MEASUREMENT AND ANALYSIS OF NATURAL VIDEO MASKED DYNAMIC DISCRETE COSINE TRANSFORM NOISE DETECTABILITY

By

JEREMY PAUL EVERT

Bachelor of Science in Mechanical Engineering
Kansas State University
Manhattan, KS, USA
2003

Master of Science in Electrical and Computer Engineering
Oklahoma State University
Stillwater, OK, USA
2010

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2015

MEASUREMENT AND ANALYSIS OF NATURAL VIDEO MASKED DYNAMIC DISCRETE COSINE TRANSFORM NOISE DETECTABILITY

Dissertation Approved:

Dr. Damon Chandler

Dissertation Adviser

Dr. Keith Teague

Dr. Guoliang Fan

Dr. Joe Cecil

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Damon M. Chandler, for his encouragement, interest, and patience. Dr. Chandler has opened my mind to a new world. Thank you for your time and kindness. I hope I can enhance the learning experience of my students in the ways you have.

I appreciate Dr. Teague, Dr. Fan, and Dr. Cecil serving on my committee. Thanks to all the professors from the College of Engineering, Architecture, and Technology for teaching important lessons and useful skills through courses offered at OSU. I am grateful for the many teachers I have had throughout my educational process.

I would also like to thank my friends from the Computational Perception and Image Quality laboratory. You have been kind partners, always offering me support and encouragement. I am in your debt for your assistance. I would also like to thank all the kind souls I have worked with in various labs during my time at OSU. I appreciate your patience and enthusiasm as we traveled down this path of learning together.

Finally, I would like to thank my family for their encouragement, support, and love. They have always been beside me to help me overcome the challenges and enjoy the resulting success. Thank you to my mother and father for forming me into the person I have become. Thank you to my wife Amanda for your unfailing love and encouragement. I am thankful for Magdalena; I appreciate you being a happy and healthy child, and I love who you are!

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: JEREMY EVERT

Date of Degree: MAY, 2015

Title of Study: MEASUREMENT AND ANALYSIS OF NATURAL VIDEO MASKED DYNAMIC DISCRETE COSINE TRANSFORM NOISE DETECTABILITY

Major Field: ELECTRICAL AND COMPUTER ENGINEERING

Abstract:

Lossy video compression lowers fidelity and can leave visual artifacts. Current video compression algorithms are guided by quality assessment tools designed around subjective data based on aggressive video compression. However, most consumer video is of high quality with few detectable visual artifacts. A better understanding of the visual detectability of such artifacts is crucial for improved video compression. Current techniques of predicting artifact detectability in videos have been largely guided by studies using no masks or using still-image masks. There is limited data quantifying the detectability of compression artifacts masked by natural videos. In this paper, we investigate the effect of natural video masks on the detectability of time-varying DCT basis function compression artifacts. We validate the findings from Watson et al. [JEI 2001], who found that as these artifacts increase in spatial and temporal frequency, detection contrast thresholds tend to increase. We extend this work by presenting compression artifacts with natural videos; when artifacts are shown with natural videos, this relationship between artifact spatial frequency and threshold is reduced or even reversed (our data suggests that some natural videos make targets easier to detect). More generally, our results demonstrate that different videos have different effects on artifact detectability. A model using target and video properties to predict target detection thresholds summarizes these results. We expand these results to examine the relationship between mask luminance, contrast, and playback rates on compression artifact detectability. We also examine how the detectability of targets that are spatially correlated with mask content differ from the detectability of uncorrelated targets. This paper's data serves to fill-in an understanding gap in natural-video masking, and it supports future video compression research.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Video compression enables digital media deliveries to the board room, the living room, and everywhere in-between. Digital video availability is a well-established part of everyday life, and the problems associated with large video files seem almost as well-established for engineers. At this time, about 300 hours of video are uploaded to youtube every minute.[1]

Understanding of video compression artifact detectability is vital for future video compression research. Aggressive but lossy compression can adversely effect the video appearance, often producing unsightly artifacts. Under-compressing files is wasteful, and may result in files too large to be useful. Because of the vast amount of video generated in the world each day, even small advances in video compression efficiency can have substantial impacts over time. It is imperative that we understand the detectability thresholds of compression artifacts.

The most common video and image artifacts come from discrete cosine transform (DCT) based compression [1, 2, 3]. In general, images in the spatial domain are transformed to the frequency domain using the DCT [4]. Individual DCT components are rounded, and then transformed back into the spatial domain [5]. Video compression is slightly more complex.

DCT basis function artifacts appear as differences between the original and compressed digital media, and are due in part to rounding of DCT components in the frequency domain. Artifact spatial frequencies in the digital media frames corre-

[1] https://www.youtube.com/yt/press/statistics.html

spond to the DCT components in the frequency domain. Various individual DCT components are rounded differently, and the amount of rounding for each DCT component should be related to artifact detectability. The detectability of DCT basis function artifacts caused by rounding should be the focus of ongoing research. Unfortunately, there is a significant gap in the understanding of video compression artifact detectability. Other than our previous work on the subject, no data has been gathered to specifically quantify the detectability of dynamic DCT noise when presented with natural videos.

This gap in understanding of artifact detectability is bounded on both sides by mountains of excellent research. On the functional side, engineers have successfully used DCT compression to reduce file sizes of images and videos. Engineers have continued research on the functional side, working with compressed videos as their benchmark. Many researchers and developers of compression algorithms use results of quality prediction algorithms as development aides. These useful tools were developed to predict more broad mean opinion scores. The opinion score is a subject's general personal assessment of a video viewing experience, usually given as a score from some scale. The score provides some assessment of how good or bad the results of the compression seem to the subject. Because personal opinions may vary, the scores of many subjects are averaged together to form a mean opinion score. Finally, because the scores are usually comparing the quality of several videos and video treatments, the scores are recorded as differential mean opinion scores, signifying how much better some videos are than others in overall quality. Several algorithms are able to provide predictions that fit the available data reasonably well, with little room for great improvement. Compression algorithm developers work to provide methods to reduce file sizes while maintaining better quality prediction algorithm scores. The direct investigation of the functional side of this problem has bore fruit quickly and abundantly, with great efficiency. This process has been successful in providing the

tools to build a vast empire of video data, with millions of videos delivered around the world everyday.

Unfortunately, these types of efforts are beginning to reach an asymptotic point in research, where greater efforts are required to provide only marginal improvements in video compression technology. Because of the size of the digital video market, even these marginal gains have significant impact around the world; however, there are other options to seek greater understanding of how to improve video processing. The other side of this gap in understanding of video compression artifact detectability is bounded by research from the areas of visual Psychology and visual Psychophysics. From this community comes the rigorous tools for the fundamental investigation of the simple question, "What is detectable to the eye?" Specifically, these tools allow the objective measurement of DCT compression artifact detectability thresholds. Our research path measures the level of contrast at which these targets are detectable, how the different spatial and temporal frequencies of the targets change detectability thresholds, and how that detectability changes from video to video. This is a much more time consuming path for data collection and analysis, requiring a great level of effort just to ensure the results are meaningful, but the outcome of such structured and dedicated research will help close the current gaps in video compression artifact detectability research.

Unfortunately, few studies have measured the relationship between compression artifact detectability and target spatial and temporal frequencies. The discrete cosine transform has seen vast applications in compression because of its effectiveness, not because of the depth of study of the eye's response to the resulting compression artifacts. One of the most complete studies was completed by Watson, Hu, and McGowan [6], who quantified the detectability of unmasked compression artifacts with temporal properties in support of their video quality metric development. They measured the detectability of an unmasked target they referred to as *dynamic DCT noise*, which

are DCT basis functions controlled for both spatial and temporal frequency. These targets are patterns of 8×8 pixel DCT basis functions, replicated to form a 256×256 pixel block, modulated in time.

The data on unmasked video compression artifacts presented by Watson, Hu, and McGowan [6] showed that unmasked compression artifact detectability thresholds followed previous trends shown with unmasked targets, unmasked compression artifacts, and unmasked targets with temporal properties. As with traditional targets, such as sine wave gratings [7], targets with higher spatial frequencies had higher detectability thresholds. Targets with higher temporal frequencies also had higher detectability thresholds [8]. Watson, Hu, and McGowan [6] described their data as generally low pass in form when plotting target detection thresholds versus either increasing spatial or temporal target frequencies. Watson, Hu, and McGowan [6] also provided a linearly separable model to summarize their data.

Although the body of related work is substantial, there have been no previous studies directly quantifying the detectability of video compression artifacts masked by natural videos. Watson, Hu, and McGowan [6] provide an excellent description of unmasked compression artifact detectability, however, they did not measure target detectability contrast threshold elevations due to presenting these targets with natural video masks. Other researchers have examined masking of compression artifacts with natural images [9, 10, 11, 12, 13]. However, these studies did not quantify the relationship between target detectability thresholds and target temporal frequencies. Nor have these related studies measured threshold elevations due to masking with natural videos, which also have temporal properties.

To help fill in this gap in knowledge, our research extends the work of Watson, Hu, and McGowan [6] to investigate how DCT basis function detectability changes when targets are presented with natural video masks. Our research applies principles from visual psychophysics to validate the detectability thresholds of dynamic DCT noise

in the unmasked condition. The data collected for this dissertation was from using the same process to extend what is known about compression artifact detectability by quantifying how target detectability thresholds change when dynamic DCT noise targets are presented in the presence of natural video masks.

Our study validates many of the previous conclusions for unmasked video compression artifacts, and extend these findings to quantify masked elevations due to presenting compression artifacts with natural video masks. Our measurements in this dissertation appear to suggest that presenting dynamic DCT noise targets with natural video masks can reduce or reverse trends seen in unmasked target research. A linear regression model summarizes our data from this dissertation for use by future researchers.

Our results represent a logical progression beyond the results of Watson, Hu, and McGowan [6], but do have notable limitations due to the time requirements of the experiments. To measure target detectability thresholds for the target spatial and temporal frequencies examined for this paper, each natural video was viewed as much as a few thousand times by each subject. As subjects complete experiments, they grow more familiar with the targets and masks, age, can have slight changes in attitude towards data collection, or even changes in vision, all of which can change a subject's results over time. Limiting the size of the dataset reduced the required collection time, and thus reduced the amount of change in each subject's results.

It should be noted that some in the visual Psychology field hold the DCT compression artifact as a radical target, and would rather this exploration be based on more controlled and more familiar targets such as the Gabor transform. Because DCT compression is so widely used, knowledge about the detectability of this target is vital. The study led by Watson provided one key stepping stone for our research to build on. The data from Watson on dynamic DCT noise is the best set of data to build off for this next step in video compression research.

There are several key differences between natural video masked dynamic DCT noise and the compression artifacts experienced by common consumers of digital videos. This target has only one spatial frequency, while most compression algorithms change multiple spatial frequencies. Dynamic DCT noise does not capture video motion prediction errors which are common with most compression algorithms. Finally, these targets are not spatially correlated with mask content. As this dissertation will show, even small changes in the target will result in changes in target detectability. Changing the dynamic DCT noise target to be spatially correlated with mask content results in slight changes in target detectability thresholds. Thus it was important that we first showed that we had properly recreated the targets used by Watson, and then presented them with masks.

The sample size for the data used in this dissertation was selected to ensure consistent and repeatable results from the volunteer subjects considering the time required to gather each data point, while providing an important next step beyond the research of Watson, Hu, and McGowan [6]. Our masked results in this paper are limited to eight 0.75 second, gray-scale videos which are four degrees of viewing tall and wide, with only a choice number of target spatial and temporal frequencies examined. Also, the data in this dissertation is restricted to detectability measurements at the threshold level using targets with only one DCT basis function. Given these limitations, this dissertation still provides a logical and meaningful extension of the findings of Watson, Hu, and McGowan [6].

We had previously shown some of our data in a conference paper; however, that work was provided without an analysis [14] . We later provided more of this work as a journal paper [15]. For the journal paper, we added data from a third subject, presented extensive analyses, and we investigated the efficacies of various models for predicting the data. For this dissertation, we have included a new set of experiments, controlling mask luminance, contrast, and playback rate. We also examined a new

type of target, correlated dynamic DCT noise, which is spatially correlated to the natural video mask content at the target spatial frequency.

Data collection followed internationally accepted principles and practices related to the ethical conduct of research involving the use of human subjects. Data collection methods were approved by the Oklahoma State University Institutional Review Board, under application number EG096. Informed consent was obtained from all subjects.

Chapter 2 provides a review and critique of related literature. Chapter 3 describes the data collection methodology. Chapter 4 presents our results and data reliability. Chapter 5 presents analysis and discussion of our data Chapter 6 presents modeling for the prediction of our results. Chapter 7 provides our results from experiments using masks controlled for luminance, contrast, and playback rate, as well as thresholds for targets spatially correlated with mask content. Chapter 8 summarizes our results and presents our conclusions.

CHAPTER 2

LITERATURE REVIEW

This chapter presents a review of literature related to the subject of natural-video masked video-compression-artifact detectability. Relevant literature provides context for the research presented in this dissertation. This chapter provides reasonable expectations for research results based on previously published articles. This chapter also details the gaps and limitations in previous research, describes which knowledge deficits this dissertation addresses, and summarizes which questions remain open for future research.

2.1 Targets with higher spatial frequencies tend to have higher unmasked thresholds

Many video compression algorithms are built on an assumed contrast sensitivity function. In general, human eyes can most easily detect targets with a spatial frequency of about one cycle per degree (c/deg) of viewing angle. One degree of viewing angle is about the width of a human finger at human arm's length. When targets have a higher spatial frequency, they are more difficult to detect. Said differently, higher spatial frequency targets are correlated with higher detection thresholds. Based on this assumed contrast sensitivity function, many compression algorithms compress lower spatial frequencies very little, and compress higher spatial frequencies aggressively. This section explains some of the research this assumption is based upon.

Comprehension of video compression artifact detectability is rooted in image compression artifact detectability [16]. Additional knowledge about compression artifact

detectability comes from research using other more controlled stimuli [7]. Our current understanding of DCT basis function artifact detectability has come from the vast research disciplines of visual psychology [17], physiology [18], and psychophysics [19]. Many of these studies focus on the detectability of various controlled visual stimuli, known as *targets*, which can be similar to DCT artifacts. An *unmasked target* is one presented against a blank background devoid of texture, and the brightness of both target and background are controlled. Human subjects identify the level of target brightness making the targets perceptible, or the *detectability threshold*.[1] The contrast between the target and background at this perceptibly level is known as the *unmasked target detectability contrast threshold*. Higher unmasked target detectability contrast thresholds signify greater differences in brightness between targets and backgrounds were necessary to make targets perceptible.

Unmasked targets, which can be similar to compression artifacts, have been shown to have detectability thresholds that vary as a function of target spatial frequency. Generally, humans have the lowest target detectability contrast thresholds for targets near one cycle per degree (c/deg), and have higher detectability contrast thresholds for targets with higher spatial frequencies. Relationships between target detectability and target spatial frequency have been measured with square-wave gratings [20], various other traditional targets [21], and even DCT basis functions [22]. Although the relationship between target detectability and target spatial frequency can be slightly altered by changing either target or background properties, in general, lower target spatial frequencies correspond to lower target detectability contrast thresholds, while higher spatial frequency targets have higher thresholds.

Campbell and Green [23] observed that as the spatial frequency of a target increases, observer sensitivity to that target decreases.[2] Campbell and Green were using

[1] Several animals have also participated in this type of objective study, including cats and monkeys.

[2] sensitivity is the inverse of the detectability threshold.

a neon-helium gas laser to produce an image on the retina. Although this method may seem a little unnatural, the results have held for many more natural viewing conditions.

In Psychophysics, understanding begins with simple and controlled stimulus, no matter how unnatural, and then builds on accepted truths towards more natural settings. Some researchers suggest that the human eye has been tuned for increased sensitivity to lower spatial frequencies because this is what typically occurs in natural scenes. Other studies have found that independent of many other variations in experiments, either by adding a mask, adding motion to the target, using only one eye instead of two, or even changing the brightness of the display itself, in general, finer details are harder for the human visual system to detect.

These related works suggest that higher spatial frequency dynamic DCT noise should have higher detectability contrast thresholds. Also, presenting a natural video mask with the targets should result in some change in target detectability contrast thresholds. However, the relationship between target detectability contrast thresholds and target spatial frequencies has not been previously quantified for dynamic DCT noise masked by natural videos.

2.2 Targets with higher temporal frequencies tend to have higher unmasked target thresholds

An important property of video, which separates it from work with images, is showing motion over time. This temporal component enables different relationships between target detectability and target temporal properties. Some of these relationships have been quantified by previous research. Expectations for unmasked video compression artifact detectability can be found in the related field of visual psychology. A traveling wave stimulus has the appearance that a sine wave is passing over the display. Examining a single point in space over time will result in a brightness that rises and

falls according to a sine wave pattern. The frequency of this pattern is measured in cycles per second or Hertz (Hz) and describes the temporal frequency of a stimulus.

Just as changing the spatial features of a stimulus from coarse to ultra-fine can increase target detectability contrast thresholds, making the target move or flicker quickly can also change detectability. There has been some disagreement and consensus about the models that explain the data over the past half century on this topic. Most of the contention in this discussion is about the narrowly tuned frequency mechanisms underlying the general shape of an overlying envelope of sensitivity. However, the data that defines this overlying envelope has been mostly consistent [24].

Two notable researchers, Kelly [25], Robson [8], and several others have shown that when the target is changed to a traveling wave, the speed of that wave can change contrast detection thresholds. In general, targets with higher temporal frequencies have higher detectability thresholds [19, 8, 26, 27]. When the temporal frequency of a stimulus gets high enough, target detectability contrast thresholds go up, no matter what the spatial frequency of the target. [1]

When target detectability thresholds increase, it is sometimes referred to as a threshold *elevation*. For instance, when a target temporal frequency is significantly increased, a significant threshold elevation will often result. Likewise, when a target spatial frequency is significantly increased, a significant threshold elevation will often result. Robson observed that the elevation due to a significant increase in target spatial frequency was largest when the target had little to no temporal frequency [8]. When target temporal frequencies increased, threshold elevations due to changing target spatial frequencies from low to high were reduced [8]. The data presented by Robson also suggests that the inverse of this relationship is true. As target spatial frequencies increase, target detectability threshold elevations due to changing target

[1] It is also important to note that when the spatial frequency is high enough, no matter what the temporal frequency, target detectability contrast thresholds increase.

temporal frequencies from low to high are also reduced [8].

Interestingly, when either the spatial or temporal frequency of the target is low enough, the stimulus is difficult to see. Kelly explored this in greater detail [25]. It is easiest for observers to see targets at a central range of temporal and spatial frequencies. High and low spatial or temporal frequencies are hard to see.

De Lange Dzn explored how flicker rate changed contrast sensitivity [28]. De Lange Dzn showed that some targets displayed a critical flicker frequency, where the target was most sensitive. One work that validated these principles was from Kelly [25], who showed a combined plot of models based on sensitivity to the coarseness and flicker rate of a stimulus.

Indeed, support for the general shape of the plots of target sensitivities at different spatial and temporal frequencies is even found in mammalian physiology. Tolhurst and Movshon [29] documented their recordings of visual neuron responses of an adult cat.

Schade [30] demonstrated a circuit meant to mimic the eye. The behavior Schade mimicked was spatiao-tempoarl sensitivity. Schade showed that a drifting sine wave was detected differently than a stationary one. The findings of Schade and de Lange Dzn are in agreement with those of Kelly.

Tolhurst [31] explored how target sensitivity changed as its temporal properties changed through different types of flicker modulation. Tolhurst found that a stationary grating was harder to see than a continuously moving grating, but a grating that had a sinusoidal modulation was the easiest of all three distortions to see. These findings were further supported by Kulikowski and Tolhurst [32].

Koenderink and van Doorn [33] made the claim that the underlying model of spatiotemporal sensitivity needs to be bimodal. Robson and Kelly had suggested that the underlying model was unimodal [8, 25]. However, Koenderink and van Doorn did not suggest that the data describing the response behavior for spatiotemporal target

detection was incorrect.

Cropper and Derrington [34] explored detection thresholds for unmasked targets with different temporal properties. Cropper and Derrington explored both detection of beat patterns and discrimination of motion direction. Cropper and Derrington found that it was easier for subjects to detect the beats than it was to determine their motion. Cropper and Derrington also found that the less time they displayed a stimulus, the harder it was for the subject for both the detection and discrimination task.

Unfortunately, few studies have measured the relationship between compression artifact detectability and target spatial and temporal frequencies. One of the most complete studies was completed by Watson, Hu, and McGowan [6], who quantified the detectability of unmasked compression artifacts with temporal properties in support of their video quality metric development. They measured the detectability of an unmasked target they referred to as *dynamic DCT noise*, which are DCT basis functions controlled for both spatial and temporal frequency. These targets are patterns of 8×8 pixel DCT basis functions, replicated to form a 256×256 pixel block, modulated in time.

The data on unmasked video compression artifacts presented by Watson, Hu, and McGowan [6] showed that unmasked compression artifact detectability thresholds followed previous trends shown with unmasked targets, unmasked compression artifacts, and unmasked targets with temporal properties. As with traditional targets, such as sine wave gratings [7], targets with higher spatial frequencies had higher detectability thresholds. Targets with higher temporal frequencies also had higher detectability thresholds [8]. Watson, Hu, and McGowan [6] described their data as generally low pass in form for both increasing spatial and temporal target frequencies. Watson, Hu, and McGowan [6] provided a linearly separable model to summarize their data.

The general expectation from previous research is that targets with higher tem-

poral frequencies should have higher detectability contrast thresholds. Additionally, targets with higher temporal frequencies should also be associatted with smaller elevations due to either masking or significant increases in target spatial frequencies. Although the work of Watson, Hu, and McGowan [6] provides data to verify for unmasked target detectability contrast thresholds, the relationships between target temporal frequency and natural video masked dynamic DCT noise contrast detectability thresholds has yet to be quantified.

2.3 Presenting masks with targets can effect target detectability

Studies with unmasked targets have provided useful guidance for compression artifact detectability. However, in video compression, artifacts are shown with natural videos. The detectability of targets presented against backgrounds with a texture, pattern, image, or video is known as *masked detectability*, and masks influence target detectability contrast thresholds [7, 35, 36, 37]. The difference in detectability thresholds between the masked and unmasked targets is known as the masked threshold *elevation*. Studies show that the mask contrast [38], mask spatial frequency, mask phase with respect to the target [7], and mask orientation with respect to the target [39] can all influence masked threshold elevations. In general, the largest changes in target detectability occur when the targets and masks are most similar. Also, masks with higher contrast cause larger elevations [38]. Targets with higher unmasked detectability thresholds tend to have smaller elevations due to masking [5]. The relationships between mask and target properties and detectability thresholds appear to hold true for masked compression artifact detectability [5].

Masks and targets more similar in temporal frequency have larger changes in target detectability due to masking [40]. Lehky [40] showed that as target temporal frequencies increased, the masks that resulted in the greatest target detectability threshold elevations were the ones that also had increased temporal frequencies. Fredericksen

and Hess reported similar results [41]. The relationship between mask and target temporal frequencies is similar to the relationship observed with targets and masks with similar spatial frequencies [38].

Previous investigations have suggested masking has many forms. One example is noise masking, where target detectability decreases because the mask corrupts the visual image [35, 42, 43, 44]. Contrast masking is where the contrast of the mask changes detection thresholds of the target [45, 38, 46, 47, 48, 49]. While exploring contrast masking, Swift and Smith [50] used two different types of experiments to measure detection thresholds and obtained different results. A further study of these differences found that the main deviations in the data could be explained by familiarity of the subjects with the masks. Familiarity with the mask or target, or lack there or, is sometimes examined under the heading of entropy masking [51, 52]. Apparent motion in a mask or target can also effect target detectability contrast thresholds [53].

Several papers exist on different types of masking [54]. However, measuring how noisy or how surprising a mask is can be a difficult task. Measuring the mask contrast is a more direct task, and has been studied in greater detail [45, 38, 48, 52, 55, 56, 57, 58, 59, 60, 61, 62].

Legge and Foley [38] explored contrast masking in human vision. When the mask has low contrast, the target is as easy to see as if no mask were present. However, as the mask contrast increases, target detectability contrast thresholds decrease, but only under certain circumstances. This is known as *facilitation*. In general, the most facilitation occurs when the mask has light contrast, and the mask and target are very similar in other measures. For example, the target used by Legge and Foley had a spatial frequency of 2 c/deg. When the mask had a spatial frequency of 1.0 c/deg, no facilitation was observed. However, when both target and mask have the same spatial frequency, there is a significant dip in the curve, signifying facilitation at some

low mask contrast levels. The area of the curve where the mask facilitates target detection is sometimes referred to as the dipper effect [35]. One observation from the work by Legge and Foley is that when the target and mask are most similar, masks with very little contrast are associated with target visibilities near the unmasked level. Increasing mask contrast slightly lowers target detectability, or causes slight facilitation. As mask contrast continues to increase, the effect of facilitation also increases up to a certain point. After that certain point, increasing mask contrast increases the target detectability contrast thresholds. That is, after a certain level of mask contrast, there is a positive correlation between mask contrast and target detectability contrast thresholds.

Swift and Smith [50] found similar results, and raised some questions about how the methods in the experiment could alter the detection thresholds. In general, target detection thresholds from most types of experiments have a similar shape. One of the seminal papers on the subject of contrast masking was by Campbell and Robson [21]. The plots of their data continue to be explored in greater detail. One example is from Bird et al. [63]. In general the findings from Bird et al. are the same as from Campbell and Robson. The reason for the exploration of the curves is a search for an explanation of why it is that the subjects generate the curves they do. And also, researchers explore the use of these types of curves to predict or understand what other related curves should look like. Some researchers even compare these types of threshold versus contrast of mask curves to similarly shaped curves resulting from different types of experiments [64].

