make the Digital video recorder a need of today's life; i.e.

- Live playback
- Dual Playback
- Dual Recording
- Record & Playback
- Playback with trickmodes and catch-live
- Play seek
- Play from different source

- Live + Record
- Recording with the same file name
- Circular Recording
- Various Recorded file edit operation : copy, crop and remove file.

1.3 Media Player (MP)

1.3.1 Introduction

Mediaplayer component is used to play different media files. The main goals in defining the media player component were:

- Simplify the APIs for playing media files
- Manage the state of the media player
- The video and audio control (configuring displays and routing the audio paths) are managed by application based on the information provided by media player. This configuration is done using interfaces implemented in application.

1.3.2 MP Architecure

Following Figure 1.4 shows the architecture of ST Mediaplayer Architecture.

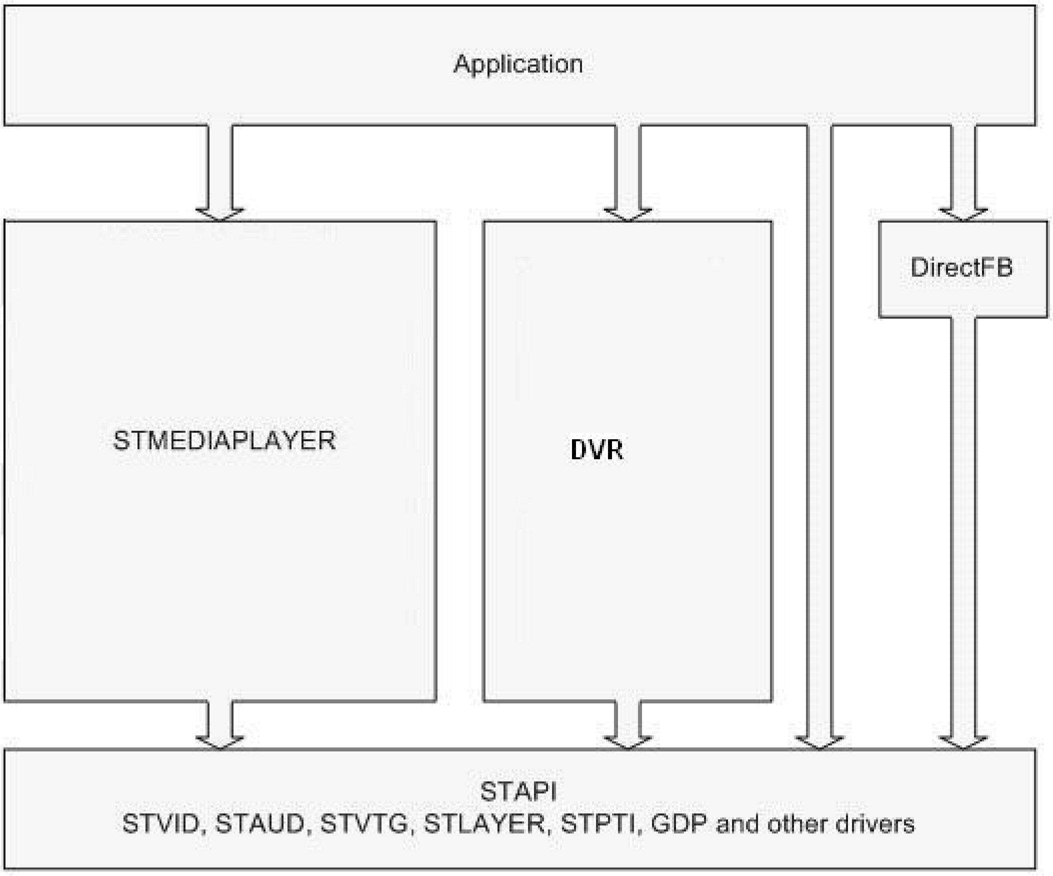


Figure 1.4: Architecture of Media Player

Formats Support

Audio formats: Mp3, WMA, AC3, WMA PRO, MPEG4 AAC, Real Cook
Video formats: MPEG1, MPEG2, MPEG4, H264, DivX, VC1, WMV

Container formats: Mpeg2 TS, Mpeg2 PS, MP4, QuickTime file format, Mobile format 3GP, AVI, AVI-DivX, ASF, RMVB, EVO, VOB, Matroska (MKV), Flash Video (FLV), MOV, WAV.

1.4 Thesis Organization

The rest of the thesis is organized as follows.

Chapter 2 Literature survey

This chapter contains the literature survey of Digital Video Broadcasting (DVB), Video Compression Standards, DVR feature like AV Sync and DivX Certification of Mediaplayer.

Chapter 3 Problem Definition and Existing Methodologies

This chapter contains Project Definition, Existing AV Sync methodology in the field of AV synchronization and stand alone Automated system test method.

Chapter 4 The Proposed Algorithm

This chapter contains AV Sync in digital set-top boxes, and Web based Automated system test method.

Chapter 5 Implementation

This chapter contains AV Sync implementation, DivX Certification, Implementation of Automated system test method Implementaion-1, Implementaion-2, Implementaion-3 and Tools & Technologies used

Chapter 6 Simulation results & Analysis

This chapter contains DivX Certification report, DVR Test-case results, and Analysis of AV Sync methods.

Chapter 7 Conclusion and Future work/scope

This chapter contains conclusion and future work.

Chapter 2

Literature survey

2.1 Digital Video Broadcasting (DVB)

2.1.1 Definition

It is a suite of internationally accepted open standard for digital TV, maintained by the DVB project, an industrial consortium with more then 300 members, and published by joined technical committee (JTC).[6]

2.1.2 Why DVB?

- Digital information is more robust as it can be coded to eliminate the error.
- Allow content manipulation i.e. storing and preprocessing, which is impossible with the analog.
- Better use of the same BW to pack more channels.

2.1.3 Core standards: DVB transmission

- **DVB-S** for satellite transmission
- **DVB-C** for cable transmission

- **DVB-T** for terrestrial transmission

2.2 Video Compression Standards

2.2.1 Why Video Compression?

As discussed in DVB to transmission of video requires digitization, increase the the bandwidth by approximately 21 times. So compression is needed.

2.2.2 MPEG overview

The MPEG encoder input sequence consists of a series of frames, each frame is a still image containing a two dimensional array of picture elements. The compression algorithm is used to reduce the data rate before transmitting the video stream over communication networks. The MPEG-2 video compression algorithm achieves very high rates of compression by exploiting the redundancy in video information. MPEG-2 removes both the temporal redundancy(inter frame) and spatial redundancy(intra frame) as shown in Figure 2.1 which are present in motion video.

In MPEG, where temporal compression is used, the current picture/ object is not sent in its entirety; instead the difference between the current picture/object and the previous one is sent. The decoder already has the previous picture/object, and so it can add the difference, or residual image, to make the current picture/object. A residual image is created by subtracting every pixel in one picture/object from the corresponding pixel in another. This is trivially easy when pictures are restricted to progressive scan, as in MPEG-1, but MPEG-2 had to develop greater complexity (continued in MPEG-4) so that this can also be done with interlaced pictures.

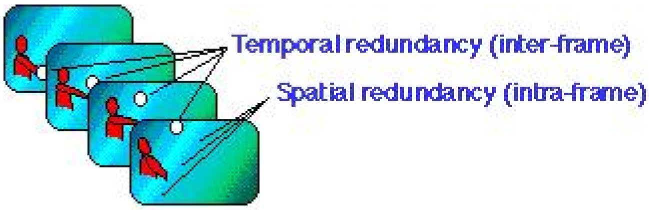


Figure 2.1: Video Redundancy

2.2.3 H.264 overview

The advent of H.264 video encoding technology has been met with great enthusiasm in the video industry. H264 has video quality similar to that of MPEG-2, but is more economical with its use of bandwidth. Being less expensive to distribute, H.264 is a natural choice of broadcasters who are trying to find cost effective ways of distributing HDTV channels and reducing the cost of carrying conventional SD channels.

H.264 over MPEG

Figure 2.2 shows the following three points that leads H264 over MPEG.

- Higher video quality at a given bit-rate
- Higher resolution
- Lower storage requirements

2.3 Audio/Video Synchronization

2.3.1 Definition:

Audio to video synchronization also known as audio video sync, audio/video sync, AV-sync, lip sync, or by the lack of it: lip sync error, lip-flap refers to the relative timing of audio (sound) and video (image) parts during creation, post-production (mixing), transmission, reception and play-back processing.[1]

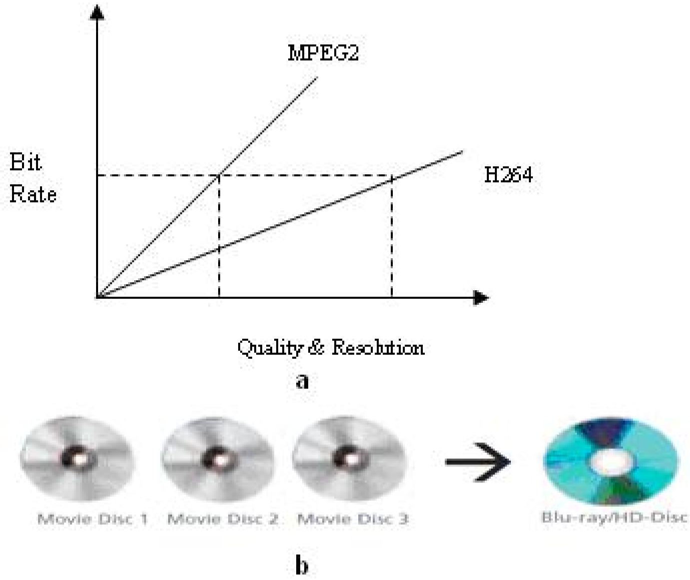


Figure 2.2: a) Video Quality and Resolution b) Storage requirements

If the audio and video corresponding to an audio-visual sequence are decoded and rendered such that the audio is matching to the lip movements in the video, then we can say that audio and video are being rendered in synchronization.[1]

2.3.2 Effects of Audio-Video Asynchrony

Laboratory experiment conducted at Stanford University on 18 adults between the ages of 19 and 45; half men, half women. People viewed three different versions:

Version #1 – Perfect audio-video synchronization
Version #2 – Audio preceded video by 2.5 video fields
Version #3 – Audio preceded video by 5 fields

of six television segments.[7] After each segment, viewers evaluated the speakers in the segments. After viewing all segments, viewers were tested for memory of the segments, and they were asked if they could detect synchronization problems in the segments.

Two different types of questions asked viewers whether they noticed problems with audio-video synchronization:

Question-1:"Do you have any comments about the segments that you just viewed?" (No mention of possible synchronization problems.)

Question-2:"Please indicate in which of the segments you thought that the timing between audio and video may have been off."

Summery of Results:

- When audio precedes video by 5 video fields, viewers evaluate people on television more negatively (e.g. less interesting, more unpleasant, less influential, more agitated, less successful). This difference is not large, but it is statistically significant.
- Viewers can accurately tell when a television segment is in perfect synch, and when it is 5 fields out of synch. Viewers cannot accurately tell the same segments are 2.5 fields out of synch.
- Even though detection is low when asynchrony is moderate (2.5 fields), viewer evaluations are still affected

Implication of Results:

- Audio - visual asynchrony does not inhibit memory for television material, its changes evaluations of television content.
- Awareness of audio-video asynchrony is not necessary for synchronization problems to have psychological effect.
- The magnitude of the effect of audio-video asynchrony is statistically small but quite large to the cost of correction.

Thus, it is required to have a perfect method for AV Synch in consumer electronics, which leads the implication results. Following section introduces the exactly what is AV Synch problem.

2.3.3 What is AV Synchronization or lip synchronization?

This part of the report contains an overview of AV Synch or Lip Synch errors.

The causes of audio to video synchronization errors in video systems, commonly known as lip sync errors, are usually quite subtle. These errors are often the result of buildup of video delays at several locations, without provision for compensating audio delay. The problem is now so pervasive that it receives attention from virtually every segment of the TV industry.

Obviously the advertisers do not want to have their commercials aired with bad lip sync. Many are now monitoring for lip sync problems and asking for "free make-ups", when they find them. Bad lip sync is also a big concern to newscasters, reporters, politicians and others who are trying to convey a message of trust and sincerity to their audience.[8] Without proper lip sync these people can be perceived by the viewers as less interesting, more unpleasant, less influential, and less successful than if the same person were viewed with proper lip sync.

