8-1-2012

Audience Image Preferences of 1080i and 720p High Definition Video Formats: An Experiment

Michael Thomas Stubben
University of Nevada, Las Vegas, michaelstubben@gmail.com

AUDIENCE IMAGE PREFERENCES OF 1080i AND 720p HIGH DEFINITION VIDEO FORMATS: AN EXPERIMENT

by

Michael Stubben

Bachelor of Arts
University of Nevada, Las Vegas
2005

A thesis submitted in partial fulfillment
of the requirements for

Master of Arts in Journalism and Media Studies

Hank Greenspun School of Journalism and Media Studies
Greenspun College of Urban Affairs
The Graduate College

University of Nevada, Las Vegas
August 2012

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Michael Stubben

entitled

Audience Image Preferences of 1080i and 720p High Definition Video Formats: An Experiment

be accepted in partial fulfillment of the requirements for the degree of

Master of Arts in Journalism and Media Studies
School of Journalism and Media Studies

Paul Traudt, Committee Chair

Gary Larson, Committee Member

Lawrence Mullen, Committee Member

Tara Emmers-Sommer, Graduate College Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

August 2012

ABSTRACT

A Study on Audience Preference With Regard to the Different High Definition Video Formats 1080i and 720p

by

Michael Stubben

Dr. Paul Traudt, Thesis Committee Chair
Associate Professor of Journalism and Media Studies
University of Nevada, Las Vegas

The two main kinds of high definition television video formats broadcast in the United States are 1080i and 720p. Both formats are claimed to have advantages and disadvantages relative to the other format. However, there has been little academic research to determine viewer preferences of these formats, and whether the stated claims about each format's strengths and weaknesses are valid. This thesis addresses this lack of research by employing an experiment to determine viewer perceptions and preferences of these two high definition formats.

The two major findings from this study pertain to viewer preferences of 1080i/720p with regard to motion, and which format is preferred at night. Statistical analysis of this study's results contradict the often accepted notion 720p is the better format to record and display video involving motion. This experiment's results showed viewers have a significant preference for 1080i when viewing video involving high in-frame movement. It was also determined viewers prefer the 720p video format to a significant level when viewing video recorded at night. Other trends such as 720p being the overall preferred high definition television format were found, but not to a statistically significant level.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my committee chair, Dr. Paul Traudt. Your support, encouragement and guidance continually helped me throughout my writing, research and studies. Your enthusiasm, patience, motivation, vast knowledge and sense of humor were of great benefit to this paper, to my research, and to me personally. I could not have picked a better thesis advisor.

I would also like to thank my other committee members, Drs. Gary Larson, Lawrence Mullen, and Tara Emmers-Sommer. Your commitment to my work, encouragement, insightful comments and tough questions made this thesis stronger.

In addition, I would like to thank UNLV-TV. Without access to their facilities I would not have been able to design and run this experiment. Without the support of engineer, Dan Grimes, this thesis would not have been possible. Dan, your willingness to help me made this experiment a reality, while your kind, sincere and friendly personality made it a joy to work with you. I am very lucky you chose to work at UNLV and that you took a vested interest in my work. In addition, I would like to state my appreciation of the UNLV librarians who skillfully assisted in my access of isolated resources.

Last but certainly not least, I would like to thank my family for their continued support. Chuck, thanks for uttering those three fateful sentences that kept this thesis from sputtering and stalling near its midpoint. Fritten, thanks for always believing in me and answering the phone whenever I needed a kind, sincere and intelligent friend. Ditto to you Jorn. Debbie you are a rock star! Thank you tremendously for the late nights, kind words, encouragement and love; I could not have done this without you.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

American broadcast television has experienced a major transformation since August 14, 1994 when the Federal Communication Commission (FCC) approved the Grand Alliance's field tests of High Definition (HD) video broadcasts (Leopold, 1994). The Grand Alliance is an advisory committee to the FCC which was created to offer advice on the standardization of advanced television (Robertson, 1995, p. 8). Although the majority of television viewed in America is standard definition television (SDTV), all of the major broadcast networks now broadcast high definition television (HDTV). Additionally, over half of all American households now own at least one television capable of displaying HDTV (HD TVs, 2010). Two types of HDTV video, 1080i and 720p, have become the main HDTV formats broadcast in America. Because 720p uses a progressive scan format, it is usually regarded by the television industry as being superior to 1080i (Péchard, Le Callet, Carnec, & Barba, 2007). This line of thinking is especially true with regard to content involving motion (Ynan, 2006, p. 45).

Although there are many other forms of technology that people employ to watch motion video that is not a traditional television (tablets, smart phones, computers, etc.) televisions remain the main way moving pictures are watched. In the second quarter of 2009, global television sales increased 26 percent when compared to the same quarter just one year before (Lillington, 2010). One reason for this increase was increased content of on demand programming. On Demand programming allows television viewers to choose what time they will watch a particular show. Comcast reported that in 2003 the company offered fewer than 1,000 On Demand choices that were viewed 200 million times. In

2011, Comcast reported offering 25,000 On Demand choices that were viewed over 20 billion times in the first half of 2011 alone. The majority of these viewings occur on home televisions via the SDTV, 1080i and 720p formats (Comcast, 2011).

Purpose of Study

This study will test which HDTV format audiences prefer when watching different HD video formats. The television industry has generally accepted the idea 720p is superior to 1080i with regard to video content, especially when capturing/displaying video involving motion (Hill, 2003, p. 40; Hoffmann, Itagaki, Wood, Hinz & Wiegand, 2008, p. 11; What You Need to Know). Conversely, it has also generally been accepted 1080i is superior to 720p with regard to capturing/displaying stationary images (Ynan, 2006, p. 45) or when the video is to be displayed across a large viewing area (Hoffmann, et al. 2008, p. 11). However, there has been little testing of these viewpoints which involved displaying the two most common HDTV formats to audiences, to verify which format they actually prefer.

Theoretical Background

Understanding the way media are delivered to, and interpreted by, people is important. This concept gained popularity in 1964 when Marshall McLuhan published *Understanding Media: The Extensions of Man*. McLuhan advocated the idea "the medium is the message" (McLuhan, 1994, p. 7). In explaining this concept, McLuhan argued all media impacts are dependent upon their delivery platform or medium. McLuhan expanded upon this idea by claiming each delivery platform used by media directly changes the way its content will be interpreted, processed and impact individuals as well as society. McLuhan asserted media should be the topic of study, not the content of the message. McLuhan felt that because technology is evolving at such a rapid and ever increasing pace "the stakes are very high, and the need to understand the effects of the extensions of man [i.e., media] become more urgent by the hour" (p. 4). McLuhan then offered a starting point for medium research, "[e]xamination of the origin and development of the individual extensions of man should be preceded by a look at some general aspects of the media, or extensions of man, beginning with the never-explained numbness that each extension brings about in the individual and society" (p. 6). In defense of his view, McLuhan compared television to film. McLuhan argues the differences in media make it harder for a viewer to determine/interpret the intended message when watching television as opposed to movies.

Phenomenologist Don Ihde (1982) expanded upon this concept of the medium being intrinsically linked to: message delivery and impact/interpretation of delivered messages. Ihde claimed any communication that does not take place face to face and occurs through the help of technology is 'mediated' by said technology. Ihde expanded this theory by asserting mediated communication is necessarily impacted by the technology mediating the communication. Since television is a communication medium, it thus takes a "mediating position" (Ihde, 1982, p. 74).

Mediation allows television to enhance some aspects of the communication process. An example of these benefits is television's ability to deliver the same message, over huge expanses of land, to large numbers of people. However, the mediating process of television also impairs communication for reasons such as television being impersonal and not accessible to everyone. McLuhan and Ihde's theories support the imperative to understand the benefits and drawbacks television technologies affects on communication and audience perception.

Zettl (1982) tried to determine how much research had been done on the topic of television aesthetics. He concluded there had been minimal research on this topic, and "these few studies revealed an astonishing inconsistency in research approach and interpretation of results" (Zettl, 1982, p. 9). Zettl also stated concern with the lack of studies on television as a medium. He speculated this was because "aesthetic potentials and requirements of the television medium have not been considered an area of serious research" (p. 9). One reason for Zettl's alarm was his belief the evolution of television created "new production techniques" which directly impacted "perceptual effects on the viewer" (p. 9). Specifically, one of the questions Zettl asked when discussing new television display formats was how "will the new screen affect the intimacy factor" (p. 9) of a viewer.

When writing about the changes new technologies have on media aesthetics, Zettl (2005) states the new technologies which "produce high-quality images" and the "wide, movie like aspect ratio of the high-definition television screen" require those creating content for these technologies to match these new technical properties with an "equally high aesthetic standard" (Zettl, 2005, p. xix). Zettl says this is important as mass communicators impact millions of unsuspecting people daily. And, people with this influence have a responsibility to not only master the skills necessary to understand the impacts they are creating, but to also stand by a moral obligation based on "genuine concern and respect for your audience" (p. 14). As a consumer of mass communication Zettl also contends one's ability to "evaluate the [previously] established production standards as well as the new ways aesthetic elements [which] are used" is essential to "media literacy—an educational discipline that has become an essential prerequisite for responsible citizenship" (p. xix). Zettl contends applied aesthetics "deals with a number of aesthetic phenomena, including light and color, space, time/motion, and sound, and our perceptual reactions to them" (p. 15). If Zettl is right, individuals' and societies' perceptual reactions to television messages are not just important factors in media aesthetics but to responsible citizenry as well. Combining this notion with McLuhan's (1994) and Ihde's (1982) contentions of message delivery platforms inherently influencing messages, the study of media delivery platforms and human impacts should be an important topic of study. This is especially true for those citizens with the power to influence others (so they act with an ethical responsibility to their audience) and those in a position to be influenced by media (to guard against being irresponsibly persuaded), for responsible citizenry in the current technology driven world.

Besides McLuhan (1994), Ihde (1982) and Zettl's (1982 & 2005) research, there has been nominal academic research on how different television formats (as media) impact society and individual viewers. Specifically important to this thesis is that audiences have never been exposed to the two main formats of HDTV broadcast in America in an experiment designed to determine the impact each format has on said viewers, and which format viewers prefer. This thesis will attempt to fill part of this research void by exposing audiences to different video clips displayed via America's two main HDTV formats, and measuring audience perceptions of these formats.

Definitions

This section starts with an explanation of terms and concepts about how television is recorded, displayed and broadcast. Subsections will provide the reader with a general introduction to standard definition (SD) video and how a television image is displayed. This information then progresses into a more complex discussion on the two main HDTV formats broadcast in America (720p and 1080i). Scanning formats, field use, frame creation and frame rates are addressed for SD video, 720p and 1080i. These topics are followed by an explanation of codecs and image impairment. This section finishes by discussing agencies which have sought to create worldwide television standards and television quality testing techniques. In doing so, the agencies: The International Electrotechnical Commission; International Telecommunication Union; Consultative Committee on International Radio; International Telegraph and Telephone Consultative Committee and the European Broadcaster Union are explained. These agencies are crucial to this field of research as they have driven and helped fund much of the research and experiments conducted to better understand viewer perceptions of video quality.

Standard definition video.