One of the more popular contrast masking models by Watson and Solomon [45] applies the concept that neurons have the ability to control their own gain depending on the inputs they get from other neurons. This model helps predict how the eye sees distortions. Specifically, they suggest that the eye can adapt to a high or low contrast stimulus to avoid saturation.

Many different aspects of contrast masking have been explored. Peli et al.[65] explored peripheral contrast thresholds. This was a study of how contrast sensitivity changes as the target is more and more offset from the center field of view. There are other ways to study the interaction of stimulus contrast and sensitivity. One such method is adaptation.

Many studies from the world of Psychophysics have focused on controlled masks and controlled targets. The advantage of such experiments is that calculations for modeling are much more straight forward. The disadvantage is that the natural world is not made up of these straight forward masks and targets. And, as researchers such as Field [66] have suggested, the human visual system has been developed for observing the natural world, and so, research on human vision should also involve the natural world. Petrov [67] showed that human eyes are optimized to be most sensitive to ecologically useful information encoded with the luminance patterns of natural scenes. As we will see, many other fundamental aspects of human vision appear tuned to be most sensitive to the natural world.

The study of masked target detectability thresholds is more similar to the study of video compression artifact detectability thresholds, however, captures of the natural world have proven to be special types of masks [68]. Natural images cause unique masked threshold elevations [69, 10, 11, 12, 13]. Generally, natural scenes cause larger threshold elevations for low spatial frequency distortions, and threshold elevations are reduced for high spatial frequency distortions, although the specific amount of masking depends on image content.

Several algorithms utilize properties of natural scenes masks during image and video compression [70, 71, 72, 73, 74, 75, 76, 77], however, little data is available to quantify the effectiveness of natural scenes as masks [78, 69]. Watson, Borthwick, and Taylor [51, 79] measured the level of compression detectable in dental images in support of image compression algorithm development. Nadenau, Reichel, and Kunt

[12] measured the relationship between image compression detection probability and the level of image compression, and used this information to evaluate masking models.

Eckstein et al. [80] described how models of contrast gain control mechanism and background random variations made it harder for humans to detect targets. Eckstein et al. explain that in many practical tasks, target detection happens against a complex and spatially varying background. As masks or backgrounds, Eckstein et al. employed samples from patient digital x-ray coronary angiograms. Eckstein et al. found that detection performance is best against a uniform background. Detection is degraded against repeated samples of structured backgrounds, and detection performance is the lowest when backgrounds are not repeated but different samples of structured backgrounds. The work of Eckstein et al. appears to suggests that the complexity of the natural video should play a part in elevating detection thresholds.

Chandler and Hemami [11] showed compression artifact detection thresholds for masked targets were significantly different from unmasked targets. The data reported by Chandler and Hemami [11] suggested unmasked compression artifact detection thresholds were monotonically increasing in relation to increasing target spatial frequencies. When the same compression artifacts were presented in the presence of a natural image mask, the relationship was mostly similar. Masked threshold elevations were significant at lower target spatial frequencies, but reduced for higher target spatial frequencies. Near a target spatial frequency of 1.15 c/deg, the masked threshold elevation was nearly one log unit for both images tested. However, at a target spatial frequency of 18.4 c/deg, the error bars of some of the subjects' detection thresholds overlap for the masked and unmasked targets.

Chandler et al. [81] showed that the spatial correlation between masking images and targets should be considered when measuring the quality of distorted images. Hemami et al. [82] reviewed some of the related literature on what research from controlled masks and targets suggest to expect in related experiments with natural

image masks. Hemami et al. also showed how using an improved masking model, based on natural images, can improve masking predictions for detection thresholds in homogeneous natural image patches.

Chandler et al. [10] examined patch based masking in natural images. This was an effort to compare an accepted masking model with human observers. They controlled the patch content, as well as the patch contrast. Chandler et al. found that increasing the contrast of the mask would increase its ability to hide wavelet subband compression artifacts. Chandler et al. also showed that the content of the mask mattered more as contrast increased. At low mask contrast, what was in the mask did not make much of a difference in threshold elevations. However, when the contrast of the mask was increased, the patches that Chandler et al. classified as textures were best at hiding the distortions, followed by the structures. The patches classified as edges did not show a strong increase in detection thresholds as contrast of the patch increased. Chandler et al. attribute this to structural masking. A follow on study by Alam et al. [69] showed that the classification of masking ability by image content in a natural image would benefit from better masking models.

Chandler and Hemami [83] measured additivity and natural image masking. Additivity, or summation, is when the combination of two targets is easier to detect than either of the targets separately. Chandler and Hemami examined both threshold detection and above threshold discrimination. They used quantized wavelet distortions as the targets. What Chandler and Hemami found is that this type of target against a natural image background produced similar results to other studies using more controlled targets and stimulus. These results were expanded later by Chandler and Hemami [11].

Webster and Miyahara [84] found that natural images produced larger masking elevations for lower spatial frequency targets. Chandler and Hemami [59] suggested a transmission model based on the ability of natural images to mask lower spatial

frequencies more than higher spatial frequency artifacts. These works were extended by Chandler and Hemami [11].

Chandler and Hemami made a comparison of masked and unmasked detection thresholds of spatially correlated distortions. For the unmasked conditions, the most sensitivity exhibited was for distortions near 1.15 c/deg. Chandler and Hemami point out that this was not the typical frequency of highest sensitivity for unmasked sine wave gratings, which is more commonly in the range of 2-6 cycles per degree. Chandler and Hemami suggested that although a sine wave grating has only one spatial frequency, the wavelet transform distortions have an octave of spatial frequencies. The results of Chandler and Hemami were more similar to other results from unmasked experiments using targets that had a range of spatial frequencies near an octave.

Chandler and Hemami [11] observed that in most controlled masks and targets, it is most difficult to see the target when it is closely matched to the mask in spatial frequency. Most of the natural images have predominantly lower spatial frequency. During the experiments, it is seen that going from an unmasked condition to a masked condition, the largest elevations are seen at low thresholds, while higher frequency targets are not as dramatically more difficult to see against natural backgrounds as compared to the unmasked condition.

The general expectations from previous research on masked target detectability is that presenting targets with masks should cause some change in target detectability contrast thresholds. Using natural videos as masks should result in larger target detectability elevations for targets with lower spatial frequencies. Increasing masked target spatial frequencies from low to high should still result in an increase in target detectability contrast thresholds, however, this elevation should be less than the elevation due to the same change in spatial frequency for unmasked targets. Similarly, increasing masked target temporal frequencies from low to high should still result in

an increase in target detectability contrast thresholds, however, this elevation should be less than the elevation due to the same change in temporal frequency for unmasked targets.

2.4 Masked target detectability has a similar response to unmasked target detectability for increasing target spatial and temporal frequencies

Watson, Solomon, Ahumada, and Peterson [5, 45, 76, 58, 77, 22] explored DCT compression artifact detectability. Ahumada and Peterson [85] provide a detection model for DCT quantization artifacts. This model is based off principles from Psychophysical findings, and has some variation to meet the specific detection curves of DCT quantization coefficients. Watson et al. [58] extend this model to incorporate differences in viewing distance. Watson et al. also explored contrast masking. They employed a DCT quantization artifact as the mask, and then used another DCT quantization artifact as the target. They found that as the contrast of the mask increased, after a certain level, thresholds would increase when the mask and target were similar in spatial frequency. However, the elevations due to masking were not as large when the target and mask were significantly different in spatial content.

The general expectation from previous research is that masked target detectability contrast thresholds should increase as target spatial frequencies increase. This was similar to the behavior

The study of static images continues to be an active area of study by prominent researchers, who constantly make significant contributions to the understanding of the human visual system. One example is the work of Ahumada and Watson [86], who have applied the concept of contrast energy to account for viewing duration required for viewing detection and discrimination in static images. As seen in sensitivity to targets by themselves, the ability to hide a target can also depend on temporal

properties of the mask. Several studies explored how changing temporal properties of the target or mask would change detection thresholds [8, 40]. A common goal in these works is the drive to find a model of temporal vision for humans.

Kelly [87] provides a summary of expectations from a few classic visual contrast sensitivity experiments. Most of these experiments were for the detection of sine wave targets [87]. An unmasked target that is not too wide, not too narrow, and flickering some, but not too much is easiest to see. However, it can be made even easier to see if there is a mask present that is similar in flicker rate and spatial frequency, and just the right amount of very light contrast. In general, the hardest thing to detect is a high spatial and temporal frequency target against a high spatial and temporal frequency mask with high contrast. [1]

Burbeck and Kelly [88] explored masked detection threshold elevations for vertical targets with horizontal gratings. Burbeck and Kelly found that at low temporal frequency, there was not much elevation in detection thresholds across spatial frequencies as the contrast of the mask increased. However, as temporal frequency increased, lower spatial frequencies showed more increase in detection elevation as contrast increased. This trend held for temporal frequencies of 1.4 Hz through 30 Hz. Pantle [89] found similar behavior when the background had a spatial frequency three times the target frequency. However, Pantle only explored two steady state conditions against one 15 Hz temporal condition. Pantles results did not show that targets were easier to detect at all contrast levels when they had flicker.

Breitmeyer et al. [90] explored how the temporal frequency of the mask changed detection thresholds. Breitmeyer et al. explored many different components of vision, such as the ability to detect targets against steady backgrounds or backgrounds flickering at 6 Hz. Breitmeyer et al. showed that humans were most sensitive to

[1] Kelly was an excellent researcher who worked in visual sciences for over forty years and made several notable contributions to the field.

flickering targets at low spatial frequencies against steady backgrounds. However, as was common with most other related studies, a high enough target spatial frequency made most targets equally difficult to see, independent of target temporal frequency.

Green [91] explored the relationships between motion and flicker through adaptation. This effort was focused on the development of a model for motion perception. Because their interest was model development, Green [91] explored such controlled tests as using one eye versus using two eyes. Adapting the eye to flicker raised detection thresholds for drifting gratings, and low frequencies behave differently than high frequencies in the spatial and temporal domain. Smith had similar findings [92].

Lehky presented a study of purely temporal mechanics [40]. Lehky showed that when target and mask spatial frequencies were the same, the temporal frequency of the mask that would cause the most elevation in detection thresholds was the one that was more similar to the temporal frequency of the target.

Henning explored masking [93]. Henning used two targets, one low spatial frequency, and one high spatial frequency. Henning used low pass and high pass noise to mask the targets. The targets were presented with temporal frequencies of 0 Hz, 2 Hz, and 10 Hz. Results were similar to other researchers using static targets. However, Henning also incorporated a drifting stimuli, where the target and mask could have the same or opposite direction. Henning observed when the target and mask drift in opposite directions, the subject thresholds are similar to the unmasked condition. Henning examined this for a drifting velocity of 2.7 degrees per second as well as at 10.9 degrees per second. At the higher drift rate, there was no significant different between the same or opposite drift direction or the unmasked condition. This suggests that for one item to mask another, they need to move with similar speed. Said another way, the relative apparent velocity of the motion for the target measured against the velocity of the mask should be low in order to lead to masking.

Hess and Snowden [94] provide an exploration of a model for vision with contrast

masking. They show that target sensitivities are highest when the mask and target are closer in flicker rate. They show that this holds across many different spatial frequencies. These findings provide further conformation of the work of Robson and Kelly [8, 25].

Eckstein et al. [95] examined noise that has spatiotemporal properties. Eckstein et al. used dynamic noise to mask a temporally modulated signal. Eckstein et al. found that as the contrast of the masking noise increased, difficulty in detecting the target also increased. Eckstein et al. made comparisons between an ideal Bayesian observer and human observers. Eckstein et al. provided additional information to both humans and the ideal observer in the form of a cue. Eckstein et al. found that providing a cue helped the performance of the humans some, but the ideal observer more. Eckstein et al. concluded that this means that the humans incorporate the cue into the task, but not in an ideal manner.

Lu and Sperling [96] use a fixed background to test detection thresholds for targets with drifting luminance or targets with modulated texture contrast. Lu and Sperling measured both detection thresholds and discrimination thresholds. Lu and Sperling showed that as the mask contrast increased, the difficulty in detecting the target also increased in a manner suggested by Webers law.

Fredericksen and Hess [41] explored the concept of stimulus energy. They confirmed that when using a noisy mask with temporal component, the most masking occurs when the target and mask are similar in temporal frequency. They also show that increased mask contrast makes detection more difficult. Masking effects are less pronounced when the mask has the lowest contrast, or the mask temporal frequency is very high or the target temporal frequency is very low. Similar results with an additional observer are shown by Fredericksen and Hess [97].

Boynton and Foley [98] explored Gabor target detection against full field sine wave maskers. Boynton and Foley showed the greatest threshold elevation when the

temporal frequency of the sinusoidal mask was twice the temporal frequency of the Gabor patch.

Meier and Carandini [99] presented an excellent summary of previous work on drifting gratings. Meier and Carandini also provided a meaningful exploration of how drifting masks hide drifting targets. They showed that drifting masks hide best when they have more contrast and are closely matched in drifting rate to the target.

Laird et al. [100] expanded on previous work by Kelly [26], where they explored contrast sensitivity as a function of velocity across the retina. Daly offered a revised version of Kelly's work [24]. These works were an extension of previous explorations from Sekuler and Ganz [101]. Laird et al. [100] explored detection of Gabor patterns to populate a two dimensional spatio-velocity contrast sensitivity function. Laird et al. explored the relationship between eye velocity on target sensitivity. Similar to the target spatial and temporal velocity relationship, if a target has enough apparent motion velocity, or if the target has a high enough spatial frequency, the human visual system sensitivity to that target is negligible.

Watson [102] extended the work of DCT compression artifacts. Watson showed that given a sufficiently high target temporal frequency, a DCT compression artifact, without masking, can have a higher target detectability contrast threshold. Watson showed that the DCT compression artifacts detectability contrast threshold trends follow the general detection trends of other controlled targets seen previously in other visual Psychophysics experiments.

The general expectation from previous works on targets and masks with temporal properties is that the masks most similar in temporal frequency to the targets will result in the greatest elevations. Targets with lower spatial frequencies should still see the largest elevations due to masking. Also, targets with higher temporal frequencies should be associated with smaller elevations due to masking.

2.5 Mask luminance has some effect on target detectability

Watanabe et al. [103] showed that the mean luminance of sine waved gratings can cause a significant difference in the detectability of unmasked sine wave targets. Ahumada and Peterson [77] show that the luminance of the target could change detection thresholds as much as the spatial frequency of the target itself. In general, there is a negative correlation between increased target luminance and increased target detectability contrast thresholds. Said differently, targets are easier to see when they are brighter.

Snowden et al.[104] explored the relationship between luminance and detection thresholds. They found that brighter targets are harder to hide, and that detection curves for low luminance targets were not always simply scaled down versions of the higher luminance curves. Snowden et al. also note that as temporal frequency of the target increased, sensitivity decreased, no matter what the luminance of the target.

Although these researchers have not examined the effects of changing mask luminance explicitly, the general expectation is that increased luminance increases visibility.

2.6 Mask temporal content has some effect on target detectability

Graham [105] explored the drift rate of gratings using adaptation. Graham found that as the velocity of the grating increased, the sensitivity to the grating decreased. This was for a spatial frequency range from close to zero to about 20 c/deg. Grahams findings also supported the idea that higher spatial frequencies are harder to see, no matter what the grating velocity.

Kelly [106, 26] presents an exploration of a stabilized stimulus. Kelly wanted to show how much difference eye motion makes. Kelly presents several figures that showed how much the eye moves or does not move with respect to the stimulus can

dramatically change sensitivities.

Kelly [106] presented the image stabilization tool, as well as some measurements of stagnant target thresholds. The data from Kelly showed that eye stabilization greatly reduced spatial contrast sensitivity. Kelly showed that increased eye motion tends to increase target detectability contrast thresholds. When the eye is free to move around a stagnant stimulus, detectability thresholds were higher. When the eye motion is taken away, or when the view of the eye is matched with the stimulus, the sensitivity to the stimulus decreases.

In the second addition of the series, Kelly [26] provides a few more details to explain the relationship of eye movement and sensitivity. Kelly showed that even with stabilized images, higher spatial frequencies made targets harder to see. Kelly also showed that higher temporal frequencies, that is faster flickering, also made targets harder to see. Kelly showed that adding a little bit of motion to a target, that is to make it drift or travel a little, made the target easier to see.

Kelly showed that for a stabilized image, there is a certain target spatial frequency that excites the eye the most. The velocity of the target is going to change what the target spatial frequency curve looks like. For very small apparent motion velocities, adding a little target temporal frequency or flicker just makes the target stand out more. After a certain point, the faster the target is moving, the lower the spatial frequency the eye is going to be most sensitive to. Similar results were found by Watanabe et al. [103].

Levinson and Sekuler [107] provided reports to support the notion that directionally selective channels in human vision are independent of contrast detectors. Watson et al. [108] did a similar study to Levinson and Sekuler, but with gratings as the target. Watson et al. showed that in the middle range of spatial and temporal frequencies, it was as Levinson and Sekuler suggested. However, at the edges, where spatial frequency is high and temporal frequency is low, the data varied slightly from

the suggestions by Levinson and Sekuler.

Van Doorn and Koendrick [109] explored masked noise moving at two different velocities. They examined how the signal to noise ratio altered with detection thresholds. They also examined how the duration of stimulus display impacted thresholds. Van Doorn and Koendrick showed that in general, faster moving targets were harder to see.

Burr and Ross [110] took the study by Van Doorn and Koendrick one step further. They explored drifting gratings and moving bars. Burr and Ross showed that at high velocities, it is hard to see high spatial frequency targets. The slower the target would move, the higher the spatial frequency they could detect.

Watson and Ahumada [111] examined how fast a capture rate on a video camera needs to be or how fast a monitor needs to refresh. Watson and Ahumada measured this through a Two Interval Forced Choice (2IFC) method where the subject would select the stimulus that looked sampled. The control stimulus was a moving line, and the target stimulus was a sampled and flashed video of that same moving line. The sampled line was presented in two methods, one where it was simply flashed on and off, and the other was a stair case motion. First, the line would be on in one point, and held there. At some time later, the line would be displayed in a new location, without any significant duration of off time in between. Watson and Ahumada use this experiment to show that there is a certain window of detectability, and inside a certain spatial and temporal response window, people can see what is going on. However, when a change occurs beyond that window, the human eye has a hard time discerning any differences. Their model suggested a critical sampling frequency, the threshold where the user could distinguish between sampled and continuous movement. Watson and Ahumada showed that this had to do with the apparent motion velocity, or how fast the object was moving in c/deg.

Daly [24] converted the stabilized detection thresholds presented by Kelly [26],

and present them in terms of unstabilized detection thresholds. Daly suggests that this is more natural for modern display viewing.

Though there is considerable related work, no experiments prior to ours were expressly designed to measure the detectability of dynamic DCT noise when masked by natural videos. The general expectation from related work is that the mask temporal content can have some influence over target detectability. However, the amount of influence the mask temporal content can have is dependent on the target spatial and temporal frequencies.

2.7 Full-reference quality assessment algorithms can be useful for predicting compression artifact detectability

Watson and Nachmias [112] explored how contrast threshold data changed with different temporal properties of gratings to understand the model of visual detection. Watson and Ahumada [111] presented a model for visual-motion sensing. Watson [73] showed a very popular model of how to optimize DCT quantization matrices to adapt to individual images. Watson and Ahumada [113] examine several models of vision, stratify them in comparison to a large set of data, known as ModelFest, and then present a model they suggest as a possible standard for contrast sensitivity predictions.

Sachs et al. [114] presented a model of human vision that suggested different channels which are sensitive to different ranges of spatial frequencies. Georgeson and Sullivan [115] propose a model, like that of Sachs et al. made up of multiple channels. However Georgeson and Sullivan add that there seems to be some feedback mechanism that inhibits some of the spatial frequencies in such a way that makes images more clear. Peli et al. [65] further expanded on this model.

Sperling [116] presented a model of contrast detection. Quick [117] proposed a popular vector-magnitude model for contrast discrimination. Legge and Foley [38, 47]

shows that discrimination thresholds follow a power law. Pelli [118] described a different model of vision based on uncertainty. This model appears to be related to entropy masking. These models are based on psychophysical data, as well as physiology. One example is from De Valois et al. [119], who made physical measurements of what spatial frequencies a primate could see.

Daly [24] presented a spatiovelocity and spatiotemporal visual model for the purpose of understanding requirements for display design. Much of this effort is based on measurements provided by Kelly [26]. Fredericksen and Hess [41] present their model for temporal vision. This model iteration builds on other models by these authors, and describes the relationship between temporal sensitivity and target spatial frequency. Lee and Blake [120] presented a model of spatiotemporal vision. They suggest that phase dependent detectability in targets is not based on local luminance, luminance changes, or contrast. They suggest that the strongest responses happen when several features of a target are synchronized in amplitude, direction, and time.

Carrasco et al. [121] used contrast sensitivity to show support for the signal enhancement model of attention. This model explained some of the mechanisms underlying covert spatial attention. The focus of [121] is to understand how attention can impact contrast sensitivity. Jarvis and Wathes [122] addressed how the visual model of vertebrates needs adjustment based on how eyes see when light is low and primarily, how the cones in the retinal provide visual responses.

Watson and Malo [123] present the idea of a Standard Spatial Observer (SSO) tuned to predict video quality measures. This did not receive as much attention as Watsons other projects, such as DCTune [74], or the model of visual contrast gain control and pattern masking [45]. However, this was billed as Watsons simplified model that could predict just noticeable distortions. Watson has presented the idea of a standard observer in other forms [124, 125, 126, 127]

Watson had presented the concept of using a biologically plausible model of human

vision to evaluate an image and then make a measurement of its masking potential. The SSO was a low-complexity metric. Watson and Malo [123] present the SSO with a few enhancements and show its ability to reproduce some of the objective quality ratings from a large database from the Video Quality Experts Group (VQEG). At that time, this model was said to be as good as the best model that the VQEG group was offering. The interesting thing is that, although this model was designed to estimate the perceptual differences between a pair of 2D contrast patterns and assess the detectability of the differences between an original and distorted sequence, Watson and Malo did not provide validation data.

The ModelFest group [128] discussed the ModelFest project. In the first paragraph of the abstract, the group explained that models for a narrow class of stimuli are popular, but they want to make more general-purpose models to improve image processing algorithms, and that the Psychological measures used in the past have been too costly to gather.

One of the more simple precursors to the Watson SSO was the simple model by Legge and Foley [38]. Legge and Foley explored how detection threshold elevations changed as masking contrast increased. Legge and Foley used controlled stimulus for the mask and target, and were able to build a frequency specific model that could predict the upper linear range of a contrast versus threshold plot based on a specific fitting parameter for each frequency. In addition, Legge and Foley found that in the upper linear range of the contrast elevation plots, a constant slope of 0.62 would suffice for all different frequencies.

Foley presented an extension of the original Legge and Foley model [129]. In this model, Foley added an inhibitory channel to the image. Teo and Heeger presented a more simplified model [130]. These were the two papers that Watson and Solomon based their work off of [45]. The standard spatial observer has many similarities to these models. An extension of these models was presented by Watson et al. [6, 131].

Watson and Malo extended these ideas in a conference paper [123]. However, Watson and Malo [123] only presented the model as a way to estimate mean opinion scores for the video quality experts group database. The basic models of Watson et al. has not been validated with sound Psychophysical experiments measuring natural video masked dct compression artifact detectability.

In another interesting turn of events, current research in the non-biologically plausible area of research has came back to this same question of estimating noticeable differences in stimuli due to compression. Dr. Alan Bovik has recently published a work with Mittal and Moorthy on visually lossless compression [132]. The goal of this effort was to use high end statistics based off of models of human vision to measure video properties, then map those video properties into estimates of distortion detectability using machine learning. In essence, they were now using a non-biologically plausible model to predict if the eye can see a distortion or not.

How all of this ties together is that now there is a strong desire to figure out exactly what the eye can and can not see. The DCT based compression artifact is still a very common place occurrence in video compression. The expectations of how this artifact should be masked has not changed. The biologically suggested model has not been verified via Psychophysics in the predictability of the detectability thresholds. And now we see the non-biologically plausible camp coming back around to the same question.

This is not to say that any of the previous researchers have done anything wrong. However, the publication history seems to suggest that without a proper Psychophysical evaluation of how video of natural scenes mask DCT distortions, the two camps are on hold. First, the world of Psychophysics does not have the data to improve the models of how the eye perceives the DCT based compression artifact when it looks at natural videos. Second, the world of Engineering does not have the data to understand which higher order statistics are necessary to include during feature

extraction from an image to properly estimate other responses with machine learning techniques. To be clear, this is the question this dissertation hopes to provide insight for.

The data, analysis, and modeling provided in this dissertation quantifies the detectability of dynamic DCT noise when masked by natural videos. The data provided by Watson, Hu, and McGowan [6] serves as a launch pad for our experiment. We begin by validating the results of Watson, Hu, and McGowan [6], and then expand on their work to examine how presenting eight natural-video masks with the targets changes detectability thresholds. Based on previous research, it is expected that masked target detectability should be similar to unmasked detectability. The detectability of targets lower in spatial and temporal frequencies should be more influenced by natural-video masking. Additionally, the biologically inspired contrast gain control model [45] should be able to provide a reasonable prediction of masked dynamic DCT noise detectability.

CHAPTER 3

METHODOLOGY

In Chapter 1, we stated there is no previous data quantifying dynamic DCT noise detectability in the presence of natural video masks. This section describes our procedure for gathering the data to help fill in that knowledge gap. Our methods generally follow those described by Watson, Hu, and McGowan [6] whenever possible. Section 3.1 describes our procedure for data collection. The targets and masks are described in Section 3.2.