A frequent problem with video delays takes place in MPEG encoding and decoding (as well as for other video delays) is the introduction of large quick video delay changes in the video. These delay changes are caused by instant changes in complexity and motion in the video image, and are often seen on network feeds.

In June 2003, the Advanced Television Systems Committee (ATSC)[9] issued a

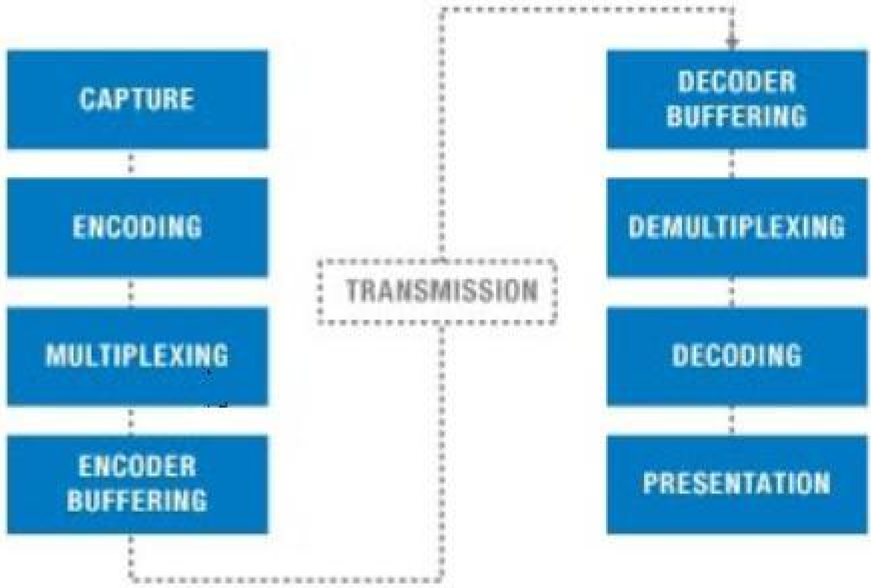


Figure 2.3: Block Diagram of Encoding, Transmission and Decoding of Audio/Video

finding on AV Synchronization is :

"One of the overarching goals of the DTV broadcasting system is to deliver audio and video in proper synchronization to the viewer. At the inputs to the DTV encoding devices, the sound program should never lead the video program by more than 15 milliseconds, and should never lag the video program by more than 45 milliseconds."

In 1993, The International Telecommunication Union (ITU) was apparently thinking along the same lines. In its Draft New Recommendation [DOC. 11/59]4 the ITU reported that errors of +20 ms (audio advanced) and -40 ms (audio delayed) were "detectable" and errors of +40 and -160 ms were "subjectively annoying". The draft recommendation suggested that an overall tolerance (for production, presentation, distribution and transmission) of +20 ms to -40 ms was appropriate.[10]

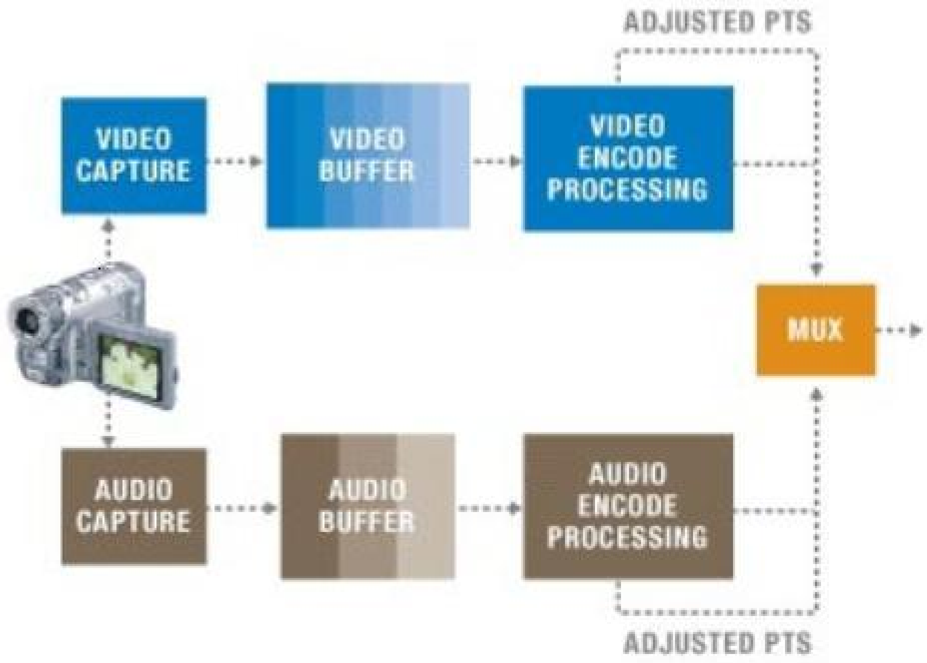


Figure 2.4: Encoding of Audio and Video

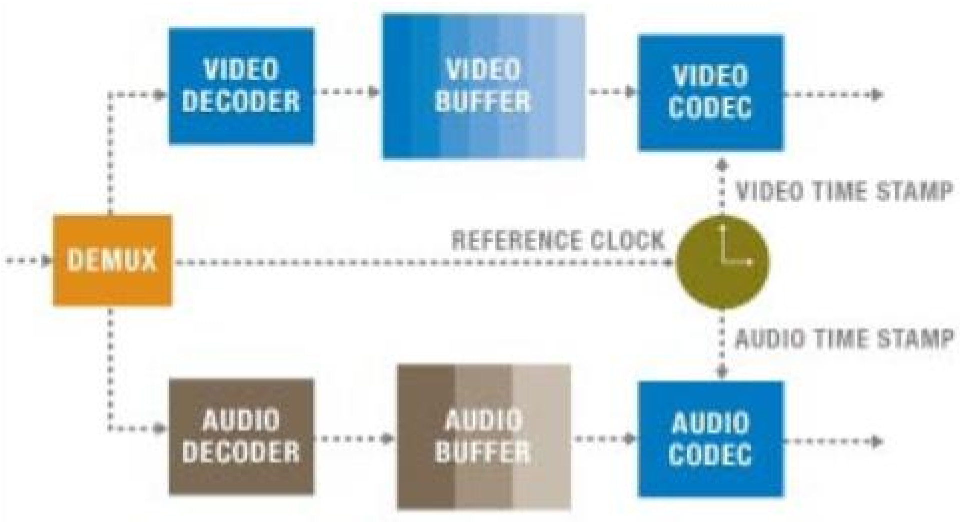


Figure 2.5: Decoding of Audio and Video

2.3.4 Introduction:

With the introduction of advanced digital delivery systems for audio and video, there is an increased awareness of the timing relationship between audio and video. Owing to advanced data compression technologies such as Dolby Digital (AC-3) for audio and MPEG-2 for video, sound is clearer and pictures are sharper. Technologies such as Digital Television (DTV), DVD, Direct Broadcast Satellite (DBS), and Digital Cable use these compression techniques to deliver extremely high-quality programming to consumers. However, it is a misalignment of these same systems that is the root cause of most audio/video synchronization problems. Perhaps this misalignment is due to a lack of understanding how the system functions as a whole.

One of the overarching goals of a digital broadcasting system is to deliver audio and video in proper synchronization to the viewer. The reason for this is each digital audio and video component in the chain, from production to reception, imposes some degree of latency on the signals passing through it. The delays imposed on the audio and video signals are typically unequal. Each component harbors the potential to cause an audio-video synchronization error at its output.

Most film editors are able to detect A/V Sync errors as short as +/- 1/2 film frame. As film is projected at 24fps in the US and 25fps in Europe, this equates to approximately +/- 20msec. It is claimed that some editors can detect even smaller errors, but this might be more accurately attributed to their familiarity with the material being viewed. Other figures abound, such as +/- 1 video frame (+/- 33-40msec). Dolby Laboratories has specified that any Dolby Digital decoder must be with the range of +5msec audio leading video to -15msec audio lagging video. This is because human perception of A/V Sync is weighted more in one direction than the other.

It is a fact that light travels much faster than sound. We are all used to seeing this proven, although as it is such a common situation many times we do not notice. For example, a basketball hitting the court in a large sports venue would appear relatively correct to the first few rows, but the further back a viewer gets, the more the sound lags behind the sight of the ball hitting the floor. The further back you get, the more the sound lags, but it still seems OK.

Now, imagine if the A/V timing was reversed. You are watching a basketball game, and the sound of the ball hitting the court arrives before the ball looks like it makes contact. This would be a very unnatural sight and would seem incorrect even if you were in the first few rows where there was just a small amount of A/V Sync error. The point is that the error is in the "wrong" direction. In summary, human perception is much more forgiving for sound lagging behind sight as this is what we are used to seeing in everyday occurrences.

The MPEG system provides the proper tools to make A/V Sync absolutely correct.[1] Each audio and video frame has a Presentation Time Stamp (PTS) that allows the decoder to reconstruct the sound and pictures in sync. These PTS values are assigned by the Multiplexer in the MPEG encoder. The decoder receives the audio and video data ahead of the PTS values and can therefore use these values to properly present audio and video in sync. It is imperative that audio and video are applied to the Dolby Digital (AC-3) and MPEG-2 encoder In Sync. A very common mistake is to calibrate the multiplexer to compensate for plant differences. Although this may work fine in the short term it should be avoided in permanent installations. Larger problems will likely result including issues with some consumer decoders.

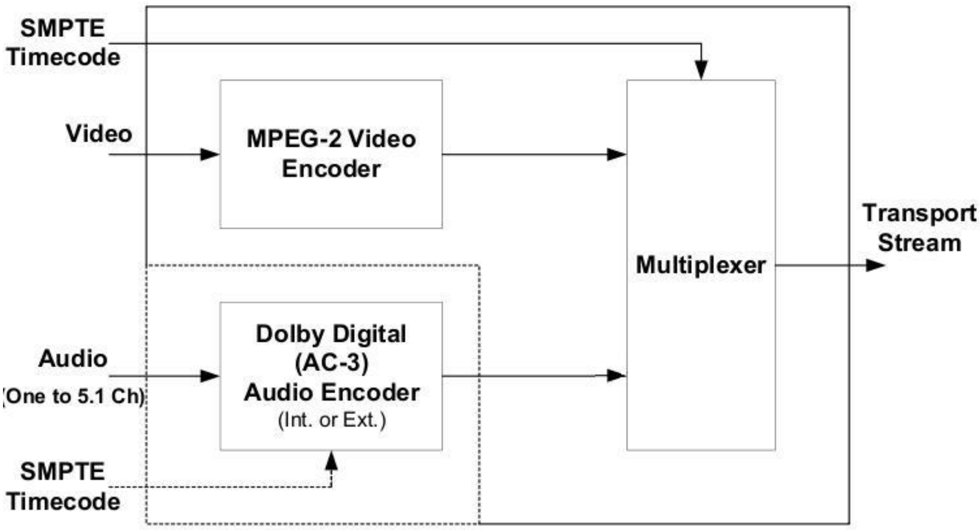


Figure 2.6: Simplified MPEG-2/Dolby Digital (AC-3) encoding system block diagram

2.4 DivX Certification of Mediaplayer

What is DivX?

DivX is a brand name of products created by DivX, Inc. (formerly DivXNetworks, Inc.), including the DivX Codec which has become popular due to its ability to compress lengthy video segments into small sizes while maintaining relatively high visual quality.[2]
The DivX codec uses lossy MPEG−4 Part 2 compression, where quality is balanced against file size for utility. It is one of several codecs commonly associated with "ripping", whereby audio and video multimedia is transferred to a hard disk and transcoded.[2]

Introduction:

The DivX AVI Format is based on the AVI 1.0 standard. In its most basic form, the DivX AVI file container complies with the AVI 1.0 standard. The file extension used by a DivX AVI file can be either of the following:

- .divx (required for DivX Certified Encoders)
- .avi

AVI is an acronym for Audio Video Interleave. It is a media container that packages audio, video, and subtitles for simultaneous playback. The AVI file format is a specific use or special case of the RIFF (Resource Interchange File Format) file structure used with applications that capture, edit, and play back audio-video sequences. The RIFF structure divides data into segments called "chunks". Each chunk is identified by a fourCC code. Every chunk in an AVI file has the following format:

```
[4 byte: 4cc code][4 byte: chunk size][Chunk data]
```

The basic file structure is a single RIFF chunk that contains sub-chunks. The most important sub-chunks are as follows:

- **'hdrl' :** The file header that contains metadata, such as height, width, and number of frames.
- **'movi' :** Contains the specific audio and video data that comprises an AVI movie.
- **'idx1' :** Contains index information of the location of chunks within the file.