Standard definition video has been broadcast in America since the 1920's (WRNY to Start, 1928, p. 13). Before June 12, 2009 video broadcast over the airwaves in America needed to adhere to the National Television Standards Committee (NTSC) format. The current version of SD video broadcast in America uses the Advanced Television Systems Committee (ATSC) format. Low power transmitters, which broadcast over the air, can still use the NTSC format. As low power transmitters are rare and only broadcast to a small fraction of the televisions in America, they will not be discussed again in this thesis. The ATSC is a conglomerate of over 20 governmental and non-governmental agencies, including the Federal Communications Commission (FCC), Institute of Electrical and Electronics Engineers, National Institute of Standards and Technology, and the Society of Motion Picture and Television Engineers. The ATSC was established to create standards for how television would be broadcast after the federally mandated digital transition which occurred June 12, 2009 (Federal Communications Commission, 2009). These ATSC broadcast standards are now enforced by the federal government. Standards set by the ATSC call for broadcasting capabilities allowing 1080i, 720p and 480i. Standard definition video is often referred to as 480i. The ‘480’ refers to how many visible, horizontal lines actually comprise the images being broadcast. There are actually 525 lines being broadcast, however, 45 of them are left blank “to allow for the screen to reset itself after the interlacing process” (Hill, 2003, p. 40). American SDTV broadcasts are displayed using a 4:3 aspect ratio (Linecker, 2004, p. 20; Loncaric et al., 2008). This means the vertical height of an SD image will only have three units of measurement, while the image’s horizontal width will have four units of measurement (A Digital Video Glossary, 1994, p. 80; Medoff & Tanquary, 1998, p. 20).

Interlaced video.

Interlacing video, which put the ‘i’ in 480i as well as 1080i, refers to how a frame of video is displayed. In traditional SD video and all other forms of interlaced video broadcast in America, each second of video is created from 29.97 individual images called frames (Medoff & Tanquary, 1998, p. 32). With interlaced video, each frame is made from two separate images called fields, resulting in a field rate of 59.94 fields per second (Linecker, 2004, p. 20; Medoff & Tanquary, 1998, p. 32). The first field in a frame of interlaced video displays all of the odd numbered horizontal lines that create one frame of video, while the second field completes the image by displaying the even numbered lines. With interlaced video, though only one individual field is shown at a time, the timing of the fields and the way they are displayed tricks the human brain into blending the two fields together to create the perception of one single image. Furthermore, by showing 29.97 frames per second (fps), the brain is again tricked and interprets the still images as a fluid moving picture.

Progressive scan video.

Video using the progressive scan format, which is what put the ‘p’ in 720p, refers to how a frame of video is displayed. As opposed to interlaced video, progressive scan refers to a format that does not use fields. Instead, progressive scan video creates a frame by displaying all of the horizontal lines sequentially (Hill, 2003, p. 40), and is broadcast in America using a frame rate of 59.94 fps (Hill, 2003, p. 40; Krauss, 2005, p. 58).

Frame rate.

Frame rate refers to how many frames are displayed in a second. Standard definition video and all other interlaced video broadcast in America is currently displayed at a frame rate of 29.97 fps. Progressively scanned video is broadcast at 59.94 fps.

High definition video.

There are two predominate HD video formats captured and broadcast in America, 720p and 1080i. Standard definition video creates a frame of video using roughly 350,000 pixels (Linecker, 2004, p. 20). A frame of video using 1080i is created using more than 2,000,000 pixels, while a frame of 720p uses approximately 920,000 pixels.

The 720p and 1080i formats are also broadcast using a 16:9 aspect ratio (Loncaric et al., 2008, p. 5).

1080i.

1080i refers to interlaced HD video displaying 1920 horizontal lines. Each of these horizontal lines are comprised of 1080 pixels (Krauss, 2005, p. 58). Because an image using the 1080i format has more horizontal lines than 720p, and has more pixels per line (720p only uses 1280 lines comprised of 720 pixels each), 1080i is generally considered better than 720p with regard to stationary images and video displayed on large monitors (Pechard, Le Callet, Carnec, & Barba, 2007). In fact, monitors "larger than 50 inches would require a higher spatial resolution than 720p/50 offers" (Hoffmann, Itagaki, Wood, Hinz & Wiegand, 2008, p. 11). Major networks which currently broadcast in 1080i include the National Broadcasting Company (commonly referred to as NBC), CBS Broadcasting Incorporated, Turner Network Television, Versus, HDnet, HBO and Showtime (Finn, 2006, p. 26).

720p.

720p describes a form of progressively scanned HD video. Video in this format is displayed using 1280 vertical lines. Each horizontal line is made of 720 pixels. The television industry usually assumes 720p delivers a superior picture to interlaced video (Loncaric et al., 2008, p. 5) and is better for recording/broadcasting quick motion due to the progressive scan format (Péchard, Le Callet, Carnec, & Barba, 2007; What You Need to Know; Ynan, 2006, p. 45). Networks which currently broadcast programming using the 720p format include: Entertainment and Sports Programming Network (often referred to as ESPN), American Broadcast Company (commonly called ABC) and Fox Broadcasting Company (Finn, 2006, p. 26).

Image impairment.

Image impairment refers to the level to which an image has been negatively altered or distorted.

Codec.

Codec is an acronym for enCOder/DECoder. Codecs work with software or hardware to compress and decompress audio and video files. This is generally done to reduce file size thus saving space and/or speeding up transmissions (PC Mag.com, 2011).

Agencies Important to Television Format Design

This section discusses agencies that are responsible for the evolution of television formats. All of the agencies mentioned have had a substantial role in the way television is currently broadcast in America or in broadcast formats previously used in this country. Additionally, all agencies mentioned are also referenced by others' research reviewed in this thesis' literature review.

International electrotechnical commission.

The International Electrotechnical Commission (IEC) is an affiliate of the International Standards Organization, which was created as a specialized United Nations agency. The IEC was the first internationally recognized organization to publish standards on subjective television tests. However, most of their suggestions are not directly applicable to this thesis as their suggestions pertain "primarily to the special requirements of in-factor testing" (Allnatt, 1983, p. 205).

International telecommunications union.

The International Telecommunication Union (ITU) is a United Nations agency which was originally founded in 1865 to create standards for the telegraph. The ITU now specializes in "information and communication technologies" (Committed to Connecting the World, 2011). In doing so, the ITU has been divided into three main areas of focus: radio communications (ITU-R); standardization; and development. The ITU is a knowledgeable and respected organization with regard to creating and offering suggestions on ways to subjectively assess the quality of television pictures. The ITU published the report: *Methodology for the subjective assessment of the quality of television pictures in 1974*. The ITU has since revised the report 13 times, with the last revision having been published in 2009.

Consultative committee on international radio.

The Consultative Committee on International Radio (CCIR) is a smaller department of the ITU-R responsible for what is known as the ITU's Study Group 11. Study Group 11 is charged with researching and making suggestions on "international technical aspects of television broadcasting" (Allnatt, 1983, p. 205). The CCIR is designed to create questions and then have researchers and professionals try and answer their questions. Results to these questions are generally published every four years. Even when results and suggestions for a question have been published it does not mean research on that topic is concluded. Of particular interest to this thesis topic are two questions published by the CCIR in 1951 which are still a topic of research today:

1. what standardized methods and means of test, independent of the television standards employed, can be used to measure accurately, and

wherever possible objectively, the deterioration introduced into monochrome and colour pictures by television, taking into account the system, the equipment and the transmission processes;

2. what are the relationships between the objective parameters of television signals and the subjective assessments of displayed picture quality? (p. 209)

International telegraph and telephone consultative committee.

The International Telegraph and Telephone Consultative Committee (CCITT) was an agency which worked with the CCIR in the late 1970's and early 1980's to research international broadcasting of both television and sound. As subjective testing methods were necessary to both these fields their past research is important to this thesis. However, in an effort to streamline results, the CCITT determined methodology "should be handled by Study Group 11 of the CCIR" (Allnatt, 1983, p. 205), making their reports of more interest to this field of study.

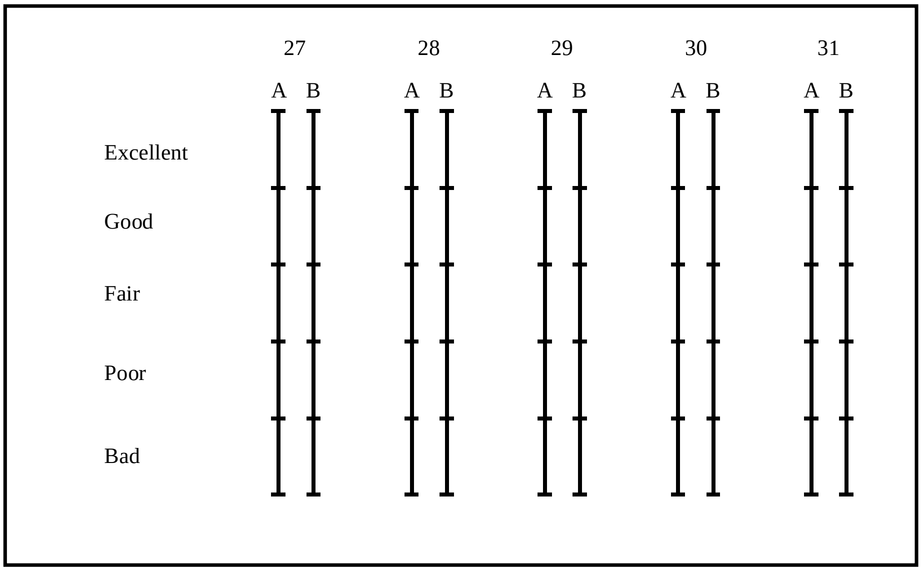
European broadcasting union.

The European Broadcasting Union (EBU) was founded in 1950. This entity's goal is "promoting cooperation between broadcasters" and facilitating "the exchange of audiovisual content" (About the EBU, 2011). Founded in 1950, the EBU claims to be the largest association of broadcasters worldwide. Their 74 active members and 36 associate members reach an audience of 650-million viewers weekly (About the EBU, 2011).

International telecommunication union recommended method.

The ITU and its subsidiary organization, ITU-R, make recommendations on different testing methods. With regard to trying to "measure the quality of systems relative to a reference" (International Telecommunication Union, 2009, p. 10) the ITU-R's recommended method is to perform a double-stimulus continuous quality scale (DSCQS). In this thesis, this is referred to as the ITU's (or ITU-R's or CCIR's) recommended method. The ITU-R's recommended method employs this DSCQS method. For the DSCQS method, assessors' are asked to view a pair of images; one is the reference image, the other is an impaired image, and rate the quality of both images. Assessors are not told which images are which and the presentation of both image types are shown in pseudo random order. When multiple assessors are used in the same testing session (described as 'variant II') the ITU-R method mandates "the pair of conditions [be] shown one or more times for an equal length of time to allow the assessor to gain the mental measure of qualities associated with them, then the pair is shown again one or more times while the results are recorded" (International Telecommunication Union, 2009, p. 13). When testing moving video, the ITU-R method recommends the video sequences should be 10 seconds in length and shown to assessors twice, only allowing assessors to vote on their second viewing. Assessor ratings should be recorded on a five-part quality-rating form using a continuous scale (see Figure 1).

Figure 1. Portion of the ITU-R recommended quality-rating for advised DSCQS method tests.



(International Telecommunication Union, 2009, p. 15)

European broadcasting union method.