3.1 Procedure and subjects

This Section describes the data collection task, describes our subjects, and explains the order of data collection. Multiple subjects participated in data collection, as discussed in Sect. 3.1.1. Subjects completed a two interval forced choice (2IFC) task, as described in Sect. 3.1.2. Subjects provided target detectability threshold estimates for various masking conditions and targets, as described in Sect. 3.1.3.

3.1.1 Subjects

Here we describe the multiple subjects used for data collection. Collecting data from humans is time consuming and can produce unreliable or unrepeatable data. Often, multiple subjects are used as a way to reduce collection time, as well as to provide a means to validate results.

Our initial data set included measurements for 297 target detectability thresholds, from a combination of nine masking conditions, nine target temporal frequencies, and

eleven spatial frequencies. For the main data set, at least three subjects completed two sets of trials for each estimate, resulting in at least six sets of trials for each estimate. The six estimates for each target detectability contrast threshold were then combined with a weighted mean, $\overline{x}$, defined [133] as

$$\overline{x} = \frac{\sum (x_i/\sigma_i^2)}{\sum (1/\sigma_i^2)}, \tag{3.1}$$

were x_i is the mean from a single set of trials, and σ_i^2 is the standard deviation of a single set of trials. The weighted standard deviation, $\sigma_{\overline{x}}^2$, is defined [133] as

$$\sigma_{\overline{x}}^2 = \quad \frac{1}{\sum (1/\sigma_i^2)}. \tag{3.2}$$

All 297 detectability threshold estimates had two subjects in common, J.E. and K.J., with a third expert subject from the CPIQ lab at Oklahoma State University. The first observer for all data was J.E., the first author, a 32-year-old male with normal vision, who was experienced in detectability threshold experiments. The second observer for all data sets was K.J., a 25-year-old female with corrected to normal vision, who was a novice subject. The third observer was one of four subjects from the CPIQ lab at Oklahoma State University. All CPIQ subjects were males, had normal or corrected to normal vision, were in their 20s or 30s, and had extensive experience in detection experiments

3.1.2 Task for the subjects

This Sect. describes the 2IFC task completed by each subject. Each subject completed two sets of 32 trial 2IFC tasks for their portion of the data set. Data collection was aided by the use of the Psychtoolbox [134, 135] and the QUEST staircase method [136].

The *2IFC task* is where a subject views one video after another, watching for the target, and makes an indication of which video contained the target. Video presentation was computer controlled. The computer would randomly select the

order of the mask only and mask plus target, and control the presentation. The subject would enter their selection by pressing keys on the computer keypad, and receive audio feedback indicating either a correct or incorrect response.

The computer utilized the presentation software `Psychtoolbox` [134, 135] and employed the `QUEST` staircase tool by Watson and Pelli [136] to adjust target contrast levels for each trial. At the beginning of the set of trials, the target had high contrast. A correct response was followed by a stimulus with lower target contrast, while an incorrect response was followed by a stimulus with a higher target contrast. Subjects completed sets of 32 trials, and the target contrast levels were adjusted in such a way that the probability of a correct answer from the subject was 75% for the next trial. For the Weibull function to estimate thresholds, β was 3.0, δ was 0.02, and γ was 0.5. At the end of the experiment, `Psychtoolbox` and `QUEST` would provide a final target detectability contrast threshold estimate as a mean with a standard deviation. Subjects completed two sets of 32 trials. If the difference in means from the two sets was over 0.5 log units, the subject was asked to complete a third set of 32 trials, after which, the two closest measurements were kept.

It should be noted that during the presentation of initial data at the Asilomar conference, there was some concern about the size of the error bars of the data. One suggestion to combat this was to use a different monitor that allowed 14 bits of control, allowing 16,384 shades of gray to be displayed, instead of the 256 used for our current setup. This is expected to be most helpful during very low contrast experiments. During experiment setup, it was observed in examining some preliminary data, that indeed, a better monitor control could have helped provide a more controlled display of the target. However, this was only for the unmasked condition, and only at contrast levels that were far below detectability. For contrast levels near the detectability threshold for most combinations of targets and masks, the LCD monitor was able to sufficiently reproduce results with acceptable repeatability.

A different option to limit the size of the standard deviation of the data from the subjects would have been to control for their data standard deviation. Each trial resulted in a mean and standard deviation. Much as the subjects were asked to complete a third trial if the first two trials were not close enough in mean, subjects could have been asked to complete a third trial if either of the first two trials had a standard deviation that was determined to be too large. This was not done; however, this would appear to be an option that would be more similar to home viewing conditions by regular consumers.

The timing of each trial was closely controlled to ensure reliable presentation of the masks and targets. Each video was 90 frames in duration, and shown at 120 frames per second, resulting in a video length of 0.75 seconds. Each set of trials began with instructions written in black text against a gray screen. This told the subject what to do, and allowed their eyes time to adapt. Each trial started with a gray screen for 0.15 seconds, followed by a the first audio cue and the first video. Between the first and second video, the gray screen would be shown again for 0.15 seconds. A second audio cue would indicate the start of the second video. After the end of the second video, a final gray screen would be presented for 2.5 seconds.

The subject could enter a response any time after the beginning of the first video, but before the end of the final gray screen. If the subject did not enter a response during the appropriate time, the trial response would be counted as incorrect. Audio feedback would indicate if the subject response was correct or incorrect on each trial.

The timing of each trial was kept short to limit the amount of time required to complete the task. The first video of a new trial would begin 0.15 seconds after the subject response for the previous trial. The subjects would often respond during the first video or start of the second video. If the subjects waited until after the video to respond, the wait was usually less than a second. This rapid presentation allowed the subjects to complete the task quickly, but did raise subject anxiety. The 2.5 seconds

allowed after the second video was available to the subject, and reduced subject anxiety [17] because they felt they had sufficient time to respond. This additional time was rarely utilized.

3.1.3 Completion order of the sets of trials

The subjects completed at least two sets of 32 trials for each of the 297 mask and target combinations. This Sect. describes the order of completion of those combinations. The goal of this completion order was to reduce subject fatigue.

Each subject was assigned a list of mask and target combinations. The subject would complete 32 trials of the 2IFC task for the first mask and target combination in the list, then proceed to the next combination in the list. After completing the list of combinations the first time, the subject would start back at the top of the list, completing a second set of the 32 trials of the 2IFC task for the first mask and target combination. After completing the list of combinations a second time, the means for each combination were compared. A new list of mask and target combinations was formed, containing the list of of combinations were the means were larger than 0.5 log units.

The list of mask and target combinations was sorted by mask, target spatial frequency, and target temporal frequency. As the subject worked through the list, the mask would change least often, and target temporal frequency would change most often. Target temporal and spatial frequencies were sorted in ascending order, and masks were sorted in alpabetical order. This ordering of target and mask combinations reduced subject fatigue and improved data repeatability [17].

One limitation of this target and mask ordering is it allows the possibility of undesirable or uncontrolled influences of learning or adaptation during the data collection process. Presenting the same mask repeatedly and slowly changing target spatial frequency could have allowed either learning or adaptation. Indeed, by the end of

data collection, a subject was likely to see every video a few thousand times.

A process to shuffle mask and target combinations to reduce the likelihood of learning or adaptation was attempted. The initial results from the shuffled list provided similar means, but with larger standard deviations. The shuffled list took longer for subjects to complete, and subjects complained of fatigue, and required shorter data collection periods. Regan suggested [17] an increase in subject fatigue can lead to decreased data reliability.

3.2 Stimuli and apparatus

This section describes the videos presented to the subjects during the 2IFC task. The stimulus videos consisted of masks and targets, while the control videos were the masks alone. The targets are described in Sect. 3.2.1, while the masks are described in Sect. 3.2.2. Section 3.2.3 describes the measurement of the contrast between the target and mask. The display apparatus is described in Sect. 3.2.4.

3.2.1 Targets

This Section describes the targets subjects watched for during the 2IFC task. For our study, the targets were DCT basis functions modulated in time, which Watson, Hu, and McGowan [6] called *dynamic DCT noise*. The same target spatial and temporal frequency was used for a set of 32 trials, while target contrast was controlled according to the process described in Sect. 3.1.

For each trial, the target was 90 frames long, and each frame was a 128×128 pixel square. Each frame was divided into 8×8 pixel blocks. Each 8×8 pixel block is formed using a DCT noise template.

The DCT noise template is a DCT basis function. The DCT noise template is defined by the DCT component for the set of trials selected from the list of mask and target combinations, as described in Sect. 3.1.3. Target DCT components included

DCT [0,0], [0,1], [0,2], [0,3], [0,5], [0,7], [1,1], [2,2], [3,3], [5,5], and [7,7].

Generation of the DCT noise template, NT, begins in the frequency domain. The initial matrix NT_0 is an 8×8 matrix of zeros, representing the 8×8 matrix of DCT components for an 8×8 pixel block, $NT_0 = 0_{8,8}$. The desired DCT component, (j, k), of NT_0 is set to one thousand, $NT_1 = NT_0|_{(j,k)=1,000}$. This matrix is converted from the frequency domain into the spatial domain, $NT_2 = DCT^{-1}(NT_1)$. [1] This matrix is then normalized to have a maximum value of one, $NT = NT_2/max(NT_2)$.

The final target, FT, is formed by combining the noise template blocks after scaling them by the Gabor function. The contrast of each block, FT is defined for each frame by the Gabor function, a multiplication of a Gaussian window and a sine wave, as

$$FT_i = G_i \times S_i \times \alpha \times NT, \tag{3.3}$$

were i is the frame number. G is a Gaussian window equal to the number of frames in length with a standard deviation of 1/3. The Gaussian window gently transitioned the target contrast up to its peak, and then gently back to zero contrast over the duration of the 90 frames of the target. [2]

S, the sine wave, allowed control of the target temporal frequency over target frame number i, according to

$$S_i = \sin\left(\frac{\pi}{2} + \left[\frac{2\pi \times Q \times i}{m}\right] + P\right), \tag{3.4}$$

were m is the frame rate of 120 frames per second. Q is the target temporal frequency of 0, 1, 2, 4, 6, 10, 12, 15, or 30 Hz, as defined for the set of trials selected from the list of mask and target combinations, as described in Sect. 3.1.3. P is a random phase value for each block, distributed from 0 to 2π. A unique value of P was assinged to each 8×8 pixel block at the start of each trial, and used for all 90 frames of the target

[1] This was completed using the inverse two dimensional DCT function `idct2` inside MATLAB.

[2] The Gaussian window was generated using the MATLAB function `gausswin`.

for that trial. The purpose of P is that multiple blocks close together spatially were not likely to change in contrast in phase with each other.

The scaling variable α allowed the contrast of the target to be set close to the level suggested by `QUEST`, as described in Sect. 3.1.2. Each mask had unique content at the target spatial frequency, and that content changed over time. Thus, each mask required a unique value of α for each target temporal frequency, target spatial frequency, and contrast level. Because of a non-linear relationship between α and target contrast, a look up table was formed before data collection, and during data collection a polynomial fit of the data helped predict the α that would provide a contrast level close to that suggested by `QUEST`.

Special considerations are necessary to calculate target spatial frequency. A typical human subject has about two degrees in their focal area. The viewing distance was controlled so that the display was viewed at 32 pixels per degree of viewing angle. Each block is only 8 pixels wide. Measured horizontally across the target, the highest spatial frequency possible would be sixteen c/deg. However, the 8×8 pixel blocks of the target also change vertically within the focal area of the subject. So to report the target spatial frequency, f_s, as a number that represents both the target horizontal frequency, f_h, and target vertical frequency, f_v, a combined spatial frequency was found according to,

$$f_s = \sqrt{f_h^2 + f_v^2}. \tag{3.5}$$

For the given apparatus, the minimum target spatial frequency is 2.8 c/deg for the target DCT [0,0], and the maximum target spatial frequency is 22.6 c/deg for the target DCT [7,7]. Figure 3.1 shows a few frames from three example unmasked targets.

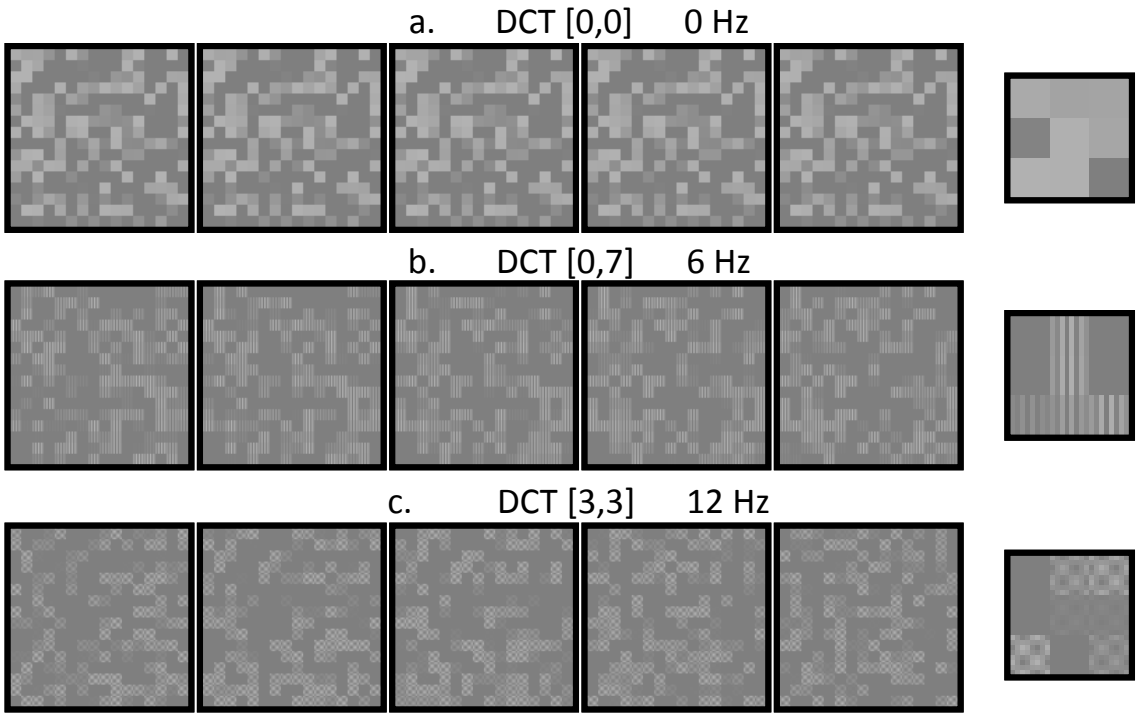

Figure 3.1: Example target frames. The top row shows frames 45 through 49 of an unmasked target, using the DCT basis function [0,0], with a target temporal frequency of 0 Hz. Note that the targets are made of small blocks, and in all five frames in row a, moving to left to right, the individual blocks don't change in contrast. The second row, b, shows frames 45 through 49 of unmasked targets, using the DCT basis function [0,7], with a target temporal frequency of 6 Hz. Observe that the distortions look like little vertical lines. Also note that across the five frames, not all 8×8 pixel blocks keep the same contrast. Finally, c, the bottom row shows frames 45 through 49 for the unmasked target, using the DCT basis function [3,3], with a target temporal frequency of 12 Hz. These distortions look more like dots than lines, and change in contrast just a little faster than those in row b. The unmasked condition used a gray source frame of with luminance of 45.1 cd/m^2, which would stand out from the rest of the background which had luminance of 43.5 cd/m^2. At the right of each row is a close up view of the upper left corner of frame 49 for each target, showing nine of the 8×8 target blocks.

3.2.2 Mask

This Sect. describes the natural video masks. An effort was made to use videos that were less controlled for content, contrast, luminance, and quality, but were more natural. Public domain high-quality color videos were subjectively chosen with a variety of content for a majority of the videos. Three videos, *Cactus*, *Kimono*, and *Flowervase* are standards videos, which are familiar to many video quality assessment researchers. These three videos are not available on the public domain, however, can be made available to researchers in this area.

Videos were converted to the bitmap file type, then converted to from three channel color videos to monochromatic videos according to

$$Z = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B, \tag{3.6}$$

were Z is the pixel input in the range of 0 to 255, and R, G, and B were the red, green, and blue channels of the initial color image. Videos were subjectively cropped to be equal in height and width around a main subject, maintaining a size larger than 128×128 pixels, and then resized down to 128×128 pixels using bicubic interpolation.[3] The native frame rates for the masks were 30 frames per second, so each frame was repeated four times to display properly at 120 frames per second.

Videos were subjectively chosen with a variety of content. Figures 3.2 shows a few frames from each of the natural video masks. Figure 3.2 presents frames from the natural videos *Waterfall*, *Cactus*, *Kimono*, *Hands*, *Timelapse*, *Lemur*, *Typing*, and *Flowervase*. The natural video *Waterfall*, has fine texture with fast vertical motion in the center, while the spray on the sides of the scene has little texture or perceptible motion [137]. *Cactus* has fine texture and fast horizontal motion [138]. The natural video *Kimono* has little texture or motion in the center of the scene, which is a female

[3]The default usage of the commands `rgb2gray` and `imresize` inside MATLAB made the monochromatic videos, and then resized them.

head and shoulders, with considerable texture and limited horizontal motion in the background [139]. Like the natural video *Kimono*, the natural video *Hands* contains familiar human features [140]. The hands counting in sign language in *Hands* have significant motion against a mostly blank background.

Figure 3.2 presented a few frames of natural videos that were captured at a natural frame rate from a fixed position, a few frames from one natural video captured with time lapsed exposure, and another with a moving camera perspective. The natural video *Timelapse* shows clouds with limited texture quickly moving across a sky devoid of texture [141]. *Lemur* shows a lemur with significant texture quickly jump into the scene against a highly textured but mostly still background of foliage and a second lemur [142]. The natural videos *Typing* and *Flowervase* contain modern man made structures. *Typing* shows hands with some texture quickly moving across a stationary keyboard background that has considerable texture [143]. The natural video *Flowervase* presents a moving camera perspective, were the viewer seems to move closer to a flower vase sitting on a table in front of wall containing a fire place [144]. The scene has significant texture, but is devoid of motion other than the changing camera perspective.

In Fig. 3.2, it can be seen that the eight masks have significantly different content. The scenes in the natural videos contain different types and levels of texture in the fore and background, as well as different types of motion. This variety of natural video types was subjectively chosen to examine if any types have significantly better or worse masking ability.

3.2.3 Contrast measurement

This Sect. describes how contrast between the target and mask was calculated. Watson, Hu, and McGowan [6] reported the log of contrast energy detection thresholds for unmasked dynamic DCT noise. Watson *et al.* described contrast energy [145]

Figure 3.2: Frames 1, 22, 45, 68, and 90 from the natural video masks. Row a shows five frames for the mask *Waterfall*. Row b is from *Cactus*, while c is *Kimono*, d is *Hands*, e is *Timelapse*, f is *Lemur*, g is *Typing*, and h is *Flowervase*.

as the integral of the square of the contrast over all dimensions in which it varies. Discrete contrast energy, CE, can be represented by

$$CE = \sum_{i=1}^{m} C_{RMS}(i)^2, \tag{3.7}$$

were m is the number of frames. C_{RMS}, the root mean square (RMS) contrast of the stimulus frame, was calculated according to

$$C_{RMS} = \frac{\left(\frac{1}{n}\sum_{j=1}^{n}(L(E_j) - \mu_{L(E)})^2\right)^{\frac{1}{2}}}{\mu_{L(I)}}, \tag{3.8}$$

where n is the number of pixels, E is the mean-offset stimulus frame, and $\mu_{L(I)}$ is the mean luminance, L, of the mask frame, I. E is calculated according to

$$E = \grave{I} - I + \frac{1}{n}\sum_{k=1}^{n} I_k, \tag{3.9}$$

were n is the number of pixels. $\grave{I}$ is defined as $\grave{I} = B + I$, where B is the target frame and I is the mask frame. [4] The luminance of the mean-offset stimulus frame, $L(E)$, and the luminance of the mask frame, $L(I)$, were calculated from the pixel values of the mean-offset stimulus frame and mask frame according to Eq. 3.10. As with the work by Watson, Hu, and McGowan [6], the final threshold was reported as the log_{10} of CE.

3.2.4 Apparatus

This section describes the physical setup for data collection. As this current research is an extension of the work by Watson, Hu, and McGowan [6], an effort was made to examine the same phenomenon. The data collection apparatus was configured to match the work by Watson, Hu, and McGowan [6] as closely as possible.

In this experiment, as well as in the work by Watson, Hu, and McGowan [6], viewing of the display was binocular with natural pupils. Data was collected in a

[4] All contrast calculations were carried out in MATLAB using double precision.

darkened room. Watson, Hu and McGowan used a cathode ray tube (CRT) display with a display frame rate of 120 Hz. The data for this paper was collected with an ACER gd235hz liquid-crystal display (LCD) with a refresh rate of 120 screens per second and a resolution of 1920×1080 pixels. Given the size and resolution of the LCD, a viewing distance of 51.5 cm was maintained to ensure subjects viewed 32 pixels in every degree of vision, as was used by Watson, Hu and McGowan.

The display was controlled with a dual-link digital visual interface (DVI) cable with 8-bit precision, allowing integer pixel-value inputs from 0 to 255. For this display, the luminance response yielded minimum and maximum luminance of 0.21 and 200 candela per square meter (cd/m^2), and luminance saturated after pixel-value inputs of 245. All videos, were clipped to ensure no pixel value was above 245. The videos were presented against a gray background with a constant luminance of 43.5 cd/m^2. For calculating contrast, stimulus pixel values, v, were mapped to luminance values via

$$L(v) = (0.084 + 0.037 \times v)^{2.41}. \tag{3.10}$$

CHAPTER 4

RESULTS FROM PRIMARY PSYCHOPHYSICAL DATA COLLECTION EXPERIMENT

This chapter presents the primary results of our research study measuring the detectability of dynamic DCT noise when masked by natural videos. This main set of data represents the weighted average of at least six target detectability contrast threshold estimates. The six estimates came from at least three subjects completing two sets of trials for each target detectability contrast threshold. Weighted averages were calculated according to Eq. 3.1 and Eq. 3.2. The chapter also includes analysis of the main data set's reliability.

Plots of masked and unmasked target detectability contrast thresholds over target spatial and temporal frequencies suggest there are noticeable and significant differences between masked and unmasked thresholds. Figure 4.1 plots target detectability thresholds over increasing target spatial frequencies at three target temporal frequencies. Fig. 4.2 plots target detectability thresholds over increasing target temporal frequencies at three target spatial frequencies. These figures show that presenting masks with targets changes target detectability contrast thresholds. Different masks resulted in different elevations in target detectability contrast thresholds. Also, changing target spatial or temporal frequency changes the effectiveness of the mask in altering target detectability contrast thresholds.

4.1 Target spatial frequency and masked target detectability

This Section discusses the relationships between target spatial frequencies and target detectability contrast thresholds. In general, targets with higher spatial frequencies have higher detectability contrast thresholds, which was expected based on previous research. Also as expected, when targets are higher in temporal frequencies, the target detectability contrast threshold elevations due to large increases in target spatial frequencies are reduced. This Sect. also shows that presenting natural video masks with targets can reduce or even reverse the effects of changing target spatial frequencies on target detectability contrast thresholds.

Figure 4.1 shows how target detectability contrast thresholds change as target spatial frequencies increase, for both masked and unmasked targets. The top row of plots in Fig. 4.1 show target detectability contrast thresholds when using DCT basis functions of [0,0], [0,1], [0,2], [0,3], [0,5], and [0,7], corresponding to target spatial frequencies of 2.8, 4.5, 6.3, 8.2, 12.2, and 16.1 c/deg, which are targets with vertical alignment. The bottom row of plots in Fig. 4.1 show thresholds for targets that have a diagonal alignment, and thus the horizontal axis of the bottom row of plots is different from the top row of plots for Fig. 4.1. The bottom row of plots in Fig. 4.1 show target detectability contrast threshold estimates when using basis functions of DCT [0,0], [1,1], [2,2], [3,3], [5,5], and [7,7], corresponding to target spatial frequencies of 2.8, 5.7, 8.5, 11.3, 17.0, and 22.6 c/deg. The first, second, and third column of plots in Fig. 4.1 show threshold estimates when using targets with temporal frequencies of 0 Hz, 6 Hz, and 30 Hz, respectively.

The solid black lines in Fig. 4.1 denote unmasked target detectability contrast thresholds. Observe from Fig. 4.1 that generally, for unmasked targets, as target spatial frequencies increase, target detectability contrast thresholds increase. This is in agreement with previous research on unmasked target detectability contrast thresholds and target spatial frequency [20, 21, 22, 6],

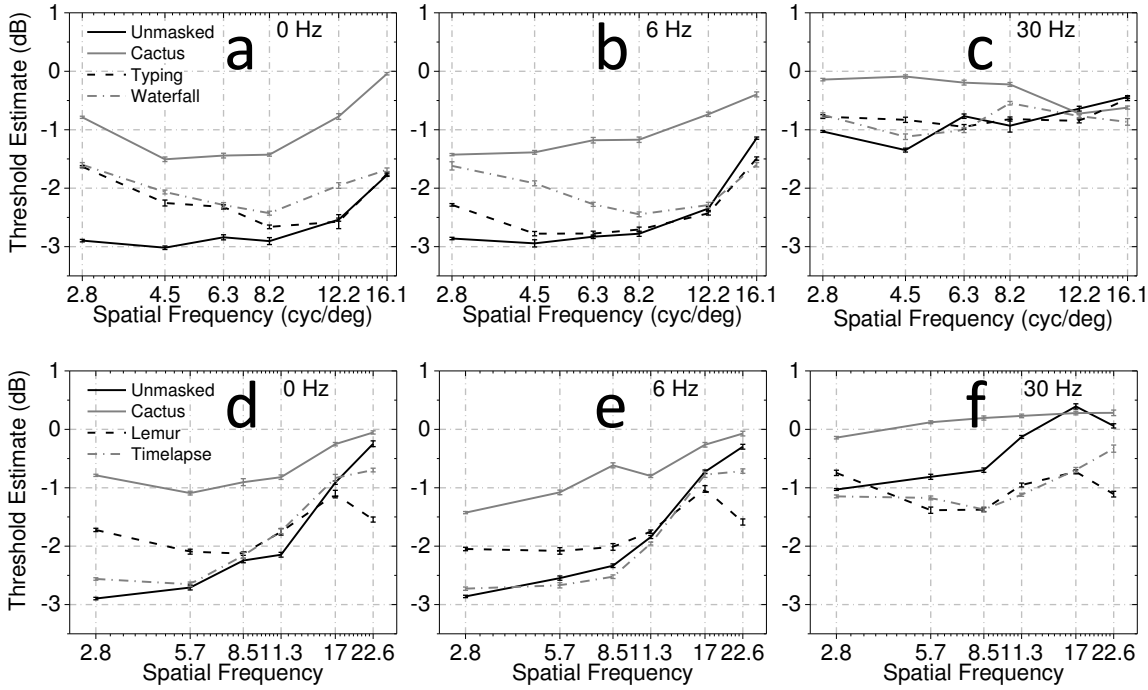

Figure 4.1: Target detectability contrast thresholds versus target spatial frequencies for masked and unmasked targets. The vertical axis shows the log_{10} of contrast energy of target detectability thresholds, the horizontal axis shows target spatial frequencies in c/deg, and the graph legend shows masking conditions used for each plot. The target temporal frequencies used in each plot are shown in the upper left hand corner of each plot.