The DivX AVI File Structure:

The DivX AVI file contains all AVI 1.0 compliant components. As shown above, the basic structure is the same as a standard AVI file: a single RIFF containing an 'hdrl' chunk, a 'movi' chunk, and an 'idx1' chunk. The differences between the AVI 1.0 standard and the DivX AVI are the components required and rules that dictate the values per field in those components. The DivX AVI defines more specific values and usages for certain file structures and fields, while ignoring some of the standard fields that are specified in AVI 1.0.

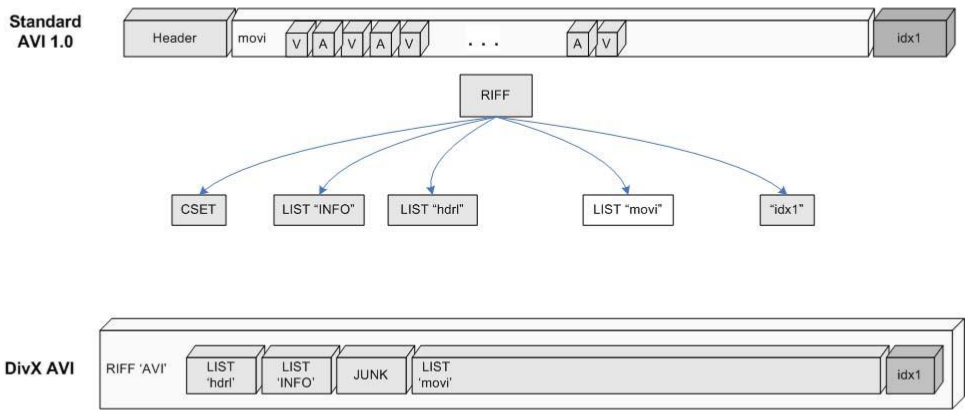


Figure 2.7: DivX format

DivX AVI requires the 'hdrl' LIST, the 'movi' LIST, and the 'idx1' chunk to be present in the file. Other chunks may exist in the DivX AVI, such as the 'INFO' and 'JUNK' chunks, but are considered optional. These optional chunks may or may not exist in a DivX AVI file, but decoder applications must be able to parse a file that contains those chunks without failure. Below is a detailed description of each chunk as shown above in Figure 2.7.

DivX Media Format (DMF):

DivX 6 expanded the scope of DivX from including just a codec and a player by adding a media container format. This optional new file format introduced with DivX 6 is called "DivX Media Format" ("DMF") (with a .divx extension) that includes support for the following DVD−Video and VOB container like features.[3]

DivX Media Format (DMF) features:

- Interactive video menus
- Multiple subtitles (XSUB)
- Multiple audio tracks

- Multiple video streams (for special features like bonus/extra content, just like on DVD−Video movies)
- Chapter points
- Other metadata (XTAG)
- Multiple format
- Partial backwards compatibility with AVI

DivX Subtitles (XSUB):

DivX, Inc. has, since DivX 6, added its own proprietary subtitle tracks that it calls "XSUB". These subtitles are not text-based like many other subtitles, instead they are bitmap (digital image) based like vobsub subtitles for DVD-Video are. And like vobsubs for DVD-Video are supposed to be, XSUB does not come in standalone files but are only embedded in .divx containers, which can be created with Dr. DivX, (Dr. DivX can actually convert/encode XSUB from vobsubs inside DVD-Video). A .divx container can contain multiple XSUB subtitles in several languages.[3]

2.5 Summary

As summery, the chapter contains what is DVB? and why it is used?, video compression standards with MPEG vs H.264, AV Sync as DVR feature and DivX certification of MP.

Chapter 3

Problem Definition and Existing Methodologies

3.1 Problem Definition

The Problem definition is to do Automation of various testing process, benchmarking of various features with the existing solutions and analysis of the ST's DVR & MP solution to support next coming features in future.

3.2 A wavelet based method for audio-video synchronization in broadcasting applications

3.2.1 Introduction:

There are many standards available for films (a photographic film used to record the sequence image) and videos and as they all are using different formats of audio/video generates problems when a conversion from one format to another is required. Since all the images are displayed, the change of frame rate induces a pitch change on the sound.

To avoid this problem, the whole soundtrack in the stream has to be processed during the duplication. This method is all about the corresponding sound transformation problem, namely the dilation of the sound spectrum without changing its duration using the method of Fourier transform and the other mathematical formula leads to Wavelet Transform.

Co-existing difference between film and video is, film contains 24 frames/sec and video (PAL video) contains 25 frames/sec; conversion between one another is required. This is not only applicable to conversion from PAL to NTSC and vice-versa since Europe uses a 50 hertz power supply, the equivalent PAL lines go out at 50 fields per second, or 25 alternating lines. And In the United States and other countries, electrical power is generated at 60 hertz, so for technical reasons the NTSC signal is also sent out at 60 'fields' per second, or 30 alternating lines.

PAL televisions only produce 25 complete frames per second, which can cause some problems with the proper display of motion. If a PAL movie is converted to an NTSC tape, 5 extra frames must be added per second or the action might seem jerky. The opposite is true for an NTSC movie converted to PAL. Five frames must be removed per second or the action may seem unnaturally slow.

3.2.2 Time/Frequency Duality:

Let's consider a time-signal s(t), and its Fourier Transform $S(\omega)$. A pitch shifting of a ratio a (i.e. all frequency is multiplied by this factor) inevitably implies a temporal

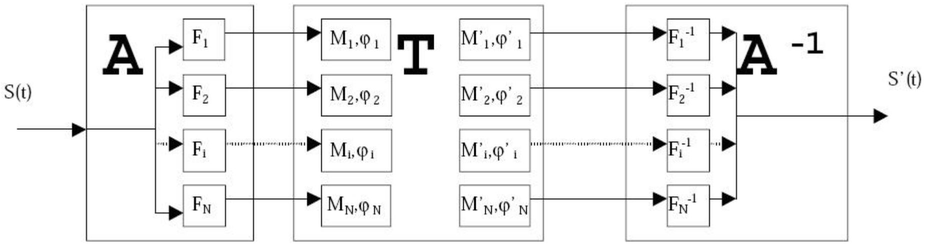


Figure 3.1: Logical block diagram of ATS method[1]

contraction of a factor 1/a:

$$s'(t) = \int S(\alpha\omega)e^{j\omega t}d\omega = \frac{1}{\alpha}s(\frac{t}{\alpha}) \tag{3.1}$$

That's what occurs when the film is projected at a different speed. Actually, a pitch shifting is equivalent to a time stretching without change of pitch, followed by a down sampling. So, using this resampling operation, it is equivalent to consider a pitch shifting or a time stretching problem[11]

3.2.3 Basic Method:

Frequency method operates in three steps:

a. **Analysis:** Given signal is decomposed by linear filtering into no of sub bands.

b. **Transformation:** Modify parameters resulted from each sub bands.

c. **Synthesis:** Recomputes the output signal from modified transformed parameters.

Using the notations in Figure 3.1, we note the output of each sub-band

$$s_i(t) = M_i(t)e^{j\phi_i(t)} \tag{3.2}$$

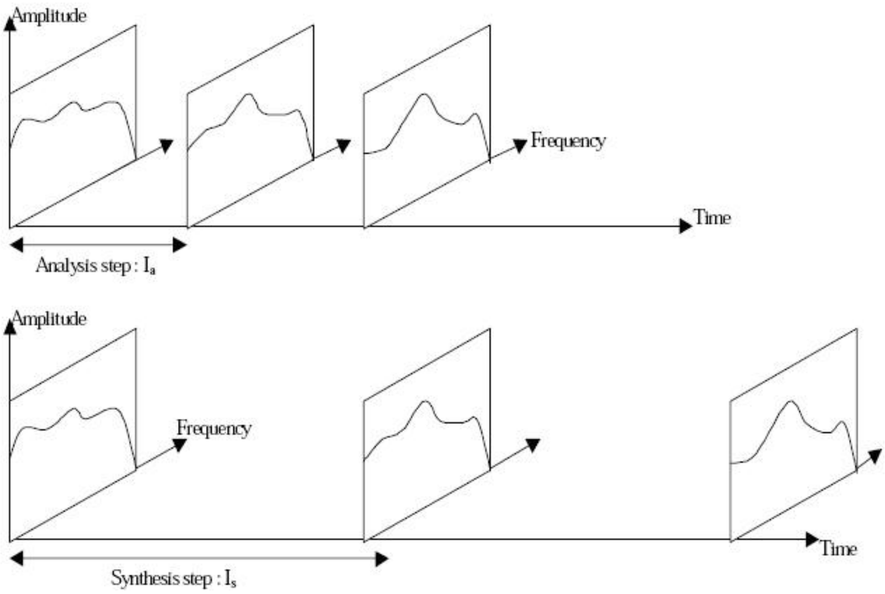


Figure 3.2: Phase Vocoder analysis/synthesis principle[1]

The pitch shifting is obtained by letting identically the modulus and by multiplying the phase of each sub-band by the ratio of transposition.

$$M_i'(t) = M_i(t)\phi_i'(t) = \alpha\phi_i(t) \tag{3.3}$$

The transposed signal is obtained by summing the real parts of the output of each sub-band.

$$S'(t) = \sum_i M_i'(t)\cos(\phi_i'(t)) = \sum_i M_i(t)\cos(\alpha\phi_i(t)) \tag{3.4}$$

Well-known example of this type of method is "The classical Phase Vocoder". The principle of this technique is: "A local spectrum is first calculated at different moments separated by an analysis constant step, and then, the output signal is resynthesized with a step according to the time modification factor (which is the same as the pitch-modification factor)"[1]

Figure 3.2 shows an example of the slowing down of the spectrum.

3.3 Audio-to-Video Synchronization by Audiovisual Correlation Analysis

3.3.1 Introduction:

When audio precedes video by 5 video frames, the satisfaction of the viewers degraded about 84%. Research shows that audio should never lead video by more than 15 ms, and should never lag video by more than 45 ms [12] Causes of AV sync drift :

- Different processing time between video and audio
- Different network transfer delay
- Drift accumulation among different processing stages, etc.

As video processing often includes many stages, even though each stage causes only minor drift, the drift can still be accumulated into an obvious one. As a solution, high quality videos include a final AV-sync adjustment stage, where a dedicated effort of human observers and a special device called an audio synchronizer are used.

Audiovisual correlation analysis is a relatively new research topic and has drawn much attention in recent years. In 1999, Hershey and Movellan[13] first introduced a method to analyze the audiovisual correlation for each pixel using Mutual Information (MI), while this method uses **Quadratic Mutual Information (QMI)**, based on kernel density estimation and can evaluate correlation for arbitrary distributions.

3.3.2 Audiovisual correlation analysis:

The most important step is to find out the correct audio-to-video drift value (AV-drift) of the AV-sync.

Audio feature:

Differential energy is used as audio feature. Input audio data are framed first. The frame duration Ta is set to be the same as the visual frame duration Tv. An overlap of the duration of Ta/2 between each pair of two successive frames is set. Finally, Hamming window is multiplied to the framed audio samples.

The logarithm energy a(t) of each audio frame is computed by

$$a(t) = \log(\frac{1}{M}\sum_{m=1}^{M} s^2(t,m)) \tag{3.5}$$

Where, $s(t,m)$ refers to the processed audio sample m in frame t. M is total length of audio.

The audio feature is defined to be the differential energy between the current and next frames, i.e.,

$$f_{a_t} = a(t+1) - a(t) \tag{3.6}$$

Verification of existence of speech is done by whether or not audio energy a(t) is beyond predefined limit. Frames violates this conditions are dropped.

Visual feature:

Better audiovisual correlation exists not in pixel values, rather between movement of photographed objects and their audio signal. So optical flow is taken as visual feature, which has two elements for each pixel: horizontal and vertical movements; Most speaking action move vertically so is used.