The EBU has a recommended methodology for testing viewer perceptions of video quality referred to as the EBU method. With regard to testing viewer perceptions of image quality using a DSCQS method, the EBU method starts with the CCIR's recommended method then aims to "specify procedures precisely enough to ensure adequate coherence" (Bernath, Kretz & Wood, 1981, p. 68). The EBU method is currently very similar to the ITU-R recommended method in that both use a double-stimulus. However, there are major differences between the two organizations' methods. The EBU method requires the assessor to first be presented with the unimpaired reference image/video, and then be shown the same image/video impaired. The assessor is then asked to rate the second image/video in relation to the unimpaired image/video. Another major difference between the ITU-R recommended method and the EBU's recommended method is that the latter uses an impairment scale (as opposed to a quality scale). Impairment and quality scales refer to a specially designed response form completed by respondents. Impairment scales are unique in that their impairment grades (or descriptors) are described using a grade of annoyance, from *imperceptible* to *very annoying*. For example a five-grade impairment scale would use these descriptors: 1) Grade 5: Imperceptible; 2) Grade 4: Perceptible but not Annoying; 3) Grade 3: Slightly Annoying; 4) Grade 2: Annoying; 5) Grade 1: Very annoying (Bernath, Kretz & Wood, 1981, p. 67). Conversely, quality scales similar to the impairment scales, use descriptors ranging from 'excellent' to 'bad'. For example a five-grade scale would use these descriptors: 1) Grade 5: Excellent; 2) Grade 4: Good; 3) Grade 3: Fair; 4) Grade 2: Poor; 5) Grade 1: Bad (International Telecommunication Union, 2009, p. 15). Otherwise, with regard to this thesis design and topic, the two methods are very similar. The ITU-R recommends using the EBU method when trying to measure system robustness (International Telecommunication Union, 2009, p. 10).

Chapter Review

This chapter discussed key terms, concepts and organizations which will prove important to the topic of viewer perceptions of different HDTV formats. This discussion started by examining SD video and how television images are displayed and developed into relevant information and terms associated with the HDTV formats broadcast in the United States. Then, after addressing image impairments and codecs, organizations seeking to create international television standards were denoted. This chapter finished with an examination of relevant testing methods recommended or discussed by these agencies.

Thesis Organization

This thesis has four subsequent chapters. Chapter 2 is a review of relevant literature. Chapter 3 discusses the hypotheses and experimental methods used to actually test which HDTV format audiences prefer with regard to image quality. Chapter 4 presents the results from the experiment. Chapter 5 discusses relevant findings from this study and concludes by presenting ideas for future research using similar methods to determine preferences of video formats.

CHAPTER 2

LITERATURE REVIEW

This chapter starts by reviewing past experimental methods and experiments used by researchers trying to determine the best way to test and rate viewer perceptions of image quality. With regard to television, subjective image assessment and corresponding testing methods were not a popular research topic before the 1980’s. In the early 1980’s that changed for multiple reasons. One reason experimental methods testing increased was the technology associated with subjective image tests (monitors, switchers, recording devices, etc.) became more economical. In addition, trying to determine better ways to assess viewer reactions to image quality gained more tangible benefits as television was evolving and gaining in popularity. Broadcasters and companies making capture and/or playback devices tried to determine which codecs were *best* with regard to viewer preferences, image quality and cost. Because of these reasons and the fact that before 1980 the ITU-R only recommended single-stimulus tests to determine viewer perceptions of image quality, this chapter analyzes research starting in 1980.

Studies Seeking to Determine the Most Accurate Method and Ranking Systems When Determining Viewer Perceptions of Visual Images

This section of the chapter progresses chronologically discussing relevant research and experiments. In doing so, this section tries to determine which testing/ranking methods are most accurate and analyzes previous research and experiments that tried to create more accurate testing/ranking methods. Next, two studies trying to determine how audio impacts viewer perceptions of images are reviewed. This literature review then examines two studies which tried to determine audience preferences between 1080i and 720p. It should be noted these studies were conducted in Europe and used slower frame rates (25/50) than video broadcast in America (29.97/59.94). In addition, relevant literature discussing the differences in HDTV formats is discussed. This thesis will build off of the experimental methods reviewed, designed and tested throughout these studies. The literature review finishes with a chapter review.

Studies from 1980's.

In 1980, Allnatt wrote an article stating television image assessment was beginning to attract "worldwide attention" (Allnatt, 1980, p. 450). Although this attention was mainly being garnered by researchers and the television industry to assess television codecs, Allnatt's article relates to this thesis. Allnatt's research dealt with ways to determine which kind of video technology was *better,* when to a typical viewer the video types being tested were often perceived as being very similar or the same. Allnatt's research showed previous work done on subjective picture assessment typically used "impairment magnitudes selected to ensure [respondents are able] to use the full range of the grading scale" (p. 450). Allnatt claimed the use of this sort of grading scale in past research led "to problems which have sometimes been overlooked" (p. 450). Allnatt cited a previous study claiming a five part impairment scale which "include categories enabling a threshold of perception to be ascertained" (p. 450) (i.e., an impairment scale) may correct some of the 'problems' created when using a 'full ranged grading scale'.

In an effort to see if this theory was correct, Allnatt re-examined the results of a previous experiment. Allnatt compared respondents' results which were recorded using a six-point quality scale (excellent, good, etc.) to respondents' results recorded using a six-point impairment scale (imperceptible, just perceptible, etc.). When comparing the results of these rating methods, Allnatt's results indicated when the experimental video image had little impairment or was distorted with blur even to a high level of impairment, the quality scale was most accurate. Allnatt's results suggest there is no research advantage to using an impairment scale in place of a quality scale. Allnatt speculates, "almost the mere fact of requiring an observer to determine whether an impairment is present diminishes his ability to judge total picture quality" (Allnatt, 1980, p. 451). Allnatt finished his paper by admitting both six-point quality scales, based on descriptors covering such a wide spectrum of perceptions, cannot accurately judge small differences in video quality. The researcher then suggested a continuous rating scale may yield more accurate results. This implementation of a ratio scale is currently suggested by the ITU-R for DSCQS methods comparing image quality.

Shortly after Allnatt's (1980) report was released, White and Allnatt (1980) published an article in the same journal. Their research was the first to conduct an experimental test of a double-stimulus quality rating method and involved comparing the results of four separate testing methods. Each method was based on respondent responses of their perceptions of image quality. In their experimental design, White and Allnatt sought to create a double-stimulus testing method similar in many aspects to the then CCIR's recommended single-stimulus method. For example, White and Allnatt ran part of the experiment using the single-stimulus quality grading scale recommended by the CCIR. Revolutionary to the field of image perception research was White and Allnatt's creation, and experimental tests using a double-stimulation method. White and Allnatt's experimental design required the researchers to test four separate testing methods: (i) Their first methodological design employed "the conventional full-range single-stimulus quality grading method" recommended by the CCIR's report 500-1; (ii) White and Allnatt's second method was the same as their first method except the grading scales were changed to the ratio-impairment scales which were each "100mm high, spaced 15mm apart" (p. 715); (iii) The researchers' third method was the same as their second method except it employed a double-stimulus technique. In addition, the space between their respondent form rating scales was reduced to 5 mm with a 10 mm space between pairs; and (iv) White and Allnatt's fourth methodological design was the exact same as their third method except it was restricted to the "three smallest" (p. 7155) impairment magnitudes (including zero).

Citing Allnatt's (1980) previous work which stated quality scales yield better results than impairment scales, White and Allnatt (1980) designed part of their experiment to use a five-part quality rating scale. Also building upon Allnatt's earlier examination of rating scales, White and Allnatt created a quality scale that was also a ratio-scale. This was done by having respondents mark their image assessment responses on a continuous line which was broken into five equally sized sections. Each section was labeled with the common descriptors of a quality-scale (excellent to bad). Although not identical, it should be noted this scoring method is very similar to the ITU-R's "quality-rating form using a continuous scale" (International Telecommunication Union, 2009, p. 15) recommended for DSCQS tests today. White and Allnatt's (1980) method (iv) results showed the effect of removing large impairments from an experiment, whether it be single-stimulus based or double-stimulus based "is only small" (White & Allnatt, 1980, p. 715).

The largest result variations between any of the experimental methods were between methods (i) and method (ii). Analysis of these results showed there was a substantial difference in results obtained by using the then recommended quality five-part impairment scale as opposed to a five-part quality scale/ratio scale. However, the researchers did not design the experiment to show which was more accurate by comparing the actual level of image impairment and the mean average of the respondents' score. The most important part of White and Allnatt's experiment is in its proof a double-stimulus method can be operated "without trouble" (White & Allnatt, 1980, p. 715) in a matter that yields empirical, testable and comparable results.

MacDiarmid and Darby (1982) conducted their own research to determine whether the double-stimulus method proposed by White and Allnatt (1980) would be better suited to determining viewer perceptions of image quality as opposed to the single-stimulus method suggested by CCIR-1 at the time. MacDiarmid and Darby conducted this research theorizing single-stimulus quality-scale (SSQS) methods should be "considered unsuitable for the evaluation of different digital coding systems, because of the difficulty in arranging tests in such a way that the observers would use the full range of the grading scale" which would "produce a grand mean score close to the Grade 3, as recommended by the CCIR" (MacDiarmid & Darby, 1982, p. 71). These researchers reviewed two experiments conducted by the Independent Broadcasting Authority (IBA) utilizing the double-stimulus quality rating method. Both of these experiments were performed with similarities to White and Allnatt's (1980) previous experiment. Although both sets of researchers were experimenting with ways to best test different television codecs, MacDiarmid and Darby's research involved double-stimulus tests using only still images displayed via monitors. Additionally, as implied by the label *double-stimulus*, all of the double-stimulus experiments discussed above were performed by simultaneously showing respondents two versions of a visual image. These images were comprised of both the original (unimpaired) reference images and impaired images. Respondents were not told which were which, and these images were shown in a pseudo random order. The experiments performed by the IBA also used five-point quality rating forms on a continuous (or ratio) scale. The IBA's form was more similar to the current ITU-R's rating form recommended for DSCQS tests today as the descriptors and continuous scales were vertical and White and Allnatt's were horizontal. Via their research and analysis of the IBA's experiment, MacDiarmid and Darby ascertained there were five advantages to the double-stimulation method because this format: (a) created a procedure which was similar to the single-stimulus quality grading method then recommended by the CCIR; (b) employed reference leads which improved accuracy while judging small impairments and produced similar, absolute results to those of the *normal* single-stimulus method; (c) used quality scales which the researchers believed were more sensitive than impairment scales when judging small impairments (as suggested by Allnatt's (1980) previous research); (d) obtained results which had highly stable results generally reflecting actual impairment-level-changes (p. 71); and (e) the IBA's experiments on subjective double-stimulus image quality and MacDiarmid and Darby's analysis on these experiments helped re-confirm White and Allnatt's (1980) previous findings which indicated double-stimulus methods were easy to use and gave "a small spread of results, in the assessment of small impairments" (p. 76).

Studies from the 1990's.

The ITU currently suggests the correct method to test the quality of coded picture images is the double-stimulus continuous quality scale method (International Telecommunication Union, 2009 p. 10). Narita (1994) conducted his research to determine if modifying the ITU-R's DSCQS method could yield better results. Specifically, Narita altered the "presentation of the test sequences" (Narita, 1994, p.7) as recommended by the ITU-R when conducting a DSCQS method experiment to determine whether an absolute scale or a relative scale would yield better results.

Narita's (1994) research compared previous tests to determine the reliability of DSCQS methods. Narita referenced research done by the CCIR in which the committee conducted statistical analysis on the difference between the actual degree of image impairment and respondents' rating of the image. This work found the discrepancies on some sequences were as small as 12 to 18 percent.

In conducting his research, Narita (1994) ran an experiment using the traditional ITU-R recommended DSCQS method and ran a similar experiment using his modified method. The original experiment, examining the un-modified DSCQS method, was conducted using an original sequence (unimpaired HD video) and versions of the sequence modified to different bit rates resulting in seven sequences with their own quantifiable level of image impairment. The respondents of this test, referred to as 'assessors', were comprised of 15 expert assessors and 20 non-expert assessors. By comparing the results of the expert assessors to those of the non-expert assessors, Narita determined the smaller the video image impairment the less accurate non-expert responses became. Narita's results also showed as image quality impairment decreased, a greater disparity was shown between the expert and non-expert assessors' results, favoring the expert's accuracy.

