Fitri N. Rahayu

Quality of Experience for Digital Cinema Presentation

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2011

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

ISBN 978-82-471-3123-7 (printed ver.)
ISBN 978-82-471-3124-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2011:277

Printed by NTNU-trykk

Abstract

Multimedia presentations of digital media services and devices are meant for human consumption and interaction. Before consumption by the user, the multimedia signal usually goes through several processing stages. Depending on the technologies, including the applied signal processing algorithms, some stages can introduce artefacts that reduce the quality of the multimedia presentation. Quality is a fundamental aspect for the design of any end-to-end multimedia signal processing architecture. A sufficiently high quality level of any multimedia presentations must be provided to the user to ensure her optimal experience. More recently, we have seen a shift of paradigm towards incorporating the user as the most important factor in the quality assessment of multimedia presentations. This shift of paradigm drives the creation of the Quality of Experience (QoE) concept. QoE depends on the user perception making it a qualitative assessment as opposed to a purely quantitative one. The definite way of assessing perception of user is by conducting a perception experiment involving human participants in a controlled environment, and this experiment must be carefully designed. Subjective quality assessment is one example of such experiment. There is another, more practical way of assessing quality from user standpoint; this utilizes perceptual-based metrics that model the human perception as closely as possible. Due to the array of current applications, it is unlikely to have a universal quality metric for assessing QoE of multimedia applications. This thesis will only focus on QoE for Digital Cinema presentations.

The thesis is composed of a paper collection; were we have classified the work in this thesis based on research questions within three main themes: QoE of still images for Digital Cinema presentation[1], QoE of motion pictures for Digital Cinema presentation, and QoE of audiovisual presentation for Digital Cinema.

In the field of QoE of images for Digital Cinema presentations, we conducted subjective image quality assessments for Digital Cinema using a methodology derived from standardized recommendations. During the assessment we collected subjective scores of the perceived image quality in a real Digital Cinema environment. We also conducted another perceptual experiment in a Digital Cinema to obtain the parameters of Multi Scale Structural Similarity (MS-SSIM) objective metric for Digital Cinema presentation. Moreover, we analysed the performance of several objective metrics including MS-SSIM with original parameters and parameters obtained from our experiment in the Digital Cinema. The results show that in the case of Digital Cinema, MS-SSIM does not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric.

In the field of QoE of motion pictures for Digital Cinema presentations, we conducted subjective motion pictures quality assessment for Digital Cinema using a careful designed experiment, which is also derived from standardized recommendations. The collected subjective data is used to analyse the performance of two compression algorithms (JPEG 2000 and AVC/H.264) for a Digital Cinema environment; the results showed that temporal compression schemes like H.264/AVC have high coding

[1] This is referred to Digital Cinema applications in Papers A-E

efficiency not only at SD resolutions, but also at high resolutions for Digital Cinema presentation. Furthermore, we performed an analysis on factors that affect visual perceived quality in a Digital Cinema using collected scores from the subjective still images and motion pictures quality assessment.

In the field of QoE of audiovisual presentation for Digital Cinema, we performed subjective experiments of audiovisual contents for Digital Cinema using also methodology derived from standardized recommendations. In addition, we investigated the multimodal effect on perceived quality in a Digital Cinema environment. A major result of our subjective visual quality assessment showed that the presence of audio (low or high quality) does not influence the visual quality judgment.

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) for partial fulfilment of the requirements for the degree of philosophiae doctor. This doctoral work has been performed at the Department of Electronic and Telecommunication, NTNU, Trondheim, with Andrew Perkis as main supervisor and with co-supervisor Touradj Ebrahimi.

In addition to scientific research, the doctorate education consists of compulsory courses equivalent of a full year studies and one year duty works. It spanned the period from July 2007 to July 2011.

This work was funded by the Centre for Quantifiable Quality of Service in Communication Systems, NTNU and the project Network Media Handling.

Acknowledgements

This work would have never been completed with the support of others. I would like to thank my two supervisors Professor Andrew Perkis and Professor Touradj Ebrahimi. They have supported and inspired me throughout my thesis with their patience and knowledge. They helped me to find and stay on the right track and gave me invaluable feedback. I value the discussion and collaboration with Dr. Ulrich Reiter, Professor Peter Svensson, and Dr. Junyong You. I would like to thank Marlon Nielsen from Midgard Media Lab, NTNU who provided me assistance during the experiments. I also would like to thank Trondheim Kino AS for allowing me to conduct experiments at Nova Kinosenter. I am thankful for technical staffs from Trondheim Kino A.S., Kurt Laumann and Knut Erik Slettum, who helped me during the experiments.

All my colleagues at the Q2S, past and present, are thanked for creating a stimulating and warm environment in which to learn and grow. A special thank also to Anniken Skotvoll for her kindness and consideration throughout my stay at Q2S. Last, but not least, I would like to thank my family and friends for their continuous support, encouragement, and understanding.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACR	Absolute Category Rating
ANSI	American National Standards Institute
AV	Audiovisual
AVC	Advanced Video Coding
CI	Confidence Interval
CSF	Contrast Sensitivity Function
D-Cinema	Digital Cinema
DCDM	Digital Cinema Distribution Master
DCI	Digital Cinema Initiatives
DCP	Digital Cinema Package
DCT	Discrete Cosine Transform
DLP	Digital Light Processing
DSM	Digital Source Master
DVI	Digital Visual Interface
ftL	foot-lambert
HAS	Human Auditory System
HD	High-definition
HD-SDI	High-definition Serial Digital Interfaces
HDTV	High-definition Television
HVS	Human Visual System
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITU	International Telecommunication Union
JPEG	Joint Photographic Experts Group
JTC	Joint Technical Commitee
JVT	Joint Video Team
LCOS	Liquid Crystal on Silicon
LSDI	Large-screen Digital Imagery
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
MXF	Material eXchange Format
NTNU	Norwegian University of Science and Technology
NTSC	National Television System Committee
PESQ	Perceptual Evaluation of Speech Quality
PSNR	Peak Signal-to-Noise Ratio
RMSE	Root Mean Square Error
SD	Standard Definition
SDSCE	Simultaneous Double Stimulus Continuous Evaluation
SMPTE	Society of Motion Picture and Television Engineers
SNR	Signal-to-Noise Ratio
SS	Single Stimulus
SSCQE	Single Stimulus Continuous Quality Evaluation

SSIM	Structural Similarity
SS-SSIM	Single Scale Structural Similarity
MS-SSIM	Multi Scale Structural Similarity
UI	User Interface
VESA	Video Electronic Standards Association
VQEG	Video Quality Expert Group
QoE	Quality of Experience
QoS	Quality of Service

1 Introduction

1.1 Motivation

As our world is becoming more and more digitized and connected every day, multimedia presentations are becoming ubiquitous. Multimedia in essence is a presentation of multiple information that may consist of images, video, graphics, audio, speech, sound, text, and even tactile content (content relating to the sense of touch) or olfactory content (content concerned with the sense of smell). These presentations are meant for human consumption and interaction. Before consumption by the user, the multimedia signal usually goes through several processing stages. Figure 1 illustrates the chain of multimedia signal processing from real world to the user. Depending on the technologies utilized in the processing stages, such as multimedia signal processing techniques, some stages can introduce artefacts and errors that reduce the quality. In light of this, optimizing the performance of each stage with the respect of what the users perceive in the signal is one of the most important challenges in this domain. Consequently, quality is a fundamental aspect for the design of any end-to-end multimedia signal processing architecture. This, as illustrated on Figure 1, requires quality assessment, which is important in each stage of multimedia signal processing. Ultimately, a sufficiently high quality level of any multimedia presentations must be provided to the users.

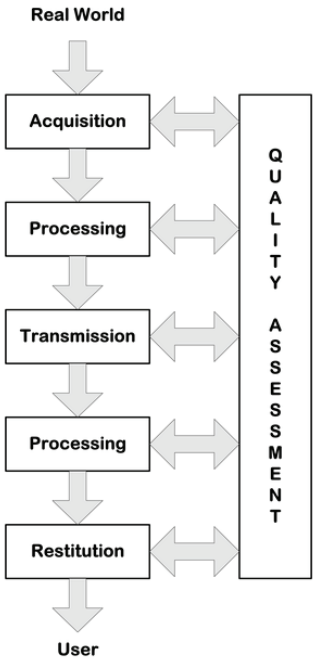


Figure 1: Chain of multimedia signal processing.

Traditionally multimedia content and service providers have addressed that issue by the notion of Quality of Service (QoS) that objectively measures and guarantees service-related characteristics from the providers perspective. Additionally, there are well-established performance standards that rely on special test signals and measurement procedures to determine signal parameters that can be related to the quality.

More recently, we have seen a shift of paradigm towards incorporating the user as the most important factor in the quality measurement; this drives the creation of the Quality of Experience (QoE) concept [1]. The International Telecommunication Union defines QoE as [2]:

> *"The overall acceptability of an application or service, as perceived subjectively by the end-user."*

The framework to assess the user's behaviour and the necessary technology is based on assessing the user experience in a consistent way, and rewarding the user's loyalty through innovative packages and new, engaging services and content delivered through their device of choice whenever and wherever they want it. These assessments are crucial for the industry and drive their innovations and investments in future new media and services. In this light, QoE can also be defined as [3]:

> *"The characteristis of the sensations, perceptions, and expectations of the people as they interact with multimedia applications through their different perceptual sensors (restricted to vision and hearing in an audiovisual context)"*

Since QoE is something that depends on the user perception, it is also a qualitative measure in addition to a quantitative one. Measuring QoE poses many challenges because QoE involves complex and numerous factors including human factors. Some studies have explored the requirements for achieving a good QoE. Developing QoE assessment methods requires also a comprehensive study on experimental design related to user experience and quality involving human participants because, intuitively, the best judge of quality are the users themselves. So far the largest body of research on QoE in multimedia presentations has focused on the perceptual visual or audio quality. It is reasonable because in multimedia presentations, regardless of the application, QoE is dominated by the quality of content which require high bandwidth and considerable processing power. From this point of view, video, image, and audio are most critical in the modelling of QoE, and the need for better understanding of the impact of audio-visual information on perceived quality is critical.

The evaluation of perceived quality is divided into two categories, subjective and objective methods. Subjective methods require human participants in a quality experiment scenario. Accordingly, the subjective methods, which are more known as the subjective quality assessments, are said to be the fundamental way of measuring perceptual quality and so far the only widely recognized method of judging perceived quality [4]. These experiments must be carefully designed in order to create significant and reliable results. In addition, performing subjective quality experiments requires significant knowledge of a number of different disciplines. For that reason subjective

evaluations are complex and time consuming. Even though the result from these methods is considered as the ground truth, faster alternative approaches are needed; subjective quality assessment cannot be utilized in real time scenario such as quality monitoring of several online video channels.

For that reason, more practical approaches to assess perceived quality are desirable. These approaches, which are called objective methods, utilize quality measurement models or metrics that take into account the human perception. Objective methods must be able to reliably measure the perceived quality as closely as possible. At the moment, we have standardised perceptual objective metrics in the field of audio [5-7]. The present-day models typically comprise of a model of the human auditory system followed by a cognitive model to estimate the human participants scoring during subjective assessment test. In the field of audio, two categories of predictive models exist: those that aim to predict a particular perceptual attribute, such as loudness, and those that aim to quantify overall performance, such as speech listening quality. With the development of speech and audio codecs, there has been a desire to evaluate the performance of codecs and associated devices. This has led to development and standardisation of various predictive models associated with speech and audio quality [8]. It is important to note, though, as with all tools, there are both correct and incorrect modes of usage. Perceptual-based audio quality objective methods normally have a particular domain of application beyond which their prediction accuracy is not known. Usage beyond the scope of application is risky and may provide misleading results. For example, one standardised model, PESQ (Perceptual Evaluation of Speech Quality) [7], is developed primarily for the assessment of narrowband speech. This model has been trained extensively with different kinds of speech stimuli, codecs, and other relevant stimuli. Applying such models to audio codecs with music is not automatically accurate and may lead to the misleading prediction of the perceived audio quality. Currently, objective methods beyond audio applications are still evolving, and there are no widely used and standardised objective models for predicting perceived visual quality yet. Additionally, predictive models for audio quality have not yet been developed for all aspects of audio perceptual evaluation, such as metric to measure the quality of spatial sound [8].

Due to the array of current applications, it is less feasible to develop a universal quality model or metric for evaluating QoE of multimedia applications. Different applications can provide different variables due to their situational context. Consequently, the scope of variables that need to be considered during the development of a universal QoE model is too large and too complex. Even in the field of audio only, so far, there exists no unified perceptual-based model for assessing audio quality that can cover all aspects of audio perception. This thesis will only focus on QoE issue for Digital Cinema presentation.

Digital Cinema is a distinct application; it is the latest and final analogue media to go digital. The motion picture industry is one of many in the media sector consisting of mature players which have an entertainment focus in common. Both broadcasting and mobile media are digital services, while the motion picture industry is currently in the process of forming standards for digitization of its complete value chain. The speed of digitization of the entire chain of cinema in the whole world is quite different with others media; it is quite slow. This is particularly evident compare to the field of television. In broadcasting, digital satellite and cable services have been available for

quite some time, and terrestrial digital TV broadcast has been introduces in a number of location around the world. Production studios, broadcasters and network providers have been installing digital video equipment at an ever increasing rate. The speedy development of digitization is also observed in photography, where digital cameras have become hugely popular worldwide within a short amount of time. Going to the movies is the end product of a long process involving a complex value chain. This value chain has developed and operated in the same manner for over 100 years. Innovations have evolved and refined the process. This includes a few major revolutions such as going from silent movies to sound and more recently the last of entertainment industries to go digital [9]. Digital Cinema requires a complete change of infrastructure in all screens worldwide. The traditional 35mm films projector needs to be replaced with a Digital Cinema server and a digital projector. The process of change is referred to as the Digital Cinema roll out which results in exhibitors (the theatres) adopting and starting to use the new technology.

So far quality has not been used explicitly to drive the Digital Cinema roll out, but it is an important factor nonetheless. The motivations for the change are complex and not solely based on quality, and not all benefits are seen by the user. But the open question still remain is whether quality plays a role in the innovation of cinema technology and adoption of Digital Cinema. It is a commonsense assumption to say that all content providers have one goal in common, the satisfied and loyal customer, buying and consuming their services and applications regardless of the technology. Being able to quantify QoE as perceived by end-users can play a major role in the success of future media services, both for companies deploying and with respect to the satisfaction of end-user that use and pay for the services [10]. Accordingly, in the context of Digital Cinema applications, QoE is a noteworthy issue to study. There are at least three main reasons to adopt Digital Cinema [3]:

- To reduce distribution costs (benefit for studios)
- To reduce piracy (benefits for studios)
- To enhance Quality of Experience (benefits for cinema goers – the users)

Digital Cinema is also a distinct application, in a sense that, it needs a special venue—a large auditorium—and very expensive equipments to screen multimedia content. Moreover, Digital Cinema is based on 4K or 2K imagery [2], a significantly higher quality not only in terms of larger pixel counts per image when compared to standard and high definition content, but also offer a higher dynamic range on the values of each pixel. These add additional distinctive factors influencing QoE assessment for Digital Cinema presentation.

This thesis presents the study of QoE issue in Digital Cinema. The main part of this thesis, Part II is a collection of five papers, Paper A-E. All the papers included are modified to fit the format of the dissertation. For the already published articles, any changes (aside from spelling errors) made are noted in the summary of the papers. Part I give an introduction to the areas of research covered in these papers.

[2] 4K is 4096x2160 resolution; 2K is 2048x1080 resolution

1.2 Research Questions

The focus of this thesis was intended to be a QoE research, specifically within formal test of subjective quality assessment for digital cinema presentation. The thesis is based on research questions within three main themes:

RQ1. QoE of still images for Digital Cinema presentation.
RQ2. QoE of motion pictures for Digital Cinema presentation.
RQ3. QoE of audiovisual presentation for Digital Cinema presentation.

During our research, the scope of the research has been narrowed down to these following issues:

- We put emphasize on alternative content beyond feature films screening in digital cinema.
- We did not consider the processing algorithm used in the digital cinema projector and media block.
- We did not consider the issue of intellectual property of the compression technology used in digital cinema, and accordingly, we take also into account the compression technology beyond JPEG 2000.

The main research works are presented in five publications, which are illustrated in Figure 2. These publications are:

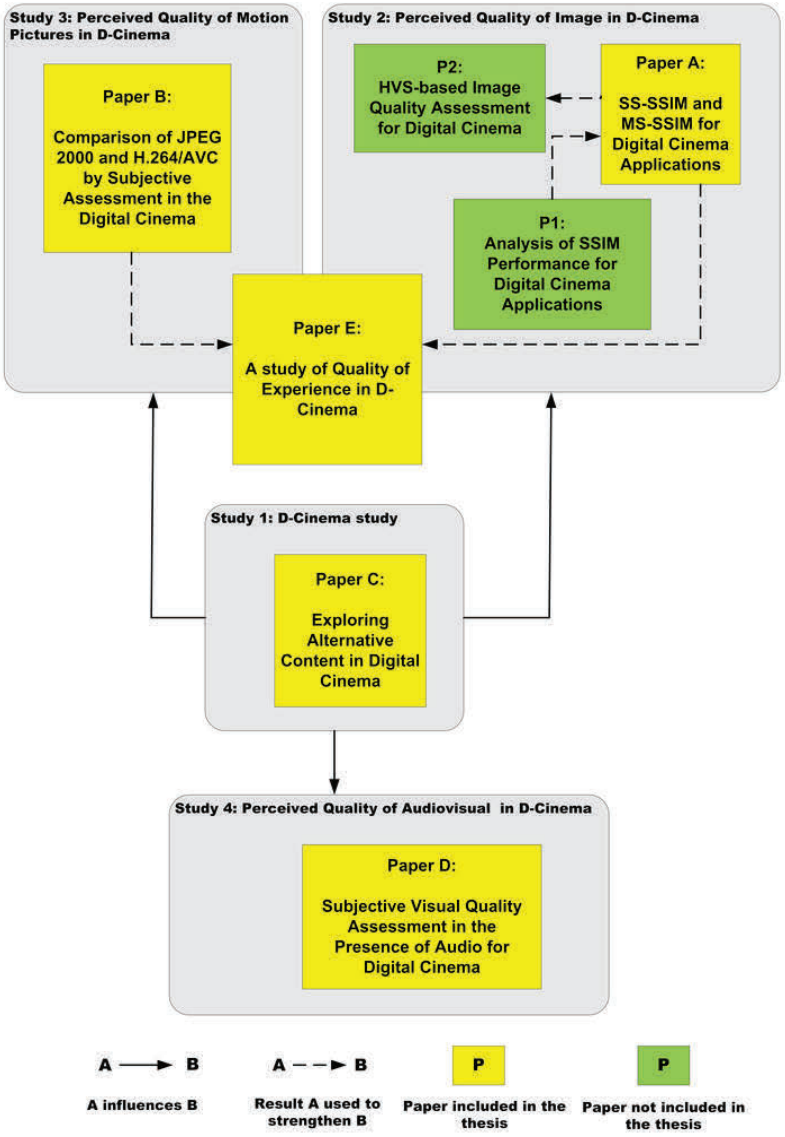


Figure 2: Connection of the written papers.

- **PAPER A**: *SS-SSIM and MS-SSIM for Digital Cinema Applications* [11]

This paper is based on RQ1 and presents our finding of RQ1. The goal of the research behind this publication is to design SS-SSIM and MS-SSIM metrics with input parameters that take into account the Digital Cinema source material characteristics and

viewing conditions. These metrics will then be utilized to measure the perceived quality of high quality digital imagery. To validate and to confirm the results, these will be compared with the PSNR metric and with a subjective evaluation/assessment carried out by human participants in a DCI specified Digital Cinema. The subjective evaluation is performed to find the correlation of the proposed metrics with how humans perceive the quality. I was the leading author of this paper, and I performed all experiments during the data gathering and was responsible for performing the analysis of the data.

The research questions covered in this paper are as follow:

RQ1.1: Protocol of subjective image quality assessment in the Digital Cinema.
RQ1.2: Parameterization of Multi Scale Structural Similarity objective metric.
RQ1.3: Performance assessment of the objective metrics.

- **PAPER B***: Comparison of JPEG 2000 and H.264/AVC by Subjective Assessment in the Digital Cinema* [12]

This paper is based on RQ2. The goal of the research behind this publication is to study the compression technologies by subjective quality assessment. Two video coding schemes with variable bit rates — JPEG 2000 and H.264/AVC — were compared in terms of perceived quality performance in a Digital Cinema environment. Consequently, the protocol to conduct a subjective motion pictures quality assessment in the Digital Cinema must be designed as well. I performed all experiments during the data gathering and was responsible for performing the analysis of the data, and I was also the leading author of this paper.

The research questions covered in this paper are as follow:

RQ2.1: Protocol of subjective motion pictures quality assessment in the Digital Cinema.
RQ2.2: Assessment of the compression algorithms based on collected subjective data

- **PAPER C***: Exploring Alternative Content in Digital Cinema* [13]

This paper is supporting the selected methods in the experiments which are conducted in Paper A, Paper B, and Paper C. The paper puts emphasize on the alternative content screening in Digital Cinema and presents the discussion about Digital Cinema business related to experimentations outside feature films. Due to our interest in screening beyond traditional feature films, the perceptual experiments conducted in Paper A and Paper B are tailored for these types of presentation. I was the leading author of this paper.

The research question covered in this paper is as follow:

RQ3: The importance of human factors represented by QoE in developing alternative content.

- **PAPER D**: *Subjective Visual Quality Assessment in the Presence of Audio for Digital Cinema* [14]

This paper is based on RQ3. The goal of the research behind this publication is to investigate whether the presence of audio with different quality levels can influence the outcome of subjective visual quality assessment in a Digital Cinema setting. We conducted subjective visual quality assessment of AV presentation for D-Cinema and

used the collected data to analyse the influence of audio to visual perceived quality. I performed all experiments during the data gathering and was responsible for performing the analysis of the data. I was the leading author of this paper.

The research questions covered in this paper are as follow:

RQ4.1: Protocol of the perceptual experiments.

RQ4.2: The multimodal effect on the visual perceived quality for Digital Cinema presentation.

- **PAPER E***: A Study of Quality of Experience in D-Cinema* [15]

This paper is based on RQ1 and RQ2. The goal of this publication is to analyze in more detail the collected subjective data obtained from the subjective visual quality assessments mentioned in Paper A and Paper B. The publication also present arguments on the importance of carefully designed subjective quality assessment. I was responsible for performing the analysis of the data and the leading author of this paper.

The research questions covered in this paper are as follow:

RQ5.1: Factors that affect the subjective visual quality assessment result in Digital Cinema.

2 Background

This section gives a short background for the work presented in this thesis. First, Digital Cinema is presented in Section 2.1 Then, subjective quality assessment is presented in Section 2.2. Section 2.3 gives a short overview of perceptual-based quality objective metrics.

2.1 Digital Cinema

The motion picture industry is one of many in the media sector consisting of mature players which have an entertainment focus in common. Currently, both broadcasting and mobile media are digital services. On the other hand, the motion picture industry is now in the process of forming standards for digitization of its complete value chain. These specifications and standards are the basis for a large scale implementation of Digital Cinema as the latest and final analogue media to go digital.

The typical, basic complete chain of digital broadcasting and mobile media is illustrated in Figure 3. The data mentioned in Figure 3 includes image, video, and audio information. Historically, the movie theatre experience has always exceeded what could be achieved by home entertainment systems. Technical improvements in the broadcasting historically influence the motion pictures industry. When the National Television System Committee (NTSC) television became widely adopted in the 1950s, it was greatly feared that this would affect the cinema negatively. The same concerns resurfaced again with the advent of colour television in the 1960s and again with advances in audio technology in the 1980s and 90s. However, the cinema was never reaching its end. Reinvention of cinema technology happened instead. There were underlying trends that the technical improvements in the broadcasting also affect the advances of cinema technology. In the recent years, the same trend happened again with the popularizing of High-definition Television (HDTV). High-definition (HD) broadcast and corresponding receiving sets have been widely available worldwide and are now mainstream for the past few years. This shows the successful rollout of HDTV. Accordingly, the advent of HDTV and technical improvements in home theatre equipments stimulates the motion picture industry to think further ahead into the future.

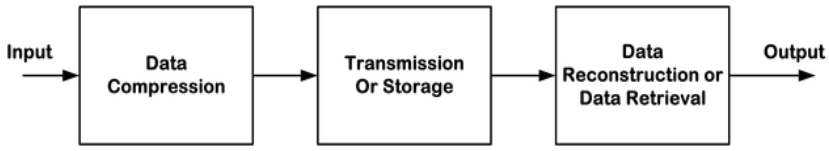


Figure 3: Data processing for transmission and storage.

Following the legacy of television and video cassettes, the cinema makes the transition from analogue to digital. Using film in motion picture industry is a robust, standardised, century-old technology, and replacing it is a complex process. With the transition,

cinema professionals, distributers, exhibitors, and cinemagoers expect a quality level and efficiency that surpass what currently exists.

2.1.1 Digital Cinema System

The digitization of the complete chain is specified by the Digital Cinema Initiative (DCI) [16] and is currently under standardization by Society of Motion Picture and Television Engineers (SMPTE). DCI was created in March 2002, as a joint venture of seven major Hollywood studios: Disney, Fox, MGM, Paramount, Sony Pictures Entertainment, Universal, and Warner Bros, and its primary purpose was to establish a voluntary specification for an open architecture for digital cinema that would ensure a high level of technical performance, reliability and quality control. DCI would also facilitate the development of business strategies to help spur deployment of digital cinema systems in movie theatres.

Figure 4 illustrates the general workflow of the digital process for digital cinema. The digital cinema system is built upon data stored in files. These files are organized around the image frames. The file is the most basic component of the system. Mastering is the stage before distribution which is represented in transport stage, and the result of mastering stage is a concept called Digital Cinema Distribution Master (DCDM). DCDM consists of image structure, audio structure, and subtitle structure. Once DCDM is compressed, encrypted, and package for distribution, it is considered to be the Digital Cinema Package (DCP). This term is used to distinguish the package from the raw collection of files known as the DCDM. Transport stage is the stage where DCP is distributed via Network, Satellite, or Physical Media. Then the exhibitor or theatre stores the obtained DCP file in the digital cinema server, which is generally part of the 2K or 4K digital cinema projector equipments; this is represented by storage stage. The projection stage includes the decrypt, extraction, and decompressed of image structure, audio structure, and subtitle structure from the DCP before screening the complete structures to the cinemagoers [17].

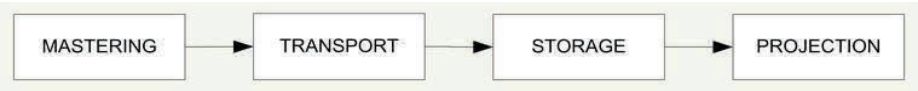


Figure 4: Digital cinema system workflow.

Figure 5 shows the elements of digital cinema system [18] that clarify further the workflow stated earlier. Content creation and then post production are processes to create DCDM (the output of post production is DCDM [17]). Acquisition, capturing the real world through camera and microphone, is part of the content creation process. The resulting content can be digital origination or need to be digitized through A-D transfer. These include the stage at which imagery must be brought into the appropriate digital environment from its original state. This stage, which is also called digital ingest, illustrated in the Figure 6. Most often today, theatrically distributed movies originate on film, but standard or high-definition video is also sometimes used; in the case of animation or visual effects films, digitally originated files are the source [18].

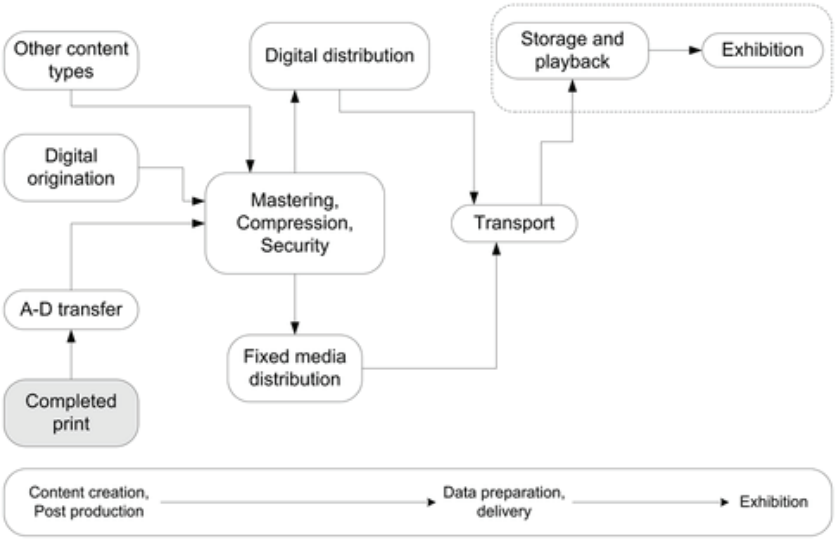


Figure 5: Digital cinema system elements [18].

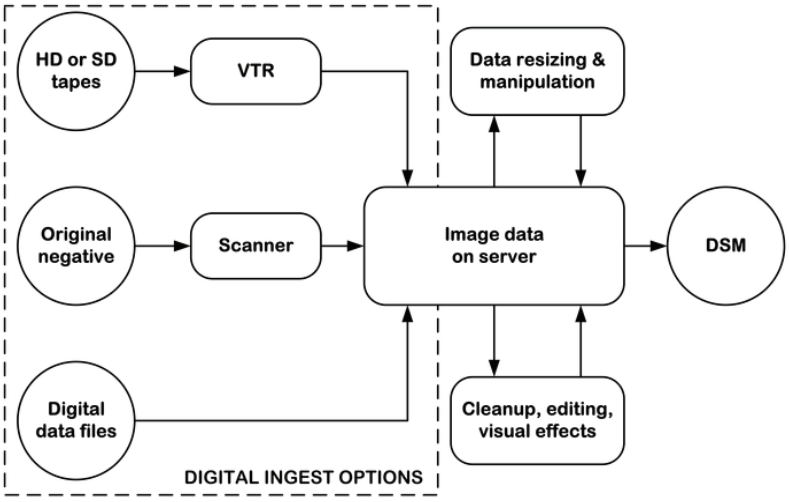


Figure 6: Digital ingest options and the DSM [18].

Post production traditionally covers the process of preparing, editing, and finishing the picture and sound; creating Digital Source Master (DSM) is part of this process. DSM can be used to convert into DCDM, and it also be used to convert to a film duplication master, a home video master, and/or a master for archival purposes. The content could come from a wide range of sources with a wide range of technical levels [17]. Figure 6 shows the processes of creating DSM.

When capturing the source material of content, one relevant factor to consider is the data size of the content. From a quality standpoint, the ideal solution is to capture the maximum that can be mathematically described. However, this also means data that bandwidth and storage are strained beyond realistic and practical limits. The practical solution remains in the required end result.

The more pixels in the picture, the finer the detail will be. This becomes particularly significant when the presentation is for large screens, particularly those from 3 to 24 meters in width. Discriminating the difference among various spatial resolutions on a CRT display can require some very close viewing; when the pixels are spread onto the large screen, the difference is magnified without having to stand so close against the screen. Most of the high-definition television production and post production equipment in place today supports resolutions up to 1920 pixels horizontally by 1080 vertically (1080p). Current Digital Cinema projectors are capable of displaying up to 2048 pixels horizontally (2K). The leap to presentation of 4096 pixels horizontally (4K) is a still larger barrier: not only is there no standardized method for recording and displaying such images, but even custom systems created to handle such data strain today's networks, disk speeds, and disk array sizes [18]. The relative comparison of spatial resolution format is illustrated in Figure 7; this figure shows approximately 1/8 of actual pixel dimensions. Trade-offs of speed and flexibility in the creative production process will often be favoured at the expense of maximum resolution.

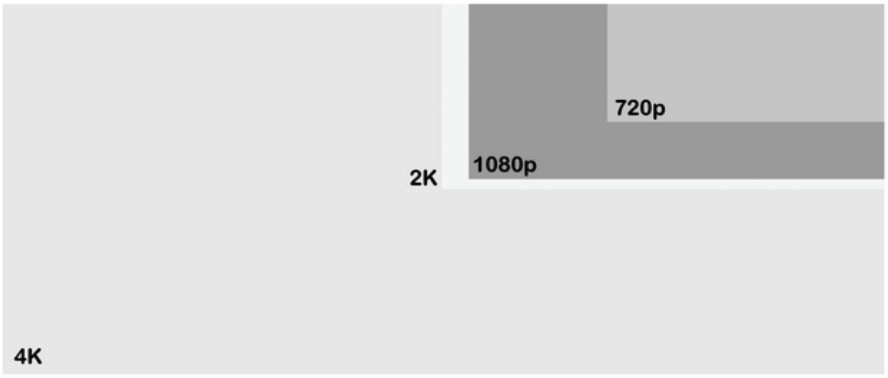


Figure 7: Comparison of spatial resolution.

Four different spatial resolutions shown in Figure 7 are common formats in digital cinema. The relative pixel comparison shown there is at 2.39:1 aspect ratio, and it is illustrated that 4K contains a huge amount of data compare to HD resolution. 4K projection contains 8.847.360 pixels for each frame, while HD projection contains 2.073.600 pixels. 720p format which is also called Standard Definition (SD) contains (only) 345.600 pixels. Nevertheless, the formats used in entertainment industry for the screening of a feature film—film production made for initial distribution in theatres—are 2K and 4K; these formats are the ones that are recommended by DCI [17]. The most common practice today is a workflow in which creative expression and image manipulation during post production including cleanup, editing, and visual effects is done at 2K from files that may have been scanned at 2K, 4K, or even 6K, and then

down-converted to 2K resolution. Computer processing power, digital storage, and network capacity need to undergo significant improvement in capability and cost effectiveness in order for the movie industry to move in 4K workflow direction [19].

Figure 8 shows the mastering processes (the creation of DCDM from DSM) and the distribution process (represented by DCP creation from DCDM) [18]. The digital cinema system uses a store-and-forward method for distribution. This allows the files to be managed, processed and transported in non-real time. After being transported to the theatre, the files are stored on file server until playback. However, during playback and projection, the digital cinema content plays out in real time. A set of DCDM files (image, audio, subtitles, etc.) contains all of the content required to provide a digital cinema feature film screening. The DCDM provides two functions, an interchange file format, and a playback format that is directly sent from the Media Block to the projector (this is referred to as DCDM*). Media Block and along with Storage are components of the theatre playback system. The Media Block is the hardware device that converts the packaged content into the streaming data that ultimately turns into the images and sound in the theatre, and whereas Storage is the file server that holds the packaged content for eventual playback. The DCDM requirements for image specified by DCI are as follows: DCDM image file format is required to be an MXF-conformant file, DCDM audio file format is required to be based on Broadcast Wave, DCDM image structure is required to support a frame rate of 24.000 Hz and a frame rate of 48.000 Hz for 2K image content only, and color encoding of DCDM is 12 bits X'Y'Z'. Furthermore, the audio requirements specified by DCI are as follow: the bit depth is 24 bits per sample, the sample rate is 48.000 or 96.000 kHz and DCP supports a channel count of 16 full-bandwidth channels [17].

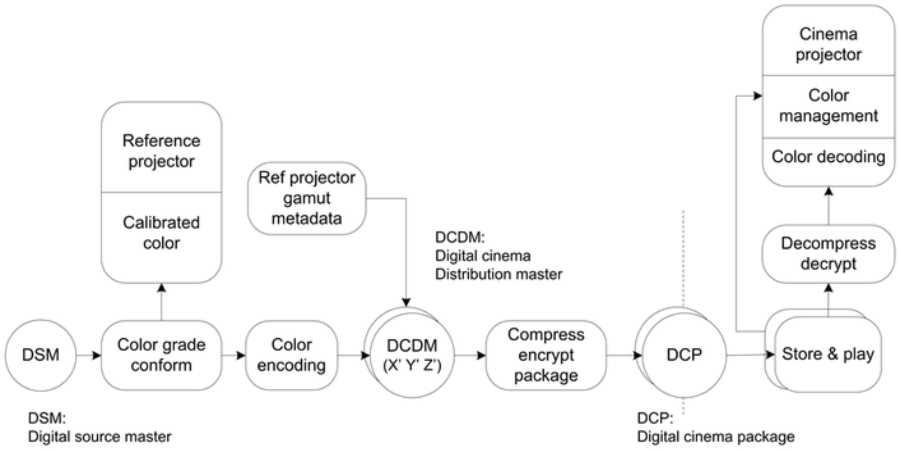


Figure 8: Digital cinema mastering and distribution process [18].

Compression for Digital Cinema uses data reduction techniques to decrease the size of the data for economical, practical delivery and storage. The 4K DCI-specified frame with an aspect ratio 1.85:1 contains 8.631.360 pixels per frame. A total bit per frame in DCDM almost reaches 40 megabytes. Consequently, 2-hour movie at 24 fps is represented for a total of nearly 7 terabytes. Such storage or transmission is physically

possible but impractical [20]. It is important to note that compression is typically used to ensure meeting transmission bandwidth or media storage limitations. This results in quality being dependent on scene content and delivered bit rate. Digital cinema image compression is much less dependent upon bandwidth or storage requirements, thus making bit rate dependent on desired visual quality rather than reverse. The compression technology chosen by DCI is JPEG 2000 [17].

Early experimental deployments have used a number of techniques, mainly proprietary. Examples include a variable bloc-sized DCT-based system from QUALCOMM and wavelet-based system from QuVis. This latter system demonstrated in early 2004 playout of a 2K presentation from a 4K compressed file. One popular standardised compression technology is MPEG-2 system, which is widely used in television. Some early experimental Digital Cinema systems were based on proprietary extension of MPEG-2. MPEG with the experts group of the ITU formed the Joint Video Team (JVT), and this team developed a new coded known as H.264 or MPEG-4 Part 10, or the MPEG Advanced Video Codec (AVC), which offers about the twice the coding efficiency of MPEG-2. The other most well-known compression standards have been developed within the Joint Technical Committee (JTC) of the International Organziation for Standardization (ISO) and the International Electrotechnical Commission (IEC). One working group within the JTC—Joint Photographic Experts Group (JPEG)—has developed standards for the compression of static images. The original JPEG was a DCT-based system designed for static images. This standard was subsequently extended in a number of proprietary systems to provide coding for motion images. JPEG 2000 also started life as compression for static images, using wavelet technology, but this time the committee also standardized the extensions necessary for motion imaging. Motion JPEG 2000 does not use temporal compression; each frame is wavelet-compressed individually as illustrated in Figure 9. However, the tools in the Motion JPEG 2000 extensions will not be used since the DCI has chosen constant-quality coding [20]

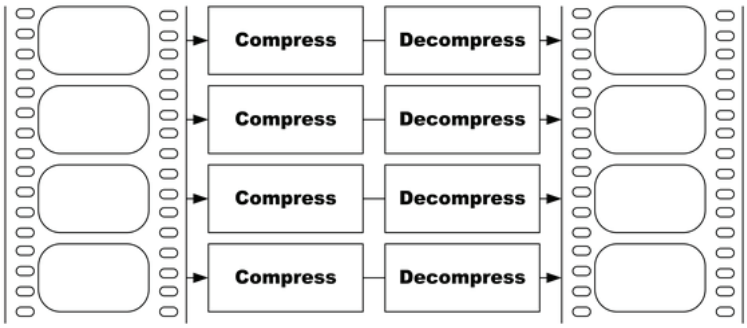


Figure 9: Intra-frame compression [20].

Digital cinema transformed the traditional methods of distributing filmed entertainment that have been employed for decades. With the advent of digital cinema, the fundamentals of distribution to theatres may introduce a profound new paradigm

shift in filmed entertainment. Distribution of digital cinema feature files can be accomplished via three main methods: through the use of optical media (typically DVD), digital storage media (tape or HDD technology), and digital distribution (both via satellite or terrestrial). Even though using optical media or digital storage media still requires physical distribution as the mechanism to transport the feature from distribute to exhibitor, those two methods are today most widely used to distribute digitally prepared features to exhibitors. Digital distribution is widely viewed as the logical platform of the future to support Digital Cinema from a mass market perspective. The use of both unicast and multicast systems via satellite, terrestrial broadband, or a combination of both, which have been used for decades for television, provides an ideal platform for digital cinema distribution. There are several options currently available and more are anticipated as bandwidth access, compression improvements, and intelligent switching networks make moving large files more reliable [21].

Projection is part of the presentation system of digital cinema. Figure 10 shows the overview of the digital form presentation in digital cinema [21]. Local ingest (loading the data manually) is still required to upload the feature onto the the display's server system. The presentation system in the Digital Cinema system includes both projector and media block. Media block is the term coined to avoid confusion of the concept of server among engineers in digital cinema community. (To a broadcaster, a server outputs a synchronous stream of content. To an information technologist, the server outputs either asynchronous or isochronous data.) The media block functional diagram for Digital Cinema is illustrated in Figure 11 [22].