Quadratic mutual information:

The audio and visual features as two random variables and compute their statistical correlation using their temporal samples. Kernel density estimation is a method to estimate the arbitrary probability density function PDF of a random variable.

Given N data points xi Where i=1, 2, , N in d dimension space. Multivariate kernel Density with Kerne KH(x) and symmetric matrix d x d bandwidth in point x is given by,

$$p(x) = \frac{1}{N} \sum_{i=1}^{N} K_H(x - x_i) \tag{3.7}$$

Where, KH() is the specified kernel function, in practice it is Gaussian function

This is called Mutual Information (MI) used by Shannon. But to have closed-form solution in this method calculates QMI. For two random variable x1, and x2 defined as C(x1, x2)

$$\log \frac{\int\int p^2(x1, x2)dx1dx2 \int\int p^2(x1)p^2(x2)dx1dx2}{(\int\int p(x1, x2)p(x1)p(x2)dx1dx2)^2} \tag{3.8}$$

subsectionAV-drift detection: To begin detection, the user first specifies a rough time span where a relatively stationary speaker is speaking. The face of the speaker must be photographed in video with the speech recorded in audio. Only data inside the time span will be used. For video segments having no part satisfying this condition, method cannot detect its AV-drift. Yet, if it has adjacent segments whose AV-drift is detected, it is possible to use this value to recover the current segment.

The optimum AV-drift is searched by first a coarse and then a fine stage. In the coarse search, drift value is quantized into integer times of the video frame duration Tv, i.e.

$$d1 = \{-L, ..., -1, 0, 1, ..., L\} \tag{3.9}$$

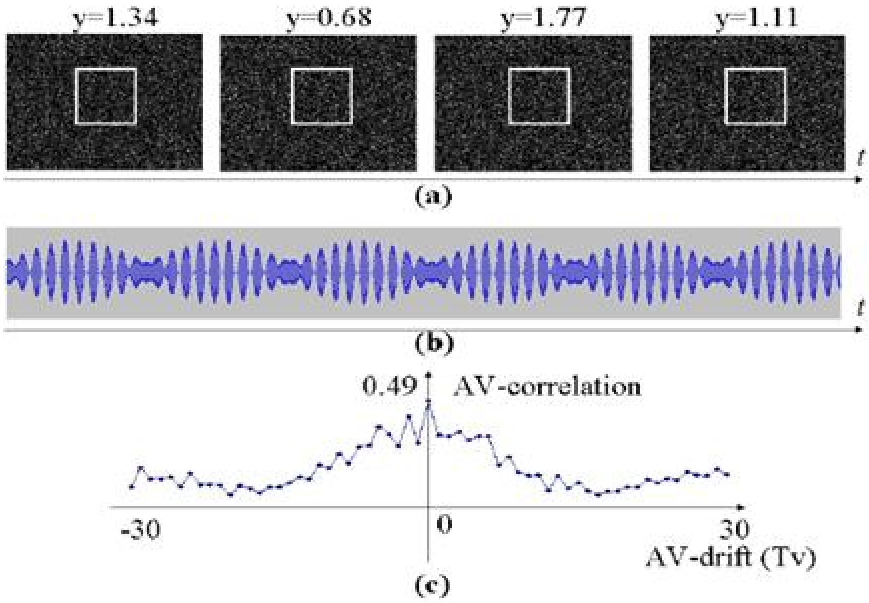


Figure 3.3: (a) shows video frames of the moving random dotted pattern with the vertical shift value y denoted above them. The white rectangle shows the speaker region adopted. (b) Shows the audio signals. (c) Plots the correlation values w.r.t. different AV-drifts [2]

Where, [-L, M] represents a pre-defined search range.

Positive AV-drift values represent how much audio lags video.

Negative ones show how much audio precedes video.

For each drift value d1, the audio data shifted temporally and compute the average audiovisual correlation inside the speaker region. The optimum value d*1 which has maximum average correlation is regarded as the found AVdrift.

3.4 Stand-alone Automated system test method

ST has its own tool for System Testing. All you need to do is just pass the parameters and have the required hardware. The testing is performed on its own and a report is created from where we can check whether the test case has been passed or not. This testing is also performed in both the environments i.e. LINUX as well as OS 21.
Once you click on to the Start Execution, the test will start on its own and certain logs will be created.
Here, after the test is completed, we generate the report. For this we set some parameters, like the OS Type, the device and the platform we are working on. Then there are certain references for which we have to give the complete path and then finally select Report Generation.
After comparing the reference logs and the logs generated during the test execution, according to the TEST PLANS, a report is generated. The report is in the form of an xls, having the Test Cases Description, their status whether PASS or FAIL.

3.5 Summary

This chapter presented work already done in the STDVR, STMP to solve the various requirements from the client and the problem definition.

Chapter 4

The Proposed Algorithm

4.1 AV Sync Methodology in Digital Set-top box

As mentioned in Chapter 3, two different methods of AV synch one is needed while format conversion which are of different frame rate format while the other is too complex to just reduce the human effort such kind of mechanisms used in costly and real time embedded syatems which required AV sync but ST's set-top boxes are cost effective and hence no complex and less costly method is the use of time stamp as described in this chapter.

4.1.1 Introduction:

To achieve this there is a two way process. First process is capturing the two streams and assigning the correct presentation time stamps on the audio and video frames/access units. Second is rendering the audio and video based on the presentation time stamps so that they appear to be in sync with each other. Details about both these processes are discussed below.[14]

a. **Lip Sync:** A visual technique to validate the AV Synch audio visual sequence (sequence of frames which is audible as well as visible).

b. **Audio Video synchronized independent of each other:**

- Audio and video are synchronized to common clock rather then each other, means capture time known as PTS assigned with respect to the common clock.
- Similarly on the decoding side each audio & video frames access unit has an associated PTS with it.
- If we make sure that common clock running on the decoder side has same frequency as that of on the encoder side we can achieve the synchronization.

4.1.2 AVSync components:

a. **Reference clock:** A common clock referred by both Audio and Video during capturing as well as rendering.

- During capture time, samples of this common clock are used to assign the PTS values for both audio and video.
- During rendering time a common clock is referenced by both audio and video renderer t o check whether it is right time to render this particular frames unit or not.

b. **PTS - presentation time stamp:** the time at which the frame is to be presented o r rendered. It's nothing but a sample of the reference clock at a particular instance.

- PTS assigned during capturing of AV data. In real time scenario Video captured at rat e of 25, 30 f/s while audio at 44.1, 96 KHz.
- Once start time is set in the reference clock, is running at 27 MHz, as soon as a frame is captured an encoder will take the sample value of the reference clock and that will be the PTS for that frame.
- This will be provided in bit stream like PES packet.

- E.g. time of reference clock is in hh:mm:ss:msec format. And let say the reference clock is 08:00:00:000 with frame rate 27 f/s. so 1st PTS would be 08:00:00:000 and 2nd would be 08:00:00:040.
- Since PTS value are decoded in the stream, if one try to play this stream when the time reference clock is 09:00:00:000, it will not play as the PTS values are based on the clock starting at 08:00:00:000.
- So rendering engine will reject these frames saying that these frames are delayed. But if we set reference clock to 08:00:00:000 then renderer will not reject it.

c. **Clock recovery:** Ideally if we set the correct start value in the reference clock then on the decoder side everything should work fine. But in real life due to slight error in crystal in the oscillator reference clock are not ideal. As a result it cause buffer overflow/underflow on receiver side if receiver does not have control over incoming data rate. To overcome this clock value is compared with PCR - program clock reference, received in bit stream. On decoder side this clock is adapted to encoder clock. This is known as Clock recovery.

- This is not happened with DVD or recorded playback as in that case we have control over input data rate.
- It is needed with transport stream.
- In this sample of encoder clock are also taken & store in the bit stream at regular intervals.
- On decoder side the start value of reference clock is set based on the first PCR value, received in the incoming bit stream.

4.1.3 Different techniques used to achieve AVSync

Q.1 How to get the clock?

Q.2 How to use the reference clock?

To answer these two questions there are different techniques.

a. Reference clock master and Audio and Video are slave : Used when input data rate is controlled and constant.

b. Keeping either Audio or Video as Master : Used when data transfer is burst mode i.e. usually faster than normal mode and no fixed input rate.

Generally audio is treated as master, since audio is continuous stream of data; a slight drop in data is noticeable. Whenever an audio frame is rendered, generally a call back is sent to the application with the PTS frame. An application compares the value of the latest rendered audio frame.

4.1.4 Algorithm Introduction:

The audio-video synchronization process guarantees continuity of displayed AV data. To initialize the process, a CPU determines whether the specified criteria of a data buffer storing received audio and video frames has been met. If the criterion is met, the CPU obtains an initial time stamp value from an initial frame, and a subsequent time stamp value from a subsequent frame. Initial and subsequent parameters are computed from these respective time stamp values, and are compared against each other. If the parameters match, the frame is valid, and corresponding audio or video frames may be decoded and displayed. If the parameters do not coincide, a recovery process is initiated. In either event, this method makes it possible to achieve audio-video synchronization for both live-playback from tuner and recorded playback from HDD modes of a digital video recorder (DVR).

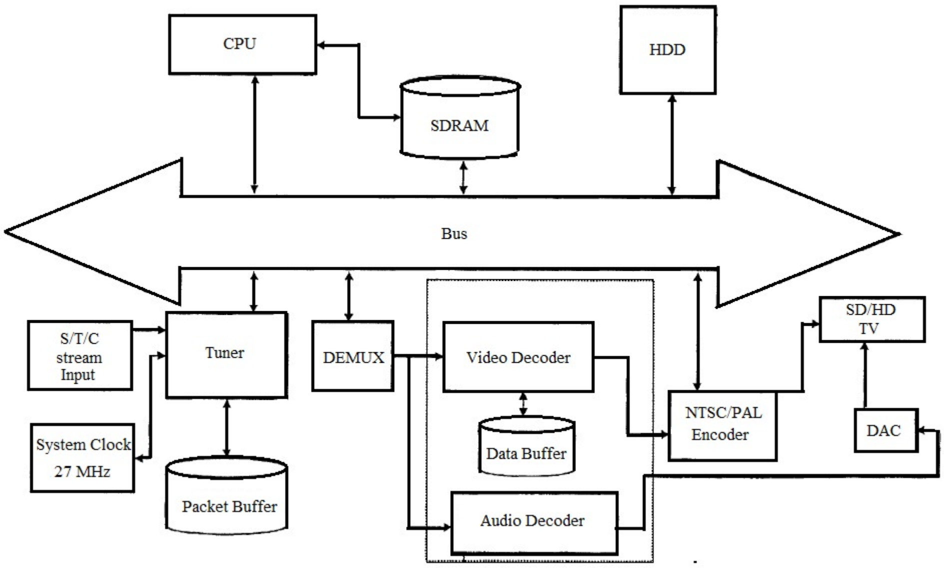


Figure 4.1: Approximate System Architecture of DVR in components like Set-top Boxes

4.1.5 System Architecture:

Following Figure 4.1 shows the system architecture of the DVR, the method of AV Synchronization in either case of DVR.