Important to this research were Narita's results which helped re-confirm the previous findings by the CCIR. Both bodies of research showed the traditional DSCQS method had a reliability score falling within the 12 to18 percent range. That is to say, respondents on average were able to judge the video sequences' video quality so accurately they would only error 12 to 18 percent away from the actual, quantifiable video quality impairment. However, Narita argued, "the meaning of 12% and 18% is obscure" (Narita, 1994, p. 10) and therefore of little value. However, as Narita claims his research shows DSCQS methods can yield reliable results, his results showing respondent responses error on average far less than 20% could make the case that the DSCQS testing methods employed by Narita are *valid* and *reliable*. Not addressing the concept of validity in his research, Narita proposed the true importance of his experiment was in determining which method had the most reliable results. Both experiments were run under the same conditions save one detail. The traditional experiment used the un-modified method employing quality scale score sheets as recommended by the ITU-R (International Telecommunication Union, 2009, p. 15). A modified example of this score sheet can be seen in Appendix 1.

Narita's (1994) second experiment used the same ratio scale but changed the identifiers to those of a traditional relative scale. After running a statistical analysis on his results, Narita determined with regard to the absolute scale "significant variations of the reference scores" (Narita, 1994, p. 13) were shown. Additionally, Narita's results showed when impairments were small the relative scale was more reliable than the absolute scale. Narita's results oppose those findings by Allnatt (1980) which found relative scales to be less reliable than absolute scales. It should be noted Allnatt's analysis was done on results obtained from an experiment not implementing a double-stimulus method.

Building upon Narita's (1994) previous research, Narita and Sugiura (1997) designed an experiment to further determine the reliability differences between the DSCQS method and SSQS method utilizing a five-grade quality scale method. Narita and Sugiura believe this work is important as their research had discovered studies with discrepant results. Narita and Sugiura's work started off by discussing the history of European subjective testing methods designed to evaluate digital television codecs. In doing so, the authors contended the DSCQS method reported in the early 1980's had certain advantages. These are basically the same five advantages previously listed by MacDiarmid and Darby (1982). Again these were: (a) the creation of a procedure, which was similar to the single-stimulus quality grading method then recommended by the CCIR; (b) employing reference leads which improved accuracy when judging small impairments and produced similar results to those of the *normal* single-stimulus method; (c) use of quality scales which the researchers believed were more sensitive than impairment scales when judging small impairments (suggested by Allnatt's (1980) previous research); (d) ability to obtain results which have stable/accurate results reflecting actual impairment-level-changes; and (e) ease of use for double-stimulus methods with only slight result impairments. Narita and Sugiura then cast doubt on some of these advantages, in particular points a-c by referencing an article and conference from the 1990's which suggested other testing methods, specifically the *modified EBU method,* was more sensitive to small impairments. Also mentioned in those articles was the suggestion the DSCQS method can be "affected significantly (with a significance level of 0.05) by the degree of the impairments in the test sequences" (p. 26).

In explaining how their experiment was conducted, Narita and Sugiura (1997) began with a description of the DSCQS method followed by a description of the SSQS method and concluded with an evaluation of their testing procedure. The DSCQS method required their assessors to mark their rankings on the score sheet recommended by the ITU-R (Appendix 1). This score sheet had five descriptors (Excellent, Good, Fair, Poor and Bad) designating an equally sized section on a continuous ratio scale. Contrariwise, assessors exposed to the SSQS method were instructed to "evaluate the absolute quality of each sequence, using the five-grade quality scale" (Narita & Sugiura, 1997, p. 27) not accompanying a ratio scale. Both score sheets used the five same descriptors. Narita and Sugiura stated the respondents for their experiment were comprised of 15 'expert' assessors and 20 'non-expert' assessors.

Narita and Sugiura (1997) compared each video sequence's bit rate (quality) with the respondents' mean video rating. Analysis of these results showed the SSQS method had more reliable responses (on average) than the DSCQS method. In addition, Narita and Sugiura's analysis of results with regard to the "sensitivity to small quality differences" determined 'the capability of [respondents] to recognize small quality differences using the SSQS method is superior to that of the absolute scale of the DSCQS method" (Narita & Sugiura, 1997, p. 30). These results are surprising as all other research reviewed for this thesis comparing DSCQS method's reliability to that of SSQS method's level of reliability had determined the DSCQS method to be a more accurate method for judging video quality. Narita and Sugiura speculated this was because "better results cannot be achieved with a continuous scale, for reasons that include the fact assessors are, in practice, no more discriminating than the discrete scale allows" (p. 31). Narita and Sugiura's research was conducted in Japan. Thus, translation between the descriptors may have been an artifact in this experiment. Additionally, the level of impairments used in their experiments ranged from minimal to very significant.

Studies researching audio impact upon image assessment.

Hands (2004) published an article describing his methodology and corresponding results for two experiments. Hands' objective for his research was to "develop a basic multimedia predictive quality metric" (Hands, 2004, p. 806). Hand performed his study by designing two experiments. The second experiment was designed around two audio-video sequences showing medium shots of people. The first experiment used these video sequences and also introduced a "high-motion sequence" (Hands, 2004, p. 806). Procedures for both experiments included the showing of video sequences to respondents, and asking them to rate the audio quality of the sequence just viewed. Next, respondents were instructed to rate the sequence's video quality. In addition, a test was run to determine the respondents' opinions of each sequence's "multimedia quality" (Hands, 2004, p. 806). Those experimental designs employed the tactic of modifying the sequences shown by adjusting the audio's bit rate as well as the video's bit rate. This created different sequences with varying levels of audio/video quality which respondents were asked to rate.

Hands ran a regression analysis on some of his results and determined "human subjects integrate audio and video quality together using a multiplicative rule" (Hands, 2004, p. 806). When shown video sequences comprised of both audio and video, respondents “integrate audio and video quality information to arrive at a single overall quality opinion” (p. 815) even when instructed to judge the video independently of the audio and vice-versa. Hands’ results built upon previous experiments by Beerends and Stemerdink (1992) which also determined audio quality does in fact impact a respondents’ impression of video quality. However, Beerends and Stemerdink’s findings did not show as great of an audio impact on respondents’ perceptions of video quality as the experiment performed by Hands. Hands theorized this was because Beerends and Stemerdink’s research did not include video sequences using high-motion video. He supported this theory by analyzing the results of his experiment. Statistical analysis showed respondents’ video perceptions were affected to a greater extent by audio when viewing the high-motion video sequences as opposed to the relatively motion-free medium shots.

Studies from the 2000’s comparing HDTV formats via audience preference.

Hoffmann, Itagaki, Wood, Hinz & Wiegand (2008) published a study intending to find a subjective way for respondents to assess the picture quality of different video formats. They compared these four video formats: (1) 1080p/50 (a possible third generation HDTV format to be used in some European countries); (2) 720p/50 (the European version of 720p which differs from the format broadcast in America as it uses a frame rate of 50 fps); (3) 1080i/25 (the European version of 1080i which differs from the format broadcast in America which uses a frame rate of 25 fps); and (4) 576/25 (SD television format broadcast in Europe). Important to this thesis is that Hoffmann et al. (2008) created a new method for audiences to compare different video formats. This method, known as the “Triple Stimulus Continuous Evaluation Scale method (TSCES)” (Hoffmann et al., 2008, p. 2), was designed to create a subjective method for audiences to compare three video formats simultaneously. Hoffman et al. listed requirements their work and the TSCES method should address. These are: allow easy, direct comparison of HDTV formats by non-export assessors; produce reliable, reproducible and statistically relevant results; show SD television- 720p, 1080i and 1080p on large and medium sized flat screen liquid crystal display televisions; and accurately measure a format’s basic quality and its impairments.

Hoffmann et al. designed their experiment by setting up three 50 inch LCD monitors in a vertical arrangement. All monitors were the same model capable of displaying a resolution of 1920 X 1080 pixels. The monitors were arranged so that a viewer with an eye height of 1.2 meters who was sitting in the center viewing position would be an equal distance from each monitor. The tests were performed with three in-sync monitors showing the same image simultaneously across the three monitors. Hence, the only difference between the images being displayed on the monitors was the video format to which the video had been down-converted. Across the four video formats listed earlier, Hofmann et al. tested seven different video scenes. These were: Crowd Run (this involved a stationary shot of marathon participants jogging by a park); Park-Joy (a pan shot of people running by a park); Princess-Run (a camera pan of one person running); Aloha-Wave (a crowd doing ‘the wave’ during a soccer game); Dancer (a stationary shot of a person dancing in a soccer field with a reflective dance costume); Police-boat (a police boat drifting on water) and Ice-Dance (an indoor pan shot of two people dancing on white ground with a background that also had detailed structures). The results of

Hoffmann et al.'s study supported the television industry's general theory that viewers prefer progressively scanned video to interlaced video (p.11). It should be noted all of the formats tested in this study used European broadcast formats which use different frame rates than video broadcast in America.

Loncaric et al. (2008) conducted a similar study to the one done by Hoffmann, Itagaki, Wood, Hinz & Wiegand (2008). Loncaric et al. also tried to determine which European HDTV format viewers preferred. Loncaric et al. examined the same HDTV formats (720p/50, 1080i/25 and 1080p/50) Hoffmann et al. (2008) studied, minus SDTV. Loncaric et al.'s study differed from Hoffman et al.'s study as it tried to determine the "optimal combination of bit rate reduction and picture quality" (Loncaric et al., 2008, p. 5). The results of the study supported Hoffmann et al.'s viewer perception findings which reported progressive scan video delivers a superior picture quality to interlaced video. Loncaric et al. suggested this is because "[o]nce interlacing is applied to an image, vertical-temporal information is lost and can never be recreated. Therefore, the 1080i/25 format provides lower picture quality than the 720p/50 format at the same bit rate" (Loncaric et al., 2008, p. 6).

Chapter Review

This literature review examined the history of experimental-image-assessment-designs. Research methods and results for many different testing methods, starting with SSQS which was originally used to assess viewer perceptions of video impairments, were reviewed. This chapter analyzed experiments conducted using DSCQS testing methods. In addition, experiments examining the reliability differences between the SSQS and DSCQS methods were evaluated. In doing so, the reliability as well as the pros and cons of employing continuous scoring methods, as opposed to discreet scoring methods, was analyzed. This chapter concluded with a review of two separate experiments run in Europe which actually exposed test respondents to both 1080i and 720p in an effort to determine their assessments of these different HDTV formats. One of these experiments even offered a new testing method designed around exposing respondents to a triple-stimulus testing method.

This chapter reviewed many studies' relevant to image assessment and applicable testing methods. How, when, and why researchers designed and used different analytical and comprehensive research methods was studied. While much of this research was conducted to determine what video format is *best*, all of the research examined seemed to lack an attempt to explain the theoretical implications of their work. As discussed in Chapter 1, Marshall McLuhan and Don Ihde both argued messages dispensed by media are directly impacted by the medium with which they are delivered. Yet, not one of the studies examined in this literature review attempted to address this issue. There was not one theory about how new television formats would impact the messages or meaning being distributed by new technologies. If McLuhan and Ihde's theories were correct, then this should be a serious area of concern for those researching image assessments as well as current and future television formats.

This literature review has shown the extent of research done to determine the *best* method to measure viewer perceptions of television images. This is an important topic of research as companies have spent billions to upgrade to, and, sustain these formats. This literature review also reported researchers are still creating/testing methods which best gauge viewer perceptions of television formats and impairments. This continued exploration and the discrepant results demonstrated by the research and experiments examined show there has yet to be an experimental design in this field which is without flaw. Most of the experimental designs examined were fairly accurate and proved to have statistically relevant results. Yet, no experimental design or scoring method was shown to enable respondents, or even expert assessors, to accurately rate a level video impairment with a high level of accuracy *and* reliability.