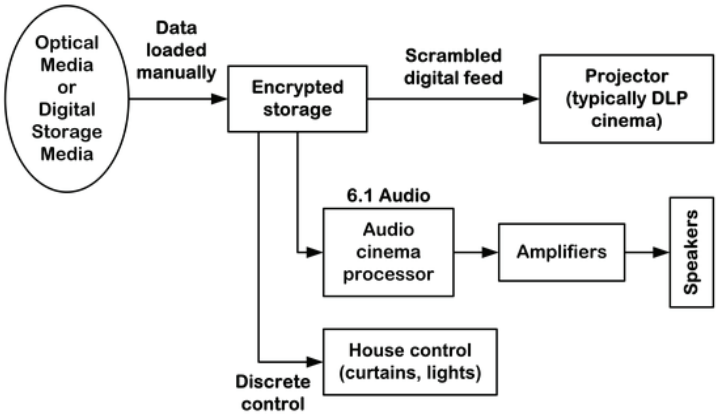


Figure 10: Overview of presenting the digital form [21].

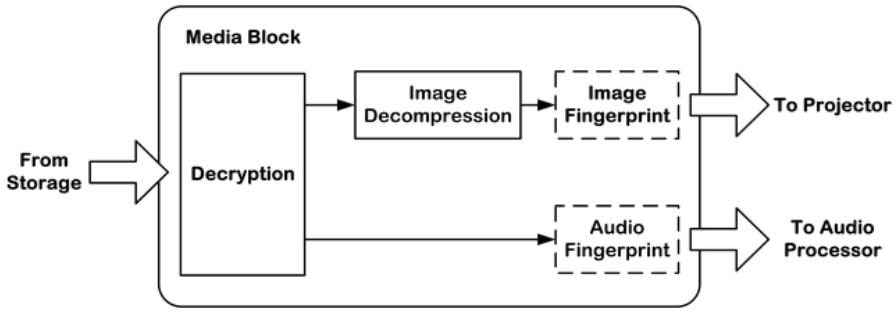


Figure 11: Media block functional diagram for Digital Cinema [22].

The media block provides essential signal processing functions of the system. Signal processing that would be performed in the media block for both image and audio might include decryption, decompression, and fingerprinting. The input to the conceptual media block is independent of transmission type, accepting synchronous or isochronous data, or possibly files. The output of a media block characteristically is a synchronous stream [22].

2.1.2 Relevant Characteristics of the Digital Cinema Projector and Theatres

The projector is one of the key elements in presenting content to the screen, and its performance depends on the effect of the viewing environment. Consequently, viewing environment and projector influences how cinemagoers perceive the content. There is no standard design for a cinema auditorium (theatre). Newer cinemas are designed with the projection booth behind the back wall of auditorium. The distance from the projector to the screen is approximately twice the width of the projection screen. The seating location of the cinemagoers is critical to how much resolution is required to satisfy them. The closer they are, the more resolution they will require [23].

For Digital Cinema projection, luminance is the measure of how bright the screen is. It is important to note that the Human Visual System (HVS) see color relationships differently depending on the brightness of the image; perception of color changes with varying image brightness. Accordingly, the screen brightness, which is the image brightness or luminance as seen by the cinemagoers, is very important for cinema. An image displayed at 6 ftL will look flat and desaturated compared to the same image at 12 ftL [23]. (Screen brightness is measured in candelas per square meter (cd/m^2) in the SI system, or foot Lamberts (ftL) in the American measurement system. The ftL is used in the motion picture industry for measuring the luminance or brightness of images on a projection screen. To convert one to another: 1 ftL = 3.426 cd/m^2.) SMPTE recommended, in SMPTE 196M, a screen luminance of 16 ftL (open gate, with no film in the 35 mm projector) in the centre of the screeen for commercial movie theatres [24]. Current practice in Digital Cinema uses 12 ftL for peak projected white, which provides

an approximate visual match to a film print running through a 35 mm projector set at 16 ftL open gate [23].

Illuminance is the light that comes from a light source that is used to illuminate an object, and for digital cinema projection, illuminance is the measure of how much light is coming form the projector and falling on the screen. The screen luminance, or brightness, is determined by the amount of light falling on the screen (illuminance) and the reflectivity of the screen. Illuminance influenced by factors associated with lamp source (lamp age, lamp type, power), how the projector is set up to correlate the aspect ratio of the screen, and other light losses. Projector light output is measure in lumens. Lumens are determined by integrating the luminous flux, or light coming out of the projections lens (measured in lux) over the total illuminated area (1 lux = 1 lumen/meter^2). American National Standard Institute (ANSI) defines a specification to calculate the useful lumens output of the projector by measuring 9 areas of the screen and integrating these measurements over the screen area which is illustrated in Figure 12 [23].

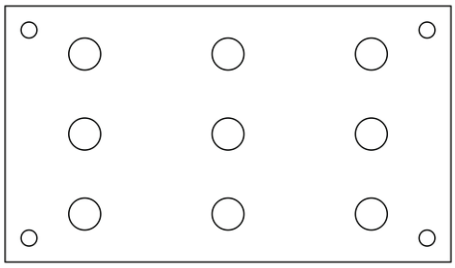


Figure 12: Measurement location for determining lumens from a projector [23].

The screen reflectivity is in theatrical situation called screen gain. Unity gain (gain=1) is compared to a Lambertian reflector, as illustrated in Figure 13. A Lambertian reflector reflects incident light equally in all directions. A higher gain than unity gain will reflect the light preferentially on axis, giving a higher on-axis reflectivity than a Lambertian reflector. This is at the cost of reflecting less light off axis. A high-gain screen will provide less brightness to observers who are off axis than to those on axis, causing non-uniform image brightness throughout the auditorium. In order to maximize the reflection from the screen, while maintaining coverage over a wide angle, the cinema theatre also must take into account the seating array factor [23].

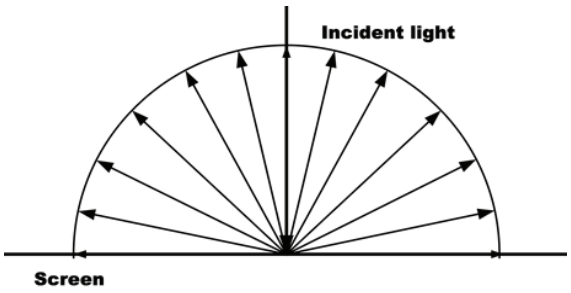


Figure 13: Lambertian reflector [23].

Conditions within the auditorium affect the look of a projected image on the screen. The theatre contributes stray lighting from aisle and exit lights that illuminate the screen. Light is also scattered back to the screen from the walls and seats. These stray sources of screen illumination compete with the projected image, resulting in degraded image. Because cinemagoers are the ones who consumed the presentation, the most important is the image quality that the cinemagoers see. Consequently, brightness is ideally measured from several positions in the theatre and averaged. In designing a projection system, the on-axis screen brightness may be calculated approximately by the following [23]:

$$Screen\_Brighness = \frac{(Screen\_gain \times Lumens)}{Screen\_area} \tag{1}$$

Contrast has an effect on the presentation seen by the cinemagoers. Consequently, determining contrast performance of the projector influence the visual perception of the cinemagoers. A low-contrast system in effect adds some light to the dark areas of the image, making the image appear milky and one-dimensional, and the detail is diminished. There are four ways to specify contrast, resulting from four methods to measure contrast: Off-to-On, ANSI, Local Area, and In Situ. Off-to-On contrast states the projectors' ability to achieve absolute black; it is measured by comparing the maximum brightness white field with minimum brightness black field. ANSI contrast compares the contrast between black and white squares in a 16-square checkerboard pattern; this useful for determining the optical system quality in terms of flare. A system with low ANSI contrast will scatter more light from the white squares to the dark squares. Local area contrast expresses the ability to maintain adequate contrast between small objects in the projected image; this is important for maintaining detail in an image. In situ contrast is determined by measuring the actual contrast achieved in a theatre; this accounts for the back scatter in the auditorium and source of stray light [23].

Digital projectors build each frame in a memory buffer and then display it for the entire frame time. There is no black time between frames. Consequently, an object in motion will appear softer, and juddering will not exist. Juddering is a visual artifact that exists in 35 mm film projection technology, in which an object in motion will appear to

judder back and forth as it moves across the screen. When 35mm film is projected at the standard speed of 24 fps using a two-bladed shutter, the projector flashes each frame twice within $1/24^{th}$ second before the next frame is moved into position, with equal time given to the black screen produced when the shutter closes. This black time between frames is the factor that causes juddering [23].

There are a number of candidate projection technologies for digital cinema. Relevant projection technologies in this thesis are Digital Light Processing (DLP) Cinema, which is developed by Texas Instruments [25], and Liquid Crystal on Silicon (LCOS), specifically SXRD, which is a liquid-crystal-based modulator technology developed by SONY [26]. DLP cinema is illustrated in Figure 14 whilst SXRD is illustrated in Figure 15. Both technologies utilize reflective modulator to manipulate the light to create color on the screen. More detail of both technologies is shown in Table 1.

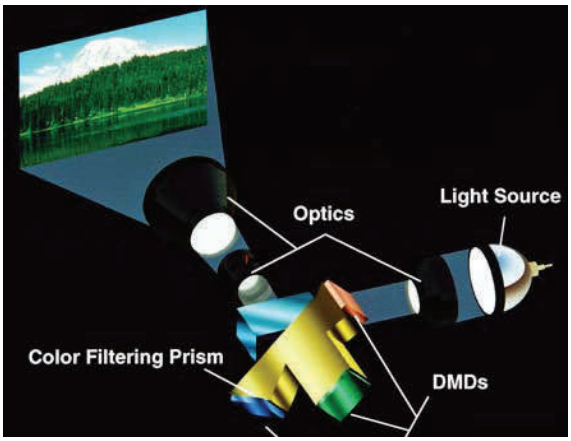


Figure 14: DLP Cinema from Texas Instruments [25].

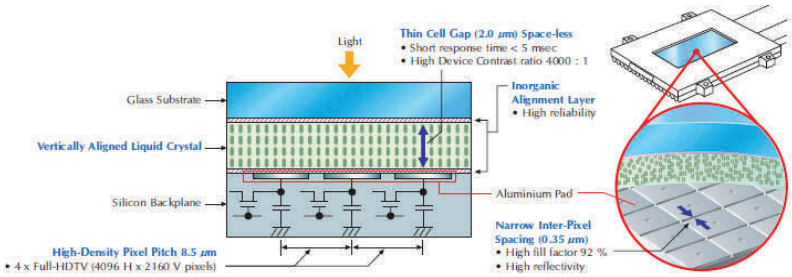


Figure 15: SXRD from Sony [26].

Table 1: DLP and LCOS Digital Cinema Technologies

Technology	Description	Advantages [23]
DLP Cinema	The modulation is done using microscopically small mirrors (DMDs) to switch the light on and off, to control the absolute amount of light that arrives at the screen.	It is very stable and uniform across the screen. Light levels do not depend on temperature or bias level differences on the modulator.
SXRD (LCOS)	The modulation is done by activating the liquid crystal (applying an electric field to the crystal gap) to control the absolute amount of light that arrives at the screen.	It is relatively cheaper to manufacture. It is inexpensive to scale up to accommodate large arrays.

There are two types presentation format in Digital Cinema; they are flat and CinemaScope (Scope). Flat is sometimes referred to as academy wide screen and has an aspect ration of 1.851:1; on the other hand, Scope has an aspect ratio of 2.39:1 [18, 23]. Projection technology can respond in three fundamentally different ways to achieve the correct aspect ratio: electronic masking, anamorphic lens, or electronic scaling in the projector. Electronic masking creates the correct aspect ratio of the image by masking the area of the modulator that is outside the desired aspect. This technique is also used to correct trapezoidal distortion caused by projecting down toward the screen. There is one unwanted effect of Scope projection using electronic masking technology; Scope reduces the usable pixels by 21%. The other alternative is using anamorphic lens which stretches the image horizontally to make it Scope by elongating the pixels as projected on the screen. Digital projectors have also the ability to scale the image electronically to make the source material fill the native imager array [23].

2.1.3 Nova Kinosenter

Nova Kinosenter is a DCI specified cinema, which is located in Trondheim, Norway and is operated by Trondheim Kino AS [27]. The perceptual experiments which will be described in the later sections were conducted in this cinema, specifically at Ullman-salen, the auditorium 1 of Nova Kinosenter.

Figure 16: Nova kinosenter, a cinema in Trondheim, Norway [27].

There are two digital cinema projectors that are relevant and utilized in the research covered in this thesis. They are: 4K Digital Cinema Projectors Sony SRX-R210 [26] (it is illustrated in Figure 17) and 2K Digital Cinema Projector Christie CP2230 [28] (it is illustrated in Figure 18). Their specifications are shown on Table 2.

Figure 17: 4K Digital Cinema projector SRX-R210 [26].

Figure 18: 2K Digital Cinema Projector CP2230 [28].

Table 2: Digital Cinema projectors specifications

	SRX-R210 [26]	CP2230 [28]
Brightness	14 ftL on 17 or 14 M wide screen with screen gain of 1.8 using a single Xenon lamp.	30.000 lumens for screen with up to 30.48 meters wide measured at screen centre.
Contrast Ratio	2000:1 (measured from a screen offering a gain of 1.0)	2100:1 Full Frame On/OFF 450:1 ANSI
Resolution	4K (4096 x 2160)	2K (2048 x 1080)

The projectors have variety of interfaces, and they support a wide variety of signal formats. SRX-210 supports images using the 12-bit X'Y'Z' signals that are stipulated in the DCI specification, and it also supports for playback from other alternative sources, such as 4:2:2 YCbCr and 4:2:0 YCbCr signal formats. The interfaces types are as follow (more details can be found in Appendix A) [26]:

- Two channels of SRLV which are used for connection to the Media Block (for 4K exhibition).
- A dual-link HD/DC-SDI input that accepts any of the following signals: SMPTE 372M dual-link HD-SDI (4:4:4), SMPTE 292M HD-SDI (4:2:2), or 12-bit (X'Y'Z' 4:4:4) signals (for 2K projection or HD projection).
- A DVI interface that accepts DVI signals for up to 2048 x 1080 at 60 Hz.

CP2230 supports two inputs of SMPTE 292M bit-serial standard and two DVI inputs (VESA DVI-D standard). All formats, which are supported by CP2230, are: at 10 bit 4:2:2 Y CbCr or lower, and DCI formats (SMPTE 428-9) at 12 bit 4:4:4 XYZ (more supported input formats can be found in Appendix A) [28]. The variety of CP2230 interfaces are illustrated in Figure 19.

Figure 19: Variety of CP2230 interfaces [28].

Research scope

Visual distortions can happen due to digital cinema projector. The factors that cause them can be classified as internal—inherently due to limitations within the projection technology—or external—due to improper use and installation of the digital cinema projector itself. To prevent such internal factors acting up, the digital projectors often offer internal processing mechanism to the image. Examples of this are electronic masking to achieve correct presentation format (such as scope format) and to prevent trapezoidal distortion or keystone distortion. Digital projectors have the ability to scale the image electronically to make the source material fill the imager array. This allows the projector to resize to accommodate the anamorphic lens. Trapezoidal distortion to the image is caused by the projection downward from the booth to the screen. This is shown in Figure 20. In 35 mm projector, this can be corrected using a trapezoidal mask in the aperture of the projector. Digital projectors apply this mask electronically; this technique is also applied by SRX-R210, which is illustrated in Figure 21. Alignment allows exhibitors or theatres to set a two screen points as well as four corner points, which provides compensation for both flat and curved screen.

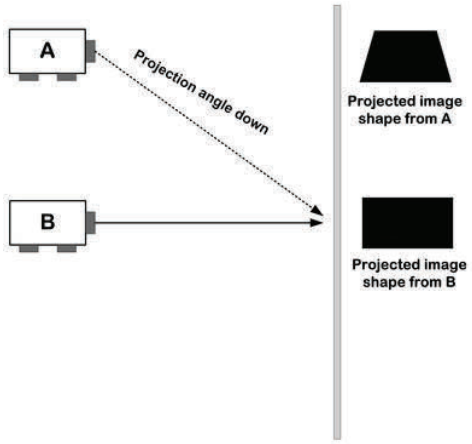


Figure 20: Projection downward from the booth introduces trapezoidal distortion.

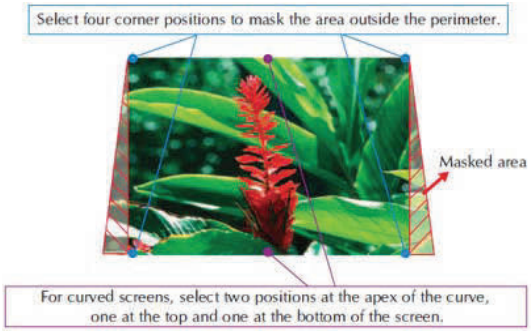


Figure 21: Image masking function to compensate for trapezoidal distortion [26].

Proper set up and calibration of the digital cinema projector installation prevents the visual distortions caused by improper use of the digital cinema projector. Digital projectors are designed to have a form factor similar to that of film projectors. The setup issues are as follow:

- Mechanical and optical. The projector is located behind the port window and optically aligned to the screen. This requires the correct lens, usually a zoom lens. To achieve the correct image size, the magnification of the lens must be designed to accommodate the screen width and throw distance. In 35 mm projectors, this has been accommodated by creating a catalogue of fixed focal lenses that vary over a wide range of magnifications ensuring that a correct lens will be available. Digital Cinema systems are providing zoom lenses to ensure that the image size is correct on the screen. The projector is unlikely to be directly on axis with the screen. This will cause some keystone distortion. Some of this may be removed through offsetting the lens. The balance will be removed by setting electronic masking to make the projected image square. The ideal position of the projector is illustrated in Figure 22.

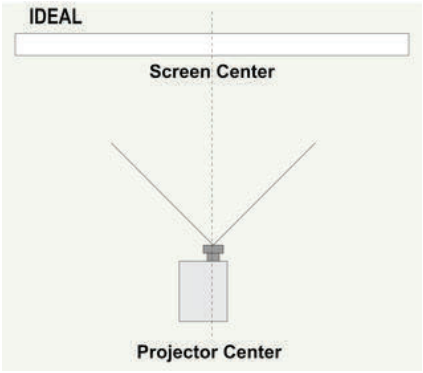


Figure 22: Ideal position of the projector.

- Color. Color calibration is required to ensure that the projector is performing to its required specification. In the case of DLP Cinema technology, the projector's primary color are measured and input back to the projector, which then calculates the required corrections automatically.
- Light level. The lamp should be adjusted to output 48 cd/m^2 on a full white image.

It is important to remember that projection is about the image, not the hardware. Optimal image quality is achieved by optimized interaction of all the factors discussed up to now. Some of these factors are particularly important to the overall image [23].

- Contrast is the most important driver of image quality. The ability to achieve good solid black affects the ability to build punch or impact into a picture
- Gamma, or transfer function, carries the dynamic information in the image.

- Color management ensures that the best color possible is presented to the screen.
- Pixel count, or resolution, ensures that the information the information in the source image can be viewed on the projector.

Auditorium 1 of Novakinosenter is DCI-specified cinema, which is used everyday as a commercial cinema in Trondheim, Norway besides as a laboratory environment for our perceptual experiments. Trondheim Kino AS, the company who run the cinema, has performed installation and regular maintenance for the operation of the cinema, including the equipments and the environment setting to meet the standard of commercial DCI-spesified cinema. Consequently, we consider the factors mentioned earlier that can affect the perception of cinemagoers have been controlled.

The projector should be considered as part of a total visual system, starting with the source material and ending at Human Visual System (HVS). The overall performance is dependent on each contributor of the system. It is possible to identify and isolate most of the factors that contribute to overall picture performance. However, in this thesis, we do not consider the processing algorithms of the digital cinema projector, such as the signal processing used to get rid of the trapezoidal distortion due to proprietary issue of processing algorithm of the digital cinema projectors. We have mentioned this in Section 1.2.

As described in Section 2.1.1, the media block is part of the digital cinema system, specifically the presentation system. Media block is the term coined to avoid confusion of the concept of server among engineers in digital cinema community. The media block functional diagram for Digital Cinema is illustrated in Figure 11. The Novakinosenter, as a DCI-specified cinema in Trondheim, utilizes media block for its feature film screening. One media block that is part of the 4K SRX-210 projector system, is LMT-100. The LMT-100 server handles DCP (Digital Cinema Packages) files that consists of image, audio, and subtitle data files, and that are wrapped into an MXF (Material eXchange Format) file. It can play back the DCP file by using advanced processing to decrypt and decode the image data, and then send it to the projector over a secure multi-pin connection system, and it can decode JPEG 2000 image data in real time for playback, regardless of whether the file was encoded at 2K or 4K resolution. In addition, it transcodes audio DCP files into AES/EBU digital audio signals, and then outputs them to the external audio processors. Up to 16 channels can be output from D-sub 25 pin or BNC connectors. The timing of the audio output can be adjusted for complete synchronization with the image, and any channel can be routed to any output to simplify installation. The LMT-100 server is controlled with the SMS (Screen Management Software). In spite of this, we do not use the media block in our experiment due to our research scope and the rigidity of the media block. We are interested in exploring the alternative content outside feature film specified by DCI. The media block can play back the file in DCP format only, which is restricted for our experiment purpose, in a sense that the media block cannot play back the uncompressed original files or other format such as images files that are not compressed by JPEG 2000. Thus, we develop off-the-shelf customized server as part of the digital cinema presentation system for our experiments by making use of the interface types offered by the projectors (both SRX-R210 and CP2230). The illustration is below.

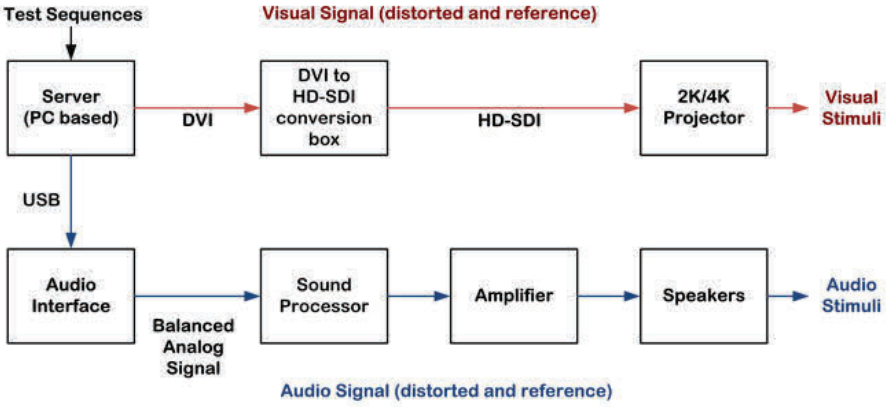


Figure 23: Presentation system being used in the experiments.

As we mentioned in section 1.2, we did not consider the issue of intellectual property of the compression technology used in digital cinema. Issue of intellectual property and licensing of the compression technologies significantly influenced the chosen technology used in the industry. In 2004 Digital Cinema Initiatives (DCI) selected JPEG 2000 as the compression technology of choice for digital cinema [17] because JPEG 2000 offers all the technical attributes desired and is also royalty free. The history of MPEG-4 licensing has left many in Hollywood opposed to adopting any MPEG standards [20]. Our approach regarding compression techniques used in digital cinema presentation is predominantly from academic standpoint. Consequently, we discuss also the compression technologies other than JPEG 2000 and the specifications described by DCI.

2.2 Subjective Quality Assessment

The definitive way of measuring perceptual quality is using human participants in controlled experiments. These methods are often called subjective quality assessments. In the field of subjective quality assessment, there are many different methodologies and rules for designing tests. A number of standards or recommendations exists that cover a wide range of topics from measurement devices through to perceptual evaluation methods for telecommunications systems. Generally, standards are developed when there is a large-scale need to address a problem within a field in an industry. The need is usually driven by several stakeholders when there are advantages from establishing a commonly agreed upon approach addressing the problem. Standards give the benefit of an agreed upon approach, developed by experts in the field from both industry and academia. In terms of perceptual assessment such as subjective quality assessment, this means that a methodology has been developed and verified as being applicable to the domain defined for that standard. Nevertheless, standards and recommendations require considerable time for being developed and also require the consent of all stakeholders involved. Besides, not all issues are standardised due to several factors. To begin with, the process is very costly and time consuming. Additionally, as not all methods are commonly needed within an industry, only the key methods that require an agreement are studied and standardised. As a result, standardised methods are not always representative of the state-of-the-art methods in the field even though there are a number of key standards that define certain aspects of perceptual assessment [8].

Test methods described in ITU have been internationally accepted for conducting subjective quality assessment [29-36]. An overview of recommendations for perceptual evaluation are illustrated in Figure 24, Figure 25, Figure 26, and Figure 27.

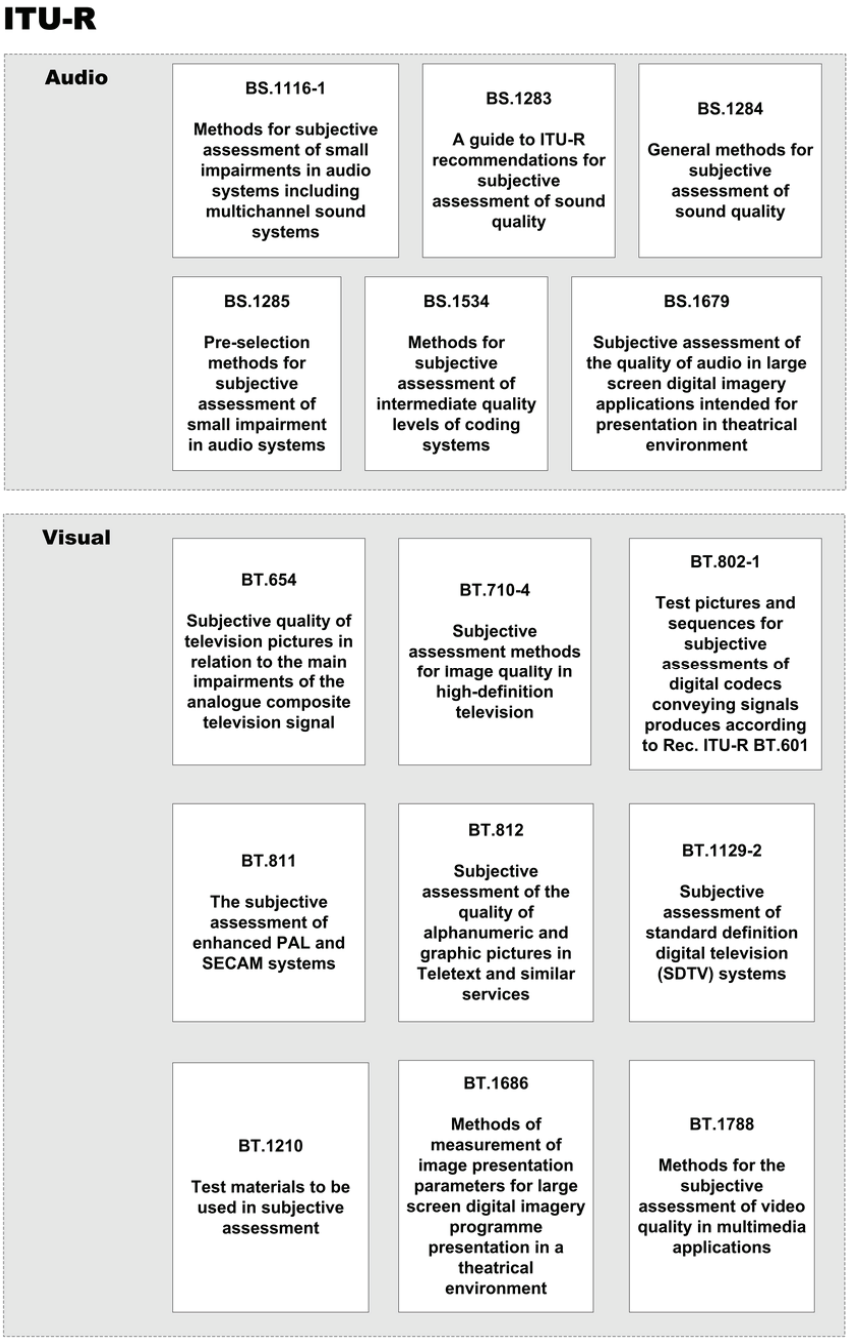


Figure 24: Summary overview of key ITU-R recommendations relating to perceptual audio and visual evaluation [8].

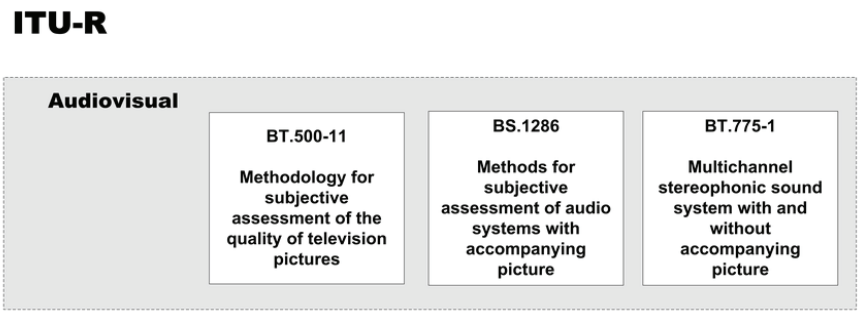


Figure 25. Summary overview of key ITU-R recommendations relating to perceptual audiovisual evaluation [8].

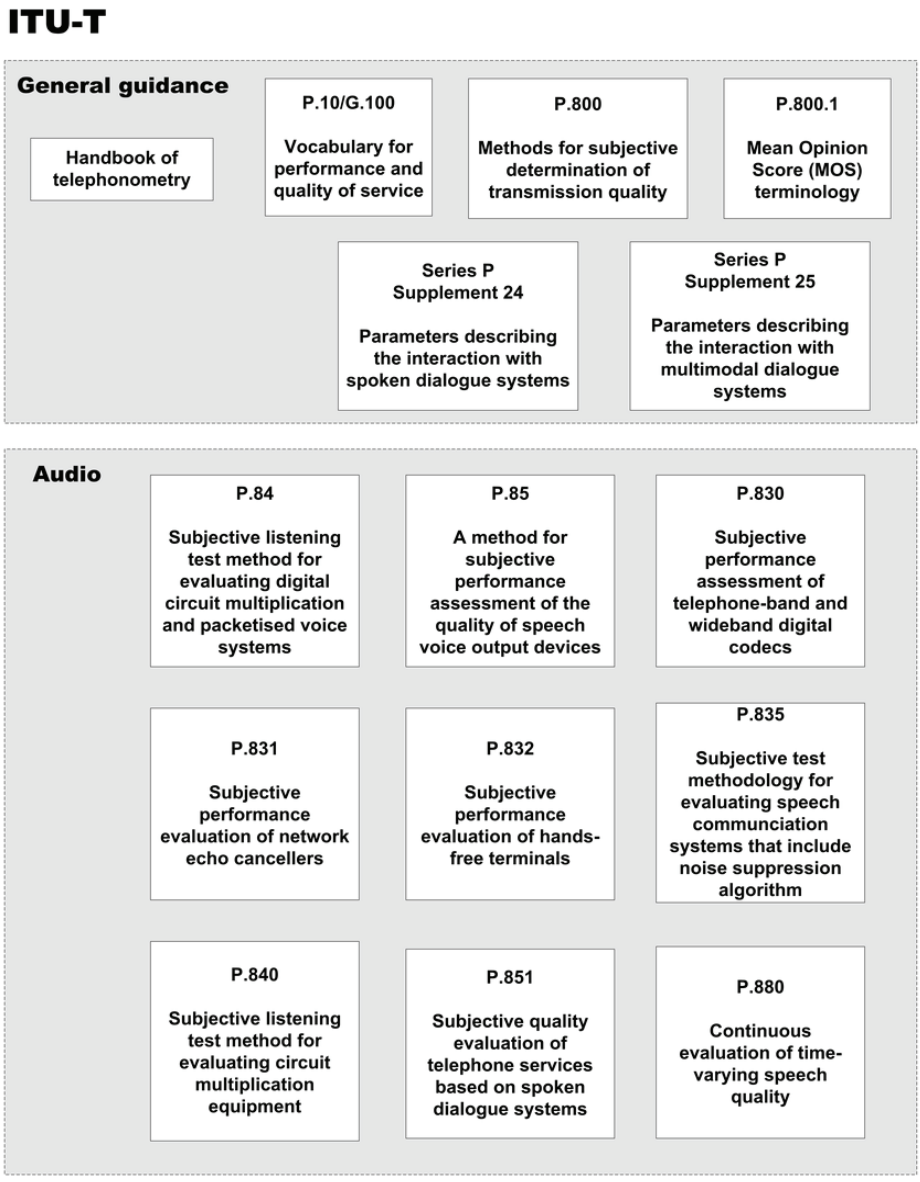


Figure 26. Summary overview of key ITU-T recommendations relating to perceptual evaluation [8].

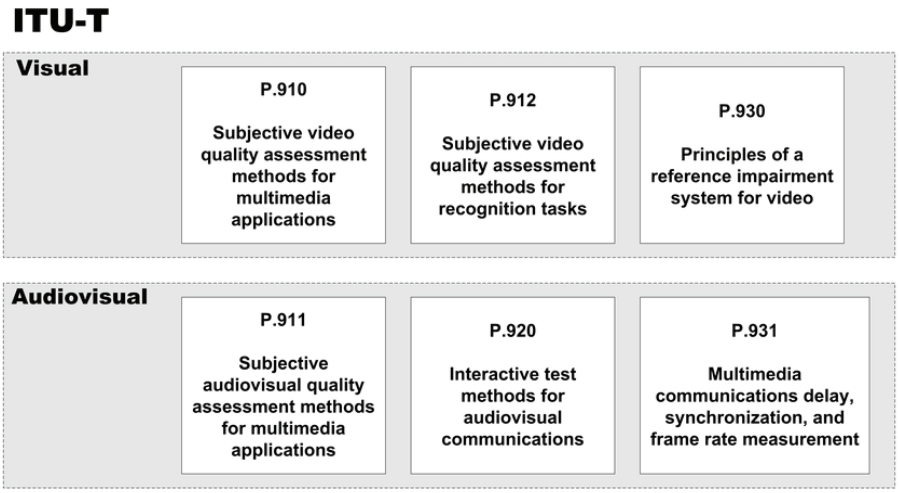


Figure 27. Summary overview of key ITU-T recommendations relating to perceptual evaluation [8].

Relevant recommendations for subjective visual quality assessment are Recommendation BT.500-11 [29] and Recommendation P.910 [35]. Recommendation BT.500 describes extensive details on methodologies for evaluation of television picture quality. The recommendation provides important information regarding influential factors related to the experiment design, such as illuminations levels, screen sizes for different resolutions and aspect ratio displays, viewing distances, and so on.

Recommendation P.910 describes non-interactive subjective assessment methods for evaluating the one-way overall video quality for multimedia applications such as videoconferencing, tele-medical applications, etc. The recommendation illustrates the characteristic of the source sequences, duration of the test sequence, content types, number of sequences, and so on. These methods can be utilized for a number of functions including algorithm selection, ranking of audio-visual system performance and evaluation of the quality level during an audio-visual connection.

Noteworthy issues in the guidelines for visual subjective quality assessment are stimulus (i.e., characteristics of viewing sequence) and types of scale used by participant to rate stimulus' quality.

The test methods can be classified into two categories based on the stimulus used in the experiment; they are double stimulus and single stimulus. In double stimulus method, the experiment's participants are presented with unimpaired reference and impaired stimulus before participants give quality rating of each (impaired) stimulus. The structure of Double Stimulus is illustrated in Figure 28. Figure 28 shows test sequence of A and B. Ar and Br are the sequences in the reference sources format while Ai is the A sequence under test condition i and Bj is the B sequence under test condition j. The participants are instructed to compare the sequence under specific test condition to its reference sequence and judge the quality of the sequence under that test condition.

It is important to note that there is a variation of this double stimulus experiment structure in which the participants do not have any idea during the experiment the order of reference sequence and the impaired/processed sequence before voting, i.e., the reference does not always in the first order.

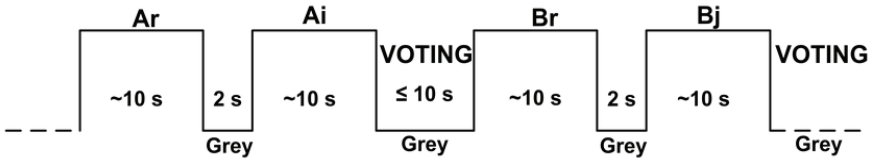


Figure 28: Double Stimulus experiment structure [30].

In single stimulus method, the participants are only presented with a single stimulus before the participants give the quality judgment of that particular stimulus. There are some variations of single stimulus method; they are structure of Single Stimulus (SS) or Absolute Category Rating (ACR) which is illustrated in Figure 29 and structure of Single Stimulus Continuous Quality Evaluation (SSCQE). Figure 29 shows test sequence A, B, and C with their respective test conditions. In this experiment structure reference is hidden among the impaired sequences. In SSCQE structure, set of test sequences is measured continuously, with the participants viewing the material only once. Then during the viewing session in the experiment, participants can also give rating throughout the entire duration. The ratings are recorded by sampling ratings at determined rate (e.g., every 0.5 seconds). This structure was developed on the idea that within the digitally coded video, the impairments may be very short-lived and the quality can fluctuate quite widely.

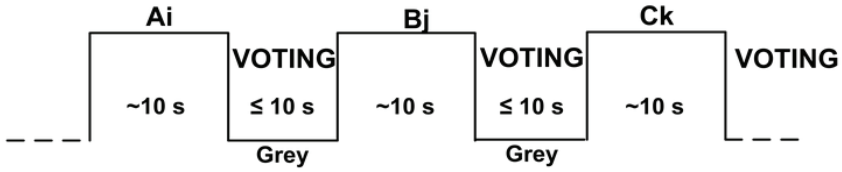


Figure 29: SS/ACR experiment structure [30].

There is another test method that is similar with SSCQE structure but involving reference sequence. This method is called Simultaneous Double Stimulus Continuous Evaluation (SDSCE) experiment structure. In SDSCE, participants view two sequences at the same time: the reference sequence and the sequence under test condition. Participants give the rating by comparing the sequence under particular test condition to the reference at the same time before judging quality the sequence under that test condition. Display format of SDSCE structure is shown in Figure 30.

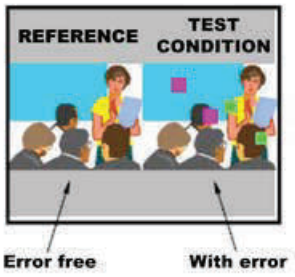


Figure 30: Display format in Simultaneous Double Stimulus Continuous Evaluation (SDSCE) [29].

Relevant recommendations for subjective audio quality assessment are Recommendation BS.1679 [36]. Recommendation BS.1679 offers a summary of the specifications when conducting subjective assessment of audio quality or audio impairment for Large Screen Digital Imagery (LSDI) applications designed for programme presentation in a theatrical environment. The recommendation is derived from ITU-R Recommendation BS.775-1 [33], ITU-R Recommendation BS.1116-1 [37], ITU-R Recommendation BS.1284 [34] and ITU-R Recommendation BS.1286 [38]. Recommendation BS.775-1 illustrates the basic physical loudspeaker configurations to be employed in domestic 5.1 multichannel sound reproduction in the form of 3/2 (3 frontal, 2 surround channels) and 3/4 (3 frontal, 4 surround channels) systems. Recommendation BS.1116-1 is a recommendation for assessing systems that introduce impairments so small as to be undetectable without rigorous control of the experimental conditions and appropriate statistical analysis. However, if the recommendation is applied for systems that introduce relatively large and easily detectable impairments, it leads to excessive use of time and may lead to less reliable results than a simpler test. This recommendation forms the base reference for the other recommendations, which may contain additional special conditions or relaxations of the requirements included in BS.1116-1. Recommendation BS.1284 offers a short guide to general requirements for performing listening tests (subjective assessment of sound quality); it outlines the experimental design including participants' selection, test methods, and statistical analysis. Recommendation BS.1286 gives information on how to conduct testing of audio systems in the presence of accompanying image, including the recommendation for different image sizes, aspect ratios, and image definitions. The use of this recommendation should be used in combination with one of the audio-only recommendations i.e., [34, 37, 39]. Relevant recommendations for subjective audiovisual quality assessment are Recommendation P.911 [30], Recommendation P.911 illustrates non-interactive subjective assessment methods for evaluating the one-way overall audiovisual for multimedia applications such as videoconferencing, tele-medical applications, etc. These methods can be utilized for a number of functions including algorithm selection, ranking of audio-visual system performance and evaluation of the quality level during an audio-visual connection. The recommendation also describes the characteristics of the source sequences, duration of the sequence, content types, number of sequences, etc. In addition it describes indications about the

relation between audio, video and audiovisual quality, as they are derived from results of tests carried out independently in different laboratories.