4.2 Web-based Automated system test method

4.2.1 Architecture

The following Figure 4.2 shows the architecture of the STAutomation Tool. To run this automation tool a server with high configuration, called Test Server is needed. Test Server is having support various supports like,

a. Authenticated Remote login support

b. Authenticated logged in user synchronization

c. External Data Storage with sufficient large capacity

d. MediaPlayer working environment support

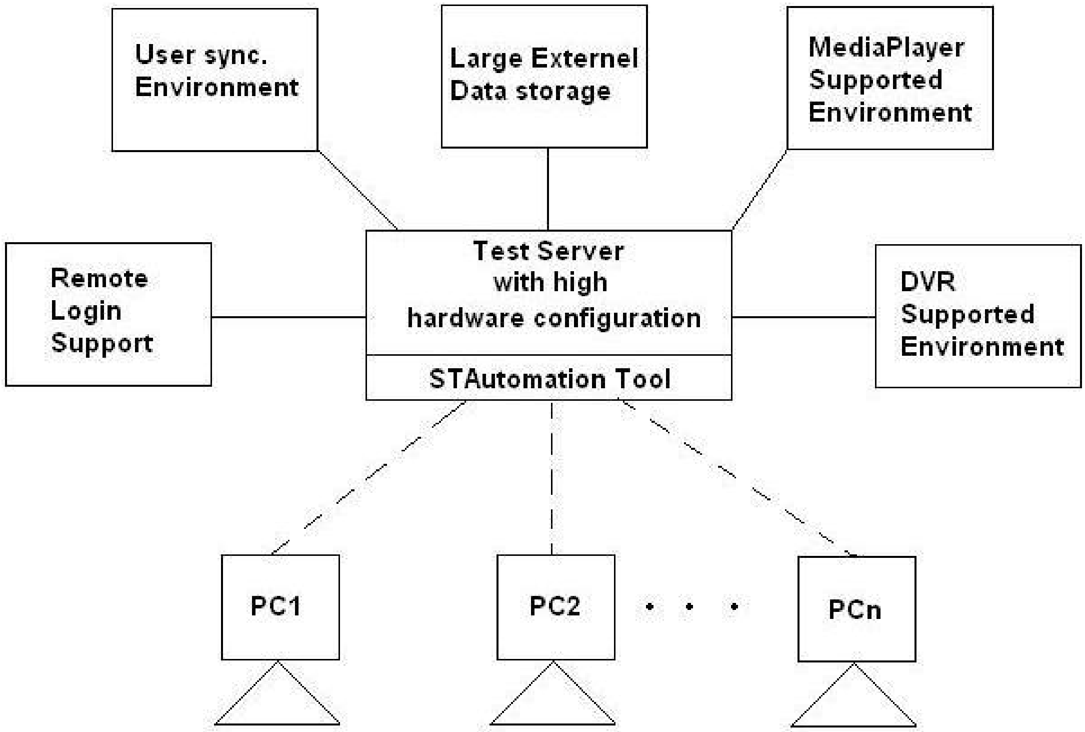


Figure 4.2: STAutomation Architecture

e. DVR working environment support

4.2.2 Data Flow Diagram

Top level context diagram, level-0 DFD of STAutomation Tool is as shown below Figure 4.3.

a. Authentication of user : Each user should be able to login into system if he/she is a valid user.

b. After login to remote server user will copy his/her release code to external data storage.

c. After copy data with label to external data storage user set the compi- lation environment. Default environment is selected using parser and user is allowed to add more options if he/she wants to add or edit the default environment selected by parser.

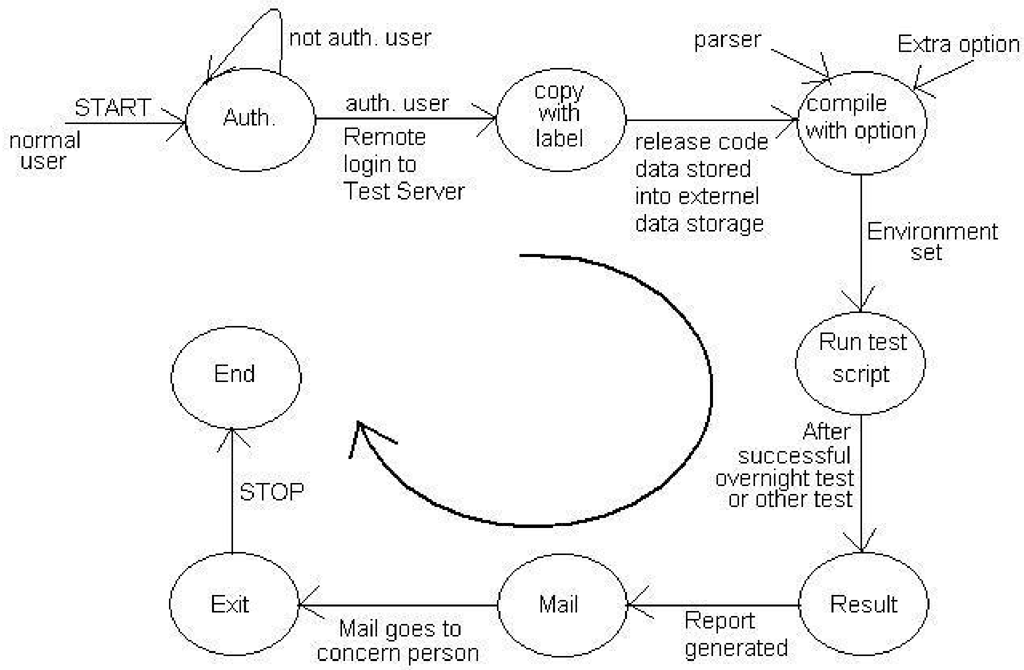


Figure 4.3: STAutomation DFD level-0 or Context Diagram

d. After setting compilation environment user run test script, which starts automatic system test which might be overnight or may be corner test depends on the user type, tester or developer respectively.

e. On a successful test as a result Report is generated.

f. Which is mailed to concern person specified as option to whom it to mailed.

4.3 Summary

By summerising this chapter, AV sych with audio as master methodology proposal, and Web based ST Automatuion Tool is mentioned.

Chapter 5

Implementation

5.1 Implementation of wavelet based AV Synch method:

Down sampling (sampled at 16 KHz) file to speed: **wavlet.m**

```
[orig,sr]=wavread('original.wav');
% 1024 samples is about 60 ms at 16kHz, a good window
y=pvoc(orig,.75,1024);
% Compare original and re synthesis
sound(orig,16000) sound(y,16000)
```

pvoc.m

```
function y = pvoc(x, r, n)
% x is original sound file r is pitch modification factor and n is
% window size for FFT.
hop = n/4;
% hann windowing method requires 25% overlap for smooth
% reconstruction
scale = 2/3;
```

```
% To get Effect of Hanns at both end is a cumulated cos^2 window.
X = scale * stft(x', n, n, hop);
% Calculate the basic STFT, magnitude scaled
[rows, cols] = size(X); t = 0:r:(cols-2);
% Calculate the new timebase samples
X2 = pvsample(X, t, hop);
% Generate the new spectrogram
y = istft(X2, n, n, hop)';
% Invert to a waveform
```

stft.m

```
function d = stft(x, f, w, h)
% Short-time Fourier transform. Returns some frames of STFT of x

s = length(x);

if length(w) == 1
  if w == 0
    % special case: rectangular window
    win = ones(1,f);
  else
    if rem(w, 2) == 0   % force window to be odd-len
      w = w + 1;
    end
    halflen = (w-1)/2;
    halff = f/2;   % midpoint of win
    halfwin = 0.5 * ( 1 + cos( pi * (0:halflen)/halflen));
    win = zeros(1, f);
    acthalflen = min(halff, halflen);
```

```
        win((halff+1):(halff+acthalflen)) = halfwin(1:acthalflen);
        win((halff+1):-1:(halff-acthalflen+2)) = halfwin(1:acthalflen);
      end
    else
      win = w;
      w = length(w);
    end

    c = 1;

    % pre-allocate output array
    d = zeros((1+f/2),1+fix((s-f)/h));

    for b = 0:h:(s-f)
      u = win.*x((b+1):(b+f));
      t = fft(u);
      d(:,c) = t(1:(1+f/2))';
      c = c+1;
    end;
```

pvsample.m

```
function c = pvsample(b, t, hop)

if nargin < 3
  hop = 0;
end [rows,cols] = size(b); N = 2*(rows-1);
if hop == 0
% default value
  hop = N/2;
```

```
end

% Empty output array
c = zeros(rows, length(t));

% Expected phase advance in each bin
dphi = zeros(1,N/2+1); dphi(2:(1 + N/2)) =
(2*pi*hop)./(N./(1:(N/2)));

% Phase accumulator
% Preset to phase of first frame for perfect reconstruction
% in case of 1:1 time scaling
ph = angle(b(:,1));

b = [b,zeros(rows,1)];

ocol = 1; for tt = t
  % Grab the two columns of b
  bcols = b(:,floor(tt)+[1 2]);
  tf = tt - floor(tt);
  bmag = (1-tf)*abs(bcols(:,1)) + tf*(abs(bcols(:,2)));
  % calculate phase advance
  dp = angle(bcols(:,2)) - angle(bcols(:,1)) - dphi';
  % Reduce to -pi:pi range
  dp = dp - 2 * pi * round(dp/(2*pi));
  % Save the column
  c(:,ocol) = bmag .* exp(j*ph);
  % Cumulate phase, ready for next frame
  ph = ph + dphi' + dp;
```

```
  ocol = ocol+1;
end
```

istft.m

```
function x = istft(d, ftsize, w, h)

s = size(d);

cols = s(2); xlen = ftsize + (cols-1)*h; x = zeros(1,xlen);

if length(w) == 1
  if w == 0
    % special case: rectangular window
    win = ones(1,ftsize);
  else
    if rem(w, 2) == 0   % force window to be odd-len
      w = w + 1;
    end
    halflen = (w-1)/2;
    halff = ftsize/2;
    halfwin = 0.5 * ( 1 + cos( pi * (0:halflen)/halflen));
    win = zeros(1, ftsize);
    acthalflen = min(halff, halflen);
    win((halff+1):(halff+acthalflen)) = halfwin(1:acthalflen);
    win((halff+1):-1:(halff-acthalflen+2)) = halfwin(1:acthalflen);
    % 2009-01-06: Make stft-istft loop be identity
    win = 2/3*win;
  end
else
```

```
    win = w;
    w = length(win);
  end

  for b = 0:h:(h*(cols-1))
    ft = d(:,1+b/h)';
    ft = [ft, conj(ft([((ftsize/2)):-1:2]))];
    px = real(ifft(ft));
    x((b+1):(b+ftsize)) = x((b+1):(b+ftsize))+px.*win;
  end;
```

5.2 Implementation of Time stamping method for AV Synch.

5.2.1 Implementation Flowchart:

AV synchronization cannot be achieved for live and playback modes without the use of additional hardware components. In a typical digital broadcast system, AV synchronization is achieved by using a System Clock Reference (SCR). The SCR is frequently embedded in the data stream and in a corresponding time stamp (TS) when the SCR is received by the system. Typically, the TS must be latched through a hardware component handling the transport stream. Therefore, for proper AV synchronization of a recorded event, these SCR and TS values are also required to be recorded, in addition to the entertainment content. This is so an inter-arrival time between the packets that are to be recorded is maintained. This adds to complexity of the system, as well as to the cost, since greater storage is required. This may result in slower system processing time. Moreover, if each frame does not have a corresponding SCR and TS therein, or the SCR and/or TS is not properly recorded, processing of