The data in Fig. 4.1 suggests the relationships between masked target detectability contrast thresholds and increasing target spatial frequencies do not always match the relationships between unmasked target detectability contrast thresholds and target spatial frequencies. In Fig. 4.1 (a) and (d), it can be seen that at low target temporal frequencies and low target spatial frequencies, the different masks caused markedly different target detectability contrast thresholds. From Fig. 4.1 (d), it can be seen that the difference in masked and unmasked target detectability contrast thresholds reduces some at higher target spatial frequencies. This is in agreement with previous research on natural image masking of compression artifacts. Chandler and Hemami [11] had shown that unmasked quantization distortion detectability was similar to other unmasked target detectability; however, when masked with natural scenes, lower target spatial frequencies had considerable elevations, while higher target spatial frequencies experienced little to now change in detectability thresholds. This lower spatial frequency elevation is most noticeable for the mask *Cactus*, in Fig. 4.1 (d), where, at lower target spatial frequencies of 2.8 c/deg, there was about a two log unit difference in target detectability contrast thresholds; however, at higher target spatial frequencies of 22.6 c/deg, the unmasked thresholds are nearly the same as the thresholds for targets presented with the mask *Cactus*.

It was not expected that some natural video masks would reduce target detectability contrast thresholds. Observe in Fig. 4.1 (b), (c), (d), (e), and (f) that there are masked target detectability contrast thresholds markedly lower than unmasked target detectability contrast thresholds. This was not expected based on previous research, and is discussed further in Sect. 5.8 and 5.9.

4.2 Target temporal frequency and masked target detectability

This Sect. details our findings on the relationships between target detectability contrast thresholds and target temporal frequencies. The plots in this Sect. show that,

as expected, targets with higher temporal frequencies have higher detectability contrast thresholds. Also as expected, when targets have higher spatial frequencies, the elevation due to changing target temporal frequencies from 0 Hz to 30 Hz is reduced. Presenting a natural video mask with the targets can reduce the effects of a large change in target temporal frequency.

Figure 4.2 shows how target detectability thresholds change as target temporal frequencies are increased, for both masked and unmasked targets. The first, second, and third columns of plots in Fig. 4.2 show contrast thresholds when using target basis functions DCT [0,0], DCT [0,7], and DCT [3,3], respectively. Like Fig. 4.1, which plotted target detectability contrast thresholds versus target spatial frequencies, Fig. 4.2 shows how target detectability contrast thresholds change over the range of target temporal frequencies from 0 to 30 Hz. The top row of plots in Fig. 4.2 presents target detectability contrast thresholds for targets masked by *Cactus*, *Flowervase*, and *Timelapse*, while the bottom row of plots presents detectability contrast thresholds for targets masked by *Hands*, *Kimono*, and *Typing*. The solid black lines in Fig. 4.2 show unmasked target detectability contrast thresholds.

As shown in Fig. 4.2, as target temporal frequencies increase, unmasked target detectability contrast thresholds increase. When target spatial frequencies are increased from 2.8 c/deg to either 16.1 c/deg or 11.3 c/deg, target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 30 Hz are reduced. Previous research on unmasked target detectability and target spatial frequencies supports these observations [19, 8, 26, 27, 6].

4.3 Data reliability and repeatability summary

As this was the first data quantifying dynamic DCT noise detectability in the presence of natural video masks, there is no existing data set to directly compare our results to. For each threshold, we had two sets of trials from each of three subjects to compare

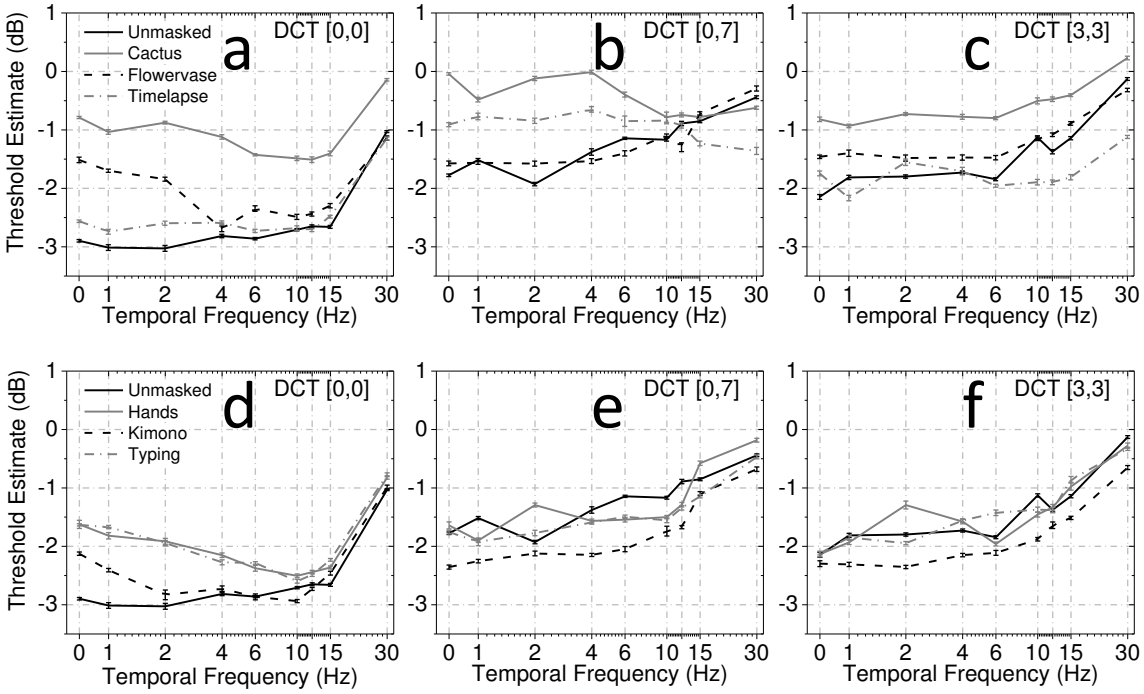

Figure 4.2: Target detectability contrast thresholds versus target temporal frequencies. The vertical axis shows the log_{10} of contrast energy of target detectability thresholds. The horizontal axis shows target temporal frequencies in Hz. The graph legend in the upper right hand corner of the left plot in each row shows the masking conditions used for each plot line for that row.

across. We determined reliability and repeatability by computing and measuring intra and intersubject agreement.

This Sect. provides a summary of intra and intersubject agreement. The detailed results of our reliability and repeatability analysis can be found later in this chapter.

Table 4.1 shows the averages of several measures of data repeatability between sets of trials for individual subjects, as well as between subjects, averaged over all target frequencies, masking conditions, and subjects. These measures assessed how well a single subject's first sets of trials matched their second sets, or, how well the results from one subject matched another. Both Pearson correlation coefficients (PCC) and Spearman rank-order correlation coefficients (SROCC) quantify the predictability of a second set of data based on a first set of data. [1] For both PCC and SROCC, the best possible score is 1. Root of the mean squared errors (RMSE) provide a summary of the differences between groups of data. [2] For RMSE, the best possible score is 0. Additionally, a linear model summarized the relationship between the first and second groups of data. [3] For this measure, an ideal slope is 1 and the best possible intercept would be 0.

As shown in Table 4.1 there is a strong agreement between the first and second runs of data for individual subjects, as well as a strong agreement between subjects, suggesting the data collected was repeatable. The agreement between the first and second run of individual subjects was stronger than the agreement between subjects. Taken together, all of the scores suggest this is a useful set of data, considering the general noisiness of data from human subjects completing psychophysics experiments.

[1] Calculations completed using the MATLAB function `corr`, using the types of `Pearson` and `Spearman`

[2] RMSE measurement was found by finding the square root of the mean of the square of the difference between the two groups of data, using the MATLAB functions `sqrt`, `mean`, and element-wise-power.

[3] The slope and intercept assessment was formed using the MATLAB function `LinearModel.fit`

Table 4.1: Average goodness of fit for intra- and intersubject agreement. The first column to the left lists the repeatability measures of the data. The second and third columns show the mean and standard deviation of the measures comparing the thresholds from first set of trials for each subject to their second set of trials. The forth and fifth columns show the mean and standard deviation of the measures comparing the thresholds of one subject to the thresholds of another subject.

	Intrasubject		Intersubject	
	mean	±	mean	±
PCC	0.93	0.06	0.87	0.08
SCOCC	0.94	0.05	0.85	0.10
RMSE	0.31	0.13	0.58	0.23
Slope	0.94	0.08	0.90	0.11
Intercept	-0.13	0.17	-0.04	0.42

Data repeatability and reliability is examined more closely in the next two sections.

4.4 Intrasubject agreement

This section examines data reliability and repeatability in greater detail. This section presents data repeatability and reliability graphically, and quantifies the agreement between the first and second experiment of each subject.

The data was broken into several groups to ease the burden on subjects during data collection, [17]. These groupings focused on relationships between target detectability contrast thresholds and either target temporal frequencies or target spatial frequencies. The relationships between target spatial frequency and target detectability are further broken down into vertical and diagonal target spatial frequencies. The four groups of data are named *Temporal 1*, *Temporal 2*, *Spatial Vertical*, and *Spatial Diagonal*. The first and second subject for all four groupings was the same, but the third

subject for each grouping was a different expert from the Oklahoma State University CPIQ lab. Intra and intersubject agreement plots and performance are broken down into these groups. Combined contrast threshold estimates are also broken down into these groups for plotting.

This Sect. details intra-subject agreement. One way to assess data repeatability is to plot the subjects results from their second set of trials against their first, as seen in Fig. 4.3. Figure 4.3 provides a scatter plot of the first and second trial from each subject, broken down by data grouping.

Observe from Fig. 4.3 the mostly linear relationship between the first and second set of trials for each subject. The plots in Fig. 4.3 suggest that the data collection process produced threshold estimates that were repeatable over time. Although Fig. 4.3 does show that some error bars were large, and that not all data fell directly into a line, overall, the second set of trials for each subject provided a reasonable match to the estimates from their first set of trials.

Graphical representations, such as Fig. 4.3, provide an efficient means to quickly assess data repeatability between sets of trials for a single subject, however, the controversial findings of this paper merit closer scrutiny of repeatability. The following Tables quantify the relationship between the first and second set of trials for individual subjects. The letters next to the subject numbers in Tables 4.2 through 4.5 correspond to the labels for Fig. 4.3. The data in Table 4.2 is for the data group *Temporal 1* and corresponds to the first row of plots in Fig. 4.3, a-c. The data in Table 4.3 is for the data group *Temporal 2* and corresponds to the second row of plots in Fig. 4.3, d-f. The data in Table 4.4 is for the data group *Spatial Vertical* and corresponds to the third row of plots in Fig. 4.3, g-i. The data in Table 4.5 is for the data group *Spatial Diagonal* and corresponds to the fourth row of plots in Fig. 4.3, j-l.

Observe from Tables 4.2, 4.3, 4.4, and 4.5 that there is significant repeatability between the first and second set of trials for each subject. Although the correla-

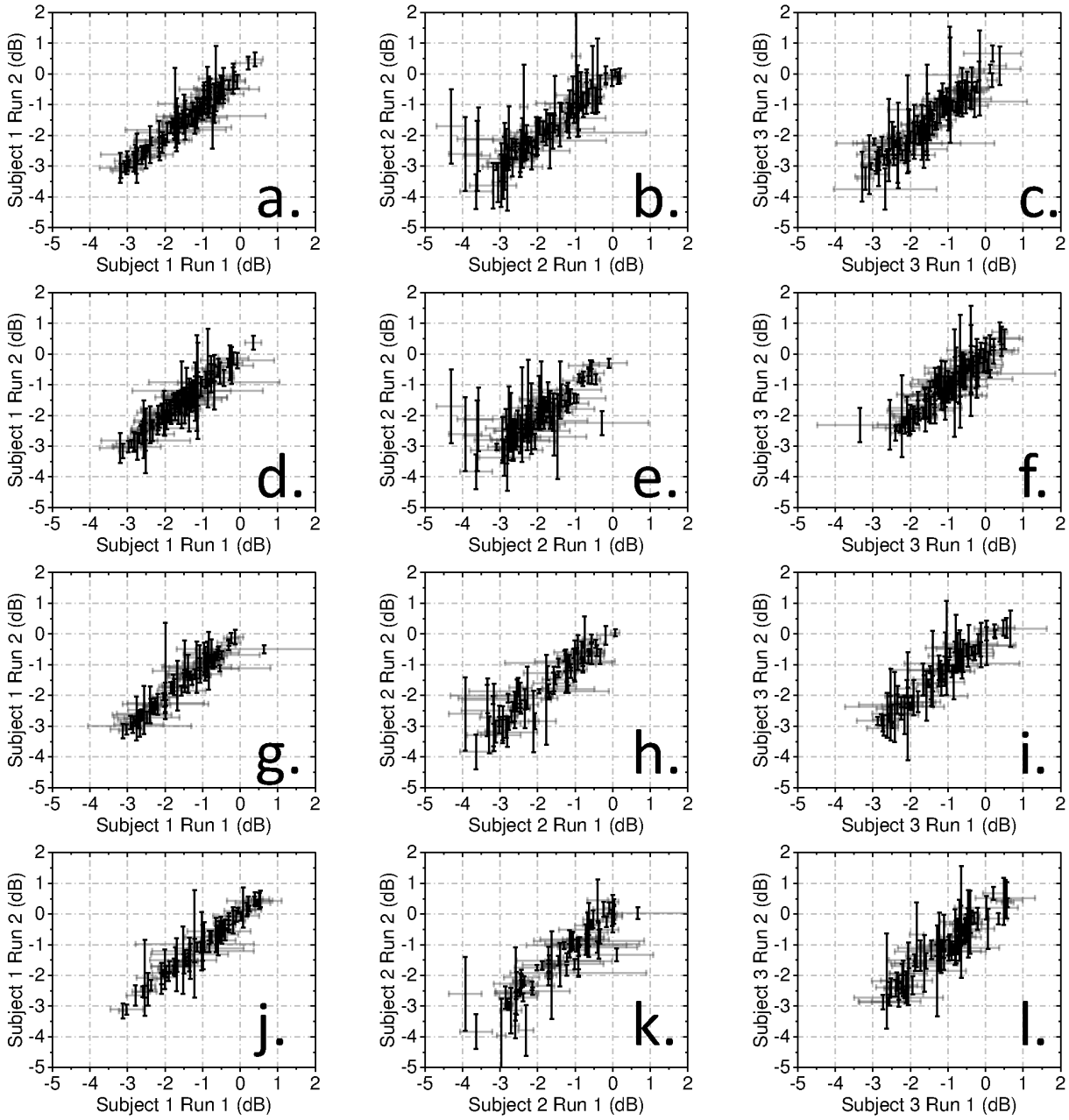

Figure 4.3: Intrasubject agreement. This figure shows how well the second trial of each subject agreed with their first trail. Plots a-c are for the data grouping *Temporal 1*. Plots d-f are for the data grouping *Temporal 2*. Plots g-i are for the data grouping *Spatial Vertical*. Plots j-l are for the data grouping *Spatial Diagonal*. Plots a, d, g, and j are for subject J.E. Plots b, e, h, and k are for subject K.J. Plot c was for M.A., f was for Y.Z., i was for P.V., and l was for T.P. Tables 4.2 through 4.5 quantify numerically how well the second set of trials from each subject agreed with their first set of trails.

Table 4.2: Intr-subject agreement for data set *Temporal 1.* The first row shows the PCC between sets of trials for subject 1 for the data group *Temporal 1.* The second row is the PCC for the second subject's trials, and the third row shows this information for the third subject's trials. The fourth and seventh rows show the SROCC and RMSE between trials for subject 1. The tenth row shows the slope of the line mapping the first set of trials from subject 1 to their second set of trials, and the eleventh row shows the intercept. The third column to the left, *Overall*, quantifies the repeatability between the first and second sets of trials for all masking conditions, while the fourth through seventh columns quantify repeatability for individual masking conditions. The letters next to the subject numbers in the second column from the left correspond to the plots in Fig. 4.3.

Intrasubject correlation		*Overall*	*Unmasked*	*Cactus*	*Kimono*	*Timelapse*
PCC	a. Subject 1	0.971	0.984	0.914	0.978	0.943
PCC	b. Subject 2	0.877	0.693	0.825	0.953	0.880
PCC	c. Subject 3	0.946	0.950	0.779	0.911	0.943
SROCC	a. Subject 1	0.960	0.987	0.924	0.977	0.844
SROCC	b. Subject 2	0.910	0.736	0.840	0.963	0.890
SROCC	c. Subject 3	0.950	0.946	0.755	0.844	0.905
RMSE	a. Subject 1	0.209	0.168	0.237	0.157	0.255
RMSE	b. Subject 2	0.482	0.742	0.359	0.257	0.426
RMSE	c. Subject 3	0.302	0.310	0.319	0.309	0.269
Slope	a. Subject 1	0.995	0.932	0.859	0.957	0.947
Intercept	a. Subject 1	-0.006	-0.155	0.016	-0.148	-0.097
Slope	b. Subject 2	0.860	0.501	0.909	1.109	0.960
Intercept	b. Subject 2	-0.230	-0.983	-0.074	0.135	-0.013
Slope	c. Subject 3	0.993	1.023	0.890	0.865	1.029
Intercept	c. Subject 3	-0.043	-0.091	-0.072	-0.212	0.022

Table 4.3: Intrasubject agreement for data set *Temporal 2*. Please see the caption of Table 4.2 for additional details.

Intra Subject Correlation		*Overall*	*Unmasked*	*Flowervase*	*Hands*	*Typing*
PCC	d. Subject 1	0.978	0.984	0.977	0.960	0.986
PCC	e. Subject 2	0.781	0.693	0.806	0.884	0.859
PCC	f. Subject 3	0.946	0.980	0.914	0.930	0.930
SROCC	d. Subject 1	0.970	0.965	0.966	0.962	0.969
SROCC	e. Subject 2	0.798	0.736	0.825	0.866	0.720
SROCC	f. Subject 3	0.955	0.938	0.939	0.960	0.868
RMSE	d. Subject 1	0.153	0.167	0.157	0.176	0.102
RMSE	e. Subject 2	0.503	0.742	0.448	0.371	0.351
RMSE	f. Subject 3	0.257	0.204	0.265	0.317	0.228
Slope	d. Subject 1	0.998	1.032	1.010	0.959	0.968
Intercept	d. Subject 1	-0.014	0.059	-0.020	-0.098	-0.033
Slope	e. Subject 2	0.713	0.501	0.819	0.997	0.855
Intercept	e. Subject 2	-0.617	-0.983	-0.460	-0.079	-0.361
Slope	f. Subject 3	0.942	0.968	0.838	0.917	1.079
Intercept	f. Subject 3	-0.106	-0.071	-0.194	-0.146	0.023

Table 4.4: Intrasubject agreement for data set *Spatial Vertical*. Please see the caption of Table 4.2 for additional details.

Intra Subject Correlation		*Overall*	*Unmasked*	*Cactus*	*Typing*	*Waterfall*
PCC	g. Subject 1	0.964	0.984	0.854	0.954	0.970
PCC	h. Subject 2	0.916	0.924	0.901	0.933	0.832
PCC	i. Subject 3	0.958	0.974	0.934	0.964	0.900
SROCC	g. Subject 1	0.967	0.948	0.920	0.963	0.944
SROCC	h. Subject 2	0.917	0.891	0.876	0.862	0.761
SROCC	i. Subject 3	0.952	0.913	0.926	0.913	0.926
RMSE	g. Subject 1	0.231	0.165	0.300	0.239	0.197
RMSE	h. Subject 2	0.418	0.405	0.280	0.334	0.587
RMSE	i. Subject 3	0.288	0.250	0.324	0.256	0.316
Slope	g. Subject 1	0.956	1.058	0.683	1.008	0.909
Intercept	g. Subject 1	-0.104	0.108	-0.302	0.066	-0.269
Slope	h. Subject 2	0.896	0.829	0.923	0.896	0.821
Intercept	h. Subject 2	-0.073	-0.348	0.026	-0.235	-0.009
Slope	i. Subject 3	0.961	1.085	0.870	0.925	1.176
Intercept	i. Subject 3	-0.151	0.126	-0.273	-0.242	0.231

Table 4.5: Intrasubject agreement for data set *Spatial Diagonal*. Please see the caption of Table 4.2 for additional details.

Intra Subject Correlation		*Overall*	*Unmasked*	*Cactus*	*Lemur*	*Timelapse*
PCC	j. Subject 1	0.990	0.989	0.981	0.947	0.993
PCC	k. Subject 2	0.909	0.933	0.879	0.773	0.849
PCC	l. Subject 3	0.951	0.969	0.964	0.947	0.915
SROCC	j. Subject 1	0.989	0.988	0.979	0.955	0.983
SROCC	k. Subject 2	0.928	0.940	0.847	0.868	0.886
SROCC	l. Subject 3	0.955	0.959	0.917	0.907	0.874
RMSE	j. Subject 1	0.132	0.166	0.119	0.132	0.105
RMSE	k. Subject 2	0.456	0.485	0.313	0.472	0.526
RMSE	l. Subject 3	0.284	0.324	0.168	0.226	0.372
Slope	j. Subject 1	0.976	0.974	0.934	0.899	0.980
Intercept	j. Subject 1	-0.018	-0.017	-0.004	-0.116	-0.051
Slope	k. Subject 2	0.961	0.923	1.062	0.775	0.798
Intercept	k. Subject 2	-0.159	-0.249	0.126	-0.557	-0.539
Slope	l. Subject 3	0.985	1.095	0.961	0.832	0.954
Intercept	l. Subject 3	0.012	0.002	-0.001	-0.118	0.035

tion coefficients for individual masking conditions for some subjects are lower than others, the average over all subjects and data groups was higher than 0.9, and the lowest correlation coefficient was 0.693 from Subject K.J. on the first data group they completed.

Taken together, the data in Fig. 4.3 and Tables 4.2, 4.3, 4.4, and 4.5 show strong agreement between the first and second trials of the subjects. Broken down by masking conditions, in general, the best agreement was for unmasked target detectability, however the difference is not large in comparison to any of the masked conditions. The best intrasubject agreement came from the first author, subject J.E., who had the most experience with the experiment. Some of the outliers from subject K.J. seen in Fig. 4.3 can also be noted in Tables 4.2, 4.3, 4.4, and 4.5. Subject K.J. was not an experienced target detection threshold subject before data collection for this paper. Additionally, although the scores for subject K.J. are not as close to perfect as the other two subjects, the performance still suggest the data is useful.

4.5 Intersubject agreement

This Sect. presents the agreement between subjects. The analysis from Sect. 4.4 has been repeated to examine the agreement of the estimates between subjects. Because each subject produced two sets of trials for each target detectability threshold estimate, the two estimates were combined into a single mean according to Eq. 3.1 for this analysis, and the standard deviation for each target detectability threshold was calculated according to Eq. 3.2. Figure 4.4 is the scatter plot of target detectability thresholds from one subject versus target detectability thresholds from another subject.