Our Experiment Design

The design of our perceptual experiments in Digital Cinema is derived from methodologies standardized by ITU, which have been described above. We selected the methodologies recommended by ITU as our foundation because they are well-established, widely-known methodologies in spite of the fact that they were developed based on telecommunication and broadcasting issue, such as television quality issues. Despite this, we will still use the same approach, as the starting point for the design of our own subjective quality assessment in digital cinema, but apply the necessary modifications to adapt to the D-Cinema environment. The main objective of the methodologies of subjective quality assessment recommended by ITU is to collect scores from participants representing the quality level of stimuli experienced by the participants in the experiments. The most common approaches produce a Mean Opinion Score (MOS) to determine the quality by averaging the quality scores. Our experiments are based on Recommendation ITU-R BT.500-11 [29] and ITU-T P.910 [35].

Because our experiments are for digital cinema, we do not literally adopt all the guidelines as there are inherent major differences between applications mentioned in the recommendations and digital cinema, such as in the following issues: viewing conditions, resolution and contrast, the source signals and test materials. A number of adaptations were made in our assessment experiments. Subjective quality assessments were conducted in a DCI-specified digital cinema theatre that is regularly maintained. Thus, we believe the viewing condition, contrast and illuminance conditions of the test environment provide realistic and representative viewing conditions. More detail of the methodologies are described in the papers included in this thesis.

2.3 Perceptual-based Quality Objective Methods

Subjective quality assessment is not practical in an application scenario which requires real time processing because it is complex and time consuming. Hence, we need automated methods that can predict the quality as it would be perceived by a user/human observer. This method is referred to as objective methods or objective perceptual quality metrics. Perceptual metrics may build a bridge between QoE and QoS parameters. The metric outcomes can be connected to human perception by relating them to MOS obtained in subjective experiments.

2.3.1 Visual Quality Metrics

Digital Cinema is an application in which the visual factor is very dominant. Hence, evaluating the perceptual visual quality is very relevant in QoE study for Digital Cinema. Currently, there is no widely used perceptual objective model to assess visual quality. The signal-to-noise ratio (SNR) is still a popular metric to assess quality

objectively. Consider Figure 31, where *f(x,y)* is the input image to a processing system such as compression system. It can also represent a process in which additive white Gaussian noise corrupts the input image. The *g(x,y)* is the output of the system. Error function *e(x,y)* is defined as the difference between the input and the output, i.e., Eq (2); it is used to measure the quality of *g(x,y)*.

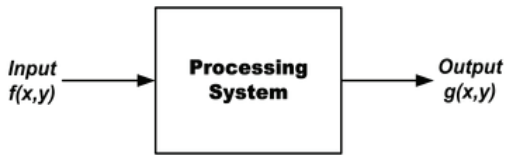


Figure 31: An image processing system.

$$e(x,y) = f(x,y) - g(x,y) \tag{2}$$

The mean square error (MSE) is defines as:

$$MSE = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} e(x,y)^2 \tag{3}$$

where *M* and *N* are the dimensions of the image in the horizontal and vertical directions.

SNR is defines as:

$$SNR = 10 \log_{10} \left[\frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x,y)^2}{MN \cdot MSE} \right], \tag{4}$$

In image and video data compression, another closely related term, *PSNR* (peak signal-to-noise ratio), which is essentially a modified version of *SNR*, is widely used. It is defined as follows.

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \tag{5}$$

The interpretation of the this is that the larger the *SNR* or *PSNR* the better the quality of the processed image, *g(x,y)*; that is, the closer the processed image *g(x,y)* is to the original image *f(x,y)* [40]. However, the HVS does not respond to visual stimuli in a straightforward way. Consequently, *SNR* or *PSNR* does not always provide us with reliable assessment of visual (image and moving images) quality.

The Radiocommunication sector (ITU-R) and the telecommunication sector (ITU-T) of ITU has been cooperating in an effort to find appropriate image and video quality measures suitable for standardisation. A group of experts known as the Video Quality Experts Groups (VQEG) [41] was formed. These experts came from both of ITU sections. VQEG performed some studies on video perceptual metrics which are reported in [42, 43]

Two general approaches have been followed in the design of objective quality metrics; they are psychophysical approach and engineering approach [44]. Metric design using psychophysical approach is basically based on incorporation of various factors of the HVS which are essential for visual perception. This can consist of modelling of frequency selectivity, contrast and orientation sensitivity, spatial and temporal masking effects, and colour perception. HVS is complex; consequently, models and metrics based on HVS can become very complex and computationally expensive too. Yet, they usually correlate very well with human perception and are usable in a wide range of applications. Fundamental work in developing visual metrics using the psychophysical approach has been performed in [45-51].

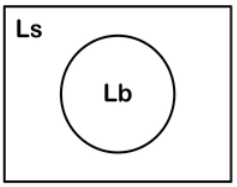
Design using the engineering approach is mainly based on image analysis and feature extraction. The extracted features and artifacts can be of different kinds such as codec parameters, content classifier, and spatial and temporal information. This does not exclude some HVS factors that are also considered in the design. While some developed metrics are derived from simple, numerical measures of single feature such as [52], some metrics can be more complex, which are based on more complex extraction and analysis algorithms, combining various measures in a meaningful way such as [53].

The HVS is extremely complex, and the current knowledge is limited mainly to low-level processes that can be divided into two major parts: the eyes and visual pathways in the brain. The eyes are the parts that capture light and convert it into signals that can be understood by the nervous system, they operate like camera device and follows similar physical principles such as applying optical focusing and a sensor (the retina) for transforming light into the information. Furthermore, the brain is the part that transmitted and processed the information [4]. One HVS characteristic that is often used in the psychophysical approach of metric development are the response of the human vision to the contrast pattern.

The HVS is very adaptive. Brightness of an object, which is also called luminance, is a reflective measurement determined by looking directly at the object. In a theatre, the filmmaker is able to present convincing images of bright sunlight using a brightness level that is less than 1% of the brightness in the actual scene. This done by creating brightness and color relationships on the screen that convinces the cinemagoers that it is bright sun. Tied to image brightness is the brightness distribution. Over time, it has become accepted that the most pleasing cinema images are brightest in the center and have a slight luminance fall off toward the edges. The optimal variance is between 15% and 25%, which is not noticeable visually but subtly draws the attention toward the center of the screen and reduces the effects of flicker in peripheral vision [23].

The function of the eyes within HVS indicates that the perceived visual quality depends on light. HVS possess a sensitivity to light characteristic, such as luminance and contrast. Luminance L is the amount of visible light leaving a point in a surface in a given direction. The standard unit to quantify the luminance is candela per square meter

(cd/m^2). The HVS is capable of adapting to an enormous range of light intensities. Light adaptation allows us to better discriminate relative luminance variations at every light level. It is important to note that human eye is more sensitive to dark gray than light gray. The response of HVS depends much less on the absolute luminance than on the relation of its local variations to the surround luminance. A typical experiment to determine this characteristic of HVS is illustrated in Figure 32. Participants of the experiments are asked to adjust *Lb*, the luminance of the middle circle, to its surrounding *Ls* so that they perceived a change in luminance intensity. This method is repeated for different surrounding which produces characteristic of human sensitivity to contrast. Contrast sensitivity is a measure of the ability to discriminate between different levels of luminance in static image [54].



Lb: background luminance

Ls: surround luminance

Figure 32: Experiment to determine luminance variation response.

By defining $\Delta L=Lb-Ls$, it was found that $\Delta L/Lb$ remains constant (c=0.02) for a variety of different luminance surrounding. This is known as the Weber-Fechner law, which can be expressed as:

$$C = \frac{\Delta L}{Lb} \tag{6}$$

The threshold contrast, which is the minimum contrast necessary for an experiment's participant to detect a change in intensity, is a function of background luminance, and it remains constant over an important range of intensities (from faint lighting to daylight) due to the adaption capabilities of the HVS which is illustrated in Figure 33 [4].

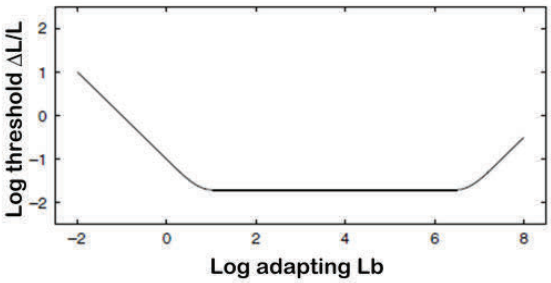


Figure 33: Weber-Fechner law [4].

The threshold contrast also depends considerably on the stimulus characteristics, i.e., color, spatial frequency and temporal frequency. Contrast sensitivity is defined as the inverse of the contrast threshold, and contrast sensitivity functions (CSF) are usually utilized to quantify the dependencies of threshold contrast *C*. Figure 34 illustrates the shape of the spatial contrast sensitivity function in an intuitive way [55].

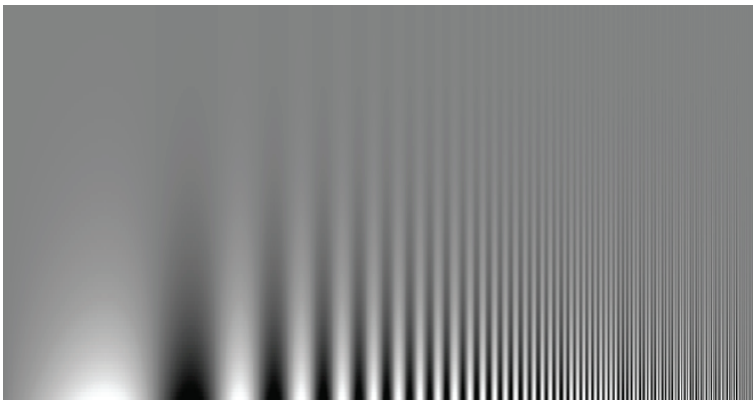


Figure 34: Campbell-Robson contrast sensitivity chart.

The information on the chart is as follow: the luminance of pixels is modulated sinusoidally along the horizontal dimension. The frequency of modulation increases exponentially from left to right; on the other hand the contrast decreases exponentially from 100% to about 0.5% from bottom to top. The minimum and maximum luminance remains constant along any given horizontal line through the image. The spatial CSF appears as the envelope of visibility of the modulated patterns. This chart show that the bars look taller in the middle of image that at the sides. Consequently, it can be concluded that the detection of contrast were not dictated by image contrast only since the alternating bright and dark bars pattern does not appear to have equal height everywhere in the image.

Visual information is processed in different pathways and channels in the visual system depending on its characteristic such as color, spatial and temporal frequency,

orientation, phase, direction of motion, and so on. These channels play an important role in explaining interactions between stimuli [4].

Color perception is based on the different spectral sensitivities of photoreceptors and the decorrelation of their absorption rates into opponent colors [4]. Generally, spectral power distribution is used to depict light. HVS perceives only lights with wavelength between 400 nm that corresponds to color of violet and 700 nm that corresponds to red. The sensitivity level depends on the angle of the incidence, and the maximum level of sensitivity is at 555 nm. Spectral sensitivity is the relative efficiency of detection of light as a function of the wavelength of the signal. This characteristic is shown in Figure 35.

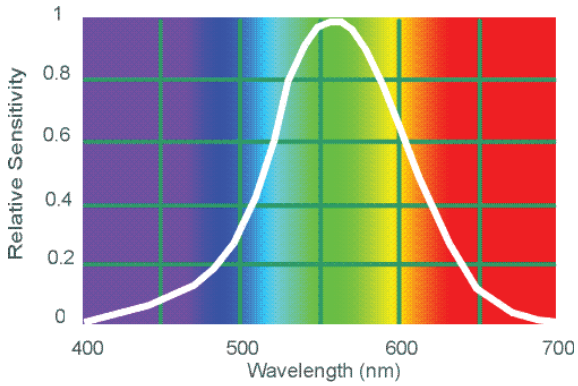


Figure 35: Spectral sensitivity of HVS.

Generic block diagram of a HVS based metric is illustrated in Figure 36 [4]. The input image or video typically undergoes color processing, which may include color space conversion and lightness transformations, a decomposition into a number of visual channels (for multi-channel models), application of the contrast sensitivity function, a model of patter masking, and pooling of the data from the different channels and locations.

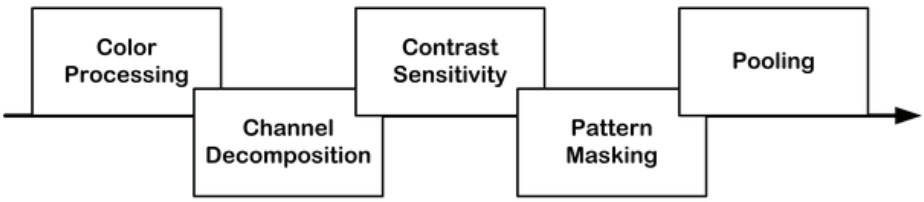


Figure 36: Generic block diagram of a vision-based quality metric [4].

One example of perceptual-based objective metric developed using engineering approach is Structural Similarity Based image quality assessment [53, 56], which is described more detail in the papers included in this thesis.

2.3.2 Audio Quality Metrics

It appears that the standardisation process for audio and speech have proceeded far ahead compare to any other applications. Unlike perceptual quality metrics for image and video, there are already standardised perceptual quality metrics for audio [6] and speech [5, 7] nowadays. The Radiocommunication Sector of the ITU focuses upon applications relating to audio for radiocommunication. Consequently, the sector considers these following features: basic audio quality and full-band audio (from 20 Hz to 20 kHz). Moreover, the ITU Telecommunications Standardisation Sector focuses upon applications relating to telecommunication, and consequently this sector has focused on these following features: speech, listening, and conversational quality and telephony bandwidth. Telephony bandwidth is classified into two bands, i.e., wideband (150-7000 Hz) and narrowband (300-3400 Hz).

With regard to speech and telecommunication applications, there have been perceptual metrics that have been developed to predict the result of speech listening quality test as performed using an ITU-T recommendation P.800 [57] absolute category rating (ACR) test method. The perceptual evaluation of speech quality (PESQ) model has been developed for narrowband telephony speech and provides high prediction accuracy in this application. In PESQ, the original and degraded signals are mapped onto an internal representation using a perceptual model. The internal representations that are used by the PESQ cognitive model to predict the perceived speech quality are calculated on the basis signal representations that use the psychophysical equivalents of frequency and intensity. The difference in this representation is used by a cognitive model to predict the perceived speech quality of the degraded signal. This perceived listening quality is expressed in terms of Mean Opinion Score (MOS). Most of the subjective experiments for developing PESQ used the ACR (Absolute Category Rating) and 5 discrete level quality scale. Figure 37 illustrates the overview of the basic philosophy used in PESQ. A computer model of the subject, consisting of a perceptual and a cognitive model, is used to compare the output of the device under test with the input, using alignment information as derived from the time signals in the alignment module [58].

In the first step of PESQ a series of delays between original input and degraded output are computed, one for each time interval for which the delay is significantly different from the previous time interval. For each of these intervals a corresponding start and stop point is calculated. The alignment algorithm is based on the principle of comparing the confidence of having two delays in a certain time interval with the confidence of having a single delay for that interval. The algorithm can handle delay changes both during silences and during active speech parts. Based on the set of delays that are found, PESQ compares the original (input) signal with the aligned degraded output of the device under test using a perceptual model. The key to this process is transformation of both the original and degraded signals to an internal representation that is analogous to the psychophysical representation of audio signals in the Human Auditory System (HAS), taking account of perceptual frequency (Bark) and loudness (Sone). This is achieved in several stages: time alignment, level alignment to a calibrated listening test, time-frequency mapping, frequency warping, and compressive loudness scaling. In PESQ, two error parameters are computed in the cognitive model, which are then combined to give an objective listening MOS [7]. The range of the

PESQ score is -0.5 to 4.5, although for most cases the output range will be a listening quality MOS-like score between 1.0 and 4.5, the normal range of MOS values found in an ACR experiment.

Recently, a wideband telephony speech version of PESQ has been standardised too in ITU-T recommendation P.862.2 [59]. The wideband extension to this includes a mapping function that allows linear comparison with MOS values produced from subjective experiments that include wideband speech conditions with an audio bandwidth of 50-7000 Hz.

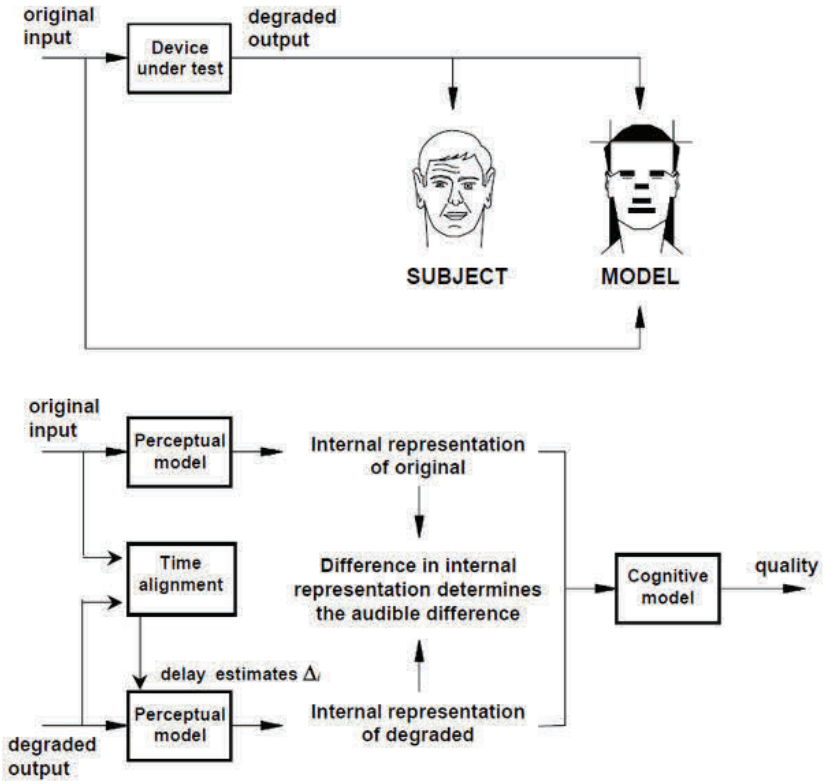


Figure 37: Overview of the basic philosophy used in PESQ [59].

There is also a model that requires no reference signal. This so-called non-intrusive model for predicting the subjective quality of narrowband telephony applications is standardised in ITU-T recommendation P.563 [60]. This model is called single-ended method for objective speech quality assessment in narrow-band telephony application and is of great interest in the monitoring of speech quality in live telephony network. The difference between non-intrusive model and the intrusive model is illustrated in Figure 38. MOS-LQO shown in the figure is MOS scores that are

applicable to a listening-only situation and are calculated by means of objective model which aims at predicting the quality for a listening-only test situation [61].

The P.563 approach is illustrated in Figure 39.

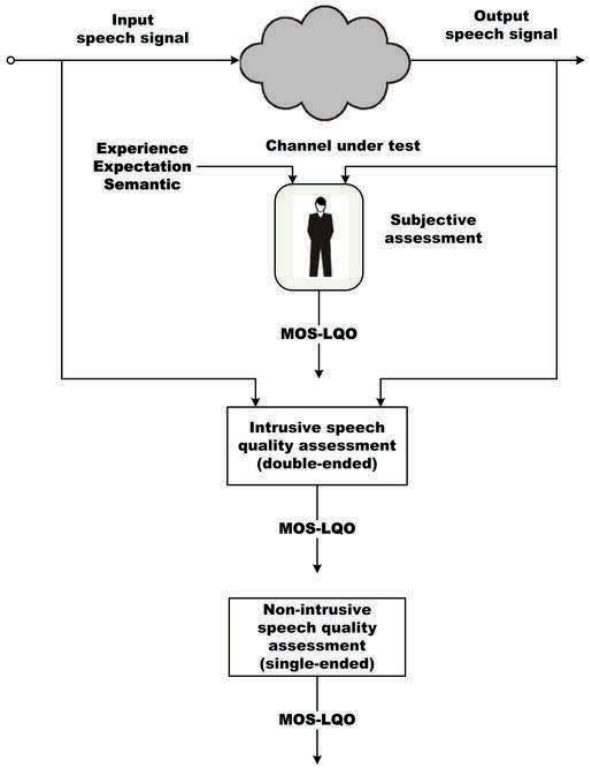


Figure 38: Non-intrusive versus Intrusive models [60].

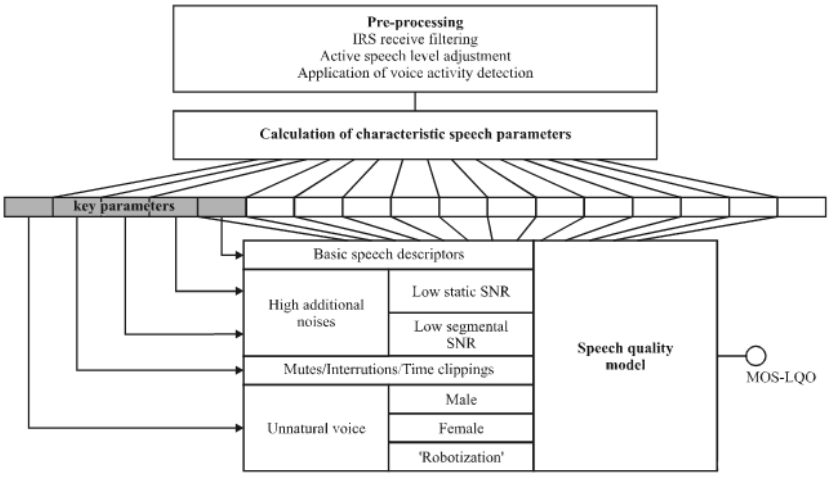


Figure 39: Block scheme of P.563.

For applications in low bit-rate audio codecs, the perceptual evaluation of audio quality (PEAQ) model has been standardised in ITU-R recommendation BS.1387 [62].

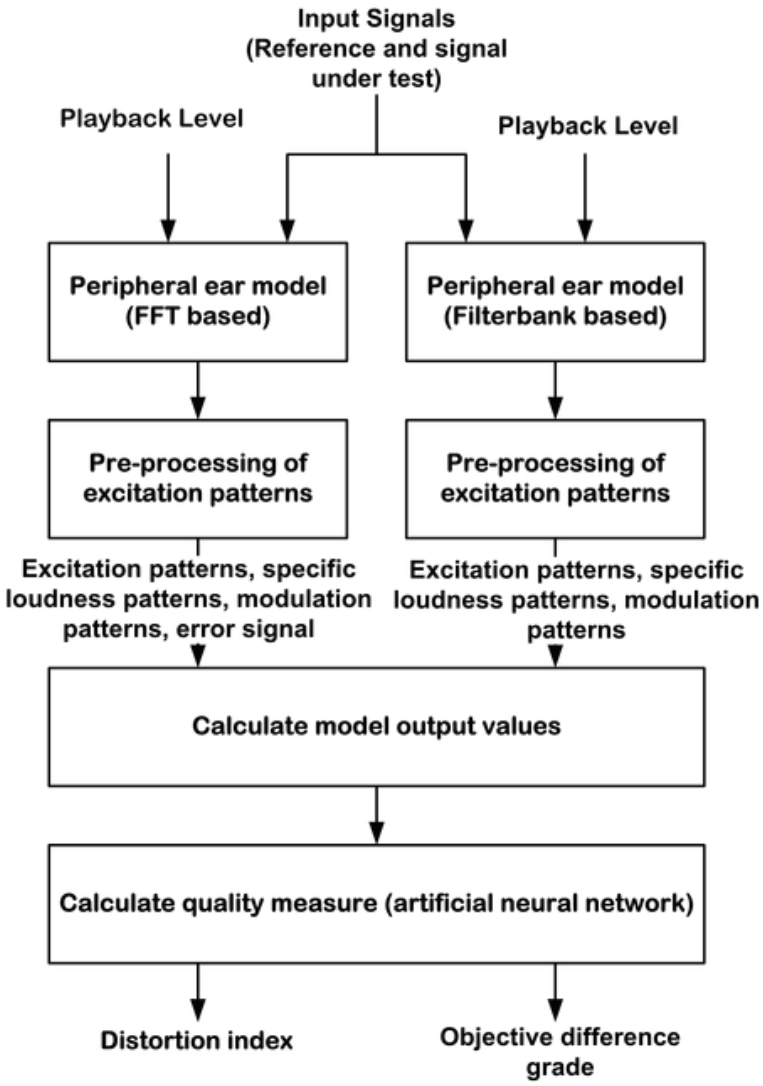


Figure 40: Generic block diagram of the measurement scheme [62].

The proposed PEAQ, as illustrated in the Figure 40, consists of peripheral ear model, several intermediate steps which is referred as pre-processing of excitation patterns, the calculation of pycho-acoustically based outputs and a mapping from a set of outputs to a single value representing basic audio quality [6].

As time passes, the frequency bandwidth of communication audio widens. Accordingly, it can be expected that there will be an increasing overlap between the application domains covered by the Telecommunications and Radiocommunication Standardisation Sector in future [8].

2.3.3 Audiovisual Quality Metrics

We rarely watch the moving picture in all its incarnations (video, television, cinema, etc) without sound. For that reason, quality models or metrics that measure entirely audiovisual quality of multimedia presentation are needed. The audiovisual quality metrics take into account both audio and visual factors in a comprehensive way; multimodal effects on the perceived quality are considered as the significant factors of the models [63]. Developing perceptual based audiovisual quality metric is started by understanding how human participants perceive audio-visual quality. This is achieved fundamentally by studying how human perceive the quality of auditory and visual stimuli, and at what stage in human perceptual process they are fused to form a single overall quality experience. The influence of video quality on perceived audio quality and the influence of audio quality on video quality of contents from broadcast audio and video and videophone (telephone) has been studied in [64]. There is a significant mutual influence between audio and video quality. The result showed that when participants are asked to judge the audio quality of audiovisual stimulus, the video quality contributed significantly to the perceived audio quality. On the other hand when participant were asked to judge the video quality of audiovisual stimulus, the audio quality has less impact on overall quality. In addition, a simple mapping from the audio and video quality to the overall audiovisual quality showed that video quality dominates the overall perceived quality in non-conversional experiments. The research also offered a metric that overall audiovisual quality can predicted from the perceived audio quality in an audio only experiment and the perceived video quality in a video only experiment.

This study [65] described a multimedia opinion model based on an objective quality assessment for audio-visual communications intended for videophone, PDA, and mobile videophone services, taking into account the mutual interaction of audio and video information. The research showed that it is important to take into account the mutual interaction of audio and video information when audiovisual quality of the multimedia opinion model/metric.

Approaches described in two previous studies are the common approaches used to develop metrics to measure perceived audiovisual quality. They started from perceptual experiments which were conducted to study perceptual audiovisual quality; these perceptual experiments consist of subjective audio quality assessment, subjective visual quality, and subjective audiovisual quality. The collected subjective data of audio quality, visual quality, and overall audiovisual quality are then utilized to develop audiovisual quality metrics. Thus, most of the study derives the metrics by determining audiovisual quality from audio quality and video quality. The most common model is [66]:

$$\begin{aligned} &AudiovisualQuality \\ &= a_o + a_1 AudioQuality + a_2 VisualQuality + a_3 (AudioQuality \cdot VideoQuality) \end{aligned} \quad (1)$$

where the parameters $\{a_1, a_2, a_3\}$ denote different weights of audio and video quality, and the multiplication factor for the overall quality. The parameter a_o is used to improve the fit. The overall audiovisual quality is influenced by several factors, but the audio quality and visual quality are the most important ones.

Another key element that contributes to the overall perceived audiovisual quality is synchronization between audio and video stimuli. Study of synchronization between audio and visual factors in a multimedia presentation has been conducted in [67]. The presentation of 'in sync' data streams of audio and video is essential to achieve a natural impression, data that is 'out of sync' is perceived as annoying, strange, and artificial. Several experiments were conducted. These includes the lip-synch issue—the temporal connection between audio and video stream for the particular case of human speaking is investigated. The test sequences used for the stimuli were news clip. This experiment showed that out of sync area spanned a skew of 80 ms between audio and video were still deemed acceptable by most casual observers. The synchronization of audio-video was taken into account in the model described here [68]. Again the model is intended for videophone applications because such applications are expected to become popular on the next-generation network (NGN). Before the model was developed, subjective quality assessment tests were conducted using PC-based point-to-point interactive videophone application to study how overall quality of the multimedia presentation characteristic depend on individual audiovisual quality, the absolute audiovisual delay and media synchronization. The framework of the model for this videophone application is illustrated in Figure 41.

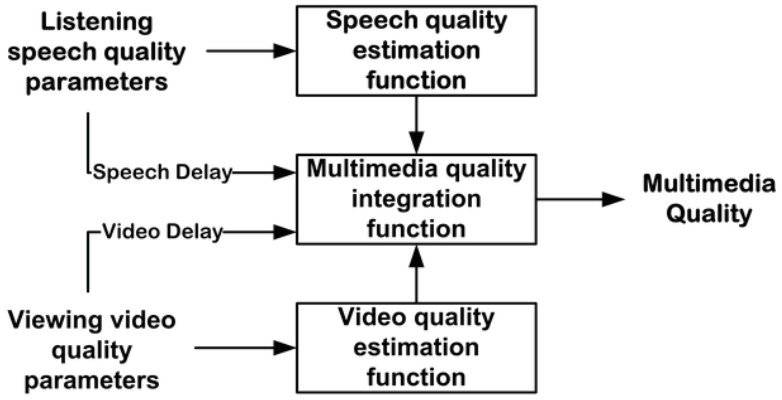


Figure 41: Framework of the model proposed by Hayashi et al [68].

The resulted model is:

$$MOS_{MM}=c_1MOS_{AV}+c_2MOS_D+c_3MOS_{AV}MOS_D+c_4, \quad (2)$$

where c_1, c_2, c_3, and c_4 are constants. This function assumes that overall multimedia quality MOS_{MM} can be estimated from audiovisual quality MOS_{AV}, degradation delay quality MOS_D, and their interaction term, and this function is also similar with the model presented with Eq (7).

3 Outline and Comments of Paper

In this section an outline of the papers included in this thesis is given. There are four published conference papers and one submitted journal paper. The first paper mainly looks into the usage of one of the available perceptual-based objective metrics—a structural similarity metric—in Digital Cinema. The second paper compares two compression algorithm—JPEG 2000 and H.264/AVC—using a subjective visual quality assessment. The third paper studies the importance of the Quality of Experience concept for developing Digital Cinemas into multi-arts venues. The fourth paper presents a study on the impact of audio content on visual perceived quality scores collected from an perceptual experiment—subjective visual quality assessment—in Digital Cinema. The last paper examines more closely the result of subjective visual quality assessments discussed in the first and the second paper. The author had a major role in research and writing in all these papers, and the author's contribution in every particular paper is specified at the end of each summary.

3.1 Paper A – SS-SSIM and MS-SSIM for Digital Cinema Applications [11]

One of the key issues for a successful roll out of digital cinema is in the quality it offers. The most practical and least expensive way of measuring quality of multimedia content is through the use of objective metrics. In addition to the widely used objective quality metric peak signal-to-noise ratio (PSNR), recently perceptual-based quality metrics such as single scale structural similarity (SS-SSIM) and multi scale structural similarity (MS-SSIM) have been asserted as good alternatives for estimation of perceived quality by taking into account the Human Visual System (HVS) characteristic. This paper studied the suitability of SS-SSIM and MS-SSIM to measure the perceived quality of images. In addition to application of these metrics using their original parameters, new parameters for MS-SSIM were obtained by taking into account the digital cinema viewing conditions, and used in this study. New parameters of MS-SSIM were obtained by performing a subjective parameterization test based on image synthesis approach for cross-scale calibration. Such tests were conducted in the same DCI-specified cinema theatre using a 12 x 5 m screen. To validate the results of these metrics, the correlation between the objective metrics and the ground truth, was investigated. The ground truth was how human participants perceive the same content in terms of quality. In the case of digital cinema content and environment, it seems that both SS-SSIM and MS-SSIM do not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric. The author wrote all of the paper. Dr. Ulrich Reiter provided an extensive support during the subjective quality assessment and in writing of the paper. Professor Touradj Ebrahimi, Professor Andrew Perkis, and Professor Peter Svensson supervised the work.

Further analyses were then conducted to the subset of collected subjective score from the experiment described in Paper A. This study was published in another conference paper titled "*Analysis of SSIM Performance for Digital Cinema Applications*"[69]. The basis of this study is due to digital cinema practice; feature film screening practice in Digital Cinema only utilized high quality imagery. Consequently, we only use high score of MOS collected from subjective quality assessment as the ground truth--we disregard the votes below fair. Based on calculated correlation coefficient values, the PSNR had the highest correlation with subjective data. However, there are no significant differences between correlation coefficients of objective metrics investigated in this paper. Hence, based on this result, there is no objective model that comes out as best performer from a statistical point of view, if we take into account only higher quality data.

As an extension of study described in Paper A, we also proposed an approach to improve the performance of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) for image quality assessment in digital cinema applications. We published this study in a conference paper titled "*HVS-based Image Quality Assessment for Digital Cinema*" [70]. Based on the particularities of quality assessment in digital cinema, some attributes of the human visual system (HVS) are taken into consideration, including the fovea acuity angle and contrast sensitivity, and combined with viewing conditions in digital cinema to select appropriate image blocks for calculating the perceived quality by PSNR and SSIM. Furthermore, as the HVS is not able to perceive all the distortions because of selective sensitivities to different contrasts, and masking always exists, we adopt a modified PSNR by considering the contrast sensitivity function and masking effects. The experimental results demonstrated that the proposed approach can evidently improve the performance of image quality metrics in digital cinema applications. Based on an intensive analysis on the mechanism of image quality assessment in a digital cinema setup, we proposed an approach for improving the performance of three image quality metrics. The images were divided into different blocks with a given size, and metrics were performed in certain blocks with high contrast levels. The mean of quality values over these blocks was taken as the image quality. The experimental results with respect to the subjective quality results in the digital cinema setup and LIVE data set demonstrated the promising performance of the proposed approach in improving the image quality metrics for digital cinema applications.

3.2 Paper B – Comparison of JPEG 2000 and H.264/AVC by Subjective Assessment in the Digital Cinema [12]

This paper studies two existing compression algorithms, JPEG 2000 and H.264/AVC in the context of Digital Cinema applications. In Digital Cinema, the use of compression is a matter of practicality; the quantity of data needed to represent high-quality imagery in its native forms is staggering. In 2005 the Digital Cinema Initiative (DCI), an organization that is a joint venture of seven major Hollywood studios, concluded on using JPEG 2000 for compression of Hollywood feature films for the large screen. By this decision DCI initiated a large roll out of digital equipment to cinemas all over the world. After digital equipment is installed, the theatre owners have the possibility to utilize this infrastructure to screen content outside of ordinary feature film screening. This produces a concept of alternative content, which also create alternative

compression algorithms used for Digital Cinema applications. We focus on such alternative content displayed using DCI specified equipment and showed on the large screen in a real theatre. Consequently, it influenced the compression techniques and parameters that were chosen and applied in this study. Subjective visual quality assessment in Digital Cinema was utilized as a mean to examine JPEG 2000 and H.264/AVC compression algorithm. The subjective score collected in a carefully designed experiment is still considered the benchmark of quality evaluation. Accordingly, we also looked into the appropriate experimental test design of conducting such assessment. We proposed a protocol to conduct subjective visual quality assessment in Digital Cinema. This includes the processing of the collected result from the test. We demonstrated that the algorithm that includes the temporal compression schemes like H.264/AVC for presentation on a large screen was very possible; the gain in bit rate that a temporal compression scheme provides, can very well be used to further increase the quality of the encoded stream. The author is responsible for the entire writing of the publication. Dr. Ulrich Reiter supervised the work and made observation in writing the publication. Marlon Nielsen assisted the author during the experiments. Professor Touradj Ebrahimi and Professor Andrew Perkis supervised the work.

3.3 Paper C – Exploring Alternative Content in Digital Cinema [13]

This paper presents the concept of alternative content. We showed that Digital Cinema business is much more that feature films. We also gave an overview on this alternative content. One of the main problems that hold back the successful roll out of D-Cinema in the market is that the theatre owners or exhibitors are those in value chain with the least benefit of the digitization. We offered an insight on how experimentation beyond traditional features films in Digital Cinema can benefit the theatre owners who embraced the digital technology. In addition, the alternative content is likely to offer new experience to the cinemagoers, it enables the cinema to become a multi-arts venue, attracting new and existing users by offering a range of products. Alternative content screening can transform the whole business of cinema exhibition into something different from what we know of today. Accordingly, we consider the Quality of Experience is a significant factor that is closely associated with further adoption of alternative content screening in Digital Cinema and is crucial for driving the innovation in Digital Cinema practice. In additions, we consider it is advantageous to use and set up a DCI-specified commercial cinema into a realistic test environment for conducting perception experiments. The author is responsible for writing the entire paper. Professor Andrew Perkis supervised the work and made many observations in the research work. Professor Touradj Ebrahimi supervised the research.

3.4 Paper D – Subjective Visual Quality Assessment in the Presence of Audio for Digital Cinema [14]

This paper investigated whether the presence of audio with different quality levels can influence the outcome of subjective visual quality assessment in a Digital Cinema

setting. We also emphasize alternative content displayed using DCI specified equipment and showed on the large screen in a real theatre. Hence, it influences the chosen stimuli used in subjective quality assessment in Digital Cinema. The stimuli used were 10 seconds long colour sequences accompanied by orchestral music at 2K resolution, 24 fps, and YCbCr 4:4:4 played on DCI certified equipment. We offered a protocol for a perceptual test by asking the experiment's participants to judge the visual quality when watching an audiovisual content in a Digital Cinema environment and examined whether the participants can neglect the presence and the quality of audio. In order to create stimuli with various visual quality levels, we encoded our data set with JPEG 2000 at different coding bit rates. We selected the bit rates of 20 Mbps, 40 Mbps, 60 Mbps, and 160 Mbps. We also incorporated four audio conditions (no audio, uncompressed audio, and two compressed conditions) for each selected bit rates of JPEG 2000. The result show that in visual only subjective quality assessment, the presence of audio (low or high quality) does not significantly influence on the visual quality judgment. The author wrote the entire paper. Both Dr. Ulrich Reiter and Dr. Junyong You made many observations in the research work and in the writing of publication. Professor Andrew Perkis and Professor Touradj Ebrahimi supervised the research work.

3.5 Paper E – A Study of Quality of Experience in D-Cinema [15]

QoE always puts the end-user at the centre of attention and it is a multidimensional concept, which consists of several objective and subjective parameters. This contributes to the difficulty of quantifying QoE. We focus on the subjective quality assessment for D-Cinema application because we believe it is an important aspect in studying QoE for D-Cinema; it is the basis to understand the perceived quality and is useful for developing a mature QoE model for D-Cinema. For this reason, subjective quality assessment for D-Cinema applications must be carefully designed. This paper offers a study of visual quality of multimedia presentations in D-Cinema applications. The study presented in this paper is an extension of Paper A and Paper B. Paper A and Paper B described the subjective visual quality assessment of images and motion pictures in a DCI-specified commercial Digital Cinema in Trondheim, Norway. Our interest is in exploring screening of alternative content using the D-Cinema equipment and environment affect the designs of the assessment. Using analysis of variance, we detected the significant differences of subjective scores among participants. Consequently, the obtained subjective scores were normalized first before MOS were computed. Our study showed that due to the different and unique digital image content and viewing conditions of D-Cinema, quality research of D-Cinema especially in the context of QoE is not really in the same category as other applications. Initial impression of our study showed that the stimulus presentation method influenced how participants used the quality scale when judging the perceived visual quality. Participants seem confident using the highest end of quality scale when judging the transparent stimuli when simultaneous double stimulus method was employed during subjective visual quality assessment of images, on the other hand participants showed hesitancy using the highest end of quality scale when judging the transparent stimuli

when single stimulus method was employed during subjective visual quality assessment of motion pictures. The results also showed that the content types influenced the subjective scores. In the assessment of motion pictures, we also showed the result of the differences between two compression algorithms. The author is responsible for writing the entire paper. Dr. Ulrich Reiter, Professor Touradj Ebrahimi and Professor Andrew Perkis supervised the research work and provided extensive support in writing the paper.