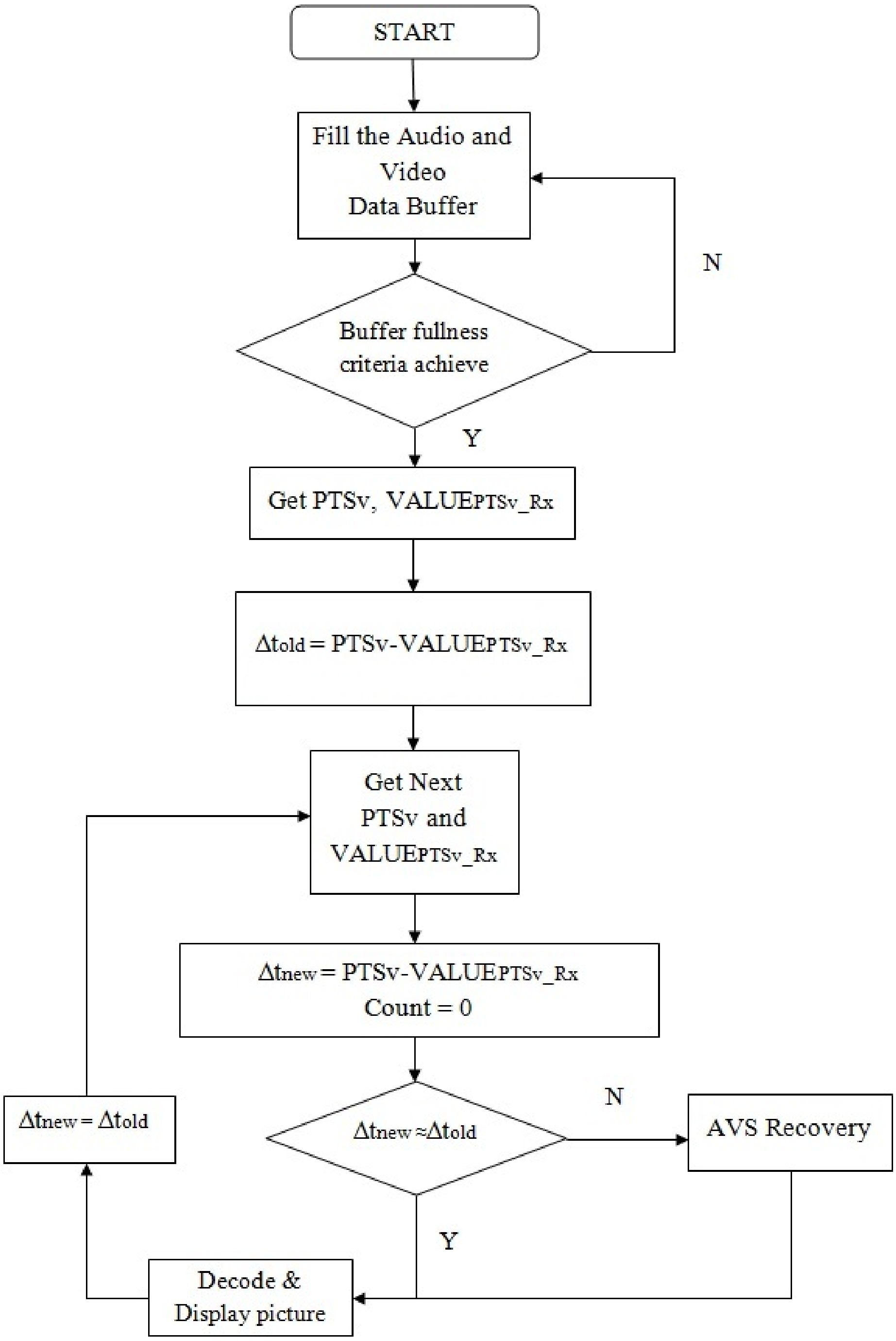


Figure 5.1: Implementation of AV Synchronization in DVR

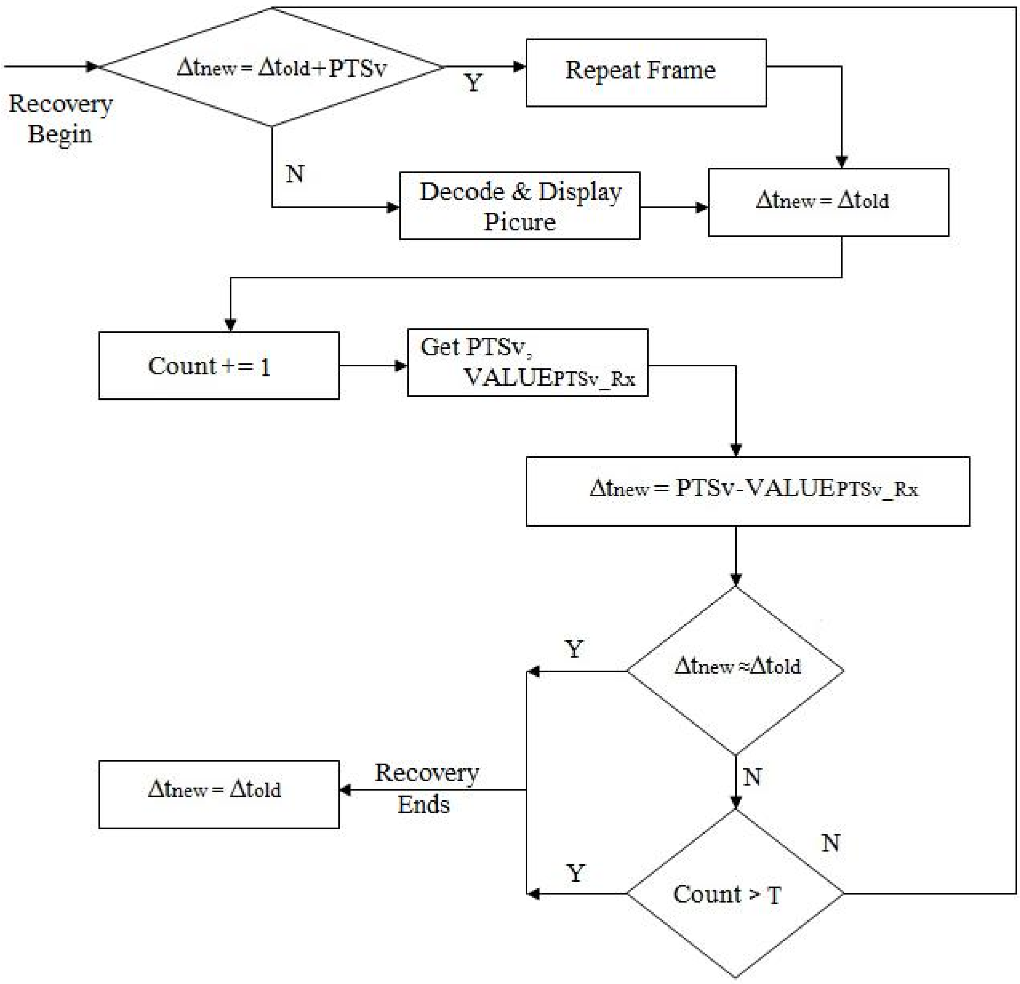


Figure 5.2: Implementation of Recover in AV Synchronization in DVR

these audio and video frames of the displayed program or event may create errors, such as a program where the audio portion lags or leads the corresponding video portion. Such is undesirable whether watching live or recorded content.

5.3 DivX Certification

5.3.1 Definition

DivX Certification is the process which offer a development and verification process that allows consumer electronics manufacturers and software publishers to integrate support for free and premium DivX video into their products.

5.3.2 Why it is required?

- Ensure interoperability with a valuable ecosystem of DivX products.
- Increase their visibility in a crowded marketplace through product differentiation.
- Raise their brand recognition among millions of DivX users worldwide.
- Give consumers access to the best quality digital media across all platforms.

5.3.3 Introduction

DivX provides a digital media licensing solution for independent publishers, aggregators and premium content providers to monetize their video content and broaden the scope of their online offerings through both rental and sale. With an established global base of more than 100 million DivX Certified consumer electronics devices, only DivX can offer consumers a high-quality digital video experience that spans multiple platforms from the computer to the television as well as on the go with mobile phones and portable devices.

5.3.4 Steps to perform

a. Download the "DivX HD Reference Player" and install it on your PC.

b. Load the DivX Certification Kit test clips in this kit to a single directory on the Reference Player hard drive.

c. Set up two TV monitors with stereo audio input/output. One TV should be a 4:3-capable unit and the other should be a High Definition 16:9 capable unit. Prepare and calibrate these two mon

d. Load the certification DVD/CD containing the appropriate test clips on the Reference Player and STMP.

e. For each category of certifications : 1. Audio/Video and Subtitle Test 2. Playback Stability Test and 3. VOD Test Instructions.

   - Select the first test clip in sequence on both the Reference Player and the STMP.
   - Play the clip simultaneously on each device.
   - Observe the playback on both Reference Player and STMP. Perform the following for each clip: Fast Forward, Fast Reverse, Pause, Resume Look for any issues or differences between the Reference Player and the STMP.
   - Use the Test Evaluation criteria for your category to evaluate Pass/Fail.
   - Record the test results in the appropriate Test Results Form

f. When you are done with the clip, restart at Step 5 above for the next clip. Repeat these tasks for each clip under each certification category in this kit.

5.3.5 Various Certifications

Logos that can be applied on media player after various divx certifications are as shown in Figure 5.3 respectively.

Figure 5.3: Various logo (a) DivX Cetrified (b) DivX Certified for HD 720p and 1080p (c) DivX Certified for DivX Plus HD

DivX Certified:

The DivX Certified Program ensures support for DivX video creation and/or playback. Devices such as DVD players, portable media players and digital televisions may be DivX Certified. DivX Certified also supports a DivX Mobile profile created specifically and optimized for mobile devices. Digital cameras, DVD recorders and software programs that create DivX video may also be DivX Certified.

DivX Certified for HD 720p:

The DivX Certified Program for HD ensures support for creation and playback of DivX video up to HD 720p and HD 1080p. Devices such as Blu−ray disc players, gaming consoles, set−top boxes and digital televisions; and recording devices and software programs that create DivX video up to HD 720p and HD 1080p may be DivX Certified for HD.

DivX Certified for DivX Plus HD:

The DivX Certified Program for DivX Plus HD provides the most comprehensive level of certification that includes support for DivX and DivX Plus video up to HD 1080p. High definition devices such as Blu−ray disc players, gaming consoles, digital televisions and set−top boxes may be DivX Certified for DivX plus HD.

5.3.6 Implementation for Certification Automated way:

There is test generator tool is used as client to send the commend to the booted board. In the back-end of Test generator one .cpp file is implemented so as to make DRM and Non-DRM certification test. The process performs as stated below :

a. Once board is booted it lock the terminal on which it gets booted.

b. Run the test generator

c. Command is defined for each functionality of MP, using which the command is sent.

d. Sent Command is internally call the according Mediaplayer function API.

e. Mediaplayer perform the desired task.

Thus by means of sequence of command is sent by means of test generator so as to make the whole process automatically.

5.4 STAutomationTool

5.4.1 Implementaion-1 & 2

```
Input: 1. Username & password.
       2. Absolute path of release code
       3. Board type & OS.
       4. Compilation options.
       5. System test configuration options.
       6. Mail recipient.
Output: test-report mailed to recipient.
START:
- read username & password and call authentication.
```

```
- if(user is authenticated)
    set all rights to access resources.
  else
    error message & login again
- various options 1. New test 2. Test in queue 3. Test
  completed.
- case(option)
  1. New test:
  - Upload release code. (1 way: DROPPED)
  - Select board type & OS.
  - Set working environment by setting various
    compilation options.
  - Set system test configuration.
  - Input mail recipient.
  - Upload release code. (2nd way: DROPPED)
  - Lock resources.
  - Run test on working setup.
  - Unlock resources
  2. Test in queue:
  - if(resources are locked)
        show the current test info.
    else
        Message "No Test in Queue"
  3. Test completed:
  - Show test completed as report generated.
- Exit END:
```

5.4.2 Implementaion-3

```
Input: 1. Username & password.
       2. Board type & OS.
       3. Compilation options.
       4. System test configuration options.
       5. Mail recipient.
Output: test-report mailed to recipient.
START:
- read username & password & call authentication.
- if(user is authenticated)
    set all rights to access resources.
  else
    error message & login again
- various options 1. New test 2. Test in queue 3. Test
  completed.
- case(option)
  1. New test:
  - Access user code from repository server.
  - Select board type & OS.
    (1.parse the env. File 2.extract options avails.
     3.show it)
  - Set working environment by setting various
    compilation options.(same as above)
  - Set system test configuration.
    (1.static display 2.save it to text file to reuse
     next time)
  - Input mail recipient.
  - Lock resources.
```

```
    - Run test on working setup.
    - Unlock resources
    2. Test in queue:
    - if(resources are locked)
          show the current test info.
      else
          Message "No Test in Queue"
    3. Test completed:
    - Show test completed as report generated.
  - Exit END:
```

5.5 Tools & Technologies used

5.5.1 Source insight:

Source Insight is a revolutionary project oriented program code editor and code browser, with built−in analysis for C/C++, C#, and Java programs, as well as other languages.[4]

5.5.2 Beyond compare:

Beyond Compare 3 is the ideal tool for comparing files and folders on your Windows or Linux system. Visualize changes in your code and carefully reconcile them.[5]

5.5.3 PuTTY:

PuTTY is a free implementation of Telnet and SSH for Win32 and Unix platforms, along with an xterm terminal emulator. It is written and maintained primarily by Simon Tatham.[6]

5.5.4 Netbean IDE:

Fully-featured Java IDE written completely in Java, with many modules available, such as: debugger, form editor, etc

5.6 Summary

This chapter summerized DivX Certification and three different implementation of STAutomationTool with the various tools & technologies used.

Chapter 6

Simulation results & Analysis

6.1 DivX Certification report

As mentioned in Chapter 5, i performed the DivX Certification and the report can not be displayed here because of the Proprietary information with DivX, inc. Pvt. Ltd. but it looks alike the report of DVR testing.