Observe from Fig. 4.4 that there is strong agreement between subjects. Because the weighted mean of each target detectability threshold is now calculated from two pieces of data, there is more confidence in the measurement, and appropriately, the

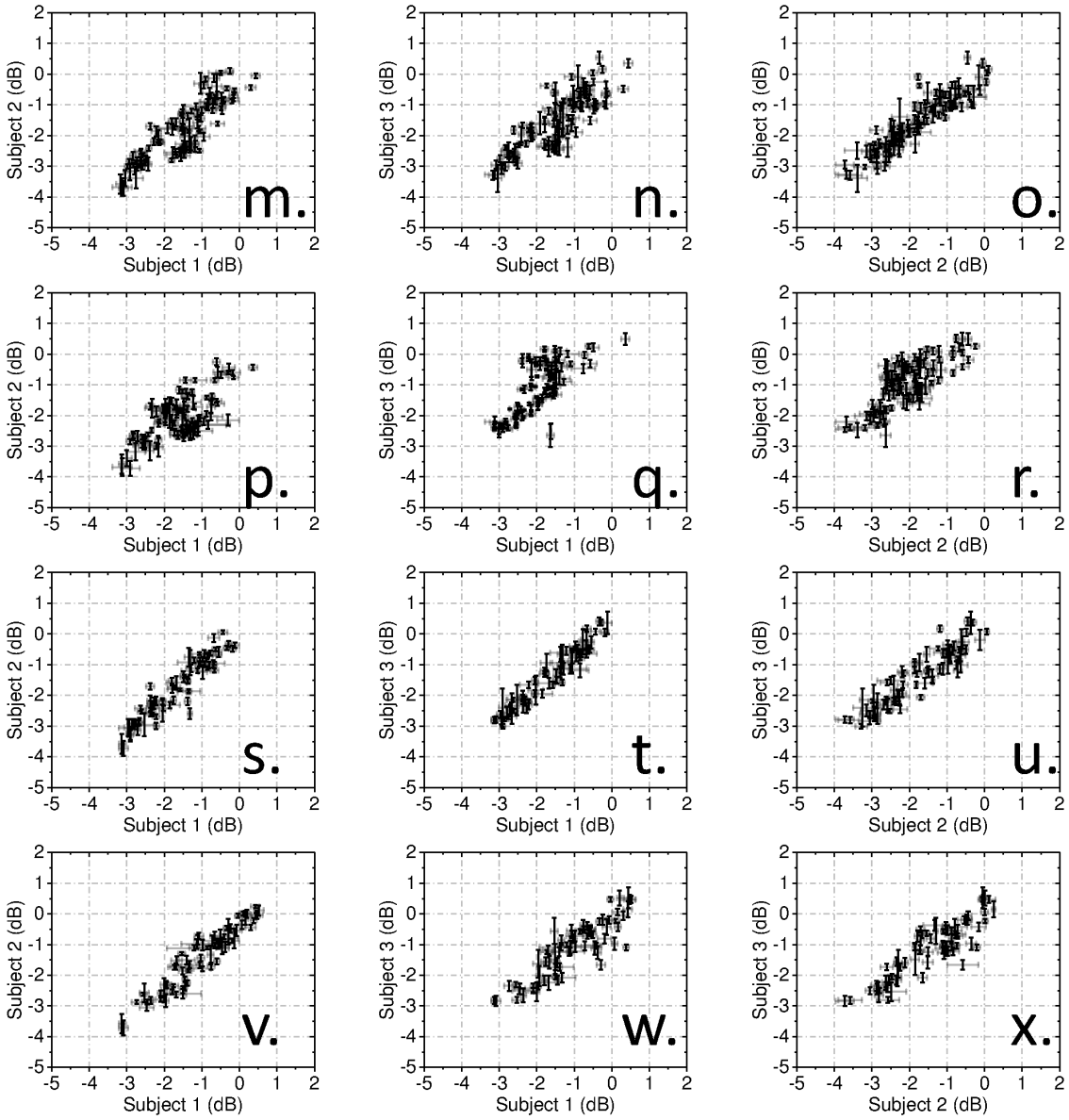

Figure 4.4: Intersubject agreement. This figure shows how well the subjects agreed with the other subjects. Each plot shows one subject's data plotted against the horizontal axis, with another subject's data plotted against the vertical axis. All data in a line of $y = x$ would represent perfect intersubject agreement. Plots m-o are for the data grouping *Temporal 1*. Plots p-r are for the data grouping *Temporal 2*. Plots s-u are for the data grouping *Spatial Vertical*. Plots v-x are for the data grouping *Spatial Diagonal*. Plots m, p, s, and v were for subject J.E vs subject K.J. Plots n, q, t, and w were for subject J.E. vs M.A, Y.Z, P.V, and T.P. Plots o, r, u, and x were for subject K.J. vs M.A., Y.Z., P.V., and T.P. The letters next to the subject numbers in Tables 4.6 through 4.9 correspond to the labels for this figure.

weighted standard deviations based on two measurements are also smaller. Also, the data shown in Fig. 4.4 appears to fall along a mostly straight line, suggesting the different subjects are in agreement with each other.

As with intra subject agreement, inter subject agreement is also assessed numerically. The following Tables quantify the relationship between individual subjects. The letters next to the subject numbers in Tables 4.6 through 4.9 correspond to the labels for Fig. 4.3. The data in Table 4.6 is for the data group *Temporal 1* and corresponds to the first row of plots in Fig. 4.4, m-o. The data in Table 4.7 is for the data group *Temporal 2* and corresponds to the second row of plots in Fig. 4.4, p-r. The data in Table 4.8 is for the data group *Spatial Vertical* and corresponds to the third row of plots in Fig. 4.4, s-u. The data in Table 4.9 is for the data group *Spatial Diagonal* and corresponds to the fourth row of plots in Fig. 4.4, v-x.

Observe from Tables 4.6, 4.7, 4.8, and 4.9 that most of the correlation coefficients are reasonably close to one when comparing entire data groups. The correlation coefficients are lower when examining agreement between subjects for individual masking conditions. The average of the mask specific correlation coefficients was about 0.8, but the lowest coefficient was a SROCC of 0.377 between subjects 1 and 3 in the data group *Temporal 1* for the mask *Cactus*. The average PCC for the mask *Cactus* for intra subject agreement was 0.89, but for inter subject agreement it was reduced to 0.73. These averages are below the averages for all masks, suggesting that there was less agreement between sets of trials for individual subjects, as well as less agreement between subjects when targets were presented with the mask *Cactus*.

Taken together, the data shown in Fig. 4.4 and presented in Tables 4.6, 4.7, 4.8, and 4.9 suggest a reasonable agreement between subjects. Although there was stronger agreement between estimates from sets of trials from individual subjects, the data presented in this subsection still suggests reasonable repeatability across subjects. The averages of the measures of repeatability presented in Tables 4.6, 4.7,

Table 4.6: Intersubject agreement for data set *Temporal 1.* See caption for Table 4.2 for additional information. For ease of reference and comparison, the letters next to the subject numbers correspond to the plots in Fig. 4.4

Inter Subject Correlation		*Overall*	*Unmasked*	*Cactus*	*Kimono*	*Timelapse*
PCC	m. Subject 1 & 2	0.871	0.818	0.631	0.793	0.871
PCC	n. Subject 1 & 3	0.843	0.805	0.481	0.775	0.838
PCC	o. Subject 2 & 3	0.916	0.930	0.692	0.863	0.919
SROCC	m. Subject 1 & 2	0.875	0.755	0.568	0.766	0.891
SROCC	n. Subject 1 & 3	0.834	0.733	0.377	0.762	0.813
SROCC	o. Subject 2 & 3	0.918	0.949	0.672	0.900	0.904
RMSE	m. Subject 1 & 2	0.552	0.637	0.472	0.586	0.495
RMSE	n. Subject 1 & 3	0.499	0.560	0.486	0.453	0.489
RMSE	o. Subject 2 & 3	0.441	0.396	0.450	0.563	0.321
Slope	m. Subject 1 & 2	0.959	0.806	0.733	0.779	0.957
Intercept	m. Subject 1 & 2	-0.355	-0.721	-0.300	-0.826	-0.361
Slope	n. Subject 1 & 3	0.879	0.790	0.456	0.804	0.857
Intercept	n. Subject 1 & 3	-0.253	-0.523	-0.314	-0.338	-0.470
Slope	o. Subject 2 & 3	0.867	0.926	0.563	0.911	0.855
Intercept	o. Subject 2 & 3	-0.016	0.063	-0.174	0.236	-0.222

Table 4.7: Inter-subject agreement for data set *Temporal 2*. See caption for Table 4.2 for additional information. For ease of reference and comparison, the letters next to the subject numbers correspond to the plots in Fig. 4.4

Inter Subject Correlation		*Overall*	*Unmasked*	*Flowervase*	*Hands*	*Typing*
PCC	p. Subject 1 & 2	0.723	0.782	0.787	0.678	0.570
PCC	q. Subject 1 & 3	0.773	0.811	0.898	0.759	0.869
PCC	r. Subject 2 & 3	0.779	0.898	0.794	0.746	0.704
SROCC	p. Subject 1 & 2	0.634	0.696	0.667	0.573	0.523
SROCC	q. Subject 1 & 3	0.703	0.647	0.846	0.747	0.798
SROCC	r. Subject 2 & 3	0.746	0.948	0.703	0.757	0.673
RMSE	p. Subject 1 & 2	0.717	0.708	0.673	0.656	0.819
RMSE	q. Subject 1 & 3	0.800	1.121	0.492	0.871	0.553
RMSE	r. Subject 2 & 3	1.202	1.432	0.974	1.211	1.147
Slope	p. Subject 1 & 2	0.728	0.764	0.726	0.790	0.607
Intercept	p. Subject 1 & 2	-0.908	-0.837	-0.893	-0.732	-1.188
Slope	q. Subject 1 & 3	0.810	0.904	0.756	1.015	0.793
Intercept	q. Subject 1 & 3	0.320	0.781	0.037	0.720	0.148
Slope	r. Subject 2 & 3	0.811	1.025	0.725	0.855	0.603
Intercept	r. Subject 2 & 3	0.702	1.419	0.359	0.787	0.207

Table 4.8: Inter-subject agreement for data set *Spatial Vertical*. See caption for Table 4.2 for additional information. For ease of reference and comparison, the letters next to the subject numbers correspond to the plots in Fig. 4.4

Inter Subject Correlation		*Overall*	*Unmasked*	*Cactus*	*Typing*	*Waterfall*
PCC	s. Subject 1 & 2	0.939	0.956	0.806	0.924	0.919
PCC	t. Subject 1 & 3	0.963	0.974	0.960	0.961	0.916
PCC	u. Subject 2 & 3	0.923	0.942	0.734	0.952	0.860
SROCC	s. Subject 1 & 2	0.931	0.961	0.827	0.870	0.858
SROCC	t. Subject 1 & 3	0.958	0.920	0.944	0.938	0.920
SROCC	u. Subject 2 & 3	0.923	0.928	0.763	0.833	0.870
RMSE	s. Subject 1 & 2	0.354	0.313	0.333	0.441	0.314
RMSE	t. Subject 1 & 3	0.371	0.443	0.411	0.274	0.330
RMSE	u. Subject 2 & 3	0.530	0.541	0.584	0.512	0.477
Slope	s. Subject 1 & 2	1.082	1.122	0.989	1.084	1.162
Intercept	s. Subject 1 & 2	0.026	0.216	-0.057	-0.124	0.215
Slope	t. Subject 1 & 3	1.056	1.126	1.353	1.024	0.909
Intercept	t. Subject 1 & 3	0.363	0.646	0.641	0.202	0.073
Slope	u. Subject 2 & 3	0.879	0.927	0.843	0.865	0.675
Intercept	u. Subject 2 & 3	0.169	0.258	0.237	0.154	-0.282

Table 4.9: Inter-subject agreement for data set *Spatial Diagonal*. See caption for Table 4.2 for additional information. For ease of reference and comparison, the letters next to the subject numbers correspond to the plots in Fig. 4.4

Inter Subject Correlation		*Overall*	*Unmasked*	*Cactus*	*Lemur*	*Timelapse*
PCC	v. Subject 1 & 2	0.944	0.961	0.903	0.727	0.942
PCC	w. Subject 1 & 3	0.883	0.956	0.656	0.804	0.807
PCC	x. Subject 2 & 3	0.894	0.974	0.719	0.674	0.835
SROCC	v. Subject 1 & 2	0.938	0.957	0.874	0.765	0.928
SROCC	w. Subject 1 & 3	0.866	0.967	0.649	0.719	0.738
SROCC	x. Subject 2 & 3	0.859	0.979	0.723	0.668	0.794
RMSE	v. Subject 1 & 2	0.506	0.584	0.374	0.539	0.505
RMSE	w. Subject 1 & 3	0.450	0.348	0.543	0.323	0.537
RMSE	x. Subject 2 & 3	0.529	0.470	0.447	0.610	0.571
Slope	v. Subject 1 & 2	1.012	1.060	0.841	0.986	0.869
Intercept	v. Subject 1 & 2	-0.377	-0.392	-0.318	-0.416	-0.595
Slope	w. Subject 1 & 3	0.875	0.942	0.703	1.104	0.823
Intercept	w. Subject 1 & 3	-0.226	-0.173	-0.286	0.176	-0.341
Slope	x. Subject 2 & 3	0.826	0.870	0.829	0.681	0.922
Intercept	x. Subject 2 & 3	0.045	0.144	-0.024	-0.131	0.179

4.8, and 4.9 are within 0.2 of ideal, with the exception of RMSE, which had an average of 0.57.

Subjects were asked to repeat sets of trials when the first two sets did not produce target detectability thresholds that were within half a log unit of each other. It is possible that the inter subject agreement scores could have been improved if subjects were also asked to repeat sets of trials when multiple subject target detectability thresholds did not match as well as desired. Even without this additional data collection step, the data in these sections still suggests that useful conclusions can be drawn from this data set.

Professor Le Callet of the IRCCyN lab with Polytech'Nantes of the University de Nantes, attended a discussion of the preliminary data collected in the development of these experiments [146]. Professor Le Callet had expressed a concern over the large error bars in the initial data, and suggested one possible way of reducing the size of the error bars would be to re-conduct the experiments employing something in line with the Cambridge Research System's Bits# stimulus processor. Because of the ability to control CRT displays down to 14 bits of brightness instead of only 8, the ability to make smaller changes in brightness of the display is clearly superior. We feel this is necessary for some later experiments to fine tune models, and vital to furthering the understanding of specific details of human vision. However, after further analysis with the current experiment setup, it appears that the range in which masked data was collected is not severely impacted by the levels of quantization for the display. Indeed, it appears that the largest error bars are only from one subject that was inexperienced at the data collection process. These large standard deviations can be decreased through running more trials per experiment, obtaining more than two experiments per subject, using more subjects, or being more selective about outlier removal. Taken together, the data in Fig. 4.4 and Tables 4.6, 4.7, 4.8, and 4.9 suggest that there was good agreement from one subject to the next. Examining the

performance measures in Tables 4.6, 4.7, 4.8, and 4.9, the strongest agreement appears to be for the data set *Spatial Vertical*, while the weakest agreement appears to be for the data set *Temporal 2*. This can be also be seen in Fig. 4.4, were the best looking data set, *Spatial Vertical*, is the third row of plots and the worst looking data set, *Temporal 2*, is the second row of plots. One possible explanation for the differences in performance may be due to the orthogonality of the thresholds in those datasets, due to the masks and targets used. Figures 4.1 and 4.2 make this point visibly. As will be discussed further in Chapter 5, changing target flicker rate from 0 Hz up to 6 Hz or even 10 Hz does not cause much of a difference in target detectability, and would not cause markedly different thresholds. The noisy data in a flat line is likely to have worse rank order correlation than data in a concave up, and more monotonically increasing line. So it is expected that the data sets *Temporal 1* and *Temporal 2* would have worse correlation than the spatial data sets, *Spatial Diagonal* and *Spatial Vertical*. Also, the mask *Cactus* is significantly different from the masks *Kimono* and *Timelapse* in masking ability, providing significant separation in target detectability thresholds. However, the masks *Flowervase*, *Hands*, and *Typing* appear to be similar in performance to the unmasked condition, resulting in these data being jumbled together. This is a possible explanation why the correlation for the data grouping *Temporal 1* was better than the grouping *Temporal 2*. Looking at the two spatial sets, the change in target spatial frequency across the data set *Spatial Diagonal* is larger than it is for the data set *Spatial Vertical*. It is possible that the larger difference in the target spatial frequency improved the separability of the data, and thus improved the correlation of the results. The differences between various masks and targets might account for the difference in overall repeatability between the four data sets.

CHAPTER 5

ANALYSIS AND DISCUSSION

This chapter presents an analysis and discussion of our results. The chapter begins with a validation of previous findings on unmasked target detectability. The unmasked data are extended by examining natural video masked target detectability. A simple linear model provides a summary of those data, as well as further analysis.

The differences between target detectability are *elevations*. Positive elevations signify the later target had a higher detectability contrast threshold. Elevations were calculated according to

$$\text{Threshold Elevation} = \overline{x}_2 - \overline{x}_1 \pm \quad \left(\frac{\sigma_2^2}{n_2 - 1} + \frac{\sigma_1^2}{n_1 - 1} \right), \tag{5.1}$$

where, n_1 and n_2 are the number of sets of trials that were used to compute the means $\overline{x}_1$ and $\overline{x}_2$. Although in some cases, more than three subjects completed two sets of trials, the value three was used for n_1 and n_2 to provide a more conservative estimate of the target detectability contrast threshold elevation standard deviations for the data presented in this chapter. When averages of target detectability contrast threshold elevations are reported, the standard deviations reported are the standard deviations of the elevations that were used to calculate those averages.

We developed a linear regression based model to predict dynamic DCT noise detectability in log units, VCT. [1] The model development loosely followed a template suggested by the excellent work by Watson, Hu, and McGowan [6]. Watson, Hu, and McGowan [6] provided an elegant separable model for unmasked target detectability that was dependent only on target temporal and spatial frequency.

[1] The coefficients for this model were found using the MATLAB tool `LinearModel.fit`.

k-fold-cross-validation, with a k of 10, was employed to find model coefficients and avoid over fitting of our limited data set. For the selection of model coefficients, our dataset was randomly divided into k equal parts. Model coefficients were found using linear regression to provide the smallest total error between the model prediction and measurements for $k-1$ parts, also known as the training set. The coefficients from training were used to predict the remaining part of the data, which is also known as the validation set. This process was repeated k times, each time withholding a separate part of the data. Multiple models were compared by their average goodness of fit scores over all k validation sets. For reporting of results, the average of the k sets of coefficients with the best performance was then used to predict the entire data set. The goodness of fit between model predictions and measured data was calculated after a non-linear transformation of the model prediction.

To emphasize the relationships between model inputs, and for easy comparison of their significance to model performance, all inputs were normalized to range from one to two. Positive model coefficients suggest that larger input values are correlated with higher target detectability. A model coefficient larger in magnitude suggests that the associated model input has more influence on target detectability.

The linear regression modeling process has many limitations, and is not the next best model of human vision for estimating all compression artifact detectability. Our data set does not sufficiently span a large range of all videos or enough types of compression artifacts to make a truly general model of masked target detectability. The form of our model was chosen because of its simplicity, which eased modification and interpretation of results. Additionally, the signs of the coefficients of the linear model describe the relationship between model inputs and target detectability. The functional models presented do not have biological plausibility, but rather, are best fits of measured target detectability, and provide a useful extension of the work by Watson, Hu, and McGowan [6].

5.1 Unmasked target detectability

This section provides our findings on unmasked target detectability. Fig. 4.1 and 4.2 show our unmasked target detectability thresholds are in line with data and suggestions from previous research. Linear regression models with normalized inputs quantify the significance of target spatial and temporal frequencies in predicting target detectability.

Our results are in line with previous research. The work by Watson, Hu, and McGowan [6] was with unmasked dynamic DCT noise, and was in line with previous research on unmasked targets, such as that by Robson [8]. The general expectation from previous research is that either higher target spatial or temporal frequencies can increase target detectability thresholds. Fig. 4.1 plots unmasked target detectability thresholds over increasing target spatial frequencies at three target temporal frequencies. Fig. 4.2 plots unmasked target detectability thresholds over increasing target temporal frequencies at three target spatial frequencies. The unmasked data in these two figures suggest our data is a confirmation of the findings of previous researchers.

Large changes in target spatial and temporal frequencies result in large changes in target detectability. Fig. 5.1 shows target elevations due to changing target spatial frequencies from 2.8 c/deg to 22.6 c/deg. When the target temporal frequency is 0 Hz, and the target is unmasked, the elevation due to changing the target from DCT [0,0] to [7,7] is 2.65 ± 0.04 log units. Fig. 5.2 shows target elevations due to changing target temporal frequencies from 0 Hz to 30 Hz. When the target is DCT [0,0], and the target is unmasked, the elevation due to changing the target temporal frequency from 0 Hz to 30 Hz is 1.87 ± 0.02 log units. These data suggest that higher target spatial frequencies are correlated with significantly higher unmasked target detectability thresholds, and that higher target temporal frequencies are also correlated with significantly higher unmasked target detectability contrast thresholds.

A simple linear model can predict most of the variation in unmasked target de-

tectability due to increases in either target spatial or temporal frequency. Watson, Hu, and McGowan [6] suggested a linearly separable model that was a function of only target spatial and temporal frequencies. However, this model had nine parameters, and the model coefficients did not clearly quantify the importance of each model input for model performance. Table 5.1 shows several linear models that can predict most or nearly all variation in unmasked target detectability.

The linear model inputs included target spatial frequency, target temporal frequency, and a third input which was the product of the first two. The data from Robson [8] suggested that at sufficiently high target temporal frequencies, target spatial frequency did not influence target detectability as much, and that the inverse of this statement was also true. This may suggest the inclusion of a third term to account for the interactions of high target frequencies, as observed by Robson. [1] Based on this observation, a third input was considered, which was target spatial frequency times target temporal frequency, $(TSF \times TTF)$.

As shown in Table 5.1 (a) and (b), the two coefficients for TSF and TTF are positive, suggesting that increasing either target spatial or temporal frequencies would also increase unmasked target detectability. Note in Table 5.1 that the coefficients for TSF were larger than the coefficients for TTF. This may suggest that TSF is a more significant target property in determining target detectability for this data set. However, the magnitude of the TTF coefficients suggest that target temporal frequencies still have a significant contribution to unmasked target detectability.

In column (e) of Table 5.1, the model coefficient for $(TSF \times TTF)$ is smaller in magnitude and negative in sign. When the target spatial or temporal frequencies are small, the contribution of this term is small. However, when both target spatial and temporal frequencies are large, this third input will have a more significant input. At

[1] Our data also confirms this observation. This interaction is examined more closely in Fig. 5.1 and 5.2.

Table 5.1: Linear models of unmasked target detectability. The top half of the table provides goodness of fit scores, and the bottom half shows the model coefficients for normalized inputs. Column (a) shows the fit scores and coefficients for a linear model with only target spatial frequency, (TSF), as an input. Column (b) is for a linear model with only target temporal frequency, (TTF) as an input. Column (c) is for a model with a combined input that is the product of target spatial frequency and target temporal frequency, ($TSF \times TTF$). Column (d) shows the fit scores and coefficients for a linear model with two inputs: TSF, and TTF. Column (e) is for a linear model with three inputs: TSF, TTF, and ($TSF \times TTF$).

	(a)	(b)	(c)	(d)	(e)
PCC	0.702	0.642	0.792	0.961	0.964
SROCC	0.689	0.583	0.721	0.958	0.961
RMSE	0.688	0.741	0.590	0.266	0.258
Constant	-4.955	-3.949	-5.568	-7.169	-7.012
TSF	2.358			2.348	2.762
TTF		1.670		1.656	2.286
($TSF \times TTF$)			3.337		-1.359

these higher target spatial and temporal frequencies, the small negative coefficient may use this input as a regulation term. This might suggest that, as suggested by Robson [8], when the targets have sufficiently high spatial frequencies, increasing target temporal frequency does not have as much influence on target detectability, and vise versa.

Observe in Table 5.1 that the worst model performance came from the models that had only one input. This suggests that a single input model is not sufficient to predict variations in target detectability. As the unmasked target detectability were gathered for multiple target spatial and temporal frequencies, it would seem reasonable that models to predict those thresholds would also need to be functions of both target spatial and temporal frequencies. When target spatial and temporal frequencies are combined as a product and used as a single input, the model prediction provided a better fit of the data than either the target spatial or temporal frequencies alone.

As shown in Table 5.1, both the two and three input models provide reasonable predictions of unmasked target detectability. It should be noted that, due to the variations from trial to trial and subject to subject, it is difficult for a model to capture all the randomness of human subjects. That being said, it appears that either the two or three input model reasonably fit the unmasked target detectability contrast threshold data. [2]

Also observe in Table 5.1 that $(TSF \times TTF)$ was important as a single input, but did not significantly improve the goodness of fit as the third model input. One measure of the importance of an input in the overall fit of model prediction to measured data is the pValue of model inputs, where the smaller a pValue is, the more significant the input to the fit of the model. For the two input model in Table 5.1

[2]The model provided by Watson, Hu, and McGowan [6] could have been tuned to provide a better fit of the unmasked data than this simple model. However, the focus of our modeling effort is to provide further analysis of the masked target detectability.

(d), the largest pValue was for the TTF coefficient, which was $8E-14$, suggesting the least significant term in the two input model still had considerable influence on the goodness of fit to the data. For the three input model in Table 5.1 (e), the largest pValue was for the $(TSF \times TTF)$ coefficient, which was $2E-02$, suggesting the least significant term in the three input model had marginal influence on the goodness of fit to the data. [3]

In summary, our unmasked data appear to be in line with results from previous researchers. Higher target spatial frequencies are associated with higher target detectability, and higher target temporal frequencies are also associated with higher target detectability. Simple models of target spatial and temporal frequencies can predict unmasked target detectability. Target spatial frequency appears to be more significant in predicting unmasked target detectability for this data set.

5.2 Masked target detectability

This section details the differences in target detectability due to presenting targets with natural video masks. Not all masks have the same effects on target detectability contrast thresholds. Simple models that do not consider mask content do not sufficiently explain all variations in masked target detectability contrast thresholds. Plots of masked and unmasked target detectability contrast thresholds over target spatial and temporal frequencies suggest there are noticeable and significant differences between masked and unmasked thresholds. Figure 4.1 plots target detectability thresholds over increasing target spatial frequencies at three target temporal frequencies. Fig. 4.2 plots target detectability thresholds over increasing target temporal frequencies at three target spatial frequencies. These figures show that presenting masks with targets changes target detectability contrast thresholds. Different masks

[3]Some suggest omitting model inputs with a pValue larger than 0.05, while other more conservative guidance suggests omitting model inputs with pValues larger than 0.01.

have result in different elevations in target detectability contrast thresholds.