CHAPTER 3

METHOD

As discussed in chapter one, Marshall McLuhan (1994) believed the way media are disseminated should be a greater topic of study than the messages media actually intend to circulate. McLuhan argued this by claiming all media messages were directly altered by the medium with which they were delivered. Specifically, McLuhan argued the characteristics of each form of medium (technology, media, etc.) had a specific numbing effect. This can be seen with his comment about the invention of movies: by the "sheer speeding up [of] the mechanical, it carried us from the world of sequence and connections into the world of creative configuration and structure. The messages of the movie medium is that of transition from lineal connections to configurations" (McLuhan, 1994, p. 12). Ihde (1982) helped to strengthen McLuhan's stance by asserting any communication that does not occur face to face is mediated by the technology utilized for its delivery. Zettl (2005) argued new technologies like HDTV create ethical responsibilities for those capturing and those disseminating HD images. Zettl also claimed a major factor influencing media aesthetics is how people react to how images are created. If McLuhan, Ihde and Zettl's theories are correct, then different HDTV formats broadcast in the United States have direct and differing impacts on the way viewers construe the messages being disseminated via these television media formats.

The previous research reviewed for this thesis determined there was no single testing method widely accepted by a majority of researchers as a *best* and/or *non-flawed* testing method when trying to evaluate viewer perceptions' of a video image. This thesis' experimental design will implement the DSCQS method recommended by the ITU. The only modification to the ITU recommended DSCQS method was a modification to the response forms filled out by respondents. The response form used for this experiment employed all of the criteria recommended by the ITU (ratio scale, quality-rating style, etc.) save for the horizontal hash marks used to separate descriptors which were removed in an effort to encourage respondents to better understand they were filling out a ratio-scale as opposed to a discreet five-part rating scale. The ITU recommended DSCQS method was implemented as many other experimenters in this field have shown this to be a valid and reliable testing method when comparing two versions of the same image. Additionally a DSCQS method was used as it is still recommended by the ITU which is the leading organization in creating, testing, endorsing and advising nations and the television industry about new television technologies. Implementation of this testing method also allowed for relevant and comparable results.

Hypotheses

This study examined viewer perceptions of 720p and 1080i, the two HDTV formats currently broadcast in America. As similar experiments have yet to be done with HDTV formats broadcast in America, the following hypotheses were examined:

> H1: There is a significant difference in mean scores of audience perception of image quality as a function of video format, with 720p preferred over 1080i.

Though 1080i has more horizontal scan lines (1920) and is comprised of more pixels (1080) than 720p (1280 lines/720 frames), H1 is predicated on the fact 1080i relies upon less fps than 720p. This frame-rate difference is created because 1080i frames are composed of two separate fields. Conversely, a 720p frame is composed of both odd and even scan lines shown at the same time. It is presumed 720p's frame rate advantage (59.94 fps as opposed to 1080i's 29.97 fps) will result in images preferred by viewers.

H2: The mean scores will show a significant and increase in perceived quality of video image as motion increases amongst the stimulus material.

As motion (and thusly temporal resolution) increases, it is assumed respondents will perceive the video as being more engaging and thus comprised of better image quality.

H3: The 720p video format will be preferred to a greater level as motion increases.

720p's progressive scan format utilizes 59.94 full frames of video every second. This use of full frames (as opposed to the 1080i format which shows two separate fields 29.97 times per second) allows for greater temporal resolution. As motion increases in the video clips, it is theorized a temporal resolution advantage will become more evident.

Research Question

R1: Will there be a statistically significant difference in audience perceived quality between day and night shots of similar scenes?

Previous research has not analyzed how outside lighting conditions impact HD viewer format preferences. By analyzing the results between clip 3 (a stationary shot of people walking on a side walk during the day) and clip 5 (a stationary shot, very similar to that of clip 3 except it was shot at night) it may become obvious this should be a future topic of study. A stationary shot was used in this case to reduce the impact of motion. Instead, the focus of this research question is to determine whether either of the two

HDTV formats are preferred when images were captured with a high rate of visual-noise in the image, in this case, gain set to a level of +18 decibels (dB).

Experimental Design

Participants were assigned to two groups. One group saw one sequence of video clips, the other group saw a different sequence. The differences in sequences involved the use of motion. One sequence was arranged from high-motion to low-motion and the other sequence was arranged from no-motion to high-motion. Respondents were asked to rate the quality of video shown simultaneously on two monitors. Unbeknownst to them, one monitor displayed a video clip captured via the 1080i/29.97 video format while the other monitor showed the same video clip captured via the 720p/59.94 format. Five video clips were arranged in an order dependent upon movement. Video clips were arranged in two sequences involving the manipulation of motion: no-motion to high-motion (clips 1-3,5 and then 4); or high-motion to no-motion (4,5 and then 3-1). This created four separate video sequences: no-motion to high-motion sequences in 720p and 1080i; plus high-motion to no-motion sequences in 720p and 1080i. This was done to avoid ordering effects.

Respondents were asked to rate the quality of each monitor's video image for each set of paired video clips using a ratio-quality scale (see Appendix 1). One video clip was shown on both monitors simultaneously. The major difference between the images shown on each monitor at any given moment was the HD format. One monitor displayed 720p while the other showed 1080i. While different video timelines were used between different test groups, only the arrangement of each video clip was altered between test groups (high-motion to low-motion, or, low-motion to high-motion). Thus, all respondents were shown the same images an equal number of times.

Per the ITU's DSCQS recommended method, each video clip was initially shown to respondents via both monitors. During this initial showing, respondents were instructed not to record any evaluations. This clip was followed by three seconds of video black. The clip was then reshown and respondents were asked to record their evaluations of each monitor's image quality on their response forms. This was followed by 30 seconds of video black displayed on both monitors. These steps were repeated with the four remaining video clips.

The experiemental design was pre-tested four times amongst 24 respondents (six per group). All of the pre-tests suggested the experiemental design worked as anticipated and resulted in testable empiracle data. No manipulation checks were run.

Respondents.

Respondents for this experiment were undergraduate students from the University of Nevada, Las Vegas. Respondents were offered extra credit in exchange for their participation in this research. Experiments were run outside of class in an effort to randomize the sample pool. The expected number of respondents in this study was 60. This would allow for approximately 30 participants per group to view/rate each video sequence. Respondents filled out signup sheets prior to participation. This allowed them to pick what time/testing session in which they would participate.

Stimulus material.

Stimulus material was comprised of five separate video clips originally recorded in both 1080i and 720p.

Clip sequences were:

Clip 1: Stationary shot of a news desk with no movement within the frame.

Clip 2: Stationary shot of two news anchors delivering a real news story from behind a desk.

Clip 3: Stationary shot of people walking down a sidewalk during the day.

Clip 4: Stationary shot of people walking down a sidewalk at night.

Clip 5: Camera-pan of people walking down a sidewalk during the day.

Measure and response form.

Respondents were asked to evaluate the picture quality of each video clip by making a pen mark on the "vertical scale" (International Telecommunication Union, 2009, p. 14) response form (see Appendix 1). These vertical scales were printed "in pairs to accommodate the double presentation of each test picture" (p. 14). These vertical lines were accompanied by five descriptors of a quality scale. The scales were printed in blue ink and the results were recorded with black ink. Ten scales were printed per page in five-groups of two. The ratio scale was used as it is recommended by the ITU-R to help alleviate any quantization errors. The response form was designed so that the number '1' corresponded to the first video clip, the '2' corresponded to the second clip and so on. Each pair was also labeled with an 'A' and 'B". The 'A' represented the monitor on the respondents' left while the 'B' represented the monitor on the respondents' right. The descriptors ranging from bad to excellent are currently the descriptors recommended by the ITU-R for DSCQS method tests (International Telecommunication Union, 2009, p. 15). The response form recommended by the ITU-R was used except the scale divisions were removed. This was done so respondents would better interpret the form as a ratio scale as opposed to an absolute scale.

Coding response forms

Completed response forms were quantified by the following method. Each vertical line was measured in millimeters (100mm per line for this experiment's response form). Each respondent's responses were also measured in millimeters. Each of these respondents' measurements were then divided by the line's total measurement. This resulted in a number which was then converted into a percentage. Example: A respondent made a mark in the 'Excellent' category 94 millimeters up a vertical response line. The equation would have been calculated as 94/100 =.94. This would be recorded as 94%.

Experimental conditions.

Unfortunately, with many experiments it is impossible to eradicate all possible forms of internal invalidity. In this experiment, efforts were made to maintain as much consistency between test groups as possible. Each group of respondents were presented with, and then read the same script, (see Appendix 2) by the same researcher. This researcher was also the sole individual who conducted/oversaw each experimental group. The researcher tried to address each testing group in the same tone and manner.

Each experimental group viewed the same monitors (two identical 16:9 46-inch Sharp PN-465u monitors). University of Nevada Las Vegas electronics engineers attempted to ensure the monitors' luminance ratios, brightness and contrast were identical. The monitors were mounted at the same height and same viewing angle (with regard to the center of the testing area). These monitors remained stationary through all testing groups (they were bolted to steel poles and remained untouched throughout the experiment).

The stimulus material was arranged in an attempt to negate any differences which might have existed between the monitors. This was done by ensuring both formats were shown on each monitor an equal number of times. Per the ITU's recommendation, testing sessions were designed to take less than half an hour and in a room with low illumination. Respondents were seated so their viewing angle was less than 30° from the monitors. Also following the ITU's recommendation, respondents were seated roughly 13.5 feet from the monitors (this number was derived by multiplying the screen height of 32-inches by 5.2, the screen height ratio advised by the ITU for monitors of this height) (International Telecommunication Union, 2009, p. 4). Additionally, as Hands (2004) and Beerends & Stemerdink's (1992) research suggested audio accompanying video images impair research respondents' abilities to accurately judge video quality, no audio was played during this experiment.

The viewing area for this experiment was designed so that respondents facing the monitors were also facing a dark corner of a television studio. The monitors were mounted close to the walls and the materials behind the monitors were black. This was done in an attempt to keep respondents from being distracted and to keep their attention focused solely upon the monitors and the respondents' response forms. A television studio was also employed as it allowed for exact and repeatable lighting during each testing session.

As mentioned previously in this section, most experimental research has the possibility of unintended stimulus impacting dependent variables. This study is no different and although many factors were controlled, developed, created and arranged to allow for as much validity and reliability as (realistically) possible, some factors should be addressed. Although respondents were instructed to rate their perceived level of image quality, there may have been other factors influencing respondent perceptions. It is possible factors in the stimulus material created artifacts in this study's results. Hands (2004) and Beerends & and Stemerdink (1992) conducted research which they claim showed respondents were influenced by audio when they were judging image quality, even after being instructed to ignore audio quality and focus solely on image quality. This suggests factors in the stimulus material such as lighting, framing, type of motion and use/lack of motion vectors may also have impacted the responses of those who took part in the experiment. Future research may want to try and determine if these variables actually are factors, and if so, how they may be negated in the future. It was intended that by showing the same images simultaneously upon identical screens, most of these factors would be minimized or expunged altogether, leaving HDTV formats as the main-to-only experimental stimulus affecting viewer perceptions. This topic will be further discussed in Chapter 5.

Analysis plan.

T-tests were utilized to answer the three research hypotheses and one research question.

CHAPTER 4

RESULTS

This section reports the results based on the data collected. This chapter also discusses the statistical analyses utilized to determine the results and subsequent findings. Although results and findings are reported in this chapter, possible causes and effects of these results are discussed in Chapter 5.