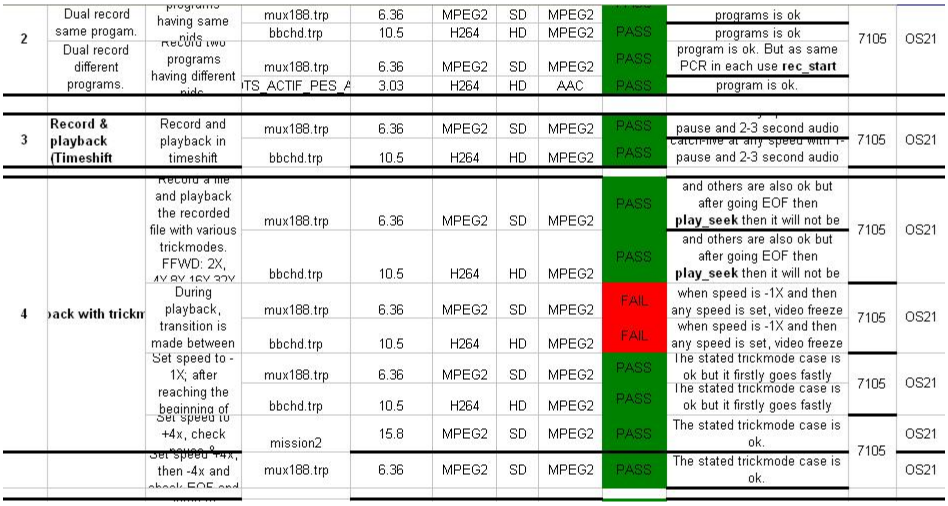
6.2 DVR Test-case results

6.2.1 Corner Testing

This is manual testing, where we have certain test cases and we try to reproduce them. For this type of testing, we need different ST boards and have to be perfomed on OS 21 and LINUX. The testing has all the test cases linked with the Use-Cases discussed above.The test report is as shown in Figure 6.1.

6.2.2 Automated System Test

Overnight Automated Test mentioned in Chapter 3 and the test report is as shown in Figure 6.2.



2	Dual record same progam.	programs having same pids	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	programs is ok	7105	OS21
			bbchd.trp	10.5	H264	HD	MPEG2	PASS	programs is ok		
	Dual record different programs.	Record two programs having different pids	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	program is ok. But as same PCR in each use **rec_start**		
			TS_ACTIF_PES_A	3.03	H264	HD	AAC	PASS	program is ok.		
3	**Record & playback (Timeshift**	Record and playback in timeshift	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	pause and 2-3 second audio	7105	OS21
			bbchd.trp	10.5	H264	HD	MPEG2	PASS	catch-live at any speed with 1- pause and 2-3 second audio		
4	**back with trickm**	Record a file and playback the recorded file with various trickmodes. FFWD: 2X, 4X 8X 16X 32X	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	and others are also ok but after going EOF then **play_seek** then it will not be	7105	OS21
			bbchd.trp	10.5	H264	HD	MPEG2	PASS	and others are also ok but after going EOF then **play_seek** then it will not be		
		During playback, transition is made between	mux188.trp	6.36	MPEG2	SD	MPEG2	FAIL	when speed is -1X and then any speed is set, video freeze	7105	OS21
			bbchd.trp	10.5	H264	HD	MPEG2	FAIL	when speed is -1X and then any speed is set, video freeze		
		Set speed to -1X; after reaching the beginning of	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	The stated trickmode case is ok but it firstly goes fastly	7105	OS21
			bbchd.trp	10.5	H264	HD	MPEG2	PASS	The stated trickmode case is ok but it firstly goes fastly		
		Set speed to +4x, check pause &	mission2	15.8	MPEG2	SD	MPEG2	PASS	The stated trickmode case is ok.	7105	OS21
		Set speed +4x, then -4x and check EOF and	mux188.trp	6.36	MPEG2	SD	MPEG2	PASS	The stated trickmode case is ok.		OS21

Figure 6.1: Corner Test Report

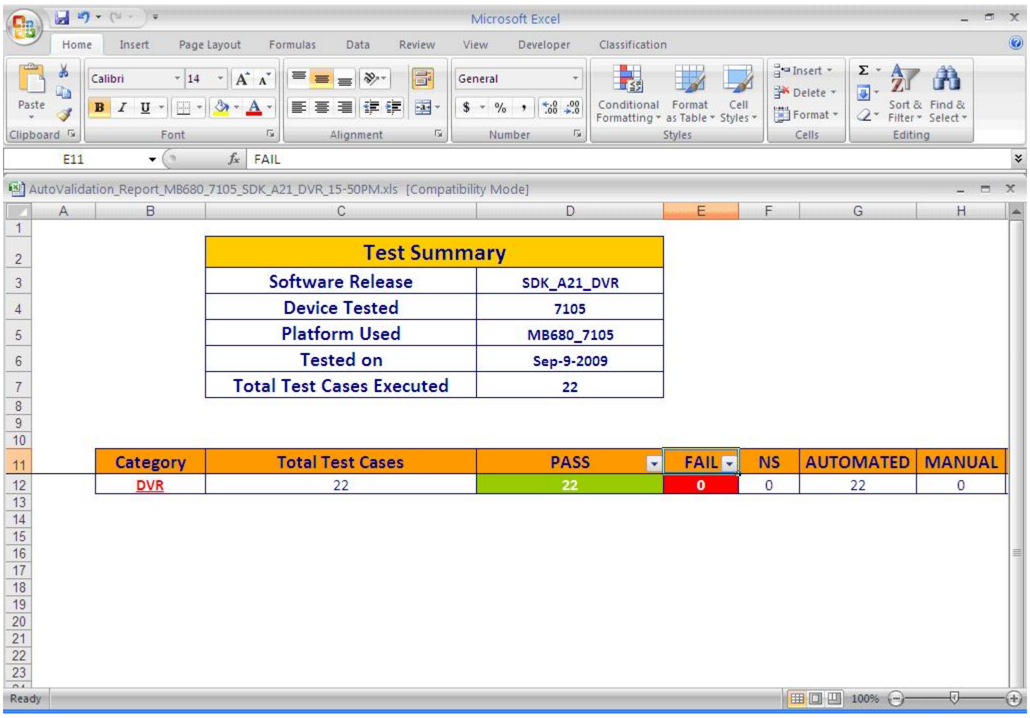


Figure 6.2: Automated System Test Report

NANV	No Audio No Video
NP	Can't play
VQ	Video Quality Not Good ,macroblock,merky..
JV	Jerky Video
NA	No Audio
NV	No Video
OK	Good
VF	Video Flicker
SH	System Hangup
CR	Crash
AN	Audio Noise
NS	Not Supported
AV	AVSync incorrect
UU	Unable to Unpause
ER	Error Look for Details
VF	Video Freeze
FN	File Not found
UI	Unable to Identify Codec,One of the Codec is invalid either Audio or Video
CF	Codec or feature Not supported
NOK	Not Good
NT	Test case not executed
N	Stream Not Tested
Y	Stream Tested

Figure 6.3: Abbreviation of test acceptance test

6.3 MP Test-case results

A very long time test of mediaplayer, for the release of the same is performed and the reports looks like below figures.

Figure 6.3 shows the abbreviation of various test cases result.

Figure 6.4 shows the various test cases performed on various streams with different format.

Figure 6.5 shows the overall result of various test cases.

Figure 6.6 shows the overall result of various test cases in the form of chart.

NO	file name	SDK				v.codec	a.codec	Audio Sample Rate	Audio Number of channels	AudioBitsPerSample	Video frame rate	Avg Byte/sec	resolution	frame type	Note	Location (Directory)
		seek	trick	play pasue play	playback & avsync											
1	_avi/STORE/AVI_RM-MPEG4_MPEG1-L3_720x480_02min13sec_bring_it_on.01.avi	SH		OK	OK	MPEG4	MP3	2	48000	0	23975	####	720x480			Streams/_AVI
2	_avi/STORE/CONT-1H.avi	ER		OK	OK	MPEG4	MP3	2	44100	0	25000	####	720x400			Streams/_AVI
3	_avi/STORE/Death_TUC.avi	SH		ER	ER	MPEG4	MP3	2	48000	0	29970	####	720x480			Streams/_AVI
4	_avi/STORE/Halo3_Announcement.avi	SH		SH	SH	MPEG4	AC3	5	48000	0	29970	####	720x304			Streams/_AVI
5	_avi/STORE/MeetTheRobinson.avi	OK		UU	OK	MPEG4	AC3	5	48000	0	29970	####	720x384			Streams/_AVI
6	_avi/STORE/SUB-13.avi	VF		OK	OK	MPEG4	MP3	2	48000	0	25000	####	720x400			Streams/_AVI
7	_avi/STORE/avi_xvid_mp3.avi	ER		OK	OK	MPEG4	MP3	2	48000	0	29970	####	640x352			Streams/_AVI
8	_avi/STORE/c06_576x320_divx505b_1mbps_mp3.avi	OK		UU	OK	MPEG4	MP3	2	44100	0	23976	####	576x320			Streams/_AVI
9	_avi/STORE/power92percent.avi	ER		OK	OK	MPEG4	MP3	2	32000	0	29971	####	640x480			Streams/_AVI
10	_avi/STORE/sb_pro_pe_02.avi	OK		UU	OK	MPEG4	MP3	2	22050	0	30000	6967	640x512			Streams/_AVI
11	_avi/STORE/AVI_DivX5_MPEG1-L3_640x352_00min25sec_DRM_Homethe	ER		ER	ER											Streams/_AVI
12	_wmv/STORE/310.wmv	SH		SH	SH	WMV3	WMA_PRO	48000	2	24	25000	####	720x426			Streams/_WMV
13	_wmv/STORE/AutoKauf.wmv	NANV		NANV	NANV	WMV3	WMAV2	48000	2	16	25000	8001	512x288			Streams/_WMV
14	_wmv/STORE/Dust_to_Glory_720.wmv	OK		UU	ER	WMV3	WMA_PRO	48000	6	16	23980	####	1280x720			Streams/_WMV
15	_wmv/STORE/KillBillVol2_720.wmv	OK		OK	ER	WMV3	WMA_PRO	48000	6	16	23976	####	1280x720			Streams/_WMV
16	_wmv/STORE/LunarSurface_HD_1080i.wmv	ER		OK	JV	WMV3	WMAV2	48000	2	16	30000	####	1440x810			Streams/_WMV
17	_wmv/STORE/O-WVM_P-02_PAL_VBR_6M_384_44_6_16.wmv	OK		OK	ER	WMV3	WMA_PRO	44100	6	16	25000	####	720x576			Streams/_WMV
18	_wmv/STORE/ScoobyDoo2_720.wmv	SH		SH	SH	WMV3	WMA_PRO	48000	6	16	23976	####	1280x720			Streams/_WMV
19	_wmv/STORE/SuperSpeedway.wmv	SH		SH	SH	WMV3	WMA_PRO	48000	6	24	23976	####	1280x720			Streams/_WMV
20	_wmv/STORE/T2_1080.wmv	OK		OK	JV	WMV3	WMA_PRO	48000	6	24	23976	####	1440x1080			Streams/_WMV
21	_wmv/STORE/T2_1280n720.wmv	OK		OK	JV	WMV3	WMA_PRO	48000	6	24	23976	####	1280x720			Streams/_WMV
22	_wmv/STORE/The_Magic_of_Flight_720_WMA_audio.wmv	ER		OK	OK	WMV3	WMAV2	48000	2	16	23976	####	1280x720			Streams/_WMV
23	_wmv/STORE/WMV_WMV3_WMA2_320x240_01min18sec_wmv_wmv3_	ER		OK	JV	WMV3	WMAV2	16000	2	16	29970	2000	320x240			Streams/_WMV
24	_wmv/STORE/WMV_WMV3_WMA2_640x480_01min30sec_conan.wmv	ER		OK	OK	WMV3	WMAV2	44100	2	16	30000	####	640x480			Streams/_WMV
25	_wmv/STORE/mv17-l-c.wmv	OK		ER	OK	WMV3	WMAV2	48000	2	16	29970	####	720x480			Streams/_WMV
26	_wmv/STORE/mv17-l-v.wmv	OK		OK	JV	WMV3	WMA_PRO	48000	2	24	29970	####	720x480			Streams/_WMV