Large changes in target spatial and temporal frequencies result in large changes in target detectability contrast thresholds, however masks reduce the effects of these changes. Fig. 5.1 shows target elevations due to changing target spatial frequencies from 2.8 c/deg to 22.6 c/deg. The effects of target spatial frequency on target detectability contrast thresholds are examined more closely in Sect. 5.3. When the target temporal frequency is 0 Hz, the average elevation due to changing the target from DCT [0,0] to [7,7] is 2.65 ± 0.04 log units for unmasked targets; however, when targets are presented with natural videos, the average elevation due to this large change in target spatial frequency is reduced to 0.93 ± 0.86 log units. Fig. 5.2 shows target elevations due to changing target temporal frequencies from 0 Hz to 30 Hz. When the target is DCT [0,0], the elevation due to changing the target temporal frequency from 0 Hz to 30 Hz is 1.87 ± 0.02 log units for unmasked targets; however, for targets presented with natural videos, the average elevation is reduced to 0.88 ± 0.31 log units. The effects of target temporal frequency on target detectability contrast thresholds are examined more closely in Sect. 5.5 .

Although the simple two and three input models were effective in explaining variations in unmasked target detectability contrast thresholds, they were not so effective in predicting masked target detectability contrast thresholds. Using the k-fold-cross-validation method to select from over twenty candidates each for two and three input models, coefficients were selected, and fit scores calculated. Table 5.2 shows model coefficients and fitness scores for two and three input models on masked and unmasked data.

Observe in Table 5.2 that neither the two or three input model provide a reasonable fit of the masked target detectability contrast thresholds. This appears to suggest that masked target detectability contrast thresholds are influenced by more than just changes in target spatial and temporal frequencies. However, some of the variation in

Table 5.2: Summary of model fit performance and model coefficients for the two and three input models on masked and unmasked data.

	Unmasked		Masked	
	2 input	3 input	2 input	3 input
PCC	0.961	0.964	0.684	0.690
SROCC	0.958	0.961	0.672	0.673
RMSE	0.266	0.258	0.554	0.550
constant	-7.169	-7.012	-4.460	-4.355
TSF coeff	2.348	2.762	1.182	1.387
TTF coeff	1.656	2.286	1.040	1.356
$(TSF \times TTF)$ coeff		-1.359		-0.703

masked target detectability contrast thresholds can be explained by changes in target spatial and temporal frequencies, suggesting the target still has a significant role in determining masked target detectability contrast thresholds.

Also shown in Table 5.2, for all models, the largest model coefficient in magnitude was for target spatial frequency. This may suggest that target spatial frequencies are more important in predicting target detectability contrast thresholds than target temporal frequencies. However, note that all coefficients are smaller for models of the masked data. This may suggest that target spatial and temporal frequencies matter less when predicting masked target detectability contrast thresholds.

5.3 Target spatial frequency and masked target detectability

This Sect. discusses the relationships between target spatial frequencies and target detectability contrast thresholds. In general, targets with higher spatial frequencies have higher detectability contrast thresholds, which was expected based on previous research. Also as expected, when targets are higher in temporal frequencies,

the target detectability contrast threshold elevations due to large increases in target spatial frequencies are reduced. This Sect. also shows that presenting natural video masks with targets can reduce or even reverse the effects of changing target spatial frequencies on target detectability contrast thresholds.

The data in Fig. 4.1 suggests the relationships between masked target detectability contrast thresholds and increasing target spatial frequencies do not always match the relationships between unmasked target detectability contrast thresholds and target spatial frequencies. In Fig. 4.1 (a) and (d), it can be seen that at low target temporal frequencies and low target spatial frequencies, the different masks caused markedly different target detectability contrast thresholds. From Fig. 4.1 (d), it can be seen that the difference in masked and unmasked target detectability contrast thresholds reduces some at higher target spatial frequencies. This is in agreement with previous research on natural image masking of compression artifacts [11]. This is most noticeable for the mask *Cactus*, in Fig. 4.1 (d), were at lower target spatial frequencies of 2.8 c/deg, there was about a two log unit difference in target detectability contrast thresholds, however, at higher target spatial frequencies of 22.6 c/deg, the unmasked thresholds are nearly the same as the thresholds for targets presented with the mask *Cactus*.

It was not expected that some natural video masks would reduce target detectability contrast thresholds. Observe in Fig. 4.1 (b), (c), (d), (e), and (f) that there are masked target detectability contrast thresholds markedly lower than unmasked target detectability contrast thresholds. This was not expected based on previous research, and is discussed further in Sect. 5.8 and 5.9.

Observe also in Fig. 4.1 that changing the target basis functions from DCT [0,0] to DCT [7,7] greatly increases unmasked target detectability contrast thresholds. The change in target basis functions from DCT [0,0] to DCT [7,7] does not cause the same increase in masked target detectability contrast thresholds. Additionally, when

target temporal frequencies are increased to 30 Hz, the change in target basis functions from DCT [0,0] to DCT [7,7] does not cause the same increase in target detectability contrast thresholds. The elevation due to changing the target basis functions from DCT [0,0] to DCT [7,7] is plotted in Fig. 5.1 against target temporal frequencies for both masked and unmasked targets.

It can be seen from Fig. 5.1 that changing the target basis functions from DCT [0,0] to [7,7] always has some effect on target detectability contrast thresholds. The solid black plots in Fig. 5.1 show these elevations are most significant when targets have low temporal frequencies and presented without masks. The other plots in Fig. 5.1 suggest that masked target detectability contrast threshold elevations due to changing target basis functions from DCT [0,0] to DCT [7,7] are less than unmasked target detectability contrast threshold elevations.

Observe also in Fig. 5.1 that when target temporal frequencies are high enough, target detectability contrast threshold elevations due to changing target basis functions from DCT [0,0] to DCT [7,7] are also reduced. Although the target detectability contrast threshold elevations for target temporal frequencies of 0 Hz and 6 Hz are similar, the contrast threshold elevations at 30 Hz in Fig. 5.1 are lower. This is in line with the findings of previous research for unmasked targets [8]. The target detectability contrast threshold elevation plots for masked targets in Fig. 5.1 appear to be reasonable extrapolations of the unmasked data. Further effects of target temporal frequencies on target detectability contrast thresholds are presented in Sect. 5.5 and 5.6.

Based on previous research and other plots of both masked and unmasked target detectability contrast threshold elevations due to changing the target basis functions from DCT [0,0] to DCT [7,7], it was not expected that this significant change in target spatial frequencies would ever cause negative elevations. However, observe in the lower left hand corner of Fig. 5.1 that when targets have temporal frequencies of 30 Hz, and

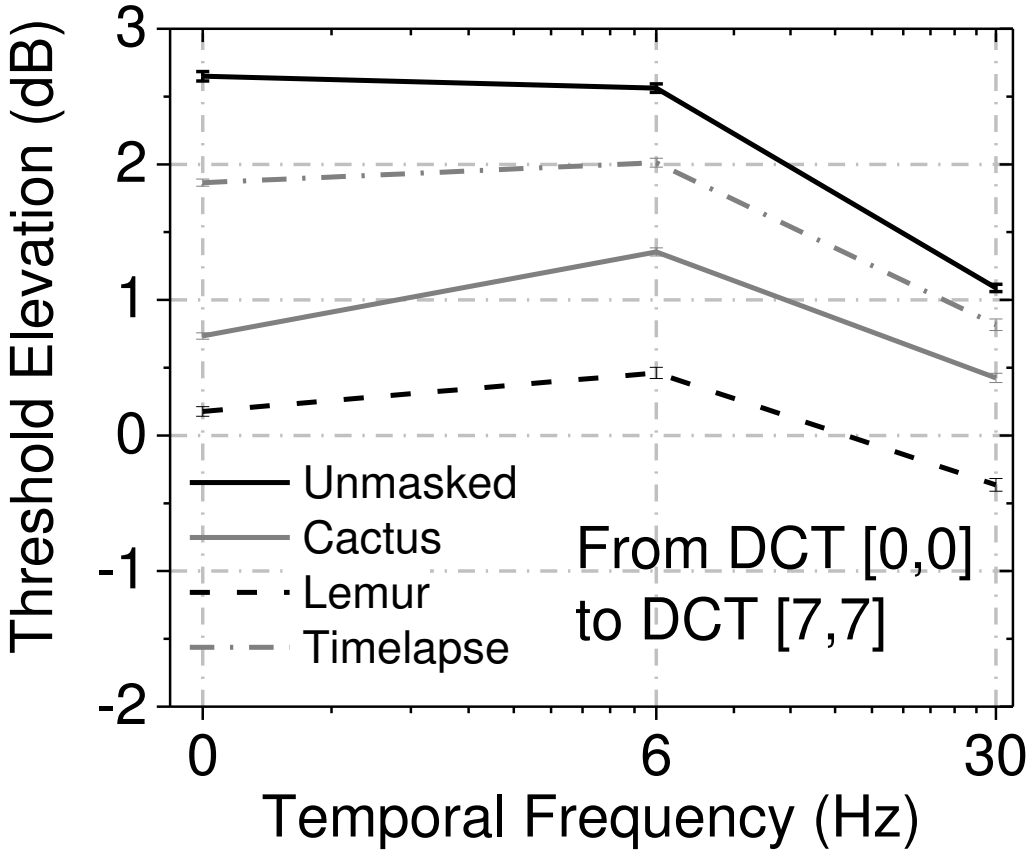

Figure 5.1: Target detectability contrast threshold elevations due to changing target basis functions from DCT [0,0] to [7,7], when masking conditions and target temporal frequencies remain constant. The vertical axis reports the target detectability contrast threshold elevations due to the change in target spatial frequencies, calculated according to Eq. 5.1. The horizontal axis shows temporal frequencies used for both the DCT [0,0] and DCT [7,7] targets. The graph legend in the lower left corner shows the masking conditions used for both the DCT [0,0] and DCT [7,7] targets for each plot line.

are presented with the mask *Lemur*, the higher spatial frequency targets have lower detectability contrast thresholds than targets with lower spatial frequencies. This can also be observed in the data presented in Fig. 4.1 (f). This was not expected based on previous research, and is discussed further in Section 5.8.

5.4 Discussion of target spatial frequencies and target detectability contrast thresholds

This Sect. discusses the relationships our data suggest between target spatial frequencies and target detectability contrast thresholds, for both the masked and unmasked targets. In Chap. 4, it was shown in Fig. 4.1 that target spatial frequencies have an effect on target detectability contrast thresholds. In general, targets with higher spatial frequencies have higher detectability contrast thresholds, and masked detectability contrast thresholds were reasonable extensions of unmasked target detectability contrast thresholds. This is in agreement with previous research on unmasked target detectability thresholds [20, 21, 22, 6], as well as masked target detectability threshold research [10].

Figure 5.1 in this Sect. detailed target detectability contrast threshold elevations due to large changes in target spatial frequencies, and presented the target detectability contrast threshold elevations due to changing DCT basis functions from [0,0] to [7,7], which changes the target spatial frequencies from 2.8 c/deg to 22.6 c/deg. When target temporal frequencies were 30 Hz, there was less of an effect on target detectability contrast thresholds due to large changes in target spatial frequencies. When the targets were presented with natural video masks, large changes in target spatial frequencies were less effective in changing target detectability contrast thresholds. Table 5.3 presents the target detectability contrast threshold elevations due to changes in target spatial frequencies from DCT [0,0] to DCT [7,7] for three target temporal frequencies for the unmasked condition, as well as averages for all masked conditions.

Table 5.3: Average target detectability contrast threshold elevations due to changes in target basis functions from DCT [0,0] to DCT [7,7]. Target detectability contrast threshold elevations are reported for the unmasked condition, while averages of the elevations are reported for masked conditions. The average was taken across all masked elevations available, and the standard deviation reported is of the elevations that were used in calculating that average.

	Unmasked		Masked average	
	elev.	±	elev.	±
0 Hz	2.65	0.04	0.93	0.86
6 Hz	2.56	0.03	1.28	0.78
30 Hz	1.09	0.03	0.29	0.60
Average	2.10	0.88	0.83	0.78

Observe in Table 5.3 that increasing target temporal frequencies to 30 Hz reduced the effectiveness of changing target spatial frequencies in influencing target detectability contrast thresholds. Robson [8] suggested that when the target detectability contrast thresholds became higher, due to higher target temporal frequencies, target spatial frequencies would matter less in determining target detectability thresholds, and that the inverse of this relationship would be true for sufficiently high target spatial frequencies. The unmasked column of Table 5.3 confirms this suggestion by Robson [8]. Table 5.3 shows this trend is continued for masked targets.

As shown in Table 5.3, presenting targets with masks can greatly reduce the change in target detectability contrast thresholds due to large changes in target spatial frequencies. This was expected based on previous research [10]. Also, Table 5.3 shows that when targets with high temporal frequencies are presented with masks, the change of target DCT basis functions from [0,0] to [7,7] makes less than half a log unit difference in target detectability contrast thresholds.

The data for this dissertation suggest that knowing the spatial frequency of DCT basis functions alone is not sufficient to predict target detectability contrast thresholds, and that mask content also needs consideration. Furthermore, the data for this dissertation supports the assumptions that there are strong connections between DCT basis functions and target detectability contrast thresholds. However, as suggested by previous research, both masking the targets and making the target temporal frequencies higher can make low spatial frequency DCT basis functions have just as high of detectability contrast thresholds as high spatial frequency DCT basis functions.

Table 5.5 showed that when the mask *Lemur* was presented with a target with a temporal frequency of 30 Hz, the DCT [7,7] target had a lower detectability contrast threshold than the DCT [0,0] target. Although this is only one point out of twelve with a negative target detectability contrast threshold elevation, it is still an interesting finding. Many compression algorithms are based on the assumption that the DCT [0,0] frequency content should be maintained with high fidelity, while the DCT [7,7] frequency content can be severely quantized. Our finding with respect to targets masked by *Lemur* with temporal frequencies of 30 Hz suggests that the assumption about what to compress more is not always accurate. Furthermore, most compression algorithms quantize much more than the highest frequency DCT basis function. Table 5.5 also showed that near this spatial frequency, threshold elevations caused by smaller target spatial frequency changes are more likely to be negative, thus not fitting previous assumptions.

In order to support the complicated field of video compression and the messy world it captures, additional research is required. Different combinations of target spatial frequencies should be measured in summation studies. Different masks should also be examined, possibly allowing for control of mask spatial content, or even chromatic studies. Future work is discussed in Chapter 8.

5.5 Target temporal frequency and masked target detectability

This Sect. details our findings on the relationships between target detectability contrast thresholds and target temporal frequencies. The plots in this Sect. show that, as expected, targets with higher temporal frequencies have higher detectability contrast thresholds. Also as expected, when targets have higher spatial frequencies, the elevation due to changing target temporal frequencies from 0 Hz to 30 Hz is reduced. Presenting a natural video mask with the targets can reduce the effects of a large change in target temporal frequency.

As shown in Fig. 4.2, as target temporal frequencies increase, unmasked target detectability contrast thresholds increase. When target spatial frequencies are increased from 2.8 c/deg to either 16.1 c/deg or 11.3 c/deg, target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 30 Hz are reduced. Previous research on unmasked target detectability and target spatial frequencies supports these observations [19, 8, 26, 27, 6].

Observe also in Fig. 4.2 that presenting natural video masks with targets results in large increases in target detectability contrast thresholds at low target temporal frequencies, however, these effects are reduced at higher target temporal frequencies. This is most noticeable for the mask *Cactus* in Fig. 4.2 (a). At 0 Hz, there is over a two log unit difference in target detectability contrast thresholds. However, at 30 Hz, unmasked target detectability thresholds are only about half a log unit less than the detectability contrast thresholds for targets presented with the mask *Cactus*. Presenting the targets with the other masks resulted in some difference in detectability contrast thresholds at low target temporal frequencies, but nearly no difference in detectability contrast thresholds for targets with temporal frequencies more than 4-6 Hz.

Figure 4.2 shows that, in general, the largest changes in target detectability contrast thresholds are due to the changes between the lowest and highest target temporal

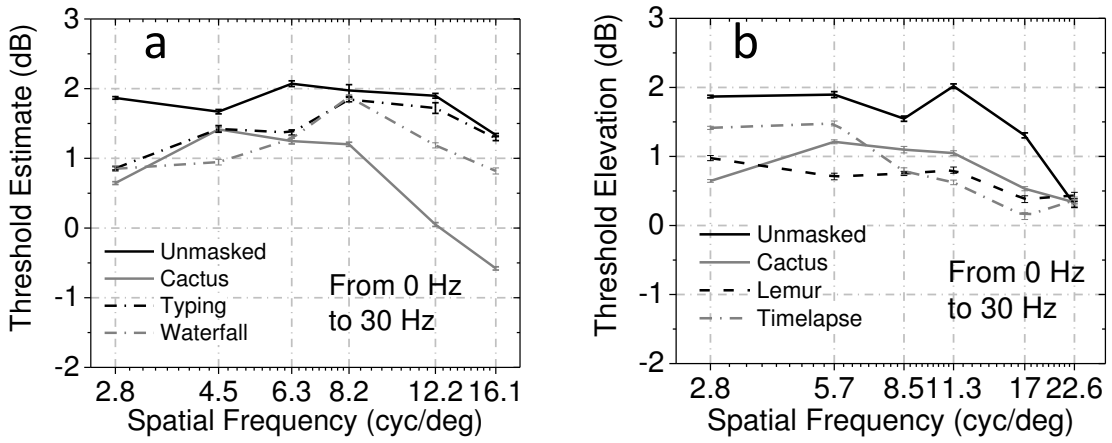

Figure 5.2: Target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 30 Hz. The vertical axis reports the target detectability contrast threshold elevations calculated according to Eq. 5.1. The horizontal axis shows the target spatial frequencies used for both the 0 Hz and 30 Hz targets. The graph legend in the upper left corner shows the masking conditions used for each line.

frequencies measured, 0 Hz and 30 Hz. However, it appears in Fig. 4.2 that either presenting targets with natural video masks, or increasing target spatial frequencies can reduce this contrast threshold elevation. Figure 5.2 shows the target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 30 Hz, while keeping masking conditions and target spatial frequencies constant. Figure 5.2 shows these target detectability contrast threshold elevations for masked and unmasked targets for the target DCT basis functions of [0,0], [1,1], [2,2], [3,3], [5,5], and [7,7]. The solid black line in Fig. 5.2 shows unmasked target detectability contrast elevations.

Observe from Fig. 5.2 that changing target temporal frequencies from 0 Hz to 30 Hz always has some effect on target detectability contrast thresholds. The solid black plot in Fig. 5.2 shows detectability contrast threshold elevations are largest when targets are presented without masks. The other plots in Fig. 5.2 show that when the

targets are presented with masks, the change in target temporal frequencies cause less of a change in target detectability contrast thresholds.

As shown in Fig. 5.2, when target spatial frequencies increase, the detectability contrast threshold elevations due to changes in target temporal frequencies are also reduced. Although the detectability contrast threshold elevations for targets using the DCT basis functions [0,0] through [3,3] are similar, the elevations for DCT [7,7] in Fig. 5.1 are lower. Previous research supports this observation for unmasked target detectability contrast thresholds [8]. The plots in Fig. 5.2 for masked target detectability contrast threshold elevations appear to be reasonable extrapolations of the unmasked data. Effects of target spatial frequencies on target detectability contrast thresholds were presented in Sect. 5.3 and are discussed further in Sect. 5.6.

5.6 Discussion of target temporal frequencies and target detectability contrast thresholds

This Sect. discusses our findings on the importance of target temporal frequencies in determining target detectability contrast thresholds. In Sect. 4.2, it was shown in Fig. 4.2 that increasing target temporal frequencies could make targets have higher detectability contrast thresholds, as long as the target temporal frequencies were sufficiently high. Figure 5.2 highlighted how changing target temporal frequencies from 0 Hz to 30 Hz caused targets to have significantly higher contrast thresholds in the unmasked condition, but this effect could be reduced by either increasing the spatial frequencies of the targets, or presenting targets with masks.

Table 5.4 shows how the effect of changing target temporal frequencies from 0 Hz to 30 Hz on target detectability contrast thresholds is effected by presenting targets with masks or increasing target spatial frequencies. The left half of Table 5.4 shows target detectability contrast threshold elevations due to greatly increasing target temporal frequencies, sorted by increasing vertical target spatial frequencies, while the right

half of the table is for diagonal target spatial frequencies. The left and right pair of of columns on each half show the differences in target detectability contrast threshold elevations for masked versus unmasked targets.

Table 5.4: Target detectability contrast threshold elevations due to changes in target temporal frequencies from 0 Hz to 30 Hz for individual target spatial frequencies for unmasked targets, as well as averaged across all masks. The average was taken across all masked elevations available, and the standard deviation is of the elevations used to calculate that average. The overall unmasked average target detectability contrast threshold elevation due to changes in target temporal frequencies from 0 Hz to 30 Hz was 1.63±0.51 log units, while the equivalent masked average was 0.98±0.57 log units.

Vertical	Unmasked		Masked average		Diagonal	Unmasked		Masked average	
c/deg	elev.	±	elev.	±	c/deg	elev.	±	elev.	±
2.8	1.87	0.02	0.88	0.31	2.8	1.87	0.02	0.88	0.31
4.5	1.67	0.03	1.26	0.27	5.7	1.90	0.04	1.13	0.39
6.3	2.07	0.04	1.30	0.06	8.5	1.55	0.04	0.88	0.19
8.2	1.97	0.08	1.64	0.38	11.3	2.02	0.03	1.27	0.49
12.2	1.90	0.03	0.99	0.85	17.0	1.30	0.04	0.35	0.20
16.1	1.33	0.02	0.78	0.92	22.6	0.31	0.04	0.38	0.05
Average	1.80	0.27	1.04	0.61		1.49	0.64	0.90	0.46

Observe from Table 5.4 that making target temporal frequencies 30 Hz makes target detectability contrast thresholds higher by more than a log unit for nearly all target spatial frequencies, and nearly a log unit for many masking conditions. Table 5.4 does show that for two high target spatial frequencies, DCT basis functions [0,7] and [7,7], this target detectability contrast threshold elevation is diminished. However, in general, making target temporal frequencies 30 Hz makes detectability contrast thresholds higher than a target detectability contrast thresholds for targets

with temporal frequencies of 0 Hz.

Recall from Sect. 4.1, Fig. 4.1 showed target detectability thresholds versus target spatial frequencies at three different target temporal frequencies. The top left plot, Fig. 4.1 (a), shows how target detectability contrast thresholds change as targets change from large blocks to vertical lines when the target temporal frequency is 0 Hz. The top right plot, Fig. 4.1 (c), shows the effects of the same change in target spatial frequencies when target temporal frequencies are 30 Hz. Observe from Fig. 4.1 (a) and (c) that, in general, targets with higher temporal frequencies have higher detectability thresholds. For the unmasked targets, the targets with a temporal frequency of 0 Hz and a basis function of DCT [0,7] had higher detectability thresholds than the DCT basis function [0,0] with a temporal frequency of 30 Hz. The targets made from single pixel wide vertical lines had higher detectability thresholds than the targets made from 8×8 pixel blocks when the blocks had a temporal frequency of 30 Hz and the lines had a temporal frequency of 0 Hz. This was not expected.

As shown in Fig. 4.1, both target spatial frequency and masking condition can still have an effect on target detectability thresholds. However, the changes due to either changing target spatial frequencies or masks are slightly reduced when the target temporal frequencies are 30 Hz. The data in this dissertation suggest that it is possible that targets with sufficiently high temporal frequencies have higher detectability contrast thresholds, independent of either target spatial frequencies or masking conditions. Although two entries in Table 5.6 (a) were negative, this was only 3.5% of the target detectability threshold elevations due to changing target temporal frequencies from 0 to 30 Hz. In general, making target temporal frequencies higher will make target detectability thresholds higher, however, the amount of change is still dependent on target spatial frequencies and mask content.

Further research is necessary to understand target temporal frequencies and target detectability thresholds. What target temporal frequencies makes compression arti-

fact detectability thresholds the highest? From Table 5.6, it is evident that increasing target temporal frequencies only slightly is not helpful, as nearly half the times targets with temporal frequencies of 6 Hz had lower detectability thresholds than targets with temporal frequencies of 0 Hz. One may ask if the limit of increasing compression artifact detectability thresholds is only limited by hardware capabilities. As with all applications of research, to truly bear fruit, such findings would eventually need to be implemented in the real world, which is often messy and complicated. The question would then become if the cost of necessary changes in video compression technology would be worth the benefit. Future work is discussed in Chapter 8.

5.7 Natural video masking and target detectability contrast thresholds

This Sect. details our findings on how detectability thresholds change when targets are presented with natural video masks. As seen in Sect. 4.1 and 4.2, in Fig. 4.1 and 4.2, presenting targets with masks can change target detectability contrast thresholds. However, this influence appears to be dependent on which masks are used, and masking effectiveness is diminished as target spatial and temporal frequencies are increased.

Figure 5.3 shows target detectability contrast threshold elevations due to presenting targets with masks for various target temporal and spatial frequencies. Figure 5.3 (a) shows how detectability contrast threshold elevations due to masking change as target temporal frequencies increase. Figure 5.3 (b) and (c) show how detectability contrast threshold elevations due to masking change as target spatial frequencies increase.