General Results

The sample size consisted of 62 respondents. All respondents attended the University of Nevada, Las Vegas and were either undergraduate or graduate students. To ensure respondent confidentiality, information such as gender and age were not collected. To help with recruitment, all respondents were given extra credit in one of the classes in which they were currently enrolled. Volunteer respondents signed up for one of 22 testing session. All testing sessions were designed to be as similar as possible and had an eight respondent maximum. The experiment was designed to test 12 respondents in a pre-test. As those initial sessions went smoothly and there were no changes to the experimental methodology, those results were treated the same as all other respondent results and were incorporated into the resulting analysis.

H1: Hypothesis 1 stated the 720p HDTV format would be significantly preferred to 1080i's video format quality. The hypothesis was not supported. A t-test used to analyze 1080i and 720p responses showed no significant difference between the respondents' preferences between the two formats (see table 1). However, when looking at the mean average of respondent perceptions' of both 1080i and 720p, it is apparent 720p was the preferred video format (but not to a significant degree).

Table I

Respondent's Perceptions of Image Quality

	M	*SD*	*t*	*df*	*d*
1080i Quality:	353.01	61.12			
720p Quality:	367.50	54.65	-1.34	61	-.24

H2: Hypothesis 2 stated there would be a significant difference in perceived video quality as video movement increased. This theory was unsupported (see Table 2). With regard to 1080i responses for the first three shots (as motion increased) respondent response averages showed a drop in quality ratings. However, contradictory to the first three shots' pattern of respondent averages, the fourth shot which had a camera pan (so everything in-frame was moving relative to the camera) was preferred the most amongst respondents. The 720p video results also did not support H2. The average respondent responses increased, then decreased, and then increased again as motion increased. These results suggest motion levels in video may not impact viewer preferences in a direct and/or linear fashion.

H3: Hypothesis 3 stated 720p would be the preferred video format as motion increases. This hypothesis was not supported. Statistical analyses showed the opposite to be true for the clip with the highest level of motion. To significant degree respondents preferred the 1080i video format better than the 720p format when rating the clip with a pan $t(61) = 2.90$, $p < .005$. These results contradict the generally accepted theory 720p is a better video format when recording/displaying video involving motion (Péchard, Le

Callet, Carnec, & Barba, 2007; What You Need to Know; Ynan, 2006, p. 45). This finding is discussed further in Chapter 5.

R1: Research Question 1 was created in an effort to determine whether there would be a statistically significant difference in audience perceived quality of similar scenes shot during the day and during the night. To test this research question, two of the shots/clips were designed to show the same street from the same location. The major difference between these two shots was that one was recorded at night and the other was recorded during the day. When comparing the mean averages between the day and night clips, respondents preferred the daytime video regardless of format (see Table 2). This could be because the night time shot was recorded using +18 dB of gain. Also of interest, there was a significant preference for the 720p format with regard to the night time video $t(61) = -2.27$, $p < .02$. This suggests video shot after dark or in low lit areas may benefit from being recorded in the 720p format as opposed to the 1080i video format. These results will be discussed further in Chapter 5.

Table 2

Video Format Preferences

Shot	Format	*M*	SD	t	df	p	d
Desk	1080i	78.71	14.94				
	720p	76.48	19.3	0.67	61	0.502	0.12
Anchors	1080i	76.26	15.96				
	720p	78.4	15.86	-0.71	61	0.479	-0.13
Street Stationery	1080i	75.54	17.12				
	720p	72.98	17.64	0.8	61	0.427	0.14
Street Pan	1080i	81.08	11.99				
	720p	74.4	14.9	2.9	61	0.005	0.49
Night	1080i	52.61	20.46				
	720p	58.61	19.34	-2.27	61	0.027	-0.3

Chapter Review

This chapter discussed the results of the experiment and the statistical analyses performed. It was shown respondents do not prefer either 720p or 1080i video formats to a significant level. The results also showed video motion does not favorably impact viewer perceptions of HD video quality. Analyses also showed 1080i may be the preferred HDTV format when watching video that is high in motion. In addition, viewers preferred to watch video recorded during the day as opposed to night and preferred the 720p format to 1080i to a significant level when watching video shot at night. The fifth and final chapter will further discuss these findings and try to reason why they occurred. In addition, strengths, weaknesses and limitations of this study will also be discussed, as well as suggestions for future research.

CHAPTER 5

DISCUSSION

This chapter discusses and analyzes the results revealed in Chapter 4. The three hypotheses' and one research question's results are analyzed. This examination progresses into a discussion about 1080i and motion, and how HDTV formats are affected by lighting conditions. The strengths and weaknesses of this study are then discussed as well as ideas for future studies.

Discussion of Reported Data

H1: Format Preferences

If H1 were correct, 720p would have been the preferred video format to a significant level. Averaging the respondent responses showed 720p was preferred to 1080i overall. However, H1 was not supported as there was not a significant difference between respondent recorded perceptions of both HDTV formats. This suggests the formats may be so similar in design that novice television raters cannot denote a significant difference between the two.

H2: Motion and Format

Hypothesis 2 speculated as video motion increased so would the overall respondent ratings of the video quality. This hypothesis was not supported. Analysis of each clip's average response scores showed little to no evidence of viewer preferences increasing as motion also increased. By analyzing mean averages it is shown viewers preferred the stationary video clip of the news desk as opposed to the clip of two anchors (who moved as they read the news and addressed each other) sitting at the same desk reading the news. However, it should be noted the highest rated video was clip 4 which involved the highest level of motion (a camera pan of a street). The statistical analysis of results for H2 show an increase in motion does not ensure viewers' video preferences will also increase.

H3: 720p and Motion

Hypothesis 3 stated respondents would prefer 720p over 1080i to a greater degree when higher levels of video motion were displayed. This hypothesis was not only unsupported but also shown to be wrong. The 1080i format was preferred to a significant level ($t(61)=2.90$, $p < .005$) over 720p when respondents rated the clip showing the highest level of motion (clip 4 which involved a camera pan). These results contradict the television industry's widely held belief that 720p's progressive scan format makes it superior to 1080i when displaying video involving motion (What You Need to Know; Ynan, 2006, p. 45). This is discussed further in the latter section of this chapter labeled '1080i and Motion'.

R1:

The only research question in this study questioned whether respondents would prefer video shot during the day or night. When comparing respondent averages, it was shown the daytime video was preferred to the nighttime video. This may be a result of the nighttime video having a great deal of 'image noise' as it was recorded with a gain level of +18 dB. However, R1 did lead to one of this study's most important findings. Analyses of results discovered respondents preferred significantly, 720p over 1080i when displaying video recorded outside at night. This finding is discussed in the latter section of this chapter entitled '720p and Low Light'.

1080i and motion.

As a camera moves, it essentially makes all objects in frame move relative to the camera. The results of this study showed respondents preferred 1080i to a statistically significant level over 720p when observing the video clips involving a camera pan. These results contradict the current belief 720p is a better format for video involving motion (Péchard, Le Callet, Carnec, & Barba, 2007; What You Need to Know; Ynan, 2006, p. 45). This may be a result of 1080i images being comprised of more pixels (more than 2,000,000) than 720p (roughly 920,000 pixels). This use of more pixels may be why viewers preferred the 1080i format with regard to the video employing high-motion. The 1080i format (theoretically) has more detail than 720p, possibly allowing 1080i to better capture/display motion. If this is in fact the case, it would contradict the television industry's belief that 720p's progressive scan format is superior to 1080i when capturing/displaying video involving motion. Considering 720p and its progressive scan format's main advantage over 1080i is often considered to be its superior ability to record/display motion (Péchard, Le Callet, Carnec, & Barba, 2007; What You Need to Know; Ynan, 2006, p. 45) this could be the most important finding of this study.

720p and low light.

This study showed viewers prefer the 720p format to the 1080i format when recording and capturing video shot at night. Because of these results, the stimulus material was re-examined. In this examination, it appears as if the 720p format's colors were more crisp and clear than that of 1080i. This may be a result of the progressive scan format showing full frames (and as a result twice as many frames) as opposed to interlaced formats which only display one single field at any given time. It may be that video recorded at low light is negatively impaired when it is interlaced, as its colors may become less crisp or bleed more than video recorded in a progressive scan format.

Weaknesses of this study.

One of the major weaknesses of this study is that it was conducted in an experimental setting. Typical television viewers do not spend the majority of their television viewing experiences trying to examine the image they are viewing, which is what respondents were asked to do for this study. By asking respondents to examine the images they are viewing, one alters the way the respondents would typically watch television. This could have resulted in unintended artifacts impairing the results of this study.

Another weakness of this study was on its reliance upon human respondents. This experiment was designed to create testable empirical data generated from subjective beings. In doing so, respondent vision was not studied or turned into its own variable. As no two eyes are the same, no two respondents saw the same image in *exactly* the same way. Factors such as color blindness, macular degeneration and cataracts would impact respondent responses. This experiment, as well as many of the studies reviewed for this thesis, treated vision as an independent variable, almost assuming that all people see the same or ignoring this issue completely. This reliance upon human respondents may have also created other variables. The stimulus material was designed to be relatively gender neutral. However, the material is probably viewed differently by different genders. Issues such as males focusing their attention on the female news anchor may have created results that were not based upon image quality.

As discussed earlier in this study, McLuhan (1994), Ihde (1982) and Zettl (1982) argued that any communication which is mediated by technology is directly impacted by said technology. And, they argued intended messages of communication are also altered by the technology mediating communication, and, each alteration differs from (in a specific way) all other forms of technology. As this study sought only to determine viewer preferences of HDTV formats, it failed to address the concerns of how these formats alter the messages being delivered by each format.

Another weakness of this experiment was in the sample size, just 62 respondents. Although research for this study's literature review did find studies of a similar nature using fewer respondents, more respondents in this study may have yielded better and more statistically significant results. Just by looking at the study's results, some trends seemed to be emerging. Analyses of the mean averages of respondent responses showed 720p was the preferred HDTV format, but not to a statistically significant level. If the sample size had been larger, this experiment may have shown that 720p was the generally preferred HDTV format to a significant level. In addition, three of the five video clips rated by respondents did not show a statistically significant format preference. However, the mean averages of respondent responses did show a preference for the 1080i format when the clips showed the news anchor desk and the daytime stationary shot of the street (clips 1 and 3). Conversely, the 720p format was preferred when displaying the news anchor desk with people reading the news from behind the desk (clip 2). Again, it should be noted none of these trends were statistically significant. However, a larger sample size may have allowed one to determine if these differences were merely coincidence or were actually based upon viewer preferences.

This study also would have benefitted from having more stimulus material. By limiting the material to five separate clips, some of the results may be misleading. For example, 1080i was preferred when showing the only clip with a pan. This makes it seem like video involving a moving camera benefits from being recorded/displayed in 1080i. However, this could be a trend created by an unidentified variable. If there were more clips in the stimulus material involving pans, there would be more evidence that this kind of motion intensive video does/does not benefit from the 1080i format. Additional stimulus material may have also allowed for analyses, which could find the motion-levels where each format benefits the most.

Strengths of this study.

Perhaps the most important aspect of this study is that it shows the ITU's recommended method is a functional, experimental design with regard to viewers rating the HDTV formats broadcast in the United States. This experiment yielded testable empirical results clearly representing viewer preferences and showed 1080i was preferred to a significant level for video displaying high-levels of motion.