4 Conclusion

The major contributions of this thesis are:

- Protocols of subjective quality assessment for images and motion pictures in Digital Cinema. The protocols are based on the methodologies described in ITU recommendations ITU-R BT.500 [29]and ITU P.910 [35]. We also provided an analysis of the protocols that we utilized.
- New parameters for MS-SSIM objective metrics were obtained by conducting subjective test that takes into account the digital cinema environment.
- Analysis of the performance of three popular perceptual objective metrics (structural similarity based metric).
- Assessment of compression technologies for alternative contents screening in digital cinema. We analysed JPEG 2000 and H.264/AVC compression algorithm based on the collected data from subjective quality assessment conducted in the digital cinema using protocols described above.
- An overview of alternative contents screening which can influence the business model of digital cinema.
- Protocol of subjective quality assessment to investigate whether the presence of audio influenced the outcome of subjective visual quality assessment in digital cinema. We also provided the analysis whether the audio influence the visual quality based on the collected subjective data from the experiments using the proposed methodology.

References

[1] R. Jain, "Quality of Experience," *IEEE Multimedia,* vol. 11, pp. 95-96, 2004.

[2] ITU-T, "Vocabulary for performance and quality of service Amendment 1: New Appendix I -- Definition of Quality of Experience (QoE)," ITU, Geneva2006.

[3] A. Perkis, F. N. Rahayu, U. Reiter, J. You, and T. Ebrahimi, "Quality of Experience for High Definition Presentations--Case:Digital Cinema," in *High-Quality Visual Experience*, ed: Springer, 2010.

[4] S. Winkler, *Digital Video Quality - Vision Models and Metrics*: John Wiley & Sons, 2005.

[5] ITU-T, "Telephone Transmission Quality. Objective Measuring Apparatus. Objective Measurement of Active Speech Level.," ITU1993.

[6] ITU-R, "Method for Objective Measurement of Perceived Audio Quality," ITU1998.

[7] ITU-T, "Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs," International Telecommunications Union2001.

[8] S. Bech and N. Zacharov, *Perceptual Audio Evaluation--Theory, Method and Application*. England: John Wiley & Sons Ltd, 2006.

[9] A. Perkis and P. V. Sychowski, "NORDIC – Norway's digital interoperability in cinemas," *Journal St. Malo: NEM - Networked and Electronic Media,* 2008.

[10] A. Perkis, "Does quality impact the business model? Case: Digital Cinema," presented at the IEEE International Workshop on Quality of Multimedia Experience, San Diego, 2009.

[11] F. N. Rahayu, U. Reiter, T. Ebrahimi, A. Perkis, and P. Svensson, "SS-SSIM and MS-SSIM for Digital Cinema Applications," in *SPIE-IS&T Human Vision and Electronic Imaging XIV*, San Jose, USA, 2009, pp. 72400P-1-72400P-12.

[12] F. N. Rahayu, U. Reiter, M. T. M. Nielsen, T. Ebrahimi, and A. Perkis, "Comparison of JPEG 2000 and H.264/AVC by Subjective Assessment in the Digital Cinema," in *2nd IEEE International Conference on Quality of Multimedia Experience*, Trondheim, Norway, 2010, pp. 112-117.

[13] F. N. Rahayu, T. Ebrahimi, and A. Perkis, "Exploring Alternative Content in Digital Cinema," in *16th IEEE International Conference on Virtual Systems and Multimedia (VSMM)*, Seoul, Korea,, 2010, pp. 295-296.

[14] F. N. Rahayu, U. Reiter, J. You, A. Perkis, and T. Ebrahimi, "Subjective Visual Quality Assessment in the Presence of Audio for Digital Cinema," in *3rd IEEE International Conference on Quality of Multimedia Experience*, Belgium, 2011.

[15] F. N. Rahayu, U. Reiter, A. Perkis, and T. Ebrahimi, "A Study of Quality of Experience in D-Cinema," *Signal Processing: Image Communication, Theory, Techniques & Applications, A publication of the European Association for Signal Processing (EURASIP) (submitted),* 2011.

[16] D. C. Initiatives, "Digital Cinema System Specification version 1.2," March 2008.

[17] DCI, *Digital Cinema System Specification version 1.2*: http://www.dcimovies.com, 2008.

[18] C. Carey, B. Lambert, B. Kinder, and G. Kennel, "The Mastering Process," in *Understanding Digital Cinema A Professional Handbook*, C. S. Swartz, Ed., ed: Focal Press, 2005.

[19] L. Silverman, "The New Post Production Workflow: Today and Tomorrow," in *Understanding Digital Cinema A Professional Handbook*, C. S. Swartz, Ed., ed: Focal Press, 2005.

[20] P. Symes, "Compression for Digital Cinema," in *Understanding Digital Cinema: A Professional Handbook*, ed: Focal Press, 2005, pp. 121-148.

[21] D. Antonellis, "Digital Cinema Distribution," in *Understanding Digital Cinema A Professional Handbook*, C. S.Swartz, Ed., ed: Focal Press, 2005.

[22] M. Karagosian, "Theatre Systems," in *Understanding Digital Cinema A Professional Handbook*, C. S. Swartz, Ed., ed: Focal Press, 2005.

[23] M. Cowan and L. Nielsen, "Projection," in *Understanding Digital Cinema A Professional Handbook*, C. S.Swartz, Ed., ed: Focal Press, 2005.

[24] SMPTE, "Standard for Luminance using 35 mm projection.," Society of Motion Pictures and Television Engineers2003.

[25] T. Instruments. (August). *DLP Technology*. Available: http://www.dlp.com

[26] SONY, "4K Digital Cinema Projectors SRX-R220/SRX-R210 Media Blok LMT-100 Screen Management System LSM-100," 2007.

[27] T. K. AS. (2011). Available: http://www.trondheimkino.no/article42891.ece

[28] Christie, "CP2230 User Manual," 2010.

[29] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva2002.

[30] ITU-T, "Subjective audiovisual quality assessment methods for multimedia applications," ITU, Geneva2000.

[31] ITU-T, "Subjective listening test method for evaluating digital circuit multiplication and packetized voice systems--telephone transmission quality subjective opinion tests.," ITU1993.

[32] ITU-T, "A method for subjective performance assessment of the quality of speech voice output devices," ITU1994.

[33] ITU-R, "Multichannel Stereophonic Sound Systems with and without Accompanying Picture," 1994.

[34] ITU-R, "Subjective Assessment of Sound Quality--A Guide to Existing Recommendations," ITU1998.

[35] ITU-T, "Subjective Video Quality Assessment Methods for Multimedia Applications," ITU1996.

[36] ITU-R, "Subjective assessment of the quality of audio in large-screen digital imagery applications intended for presentation in a theatrical environment," ITU2004.

[37] ITU-R, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," ITU1997.

[38] ITU-R, "Methods for the subjective assessment of audio systems with accompanying picture. ," ITU1998.

[39] ITU-R, "Pre-selection methods for the subjective assessment of small impairments in audio systems," ITU1998.

[40] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering--Fundamentals, Algorithms, and Standards*: CRC Press, 2000.

[41] VQEG. (2011). *Video Quality Experts Group*. Available: http://www.vqeg.org/

[42] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment," VQEGMarch 2000.

[43] VQEG, "Final report from the Video Quality Experts Group on the validation of objective model of video quality assessment, phase II," VQEG2003.

[44] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding*: CRC Press, 2006.

[45] J. Mannos and D. Sakrison, "The effect of a visual fidelity riterion on the encoding of images," *IEEE Trans. On Inf. Theory,* vol. 20, pp. 525-536, July 1974.

[46] F. Lukas and Z. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. On Commun,* vol. 30, pp. 1679-1692, July 1982.

[47] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *Proc. Of SPIE Human Vision, Visual Processing, and Digital Display III,* vol. 1666, pp. 2-15, August 1992.

[48] J. Lubin, "A visual discrimination model for imaging system design and evaluation," *Vision Models for Target Detection and Recognition, World Scientific,* pp. 245-283, 1995.

[49] J. B. Martens, "Multidimensional modelling of image quality," *Proc of the IEEE,* vol. 90, pp. 133-153, January 2002.

[50] P. C. Teo and D. J. Heeger, "Perceptual image distortion," *in Proc. of IEEE Int. Conf. On Image Processing,* vol. 1, pp. 982-986, November 1994.

[51] A. B. Watson and J. A. Salomon, "Model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America,* vol. 14, pp. 2379-2391, September 1997.

[52] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans on Commun,* vol. 43, pp. 2959-2965, December 1995.

[53] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," in *IEEE Asilomar Conference on Signals, Systems and Computers*, Asilomar, 2003.

[54] B. Girod. (August). *Human Visual Perception - topics*. Available: http://www.stanford.edu/class/ee368b/Handouts/09-HumanPerception.pdf

[55] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of grating," *Journal of Physiology,* vol. 197, pp. 551-566, 1968.

[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, p. 13, April 2004.

[57] ITU-T, "Methods for subjective determination of transmission quality," International Telecommunications Union1996.

[58] J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier, "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment part II--psychoacoustic model," *Journal of the Audio Engineering Society,* vol. 50, pp. 765-778, October 2002.

[59] ITU-T, "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs.," International Telecommunications Union2007.

[60] ITU-T, "Single-ended method for objective speech quality assessment in narrow-band telephony applications.," International Telecommunications Union2004.

[61] ITU-T, "Mean Opinion Score (MOS) terminology," International Telecommunications Union2006.

[62] ITU-R, "Method for objective measurements of perceived audio quality," International Telecommunications Union2001.

[63] S. Winkler, "Video Quality and Beyond," in *15th European Signal Processing Conference (EUSIPCO 2007)*, Poznan , Poland, 2007, pp. 150-153.

[64] J. G. Beerends and F. E. D. Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *Journal of the Audio Engineering Society,* vol. 47, 1999.

[65] N. Kitawaki, Y. Arayama, and T. Yamada, "Multimedia opinion model based on media interaction of audio-visual communications," in *4th International Conference on Measurement of Speech and Audio Quality in Network*, Prague, Czech Republic, 2005, pp. 5-10.

[66] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: A survey," *Signal Processing: Image Communication* vol. 25, pp. 482-501, 2010.

[67] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE J. Selected Areas in Communication,* vol. 14, pp. 61-72, 1996.

[68] T. Hayashi, K. Yamagishi, T. Tominaga, and A. Takahashi, "Multimedia quality integration function for videophone services," presented at the IEEE Int. Conf. Global Telecommunication 2007.

[69] F. N. Rahayu and U. Reiter, "Analysis of SSIM Performance for Digital Cinema Applications," in *1st IEEE QoMEX*, San Diego, USA, 2009.

[70] J. You, F. N. Rahayu, U. Reiter, and A. Perkis, "HVS based Image Quality Assessment for Digital Cinema," in *SPIE Electronic Imaging: Image Quality and System Performance VII*, San Jose, USA, 2010.

Paper A: SS-SSIM and MS-SSIM for Digital Cinema Applications

A

Fitri N. Rahayu, Ulrich Reiter, Touradj Ebrahimi, Andrew Perkis, and Peter Svensson

Appeared in
Proceedings SPIE-IS&T Human Vision and Electronic Imaging XIV Vol. 7240, San Jose, USA, 2009

Abstract

One of the key issues for a successful roll out of digital cinema is in the quality it offers. The most practical and least expensive way of measuring quality of multimedia content is through the use of objective metrics. In addition to the widely used objective quality metric peak signal-to-noise ratio (PSNR), recently other metrics such as single scale structural similarity (SS-SSIM) and multi scale structural similarity (MS-SSIM) have been claimed as good alternatives for estimation of perceived quality by human subjects. The goal of this paper is to verify by means of subjective tests the validity of such claims for digital cinema content and environment.

1 Introduction

The motion picture industry is one of the many players in the media industry. Both broadcasting and mobile media have successfully completed their transition to fully digital services, while the motion picture industry is currently in the process of forming standards for digitization of its complete value chain. These specifications and standards are the basis for a large scale implementation of digital cinema as the latest and final analogue media to go digital. The digitization is specified by the Digital Cinema Initiative (DCI) and is currently under standardization by SMPTE [1]. One of the key issues for a successful roll out of digital cinema in the market is in the service assurance of the quality it offers. The ultimate measure of a service is how an end-user perceives its performance. Hence, the best way of measuring perceived quality is to rely on human subject assessment, in a controlled environment. This is referred to as subjective quality assessment.

Performing subjective assessments is time consuming, expensive, and complex. Furthermore, it does not lend itself to real-time environments. As an alternative, objective measurement methods (objective metrics) have been developed to predict the perceived quality of human subjects. Among objective metrics proposed to estimate perceived quality, Wang and Bovik introduced the structural similarity quality paradigm (SSIM) based on the assumption that the human visual system is highly adapted for extraction of structural information from a scene [2]. They argue that a measure of structural similarity can provide a good approximation of perceived quality. In their experiments, SSIM has shown a good correlation with perceived quality, outperforming traditional metrics such as peak-signal-to-noise ratio (PSNR). Multi-scale structural similarity (MS-SSIM) is proposed to supply more flexibility when compared to the single-scale method (SS-SSIM), by taking into account variations in viewing conditions [3]. These metrics have been used to measure perceived image quality in digital cinema.

In this paper, we report the results of a study to assess the suitability of SS-SSIM and MS-SSIM to measure the perceived quality of images from DCI Standard Evaluation Material (StEM) [4]. In addition to application of these metrics using their original parameters, new parameters for MS-SSIM were obtained by taking into account the digital cinema viewing conditions, and used in this study. To validate the results of these metrics, we investigated the correlation between the objective metrics and the ground truth, i.e. how human subjects perceive the same content in terms of quality. To this end, a subjective quality assessment was carried out in a DCI-specified movie theatre in Trondheim, Norway.

The paper is structured as follows. First, an overview of single scale structural similarity (SS-SSIM) and multi scale structural similarity (MS-SSIM) is given in Section 2. The subjective quality assessment in the DCI-specified movie theater environment and its results are provided in Section 3. The parameterization of MS-SSIM for digital cinema content and environment is presented in Section 4. Next, the test results are discussed in Section 5. Finally, we draw some conclusions in Section 6.

2 Overview of Structural Similarity Measures

Natural image signals are highly structured. Their pixels exhibit strong dependencies, which carry important information about the structure of the objects in the visual scene [2]. The human visual system is highly adapted to extract structural information. It is therefore assumed that the measurement of structural information changes provides a good estimation of the perceived image distortion [2]. Suppose x and y were two image signals; if one of the signals had perfect quality, then the similarity measure could be utilized to measure quantitatively the quality of the second signal.

Structural information in an image is defined as attributes that represent the structure of objects in the scene, which are independent of the illumination. Accordingly, the information of structure is independent of the average luminance and contrast. The quality assessment uses local luminance and contrast because luminance and contrast can vary across a scene. The similarity measurement system is based on three comparisons: luminance, contrast, and structure. The luminance comparison function l(x,y) is estimated as a comparison of the mean intensity of two discrete signals, □x and □y, which is defined by

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{A.1}$$

where the constant C_1= 6.50 [2] is included to avoid instability when μx2+ μy2 is very close to zero. The contrast comparison function c(x,y) takes a similar form, based on the standard deviation of the two signals, □x and □y,

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \tag{A.2}$$

Here again, the constant C_2= 58.52 [2] is included to avoid instability when $\sigma x^2+\sigma y^2$ is very close to zero. The third comparison — the structure comparison function s(x,y) — is defined as follows:

$$s(\mathbf{x},\mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \tag{A.3}$$

To avoid instability when σxσy is very close to zero, a constant C_3= 29.26 [2] is incorporated. The general form of the Structural SIMilarity (SSIM) index between signals x and y becomes:

$$SS-SSIM(\mathbf{x},\mathbf{y}) = \left[l(\mathbf{x},\mathbf{y})\right]^{\alpha} \cdot \left[c(\mathbf{x},\mathbf{y})\right]^{\beta} \cdot \left[s(\mathbf{x},\mathbf{y})\right]^{\gamma} \tag{A.4}$$

where α > 0, β > 0, γ > 0 are parameters used to adjust the relative importance of the three components. In the Single-Scale Structural SIMilarity (SS-SSIM) approach, the parameter values are set to α = β = γ = 1.

The perceivability of image details depends on the sampling density of the image signal and the distance of the image plane from the observer. When these factors vary, the subjective evaluation of a given image varies too. The single scale method does not incorporate such factors. For this reason, a multi scale variant of structural similarity has been developed to incorporate image details at different resolutions [3]. Taking the reference and the distorted image signals as input, the method iteratively applies a low-pass filter and down-samples the filtered image by a factor of 2. For example, at the j-th scale, the reference and the distorted image signals are low-pass filtered and down-sampled 2 j-1 times.

The overall computation is acquired by combining the measurement at different scales using

$$MS-SSIM(\mathbf{x},\mathbf{y}) = \left[l_M(\mathbf{x},\mathbf{y})\right]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(\mathbf{x},\mathbf{y})\right]^{\beta_j} \left[s_j(\mathbf{x},\mathbf{y})\right]^{\gamma_j} \quad \text{(A.5)}$$

in which the original image is indexed as scale 1. This definition also comprises the single scale measurement as the special case M = 1.

The exponents αM, βj, and γj are used to adjust the relative importance of the three components, and M is set to 5. Based on a subjective parameterization test [3], the resulting parameters are $\beta 1 = \gamma 1 = 0.0448$, $\beta 2 = \gamma 2 = 0.2856$, $\beta 3 = \gamma 3 = 0.3001$, $\beta 4 = \gamma 4 = 0.2363$, and $\alpha 5 = \beta 5 = \gamma 5 = 0.1333$, respectively.

The SS-SSIM metric was originally tested by Wang et al. [2] by measuring the quality of 29 images from the LIVE database [5]. These consisted of 24-bits/pixel RGB color images (typically 768 x 512 or similar size), compressed using JPEG or JPEG 2000. Then, to provide a quantitative evaluation of the performance of SS-SSIM, PSNR and SS-SSIM were compared to results obtained by subjective quality evaluation of the same data by human observers. The result was that SS-SSIM performs better than PSNR [2]. To test the MS-SSIM, the metric was also used to measure the quality of images from the LIVE database, and its performance was compared to subjective evaluation data by human observers, concluding that it outperforms both SS-SSIM and PSNR [3].

Digital cinema applications are based on motion pictures with significantly higher quality when compared to standard and high definition content. Furthermore, this content is only watched in a specific environment — a movie theatre. SS-SSIM and MS-SSIM metrics were not originally intended for high quality images such as those in digital cinema imagery. They take into account only the luminance component leaving out the color components of the pictures, and these metrics also overlook the motion/frame rate, which is a significant characteristic of digital cinema source materials. In addition, the original MS-SSIM parameters had been obtained from subjective parameterization tests conducted in a certain viewing environment, which was significantly different from the viewing environment in a movie theatre. Despite these constraints, SS-SSIM and MS-SSIM have a potential to be utilized as objective metrics for measuring the perceived visual quality of digital cinema applications. Thus, in this paper we study the suitability of these objective metrics for use in digital cinema applications. However, in this initial study, we also limit ourselves to the luminance component, and neglect the motion aspect.

In this paper, we do not use test images from the LIVE database as in the original study, because those images are not suitable for the digital cinema applications. Due to the specific digital cinema viewing environment, it was necessary to obtain subjective scores data from a group of human observers with a subjective quality assessment conducted in a real movie theatre.

3 Subjective Quality Assessment in Movie Theatre and Its Protocol

In the field of subjective evaluation, there are many different methodologies and rules to design a test. The test recommendations described by the ITU have been internationally

accepted as guidelines for conducting subjective assessments. Recommendation ITU-R BT500-11 [6] provides a thorough guideline for the test methods and the test conditions of subjective visual quality assessments. Important issues include characteristics of the laboratory set up, stimulus viewing sequence, and rating scale. Another important guideline relevant to this work is recommendation ITU-R BT.1686 [7]; it provides recommendations on how to perform on-screen measurements of the main projection parameters of large screen digital imagery applications, based on presentation of programs in a theatrical environment.

3.1 Laboratory Set Up

The evaluation described in this paper has been conducted at a commercial movie theatre in Trondheim, Norway. The DCI-specified cinema set up is considered to provide ideal viewing conditions. Figure A.1 shows a view of the auditorium. Table A.1 gives the specifications of the movie theatre.

The digital cinema projector used is a Sony CineAlta SRX-R220 4K projector, one of the most advanced projectors in digital cinema installations around the world (for more details on this projector see [8, 9]). Projector installation, calibration, and maintenance have been performed by Sony Corporation. Therefore, it did not seem necessary to perform any additional measurement of contrast, screen illumination intensity and uniformity, or any other measurements recommended in [7].

In order to reproduce a movie theatre experience, the assessment was conducted in the same conditions as when watching a feature film, i.e. in complete darkness. To illuminate the subject's scoring sheets during the subjective assessment without affecting the projected images perception, small low-intensity lights were attached to the clipboard used by each subject for voting.

Figure A.1: Ullman auditorium of Nova Kinosenter.

Table A.1 Ullman auditorium specifications.

DISPLAY		HALL		PROJECTOR	
Screen (H x W)	5 x 12 m	Number of Seats	440	Type	Sony SRX-R220
Projection Distance	19 m	Number of Wheelchair Seats	3		
Image Format	WS 1:1.66	Width	18.3 m		
	WS 1:1.85	Floor area	348 m^2		
	CS 1:2.35	Built Year	1994		

The physical dimensions of the screen are 5 meters by 12 meters (H x W); as a result the observation at 1H is equal to observation at 5 meters from the screen. To get a viewing distance of 1H, subjects must be seated in the front rows of the theatre. However, this location is not optimal because the point of observation is too close to the lower border of the screen, and is uncomfortable for the subjects. For this reason, a viewing distance of 2H was selected [10]. Consequently, the test subjects' seats were located in the 6th row from the screen as illustrated by the cross mark in Figure A.2. In order to ensure a centralized viewing position, only five seats located in the 6^{th} row from the screen were used by subjects during the evaluation. The location of these seats is illustrated by the cross mark in the Figure A.3 .

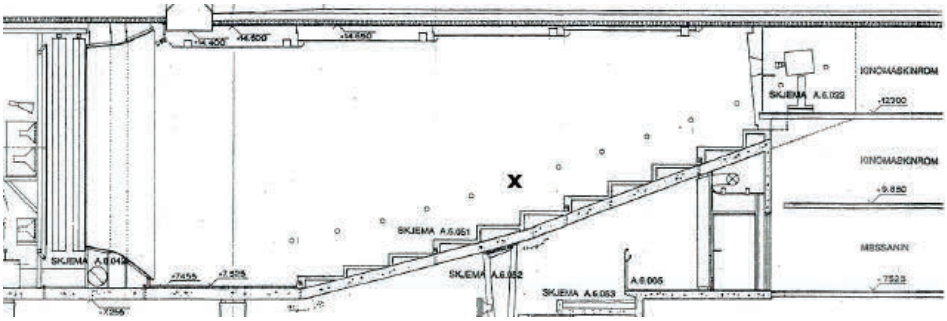


Figure A.2: Ullman auditorium of Nova Kinosenter (side view).

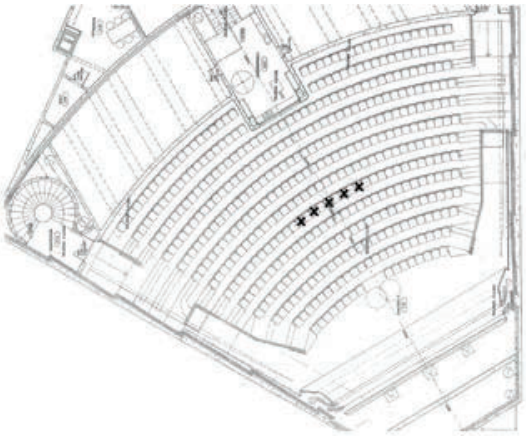


Figure A.3: Ullman auditorium of Nova Kinosenter (top view).

3.2 Test Materials

The digital cinema specification [11] provides guidance for selecting test materials for the subjective assessments' stimuli. Digital cinema is based on 2K or 4K imagery, which is a significantly higher quality in terms of larger pixel counts per image when compared to standard and high definition content, respectively. In order to comply with the DCI specifications, the stimuli used in the assessment were images taken from the DCI Standard Evaluation Material (StEM) [4]. From these, six 2K images were selected. Because we only take into account the luminance component of images in this study, the luminance component was extracted from each image resulting in six gray scale 2K images.

The subjective assessment was performed by examining a range of JPEG 2000 compression errors introduced by varying bit rates. In the design of a formal subjective test, it is recommended to maintain a low number of compression conditions in order to allow human subjects an easier completion of their evaluation task. Accordingly, 8 different conditions were applied to create 8 processed images from each source image. The selected conditions covered the whole range of quality levels, and the subjects were able to note the variation in quality from each quality level to the next. This was verified prior to the subjective quality assessment with a pilot test that involved expert viewers in order to conclude the selection of the final 8 bit rates. As a result of the pilot test, the selected bit rates were in the range of 0.01 to 0.6 bits/pixel. To create 48 processed gray scale images, 6 source images were compressed using the KAKADU software version 6.0, with the following settings [12]: codeblock size of 64x64 (default), 5 decomposition levels (default), and switched-off visual frequency weighting.

3.3 Test Methods and Conditions

There are several stimulus viewing sequences methods described in Recommendation ITU-R BT.500-11 [6]. They can be classified into two categories: single stimulus (the

subjects are presented with a sequence of test images and are asked to judge the quality of each test image) and double stimulus (the subjects are presented with the reference image and the test image before they are asked to judge the quality of the test image). The presentation method of single stimulus is sequential, whereas the presentation method of double stimulus can be sequential and simultaneous (side by side). The decision on which test method to use in a subjective assessment is crucial, because it has a high impact on the difficulty of the test subjects' task. The pilot test prior to the main subjective assessment was also conducted to compare sequential presentation and simultaneous presentation. Differentiating between levels of high quality images requires a test method that possesses a higher discriminative characteristic. Our pilot test indicated that the simultaneous (side by side) presentation had a higher discriminative characteristic than the sequential presentation order. Therefore, the subjective quality assessment uses the Simultaneous Double Stimulus test method, in which the subjects are presented with the reference image and the distorted test image displayed side by side on the screen. Figure A.4 illustrates the display format in this method.

Figure A.4: Display format of Simultaneous Double Stimulus.

The reference image is always shown on the left side of the image and the distorted image is shown on the right side. Test subjects grade the quality of the distorted image on the right hand side by comparing it to the reference image on the left.

The quality scale is the tool that the human subjects utilize to judge and to report on the quality of the tested images. One of the most popular quality scales in the subjective quality assessment research field is the 5 point quality level. Here, a 10 point quality scale was chosen, because the pilot test had shown that eight different quality levels could be clearly differentiated. Also, selecting a finer scale seemed to be advantageous due to the higher quality of test images used, in which a finer differentiating quality is suitable [10]. The test used a discrete quality grading scale, which implies that the subjects are forced to choose one of the ten values and nothing in between. The quality grading scale, which is illustrated in Figure A.5, refers to "how good the picture is".

Figure A.5: Ten point quality scale and presentation structure of the test.

The test was conducted as a single session. Each of the 48 processed images and the 6 reference images were presented for a period of 10 seconds; subjects evaluate each presented image once. Subjects then needed to vote on their questionnaire sheet before the next image was presented, and they were given 5 seconds to cast their vote. The presentation structure of the test is illustrated in Figure A.5. The total session length was 15 minutes. Prior to the main session, a training session was conducted. Subjects were informed about the procedure of the test, how to use the quality grading scale, and the meaning of the designated English term related to the distortion scale of the image. During the training session, a short pre-session was run in which 19 images were shown to illustrate the range of distortions to be expected. The order of the main session was randomized, meaning that the six images and eight processing levels were randomized completely. Four to five subjects participated at the same time, and six such rounds were needed to include all subjects (see next section). The images presentation orders for each six rounds were different.

3.4 Subjects

A proper evaluation of visual quality requires human subjects with good visual acuity and high concentration, e.g. young persons such as university students. 29 subjects (10 female, 19 male) participated in the evaluation tests performed in this work. 27 of them were university students. Some of the subjects were familiar with image processing. Their age ranged from 21 to 32 years old. All subjects reported that they had normal or corrected to normal vision.

3.5 Subjective data analysis

In this section, the Mean Opinion Score (MOS) result from the subjective image quality assessment is analyzed and will be used in the next section to evaluate the performance of the objective metrics. Before processing the resulting data, post-experiment subject screening was conducted to exclude outliers using the method described by VQEG [13]. In addition to using this method, the scores of each subject on reference images were also examined. As a result, one subject was excluded because this subject showed randomness due to scoring low for the quality of reference images. Then the consistency level for each of the remaining 28 subjects was verified by comparing his/her scores for each of the 48 processed images to the corresponding mean scores of those images over

all subjects. The consistency level was quantified using Pearson's correlation coefficient r, and if the r value for one subject was below 0.75, this subject would be excluded [13]. Here, the value of r for each subject was ≥ 0.9. Hence, data from all remaining 28 subjects was considered.

All data was then processed to obtain the Mean Opinion Score (MOS) by averaging the votes for all subjects. Figure A.6 illustrates the MOS results. In addition, the Standard Deviation and the 95% Confidence Intervals (CI) were computed (based on a normal distribution assumption). From a statistical point of view, no overlap with the 95% CI provides a strong indication of the existence of differences between adjacent MOS values. MOS values of every tested image shown with its 95% Confidence Interval are illustrated in Figure A.7.

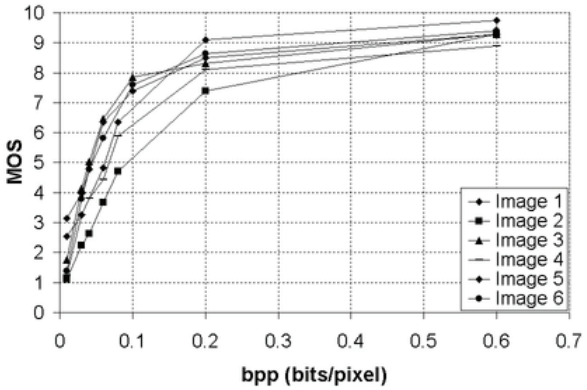


Figure A.6: MOS score vs. bit rate.

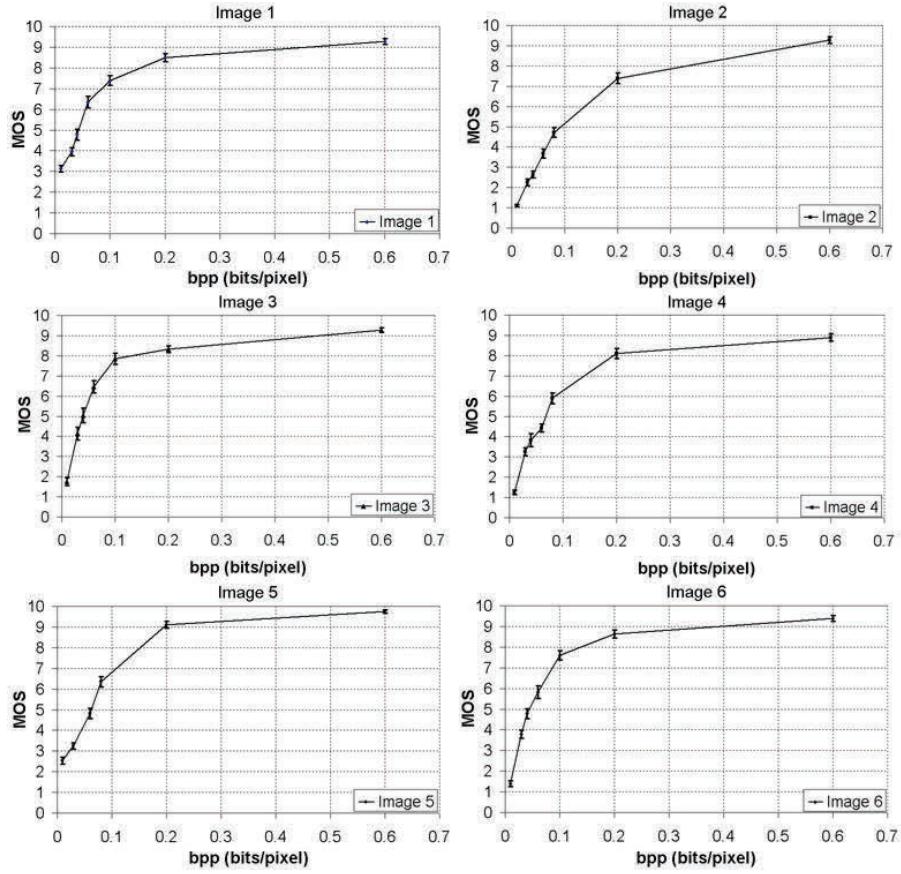


Figure A.7: MOS score of each image vs. bit rate.

The behavior of a codec is generally content dependent, and this can be observed in Figure A.6. As an example, for the lowest bit rate subjects score higher for Images 1 and 5 when compared to other images; these two images show a close up face, which typically has low spatial complexity characteristics. Furthermore, Image 2, which depicts a crowd and has high spatial complexity, tends to have the lowest score of all the images except for the highest bit rate.

4 Parameterization of MS-SSIM for Digital Cinema Applications

In this work, the test methodology based on an image synthesis approach for cross-scale calibration was also used to estimate better parameters for MS-SSIM evaluation metric

[3]. Such tests were conducted in the same DCI-specified movie theatre mentioned in the previous section, using a 12 x 5 m screen. For this purpose, a table of distorted images was synthesized as illustrated in Figure A.8. Each image in the table is associated with a specific distortion level defined by MSE and a specific scale. Each distorted image is created by randomly adding a white Gaussian noise to the original image, while constraining MSE to be fixed and restricting the distortions to occur only in the specified scale. Similar to the original method, 5 scales and 12 distortion levels were used, resulting in 60 images, as depicted in Figure A.8. Even though the images in each column have the same MSE, the visual qualities of images located in different rows (scales) are different. This provides an indication that the distortions at different scales have different significance in terms of perceived visual quality. Ten original 128x128 images with different type of content were used to create ten sets of distorted image tables.

Figure A.8: Demonstration of the table of distorted images. Images in the same column have the same MSE. Images in the same row have distortions only in one specific scale. Each subject was asked to select a set of images, one from each scale, exhibiting similar visual qualities. As an example, one subject chose the marked images.

As in the original method [3], subjective tests were conducted with 8 subjects. Each subject was shown the ten sets of test images, one set at a time. The viewing distance was fixed at 2H (10 m) similar to the subjective quality assessment. The subject was asked to compare the quality of the images across scales and select a set of images, one from each of the five scales (shown as rows in Figure A.8) exhibiting similar qualities. Marked images in Figure A.8 are one of the selected set of images perceived as having similar qualities by one subject. The positions of the selected images in each scale were recorded and averaged over all test images and all subjects. The results were then normalized and used in the calculation of exponents in Equation A.5.

The obtained parameters are $\beta 1 = \gamma 1 = 0.1587$, $\beta 2 = \gamma 2 = 0.2329$, $\beta 3 = \gamma 3 = 0.2298$, $\beta 4 = \gamma 4 = 0.2008$, and $\alpha 5 = \beta 5 = \gamma 5 = 0.1778$, respectively.

5 Results

In this section, performances of objective models (PSNR, SS-SSIM, MS-SSIM using original parameters, and MS-SSIM using the new parameters) are evaluated. This is achieved by statistically comparing the objective measurement data from each model with the subjective data. The scatter plots of raw MOS versus model predictions are illustrated in Figure A.9. A predicted MOS score resulting from the non-linear mapping function—using cubic polynomial as suggested by VQEG [14]—is also shown in each plot.

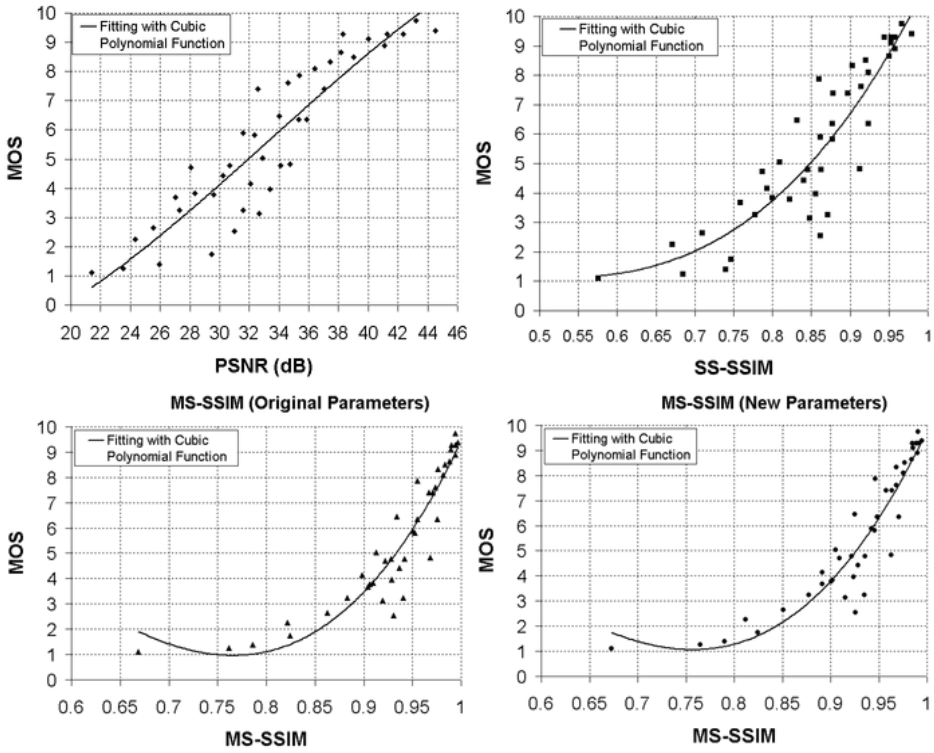


Figure A.9: Scatter plots of MOS vs. model predictions.

The linear Pearson's correlation coefficient for each metric according to the corresponding raw MOS scores is computed. The respective correlation coefficients are reported in Table A.2. Figure A.10 shows the Pearson's correlations and their associated 95% confidence intervals for each metric.

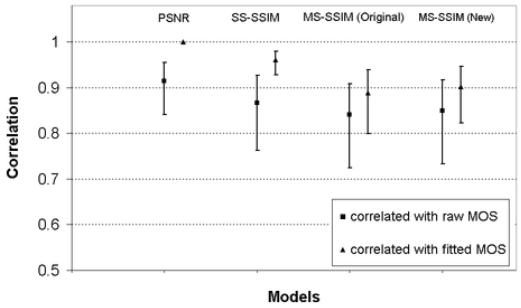


Figure A.10: Pearson's correlation coefficient.

The subjective rating data are often compressed at the ends of the rating scale. Applying a non-linear mapping step using cubic polynomial mapping function is recommended in [13, 14] before proceeding with any performance evaluation. The computed Pearson's correlation coefficient between objective measurement data and the fitted MOS value is also reported in Table A.2, and the correlations and their associated 95% confidence intervals are illustrated in Figure A.10.

To check the significance of the difference between the correlation coefficients, the statistical significance test is conducted. No significant difference between coefficients is used as H_0 hypothesis. The test uses the Fisher-z transformation. The normally distributed statistics Z_N is determined for each comparison and compared against the 95% t-Student value for the two-tail test — t(0.05)=1.96. The calculated Z_N for each correlation coefficient comparison is shown in

Table A.3. If Z_N is higher than 1.96, there is a statistically significant difference with 0.05 significance level between correlation coefficients. All calculated Z_N values based on raw MOS are lower than 1.96, which means statistically, there are no significant differences between correlation coefficients of all models. However, calculation based on fitted MOS yields values higher than 1.96, except for correlation comparison between MS-SSIM (using original parameters) and MS-SSIM (using new parameters). It means that there are statistical differences between correlation coefficient results of PSNR versus the other models, and SS-SSIM versus the other models.

Table A.2: Correlation coefficients.

Objective Model	**Pearson** (based on raw MOS)	**Pearson** (based on fitted MOS)
PSNR	0.91	0.99
SS-SSIM	0.87	0.96
MS-SSIM (original parameter)	0.84	0.88
MS-SSIM (new parameters)	0.85	0.90

Table A.3: Significance of the difference between correlation coefficients.

Models Comparison	Z_N (based on raw MOS)	Z_N (based on fitted MOS)
PSNR vs. SS-SSIM	1.03	10.66
PSNR vs. MS-SSIM	1.45	13.03
PSNR vs. MS-SSIM (new parameters)	1.32	12.74
SS-SSIM vs. MS-SSIM	1.92	2.37
SS-SSIM vs. MS-SSIM (new parameters)	1.68	2.07
MS-SSIM vs. MS-SSIM (new parameters)	0.13	0.29

6 Conclusion

Based on raw MOS data, there are no significant differences between correlation coefficients of objective metrics investigated in this paper. Hence, based on this result, there is no objective model that comes out as best performer from a statistical point of view.