Observe in Fig. 5.3 that at low target spatial and temporal frequencies, target detectability contrast threshold elevations due to presenting targets with masks are large. However, Fig. 5.3 (c) also shows that when the target spatial frequencies are high enough, masking the target results in reduced elevations, or can even make

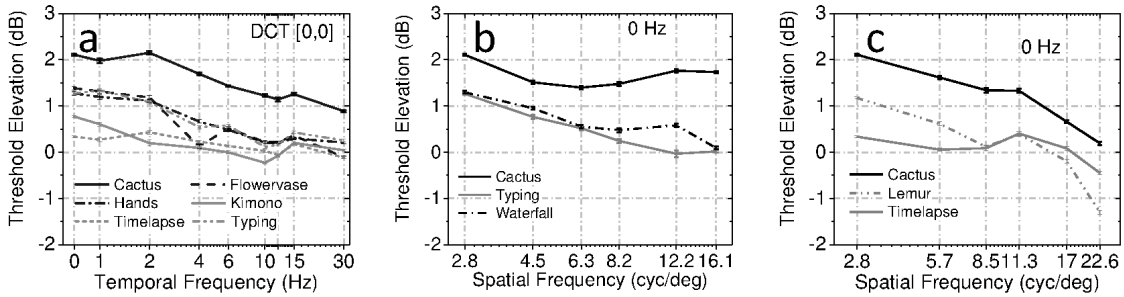

Figure 5.3: Target detectability contrast threshold elevations due to masking for various target temporal and spatial frequencies. (a) shows target detectability contrast threshold elevations due to masking for DCT basis function [0,0] targets for six different masks at temporal frequencies of 0, 1, 2, 4, 6, 10, 12, 15, and 30 Hz. (b) shows the target detectability contrast threshold elevations due to presenting three different masks with targets using DCT basis functions of [0,0], [0,1], [0,2], [0,3], [0,5], and [0,7] and temporal frequencies of 0 Hz. (c) shows the target detectability contrast threshold elevations due to presenting three different masks with targets using DCT basis functions of [0,0], [1,1], [2,2], [3,3], [5,5], and [7,7] and temporal frequencies of 0 Hz. The vertical axis reports the target detectability contrast threshold elevations calculated according to Eq. 5.1. The legend for each plot is in the lower left corner, and shows the masking conditions used for each plot line.

the target detectability threshold elevations negative, as with the mask *Lemur*. A reduction in target detectability contrast thresholds due to presenting a target with a mask is known as facilitation. Although the mask *Cactus* makes a two log unit target detectability threshold elevation when the DCT basis function is [0,0], it makes nearly no difference when the basis function is changed to [7,7]. These results were expected based on previous research [11].

As shown in Fig. 5.3 (a), when target temporal frequencies are high enough, masking targets makes little difference in detectability contrast thresholds. Above target temporal frequencies of 4 Hz - 6 Hz, many of the masks have little effect on target detectability contrast thresholds. At target temporal frequencies of 30 Hz, the mask *Cactus* makes about half the difference it makes at 0 Hz for target detectability contrast threshold elevations. Changes in target detectability contrast thresholds due to natural video masks are discussed further in Sect. 5.9.

5.8 Masked target detectability contrast thresholds that were not expected based on previous research

This Sect. discusses our results that did not meet expectations based on previous research. Previous research suggested that targets higher in spatial frequency and temporal frequency should have higher detectability contrast thresholds. For unmasked targets, these assumptions were generally true, however, presenting masks with targets sometimes reduced or even reversed these trends. Additionally, previous research suggests presenting targets with natural videos should make target detectability contrast thresholds higher, however, our data suggests this was not always the case.

5.8.1 Negative target detectability contrast threshold elevations due to increased target spatial frequencies

Table 5.5 is a summary of negative target detectability contrast threshold elevations due to increasing target spatial frequencies. Table 5.5 provides the average of the differences in target detectability contrast thresholds when targets with higher spatial frequencies had lower detectability contrast thresholds than targets with lower spatial frequencies. These were target detectability contrast threshold elevations calculated for changes in DCT basis function from [0,0], with spatial frequencies of 2.8 c/deg, to [7,7], [0,7], and [3,3], which correspond to spatial frequencies of 22.6 c/deg, 16.1 c/deg, and 11.3 c/deg.

Table 5.5: Negative target detectability contrast threshold elevations due to changing target basis function from DCT [0,0]. The first column to the left signifies what the DCT basis functions were changed to. The second column to the left shows the average of only the negative target detectability contrast threshold elevations due to changes in target spatial frequencies. This average was over all target temporal frequencies and masking conditions. The third column tells the fraction of negative contrast threshold elevations out of the total population for each change in spatial frequencies, and the fourth column gives this fraction as a percent for ease of comparison. What is noteworthy is that there was one case were changing target spatial frequencies from 2.8 c/deg to 22.6 c/deg made the targets have lower detectability contrast thresholds. This special case occurred when target temporal frequencies were 30 Hz, and the targets were shown with the mask *Lemur*.

	elevation	count	percent
DCT [7,7]	-0.36 ± 0.05	(1/12)	8.3%
DCT [0,7]	-0.17 ± 0.13	(10/66)	15.2%
DCT [3,3]	-0.19 ± 0.19	(9/66)	13.6%

Observe from Table 5.5 that for the three increases in target spatial frequencies examined, there was always at least one case were target detectability contrast thresholds were decreased. Table 5.5 shows that for a majority of the data, targets with higher spatial frequencies will have higher detectability contrast thresholds. Additionally, on average, the negative target detectability contrast threshold elevations are small. Also shown in Table 5.5, when the difference in target spatial frequencies is smaller, the probability of finding a negative elevation is higher.

5.8.2 Negative target detectability threshold elevations due to increased target temporal frequencies

This subsection describes the exceptions to the expectation that increased target temporal frequencies results in higher target detectability contrast thresholds. Table 5.6 is a summary of negative target detectability contrast threshold elevations due to increased target temporal frequencies. Table 5.6 provides the average target detectability contrast threshold elevations of the events where higher temporal frequency targets had lower detectability contrast thresholds.

Observe from Table 5.6 that very few targets have lower detectability contrast thresholds when temporal frequencies are changed from 0 Hz to 30 Hz. However, comparing target temporal frequencies of 6 Hz with target temporal frequencies of 0 Hz, keeping target spatial frequencies and masking conditions constant, 42.1% of the higher temporal frequency targets had lower detectability contrast thresholds. For the 24 targets with higher temporal frequencies and lower detectability contrast thresholds, out of a population of 57, the average of the negative target detectability contrast threshold elevations only was -0.28$\pm$0.28 log units, but when examining all target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 6 Hz, the average elevation was -0.02$\pm$0.31 log units. To say it differently, nearly half of the time when the target temporal frequencies are changed

Table 5.6: Negative target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz. The first column to the left signifies what target temporal frequencies were changed to. The second column to the left shows the average of only the negative target detectability contrast threshold elevations due to changes in target temporal frequencies. This average was over all target spatial frequencies and masking conditions. The third column from the left provides the fraction of negative target detectability contrast threshold elevations out of the total population available for the changes in target temporal frequencies, and the fourth column from the left provides this fraction as a percent for ease of comparison. The only negative target detectability contrast threshold elevations due to changing target temporal frequencies from 0 Hz to 30 Hz were for the DCT basis functions [0,7] shown with the masks *Cactus* and *Timelapse*, for elevations of -0.58±0.02 log units and -0.45±0.05 log units.

	elevation	count	percent
30 Hz	-0.51 ± 0.09	(2/57)	3.5%
6 Hz	-0.28± 0.28	(24/57)	42.1%

from 0 Hz to 6 Hz, there is a decrease in detectability thresholds. On average, looking at all elevations due to changing target temporal frequencies from 0 Hz to 6 Hz, the positive elevations are so small that they are essentially canceled out by the negative elevations, and the net result is the average elevation due to increasing the target temporal frequency from 0 Hz to 6 Hz is nearly zero. The largest negative target detectability contrast threshold elevation due to increasing target temporal frequency from 0 Hz to 6 Hz was -0.83±0.05 log units for the DCT basis function [0,0] presented with the mask *Flowervase*. This was also discussed in Sect. 5.6.

5.8.3 Negative target detectability threshold elevations due to presenting targets with masks (Facilitation)

This Sect. summarizes the exceptions to the expectation that presenting targets with masks should make target detectability contrast thresholds higher. Table 5.7 lists the events when masked targets had lower detectability contrast thresholds than unmasked targets, sorted by mask, presented with the average of all the negative target detectability contrast threshold elevations associated with each mask. The bottom of Table 5.7 presents the average of all masking conditions. Table 5.7 shows, by mask, the portion of the population available for comparison, the population with negative target detectability contrast threshold elevations, that measure as a percentage, and the average of all the negative contrast threshold elevations for each mask.

Observe from Table 5.7 that most of the time, presenting a mask with a target will make target detectability contrast thresholds higher. However, Table 5.7 also shows that every mask makes some target detectability contrast thresholds lower. The mask *Cactus* rarely makes target detectability contrast thresholds lower, while the mask *Kimono* makes most of the target detectability contrast thresholds lower. Overall, there is about a one in three probability that presenting targets with masks will make

Table 5.7: Negative target detectability contrast threshold elevations due to presenting targets with masks, sorted by mask. For each mask, the average of all negative target detectability contrast threshold elevations due to presenting targets with masks are reported in the second column, along with the standard deviation of the elevations used in that average calculation in the third column. The fourth and fifth column report the negative target detectability contrast threshold elevation count over the total population available for comparison for that mask. The right column reports the percentage of the population with negative target detectability contrast threshold elevations by mask. The average negative target detectability contrast threshold elevations due to changing masking condition for all masking conditions, as well as that population as a fraction and percentage are at the bottom of the table.

	Average	Count	Percentage
Cactus	-0.13 ±0.05	3/51	5.9%
Flowervase	-0.17 ±0.13	7/27	25.9%
Hands	-0.27 ±0.12	9/27	33.3%
Kimono	-0.43 ±0.26	21/27	77.8%
Lemur	-0.83 ±0.42	9/18	50.0%
Timelapse	-0.43 ±0.33	20/39	51.3%
Typing	-0.20 ±0.14	16/39	41.0%
Waterfall	-0.31 ±0.16	4/18	22.2%
Overall Average	*-0.38*	±0.31	
Total count	89/246	36.2%	

target detectability contrast thresholds lower.

5.9 Discussion of natural video masking and target detectability contrast thresholds

This Sect. discusses our findings on changing target detectability contrast thresholds by presenting targets with natural video masks. The mask *Cactus* appears to be most effective in increasing target detectability contrast thresholds. The data presented in Fig. 4.1 shows that the natural video *Cactus* can often make target detectability contrast thresholds higher as in Fig. 4.1 (d). However, Fig. 4.1 (d) also shows that the abilities of masks to change target detectability contrast thresholds are also dependent on the targets' spatial frequencies and temporal frequencies.

Observe also from Fig. 4.1 (a) that natural videos can make a significant difference in target detectability contrast thresholds, and that difference changes from mask to mask, as suggested by Chandler and Hemami [11]. Figure 4.1 (d) shows that masking effectiveness in elevating target detectability contrast thresholds is also dependent on target spatial and temporal frequencies. In Sect. 5.8.3, Table 5.7 showed that nearly half the natural videos made target detectability contrast thresholds lower about half the time.

The range of effectiveness in raising target detectability contrast thresholds by presenting targets with masks is shown in Fig. 5.4. Figure 5.4 shows the average target detectability contrast threshold elevations due to presenting targets with masks, averaged across target spatial and temporal frequencies. Figure 5.4 (a) shows the average target detectability contrast threshold masking elevations averaged across all masks and all target spatial frequencies, and how those elevations change as the target temporal frequencies increase. Figure 5.4 (b) shows the average target detectability contrast threshold masking elevations averaged across all masks and all target temporal frequencies, and how those elevations change as the target basis functions change

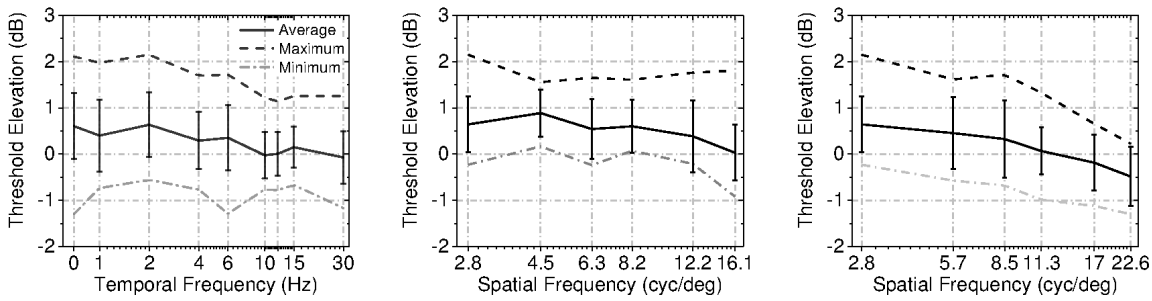

Figure 5.4: Average target detectability contrast threshold elevations due to changes in masking conditions sorted by target frequency. Plot (a) shows target detectability contrast threshold elevations due to masking for targets at different temporal frequencies. The average represents elevations across all masks and all target spatial frequencies at each target temporal frequency. The maximum and minimum plot represents the outer most target detectability contrast threshold elevations due to masking, across all masks and target spatial frequencies for each target temporal frequency. Plots (b) and (c) represent the same information, only grouped by target spatial frequencies and examined across all masks and target temporal frequencies. Plot (b) examines the vertically oriented targets, while plot (c) represents the diagonally oriented targets.

from blocks to lines, while plot (c) shows elevation changes as targets go from blocks to dots.

Observe from Fig. 5.4 that masks have a range of effects on target detectability contrast thresholds and that different natural videos do have different masking capabilities. Figure 5.4 (a) shows that masks can make target detectability contrast thresholds a little lower or much higher at about any target temporal frequency, however the average target detectability contrast threshold elevations above target temporal frequencies of 6 Hz is close to zero. Figure 5.4 (b) and (c) show that masks can make target detectability contrast thresholds a little lower or much higher for any DCT basis function. Figure 5.4 (c) may suggest that as the targets become higher

in spatial frequencies, the masks are less likely to raise target detectability contrast thresholds, or may be more likely to lower target detectability contrast thresholds. This is quantified more clearly in Table 5.8, which shows at what target spatial and temporal frequencies negative elevations were likely to occur.

Table 5.8 shows the averages of the negative target detectability threshold elevations due to presenting targets with masks, sorted by target spatial and temporal frequency. The top of Table 5.8 shows the negative threshold elevations due to masking sorted by target temporal frequency. The two lower parts of Table 5.8 show the negative threshold elevations due to masking sorted by target spatial frequency, with the middle part sorting the data across vertical targets, and the lower part of Table 5.8 sorting the data by diagonal targets.

Observe from Table 5.8 that targets higher in spatial or temporal frequencies are more likely to have lower detectability contrast thresholds when targets are presented with natural videos. This appears to suggest that when unmasked target detectability contrast thresholds are high, due to either sufficiently high target spatial or temporal frequencies, presenting them with natural videos makes their detectability thresholds lower. This has not been suggested by previous research, and encourages further investigation. Future work is discussed in Chapter 8.

Table 5.8: Negative target detectability contrast threshold elevations due to presenting targets with masks, averaged across masks for individual target spatial and temporal frequencies. Table 5.8 shows, as a percentage, how many negative elevations were associated with each target spatial or temporal frequency. The elevations are averaged across all masks for each target frequency.

	Average	Count	Percentage
0 Hz	-0.45 ±0.46	6/46	13.0%
1 Hz	-0.32 ±0.24	8/18	44.4%
2 Hz	-0.30 ±0.22	3/18	16.7%
4 Hz	-0.35 ±0.26	5/18	27.8%
6 Hz	-0.33 ±0.34	16/46	34.8%
10 Hz	-0.40 ±0.25	9/18	50.0%
12 Hz	-0.30 ±0.26	10/18	55.6%
15 Hz	-0.39 ±0.16	5/18	27.8%
30 Hz	-0.44 ±0.36	27/46	58.7%
	Average	Count	Percentage
DCT [0,0]	-0.10 ±0.07	6/60	10.0%
DCT [0,1]	0.00 ±0.00	0/9	0%
DCT [0,2]	-0.21 ±0.04	2/9	22.2%
DCT [0,3]	0.00 ±0.00	0/9	0%
DCT [0,5]	-0.11 ±0.07	5/9	55.6%
DCT [0,7]	-0.40 ±0.24	30/57	52.6%
	Average	Count	Percentage
DCT [0,0]	-0.10 ±0.07	6/60	10.0%
DCT [1,1]	-0.35 ±0.22	3/9	33.3%
DCT [2,2]	-0.51 ±0.28	3/9	33.3%
DCT [3,3]	-0.36 ±0.27	27/57	47.4%
DCT [5,5]	-0.48 ±0.49	6/9	66.7%
DCT [7,7]	-0.84 ±0.46	6/9	66.7%

CHAPTER 6

MODELING

This chapter presents a modeling effort to predict our measured detectability contrast thresholds for natural-video masked dynamic DCT noise. A simple linear regression model provides a summary of those data. These results were also compared against predictions from full reference image and video quality algorithm predictions.

6.1 No-reference linear regression modeling of masked target detectability with a single measure of mask content

This Sect. examines the influence of mask content in predicting target detectability contrast thresholds. In Chap. 4, it was shown that presenting targets with masks had an effect on target detectability contrast thresholds. In Sect. 5.2 , it was shown that target spatial and temporal frequencies mostly define unmasked target detectability contrast thresholds. However, the variations in masked target detectability contrast thresholds were not as fully explained by target spatial and temporal frequencies.

This Sect. details how single measures of mask content can be used to improve fit performance of linear regression models predicting masked target detectability contrast thresholds. Table 6.1 lists several measures of mask content which were considered for inputs to the linear model. [1] These were either standard measures of mask content, common in many image and video processing tools, or simple extensions

[1]Note that the four DCT band specific measures all have the same time. For more efficient calculations, the four measurements were calculated at the same time, and the resulting time required for the calculation was divided equally between the calculations.

or modifications of these common measures.

Observe in Table 6.1 that seventeen measures of mask content are spatial, while only four are temporal. This may suggest the larger body of research on the spatial content of images. Several measures are defined in other literature [69]. The three temporal statistics are simple variations of spatial statistics often used to measure image content. Instead of calculating standard deviation, skewness, and kurtosis of a two dimensional frame for the spatial measures, these measures were calculated on the one dimensional temporal luminance of single pixels. The two recently published sharpness measures have been shown to be useful measures of image content, and are available from the CPIQ lab home page [147, 148]. The Magnitude spectra slope and intercept represent the amount of mask content at different spatial frequencies. The DCT band measurements were found by converting the frames to the frequency domain and removing all content that was not in either the DCT basis function of the target, or one more or less than the horizontal or vertical components of the target for the nearest neighbor case, then returning the frame to the spatial domain and calculating either RMS contrast or kurtosis.

To collapse these measurements into single numbers describing mask content, we calculated either the mean, 2-norm, 5-norm, or maximum of individual measurements. [2] As shown in Table 6.1, the list of candidates contained both spatial and temporal measures of video content. Spatial measures were calculated on a frame by frame basis, while temporal measures were calculated on a pixel by pixel basis. The exception is the VQEG Temporal Perceptual Information, where the calculation is based on the standard deviation of the difference image [149]. This measure is then collapsed like other spatial measurements.

Each measurement, after collapsing, was evaluated as a model input after some treatment. In Fig. 5.3, the data appears to suggest that when target spatial or

[2]Selecting the maximum value out of a set of measurements was suggested by VQEG. [149]

Table 6.1: List of video measurements explored as additional inputs for a linear regression model to predict masked target detectability thresholds. The right column presents the processing time in seconds to calculate each measurement on all 90 frames of each of the eight masks on a standard desktop computer.

Measurement name	Time (sec)
VQEG Spatial Perceptual Information	2.50
Spatial Standard Deviation	1.58
Spatial Skewness	2.78
Spatial Kurtosis	2.74
Spatial Edge Density	15.33
Spatial Entropy	1.73
Spatial Local Entropy	17.69
Spatial Magnitude Slope	7.77
Spatial Magnitude Intercept	7.75
Spatial FISH Sharpness	4.52
Spatial S3 Sharpness	918.54
Spatial Michaelson Contrast	1.37
Spatial RMS Contrast	1.58
Spatial DCT Band RMS Contrast	108.05
Spatial DCT Band Kurtosis	108.05
Spatial DCT Band RMS Contrast Nearest Neighbor	108.05
Spatial DCT Band Kurtosis Nearest Neighbor	108.05
VQEG Temporal Perceptual Information	1.65
Temporal Standard Deviation	2.54
Temporal Skewness	6.25
Temporal Kurtosis	6.22

temporal frequencies increase, masking the targets appears to have less effect on detectability contrast thresholds. Considering this, measurements of video content were evaluated as inputs after dividing them by either target spatial frequency alone, or the sum of the target temporal and spatial frequency. [3] The other two model input treatments were either as is, or squared.

Video content measures, collapsing methods, and measurement treatments, were included in a four input model, using one video content measure in addition to TSF, TTF, and $(TSF \times TTF)$. Coefficients for these models were found using a single pass of the k-fold-cross-validation method, and the resulting average coefficients were used to predict all masked target detectability contrast thresholds. Goodness of fit for the various models is reported in Appendix B.

From Appendix B, the single mask content measure resulting in the best fit of the measured data was spatial standard deviation. Table 6.2 repeats the spatial standard deviation table from Appendix B. Table 6.2 shows the PCC, SROCC, and RMSE between model predictions and target detectability contrast thresholds for a four input model for all collapsing methods and measurement treatments.

Observe in Table 6.2 that adding a single measure of mask content could make a significant improvement to the agreement between model predictions and measured data. However, this improvement is dependent on both how the measurement was collapsed over time, as well as how the measurement was treated before inclusion in the model. The data in Table 6.2 also suggests that the addition of one mask content measure may not fully explain all variations in masked target detectability contrast thresholds.

As shown in Table 6.2, changing how the frame by frame measurements were collapsed into a single measurement did not make a significant difference in overall model

[3] The sum, and not the product, of the target temporal and spatial frequencies was used as the denominator to avoid division by zero.

Table 6.2: Goodness of fit between masked target detectability contrast thresholds and no reference linear regression model predictions using model inputs of TSF, TTF, $(TSF \times TTF)$, and the mask measurement of video spatial standard deviation. The first column identifies how the measure was collapsed. For spatial standard deviation, the measure of each frame was found, and then this measurement was collapsed over time by calculating either the mean, 2-norm, 5-norm, or maximum of all individual frame measurements. The third column lists the fitness scores when the measurement was considered as the fourth input to the linear model without any additional treatment. The fourth column lists the fitness scores when the measurement was squared before inclusion in the model. The fifth column lists fitness scores when the measurement was divided by target spatial frequency before inclusion. The sixth column lists fitness scores when the measurement was divided by the sum of the target spatial frequency and target temporal frequency.

Video Spatial Standard Deviation

Clock Time (sec): 1.58

		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.81	0.76	0.74	0.70
	SROCC	0.81	0.75	0.74	0.70
	RMSE	0.44	0.50	0.51	0.54
2-Norm	PCC	0.81	0.76	0.74	0.70
	SROCC	0.80	0.75	0.74	0.70
	RMSE	0.44	0.50	0.51	0.54
5-Norm	PCC	0.81	0.76	0.74	0.70
	SROCC	0.80	0.75	0.74	0.70
	RMSE	0.45	0.50	0.51	0.54
Max	PCC	0.80	0.75	0.73	0.69
	SRCOO	0.79	0.74	0.73	0.69
	RMSE	0.46	0.50	0.52	0.55

performance. This can also be seen in Appendix B. Taking the average PCC scores across all video measurements and all measurement treatments, using the average to collapse measurements provided correlations that were better than using the 2-Norm method by 0.002, 5-Norm method by 0.003, and maximum method by 0.006.

Observe in Table 6.2 that changing the measurement treatment can change the ability of the model to use the measurement to predict target detectability contrast thresholds. For the video spatial property of frame standard deviation, squaring the measurement was the least detrimental. However, when frame spatial standard deviation was first collapsed over time by finding selecting the maximum measurement, then divided by the sum of target spatial and temporal frequencies, the result was a model prediction that was little better than the prediction using only target properties. As shown in Appendix B, looking at the average PCC scores across all video measurements and all collapsing methods, using the measurement without a treatment was better than squaring the measurement by 0.003, better than dividing by the target spatial frequency by 0.022, and better than dividing by the product of the target spatial and temporal frequencies by 0.017.

Some useful insights can be found by examining this four input model more closely, after all model inputs were normalized. The k-fold-cross-validation method was used to compare more than twenty versions of three and four input linear models that included the video content measure of spatial standard deviation averaged over time as an input. The best coefficients and fit scores from this comparison are listed in Table 6.3. [4] Table 6.3 shows model fit performance on masked and unmasked target detectability contrast thresholds, and provides the coefficients for normalized inputs for those linear models. The data in Table 6.3 suggests that information about mask content may be useful in predicting masked target detectability contrast thresholds.

[4]There was a slight improvement in model fit performance scores in comparison to Table 6.2 when there were more than 20 models of the same form to choose from.

Table 6.3: Summary of no reference linear regression model coefficients and goodness of fit between model predictions and measured data for both masked and unmasked target detectability. The variable P in the two right columns signifies the mask property measurement of video spatial standard deviation. Note that the two and three input models using only target property information can explain most of the variation in unmasked target detectability thresholds. Also note that the four input model that includes video spatial standard deviation as an input does well for explaining most of the variation in masked target detectability thresholds. However, the four input model does not perform as well in predicting masked thresholds as the two input model does in prediction unmasked thresholds.

		Unmasked		Masked			
	input count	2	3	2	3	2 + P	3 + P
fit	PCC	0.961	0.964	0.684	0.690	0.812	0.818
	SROCC	0.958	0.961	0.672	0.673	0.806	0.807
	RMSE	0.266	0.258	0.554	0.550	0.444	0.437
coefficient	constant	-7.169	-7.012	-4.460	-4.355	-2.901	-2.793
	TSF	2.348	2.762	1.182	1.387	1.154	1.364
	TTF	1.656	2.286	1.040	1.356	1.076	1.397
	$(TSF \times TTF)$		-1.359		-0.703		-0.714
	P					-1.079	-1.080

Observe from Table 6.3 that the target property coefficients are similar for all models of masked target detectability contrast thresholds. However, the model coefficients for the mask property inputs are negative in sign. This suggests that as the average frame standard deviation increases, masked target detectability contrast thresholds are going to decrease. Said differently, this appears to suggests that videos that have a more narrow distribution of brightness are more likely to cause masked target detectability contrast thresholds to be higher. This seems somewhat counter intuitive. The mask *Cactus* has a larger average frame standard deviation than the blank gray frame used for the unmasked condition, however, target detectability contrast thresholds for targets presented with the mask *Cactus* were generally higher.