Introduction | Index | Chart | BugList | WMA | AVI | TS | WMV | MP4 & MOV | MKV | RMVB | MPG | FLV | 3GP,VOB & VRO | MP3 | WAV | LPCM | ETCs | She

Figure 6.4: Acceptance test on various streams

FileType	No of Files	Bugs reported	Not playing	Not supported	Review bug	fix available	Remaining
WMA	165	1	28				
AVI	249	11	10				
TS	40	0	0				
WMV	143	14	1				
MOV&MP4	333	11	17				
MKV	13	0	3				
RMVB	44	0	0				
MPG	141	13	0				
VOB	3	0	0				
VRO	4	0	1				
3GP	5	0	1				
MP3	181	0	0				
FLV	20	5	2				
WAV	46	1	4				
Total	**1387**	**60**	**67**				

Figure 6.5: Summery of Acceptance test

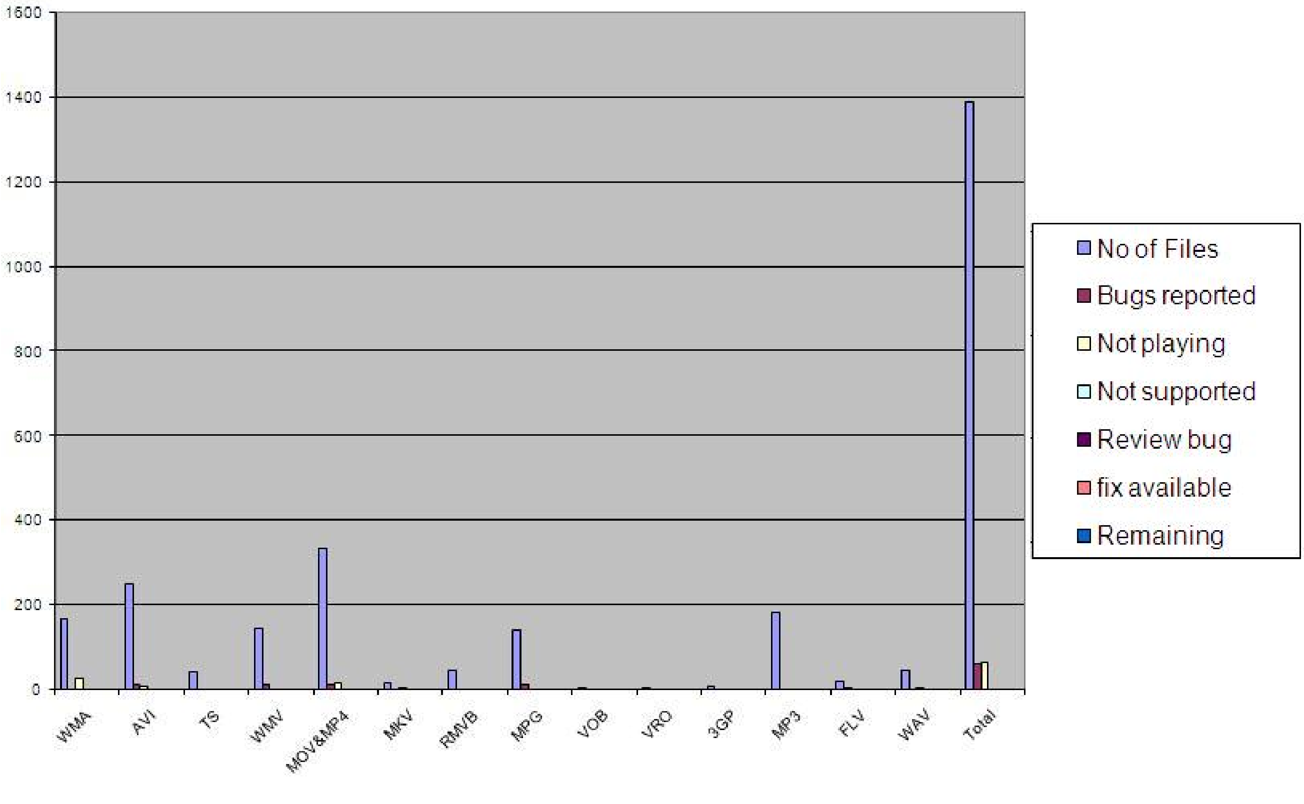


Figure 6.6: Acceptance test on chart

6.4 Analysis

6.4.1 AV Sync:

Wavelet based method:

Advantage:

a. No operation needed with the video signal as in this case just sound transformation need to perform according to the frame rate of video signal.

b. Though the resulting signal is longer than the original one, but contains the same frequencies. We just have to resample this signal to obtain the desired transposed signal.

Limitation: This method gives good results as long as the following assumption made on the input signal is respected: It must be a linear combination of a fixed number of sinusoids, slowly modulated in amplitude and in frequency.

Following Table I shows few absolute values of input audio, output audio and their Difference in digitally quantized form.[3]

Audio visual correlation method:

Advantage:

a. Use of audiovisual correlation analysis to recover AVsync for video segments with little human effort.

b. The development of a measure of audiovisual correlation value using kernel density estimation and QMI.

Assumption: Given a video segment, Assumption is made that its AV-drift is a constant. Considering the reasons resulting in the drift, this assumption stands for most situations. For the drifts that are not completely constant, such as drifts caused

Before WT	After WT	Difference
0.0407	0.0272	0.0135
0.1607	0.1071	0.0536
0.25	0.1667	0.0833
0.1647	0.1098	0.0549
0.0356	0.0237	0.0119
0.211	0.1406	0.0704
0.2391	0.1594	0.0797
0.1003	0.0669	0.0334
0.1085	0.0723	0.0362
0.2416	0.1611	0.0805
0.206	0.1373	0.0687
0.0266	0.0178	0.0088
0.1714	0.1143	0.0571
0.2497	0.1665	0.0832
0.1538	0.1025	0.0513
0.0496	0.033	0.0166

Table I: Difference between Values before Wavelet Transformed applied and After Wavelet Transformed applied on Audio Signal

by sudden network transfer delays, AV-drift is considered to be piecewise constant. The assumption still holds if we divide these videos into segments to process.

Disadvantage: To complex computation just reduce the effort of mankind time is taken by the method to calculate a simple AV Drift is too much high. So its better to have manual calculation.

Time stamping method:

This method offers several advantages. System complexity and costs are reduced since no additional hardware components such as an SCR are needed for synchronization. Since an SCR is not required, AV synchronization of both live and recorded content can be done in an identical fashion, as the algorithms may be used for both live and recorded content.

Additionally, since little processing power is wasted in synchronizing audio and video

Video PTS Value	Consecutive Diff/90	TIMENOW	TIME DIFF	ERROR
24000	41.66666667	2179	40	1.66666667
27750	41.66666667	2219	40	1.66666667
31500	41.66666667	2259	40	1.66666667
35250	41.66666667	2299	40	1.66666667
39000	41.66666667	2339	40	1.66666667
42750	41.66666667	2379	40	1.66666667
46500	41.66666667	2419	40	1.66666667
50250	41.66666667	2459	40	1.66666667

Table II: Time stamp value w.r.t video PTS

frames, a greater amount of processing power is available to perform encryption.

Testcase: Initial AV out of sync for 8−10 seconds At the time of recorded playback, there was a problem in AV Sync of 8 to 10 seconds. So this problem was analyzed and solved by means of this method of AV sync implementation. As shows in Table IIthe statistics of the analyzed stream.

How Analyzed:

a. take system clock as reference : time0, time1, ... : using TimeNow function.

b. take the Video PTS value : PTS0, PTS1, ... : using STVIDPictureInfos structure

Difference of PTS and Difference of PTS values should be almost same, then we can call PTS is well distributed.

Solved: Done by changing No of audio waiting events from 10 to 20 because it is very low bit rate H264 stream

Implementation-1 & 2	**Implementation−3**
Implementation1 is dropped because of the blocking of user while uploading the tree and 2 dropped because of still it is taking too much time and user cant update the code at the same time. If he/she update it then still he/she has to upload again so time consuming process	Implementation−3 is more powerful then 1&2 as far as the time concern but still the problem is how to access the code of particular user from the clear case. This process is under development.

Table III: Comparison of different implementation of STAutomationTool

6.4.2 MP related:

DivX Certification:

The DivX certification has been performed for HD 720p and HT 3.0 and We had 7 and 5 DVDs of streams respectively which are to be played to our Mediaplayer 1.0 and it is to be successfully played, then and then we can put a logo like shown in Figure 5.3. And we got the same now as this is completed successfully.[7]

Acceptance Test:

More then 1600 streams need to be test to accept the MP Release and each takes to pass through all the testcase mentioned above 2-3 min.

8000 min/60 = 67 hours !!!! Approximately 3 full days.!!!

6.4.3 STAutomation:

Comparison

Before this tool it was a task taking time of Avg time per user: more then 8-10 hours.

- Now this tool made it to 6-8 hours. Because of utilizing the ideal resources during night as this will be running overnight and will be executing the test of various users one by one.

- Save the resources because everyone did not have to have a separate hardware setup on each-ones machine.
- No argument like it was successful with my setup as the same setup is going to use for any user.

6.5 Summary

This chapter shows the various results of test case as xls test-report of DVR and the analysis related to DVR, MP and STAutomationTool.

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

In this thesis, study of Set Top Box and Digital Video Recorder data flow is done. This thesis also includes the study of different video compression standards like MPEG and H.264. All the features of Digital Video Recorder and their data flow is studied. After understanding the ST Toolset flow for compiling, linking and debugging, the test suite has been developed for DVR stack and implement this suite in DVR software and verify this for different test cases on different STB and the same is done with media player so as to earn DivX Certification.

Devised web based automated features test of STDVR and has reduced the overall time of system test.

DivX certification helps ST to put a logo that STMP supports DivX format for 720p and HT 3.0.

Implementation of various AV Sync method gives the ease in processing AV playback and so as to reduces the overall cost.

7.2 Future Scope

The future work leads to the AV synchronization problem in the area of **Dual Video (HD & SD) with Audio** in RTOS environment.

Website References

[1] http://en.wikipedia.org/wiki/Audiotovideosynchronization

[2] http://en.wikipedia.org/wiki/DivX

[3] http://www.divx.com/en/partner/certified-programs/program-details

[4] http://www.sourceinsight.com/

[5] http://www.scootersoftware.com/

[6] http://www.chiark.greenend.org.uk/sgtatham/putty/

[7] http://www.tradingmarkets.com/news/.../divx_
st-media-player-on-stmicroelectronics-sti7105-earns-divx-certification-88217
html

References

[1] G. Pallone, “A wavelet based method for audio-video synchronization in broadcasting applications,” December 1999.

[2] Y. Liu, “Recovering audio-to-video synchronization by audiovisual correlation analysis,” 2008.

[3] P. C. Jain, “Introduction to set-top box,” 2008. Section Manager S&M Division STMicroelectronics Pvt. Ltd.

[4] B. Sundareshan, “Digital set top box (stb) - open architecture/interoperability issues,” 1999. Consultant, Broadcast Engineering Consultants India Limited.

[5] F. Xiangdong, “A personal video recording system using h.264 and mpeg-2 transport stream,” Jan 10-14 2007. International Conference on Consumer Electronics.

[6] D. Kucera, “Dvb overview,” November 2002.

[7] B. Reeves, “Effects of audio-video asynchrony on viewer’s memory, evaluation of content and detection ability,” October 1993.

[8] J. C. Cooper, “A short tutorial on lip sync errors,” March 2002.

[9] R. Hoffner, “Ats recommends av sync rules by technology corner,” July 2004.

[10] *International Telecommunication Union Document 11A/47-E*, 13 October 1993.

[11] J. Larocher, “Time and pitch scale modifications of audio signals,” 1998. Applications of DSP to Audio and Acoustics.

[12] R. Hoffner, “Relative timing of sound and vision for broadcast operations,” 2003. Advanced Television Systems Committee report.

[13] J. Hershey and J. Movellan, “Audio vision: Using audiovisual synchrony to locate sounds,” 1999.

[14] B. Soni, “Audio/video synchronization,” 2009. STMicroelectronics Pvt. Ltd.