However, based on fitted MOS data, the differences between PSNR and other objective metrics are significant. Hence, if the non-linear mapping function using cubic polynomial is applied first, the PSNR has the best performance because it correlates best with the subjective data. The differences between correlation of SS-SSIM with correlation of two versions of MS-SSIM (using original and new parameters) are also statistically significant. Thus, SS-SSIM performs second.

These results show that in the case of digital cinema content and environment, it seems that both SS-SSIM and MS-SSIM do not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric.

7 Acknowledgement

The authors would like to acknowledge various help, fruitful inputs, and valuable discussions from the following individuals: Vittorio Baroncini (FUB), Francesca de Simone (EPFL), Marlon Thomas M. Nielsen (NTNU), and Zhou Wang (University of Waterloo). Subjective tests carried out in this work were possible thanks to Trondheim Kino AS by putting at disposition the movie theatre Nova Kinosenter.

8 Reference

[1] SMPTE. (2008). *Society of Motion Picture and Television Engineers*. Available: http://www.smpte.org/home/

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, p. 13, April 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," in *IEEE Asilomar Conference on Signals, Systems and Computers*, Asilomar, 2003.

[4] DCI. (2008). *DCI Digital Cinema Initiatives*. Available: http://www.dcimovies.com

[5] H. R. Sheiks, Z. Wang, A. C. Bovik, and L. K. Cormack. *Image and video quality assessment research at LIVE*. Available: http://live.ece.utexas.edu/research/quality/

[6] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva2002.

[7] ITU-R, "Methods of measurement of image presentation parameters for LSDI programme presentation in a theatrical environment," Geneva2004.

[8] SONY, "4K Digital Cinema Projectors SRX-R220/SRX-R210 Media Blok LMT-100 Screen Management System LSM-100," 2007.

[9] SONY. (2008). *SRX-R220*. Available: http://www.sony.co.uk/biz/product/4k-digital-cinema/srx-r220/overview

[10] V. Baroncini, "Title," unpublished|.

[11] D. C. Initiatives, "Digital Cinema System Specification version 1.2," March 2008.

[12] D. Taubman, "Kakadu Software," 6.0 ed, 2008.

[13] VQEG, "Multimedia Group Test Plan Version 1.21," March 2008 2008.

[14] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase 1," September 2008 2008.

Errata

Page 72; Figure A.7.

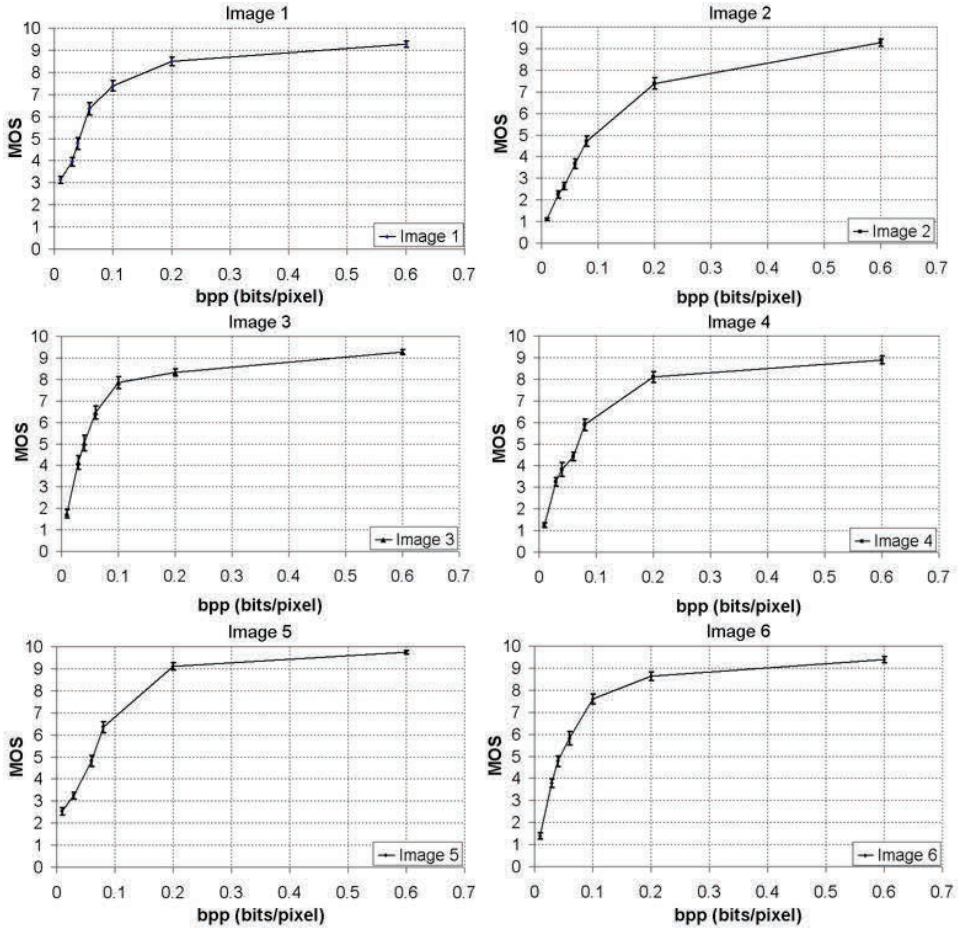


Figure A.7: MOS of each image vs. bit rate

Figure A.7 illustrates the obtained Mean Opinion Score (MOS) by averaging the votes of all subjects along with its 95% Confidence Interval. However, the former figure shows an incorrect range of 95% Confidence Interval. The range illustrated half of the intended 95% Confidence Interval. The model illustrated by Equation A.6 represented 95% Confidence Interval shown in the Figure A.7.

$$\bar{x} \pm \left[\frac{1.96(\frac{\sigma}{\sqrt{n}})}{2} \right] \tag{A.6}$$

The corrected illustration of the calculated 95% Confidence Interval is shown in figure A2.1. The 95% Confidence Interval is based on the Equation A.7.

$$\bar{x} \pm A(\frac{\sigma}{\sqrt{n}}), \tag{A.7}$$

with A is the critical point based on two sided t-distribution with 27 degrees of freedom. The value of A is $A = t_{0.05,27} = 2.05$.

The correction is illustrated in the figure below.

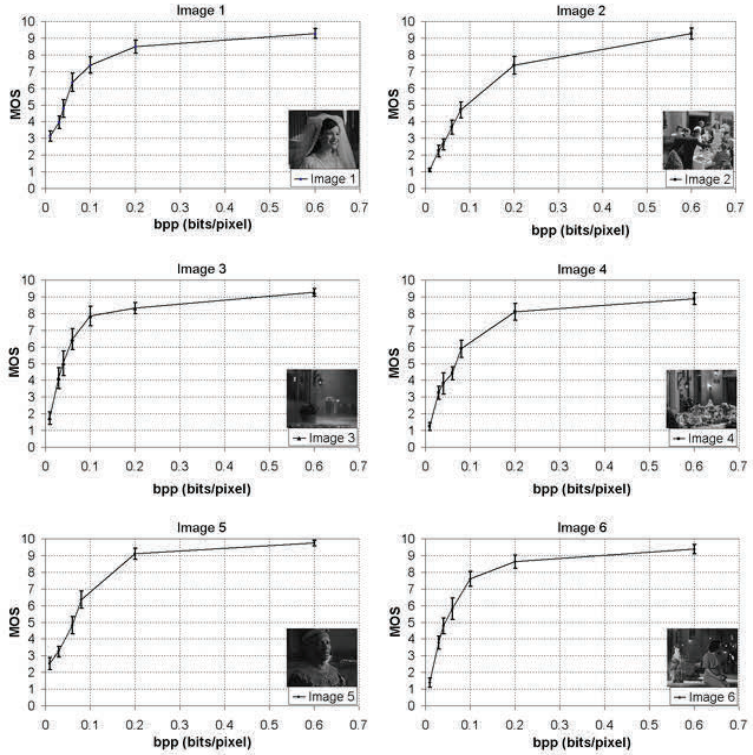


Figure A.7: MOS of each image vs. bit rate.

Page 74; Figure A.9

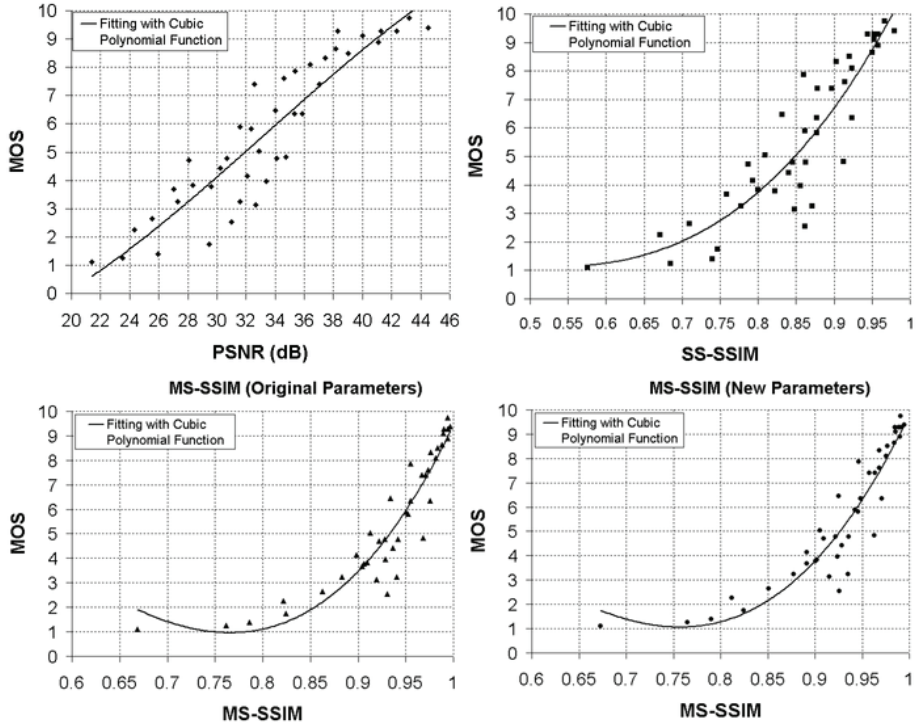


Figure A.9: Scatter plots of MOS vs. model prediction

Figure A.9 that demonstrates scatter plots of MOS vs. model prediction have erroneous representation of fitting with cubic polynomial function due to erroneous calculation during subjective data processing. The incorrect calculation of non linear mapping especially yielded the extreme value of correlation of one objective model (PSNR). Based on VQEG recommendation, non linear mapping that has been found to perform well empirically is cubic polynomial function as shown in Equation A.8. The weightings a, b and c and the constant d are obtained by fitting the function to the data

$$MOSp = ax^3 + bx^2 + cx + d \tag{A.8}$$

The correction is illustrated in the figure below.

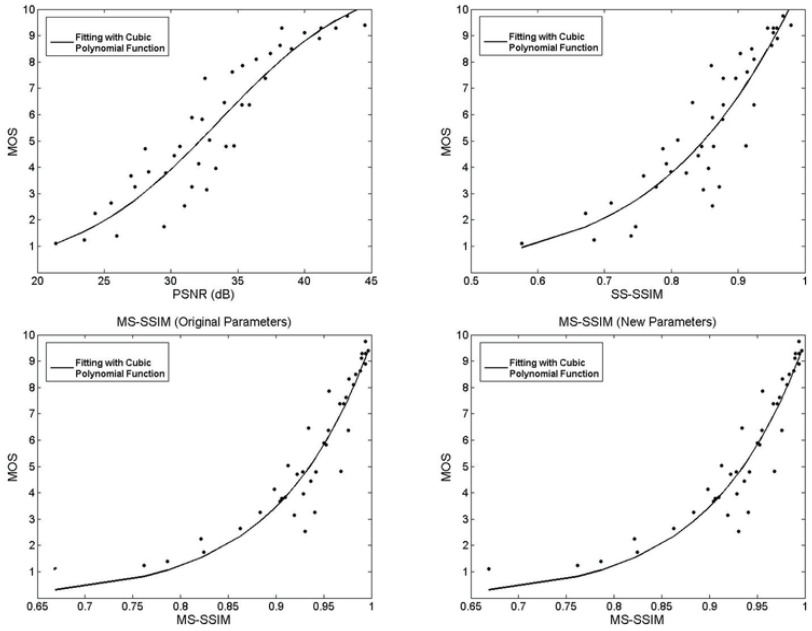


Figure A.9: Scatter plots of MOS vs. model prediction.

Page 75; Table A.2.

Table A.2**:** Correlation coefficients.

Objective Model	**Pearson** (based on raw MOS)	**Pearson** (based on fitted MOS)
PSNR	0.91	0.99
SS-SSIM	0.87	0.96
MS-SSIM (original parameter)	0.84	0.88
MS-SSIM (new parameters)	0.85	0.90

Table A.2 shows the Pearson's correlation coefficient calculated from the erroneous scores. Using the corrected scores, the new correlation coefficient of each model is recalculated. The calculation of the Pearson's correlation coefficient is based on Equation A.9.

$$R = \frac{\sum_{i=1}^{N} (Xi - \overline{X}) * (Yi - \overline{Y})}{\sqrt{\sum (Xi - \overline{X})^2} * \sqrt{\sum (Yi - \overline{Y})^2}} \qquad \text{(A.9)}$$

The accurate calculated correlation coefficients are shown Table below.

Table A.2: Correlation coefficients.

Objective Model	**Pearson** (based on raw data)	**Pearson** (based on fitted data)
PSNR	0.91	0.92
SS-SSIM	0.87	0.9
MS-SSIM (original parameter)	0.84	0.95
MS-SSIM (new parameters)	0.85	0.94

Page 75; Figure A.10.

Original figure:

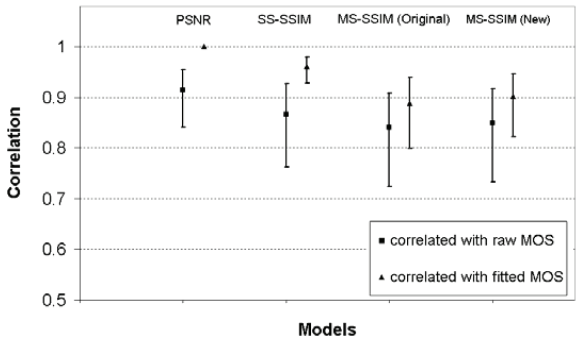


Figure A.10: Pearson's correlation coefficient

Figure A.10 illustrated the correlation coefficient based on the calculated values from the Table A.2. The correlation coefficients are shown with the 95% Confidence Interval. The new illustration of accurate Pearson's correlation coefficients with the 95 % Confidence Interval is shown on Figure below

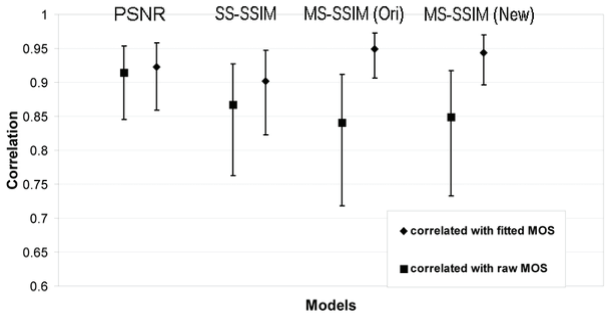


Figure A.10: Pearson's correlation coefficient.

Page 76; Table A.3.

Table A.3: Significance of differences between correlation coefficients

Models Comparison	**Z_N** (based on raw MOS)	**Z_N** (based on fitted MOS)
PSNR vs. SS-SSIM	1.03	10.66
PSNR vs. MS-SSIM	1.45	13.03
PSNR vs. MS-SSIM (new parameters)	1.32	12.74
SS-SSIM vs. MS-SSIM	1.92	2.37
SS-SSIM vs. MS-SSIM (new parameters)	1.68	2.07
MS-SSIM vs. MS-SSIM (new parameters)	0.13	0.29

Table A.3 shows the significant differences between correlation coefficients illustrated by Figure A.10. Based on the recalculated correlation coefficients, we then also need to recalculate the significant differences between new correlation coefficients. The new values are shown in the table below.

Table A.3: Significance of differences between correlation coefficients

Models Comparison	**Z_N** (based on raw MOS)	**Z_N** (based on fitted MOS)
PSNR vs. SS-SSIM	1.03	0.5
PSNR vs. MS-SSIM	1.45	0.95
PSNR vs. MS-SSIM (new parameters)	1.32	0.7
SS-SSIM vs. MS-SSIM	0.43	1.5
SS-SSIM vs. MS-SSIM (new parameters)	0.3	1.25
MS-SSIM vs. MS-SSIM (new parameters)	0.13	0.24

Page 75; 5. Result

Original text:
"All calculated Z_N values based on raw MOS are lower than 1.96, which means statistically, there are no significant differences between correlation coefficients of all models. However, calculation based on fitted MOS yields values higher than 1.96, except for correlation comparison between MS-SSIM (using original parameters) and MS-SSIM (using new parameters). It means that there are statistical differences between

correlation coefficient results of PSNR versus the other models, and SS-SSIM versus the other models."

Correction:
"All calculated Z_N values are lower than 1.96, which means statistically, there are no significant differences between correlation coefficients of all models."

Page 76; 6. Conclusion

Original text:
"Based on raw MOS data, there are no significant differences between correlation coefficients of objective metrics investigated in this paper. Hence, based on this result, there is no objective model that comes out as best performer from a statistical point of view.
However, based on fitted MOS data, the differences between PSNR and other objective metrics are significant. Hence, if the non-linear mapping function using cubic polynomial is applied first, the PSNR has the best performance because it correlates best with the subjective data. The differences between correlation of SS-SSIM with correlation of two versions of MS-SSIM (using original and new parameters) are also statistically significant. Thus, SS-SSIM performs second.
These results show that in the case of digital cinema content and environment, it seems that both SS-SSIM and MS-SSIM do not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric."

Correction:
"Based on collected MOS data, there are no significant differences between correlation coefficients of objective metrics investigated in this paper. Hence, based on this result, there is no objective model that comes out as best performer from a statistical point of view.
These results show that in the case of digital cinema content and environment, it seems that both SS-SSIM and MS-SSIM do not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric.

1 Introduction

The motion picture industry is one of the many players in the media industry. Both broadcasting and mobile media have successfully completed their transition to fully digital services, while the motion picture industry is currently in the process of forming standards for digitization of its complete value chain. These specifications and standards are the basis for a large scale implementation of digital cinema as the latest and final analogue media to go digital. The digitization is specified by the Digital Cinema Initiative (DCI) and is currently under standardization by SMPTE [1]. One of the key issues for a successful roll out of digital cinema in the market is in the service assurance of the quality it offers. The ultimate measure of a service is how an end-user perceives its performance. Hence, the best way of measuring perceived quality is to rely on human subjects assessment, in a controlled environment. This is referred to as subjective quality assessment.

Performing subjective assessments is time consuming, expensive, and complex. Furthermore, it does not lend itself to real-time environments. As an alternative, objective measurement methods (objective metrics) have been developed to predict the perceived quality of human subjects. Among objective metrics proposed to estimate perceived quality, Wang and Bovik introduced the structural similarity quality paradigm (SSIM) based on the assumption that the human visual system is highly adapted for extraction of structural information from a scene [2]. They argue that a measure of structural similarity can provide a good approximation of perceived quality. In their experiments, SSIM has shown a good correlation with perceived quality, outperforming traditional metrics such as peak-signal-to-noise ratio (PSNR). Multi-scale structural similarity (MS-SSIM) is proposed to supply more flexibility when compared to the single-scale method (SS-SSIM), by taking into account variations in viewing conditions [3]. These metrics have been used to measure perceived image quality in digital cinema.

In this paper, we report the results of a study to assess the suitability of SS-SSIM and MS-SSIM to measure the perceived quality of images from DCI Standard Evaluation Material (StEM) [4]. In addition to application of these metrics using their original parameters, new parameters for MS-SSIM were obtained by taking into account the digital cinema viewing conditions, and used in this study. To validate the results of these metrics, we investigated the correlation between the objective metrics and the ground truth, i.e. how human subjects perceive the same content in terms of quality. To this end, a subjective quality assessment was carried out in a DCI-specified cinema in Trondheim, Norway.

The paper is structured as follows. First, an overview of single scale structural similarity (SS-SSIM) and multi scale structural similarity (MS-SSIM) is given in Section 2. The subjective quality assessment in the DCI-specified cinema environment and its results are provided in Section 3. The parameterization of MS-SSIM for digital cinema content and environment is presented in Section 4. Next, the test results are discussed in Section 5. Finally, we draw some conclusions in Section 6.

2 Overview of Structural Similarity Measures

Natural image signals are highly structured. Their pixels exhibit strong dependencies, which carry important information about the structure of the objects in the visual scene [2]. The human visual system is highly adapted to extract structural information. It is therefore assumed that the measurement of structural information changes provides a good estimation of the perceived image distortion [2]. Suppose x and y were two image signals; if one of the signals had perfect quality, then the similarity measure could be utilized to measure quantitatively the quality of the second signal.

Structural information in an image is defined as attributes that represent the structure of objects in the scene, which are independent of the illumination. Accordingly, the information of structure is independent of the average luminance and contrast. The quality assessment uses local luminance and contrast because luminance and contrast can vary across a scene. The similarity measurement system is based on three comparisons: luminance, contrast, and structure. The luminance comparison function l(x,y) is estimated as a comparison of the mean intensity of two discrete signals, x and y, which is defined by

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{A.1}$$

where the constant C_1= 6.50 [2] is included to avoid instability when μx2+ μy2 is very close to zero. The contrast comparison function c(x,y) takes a similar form, based on the standard deviation of the two signals, x and y,

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \tag{A.2}$$

Here again, the constant C_2= 58.52 [2] is included to avoid instability when $\sigma_x^2+\sigma_y^2$ is very close to zero. The third comparison — the structure comparison function s(x,y) — is defined as follows:

$$s(\mathbf{x},\mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \tag{A.3}$$

To avoid instability when σxσy is very close to zero, a constant C_3= 29.26 [2] is incorporated. The general form of the Structural SIMilarity (SSIM) index between signals x and y becomes:

$$SS - SSIM(\mathbf{x},\mathbf{y}) = \left[l(\mathbf{x},\mathbf{y})\right]^{\alpha} \cdot \left[c(\mathbf{x},\mathbf{y})\right]^{\beta} \cdot \left[s(\mathbf{x},\mathbf{y})\right]^{\gamma} \tag{A.4}$$

where α > 0, β > 0, γ > 0 are parameters used to adjust the relative importance of the three components. In the Single-Scale Structural SIMilarity (SS-SSIM) approach, the parameter values are set to α = β = γ = 1.

The perceivability of image details depends on the sampling density of the image signal and the distance of the image plane from the observer. When these factors vary, the subjective evaluation of a given image varies too. The single scale method does not incorporate such factors. For this reason, a multi scale variant of structural similarity has been developed to incorporate image details at different resolutions [3]. Taking the reference and the distorted image signals as input, the method iteratively applies a low-pass filter and down-samples the filtered image by a factor of 2. For example, at the $_{\text{j-th}}$ scale, the reference and the distorted image signals are low-pass filtered and down-sampled by a factor of 2 j-1 times.

The overall computation is acquired by combining the measurement at different scales using

$$MS-SSIM(\mathbf{x},\mathbf{y}) = \left[l_M(\mathbf{x},\mathbf{y})\right]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(\mathbf{x},\mathbf{y})\right]^{\beta_j} \left[s_j(\mathbf{x},\mathbf{y})\right]^{\gamma_j} \quad \text{(A.5)}$$

in which the original image is indexed as scale 1. This definition also comprises the single scale measurement as the special case M = 1.

The exponents α_M, β_j, and γ_j are used to adjust the relative importance of the three components, and M is set to 5. Based on a subjective parameterization test [3], the resulting parameters are $\beta_1 = \gamma_1 = 0.0448$, $\beta_2 = \gamma_2 = 0.2856$, $\beta_3 = \gamma_3 = 0.3001$, $\beta_4 = \gamma_4 = 0.2363$, and $\alpha_5 = \beta_5 = \gamma_5 = 0.1333$, respectively.

The SS-SSIM metric was originally tested by Wang et al. [2] by measuring the quality of 29 images from the LIVE database [5]. These consisted of 24-bits/pixel RGB color images (typically 768 x 512 or similar size), compressed using JPEG or JPEG 2000. Then, to provide a quantitative evaluation of the performance of SS-SSIM, PSNR and MS-SSIM were compared to results obtained by subjective quality evaluation of the same data by human observers. The result was that SS-SSIM performs better than PSNR [2]. To test the MS-SSIM, the metric was also used to measure the quality of images from the LIVE database, and its performance was compared to subjective evaluation data by human observers, concluding that it outperforms both SS-SSIM and PSNR.

Digital cinema applications are based on motion pictures with significantly higher quality when compared to standard and high definition content. Furthermore, this content is only watched in a specific environment — a movie theatre. SS-SSIM and MS-SSIM metrics were not originally intended for high quality images such as those in digital cinema imagery. They take into account only the luminance component leaving out the color components of the pictures, and these metrics also overlook the motion/frame rate, which is a significant characteristic of digital cinema source materials. In addition, the original MS-SSIM parameters had been obtained from subjective parameterization tests conducted in a certain viewing environment, which was significantly different from the viewing environment in a movie theatre. Despite these constraints, SS-SSIM and MS-SSIM have a potential to be utilized as objective metrics for measuring the perceived visual quality of digital cinema applications. Thus, in this paper we study the suitability of these objective metrics for use in digital cinema applications. However, in this initial study, we also limit ourselves to the luminance component, and neglect the motion aspect.

In this paper, we do not use test images from the LIVE database as in the original study, because those images are not suitable for the digital cinema applications. Due to the specific digital cinema viewing environment, it was necessary to obtain subjective scores data from a group of human observers with a subjective quality assessment conducted in a real movie theatre.

3 Subjective Quality Assessment in Movie Theatre and Its Protocol

In the field of subjective evaluation, there are many different methodologies and rules to design a test. The test recommendations described by the ITU have been internationally

accepted as guidelines for conducting subjective assessments. Recommendation ITU-R BT 500-11 [6] provides a thorough guideline for the test methods and the test conditions of subjective visual quality assessments. Important issues include characteristics of the laboratory set up, stimulus viewing sequence, and rating scale. Another important guideline relevant to this work is recommendation ITU-R BT.1686 [7]; it provides recommendations on how to perform on-screen measurements of the main projection parameters of large screen digital imagery applications, based on presentation of programs in a theatrical environment.

3.1 Laboratory Set Up

The evaluation described in this paper has been conducted at a commercial digital cinema Nova Kinosenter in Trondheim, Norway. The DCI-specified cinema set up is considered to provide ideal viewing conditions. (Figure A.1) shows a view of the auditorium. Table A.4 gives the specifications of the movie theatre.

Figure A.1: Ullman auditorium of Nova Kinosenter.

The digital cinema projector used is a Sony CineAlta SRX-R220 4K projector, one of the most advanced projectors in digital cinema installations around the world (for more details on this projector see [8, 9]). Projector installation, calibration, and maintenance have been performed by Trondheim Kino AS. Therefore, it did not seem necessary to perform any additional measurement of contrast, screen illumination intensity and uniformity, or any other measurements recommended in [7].

In order to reproduce a movie theatre experience, the assessment was conducted in the same conditions as when watching a feature film, i.e. in complete darkness. To illuminate the subject's scoring sheets during the subjective assessment without

affecting the projected images perception, small low-intensity lights were attached to the clipboard used by each subject for voting.

Table A.4: Ullman auditorium specifications.

DISPLAY		HALL		PROJECTOR	
Screen (H x W)	5 x 12 m	Number of Seats	440	Type	Sony SRX-R220
Projection Distance	19 m	Number of Wheelchair Seats	3		
Image Format	WS 1:1.66	Width	18.3 m		
	WS 1:1.85	Floor area	348 m^2		
	CS 1:2.35	Built Year	1994		

The physical dimensions of the screen are 5 meters by 12 meters (H x W); as a result the observation at 1H is equal to observation at 5 meters from the screen. To get a viewing distance of 1H, subjects must be seated in the front rows of the theatre. However, this location is not optimal because the point of observation is too close to the lower border of the screen, and is uncomfortable for the subjects. For this reason, a viewing distance of 2H was selected [10]. Consequently, the test subjects' seats were located in the 6th row from the screen as illustrated by the cross mark in Figure A.2. In order to ensure a centralized viewing position, only five seats located in the 6^{th} row from the screen were used by subjects during the evaluation. The location of these seats is illustrated by the cross mark in the Figure A.3.

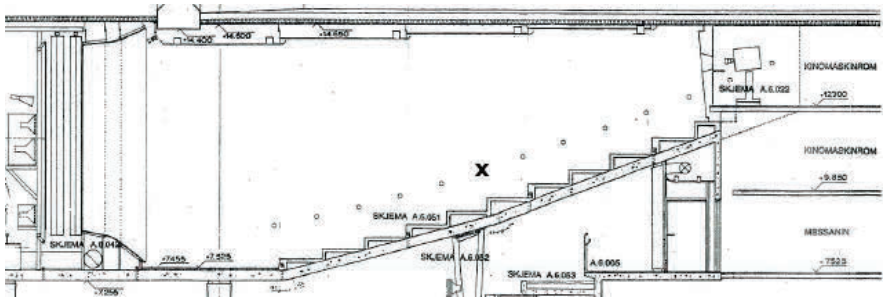


Figure A.2: Ullman auditorium of Nova Kinosenter (side view).

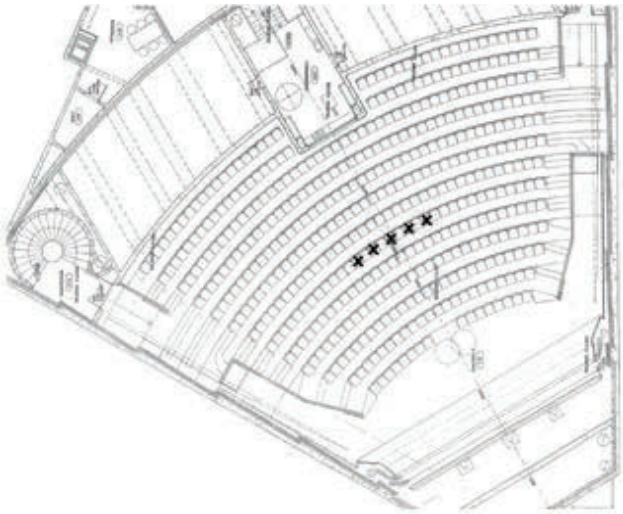


Figure A.3: Ullman auditorium of Nova Kinosenter (top view).

3.2 Test Materials

The digital cinema specification [11] provides guidance for selecting test materials for the subjective assessments' stimuli. Digital cinema is based on 2K or 4K imagery, which is a significantly higher quality in terms of larger pixel counts per image when compared to standard and high definition content, respectively. In order to comply with the DCI specifications, the stimuli used in the assessment were images taken from the DCI Standard Evaluation Material (StEM) [4]. From these, six 2K images were selected. Because we only take into account the luminance component of images in this study, the luminance component was extracted from each image resulting in six gray scale 2K images.

The subjective assessment was performed by examining a range of JPEG 2000 compression errors introduced by varying bit rates. In the design of a formal subjective test, it is recommended to maintain a low number of compression conditions in order to allow human subjects an easier completion of their evaluation task. Accordingly, 8 different conditions were applied to create 8 processed images from each source image. The selected conditions covered the whole range of quality levels, and the subjects were able to note the variation in quality from each quality level to the next. This was verified prior to the subjective quality assessment with a pilot test that involved expert viewers in order to conclude the selection of the final 8 bit rates. As a result of the pilot test, the selected bit rates were in the range of 0.01 to 0.6 bits/pixel. To create 48 processed gray scale images, 6 source images were compressed using the KAKADU software version 6.0, with the following settings [12]: codeblock size of 64x64 (default), 5 decomposition levels (default), and switched-off visual frequency weighting.

3.3 Test Methods and Conditions

There are several stimuli viewing sequence methods described in Recommendation ITU-R BT.500-11 [6]. They can be classified into two categories: single stimulus (the subjects are presented with a sequence of test images and are asked to judge the quality of each test image) and double stimulus (the subjects are presented with the reference image and the test image before they are asked to judge the quality of the test image). The presentation method of single stimulus is sequential, whereas the presentation method of double stimulus can be sequential and simultaneous (side by side). The decision on which test method to use in a subjective assessment is crucial, because it has a high impact on the difficulty of the test subjects' task. The pilot test prior to the main subjective assessment was also conducted to compare sequential presentation and simultaneous presentation. Differentiating between levels of high quality images requires a test method that possesses a higher discriminative characteristic. Our pilot test indicated that the simultaneous (side by side) presentation had a higher discriminative characteristic than the sequential presentation order. Therefore, the subjective quality assessment uses the Simultaneous Double Stimulus test method, in which the subjects are presented with the reference image and the distorted test image displayed side by side on the screen. Figure A.4 illustrates the display format in this method.

Figure A.4: Display format of Simultaneous Double Stimulus.

The reference image is always shown on the left side of the image and the distorted image is shown on the right side. Test subjects grade the quality of the distorted image on the right hand side by comparing it to the reference image on the left.

The quality scale is the tool that the human subjects utilize to judge and to report on the quality of the tested images. One of the most popular quality scales in the subjective quality assessment research field is the 5 point quality level. Here, a 10 point quality scale was chosen, because the pilot test had shown that eight different quality levels could be clearly differentiated. Also, selecting a finer scale seemed to be advantageous due to the higher quality of test images used, in which a finer differentiating quality is suitable [10]. The test used a discrete quality grading scale,

which implies that the subjects are forced to choose one of the ten values and nothing in between. The quality grading scale, which is illustrated in Figure A.5, refers to "how good the picture is".

Figure A.5: Ten point quality scale and presentation structure of the test.

The test was conducted as a single session. Each of the 48 processed images and the 6 reference images were presented for a period of 10 seconds; subjects evaluated each presented image once. Subjects then needed to vote on their questionnaire sheet before the next image was presented, and they were given 5 seconds to cast their vote. The presentation structure of the test is illustrated in Figure A.5. The total session length was 15 minutes. Prior to the main session, a training session was conducted. Subjects were informed about the procedure of the test, how to use the quality grading scale, and the meaning of the designated English term related to the distortion scale of the image. During the training session, a short pre-session was run in which 19 images were shown to illustrate the range of distortions to be expected. The order of the main session was randomized, meaning that the six images and eight processing levels were randomized completely. Four to five subjects participated at the same time, and six such rounds were needed to include all subjects (see next section). The images presentation orders for each six rounds were different.

3.4 Subjects

A proper evaluation of visual quality requires human subjects with good visual acuity and high concentration, e.g. young persons such as university students. 29 subjects (10 female, 19 male) participated in the evaluation tests performed in this work. 27 of them were university students. Some of the subjects were familiar with image processing. Their age ranged from 21 to 32 years old. All subjects reported that they had normal or corrected to normal vision.

3.5 Subjective data analysis

In this section, the Mean Opinion Score (MOS) result from the subjective image quality assessment is analyzed and will be used in the next section to evaluate the performance of the objective metrics. Before processing the resulting data, post-experiment subject

screening was conducted to exclude outliers using the method described by VQEG [13]. In addition to using this method, the scores of each subject on reference images were also examined. As a result, one subject was excluded because this subject showed randomness due to scoring low for the quality of reference images. Then the consistency level for each of the remaining 28 subjects was verified by comparing his/her scores for each of the 48 processed images to the corresponding mean scores of those images over all subjects. The consistency level was quantified using Pearson's correlation coefficient r, and if the r value for one subject was below 0.75, this subject would be excluded [13]. Here, the value of r for each subject was ≥ 0.9. Hence, data from all remaining 28 subjects was considered.

All data was then processed to obtain the Mean Opinion Score (MOS) by averaging the votes for all subjects. Figure A.6 illustrates the MOS results. In addition, the Standard Deviation and the 95% Confidence Intervals (CI) were computed (based on a normal distribution assumption). From a statistical point of view, no overlap with the 95% CI provides a strong indication of the existence of differences between adjacent MOS values. MOS values of every tested image shown with its 95% Confidence Interval are illustrated in Figure A.7.

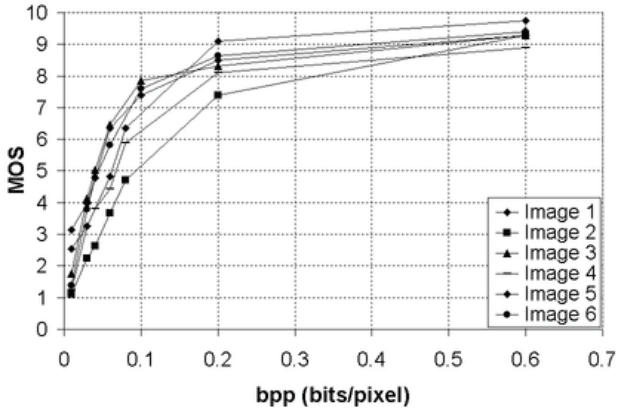


Figure A.6: MOS score vs. bit rate.

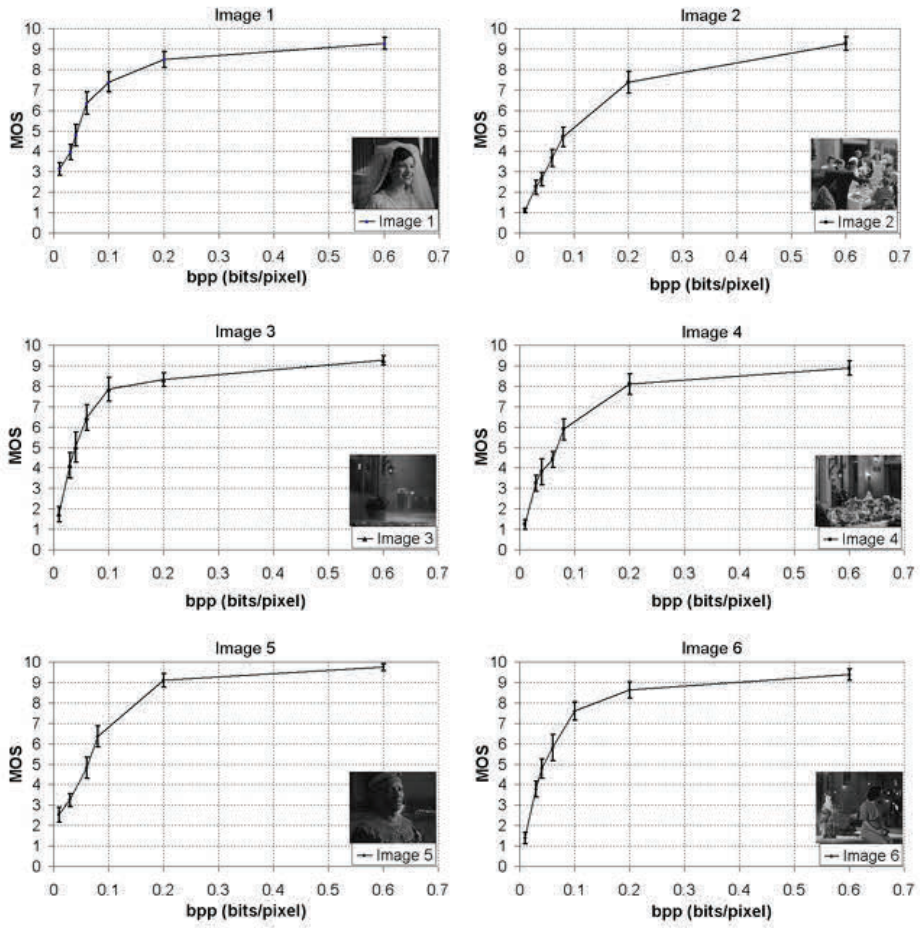


Figure A.7: MOS score of each image vs. bit rate.

The behavior of a codec is generally content dependent, and this can be observed in Figure A.6. As an example, for the lowest bit rate subjects score higher for Images 1 and 5 when compared to other images; these two images show a close up face, which typically has low spatial complexity characteristics. Furthermore, Image 2, which depicts a crowd and has high spatial complexity, tends to have the lowest score of all the images except for the highest bit rate.