Another strength of this study was that the experiment was conducted in a television studio. This allowed for more control than most classrooms would permit, and allowed for uniformity of testing conditions. By using a studio, lighting conditions were recorded on a digital lighting console which duplicated lighting conditions for each session. This lighting design removed any reflections/flare from the overhead lights. In addition, lighting around the studio was deliberately set to eliminate distractions. The studio's temperature was also maintained such that all sessions would be conducted at the same temperature. Sessions were limited to a maximum of eight respondents so that respondents would be within the ITU's recommended distance and angle for DSCQS experiments.

Although the ITU's method was followed, it was deliberately modified in one way. The response form recommended by the ITU was a quality ratio-scale with small, horizontal hash marks at each of the five descriptors. These horizontal hash marks were removed from the response forms in this experiment (see Appendix I). This was done in an effort to have respondents better interpret the forms as ratio-scales and not ordinal-scales. Previous research has shown quality ratio-scales yield more reliable results (Allnatt, 1983, p. 205; White & Allnatt, 1980; MacDiarmid & Darby, 1982) than ordinal-scales or impairment-scales. However, these researchers did mention respondents still failed to complete all quality ratio forms correctly. For this experiment, all response forms appeared to have been filled out correctly. This may have been due to a combination of using the modified ITU's form which is fairly straight forward and intuitive, reading the script to respondents (Appendix 2), and the lack of horizontal hash marks. Future researchers may want to also remove these hash marks in an effort to prevent respondents from interpreting ratio-scales response forms as ordinal-scale response forms.

Suggestions for future research

The findings that 1080i was preferred significantly over 720p when respondents rated video involving high-motion (clip 4), were derived by performing statistical analyses to determine whether H3 was valid. These findings directly contradict the television industry's belief that 720p is a better HDTV format for capturing/displaying video involving motion (Péchard, Le Callet, Carnec, & Barba, 2007; What You Need to

Know; Ynan, 2006, p. 45). Future researchers may want to try and determine which format is really the *best* at recording/displaying motion. By using double stimulus experimental designs, future research could help to validate or contradict the novel findings of this study. In addition, the strengths and weaknesses of each format with regard to motion should be researched. Future research could make use of more variables than those used in this study, such as: slow pans, quick pans and different speed tilts combined with other variables such as: subjects that are not moving, subjects moving slowly, subjects that are moving quickly. By doing so, researchers may be able to show one HDTV format is better at capturing/displaying certain scenarios (like a hockey game involving quick subject-motion and quick camera pans) while other formats are better for recording other events (such as a curling competition comprised mostly of slow subject-motion and slow camera pans). Additional research would better enable the television industry to use the *correct tools* when recording video leading to better image quality and a *better* viewing experience for the viewer.

The finding that 720p was preferred to 1080i when recording video outside after the sun went down was unexpected, interesting, and could lead to other studies showing 720p is the better format for recording/displaying video at night. However, as only one nighttime clip was analyzed, more research is necessary before one can state viewers prefer the 720p format to that of 1080i when viewing video recorded at night. These results alluding to 720p's superiority at night may be a misinterpretation of results. As only one clip was analyzed at night, the results could show respondents preferred 720p not because of the lighting conditions, but because one format is superior when rating other factors than light levels, such as: the video was recorded using +18 dB's of gain, or because the respondents liked headlights better at night when recorded in 720p as opposed to 1080i. The limited nighttime stimulus material makes it crucial that there is more research conducted on the subject before one can determine whether any HDTV format is superior to the other when displaying video at night.

If this type of future research is conducted, it may try to determine when, where and why light levels affect each format. Researchers could test images captured under different lighting conditions in an experimental setting. This research could determine which HDTV formats are *best* for varying lighting conditions. It would be interesting to see if an experimental design could be created which determines the luminance level where one HDTV format stops being the *best* and is out performed by another, relative to viewer preferences.

By focusing their research on HDTV formats and lighting, researchers could also try and determine viewer format preferences when record device settings are altered. This could allow for results showing which format is preferred by viewers when shooting low light video under different settings such as: large/small f-stops, different gain dB settings, long/short lenses, polarized filters, etc. This could give camera operators and businesses the ability to better choose which gear they use in different situations, in an effort to deliver the best possible picture to their customers.

Further analysis of HDTV formats could positively affect the television industry and its viewers. Analysis of how television formats are impacted by different variables (such as lighting and motion) allows one to better choose which technologies in which to invest and should allow viewers to receive a more preferred image. This area of study could also impact the design of future television technologies.

Conclusion

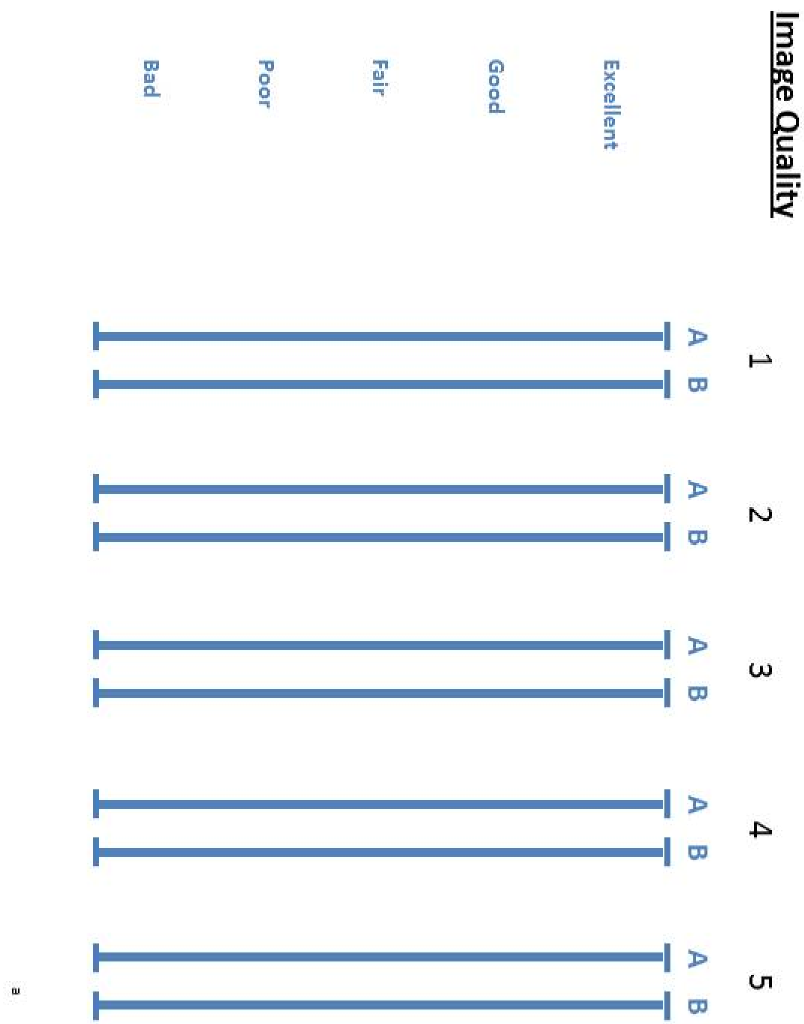
This experiment and its design joins a long list of previous studies which showed the DSCQS method is a valid way of having *regular people* rate video images which are similar. In doing so, it was the first study to compare the two main types of HDTV formats broadcast in the United States, via an academic setting. The main purpose of this study was to assess viewer preferences of the two main HDTV formats broadcast in America, 1080i and 720p. Specifically, this study tried to determine whether viewers generally prefer one HD format to the other; whether motion impacts viewer preferences of video quality; and if these differences in motion benefit a specific HDTV format. According to the data collected, the average respondent did not prefer one format over the other to a significant level. In the stimulus video clips, increases in motion failed to show a significant increase in respondent ratings. These results indicate the level of motion shown in a video clip does not directly impact viewers' perceptions of said video quality. However, data analyses showed the clip involving the highest level of motion (a camera pan of a busy street) showed respondents preferred 1080i to a significant level. These results directly contradict the television industry's widely held belief that 720p's progressive scan format is better suited to capturing/recording video involving high levels of motion (Péchard, Le Callet, Carnec, & Barba, 2007; What You Need to Know; Ynan, 2006, p. 45). Perhaps 1080i's design which uses over twice as many pixels as 720p allows it to overcome its reliance upon interlaced video, and actually capture/record certain types of motion better than 720p.

When analyzing the collected data to determine whether viewers prefer video shot at night or during the day, analyses showed viewers preferred the daytime shot more (perhaps a result of the night time video being shot with a gain level of +18 dB). Surprisingly, and perhaps of more importance is the data analyses which showed viewers preferred the night time video which was recorded/displayed in the 720p format as opposed to the 1080i HDTV format.

As this study had design flaws (namely small sample size and limited variation in stimulus materials), the results of this study are limited. However, these findings may be valued as a reason to doubt the television industry's relatively untested theory that progressive scan video is better at capturing/displaying video with high levels of motion. This study's findings could also be used as a starting point for future research. Researchers may want to try and re-validate these findings as well as explore them further by trying to determine which kinds/forms of motion, and lighting conditions, make viewers prefer one HDTV format over the other.

Appendix I

Example of Response Form



Appendix II

Script Read to Respondents Before the Experiment was Administered and Attached as a Cover Page to the Response Forms

Please do not communicate with others while this experiment is being conducted. Please read this cover page material to yourselves as I read it out loud for everybody. You are about to be shown five, separate, video clips. You will actually see two sets of video clips, side by side, on the two screens you see before you. Each clip will be shown twice. During the first showing, just watch the clips and think about each screen's video quality. The clips will then be re-shown. We will ask that during this second showing you record what you think about each video's quality.

Any Questions?

Now, turn the page and look at your response form. You will notice there are five pairs of vertical lines. Each pair is numbered, 1, 2, 3, and so on. The number 1 corresponds to the first video clip, the 2 corresponds to the second video clip, and so on. Each pair is also labeled A and B. The A represents the screen on your left, and the B represents the screen on your right. A is left screen. B is right screen.

You may have noticed there are five descriptors ranging from bad to excellent on the left side of your response form. These are there to help you determine where you make your mark representing how you interpret each monitor's image quality. Note also that the far left portion of the response form shows the words "Image Quality" and below it are a series of descriptors ranging from "Excellent to Bad". So, what we want you to do is watch a video clip on both monitors, and then rate the image quality of the images on both the left and right screens. Your responses should be recorded as a distinct horizontal line upon the appropriate vertical line. The higher you make your mark the better you are rating that particular monitor's image quality.

Any Questions?

So to review. You will watch a video clip on both monitors once, this clip will then be replayed and you will rate the image quality of both monitors. The higher the mark the better the image quality. The left monitor will be recorded on the 'A' column lines and the monitor on your right will be recorded on the 'B' column lines. Next, we will show you a second clip. This will also be re-shown once and you should record your image assessments during or immediately after this second showing. This will be repeated until all five clips have been shown and rated. Any Questions?

Appendix III

Script Read For Recruitment

I would like to thank [Professor's Name] for letting me take some of her/his class time to talk to you. My name is Michael Stubben, and I am a graduate student here at UNLV.   I am conducting an experiment in which I need volunteers. In exchange for helping me out and participating in my experiment, [Professor's name] has been generous enough to offer two-percent extra credit in this class to each of you who qualify, and participate, in my experiment. To qualify for participation in this experiment, you must be at least 18 years of age. In addition, if a doctor has determined you are legally blind you do not qualify for this experiment.

Participation is completely voluntary and basically involves analyzing television for no more than 40 minutes. There are multiple days and times you can participate. If you would like to volunteer and help me out you, need to sign yourself up on the sign in sheet outside of Dr. Traudt's office, GUA-2140. You can only register and get credit for one session. You must complete the entire session for credit…no partial credit will be allotted.