4 Parameterization of MS-SSIM for Digital Cinema Application

In this work, the test methodology based on an image synthesis approach for cross-scale calibration was also used to estimate better parameters for MS-SSIM evaluation metric [3]. Such tests were conducted in the same DCI-specified movie theatre mentioned in the previous section, using a 12 x 5 m screen. For this purpose, a table of distorted images was synthesized as illustrated in Figure A.8. Each image in the table is associated with a specific distortion level defined by MSE and a specific scale. Each distorted image is created by randomly adding a white Gaussian noise to the original image, while constraining MSE to be fixed and restricting the distortions to occur only in the specified scale. Similar to the original method, 5 scales and 12 distortion levels were used, resulting in 60 images, as depicted in Figure A.8. Even though the images in each column have the same MSE, the visual qualities of images located in different rows (scales) are different. This provides an indication that the distortions at different scales have different significance in terms of perceived visual quality. Ten original 128x128 images with different type of content were used to create ten sets of distorted image tables.

Figure A.8: Demonstration of the table of distorted images. Images in the same column have the same MSE. Images in the same row have distortions only in one specific scale. Each subject was asked to select a set of images, one from each scale, exhibiting similar visual qualities. As an example, one subject chose the marked images.

As in the original method [3], subjective tests were conducted with 8 subjects. Each subject was shown the ten sets of test images, one set at a time. The viewing distance was fixed at 2H (10 m) similar to the subjective quality assessment. The subject was asked to compare the quality of the images across scales and select a set of images, one from each of the five scales (shown as rows in Figure A.8) exhibiting similar qualities. Marked images in Figure A.8 are one of the selected set of images perceived as having similar qualities by one subject. The positions of the selected

images in each scale were recorded and averaged over all test images and all subjects. The results were then normalized and used in the calculation of exponents in Equation A.5.

The obtained parameters are $\beta_1 = \gamma_1 = 0.1587$, $\beta_2 = \gamma_2 = 0.2329$, $\beta_3 = \gamma_3 = 0.2298$, $\beta_4 = \gamma_4 = 0.2008$, and $\alpha_5 = \beta_5 = \gamma_5 = 0.1778$, respectively.

5 Results

In this section, performances of objective models (PSNR, SS-SSIM, MS-SSIM using original parameters, and MS-SSIM using the new parameters) are evaluated. This is achieved by statistically comparing the objective measurement data from each model with the subjective data. The scatter plots of raw MOS versus model predictions are illustrated in Figure A.9. A predicted MOS score resulting from the non-linear mapping function—using cubic polynomial as suggested by VQEG [14]—is also shown in each plot.

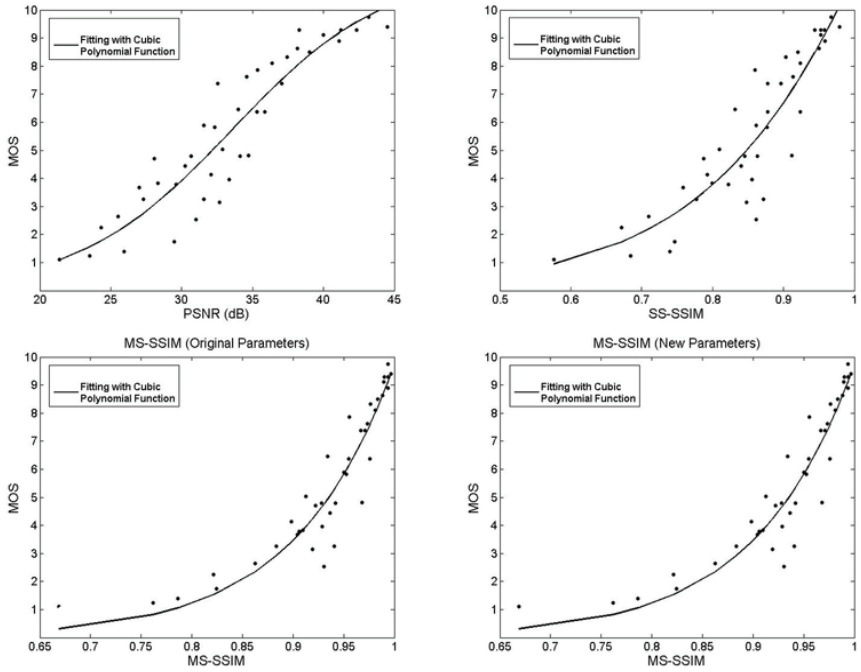


Figure A.9: Scatter plots of MOS vs. model predictions.

The linear Pearson's correlation coefficient for each metric according to the corresponding MOS scores is computed. The respective correlation coefficients are reported in Table A.5. Figure A.10 shows the Pearson's correlations and their associated 95% confidence intervals for each metric.

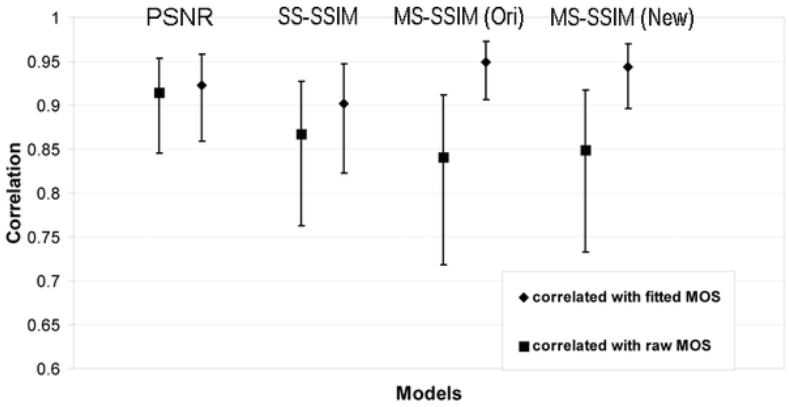


Figure A.10: Pearson's correlation coefficient.

The subjective rating data are often compressed at the ends of the rating scale. Applying a non-linear mapping step using cubic polynomial mapping function is recommended in [13, 14] before proceeding with any performance evaluation. The computed Pearson's correlation coefficient between objective measurement data and the fitted MOS value is also reported in Table A.5, and the correlations and their associated 95% confidence intervals are illustrated in Figure A.10.

To check the significance of the difference between the correlation coefficients, the statistical significance test is conducted. No significant difference between coefficients is used as H_0 hypothesis. The test uses the Fisher-z transformation. The normally distributed statistics Z_N is determined for each comparison and compared against the 95% t-Student value for the two-tail test — t(0.05)=1.96. The calculated Z_N for each correlation coefficient comparison is shown in Table A.3. If Z_N is higher than 1.96, there is a statistically significant difference with 0.05 significance level between correlation coefficients. All calculated Z_N values are lower than 1.96, which means statistically, there are no significant differences between correlation coefficients of all models.

Table A.5: Correlation coefficients.

Objective Model	**Pearson** (based on raw MOS)	**Pearson** (based on fitted MOS)
PSNR	0.91	0.92
SS-SSIM	0.87	0.9
MS-SSIM (original parameter)	0.84	0.95
MS-SSIM (new parameters)	0.85	0.94

Table A.6: Significance of the difference between correlation coefficients.

Models Comparison	Z_N (based on raw MOS)	Z_N (based on fitted MOS)
PSNR vs. SS-SSIM	1.03	0.5
PSNR vs. MS-SSIM	1.45	0.95
PSNR vs. MS-SSIM (new parameters)	1.32	0.7
SS-SSIM vs. MS-SSIM	0.43	1.5
SS-SSIM vs. MS-SSIM (new parameters)	0.3	1.25
MS-SSIM vs. MS-SSIM (new parameters)	0.13	0.24

6 Conclusion

Based on collected MOS data, there are no significant differences between correlation coefficients of objective metrics investigated in this paper. Hence, based on this result, there is no objective model that comes out as best performer from a statistical point of view.

These results show that in the case of digital cinema content and environment, it seems that both SS-SSIM and MS-SSIM do not exhibit the same type of performance that has been reported in the literature, when compared to PSNR metric.

7 Acknowledgements

The authors would like to acknowledge various help, fruitful inputs, and valuable discussions from the following individuals: Vittorio Baroncini (FUB), Francesca de Simone (EPFL), Marlon Thomas M. Nielsen (NTNU), and Zhou Wang (University of Waterloo). Subjective tests carried out in this work were possible thanks to Trondheim Kino AS by putting at disposition the movie theatre Nova Kinosenter.

8 Reference

[1] SMPTE. (2008). *Society of Motion Picture and Television Engineers*. Available: http://www.smpte.org/home/

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, p. 13, April 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," in *IEEE Asilomar Conference on Signals, Systems and Computers*, Asilomar, 2003.

[4] DCI. (2008). *DCI Digital Cinema Initiatives*. Available: http://www.dcimovies.com

[5] H. R. Sheiks, Z. Wang, A. C. Bovik, and L. K. Cormack. *Image and video quality assessment research at LIVE*. Available: http://live.ece.utexas.edu/research/quality/

[6] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva2002.

[7] ITU-R, "Methods of measurement of image presentation parameters for LSDI programme presentation in a theatrical environment," Geneva2004.

[8] SONY, "4K Digital Cinema Projectors SRX-R220/SRX-R210 Media Blok LMT-100 Screen Management System LSM-100," 2007.

[9] SONY. (2008). *SRX-R220*. Available: http://www.sony.co.uk/biz/product/4k-digital-cinema/srx-r220/overview

[10] V. Baroncini, "Title," unpublished|.

[11] D. C. Initiatives, "Digital Cinema System Specification version 1.2," March 2008.

[12] D. Taubman, "Kakadu Software," 6.0 ed, 2008.

[13] VQEG, "Multimedia Group Test Plan Version 1.21," March 2008 2008.

[14] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase 1," September 2008 2008.

Paper B: Comparison of JPEG 2000 and H.264/AVC by Subjective Assessment in the Digital Cinema

B

Fitri N. Rahayu, Ulrich Reiter, Marlon T.M. Nielsen, Touradj Ebrahimi, and Andrew Perkis

Appeared in
Proceedings of 2nd IEEE International Conference on Quality of Multimedia Experience, Trondheim, Norway, June, 2010, ISBN: 978-1-4244-6958-1, pages 112-117

Abstract

Two video coding schemes with variable bit rates — JPEG 2000 and H.264/AVC — were compared in terms of perceived quality performance in a Digital Cinema environment. In this paper we describe in detail the procedure for a subjective quality assessment. The stimuli used were 10 seconds long color sequences at HD resolution (1920x1080 progressive), 30 fps, and YCbCr 4:2:0. We do not consider DCI-specified content, but rather exploring the quality of rates suitable for alternative content. The results show that temporal compression schemes like H.264/AVC can play out their high coding efficiency not only at SD resolutions, but also at high resolutions and at high bit rates around 31 Mbps.

1 Introduction

In Digital Cinema, the use of compression is a matter of practicality. The quantity of data needed to represent high-quality imagery in its native uncompressed form is prohibitive [1]. The Digital Cinema Initiative (DCI) has released a specification for Digital Cinema format files [2]. Based on this specification, the size of one frame of a Digital Cinema Distribution Master (DCDM) can be up to almost 40 megabytes (4K, 12 bit per component), such that two hours of a 24 fps movie can amount to a total of close to 7 terabytes. This is a very huge amount of data that makes storage and transmission of such data physically possible, but rather impractical with today's storage and transmission technology. Hence, Digital Cinema cannot become a practical form of business without significantly reducing the quantity of data. In this paper, we study two existing compression algorithms, JPEG 2000 and H.264/AVC, in the context of Digital Cinema applications.

Evaluation of compression technology for digital still image or moving pictures is part of the acceptance process in the international standardization community [3]. The evaluation generally consists of comparative studies to test the compression efficiency attained by the coding algorithm, computational complexity, and additional features and functionality. Performing subjective quality assessments is one of the means to study the compression technology. Subjective quality assessments are needed to evaluate the visual quality of compressed images or moving images at a certain number of bits used to represent the compressed items, along with computational complexity. Motivated by the fact that conducting a subjective quality assessment is time consuming and not necessarily straightforward, the research community has developed several objective metrics to model how humans perceive the quality of images or moving pictures. However, the existing objective metrics for predicting the perceived quality are limited. The subjective score collected in a carefully designed experiment is still considered the benchmark of quality evaluation.

In subjective quality assessments, a group of human participants is asked to watch a set of moving pictures with varying quality, and to rate the perceived quality on a pre-defined scale. From these ratings, a MOS (Mean Opinion Score) can be obtained by averaging the collected ratings, assuming that they follow a Gaussian distribution. In order to obtain a meaningful MOS, a proper and systematic procedure must be applied to the experiment and the collected subjective ratings. Currently, there are some recommendations issued by international standardization bodies concerning the procedure of conducting subjective visual quality assessments. However, at present there are no existing recommendations specifically directed towards subjective visual quality assessments in the Digital Cinema environment.

In this paper, we present a detailed procedure for subjective visual quality assessment of high quality moving images in the Digital Cinema. The stimuli are moving images compressed with two different algorithms, JPEG 2000 and H.264/AVC. The collected subjective data is then used to analyze and study the performance of JPEG 2000 and H.264/AVC in the Digital Cinema.

The test conditions of our experiment, including a description of test environment, dataset and configuration of coding algorithms, is described in detail in section 2. The

test methodology employed in the subjective assessment, including test design and analysis of subjective data, is presented in section 3. The results, including a discussion, are presented in section 4. Finally, section 5 summarizes the conclusions.

2 Test Conditions

2.1 Test Environment

The subjective quality assessment of JPEG 2000 and H.264/AVC described here was conducted at Nova 1, a DCI-specified cinema in Trondheim, Norway. As the cinema is in daily commercial use, it is considered a meaningful test environment for subjective quality assessments. A DCI-specified cinema is also considered to provide ideal viewing conditions. Table B.1 summarizes the specifications of the test environment.

Table B.1: Test environment specifications.

DISPLAY	
Screen (H x W)	5 x 12 m
Projection Distance	19 m
Image Format	WS 1:1.66
	WS 1:1.85
	CS 1:2.35
HALL	
Number of Seats	440
Width	18.3 m
Floor area	348 m^2
Built Year	1994

The digital cinema projector used in the experiment was a Sony CineAlta SRX-R220 4K projector [4]. Calibration and maintenance of the projector are regularly performed by Trondheim Kino. For that reason, measurement of contrast, screen illumination intensity and uniformity, or any other measurement were not considered necessary.

Although the cinema could obviously accommodate all 20 subjects at once, we designed the experiment to allocate only five subjects per session. The main reason was to avoid influence of two additional factors: the distance of subjects to the cinema screen, and the viewing angle. We chose the viewing distance (10 meters) to be 2 times the height of the screen. This resulted in subjects being placed in the 6^{th} row. In order to maintain a centralized viewing condition for all subjects, only 5 seats were allocated for subjects in this row. The subjects' exact position during the experiment is illustrated in Figure B.1and Figure B.2.

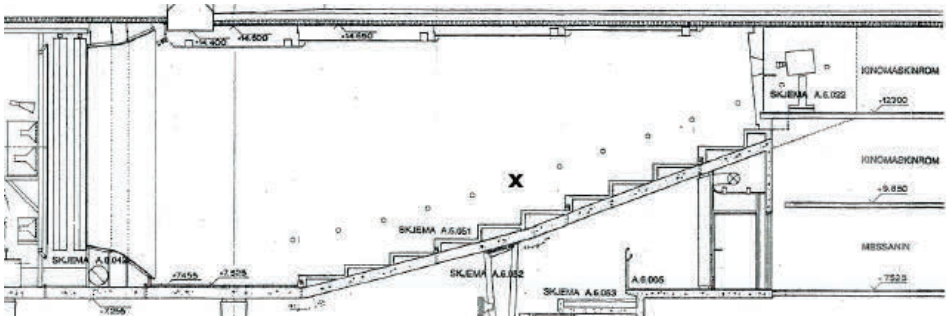


Figure B.1: Subject located at the 6th row from the screen.

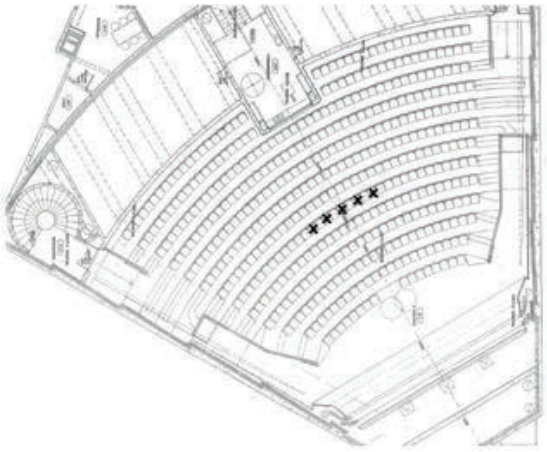


Figure B.2: Subjects' position at the 6th row.

In order to reproduce the cinema viewing experience, the assessment was conducted under the same conditions as when moviegoers watch a feature film, i.e. in complete darkness. To illuminate the subject's scoring sheets during the subjective assessment without affecting the projected images' perception, small low-intensity lights were attached to the clipboard used by each subject for voting.

2.2 Data Set

The data set was taken from the SVT High Definition Multi Format Test Set [5], EBU database [6], NRK [7], and NTIA/ITS database [8]. The dataset format was HD 1920x1080 progressive and converted into 30 fps, YCbCr 4:2:0 format using VirtualDub [9]. The whole set of test sequences was split into a training set of two sequences from the NTIA/ITS database (Aspen and RedKayak), a testing set of six sequences from SVT (CrowdRun, DucksTakeOff, OldTownCross, IntoTree, and

ParkJoy) and EBU database (Dancer), and a dummy set of one sequence from NRK used for the stabilization stage. The dataset used for training and the dummy sequence are illustrated in Figure B.3. The set of test sequences is shown in Figure B.4.

Figure B.3: Training and dummy set.

Figure B.4: Test set. From top left to bottom right: CrowdRun, Dancer, DucksTakeOff, OldTownCross, IntoTree, and ParkJoy.

2.3 Codecs

For the lossy compression of high resolution (HD 1920 x 1080 progressive at 30 fps) videos, two different codecs were considered. These were JPEG 2000 and H.264/AVC. Different coding bit rates were selected for the test.

2.3.1 JPEG 2000

JPEG 2000 is a wavelet-based compression scheme for still images and image sequences such as those in Digital Cinema [10]. For JPEG 2000 coding, the Kakadu version 6.0 [11] was used for the implementation. One configuration of encoding with the following parameters was used: codeblock size 64x64 (default), one tile per frame (default), 5 decomposition level (default), and visual frequency weighting factor as recommended for Digital Cinema environment in [12].

2.3.2 H.264/AVC

Table B.2: H.264/AVC encoding parameters

Reference software	JM 16.1
Profile	High (FREXT Profile)
Number of frames	300
Chroma format	4:2:0
GOP structure	IBPBPBPBPBPBP
Number of reference frames	2
Slice mode	off
Rate control	Enabled (initial QP=30)
Macroblock partitioning for motion estimation	Enabled
Motion estimation algorithm	Fast full search (default)
Early skip detection	Disabled
Selective intra mode decision	Disabled

H.264/AVC is the latest motion-compensation-based compression scheme for video [13]. For H.264/AVC coding, the JM version 16.1 [14] was used for the implementation. One configuration of encoding with parameters depicted in

Table B.2 was utilized.

2.4 Description of hardware

A PC-based server was used to play back the stimulus. All the compressed stimuli are decoded first before the experiment was carried out. The output interface of the server was a DVI connector, whereas the input interface of the projector is HD-SDI. Therefore, a DVI to HD-SDI scaler was used to bridge the two different types of interface. We carefully set the DVI to HD-SDI scaler so that it didn't do any further unnecessary processing (such as resolution transformation) to the stimulus.

3 Test Methodologies

In the field of subjective evaluation, there are many different methodologies and rules to design a test. The test recommendations described by the ITU have been internationally accepted as guidelines for conducting subjective assessments.

3.1 Presentation Method and Scale

We adopted the single stimulus method described in recommendation ITU-R BT.500-11 [15].

3.2 Training

In the beginning of each test session, an instruction sheet was provided to each subject to give a brief introduction to the task. This was followed by an extended oral explanation. This included a definition of the English scale terms (see Figure B.5), which were related to the quality range of stimuli presented on the screen. Subjects were

also given the opportunity to ask questions regarding their task. Then a training session was conducted in order to familiarize the subjects with the assessment procedure. The training session lasted around five minutes.

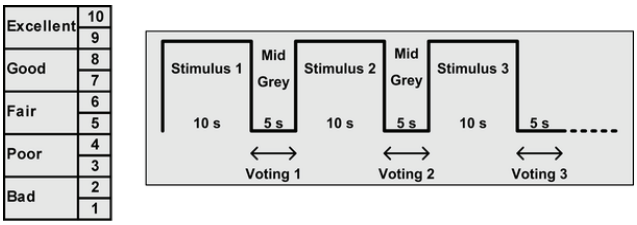


Figure B.5: Presentation method and scale.

3.3 Test and Test Subjects

The test was conducted as a single session. The six test sequences shown in Figure 4 were compressed at various bitrates, resulting in 62 different test conditions which were presented in randomized order. Additionally, five dummy conditions were included at the beginning of the test session to stabilize the subjects' ratings. Thus, there were a total of 67 test conditions for a single session, which lasted around 17 minutes.

20 subjects (8 females, 12 males), who reported to have normal or corrected to normal vision, participated in the experiment. Prior to the experiment, all subjects were screened for color blindness. Half of the subjects were familiar with image processing and compression artifacts. Subjects' age ranged from 21 to 47 years. Two outliers were detected and discarded.

3.4 Statistical Analysis of the Collected Data

The statistical analysis of the assessment data is based on the following model:

$$m_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \text{(B.1)}$$

Here, m_{ij} is a score obtained from subject i after scoring stimulus j; μ is the overall mean score computed across all subjects and stimuli; α_i is the subject effect; β_j is the effect of specific stimulus j; ε_{ij} is an experimental error caused by uncontrollable variables [16].

3.4.1 Distribution of data

Distribution of the collected score can be analyzed for each subject, across different test conditions, or for each test condition across different subjects. We used a Shapiro-Wilk test to verify the normality of distributions. The result showed that, as expected, score distributions for each subject across different test conditions were not normally distributed. However, the majority (75%) of the score distributions for each test condition across subjects were normal or close to normal (mean p-value equal to 0.073). The results validate the processing applied to the data which is explained in the next subsections.

3.4.2 Offset correction

Based on the model given in Eq. (B.1), it is relevant to verify if there are significant differences between the ways subjects used the rating scale when scoring the stimulus. To verify how subjects used the rating scale, first, we have to check the distribution of the raw score collected in the subjective assessment. Again, we used a Shapiro-Wilk test to verify the normality of distribution, and the computed p-value was equal to 0, which indicates that the distribution was not normal. Hence, it was not suitable to use a parametric test like ANOVA (analysis of variance) to investigate whether variations of scores across the subjects were large. Instead, we used a non-parametric Kruskal-Wallis analysis of variance [17] . This test was performed on the raw scores across the subjects. The differences between subjects were significant, with $H(19) = 78.354$, and $p<0.05$. This indicates that indeed there were large variations in means of subjective scores among subjects, i.e. there were significant differences between the ways subjects used the rating scale to judge the quality of the stimulus. Consequently, a subject-to-subject correction was applied by normalizing all the scores according to an offset mean correction [16].

3.4.3 Outlier detection and removal

An outlier detection was performed according to the guidelines described in section 2.3.1 of annex 2 of recommendation ITU-R BT. 500-11 [15]. Two outliers were detected out of 20.

3.4.4 Mean Opinion Score (MOS)

After discarding the outliers, the MOS was then computed for each test condition, together with the 95% confidence interval. The confidence interval for each MOS was computed using the Student's t-distribution.

4 Results and Discussion

Figure B.6 illustrates the results (MOS vs. bit rate) for each test sequence. It can be seen that for the same bit rate value for each test sequence, H.264/AVC encoded sequences received a higher MOS compared to the JPEG 2000 encoded ones. This is apparent for all selected bitrates of all contents. This result was anticipated, since only the H.264/AVC algorithm employs motion estimation. Motion estimation provides a considerable level of temporal compression that is capable of providing significant improvement in coding gain without loss of perceived quality. Consequently, there is a significant impact on the coding gain of H.264/AVC compared to the coding gain of JPEG 2000. The latter does not exploit at all the redundancy of temporal information of a motion picture sequence. Rather, it treats each frame as a separate entity, and compresses without any reference to other frames.

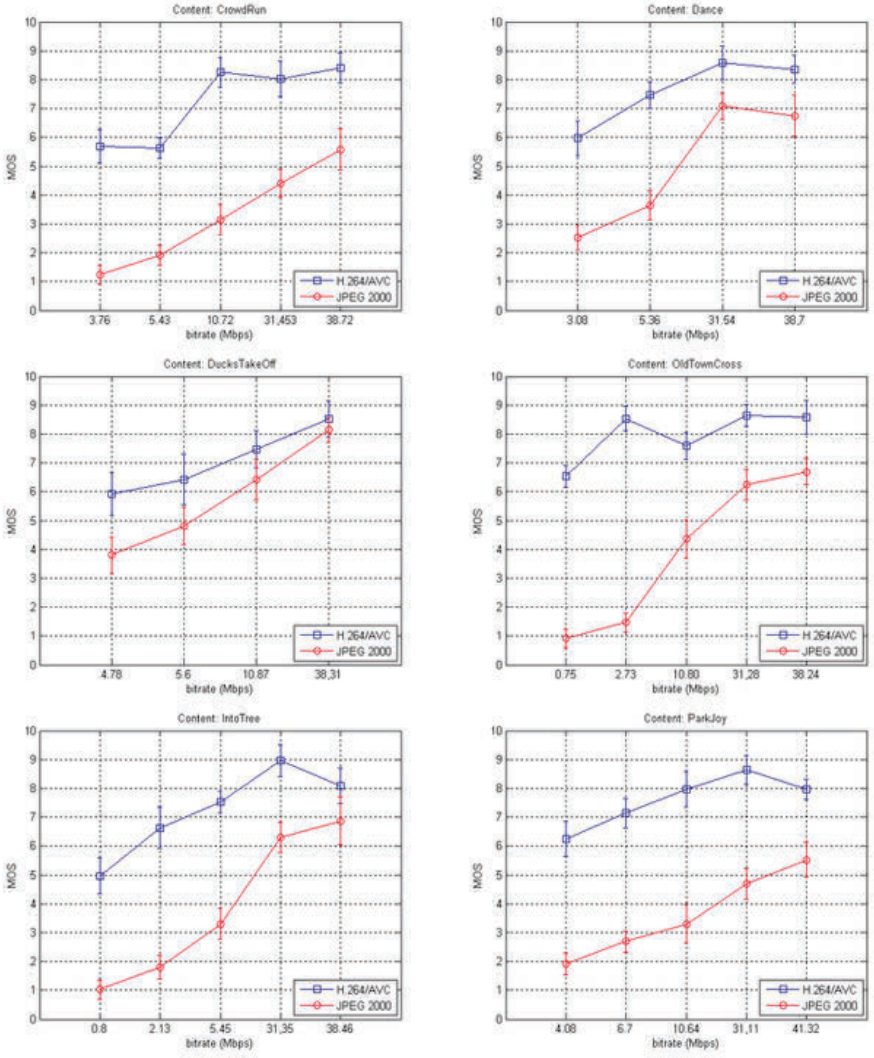


Figure B.6: MOS vs. bit rate for both codecs across test sequences.

Temporal compression based on a motion estimation algorithm is at times criticized in the context of Digital Cinema [1]. As the motion estimation feature of H.264/AVC allows for very high levels of compression, it is often used in applications that need very low bitrates to work properly, e.g. on free video sharing websites. Consequently, at the low bit rates used there, major artifacts appear in the content. Yet, it is important to notice that it is not the motion estimation itself that causes additional artifacts – rather, these are caused by the low bit rates necessary in the aforementioned applications. The results illustrated in Figure B.6 show that a MOS higher than 8 — a value that corresponds to good perceived quality — can even be attained by content

with bit rates lower than 100 Mbps when using the H.264/AVC codec. 100 Mbps is the lowest bit rate value used in commercial Digital Cinema practice.

Interestingly, MOS for H.264/AVC encoded material seems to peak around 31 Mpbs. A further increase in perceived quality could only be noted for one of the sample contents ('CrowdRun'), and it is not statistically significant.

5 Conclusions

Making a final conclusion on which is the best codec providing the better performance solely based on the MOS as computed from the subjective data in this experiment is not possible. Such a comparison would have the nature of an apples and oranges comparison, because H.264/AVC-encoded content includes P and B frames besides the I frames, whereas JPEG 2000-encoded content only includes I frames. Furthermore, our study was performed using 4:2:0 format content which is not included in the DCI specifications. We do not consider DCI-specified content but rather exploring the quality of rates suitable for alternative content, such as current screening of opera from New York Metropolitan in Nova 1. Based on our study, we can state that the usage of H.264/AVC in a Digital Cinema environment, i.e. a presentation of high quality content on a large screen using temporal compression, is very well possible. The assumption stated elsewhere [1] that temporal compression introduces major artifacts, could not be substantiated. Instead, the gain in bit rate that a temporal compression scheme provides, can very well be used to further increase the quality of the encoded stream, resulting in higher MOS at equal bit rates when compared to JPEG 2000.

As an extension of this work, we are planning to perform subjective visual quality assessments of DCI-specified content, such as the DCI Standard Evaluation Material (StEM) encoded with JPEG 2000 and H.264/AVC. Also, in future assessments we wish to take into account the multimodal factor by adding audio, i.e. conducting subjective audiovisual quality assessments in the Digital Cinema environment.

References

[1] P. Symes, "Compression for Digital Cinema," in *Understanding Digital Cinema: A Professional Handbook*, ed: Focal Press, 2005, pp. 121-148.

[2] DCI. (2008). *DCI Digital Cinema Initiatives*. Available: http://www.dcimovies.com

[3] F. D. Simone, L. Goldmann, V. Baroncini, and T. Ebrahimi, "Subjective evaluation of JPEG XR image compression," 2009.

[4] SONY, "4K Digital Cinema Projectors SRX-R220/SRX-R210 Media Blok LMT-100 Screen Management System LSM-100," 2007.

[5] SVT. (2006). *The SVT High Definition Multi Format Test Set*. Available: http://tech.ebu.ch/docs/hdtv/svt-multiformat-conditions-v10.pdf

[6] EBU. Available: http://www.ebu.ch/fr/technical/hdtv/test_sequences.php

[7] NRK. Available: http://nrk.no/about/

[8] NTIA/ITS. Available: ftp://vqeg.its.bldrdoc.gov/HDTV/NTIA_source/

[9] "VirtualDub," ed.

[10] ITU, "Information technology – JPEG 2000 image coding system: Core coding system," 2002.

[11] D. Taubman, "Kakadu Software," 6.0 ed, 2008.

[12] D.-C. AHG, "Guidelines for digital cinema applications," JPEGOctober 2009 2009.

[13] ITU-T, "Advanced video coding for generic audiovisual services," 2005.

[14] *JM version 16.1*. Available: http://iphome.hhi.de/suehring/tml/

[15] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva2002.

[16] E. D. Gelasca, "Full-reference objective quality metrics for video watermarking, video segmentation and 3D model watermarking," Ph.D, EPFL, Lausanne, 2005.

[17] H. Coolican, *Research methods and statistics in psychology*. London: Hodder Arnold, 2004.

Paper C: Exploring Alternative Content in Digital Cinema

Fitri N. Rahayu, Touradj Ebrahimi, Andrew Perkis

Appeared in
Proceedings 16th IEEE International Conference on Virtual Systems and Multimedia (VSMM), Seoul, Korea, 2010, ISBN: 978-1-4244-9025-7, page 295-296

C

Abstract

Digital Cinema (D-Cinema) business is much more than feature films. Experimentation by showing programming other than feature films has been carried out since the early days of D-Cinema. Alternative content in Digital Cinema enables the cinema to become a multi-arts venue, attracting new and existing users by offering a range of vary products. We give an overview on this alternative content. Human factors issue—Quality of Experience (QoE), which is closely associated with the adoption of alternative content in D-Cinema, is also discussed.

1 Introduction

The motion picture industry (cinema) is the last of entertainment industry to go digital [1]. Currently, the motion picture industry is in the process of forming standards for digitization of its complete value chain. This process of change is referred to as Digital Cinema (D-Cinema) roll out. The digitization is specified by Digital Cinema Initiative (DCI) and is currently under standardization by The Society of Motion Picture and Television Engineers (SMPTE). D-Cinema demands a complete change of infrastructure in one of the players—the exhibitors or the owners of the screens—by adopting the new technology. The traditional 35 mm films projector needs to be replaced with a D-Cinema server and a digital projector. One of the main problems that hold back the successful roll out of D-Cinema in the market is that the exhibitors are those in value chain with the least benefit of the digitization. However, there are optimistic prospect to embrace D-Cinema pushed by two technical reinventions: the rebirth of 3D and the possibility for exhibitors to screen alternative content [2].

Other Digital Stuff (ODS)—the reference term for alternative content in D-Cinema application—is one of the issues that plays as one of change agents who contribute to the business model innovation and may transform the whole business of cinema exhibition into something different from what we know of today. It includes the creation of new services and business with high quality imagery content for the big screen, which maximizes business profit for exhibitors who adopt D-Cinema.

ODS has been slow to develop, and the value of the ODS sector is relatively low compared to traditional exhibition, the feature films. Nevertheless, it has potential. Norwegian cinema in Trondheim which is operated by Trondheim Kino AS regularly screens ODS to public and receives regular revenue streams from it.

Even though the alternative content has been experimented since the early days of D-Cinema technology, it is still a minority activity for most exhibitors, and it contributes to the lackluster popularity of ODS exploration and implementation. Exploring alternative content in D-Cinema is not a widely discussed subject at this time. In this paper, we present and discuss cases beyond traditional use of D-Cinema and its venue, including what has been done at Nova Kino, a commercial DCI-specified cinema in Trondheim, Norway. We also describe important human factors issue that contribute to the successful of ODS adoption—Quality of Experience (QoE), which is the perspective of the users who consume the screened content.

The rest of this paper is organized as follow. Section 2 provides a description of alternative content in D-Cinema. We present the QoE as one important human factors issue in D-Cinema in Section 3. Finally, we give a short outlook and conclude the paper.

2 Alternative Content in Digital Cinema

ODS covers all content screened in the cinema that is not feature film. This includes advertising, live events (sport, music), educational and gaming [2]. The exploitations of ODS at Trondheim cinema have ranged from Laparoscopic surgery, live stream music concert from various genres (including heavy metal and opera), and live stream of World Cup 2010 matches in 3D.

Midgard Media Lab, NTNU and The Operation Room of the Future at St. Olavs hospital in Trondheim together with several industries partners successfully transmitted live HD surgical images from operation rooms to 4K cinema projectors at Nova Kino 1 [3]. 200 medical doctors from around the world experienced the successful live event in the cinema. Nova Kino 1 also regularly screens opera transmitted from The New York Metropolitan Opera [4]. Other musical genre performance which had been screened live is heavy metal concert; live HD concert of four popular heavy metal bands was broadcasted from Bulgaria to several digital cinemas including Nova Kino 1.

D-Cinema can also be exploited for gaming event. The unique atmosphere of the cinema auditorium, large screen display and sound reproduction in the cinema can provide users, especially gamers, a unique out-of-home experience. Multi-player gaming events using D-Cinema had been organized before [5]. It shows the potential possibility of integrating multi-player games genre, including serious games such as WON (World of NTNU) [6] with D-Cinema as a new variety of ODS.

Figure C.1: From top to down: image from the OR shot with SONY HDC-X300K HD Camera and patient's stomach tissue shot with Olympus EndoEye HD-TV Video .

3 Quality of Experience

Development of ODS will certainly revitalize the cinema experience. Digital multimedia presentations especially in the case of D-Cinema are meant for human consumption and are powerful expressions. In the case of D-Cinema, the screening provides a rich experience to the cinemagoers as users. However, before the content is

consumed by the users, it usually goes through many processing stages. Each stage may introduce artifacts that degrade the cinema experience. Consequently, quality assessments of multimedia presentations in the D-Cinema system are essential.

The focus of quality measurement has recently shifted towards Quality of Experience (QoE), which is more related to how end users (or cinemagoers) experiences, perceives, and values multimedia presentation. Users are put at the centre of attention considering that the industry is there to serve the users and must understand their needs and perception in offered products and services. The expectations and technical comfort levels of the users have evolved in terms of complexity, as users are increasingly embracing advanced technologies which fit their lifestyle (leisure, work and education). The framework to assess the user's behavior and the necessary technology management is based on assessing the user experience in a consistent way, and rewarding the user's loyalty through innovative packages and new engaging services. Thus, QoE assessments are crucial for driving the innovations in D-Cinema industry [2].

To facilitate the research of QoE of digital multimedia presentation, a controlled environment, such as laboratory, is needed for a range of perception experiments involving human participants. We participate in the research of QoE in D-Cinema by exploiting a commercial DCI-specified cinema in Trondheim (Nova Kino 1) as a realistic test environment for numerous experiments, and we had made several contributions on visual perceptual quality of high quality imagery in D-Cinema [7], [8], [9].

4 Conclusion

In this paper, we described the issue of ODS—the alternative content in D-Cinema and how its innovation and exploration tightly connected to a human factors issue—QoE. Our experience on using a commercial DCI-specified cinema as a laboratory for QoE experiments indicates that understanding the whole experience of cinemagoers and assessing its quality by subjective experiments will need novel methodologies, such as methodology that takes into account multi-modal factors. Setting up a DCI-specified commercial cinema into a test environment for perception experiments using novel methodologies is a challenge; but we believe the process plays a role in exploration for novel alternative content in D-Cinema.

5 References

[1] A. Perkis and P. V. Sychowski, "NORDIC – Norway's digital interoperability in cinemas," *Journal St. Malo: NEM - Networked and Electronic Media,* 2008.

[2] A. Perkis, F. N. Rahayu, U. Reiter, J. You, and T. Ebrahimi, "Quality of Experience for High Definition Presentations--Case:Digital Cinema," in *High-Quality Visual Experience*, ed: Springer, 2010.

[3] M. M. Lab, "Report from Live Transmission of HD-SDI Operation Images to 4K Projector," NTNU, Trondheim2006.

[4] (7 July 2010). *Opera høsten 2010 og våren 2011*. Available: http://www.trondheimkino.no/incoming/article240119.ece

[5] (2009, 7 July 2010). *Sony 4K Technology Brings Multi-Player Gaming to Digital Cinema Theatres*. Available: http://www.dcinematoday.com/dc/PR.aspx?newsID=1551

[6] J. Xu, J. Puig, J. Lomeland, and A. Perkis, "A Serious Game for Both University Recruitment and Research Platform," presented at the Games: Design and Research Conference, Volda, Norway, 2010.

[7] F. N. Rahayu and U. Reiter, "Analysis of SSIM Performance for Digital Cinema Applications," in *1st IEEE QoMEX*, San Diego, USA, 2009.

[8] F. N. Rahayu and U. Reiter, "Comparison of JPEG 2000 and H.264/AVC by Subjective Assessment in the Digital Cinema," in *2nd IEEE QoMEX*, Trondheim, 2010.

[9] J. You, F. N. Rahayu, U. Reiter, and A. Perkis, "HVS based Image Quality Assessment for Digital Cinema," in *SPIE Electronic Imaging: Image Quality and System Performance VII*, San Jose, USA, 2010.

Paper D: Subjective Visual Quality Assessment in the Presence of Audio for Digital Cinema

Fitri N. Rahayu, Ulrich Reiter, Junyong You, Andrew Perkis, Touradj Ebrahimi

Appeared in
Proceedings 3rd IEEE International Conference on Quality of Multimedia Experience, Mechelen, Belgium, September 2011.

D

Abstract

In this paper, we investigate whether the presence of audio with different quality levels can influence the outcome of subjective visual quality assessment in a Digital Cinema setting. We asked the participants to judge the visual quality when watching an audiovisual content in a Digital Cinema environment and investigated whether the participants can neglect the presence and the quality of audio. The stimuli used were 10 seconds long color sequences accompany with orchestral music at 2K resolution, 24 fps, and YCbCr 4:4:4 played on DCI certified equipment. The result show that in visual only subjective quality assessment, the presence of audio (low or high quality) does not significantly influence on the visual quality judgment.

1 Introduction

Performing subjective visual quality assessments is one of the means to study the visual quality as perceived by the end user. One use of this practice is to study the quality degradation introduced by content compression. Subjective quality assessments are needed to evaluate the visual quality of compressed images before or after content delivery over a network. The subjective score collected in a carefully designed experiment is still considered the ground truth of quality evaluation

In subjective video quality assessments, a group of human participants are asked to watch a set of visual stimuli with varying quality, and to judge the perceived quality. One way of judgment is by giving a rating on a pre-defined scale. From these ratings, a MOS (Mean Opinion Score) can be obtained by averaging the collected ratings. In order to obtain a meaningful MOS, a proper and systematic procedure must be applied to the experiment and the collected subjective ratings. Currently, there are some recommendations issued by international standardization bodies concerning the procedure of conducting subjective visual quality assessments. These recommendations include the use of various visual stimuli to determine the perceived visual quality.

During the subjective quality assessment, many factors influence the judgment of stimulus quality. Figure D.1 illustrates a holistic model of participant who participates in a subjective audio quality experiment as proposed by Zielenski, Rumsey, and Bech [1]. The hearing block corresponds to the properties of a listener. Perception block corresponds to the cognitive processes that make the listener able to describe and distinguish the sound in terms of its basic characteristics and attributes. The judgment block, which represents judgmental processes that are responsible for the assessment of sound in terms of its character, is considered as the main component of the holistic model. The last block, mapping block, relates to the processes engaged in the conversion of internal judgment into the quantifiable response.

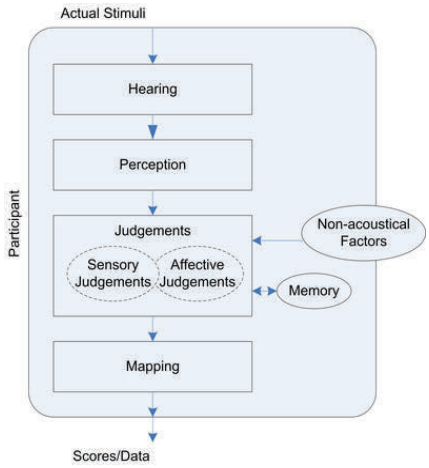


Figure D.1: Holistic model of listener [1].

The model in Figure D.1 can also represent a participant in a subjective visual quality assessment. The judgment block shows how the non-acoustical factors can have an influence on the judgment process. If we use this model to represent the participant‘s judgment during the subjective visual quality assessment, it indicates that the non-visual factor can have an influence on the judgment process.

Several studies on subjective audiovisual quality assessment showed that multimodal context influences how participants perceived the quality of audiovisual content; visual quality affects the perceived audio quality and vice versa [2, 3].

In subjective visual quality assessment, the participants are asked to judge the quality of a series of visual stimulus. However, single modality stimulus (e.g., visual only content) is rarely presented in the commercial Digital Cinema applications. When watching a movie and asked to judge the quality of the picture, is it possible for the participants to neglect the presence and the quality of the audio?

In this paper, we would like to investigate whether the presence of audio with different quality levels can influence the perceived visual quality of participants in a Digital Cinema setting. We present a detailed procedure for subjective visual quality assessment of high quality audiovisual content presented on a DCI certified D-Cinema screen in a commercial cinema. The stimuli are moving images compressed with JPEG 2000. Some stimuli are audiovisual contents with the audio stimulus compressed with MPEG Audio Layer III.

The test conditions of our experiment, including a description of test environment, dataset and configuration of coding algorithms, is described in detail in section 2. The test methodology employed in the subjective assessment, including test design and analysis of subjective data, is presented in section 3. The results, including a discussion, are presented in section 4. Finally, section 5 summarizes the conclusions.

2 Test Conditions

2.1 Test Environment

The subjective visual quality assessment described here was conducted at Nova 1 – the Liv Ullman theater, a DCI-specified cinema in Trondheim, Norway. As the cinema is in daily commercial use, this is a realistic and meaningful test environment for subjective quality assessments.

Figure D.2: Participants located at the 6th row from the screen.

Table D.3 summarizes the specifications of the test environment.

Although the cinema could obviously accommodate all participants at once, we designed the experiment to allocate a maximum of five participants per session. The main reason was to avoid influence of two additional factors: the distance of participants to the cinema screen, and the viewing angle. We chose the viewing distance (10 meters) to be 2 times the height of the screen. This resulted in participants being placed in the 6th row. In order to maintain a centralized viewing condition for all

participants, only 5 seats were allocated for participants in this row. The participants' exact position during the experiment is illustrated in Figure D.2.

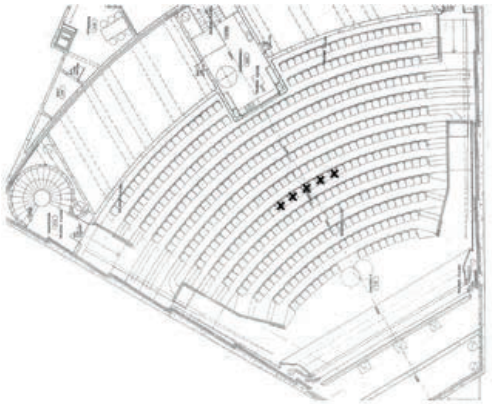


Figure D.2: Participants located at the 6th row from the screen.

Table D.3: Test environment specifications.

DISPLAY	
Screen (H x W)	5 x 12 m
Projection Distance	19 m
Image Format	WS 1:1.66
	WS 1:1.85
	CS 1:2.35
HALL	
Number of Seats	440
Width	18.3 m
Floor area	348 m^2
Built Year	1994

In order to reproduce the cinema viewing experience, the assessment was conducted under the same conditions as when moviegoers watch a feature film, i.e. in complete darkness. To illuminate the participant's scoring sheets during the subjective assessment without affecting the projected images' perception, small low-intensity lights were attached to the clipboard used by each subject for voting.

2.2. Data Set

The data set was 10 seconds audiovisual sequences taken from the DCI Standard Evaluation Material (StEM) [4]. StEM test materials include a lot of 12 minutes audiovisual sequences. We selected some scenes for our data set. The audio information in our data set is orchestral music. The data set format was 2K, 24 fps, and YCbCr 4:4:4. The whole set of test sequences taken from StEM was split into a training set containing two sequences, a testing set with four sequences and a dummy set used for the stabilization stage. The dataset used for training and the dummy sequence are illustrated in Figure D.3. The set of test sequences is shown in Figure D.4. StEM test material is a 12 minutes audiovisual sequence.

Figure D.3. Training and dummy set.

Figure D.4: Test set. From top left to bottom right: Sequence 1, Sequence 2, Sequence 3, and Sequence 4.

2.3 Stimuli

In order to create stimuli with various visual quality levels, we encoded our data set with JPEG 2000 at different coding bit rates. We selected the bit rates of 20 Mbps, 40 Mbps, 60 Mbps, and 160 Mbps. We also incorporated four audio conditions (no audio, uncompressed audio, and two compressed conditions) for each selected bit rates resulting in a total of 64 stimuli with different levels of quality. In commercial Digital Cinema applications, the audio signal is not compressed. However, in our experiment, we compressed the audio signal in order to degrade the audio quality by introducing the compression error so that we had audio stimuli with different quality levels. We used the MPEG Audio Layer III compression algorithm. We chose this algorithm because it is a widely used compression algorithm for audio.

2.3.1 JPEG 2000

JPEG 2000 is a wavelet-based compression scheme for still images and image sequences such as those in Digital Cinema [5]. For JPEG 2000 coding, the Kakadu version 6.0 software [6]was used to encode our data set. The configuration of encoder used in our experiment is illustrated in Table D.4.

Table D.4: JPEG 2000 encoding parameters.

Reference software	Kakadu version 6.0
Codeblock size	64x64 (default)
Decomposition level	5 level (default)
Number of tile	One tile per frame (default)
Visual frequency weighting factor	As recommended for D-Cinema environment [7]
Bit rate selection	20 Mbps, 40 Mbps, 60 Mbps, 160 Mbps

2.3.2 MPEG Audio Layer III

MPEG Audio Layer III is a digital audio encoding format for lossy audio compression that was designed by the Moving Picture Experts Group as part of its MPEG-1 standard and later extended in MPEG-2 standard. We used the LAME encoder software [8] licensed under the LGPL to encode our data set. The configuration of LAME encoder used for our experiment is given in Table D.5.

Table D.5: MPEG Audio Layer III encoding parameters.

Reference software	LAME 3.98.4
Encoding modes	Constant Bitrate (CBR)
Sample rate	12 kHz
Total channels	2 (Stereo)
Bit rate selection	8 kbps and 24 kbps

2.4 Description of Hardware

A PC-based server was used to play back the stimuli. All the compressed stimuli are decoded first before the experiment was carried out. The output interface of the server was a DVI connector, whereas the input interface of the projector is HD-SDI. Therefore, a DVI to HD-SDI conversion box was used to bridge the two different types of interface. We carefully set this box so that it did not do any further unnecessary

processing (such as resolution transformation) to the stimulus. Figure D.5 illustrates the hardware set up of the experiment.

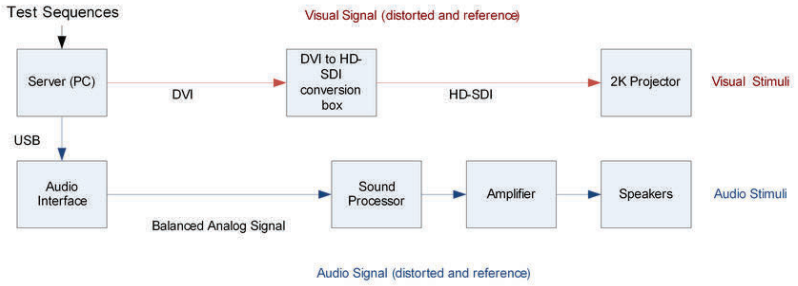


Figure D.5: Hardware illustration of the experiment.

3 Test Methodologies

In the field of subjective evaluation, there are many different methodologies and rules to design a test. The test recommendations described by the ITU have been internationally accepted as guidelines for conducting subjective assessments. However, at present there are no existing recommendations specifically directed towards subjective visual quality assessments in the Digital Cinema environment. In our experiment we adopted existing test methodology based on ITU recommendations and modified it to suit our test environment.

3.1 Presentation Method and Scale

We adopted the double stimulus method described in recommendation ITU-T P.911 [9] and used a ten point discrete quality scale representing bad, poor, fair, good, and excellent quality. The presentation method and the scale are illustrated in Figure D.6.

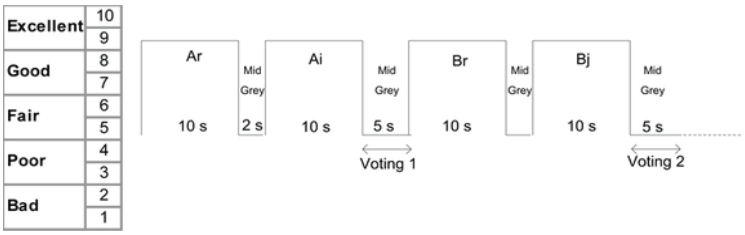


Figure D.6: Scale and Presentation Method (Ai is sequence A under test condition i; Ar, Br are sequences A and B in the reference source format; Bj is sequence B under test condition j).

3.2 Training and Test

In the beginning of each test session, an instruction sheet was provided to each participant to give a brief introduction to the experiment. This was followed by an extended oral explanation. We asked the participants to judge only the visual quality of the sequences on the screen. This included a definition of the English scale terms, which were related to the quality range of stimuli presented on the screen. Participants were also given the opportunity to ask questions regarding their task. Then a training session was conducted in order to familiarize the participants with the assessment procedure. The training session lasted around ten minutes.

3.3 Test and Test Subjects

The test was conducted as a single session with a break of three minutes in the middle of the session. The break was intended so that the participants can recharge to avoid losing concentration during the last 15 minutes of stimuli presentation due to fatigue. During the break participants were presented with relaxing music and images. In addition to the 8 dummy sequences used for stabilization phase in the beginning of assessment, the four test sequences shown in Figure D.4 were compressed at various bit rates (4 audio conditions and 4 visual conditions), as introduced in Table D.4 and Table D.5, resulting in 64 test stimulus. Thus, there were a total of 72 test conditions for a single session, which lasted around 30 minutes.

A total of 15 participants (4 females, 11 males) participated in the experiment. Prior to the experiment, all subjects were screened for visual acuity and color blindness.

We conducted three sessions, and the test conditions are randomized for each session.

3.4 Statistical Analysis of the Collected Data

The statistical analysis of the assessment data is based on the following model:

$$m_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

Here, m_{ij} is a score obtained from participant i after scoring stimulus j; μ is the overall mean score computed across all participants and stimuli; α_i is the participant effect; β_j is the effect of specific stimulus j; ε_{ij} is an experimental error caused by uncontrollable variables [10].

3.4.1 Distribution of Data

Distribution of the collected score can be analyzed for each participant, across different test conditions, or for each test condition across different participants. We used a Shapiro-Wilk test to verify the normality of distribution. Since the selection of encoding parameter to create all stimuli is not based on normal distribution, the subjective data result showed that, as expected, score distributions for each participant across different test conditions were not normally distributed. However, the majority (92.2%) of the

score distributions for each test condition across participants were normal ($p>0.05$). The results validate the processing applied to the data which will be explained in the next subsections.

3.4.2 Offset Correction

Based on the model given in Eq. (1), due to participant effect, it is relevant to verify if there are significant differences between the ways participants used the rating scale when scoring the stimulus. To verify how participants used the rating scale, first, we have to check the distribution of the raw score collected in the subjective assessment. Again, we used a Shapiro-Wilk test to verify the normality of distribution, and the computed p-value indicates that the distribution was not normal ($p<0.05$). Skewness can also provide an indication of the distribution. The calculated skew value (0.6) is more than twice the standard error (0.08); it again indicates that the distribution indeed is not normal. Hence, it was not suitable to use a parametric test like ANOVA (analysis of variance) to investigate whether variations of scores across the participants were large. Instead, we used a non-parametric Kruskal-Wallis analysis of variance [11]. This test was performed on the raw scores across the participants. The differences between participants were significant, with $H(14) = 83.5$, and $p<0.05$. This indicates that indeed there were large variations in means of subjective scores among participants, i.e. there were significant differences between the ways participants used the rating scale to judge the quality of the stimulus.

Consequently, a participant-to-participant correction was applied by normalizing all the scores according to an offset mean correction as follow [10]:

$$\hat{m}_{ij} = \frac{1}{\hat{g}_i}(m_{ij} - \hat{b}_i) \tag{2}$$

where $\hat{g}_i$ is the corrected gain, $\hat{b}_i$ is the corrected offset, and $\hat{m}_{ij}$ is the normalized score. The offsets are estimated using the mean of all measurement made by each subject:

$$\hat{b}_i = \frac{1}{J}\sum_{j=1}^{J} m_{ij} - \mu \tag{3}$$

The corrected gains are estimated using this following model as shown in Eq. (4),

$$\hat{g}_i = \frac{1}{K}\max_{j \in J}(m_{ij}) \tag{4}$$

where K is equal to the upper end of the scale (10) in our experiment.

3.4.2 Outlier Detection and Removal

An outlier detection was performed according to the guidelines described in section 2.3.1 of annex 2 of recommendation ITU-R BT. 500-11 [12]. We did not detect any outliers. Hence we included the data from all participants.

3.4.3 Mean Opinion Score (MOS)

The MOS was then computed for each test condition, together with the 95% confidence interval. The confidence interval for each MOS was computed using the Student's t-distribution as shown in Eq (5).

$$CI = t_{0.05,14} \frac{\sigma}{\sqrt{n}} \tag{5}$$

4 Results and Discussion

Figure D.7 illustrates the MOS results for each selected bit rate within its 95% confidence interval. We show the result for each sequence because the non-parametric Kruskal-Wallis analysis of variance results shows that there are significant differences of scores between different sequences/contents [H(3) = 49.516, and p<0.05].

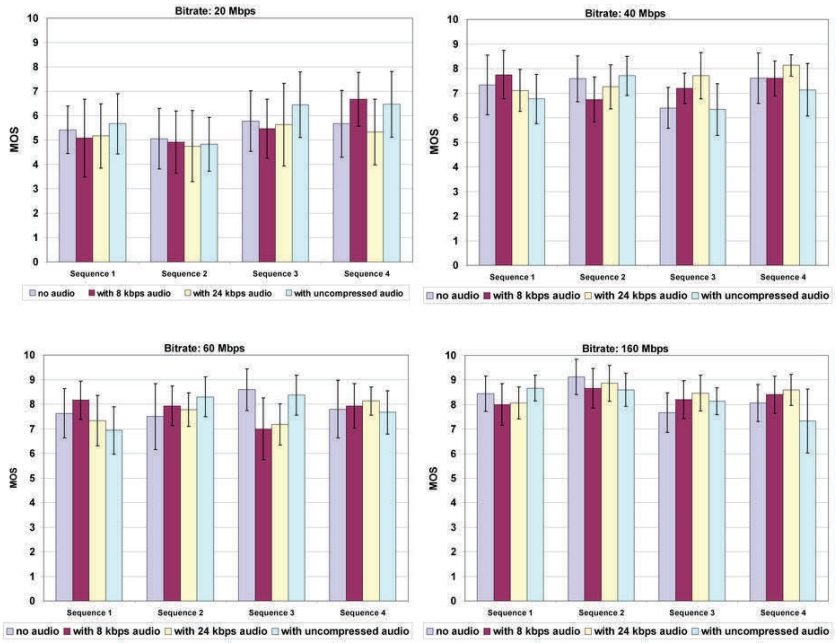


Figure D.7: MOS results for each selected JPEG 2000 coding bitrate.

For each sequence, the figure illustrates that there are overlaps of scores within 95% confidence interval at each bitrate for different audio condition, which indicates that the presence of audio did not significantly contribute to subjective visual quality during the test. Further analysis using non-parametric Kruskal-Wallis analysis of variance, as shown in Table D.6, shows that there are no significant score differences across different audio conditions. This result supported by a similar study in a different application [2] .

Table D.6: Result of non-parametric test

N=208		
Sequence #	**H(3)**	**Significance value**
1	1.641	p=0.650
2	1.212	p=0.750
3	1.478	p=0.687
4	1.805	p=0.614

5 Conclusions

During visual only subjective quality assessment in Digital Cinema using the test methodology based on ITU recommendations, the presence of audio (low or high quality) does not affect the visual quality judgment even though perceived visual quality can be influenced by non visual quality factors. When the participants were specifically asked to judge only the visual quality, their perception which corresponds to the cognitive processes during the experiment is able to disregard the basic characteristics and attributes of the audio even though the presence of the audio is prominent especially during low quality audio stimuli

6 References

[1] S. Zielinski, F. Rumsey, and S. Bech, "On Some Biases Encountered in Modern Audio Quality Listening Tests—A Review," *Journal of the Audio Engineering Society,* vol. 56, June 2008 2008.

[2] J. G. Beerends and F. E. D. Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *Journal of the Audio Engineering Society,* vol. 47, 1999.

[3] D. S. Hands, "A Basic Multimedia Quality Model," *IEEE Transactions on Multimedia,* vol. 6, December 2004 2004.

[4] D. D. C. Initiatives. *Digital Cinema Initiatives StEM Access Procedures*. Available: http://www.dcimovies.com/StEM/

[5] ITU, "Information technology – JPEG 2000 image coding system: Core coding system," 2002.

[6] *JM version 16.1*. Available: http://iphome.hhi.de/suehring/tml/

[7] D.-C. AHG, "Guidelines for digital cinema applications," JPEGOctober 2009 2009.

[8] "LAME encoder," ed.

[9] ITU-T, "Subjective audiovisual quality assessment methods for multimedia applications," ITU, Geneva2000.

[10] E. D. Gelasca, "Full-reference objective quality metrics for video watermarking, video segmentation and 3D model watermarking," Ph.D, EPFL, Lausanne, 2005.

[11] H. Coolican, *Research methods and statistics in psychology*. London: Hodder Arnold, 2004.

[12] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva2002.

Paper E: A Study of Quality of Experience in D-Cinema

Fitri N. Rahayu, Ulrich Reiter, Touradj Ebrahimi, and Andrew Perkis

Submitted to
Signal Processing: Image Communication, Theory, Techniques & Applications, A publication of the European Association for Signal Processing (EURASIP), ISSN: 0923-5965, 2011

E

Abstract

Quality assessment is an important matter in any multimedia presentation situation, including Digital Cinema (D-Cinema). The fundamental way of assessing quality is by using the Quality of Experience (QoE) concept, assessing how cinemagoers perceive the quality of the multimedia presentation. This paper will offer a study of visual quality of multimedia presentations in D-Cinema applications. We performed subjective visual quality assessment of images and motion pictures in a DCI-specified commercial Digital Cinema in Trondheim, Norway, using methodologies from standardized recommendations. Our interest is in exploring screening of alternative content using the D-Cinema equipment and environment. The designs of our subjective quality assessment took into account contents, bit rates and compression techniques used for alternative content but not used for screening the feature film. Using analysis of variance, we detected the significant differences of subjective scores among participants. Consequently, the obtained subjective scores were normalized first before MOS were computed. The result indicated that stimulus presentation method in the experiment influenced the human participants when judging transparent test sequences. The results also showed that the content types influenced the subjective scores.

1 Introduction

Multimedia in essence is a presentation of multiple information that may consist of image, video, graphics, audio, speech, sound, text, and even tactile content (content relating to the sense of touch) or olfactory content (content concerned with the sense of smell). In every multimedia presentation, including those in D-Cinema, content consumed by users has gone through several stages that incorporate specific multimedia signal processing methods. Every stage also introduces specific artifacts that might degrade quality. For this reason, it is important to evaluate content before it is finally consumed, such that a sufficiently high quality level is guaranteed. The long-established quality assessment approaches such as those in the Quality of Service (QoS) framework depend on metrics that only take into account factors from multimedia signals and network operations. For instance, the most widely used metric to evaluate image quality is PSNR—a metric that is solely based on an arithmetic pixel to pixel comparison between an original image and a distorted or processed version of it [1]. More recently, we have seen a paradigm shift towards incorporating the user as the most important element in the quality measurement of multimedia presentations. This shift has been a driver behind creation of the concept of Quality of Experience (QoE) [2, 3]

QoE is a multidimensional concept [2]. It consists of several objective and subjective parameters which contribute to the difficulty of quantifying it. Currently, there is no widely accepted metric that can measure QoE, and design of such metrics remains a challenging research topic. Furthermore, in addition to objective variables connected to a certain multimedia content, many variables that influence QoE can also originate from socio-cultural (e.g. age, sex, nationality) and psychological factors (e.g. expectation, social context). So far, the influences from subjective factors have been largely explored by introspection and intuition, but this approach suffers from a lack of validity, generality, and precision.

The techniques of psychologically oriented, controlled experimentation with groups of human participants can lead to a deeper understanding of the fundamentals of user experience in multimedia applications. The basic steps of a controlled experimentation is as follows: recognition of practical problems, coherent statement of testable hypotheses, manipulation of few independent variables, assessment of particular dependent variables, careful selection of experiment's participants, design of careful tasks for participants to perform, application of statistical tools, and finally, interpretation of results [4].

Subjective quality assessment is one approach of controlled experimentation to understand the effect of particular dependent variables on the perceived quality. The quality judgments are obtained from groups of human participants who are presented with stimuli, which consist of certain multimedia presentations of varying quality by manipulation of independent variables, in a controlled laboratory environment. This approach is a principal assessment method, in such a way that it is regularly carried out in the industry nowadays. For example, it is important for broadcasters to utilize human experts to evaluate varying range of encoded contents before they are transferred to be viewed by customers [5]. It is also important in the acceptance process of a new technology—testing a novel compression algorithm will certainly include evaluations that incorporate subjective quality assessments that are performed in laboratory

environments [6]. In addition, developing any perceptual quality metric, i.e.: a metric that takes into account human/user factor, certainly depends on the subjective data—data that is collected from subjective quality assessment [7]. For this reason, subjective quality assessment is one significant stage toward development and acceptance of any QoE metric, and a noteworthy stage requiring special care and attention.

Needless to say, the scientific method based on controlled experiments in a laboratory environment has its limitations. Finding adequate participants is sometimes difficult and laboratory conditions may distort the situation so much that the conclusions become unrealistic for the underlying application [8].

Different applications can provide different variables due to their situational context. To give an example, the ways a user experiences a multimedia presentation on a mobile device versus a TV set are likely to be very different. From a situational point of view, there are more variables in a mobile application when compared to TV application, due to wider scenarios such as interactivity and outdoor use. In multimedia presentations, regardless of the application, QoE is dominated by the quality of content which require high bandwidth and considerable processing power. From this point of view, video, image, and audio are most critical in the modeling of QoE, and the need for better understanding of the impact of audio-visual information on perceived quality is critical. In this paper, we are going to illustrate our approach to understand QoE, focusing on the impact of visual information in D-Cinema.

The paper reports findings while conducting perceptual experiments to study QoE in D-Cinema. We believe that carefully designed experimentations, such as subjective quality assessment, are among important steps toward QoE model development in D-Cinema. We present issues that need to be addressed before conducting subjective quality assessment in D-Cinema environment, and discuss results with more emphasis on the content types and seating positions of participants during experiments. This paper is organized as follow. Section 2 discusses the subjective quality assessment in D-Cinema. Section 3 provides the analysis and processing of the subjective data, and section 4 concludes the paper.

2 Subjective quality assessment in D-Cinema

In 2005 the Digital Cinema Initiative [9] concluded on using JPEG 2000 for compression of Hollywood feature films for the large screen. By this decision DCI initiated a large roll out of digital equipment to cinemas all over the world. Once digital equipment is installed it opens a whole new world for the theater owner to utilize this infrastructure outside of ordinary feature film screening. This paved way for the concept of alternative content, defines as everything else than the feature film and also opening for alternative compression algorithms. Our paper focuses on such alternative content displayed using DCI specified equipment and showed on the large screen in a real theater.

Perceptual experiment design to collect subjective data for understanding overall QoE can be quite complex since the variables can be difficult to identify. In addition, some of these variables also provide cross-contextual and cross-modal effects. For this

reason, it is difficult if not even impossible to design and to conduct only one experiment to resolve the QoE measurement problem.

In the exploration of QoE in D-Cinema, there are many research questions to start from. For example, in our early investigations, some of the research questions that we were interested in pursuing were as follows:

- Which is the best video compression algorithm from the point of view of the cinemagoers?
- What are the impacts of each modality, namely, audio and visual, on the overall quality of D-Cinema presentations?
- Does the perceived quality depend on the content type, and how?

Based on the research question along with its developed hypotheses [4], we first identify dependent and independent variables for perceptual experiments. The dependent variables are the answers we seek from human participants in experiments, while independent variables are the factors whose values are controlled and varied in experiments. Next, the other related variables must be identified and controlled in order to prevent them from becoming confounding variables that taint experimental results.

There are standardized methodologies recommended for perceptual experiments, but they are only applicable in some cases. These recommended methodologies are at times too restrictive as they mostly rely on traditional use cases from telecommunications and broadcasting. The main intention of such methodologies is to collect scores or data, representing the quality level. The most common approaches produce a Mean Opinion Score (MOS) to determine the quality. The MOS value is obtained by averaging quality scores from participants in perceptual experiments. The scores are based on direct scaling techniques using interval scale, e.g., a 5-point scale for quantification of the perceived quality representing 5 levels of quality (poor, bad, fair, good, and excellent). Currently, MOS obtained from subjective visual quality assessment is considered a significant way of representing the visual quality of images or videos. There are sets of established recommended methodologies from ITU to perform subjective visual quality assessment [10, 11]. These recommendations rely on scenarios driven from telecommunication issues in television transmission, when watched on a CRT monitor. Despite this, we will still use the same approach, as the starting point in the design of our own subjective quality assessments in D-Cinema, but apply the necessary modifications to adapt to D-Cinema environment.

2.1 Subjective visual quality assessment

Our experiments are based on Recommendation ITU-R BT.500-11 [10], a widely used standard for subjective quality assessment. The document provides a thorough guideline describing the methodology for the subjective assessment of the quality of television pictures; it includes general viewing conditions, testing environment (laboratory or home), monitor resolution, monitor contrast, test materials selection in terms of anchoring or conditions, test methods, etc. Because our experiments are for D-Cinema,

we cannot literally adopt all the guidelines from ITU-R BT.500 as there are inherent and major differences between television and D-Cinema applications, particularly in the following subjects:

- Viewing conditions

The recommendation provides two alternatives for viewing conditions, i.e.: laboratory environment and home environment. Laboratory environment is intended to provide critical conditions to check the systems; on the other hand home viewing environment is intended to provide a means to evaluate quality at the consumer side of the TV chain. Viewing conditions also include the distance of participants from the display, which is expressed as a function of the projected frame size. There are considerable differences between TV and D-Cinema viewing environments and the screen sizes. Hence, adopting the values related to the viewing conditions and participants distances exactly as stated in the recommendation is not suitable.

- Resolution and contrast

These are related to the required conditions of luminance operating range for subjective assessments. Monitor contrast is strongly influenced by the environment illuminance. However, important to note that conditions described in recommendation BT.500 are based on the use of CRT monitors which are actually far away from D-Cinema. There is significant and obvious differences between stimuli viewed on a CRT monitor and stimuli viewed on a large screen (5 meter x 10 meter) projected by a 2K or 4K projector. For this reason we do not adopt BT.500 recommendations with this regard. However, it is important to remember that illumination, contrast, and resolution issues indeed have influence on the rating of perceived quality; they are considered as the external variables in our experiments. In D-Cinema framework, equivalents can be drawn as variables of screen illumination and screen uniformity. We need to control these external variables since we only want values of independent variables to affect dependent variable—the perceived visual quality. Then, the issue is to measure values of these external variables to confirm whether they are in the range of realistic viewing conditions in D-Cinema. If the measured values are not realistic for D-Cinema applications, we must be able to control them somehow. The installation, regular calibration, and regular maintenance of the 4K projector used in our experiment, were performed by Trondheim Kino AS [12], which indicates that the screen illumination and the screen uniformity values in our experiment were within the range of D-Cinema practices.

- The source signals and test materials.

The reference materials are often original undistorted materials that are considered perfect and used as references. The absence of important distortions in the reference part of the presentation pair is crucial to obtain stable results. Based on the DCI (Digital Cinema Initiatives) specifications [9], D-Cinema is based on 2K or 4K imagery, which is a significantly higher quality in terms of larger pixel counts per image when compared to content mentioned in the recommendation BT.500. However, it is important to note that DCI is a joint venture of major Hollywood studios which primary

purpose is to establish and document voluntary specifications for digitization of all chains in the motion picture industry. Hence, the specifications described in DCI document are tailored to one specific traditional presentation: feature film screening. In our purpose, we also interested in the case beyond feature film. We take into account the opportunity of applying alternative contents in Digital Cinema applications. Thus, we also utilized source signals and test materials based on HD contents.

A number of adaptations were made in our assessment experiments. Subjective quality assessments were conducted in a commercial, DCI-specified D-Cinema theater in Trondheim, Norway. Thus, we believe the viewing conditions in the auditorium of an actual cinema provide realistic and representative viewing conditions. In our experiments, we set up the test as if the participants were watching a feature movie, i.e. in complete darkness. One issue came up in this situation; due to darkness, participants had difficulties marking down their quality ratings on the score sheets provided to them. For this reason, we provided them with clipboards and small intensity reading lamps.

Regarding the test materials and the source signal for our stimuli, we used DCI-specified test images StEM (Standard Evaluation Material) [13] and HD video contents. StEM is a 15-minutes 2K audiovisual test material. We selected nine frames from different type of scene in StEM to be utilized as stimuli of test images. Because we only took into account the luminance component of images in our study, the luminance component was extracted from each image resulting in nine gray scale 2K images. The whole set of test images was then split into a training set of two images, a dummy set of one image, and a testing set of six images. The images used for training and dummy sequence are illustrated in Figure E.1. The set of test images is shown in Figure E.2.

Figure E.1: Training and dummy set of subjective quality assessment of image.

Figure E.2 Test set of subjective quality assessment of image.

HD video contents were also used as test materials in our subjective visual quality assessment of motion pictures because we were also interested in exploring the perceived quality of alternative contents, which in D-Cinema practices are known as ODS (Other Digital Stuff). The data set was taken from the SVT High Definition Multi Format Test Set [14], EBU database [15], NRK [16], and NTIA/ITS database [17]. The dataset format was HD 1920x1080 progressive and converted into 30 fps, YCbCr 4:2:0 format using VirtualDub [18]. The whole set of test sequences was split into a training set of two sequences from the NTIA/ITS database (Aspen and RedKayak), and a testing set of six sequences from SVT (CrowdRun, DucksTakeOff, OldTownCross, IntoTree, and ParkJoy), EBU database (Dancer), as well as a dummy set of one sequence from NRK used for stabilization. The dataset used for training and the dummy sequence are illustrated in Figure E.3. The set of test sequences is shown in Figure E.4.

Figure E.3: Training and dummy set of subjective quality assessment of motion pictures.

Figure E.4: Test set. From top left to bottom right: CrowdRun, Dancer, DucksTakeOff, OldTownCross, IntoTree, and ParkJoy.

The stimuli of the experiments consist of test sequences with varying degree of quality. We selected the compression algorithm as a source of visual degradation because compression is an influential issue in D-Cinema; the quantity of data needed to represent high-quality imagery for D-Cinema in its native uncompressed form is staggering so that D-Cinema cannot become a practical form of business without significantly reducing the quantity of data [19]. There are several standardized compression techniques relevant to D-Cinema. Early experimental deployments have used a bloc-sized DCT-based system from QUALCOMM and a wavelet-based system from QuVis [19].

The most well-known compression standards for images have been developed within the Joint Technical Committee (JTC) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). Two working groups within the JTC are responsible for a great deal of compression technology. The Joint Photographic Experts Group (JPEG) is well known for developing standards for the compression of static images, and the Moving Pictures Experts Group (MPEG) is well known for its standards for compressing video. MPEG-2 is the most widely deployed compression system, being the basis for digital television systems. D-Cinema is not video, but both D-Cinema and video are moving image sequences, and the same techniques are applicable to both, even though the parameters may be very different. Proprietary extensions of MPEG-2 had also been utilized in early deployment of D-Cinema. Currently, there is a latest codec known as H.264/MPEG Advanced Video Codec (AVC) that offers twice the coding efficiency of MPEG-2. This popular motion-compensation-based compression scheme has potential for D-Cinema applications; it can be employed for use beyond traditional feature film screening due to high coding efficiency compare to MPEG-2. JPEG 2000 was created by the Joint Photographic Expert Group (JPEG) committee with the intention of superseding their original discrete cosine transform-based JPEG standard. In 2004, DCI selected JPEG 2000, a wavelet-based compression scheme for still images and image sequences, as the technology choice for digital cinema picture tracks [19].

The subjective image quality assessment was performed by examining a range of JPEG 2000 compressed contents by varying bit rates. Compression of image is generally implemented to ensure meeting transmission bandwidth or media storage limitations. In motion pictures industry practice, image compression much less dependent upon bandwidth or storage requirement, in that way making bit rate more

dependent on desired image quality. DCI system specification selected a maximum bit rates of 250 Mbits/sec [9]. However, since we are also interested in exploring alternative content for D-Cinema applications, we are interested in exploring and studying the perceived quality of bit rates that are significantly lower than bit rates used in motion pictures practice. In the design of a formal subjective test, it is recommended to maintain a low number of evaluation conditions in order to allow human participants an easier completion of their assessment task. Accordingly, 8 different conditions were applied to create 8 processed images from each source image. The selected conditions covered the entire range of quality levels, and the subjects were able to note the variations in perceived quality from each level to the next. This was verified prior to the subjective quality assessment with a pilot test that involved expert viewers in order to conclude the selection of the final bit rates. As a result of the pilot test, the selected bit rates were in the range of 0.01 to 0.6 bits/pixel. To create 48 processed gray scale images, 6 source images were compressed using the KAKADU software version 6.0 [20], with the following settings: codeblock size of 64x64 (default), 5 decomposition levels (default), and switched-off visual frequency weighting.

For the subjective motion pictures quality assessment, two codecs were considered. These were JPEG 2000 and H.264/AVC [21]. Even though H.264/AVC is not used in motion picture industry practice, we wanted to study the utilization of H.264/AVC in D-Cinema environment. For JPEG 2000 coding, the Kakadu version 6.0 was used. One configuration was used for encoding with parameters depicted in Table E.7. For H.264/AVC coding, the JM version 16.1 [22] was used. One configuration was used for encoding with parameters depicted in Table E.8. We also take into account alternative content screening practice, such as live transfer screening in D-Cinema, in which optimizing the bit rate of transferred content is useful factor. Consequently, we believe understanding perceived quality of lower-than-typical bit rates of compression techniques are relevant; this includes the issue how participants perceived the transparency of HD content. For this reason, the selected rate conditions were much lower than the typical bit rates used in feature film screening that reliably produce transparency and designed so that the participants were able to distinguish the differences in perceived quality from each level to the next. Due to the recommendation of maintaining a low number of evaluation conditions so that it is easier for participants to complete the assessment task, we applied four to five different bit rates for both compression techniques from each source of HD content.

Table E.7: JPEG 2000 encoding parameters.

Reference software	Kakadu version 6.0
Codeblock size	64x64 (default)
Decomposition level	5 level (default)
Number of tile	One tile per frame (default)
Visual frequency weighting factor	As recommended for D-Cinema environment [23]

We adopted the simultaneous double stimulus method for presenting the stimulus to participants in subjective image quality assessment and the single stimulus with hidden reference method [10] for presenting the stimulus to participants in subjective

motion pictures quality assessment. Figure E.5 illustrates the presentation method and scale used.

Table E.8: H.264/AVC encoding parameters.

Reference software	JM 16.1
Profile	High (FREXT Profile)
Number of frames	300
Chroma format	4:2:0
GOP structure	IBPBPBPBPBPBP
Number of reference frames	2
Slice mode	off
Rate control	Enabled (initial QP=30)
Macroblock partitioning for motion estimation	Enabled
Motion estimation algorithm	Fast full search (default)
Early skip detection	Disabled
Selective intra mode decision	Disabled

Although the cinema could obviously accommodate all 20 participants at once, we designed the experiment in order to allocate only five participants per session. The main reason was to avoid influence of two additional factors: the distance of participants to the cinema screen, and the viewing angle. We chose the viewing distance (10 meters) to be 2 times the height of the screen. This resulted in subjects being placed in the 6th row. In order to maintain a centralized viewing condition for all participants, only 5 seats were allocated for participants in this row. The participants' exact position during the experiments is illustrated in Figure E.6 and Figure E.7.

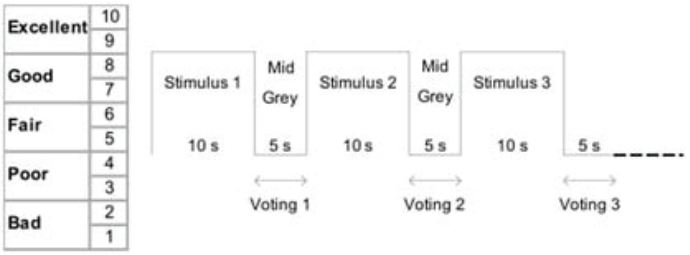


Figure E.5: Presentation method and scale.

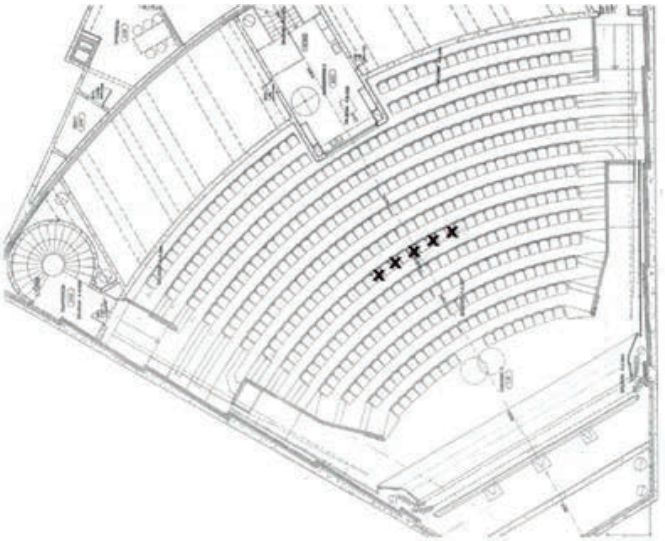


Figure E.6: Participants' position at the 6th row.

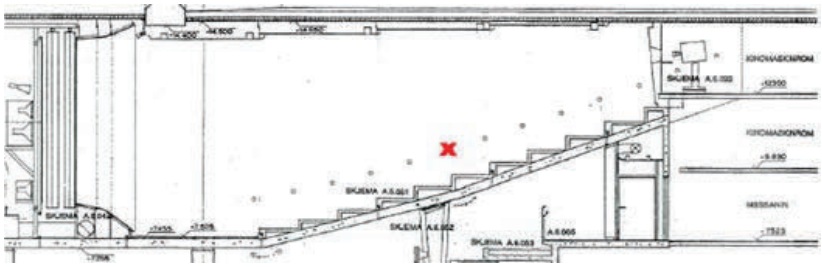


Figure E.7: Participants located at the 6th row from the screen.