Again I would like to re-iterate: I would like you to volunteer for my thesis experiment. Participation is completely voluntary; there will be no bias for not participating. By successfully participating in this experiment, you will receive two-percent extra credit in this class. If this extra credit is offered to you in another class, you can only receive this credit in one of the classes. Therefore, you must choose which class you want the credit for at the time of the experiment…you will not receive credit in both. Participation will take less than 40 minutes. If you are interested you must sign up for a time slot on the sign up sheet outside of office GUA-2140, Dr. Traudt's office. Any questions?

For those of you who do not qualify for participation in this experiment, those under 18 years of age, and/or those with a high level of vision impairment, and/or any of you who do not want to participate in this experiment, there is another alternative exercise which will earn you the same amount of extra credit. To receive that credit you will need to find a recent article at the library that deals with television technology. You will then write a two-page, double-spaced paper discussing that article. Your paper should follow APA guidelines and should have the proper citation and a link to the article at the top of the page. Please let me know if you have any questions.

Appendix IV

Research Approval

UNLV

UNIVERSITY OF NEVADA LAS VEGAS

Social/Behavioral IRB – Expedited Review
Approval Notice

NOTICE TO ALL RESEARCHERS:

Please be aware that a protocol violation (e.g., failure to submit a modification for any change) of an IRB approved protocol may result in mandatory remedial education, additional audits, re-consenting subjects, researcher probation, suspension of any research protocol at issue, suspension of additional existing research protocols, invalidation of all research conducted under the research protocol at issue, and further appropriate consequences as determined by the IRB and the Institutional Officer.

DATE: February 15, 2012

TO: **Dr. Paul Traudt**, Journalism and Media Studies

FROM: Office of Research Integrity - Human Subjects

RE: Notification of IRB Action

Protocol Title: **Audience Image Preferences of 1080i and 720p High Definition Video Formats: An Experiment**

Protocol #: 1201-4004M

Expiration Date: February 14, 2013

This memorandum is notification that the project referenced above has been reviewed and approved by the UNLV Social/Behavioral Institutional Review Board (IRB) as indicated in Federal regulatory statutes 45 CFR 46 and UNLV Human Research Policies and Procedures.

The protocol is approved for a period of and expires February 14, 2013. If the above-referenced project has not been completed by this date you must request renewal by submitting a Continuing Review Request form 30 days before the expiration date.

PLEASE NOTE:
Upon approval, the research team is responsible for conducting the research as stated in the protocol most recently reviewed and approved by the IRB, which shall include using the most recently submitted Informed Consent/Assent forms and recruitment materials. The official versions of these forms are indicated by footer which contains approval and expiration dates.

Should there be *any* change to the protocol, it will be necessary to submit a **Modification Form** through ORI - Human Subjects. No changes may be made to the existing protocol until modifications have been approved by the IRB. Modified versions of protocol materials must be used upon review and approval. Unanticipated problems, deviations to protocols, and adverse events must be reported to the ORI – HS within 10 days of occurrence.

If you have questions or require any assistance, please contact the Office of Research Integrity - Human Subjects at IRB@unlv.edu or call 895-2794.

Appendix V

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

INFORMED CONSENT

Department of Journalism and Media Studies

TITLE OF STUDY: Audience Image Preferences of 1080i and 720p High Definition Video Formats: An Experiment

INVESTIGATOR(S): Dr. Paul Traudt, Principal Investigator; Michael Stubben, Associate Investigator

CONTACT PHONE NUMBER: 702-610-7702

Purpose of the Study
You are invited to participate in a research study. The purpose of this study is to determine which video image viewers prefer.

Participants
You are being asked to participate in the study because you fit this criteria: You are a student of higher education.

Procedures
If you volunteer to participate in this study, you will be asked to do the following: Watch multiple video clips being shown simultaneously on two monitors: Then rate your impression of each clip's image quality.

Benefits of Participation
There *may/may not* be direct benefits to you as a participant in this study. However, we hope to learn which video image viewers prefer.

Risks of Participation
There are risks involved in all research studies. This study includes only minimal risks. You may feel uncomfortable when asked to rate the different clip's image.

Cost /Compensation
There *may not* be financial cost to you to participate in this study. The study will take *40 minutes* of your time. You *will not* be compensated for your time.

Contact Information

If you have any questions or concerns about the study, you may contact Dr. Traudt at **895-3647.** For questions regarding the rights of research subjects, any complaints or comments regarding the manner in which the study is being conducted you may contact **the UNLV Office of Research Integrity – Human Subjects at 702-895-2794 or toll free at 877-895-2794 or via email at IRB@unlv.edu.**

Voluntary Participation

Your participation in this study is voluntary. You may refuse to participate in this study or in any part of this study. You may withdraw at any time without prejudice to your relations with the university. You are encouraged to ask questions about this study at the beginning or any time during the research study.

Confidentiality

All information gathered in this study will be kept completely confidential. No reference will be made in written or oral materials that could link you to this study. All records will be stored in a locked facility at UNLV for 3 years after completion of the study. After the storage time the information gathered will be destroyed.

Participant Consent:

I have read the above information and agree to participate in this study. I am at least 18 years of age. A copy of this form has been given to me.

____ __________________

Signature of Participant Date

____

Participant Name (Please Print)

Participant Note: Please do not sign this document if the Approval Stamp is missing or is expired.

REFERENCES

A digital video glossary. (1994, March 7). *InfoWorld*, Retrieved September 27, 2009, from Military & Government Collection database.

About the EBU. (n.d.) Retrieved October 4, 2011, from http://www.ebu.ch/en/about/index.php

Advanced Television Systems Committee Incorporated. (2008, July 29). *ATSC Standard: Video System Characteristics of AVC in the ATSC Digital Television System, 10.*

Allnatt, J. W. (1980). Subjective assessment method for television digital codecs. *Electronics Letters*, *16*(12), 450-451.

Allnatt, J. (1983). *Transmitted-picture assessment*. New York: Wiley-Interscience Publication.

Beerends, J. G., & Stemerdink, J. (1992). A perceptual audio quality measure based on a psychoacoustic sound representation. *Journal of Audio engineering Society, 40,* 963-974.

Bernath, K., Kretz, F., & Wood, D. (1981). The EBU method for organizing subjective tests of television picture quality.*EBU Review-Technical, 186,* 66-75.

Colucci, D. (1993, August 2). ‘Grand Alliance’ to develop HDTV standard. *Design News*, *49*(15), 48.

Comcast Corporation. (2011, May 25) Comcast announces milestone in TV viewing: Consumers watch 20 billion entertainment choices on demand. *Business Wire*.

Committed to Connecting the World. (2011). *International Telecommunication Union*. Retrieved October 3, 2009, from http://www.itu.int/en/about/Pages/default.aspx

Federal Communications Commission. (n.d.). *What You Need to Know About Digital TV Sets*. Retrieved October 2, 2009, from http://www.dtv.gov/needtoknow.html

Finn, N. (2006, October). Rival formats stake out turf. *Television Week, 25*(41), 26.

Frames and fields. (2008, January). *Broadcast Engineering*, *66*. Retrieved September 22, 2009, from Film & Television Literature Index database.

Gaps persist between number of HDTV owners, HD viewers. (2007, Dec 11). Retrieved September 22, 2009, from Film & Television Literature Index database.

Hands, D. S., (2004). A basic multimedia quality model. *IEEE Transactions on Multimedia, 6*(6), 806-816.

HD TVs now the majority, but HD viewing lags behind. (2010, Nov 8). Retrieved October 6, 2011, from http://blog.nielsen.com/nielsenwire/media_entertainment/hd-tvs-now-the-majority-but-hd-viewing-lags-behind/

Hill, J. (2003, May). HDTV? 1080i? 720p? Say what?. *Presentations*, *17*(5), 40.

Hoffmann, H., Itagaki, T., Wood, D., Hinz, T., & Wiegand, T. (2008, March). A novel method for subjective picture quality assessment and further studies of HDTV formats. *IEEE Transactions on Broadcasting*, *54*(1), 1-13.

Ihde, D. (1982). *Interpersonal communication: Essays in phenomenology and hermeneutics.* J. J. Pilotta, (Ed.). Washington D.C.: University Press of America.

International Telecommunication Union. (2009). Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT.500-12.*

Krauss, J. (2005, December). Here comes 1080p—maybe. *CEDMagazine.com*, *31*(13), 58-1.

Leopold, G. (1994, September 26). HDTV aces field trials. *Electronic Engineering Times*. Retrieved September 22, 2009, from Academic Search Premier database.

Lillington, K. (2010, September 24). Do not adjust your screens. *Irish Times*.

Linecker, A. (2004, November). Meeting HD. *Macworld*, *21*(11), 20.

Loncaric, M., Tralic, D., Brzica, M., Vukovic, J., Lovrinic, J., Dumic, E., & Grgic, S. (2008). *Testing Picture Quality in HDTV Systems*. ELMAR, 2008, 50th International Symposium (*1)*5-8, ISBN: 978-1-4244-3364-3.

MacDiarmid, I. F., & Darby, P. J. (1982, April). Double-stimulus assessment of television picture quality. *EBU Review (162)*, 71-78.

McLuhan, M., & Lapham, L. H. (1994). *Understanding media: the extensions of man* (1st MIT ed.). London: The MIT Press.

Medoff, N. J., & Tanquary, T. (1998). *Portable video: ENG and EFP* (3rd ed.). Boston: Focal Press.

Narita, N. (1994). Subjective-evaluation method for quality of coded images. *IEEE Transaction on Broadcasting*, *40*(1), 7-13.

Narita, N., & Sugiura, Y. (1997). On an absolute evaluation method of the quality of television sequences. *IEEE Transaction on Broadcasting, 43*(1), 26-35.

PC Mag.com. Retrieved October 3, 2011, from http://www.pcmag.com/encyclopedia_term/0,2542,t=codec&i=39939,00.asp#fbid=ZpipyFNly-S

Péchard, S., Le Callet, P., Carnec, M., Barba, D. A new methodology to estimate the impact of H.264 artifacts on subjective video quality. (2007). *International*

Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM. hal-00275334, (1) Apr 23, 2008).

Robertson, J. (1995, November 13). FCC approves HDTV design. *Electronic Buyers' News, 981*, 8.

Robin, M. (2001, February). Digital standards defined. *Broadcast Engineering*, *43*(2), 34.

What You Need to Know About the Digital TV Transition. Retrieved September 18, 2009, from http://www.dtv.gov/glossary_P.html

White, T. A. & Allnatt, J. W. (1980). Double-stimulus quality rating method for television digital codec's. *IEEE Electronics Letters*, *16*(18), 714-716.

WRNY to Start Daily Television Broadcasts; Radio Audience Will See Studio Artists. (1928, August 13). *The New York Times*, 13.

Ynan, D. (2006, November). The lowdown on high-definition tv. *PC World*, *24*(11), 45.

Zettle, H. (1982, August). Motion picture and video production. *Spectra*, *18*(7), 9.

Zettle, H. (2005). *Sight sound motion: Applied media aesthetic* (4th SFSU ed.). Belmont, Ca: Thomson Wadsworth .

VITA

Graduate College
University of Nevada, Las Vegas

Michael T. Stubben

Degrees:

Bachelor of Arts, Communications Studies (Emphasis in Broadcast Journalism)
University of Nevada, Las Vegas

Thesis Title: Audience Image Preferences of 1080i and 720p High Definition Video Formats: An Experiment

Thesis Committee:
Chairperson, Paul Traudt, Ph. D.
Committee Member, Gary Larson, Ph. D.
Committee Member, Lawrence Mullen, Ph. D.
Committee Member, Tara Emmers-Sommer, Ph. D.
Graduate Faculty Representative, Tara Emmers-Sommer, Ph. D.