3 Processing and analysis of subjective data

3.1 Processing subjective data

In order to have a meaningful result from the scores collected from participants in the experiments, they should be processed based on principle derived from statistics. The data collected from subjective quality assessment can be described by the Eq. (1)

$$m_{i,j} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j}$$
$$\varepsilon_{i,j} \sim N(0,\sigma^2) \quad (1)$$

Here, $m_{i,j}$ is a quality score obtained from participant i after scoring stimulus j; μ is the overall mean of collected scores across all participants and all stimuli; α_i is the participant effect; β_j is the specific treatment or stimulus j effect; $\varepsilon_{i,j}$ represents the experimental error.

Based on this model, it is not wise to estimate the MOS (Mean Opinion Score) by averaging raw scores—the exact scores obtained from all participants—of each stimulus because we disregard the participant effect α_i (e.g., how each participant uses the quality scale to score, etc.) and the experimental error $\varepsilon_{i,j}$. Figure E.8 and Figure E.9 shows how participants' scores can vary. Consequently, we believe that it is important to statistically process the scores before estimating the MOS. Figure E.10 illustrates the stages that were performed on our data. The first stage of subjective data analysis is a process called descriptive analysis [4]. The purpose of descriptive analysis is to check the statistical assumptions, such as the data distribution and the homogeneity of variance across tested combinations of independent variables.

3.1.1 Processing of scores in subjective image quality assessment

We observed the skewness value and compared it to its standard error, in which skewness value of the data (0.048) is not greater than twice the value of its standard error (0.066); it indicated the data distribution is not departing from normal distribution [24].

Based on the model given in Eq. (1), it is relevant to analyze the effect on scoring by each participant. In other world, checking whether there are significant differences between the ways participants used the quality scale when scoring stimuli. This can be achieved using a parametric test like ANOVA (analysis of variance). However, ANOVA is only suitable for data that is normally distributed and whose variance is homogeneous.

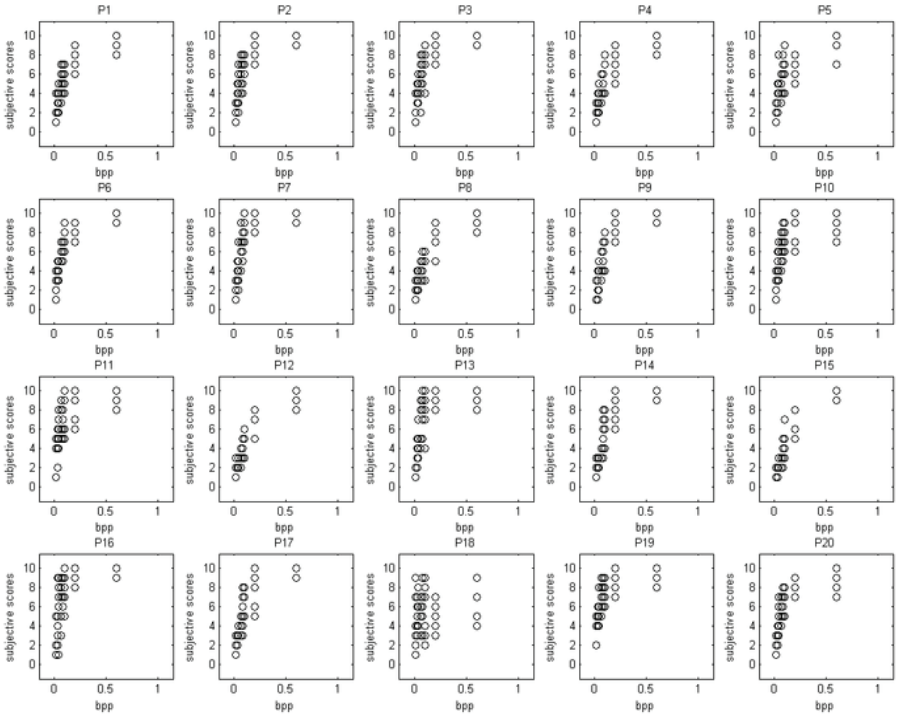


Figure E.8: Illustration of scores variations among twenty participants in subjective quality assessment of still images.

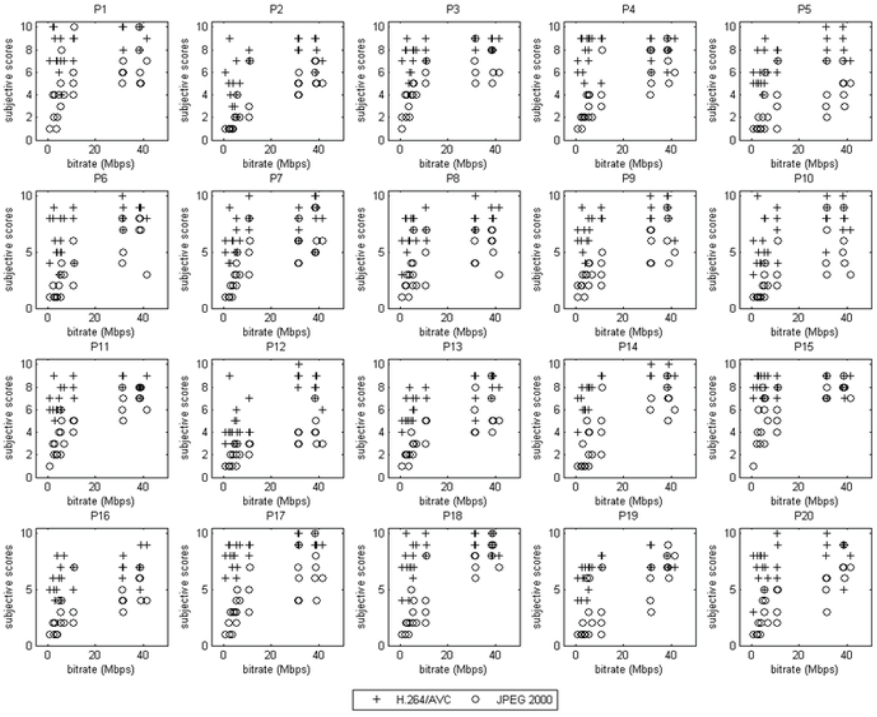


Figure E.9: Illustration of the scores variation among twenty participants in subjective quality assessment of motion pictures.

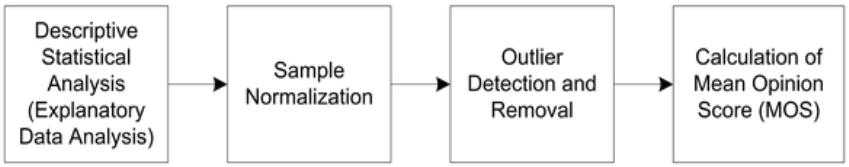


Figure E.10: Process stage of data analysis.

We performed Levene test to check the homogeneity of variances of the scores. The result [$F(28,1363)=2.246$, $p<0.05$] showed that the null hypothesis in which the variances are homogeneous is rejected. This indicates that the variance of scores collected from the experiments was not homogeneous.

We can use the non-parametric Kruskal-Wallis analysis of variance instead. This test was performed on the raw scores across the participants (the participant is the independent variable). The differences between participants were significant, with $H(28) = 85.089$, and $p<0.05$. This indicates that indeed there were significant variations

of subjective scores among participants which can be caused by differences in the way participants used the quality scale to judge the quality of the stimuli. We can also perform alternative parametric test Welch and Brown-Forsythe test [25] which do not require a homogeneous variance assumption to indicate the significant variations between participants scores. The result of Welch test [$F(28)= 3.098$, $p<0.05$] and Brown-Forsythe test [$F(28) = 3.123$, $p<0.05$) confirm the result of the non-parametric test.

Consequently, a participant-to-participant correction was applied by normalizing all the scores according to an offset mean correction [26], which is given by Eq. (2) and Eq. (3).

$$\hat{m}_{ij} = (m_{ij} - \hat{b}_i) \tag{2}$$

$$\hat{b}_i = \frac{1}{J}\sum_{j=1}^{J} m_{ij} - \mu \tag{3}$$

$\hat{m}_{ij}$ is the normalized score of participant i for stimulus j; m_{ij} is the score of subject i for stimulus j; $\hat{b}_i$ is the corrected offset for participant i; μ is the average score across all participants and all stimuli.

An outlier detection was performed according to the guidelines described in section 2.3.1 of annex 2 of recommendation ITU-R BT. 500-11 [10]. One outlier was detected out of 29 participants.

3.1.2 Processing of scores in subjective quality assessment of motion pictures

We used a Shapiro-Wilk test to verify the normality of distribution of scores obtained from participants. The computed p-value was equal to 0, which indicates that the distribution was not normal. We checked the skewness value (0.439) to verify the result of Shapiro-Wilk test, and it was larger than twice of its standard error value (0.070) that indicates indeed a non normal distribution.

We used a non-parametric Kruskal-Wallis analysis of variance [24] instead of parametric test since the normality distribution assumption of the subjective scores was rejected.

The differences between participants were significant, with $H(19) = 78.354$, and $p<0.05$. This indicates that indeed there were large variations in means of subjective scores among participants, i.e. there were significant differences between the ways participants used the rating scale to judge the quality of the stimuli. Figure E.9 illustrates these differences. Consequently, a participant-to-participant correction was applied by normalizing all the scores according to offset mean correction, given by Eq. (2).

Then, an outlier detection was applied to the normalized scores according to the guidelines described in section 2.3.1 of annex 2 of recommendation ITU-R BT. 500-11 [10]. Two outliers were detected out of 20.

3.2 Procedures for estimation of mean opinion scores

MOS was calculated by averaging the scores of each stimulus from participants who were not outliers. The MOS with its 95 % confidence interval for all stimuli from subjective image quality assessment are shown in Figure E.11. We investigated if there were significant differences between different stimuli (bpp values) in order to check if the differences between bpp values were statistically meaningful. Levene's test results [$F(7)=6.5$, $p<0.05$] indicate that the variances of the data samples are not homogeneous. Because of the Levene's test results, we used an alternative analysis of variance Welch and Brown-Forsythe test. The results of Welch test [$F(7)=662.6$, $p<0.05$] and Brown-Forsythe test [$F(7)=561.62$, $p<0.05$] indicate that there are indeed significant differences, and these significant differences exist between scores for all bpp values as shown by post-hoc analysis using Games-Howell test. The MOS with its 95 % confidence interval for all stimuli for two different types of compression technologies— JPEG 2000 and H.264/AVC— from subjective motion pictures quality assessment are shown in Figure E.12. Using non-parametric test (Kruskal-Wallis test), the results show there are significant differences on the collected subjective data between two compression techniques [$H(1) = 406.213$), $p<0.05$]; Figure E.13 also supports the results of Kruskal-Wallis test.

There is a high expectation demanded from D-Cinema applications; they deliver high quality multimedia presentation to the cinemagoers. Consequently, the issue of transparency in the perceived quality of visual presentation is important. With regard to the subjective scores obtained from subjective visual quality assessment, it is reasonable to assert that the transparency of visual stimulus is reached when the users perceived the visual quality as excellent which is expressed in the score of 9 or 10. In the subjective image visual quality assessment, scores of all references (uncompressed contents) reach transparency region, which was expected. We observed that for images compressed at 0.6 bits per pixel, the transparency region is reached consistently. As shown in Figure E.11, the higher the JPEG 2000 bit rates, the higher MOS value is reached, which was expected, even though the MOS values differed for different contents. For Image 3 and 5, the MOS of reference image are lower than the MOS of images compressed at 0.6 bits per pixel though the scores are still within the transparency region. However, if we look closely how the MOS differences between reference and 0.6 bpp rate of these two images, the subjective scores of these two conditions are overlap within 95% confidence interval, which suggested that the MOS differences are not significant.

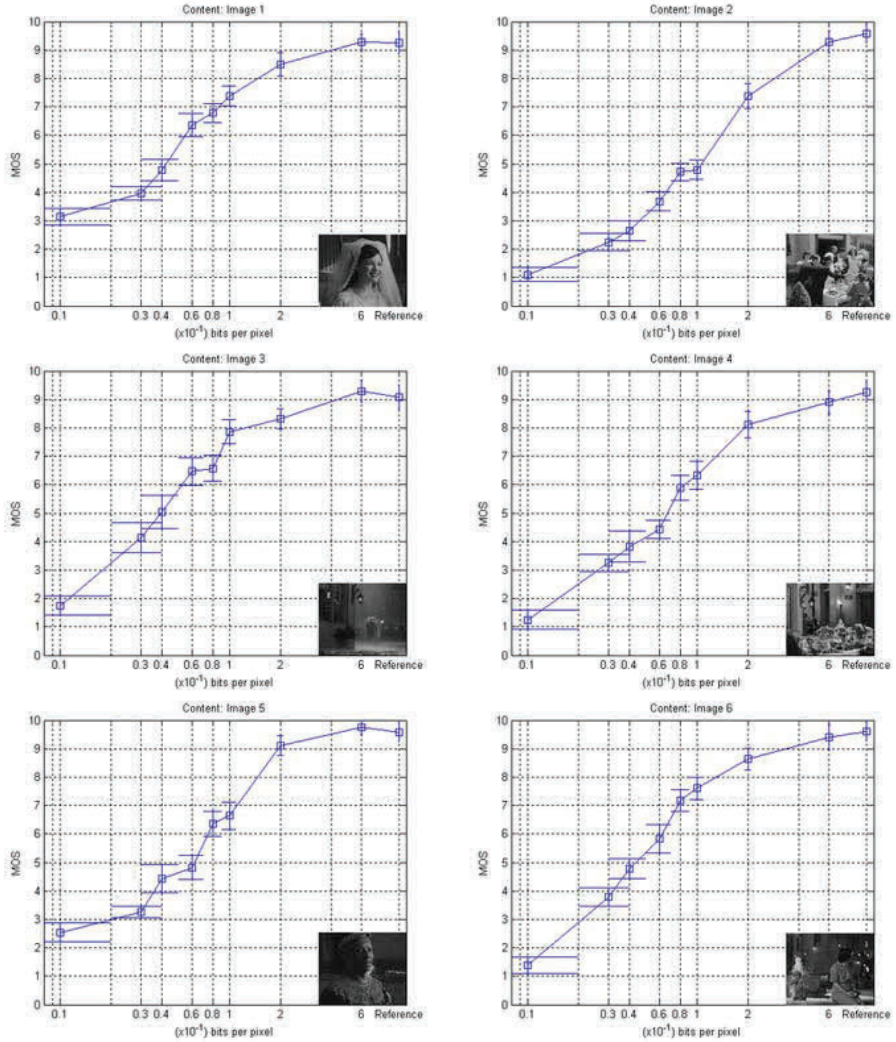


Figure E.11: Computed MOS of each stimulus with its 95 % confidence interval from subjective image visual quality assessment.

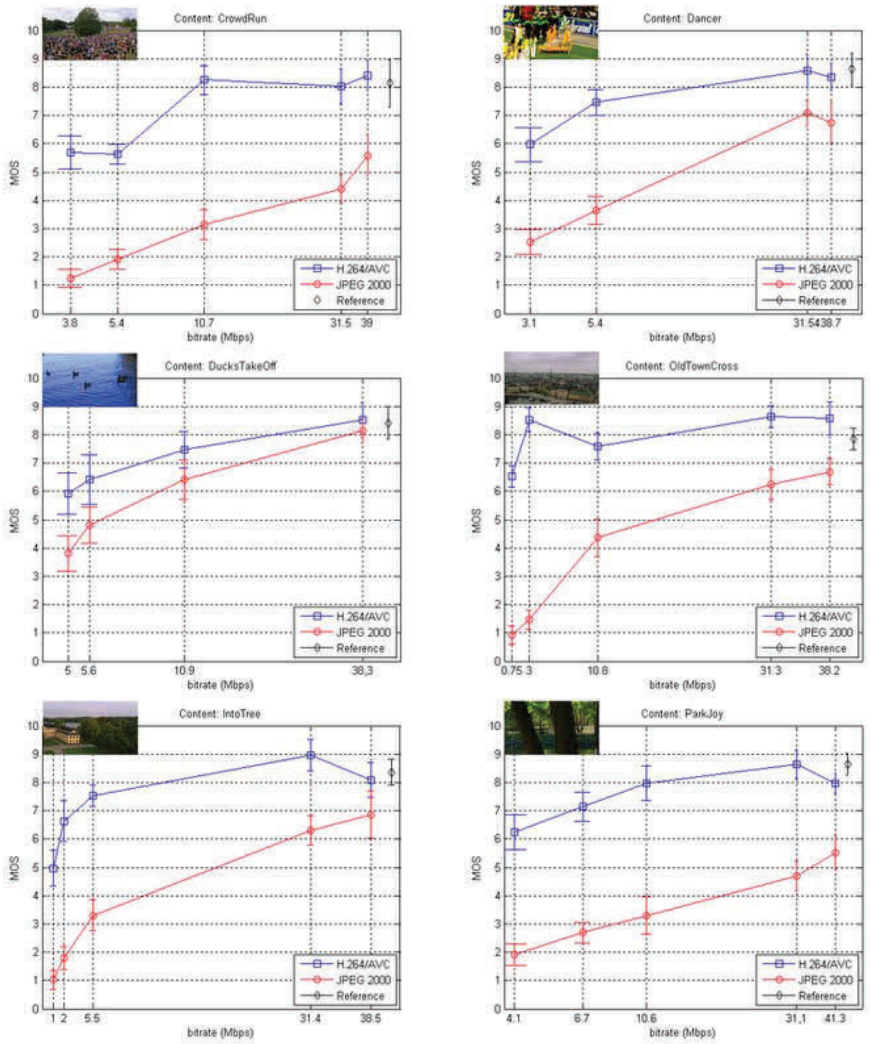


Figure E.12. Computed MOS of each stimulus with its 95 % confidence interval from subjective visual quality assessment of motion pictures.

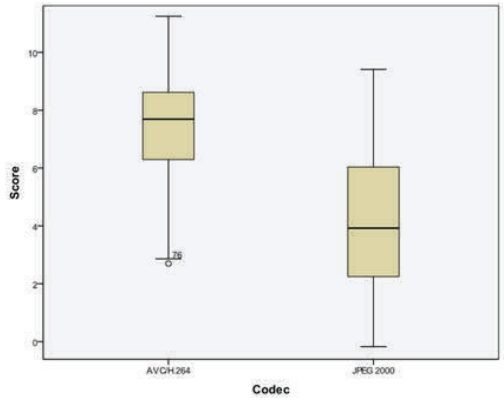


Figure E.13: The boxplot of scores from subjective quality assessment of motion pictures in D-Cinema grouped by different codecs.

In the subjective visual quality assessment of motion pictures, the behavior of the collected subjective scores is rather distinctive, notably in the following issues:

1. Transparency

The transparency region of 9 or 10 score is not reached; there is no excellent quality representation of the obtained MOS from all compressed condition—both for JPEG 2000 and AVC/H.264. In spite of this, it is vital to see that participants did not judge the quality of the reference (uncompressed) conditions, which are transparent contents, as excellent. It appears that the participants experienced high ambivalence utilizing the highest end of quality scale when judging the quality of the stimulus, which is different with what occurred at the subjective image quality assessment. The significant difference between the subjective image quality assessment and the subjective motion pictures quality assessment beyond the stimulus factor is the presentation test method in the experiments. In the subjective image quality assessment, we used simultaneous double stimulus in which the stimulus was always displayed alongside the uncompressed sequence as seen in Figure E.15. Consequently, the participants always had comparison of the reference when judging the quality of stimulus which made them use the highest end of quality scale because the participant know what to expect regarding the excellent quality of visual presentation. On the other hand, we used single stimulus with hidden reference method in the subjective quality assessment of motion pictures. The participants were not informed that there were some uncompressed (perfect) sequences during the experiments; they were only asked to judge the quality of short sequence of HD presentation series. Due to the absence of reference point during the judgment of each stimulus, the participants possibly would use their own conceptual reference point from their expectation and experience from watching feature films in the cinema which resulted in their hesitancy to give the highest score during viewing sequence. MOS values of the uncompressed transparent contents which in theory should attain excellent quality (score of 9 or 10) only reached the good quality (score of 8). Accordingly, we can infer that the level of transparency region obtained from our subjective motion pictures quality assessment is lower compared to the level obtained

from our subjective image quality assessment; in motion pictures case score of 8 is already representing the value of transparent content.

2. Compression algorithm
It can be seen that for the same bit rate value for each test sequence, H.264/AVC encoded sequences received a higher MOS compared to the JPEG 2000 encoded ones. This is apparent for all selected bitrates of all contents. This result was anticipated, since only the H.264/AVC algorithm employs motion estimation. Motion estimation provides a considerable level of temporal compression that is capable of providing significant improvement in coding gain without loss of perceived quality. Consequently, there is a significant impact on the coding gain of H.264/AVC compared to the coding gain of JPEG 2000. However, making a final conclusion on which is the best compression algorithm providing the better performance solely based on the MOS as computed from the subjective data in this experiment is not possible. Such a comparison would have the nature of an apple and orange comparison. But it is important to note that the MOS of H.264/AVC indicates it is possible to employ a compression algorithm that take into account temporal compression such as H.264/AVC for presentation in D-Cinema, i.e.: for ODS application, such as live transfer event screening in D-Cinema in which high coding gain is a more appealing feature.

3. Perceived quality
There are repeated patterns shown in the higher end spectrum of perceived visual quality in the contents of Dancer, IntoTree, and ParkJoy. The highest MOS of perceived visual quality are not obtained by the highest bit rates of H.264/AVC. In the content of IntoTree, the uncompressed content did not even get the highest MOS for its visual quality. We knew that the uncompressed content is a transparent content, and these results happened in the transparency region (score of 8 or above) where the MOS of reference contents were reached. One possible explanation is that up to certain bit rate (especially in the context of H.264/AVC), saturation, in which the participants could not differentiate any visual quality differences among test conditions, is happening due to transparency; the participants did not perceive any artifacts from compression error. In addition, the lack of using the highest end quality scale during judgment cause the MOS score to not extend to the highest values. The saturation could also explain the pattern shown beyond Dancer, IntoTree, and ParkJoy, i.e.: in CrowdRun content, saturation is even reached at 10.7 Mbps. In the content of OldTownCross, in which the uncompressed even get MOS of lower than 8, this type of pattern even already arrived at 3 Mbps.

3.4 Content types

From Figure E.11 and Figure E.12 it can be observed that there are different MOS values for different stimuli (bpp or bit rate value). It is essential to further investigate the reasons behind these differences and understand their origins. Therefore, here, we investigate whether there are significant differences in subjective quality scores for different contents. The parametric test can be used to test these differences. First, we need to check whether the data distribution of the corrected score is normal. Skew value (0.02) of the corrected score is less than twice its standard error (0.07) which indicates

that we can apply a parametric test on our corrected scores since the distribution is not seriously departing from normal. Levene's test on our data [$F(5)=2.7$, $p<0.05$] indicates the variances are heterogeneous. Because of the Levene's test results, instead of using ANOVA to test the differences of scores between image types, we use alternative analysis of variance, Welch and Brown-Forsythe test. The results of Welch test [$F(5)=15.9$, $p<0.05$] and Brown-Forsythe test [$F(5)=16.16$, $p<0.05$] indicate that there are indeed significant differences of scores for different contents. The post-hoc analysis of Games-Howell shows that scores from Image 2 and Image 4 are significantly different in other images.

There are different characteristics in the selected content types in our experiments, and this may have influenced how subjects perceived the quality in different images. ITU-T Recommendation P.910 [11] which contains recommendations for subjective quality assessment methods for multimedia applications states that particular characteristics—spatial and temporal perceptual information of the scenes—are critical parameters. Spatial perceptual Information (SI) is based on the Sobel filter, which is used for edge detection. Each luminance value of the frame is processed by Sobel filter. Then, standard deviation in each filtered frame is computed. The maximum value in the time series is the SI value. Temporal perceptual Information (TI) is based on the motion difference feature which is the difference between the luminance values of successive frames. The TI of the content is computed as the maximum standard deviation of the differences over time. The measured SI values of our test images are illustrated in Figure E.14.

Based on the calculated SI, we can categorize our image contents into low SI (SI < 50), which consists of Images 1, 3, 5, and 6, and high SI (>50), which consists of Image 2 and 4. Hence, the spatial complexity of the content indicates the reason behind the significant differences of scores of Images 2 and 4, when compared to other images.

We also performed the same spatial and temporal analysis on the motion pictures stimulus from our subjective quality assessment of motion pictures. The spatial information (SI) and temporal information (TI) of the motion pictures are also illustrated in Figure E.14.

We also would like to test whether there are differences on the subjective score among different motion pictures. The analysis was performed on the corrected scores. The results of Shapiro-Wilk test ($p<0.05$) and distribution skew value, in which the skew value (0.4) is more than twice its standard error (0.16), indicate that the corrected scores are not normally distributed. Consequently, using parametric test is not suitable. The results of Kruskal Wallis test show that there are significant differences of the scores between different contents [$H(5)=23.33$ and $p<0.05$].

Based on the calculated SI and TI indexes, the test material is divided into 4 category—Low SI and Low TI, High SI and Low TI, High SI and High TI, and Low SI and High TI, by dividing the area illustrated Figure E.14 into four separate regions. Using non parametric testing, the Kruskal Wallis test, there are significance differences on the collected subjective data between different categories [$H(3)=8.049$ and $p<0.05$].

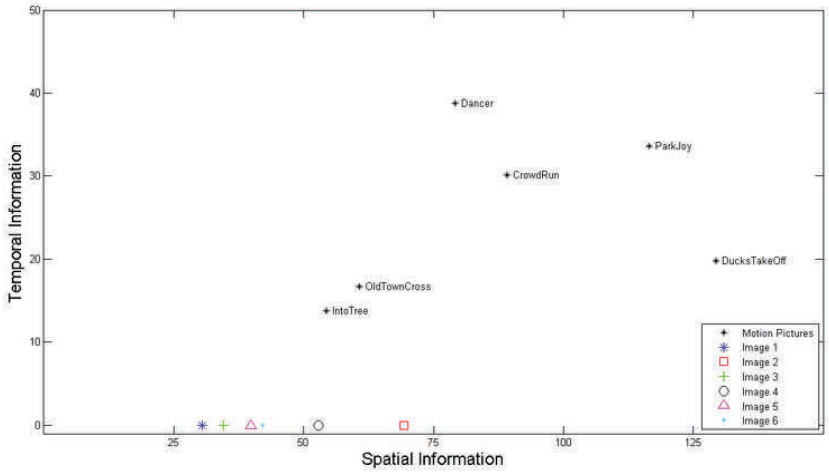


Figure E.14: Spatial information and temporal information of the test sequences of subjective visual quality assessment.

Figure E.15: Presentation of the stimulus in the subjective image quality assessment.

3.5 Participant's position

In our image quality assessment in D-Cinema, we used simultaneous double stimulus method: the reference image was always presented with the processed image as illustrated in Figure E.15. The participants had to judge the quality of the processed image on the right hand side of the screen and compared it to the reference image on the left hand side of the screen. In the human visual system (HVS), eye movement is typically divided into fixation and saccade. Fixation is the maintenance of the visual gaze on a single location. Saccade refers to a rapid eye movement. Humans do not look at a scene in fixed steadiness, instead, the human fovea sees only the central 2º of visual angle in the visual field and fixed on this target, then moves to another target by saccadic eye movements [27]. Saccades to an unexpected stimulus normally take about 200 milliseconds to initiate, and then last about 20-200 milliseconds, depending on their amplitude (20-30 milliseconds is typical in language reading). In image quality

assessment, quality evaluation is taking place during eye fixation when the fovea can perceive the visual stimulus with maximum acuity [28]. Thus, when viewing a picture on a large screen in the D-Cinema, participants cannot see an entire image at once and evaluate distortions in all regions in this picture. Due to these factors (HVS characteristic and the stimulus presentation method) the viewing angle of the participants can have an impact on the quality judgment.

The seat location of the participants during experiments is a factor that determines the viewing angle of participants. Hence, position of the participants can have an effect on the subjective scores. Figure E.16 illustrates the box plot of the subjective scores from our image quality assessment in the cinema grouped by the position of the participants.

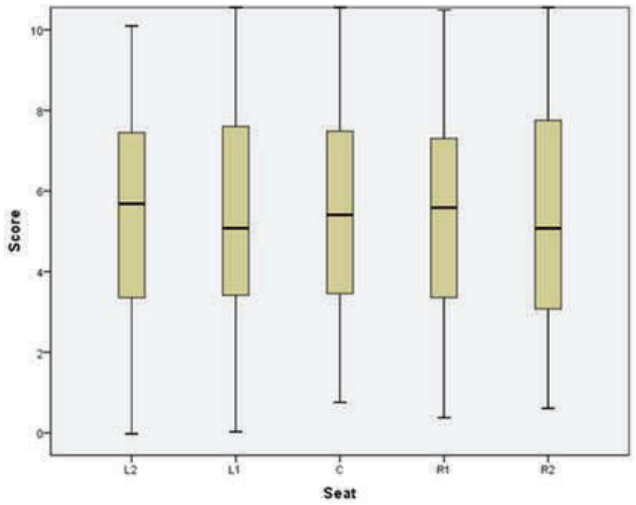


Figure E.16: The scores of subjective image quality assessment in D-Cinema grouped by 5 different positions of the participants.

Figure E.16 indicates that the participants' positions do not have significant impact on the collected quality score. The results of Welch [F(4)=0.434, p=0.789] and Brown-Forsythe [F(4)=0.430, p=0.787] tests support this observation. Further analysis, using two-way ANOVA with participant position and content type as dependent variables also shows that the position does not have significant impact on the subjective scores [F(4)=0.450, p=0.773], and so does the interaction between content type and position [F(20)=0.234, p=1].

Figure E.17 which shows the box plot of the scores from subjective quality assessment of motion pictures grouped by participant's position also indicates that the position of the participants do not have a significant influence on the quality scores. Further analysis using non parametric testing, the Kruskal Wallis test, [H(4)=0.348 and p=987] confirms the same.

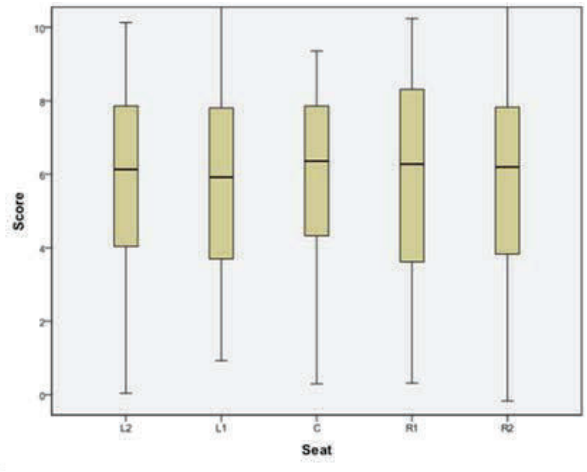


Figure E.17: The scores of subjective quality assessment of motion pictures in D-Cinema grouped by 5 different positions of the participants.

4 Conclusion

QoE always puts the end-user at the centre of attention and it is a multidimensional concept, which consists of several objective and subjective parameters. This contributes to the difficulty of quantifying QoE. In this paper we focus on the subjective quality assessment for D-Cinema application because we believe it is an important aspect in studying QoE for D-Cinema; it is the basis to understand the perceived quality and is useful for developing a mature QoE model for D-Cinema. For this reason, subjective quality assessment for D-Cinema applications must be carefully designed. We applied existing recommendations methodology for the assessment and made several proper adjustments. Our study showed that due to the different and unique digital image content and viewing conditions of D-Cinema, quality research of D-Cinema especially in the context of QoE is not really in the same category as other applications. In our study, we focus on the alternative content scenario in D-Cinema. Consequently, it influences the experimentation of selected parameters for stimuli. Before MOS value is computed, thorough statistical processing of obtained subjective data is conducted. The results show that content type is a significant factor that influences the perceived visual quality for both image and motion pictures quality. In our assessment of motion pictures, we also showed the result of the differences between two compression algorithms. Initial impression of our study showed that the stimulus presentation method influenced how participants used the quality scale when judging the perceived visual quality. Participants seem confident using the highest end of quality scale when judging the transparent stimuli when simultaneous double stimulus method was employed during subjective visual quality assessment of images, on the other hand participants showed hesitancy using the highest end of quality scale when judging the transparent stimuli when single stimulus method was employed during subjective visual quality assessment of motion pictures.

QoE is affected by several factors including perception, sensations, and expectations of users as they consume digital content presented to them using their perceptual sensors. In D-Cinema the main perceptual sensors are sight and hearing; consequently, perceived audiovisual quality in D-Cinema are integral in QoE research. As an extension of this work we performed subjective audiovisual quality assessment to study the mechanisms of QoE in D-Cinema [29]

5 References

[1] J.-R. Ohm, *Multimedia Communication Technology: Representation, Transmission and Identification of Multimedia Signals*: Springer, 2004.

[2] R. Jain, "Quality of Experience," *IEEE Multimedia,* vol. 11, pp. 95-96, 2004.

[3] ITU-T, "Vocabulary for performance and quality of service Amendment 1: New Appendix I -- Definition of Quality of Experience (QoE)," ITU, Geneva2006.

[4] S. Bech and N. Zacharov, *Perceptual Audio Evaluation--Theory, Method and Application*. England: John Wiley & Sons Ltd, 2006.

[5] W. Hoeg, L. Christensen, and R. Walker, "Subjective assessment of audio quality -- the means and methods within the EBU," European Broadcasting Union, Geneva1997.

[6] T. Oelbaum, V. Baroncini, T. K. Tan, and C. Fenimore, "Subjective Quality Assessment of the Emerging AVC/H264 Video Coding Standard," presented at the International Broadcasting Convention, Amsterdam, 2004.

[7] D. S. Hands, "A Basic Multimedia Quality Model," *IEEE Transactions on Multimedia,* vol. 6, December 2004 2004.

[8] B. Shneiderman, C. Plaisant, M. Cohen, and S. Jacobs, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 5 ed.: Addison-Wesley, 2010.

[9] DCI, *Digital Cinema System Specification version 1.2*: http://www.dcimovies.com, 2008.

[10] ITU-R, *Methodology for the subjective assessment of the quality of television pictures*. Geneva: Tech. Rep. ITU-R BT.500-1, 2002.

[11] ITU-T, *Subjective video quality assessment methods for multimedia applications*. Geneva: Tech. Rep. ITU-T P.910, 09/99.

[12] T. K. AS. Available: http://www.trondheimkino.no/article42891.ece

[13] DCI, *Digital Cinema Initiatives StEM Access Procedures*: http://www.dcimovies.com/StEM/.

[14] SVT. (2006). *The SVT High Definition Multi Format Test Set*. Available: http://tech.ebu.ch/docs/hdtv/svt-multiformat-conditions-v10.pdf

[15] EBU. Available: http://www.ebu.ch/fr/technical/hdtv/test_sequences.php

[16] NRK. Available: http://nrk.no/about/

[17] NTIA/ITS. Available: ftp://vqeg.its.bldrdoc.gov/HDTV/NTIA_source/

[18] VirtualDub. Available: http://www.virtualdub.org

[19] P. Symes, "Compression for Digital Cinema," in *Understanding Digital Cinema: A Professional Handbook*, ed: Focal Press, 2005, pp. 121-148.

[20] Kakadu. Available: http://www.kakadusoftware.com/

[21] ITU-T, "Advanced video coding for generic audiovisual services," 2005.

[22] JM. Available: http://iphome.hhi.de/suehring/tml/

[23] D.-C. AHG, "Guidelines for digital cinema applications," JPEGOctober 2009 2009.

[24] H. Coolican, *Research methods and statistics in psychology*. London: Hodder Arnold, 2004.

[25] J. F. Reed and D. B. Stark, "Section I. Methodology: Robust alternatives to traditional analysis of variance: Welch W*, James J_I*, James$_{II}$*, Brown-Forsythe BF*," *Computer Methods and Programs in Biomedicine, Elsevier Science Publishers B.V. ,* vol. 26, pp. 233-238, 1988.

[26] E. D. Gelasca, "Full-reference objective quality metrics for video watermarking, video segmentation and 3D model watermarking," Ph.D, EPFL, Lausanne, 2005.

[27] G. A. Carpenter, *Movements of the eyes*. London: Pion, 1988.

[28] D. Burr and M. C. Morrone, "Selective suppression of the magnocellular visual pathway during saccadic eye movements," *Nature 371 (6497),* pp. 511–513, 1994.

[29] F. N. Rahayu, U. Reiter, J. You, A. Perkis, and T. Ebrahimi, "Visual Subjective Quality Assessment with the Presence of Audio Stimulus in Digital Cinema," presented at the Proceeding of Third International IEEE Workshop on Quality of Multimedia Experience, Mechelen, Belgium, 2011.

Appendix A: Interfaces Supported by Cinema Projectors

Table I: The Interfaces Supported by SRX-R210 Projector

Resolution	Remarks
1024 x 768 at 60 Hz (XGA)	VESA
1280 x 960 at 60 Hz (SXGA1)	VESA
1280 x 1024 at 60 Hz (SXGA2)	VESA
1400 x 1050 at 60 Hz (SXGA+)	VESA
1600 x 1200 at 60 Hz (UXGA)	VESA
2048 x 1080 at 60 Hz (DC)	
1920 x 1080 at 24 Hz (HD)	
2048 x 1080 at 24 Hz (DC)	
1920 x 1200 at 59.95 Hz Reduced Blanking (WUXGA)	VESA
1920 x 1080 at 60 Hz (HD)	EIA/CEA-861B
2048 x 1080 at 48 Hz (DC)	

Table II: The Interfaces Supported by CP2230 Projector

Source Standard	Original Source Resolution	Vertical Frequency (Hz)	Scan Type	Display Frame Rate (Hz)
SMPTE 296M	1280 x 720	23.98/24	Progressive	23.98/24
SMPTE 296M	1280 x 720	25	Progressive	25
SMPTE 296M	1280 x 720	29.97/30	Progressive	29.97/30
SMPTE 296M	1280 x 720	48	Progressive	48
SMPTE 296M	1280 x 720	50	Progressive	50
SMPTE 296M	1280 x 720	59.94/60	Progressive	59.94/60
SMPTE 296M	1280 x 720	100	Progressive	100
SMPTE 296M	1280 x 720	120	Progressive	120
SMPTE 274M	1920 x 1080	23.98/24	Progressive	23.98/24
SMPTE 274M	1920 x 1080	25	Progressive	25
SMPTE 274M	1920 x 1080	29.97/30	Progressive	29.97/30
SMPTE 274M	1920 x 1080	48	Progressive	48
SMPTE 295	1920 x 1080	50	Progressive	50
SMPTE 274M	1920 x 1080	59.94/60	Progressive	59.94/60

	Original Source Resolution	Vertical Frequency (Hz)	Scan Type	Display Frame Rate (Hz)
SMPTE 274M	1920 x 1080	23.98/24	Interlaced	11.99/12
SMPTE 274M	1920 x 1080	25	Interlaced	12.5
SMPTE 274M	1920 x 1080	29.97/30	Interlaced	14.985/15
SMPTE 274M	1920 x 1080	48	Interlaced	24
SMPTE 295	1920 x 1080	50	Interlaced	25
SMPTE 274M	1920 x 1080	59.94/60	Interlaced	29.97/30
SMPTE 274M	1920 x 1080	100	Interlaced	50
SMPTE 274M	1920 x 1080	120	Interlaced	60
SMPTE RP 211	1920 x 1080	23.98/24	Progressive (sF)	23.98/24
SMPTE RP 211	1920 x 1080	25	Progressive (sF)	25
SMPTE RP 211	1920 x 1080	29.97/30	Progressive (sF)	29.97/30
	640 x 480	23.98/24	Progressive	23.98/24
	640 x 480	25	Progressive	25
	640 x 480	29.97/30	Progressive	29.97/30
	640 x 480	48	Progressive	48
	640 x 480	50	Progressive	50
	640 x 480	59.94/60	Progressive	59.94/60
	640 x 480	100	Progressive	100
	640 x 480	120	Progressive	120
	720 x 525	23.98/24	Interlaced	11.99/12
	720 x 525	25	Interlaced	12.5
	720 x 525	29.97/30	Interlaced	14.985/15
	720 x 525	48	Interlaced	24
	720 x 525	50	Interlaced	25
	720 x 525	59.94/60	Interlaced	29.97/30
	720 x 525	100	Interlaced	50
	720 x 525	120	Interlaced	60
DCI	2048 x 1080	24	Progressive	24
DCI	2048 x 1080	48	Progressive	48