Figure 6.1 shows the scatter plot of target detectability contrast thresholds over average video spatial standard deviation. Observe from Fig. 6.1 that targets shown with masks having higher average video spatial standard deviation tend to have lower masked target detectability contrast thresholds.

As shown in Fig. 6.1 (a), average video spatial standard deviation by itself does not provide a very clear explanation of the variations in target detectability contrast thresholds. It is possible that this is in part due to variations in masked target detectability contrast thresholds due to changes in target spatial and temporal frequencies. However, when only one target spatial and temporal frequency is considered, the relationship between target detectability contrast thresholds and average video spatial standard deviation is not much more evident. Figure 6.1 (b) shows the scatter plot of target detectability contrast thresholds over average video spatial standard deviation for only one target, DCT [0,0], with a temporal frequency of 0 Hz. Note that target detectability for this target spatial and temporal frequency should be the most influenced by mask content.

Observe from Fig. 6.1 (b) that for a single target spatial and temporal frequency, target detectability contrast thresholds do not appear to be clearly dependent on

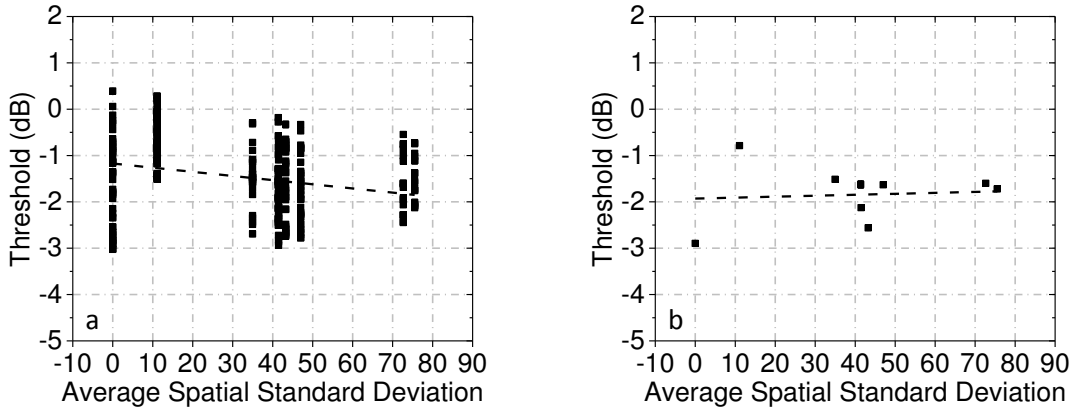

Figure 6.1: Target detectability contrast threshold estimates plotted over average video spatial standard deviation. Plot (a) shows thresholds for all targets, while (b) is for the target DCT [0,0] at 0 Hz. The equation for the line in (a) is Threshold = -0.009 $\times$ average spatial standard deviation + -1.17, and the adjusted R^2 of this model to the data was 0.06. The equation for the line in (b) is Threshold = 0.002 $\times$ average spatial standard deviation + -1.93, and the adjusted R^2 of this model to the data was -0.14.

average video spatial standard deviation. The R^2 values shown in both figures are quite low, and the slope of the trend lines through the data is also quite small. The data in Fig. 6.1 appear to suggest that, although including average video spatial standard deviation as a fourth input to the linear model improves the model fit of masked target detectability contrast thresholds, the mask content measurement by itself may not be an effective predictor of masked target detectability contrast thresholds.

6.2 Complexity analysis of no-reference linear regression modeling of masked target detectability with multiple measures of mask content

This Section shows how well models including target properties and multiple measures of video content can predict measured masked target detectability contrast thresholds. This is an extension of the linear model loosely fashioned after the work of Watson, Hu, and McGowan [6]. A form of the greedy algorithm was employed to select additional mask content measures as model regressors that would most increase goodness of fit, including higher correlations coefficients and lower prediction errors. The starting point for the summary model was a three input model, with inputs of target spatial frequency, target temporal frequency, and average video spatial standard deviation. The $n+1$ model used the previous inputs, as well as the additional input that most improved the model prediction goodness of fit. Table 6.4 reports PCC, SROCC, and RMSE, as well as an overall fitness score, OFS_t, for models with two to fourteen inputs.

The time weighted overall fitness score, OFS_t, was a cost function, with discounts for higher correlation scores, and penalties for larger RMSE and processing time. The time weighted cost function was defined as: $OFS_t = 2 - PCC - SROCC + RMSE + \omega \times EFR$, where ω was a fitting factor of 1/1000, and EFR was the effective frame processing frequency of video content measure calculations. The parameter ω was

subjectively chosen to benefit models who had an EFR above the frame rate at which experiment stimuli were shown, which was 120 Hz. OFS_t represents the cost for performance, where the goal is the best ratio of performance to cost. OFS_t could be reduced by having a larger PCC, larger SROCC, smaller RMSE, or smaller required processing time for calculations.

Observe in Table 6.4 that every additional measure of mask content improved the goodness of fit of the model predictions to the masked target detectability contrast threshold data. When the number of model inputs was smallest, the improvement due to adding another term was larger. However, as the number of inputs grew, the improvement due to additional terms was reduced. The diminishing returns on model performance improvement for increasing model complexity is shown more clearly in Fig. 6.2.

Also shown in Table 6.4, the best overall fitness score, OFS_t, came from the four input model. The four inputs for this model were the target spatial frequency, target temporal frequency, and two different collapsing methods and treatments of the video content measurement spatial standard deviation. Model inputs were selected by choosing the input resulting in the largest increase in PCC, SROCC, and RMSE, and the first two measures of video content resulting in the largest improvement happened to both be variations of spatial standard deviation. Because only the processing time due to the frame by frame measurement was considered, no additional time penalty was added for the collapsing or treatment of measures. The model prediction improved without a decrease in the effective frame rate. Because the remaining additions to the model came at a significant measurement calculation time penalty, the four input model had the most preferred OFS_t. Figure 6.2 shows how the model fit of the measured data improved as the number of model inputs increased.

Figure 6.2 plots two measures of model performance over model complexity. The first measure is the combination of PCC + SROCC - RMSE. The second measure is

Table 6.4: Goodness of fit for predictions from no reference linear regression models with 2 to 14 inputs. The left column shows the number of inputs for the no reference linear regression model, beginning with two target property inputs, TSF and TTF. Using the greedy method to chose the next model input that would most increase goodness of fit, the model was grown by adding one video content measure at a time as an additional model input. Each additional input is listed in the second column from the left. The third column from the left lists video content measurement type, either spatial, (s), or temporal (t). The fourth column from the left lists the measurement collapsing method. The fifth column from the left lists the regressor treatment. The four right columns list the goodness of the model fit. These were found by first using a single pass through the k-fold-cross-validation method to find one set of model coefficients, and then using those coefficients to fit the model to the entire data set.

	video property	type	collapse	treatment	PCC	SROCC	RMSE	OFS_t
2	n/a		n/a	n/a	0.684	0.672	0.554	n/a
3	std. deviation	s	mean	none	0.812	0.806	0.444	0.370
4	std. deviation	s	5 norm	squared	0.859	0.858	0.388	0.215
5	edge density	s	mean	/(SF+TF)	0.887	0.890	0.350	0.531
6	std. deviation	t	max	squared	0.892	0.896	0.343	0.518
7	VQEG P I	t	2 norm	/SF	0.897	0.900	0.336	0.505
8	kurtosis	s	5 norm	/SF	0.905	0.911	0.323	0.477
9	std. deviation	s	2 norm	squared	0.907	0.913	0.320	0.471
10	Michaelson cont.	s	mean	/(SF+TF)	0.908	0.913	0.318	0.468
11	edge density	s	max	none	0.910	0.915	0.314	0.460
12	DCT band RMSC	s	mean	squared	0.911	0.916	0.313	0.480
13	DCT band RMSC	s	max	/SF	0.913	0.917	0.310	0.475
14	skewness	t	mean	/(SF+TF)	0.914	0.918	0.309	0.472

OFS_t, the cost function that incorporates goodness of fit and processing time.

Observe in Fig. 6.2 that the plot of goodness of fit versus model complexity appears to be somewhat asymptotic. This may suggest that there is no combination of model inputs that would provide the ideal PCC of 1, SROCC of 1, and RMSE of 0. This may be due to the lack of proper model inputs. This may also be due to noise in the collected data from human subjects. This may also indicate that the form of the model being used is incorrect, and may require the consideration of other interactions between measures of mask content and target properties.

Also shown in Fig. 6.2, the plot of OFS_t versus model complexity also appears to be somewhat asymptotic. The calculation of frame by frame edge density took about fifteen seconds for all eight videos. Edge density was the third additional property, and because of the significant decrease in the EFR, after four inputs, OFS essentially became 2-PCC - SROCC + RMSE.

The selection of the best summary model still remains an open question. Table 6.3 presents four possible equations to summarize the masked target detectability contrast thresholds. Depending on the application, this list of model inputs and coefficients may be most useful. Table 6.4 presented a list of inputs that would result in the fastest improvement in the goodness of fit of the model prediction to the measured data. Although the list in Table 6.4 may provide a reasonable fit of the measured data, the greedy algorithm selections based on PCC, SROCC, and RMSE only may not provide the most efficient summaries of the measured data.

Using a modification of the greedy method to select additional inputs based on largest improvements to the OFS_t would provide a slightly more efficient list of inputs. The data in Table 6.4 appears to suggest the first four inputs should be target spatial frequency, target temporal frequency, average spatial standard deviation, and the 5-norm of spatial standard deviation squared. Using spatial standard deviation as the fifth term allows an improvement of model fit to the measured data without

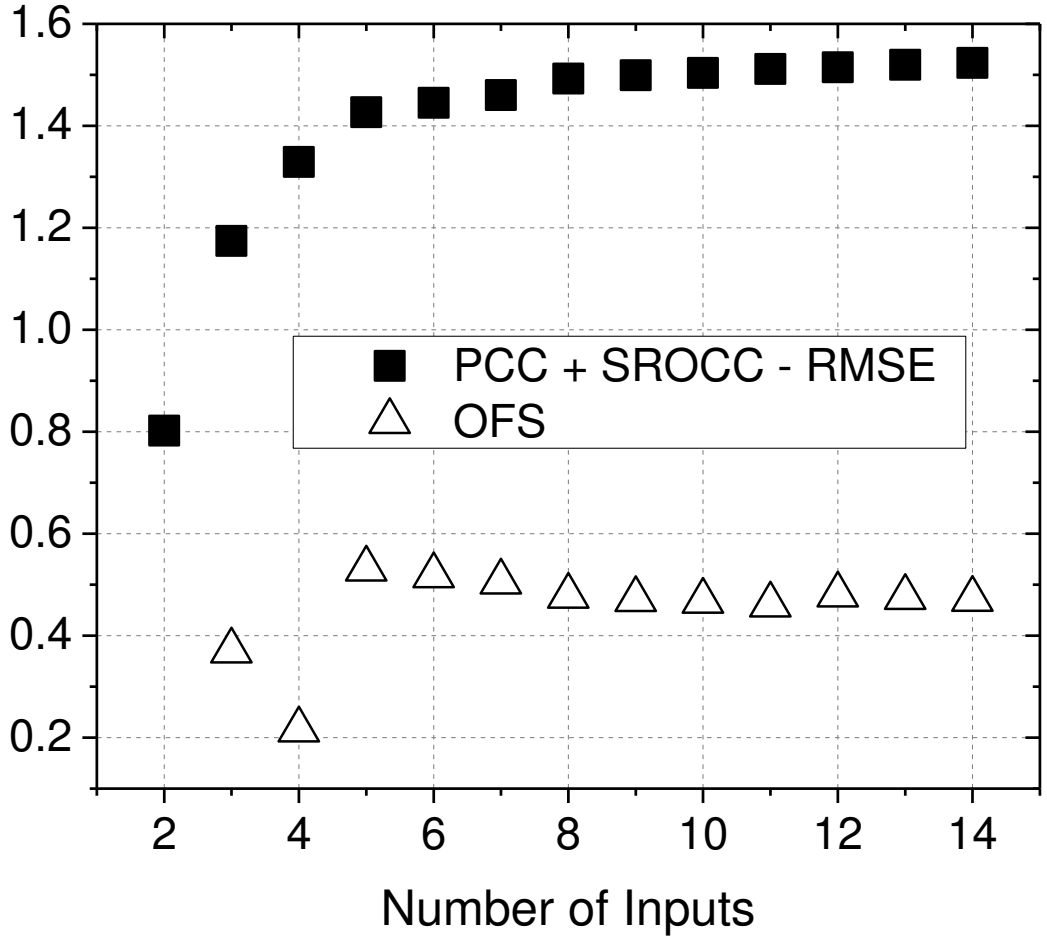

Figure 6.2: OFS_t and goodness of fit versus model complexity. The vertical axis represents either the OFS_t or the sum of the PCC and SROCC minus the RMSE for the goodness of fit of each model. The horizontal axis shows the number of model inputs.

calculating a new measure of video content. Using the maximum frame spatial standard deviation divided by the sum of target spatial and temporal frequencies as the fifth model input provides a PCC of 0.882, SROCC of 0.879, and RMSE of 0.359, and an OFS_t of 0.143.

Given that every collapsing method and measurement treatment provides a slightly different input vector, and thus would result in a slight improvement in goodness of fit, the next additional inputs are likely to be only slight variations of average spatial standard deviation. [1] This is not to say that spatial standard deviation is the best measure of video content to predict variations in masked target detectability contrast thresholds. [2] This data may only suggest that spatial standard deviation could be a useful tool in a more detailed examination of masked target detectability contrast thresholds. Also, the repetition of a single measure of mask content with different collapsing methods and measurement treatments might suggest that collapsing methods and measurement treatments may provide useful information about the human visual system for later consideration.

A better summary of the data might come from a choice that is somewhere between the model with the best PCC, SROCC, and RMSE scores and the model that makes most efficient use of video content measures. A better combination of model inputs could have come from an optimization search that could change all model inputs at once, that was not limited to only adding one input at a time. Other measures of mask content not included in Table 6.1 may also be more useful in quantifying mask content. Other forms of models, such as neural networks, could have provided better fits of the data. However, even the perfect selection of model form and the best set

[1] The sixth input suggested by the OFS greedy method was the maximum frame spatial standard deviation divided by target spatial frequency, and the resulting six input model provided a PCC of 0.882, SROCC of 0.879, and RMSE of 0.358, and an OFS_t of 0.142.

[2] It should be noted that other measures, such as RMS contrast, have some commonality in calculations with spatial standard deviation.

of inputs should not be tuned to provide an exact fit of the data, given the natural variations that occur in such data, collected from experiments with human subjects.

6.3 Summary of masked target detectability with a no-reference linear regression model

The first seven inputs listed in Table 6.4 appear to provide a reasonable fit of the data. The effective frame rate to calculate these measures is more than 30 Hz on a typical desktop machine. The list includes measures of both mask spatial and temporal content. This list does not define the only measures important to classifying mask content. Using the k-fold cross validation method to chose between at least twenty different sets of coefficients for this model form, the set of coefficients providing the best fit were

$$VCT_{masked} = -5.801 + 1.703 \times TSF + 1.316 \times TTF + ...$$

$$-3.084 \times P_1 + 2.299 \times P_2 + 1.322 \times P_3 + -0.270 \times P_4 + 0.476 \times P_5, \qquad (6.1)$$

where P_1 throup P_5 are the first five properties listed in Table 6.4.

Observe the coefficients in Eq. 6.1. The coefficients for TSF and TTF are positive, suggesting that larger target spatial or temporal frequencies predict higher masked target detectability contrast thresholds. The coefficients for P_1 and P_2 are opposite in sign. P_1 and P_2 were different collapsing methods and treatments of spatial standard deviation. This may suggest that the modeling process is using the minor differences in these two vectors to balance each other out. Perhaps if these two measurements were combined into a single input, the coefficient of this single term would be smaller. Note also from Table 6.3 that when only one spatial standard deviation input was included, the coefficient for spatial standard deviation was smaller than the coefficients for the target property terms.

Also shown in Eq. 6.1, for the five coefficients for the video content measures, two coefficients are negative. This might also suggest that the modeling process may not provide coefficients that clearly signify the importance of the individual inputs in predicting masked target detectability thresholds. The modeling process might only be using the small differences in video content measures to make slight improvements in the fit of the model, without providing explicit identification of key predictors for masked target detectability contrast thresholds.

For all video property calculations, each natural video mask pixel value was first converted to luminance using Eq. 3.10. P_1 and P_2 were based on video spatial standard deviation, $VSSD$, which was calculated on a frame by frame basis according to

$$VSSD = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{\frac{1}{2}}, \tag{6.2}$$

were n is the number of pixels in each frame, and $\bar{x}$ was the average luminance of each frame of the mask. [1] For P_1, this value was then average standard deviation over all frames. For P_2, the 5-norm of the standard deviation from all frames for each mask was then squared.

P_3 was video spatial edge density, $VSED$, which was calculated on a frame by frame basis. To calculate $VSED$, pixels belonging to edges were identified using Canny edge detection. [2] For P_3, the average $VSED$ over all the frames was divided by the sum of target spatial and temporal frequency.

P_4 was the standard deviation of the luminance over time for all pixels. [3] Temporal

[1] Standard deviation of the frames was calculated using the MATLAB function `std2`.

[2] The Canny edge detection was performed using the MATLAB function `edge`, employing the `Canny` method, with sensitivity thresholds of [0.08, 0.2], and a standard deviation of the Gaussian filter of 4.5.

[3] The luminance of each pixel over the duration of the 90 frames was converted into a vector. The standard deviation of this vector was calculated using the MATLAB function `std`.

standard deviation was as defined by

$$VTSD = \left(\frac{1}{q} \sum_{i=1}^{q} (PL_i - \bar{PL})^2 \right)^{1/2}, \tag{6.3}$$

were q was the number of pixels in each frame, PL was the luminance of each pixel in each frame, and $\bar{PL}$ was the average luminance of each pixel over all frames of the mask. For P_4, the maximum $VTSD$ for each mask was squared.

P_5 was the video quality experts group measure of temporal perceptual information, $VQEG\ P\ I_t$. This measure is the standard deviation of the pixel luminance of each difference frame, where the difference frame is the next frame minus the current frame in a sequence of frames for a video [149]. For P_5, the 2 norm of all these measurements for each mask was then divided by target spatial frequency. [4]

The information in Table 6.4 suggests a strong correlation between several model predictions and measured target detectability contrast thresholds. From Table 4.1, the PCC from one subject to the next on average was 0.87±0.08. Table 6.4 shows the PCC for models with five or more inputs to be above that score. The SROCC from one subject to the next on average was 0.85±0.10, and models with four or more inputs had better scores for SROCC. The RMSE from one subject to the next on average was 0.58±0.23, and all models of masked data had better scores for RMSE. This comparison suggests that there is as much agreement between subjects as there was agreement between the modeled and measured data. This may suggest that the proposed seven input model may be at the upper limit of correlation and lower limit of prediction error that can be justified by this set of target detectability contrast thresholds.

Although this model provides a PCC of 0.897, SROCC of 0.900, and RMSE of 0.336, it is possible that not all model inputs significantly improve the prediction's fit

[4]Before normalization, the maximum measure for P_1 was 75.5, and the minimum was 11.08. The maximum and minimum for P_2 were 5,743.71 and 164.30, for P_3 were 394.25 and 7.19, for P_4 were 48.29 and 6.28, and for P_5 were 6.74 and 0.10.

of modeled data. The pValue for P_4 was 0.004, and the pValue for P_5 was 0.07, and the coefficients for P_4 and P_5 were also the smallest in Eq. 6.1. This suggests these two additions to the equation were the least significant in matching the measured data. This can also be seen in Table 6.4, were the first five additional model inputs resulted in increases of PCC of 0.128, 0.048, 0.028, 0.005, and 0.004 respectively. This may suggest that the first five terms provide as good of a summary of the data as this form of model can provide. However, this may suggest that target properties are most important in predicting target detectability contrast thresholds, while spatial mask content is still significant, but less important, and finally mask temporal content may not be significant in predicting masked target detectability contrast thresholds.

Figures 6.3 and 6.4 show how well the model predictions matched masked target detectability contrast thresholds. Figure 6.3 demonstrates how predictions and measurements of masked target detectability change as target spatial frequencies increase. The horizontal axis for the top row of plots in Fig. 6.3 are target spatial frequencies, increasing from DCT [0,0] to DCT [0,7], representing targets with vertical orientation. The horizontal axis for the bottom row of plots in Fig. 6.3 are target spatial frequencies, increasing from DCT [0,0] to DCT [7,7], representing targets with diagonal orientation.

Observe in Fig. 6.3 that the seven input model provides a reasonable prediction of the measured target detectability thresholds. The shape and elevation of many of the plots in Fig. 6.3 appear to be similar for both measurements and predictions. Note that the model predicts a significant difference in thresholds when masks are presented with the mask *Cactus*. This is in line with observations from Sections 4.1 and 4.2.

Also shown in Fig. 6.3, there is noticeable separation between plots of predictions and measured data in all three plots. At some target spatial and temporal frequencies, predictions are quite good. However, in Fig. 6.3 (c) and (f), there appears to be

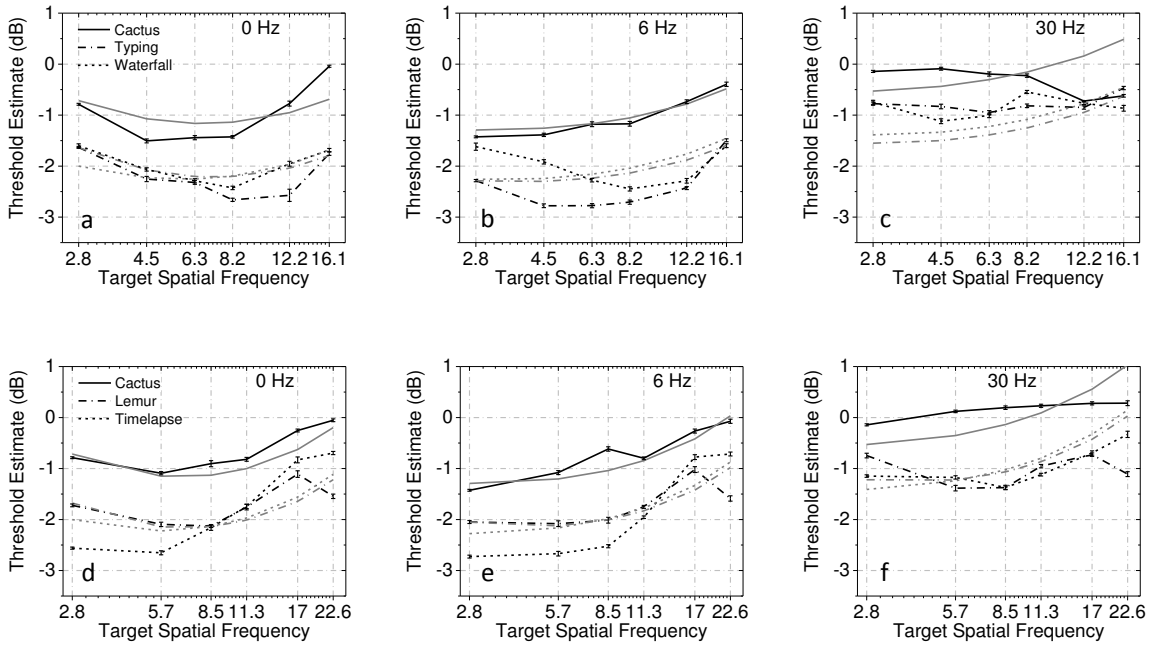

Figure 6.3: Measured and modeled target detectability contrast thresholds plotted over target spatial frequency, measured in c/deg. This figure shows the measured target detectability contrast thresholds in black and modeled threshold estimates in gray. The specific target temporal frequency examined for each plot is listed in the upper right corner of each plot. The model predictions come from the model described in Eq. 6.1.

nearly a log unit of difference between the predicted and measured thresholds for the masking condition *Cactus* at temporal frequencies of 30 Hz and spatial frequencies of 16.1 cyc/deg and 22.6 cyc/deg. The seven input model does not perfectly predict all target spatial frequencies for any masking condition. However, the model was able to correctly predict some differences in the masking abilities of these natural videos.

Figure 6.4 plots measured and modeled target detectability contrast thresholds versus target temporal frequencies. Thresholds for three masking conditions are plotted for three target basis functions, DCT [0,0], [0,7], and [3,3] in Fig. 6.4 (a) and (d), (b) and (e), and (c) and(f) respectively. The axis in the top and bottom rows of Fig. 6.4 are the same. The plots for Fig. 6.4 were split into two rows to more clearly display differences between masking conditions.

Observe in Fig. 6.4 that the model predictions appear to be in line with measured masked target detectability contrast thresholds. The plots in Fig. 6.4 suggest the model does not perfectly predict target detectability contrast thresholds for any one masking condition or target temporal or spatial frequency. Rather the model provides a reasonable prediction of measured data for all target properties and masking conditions.

Also shown in Fig. 6.4, there are significant differences in model predictions and measured thresholds. In Fig. 6.4 (b), for a target temporal frequency of 30 Hz, the prediction for a threshold masked by *Cactus* is nearly a long unit higher than the measurement. Some of the jagged plots of measured target detectability contrast thresholds may suggest that some of these differences may be due to noise from the data collection process. However, it is also possible that the seven input model is not sufficient to adequately capture all the interactions between target properties and mask content measures to properly predict masked target detectability contrast thresholds.

As shown in Fig.s 6.3 and 6.4, the model predictions do not appear to match any

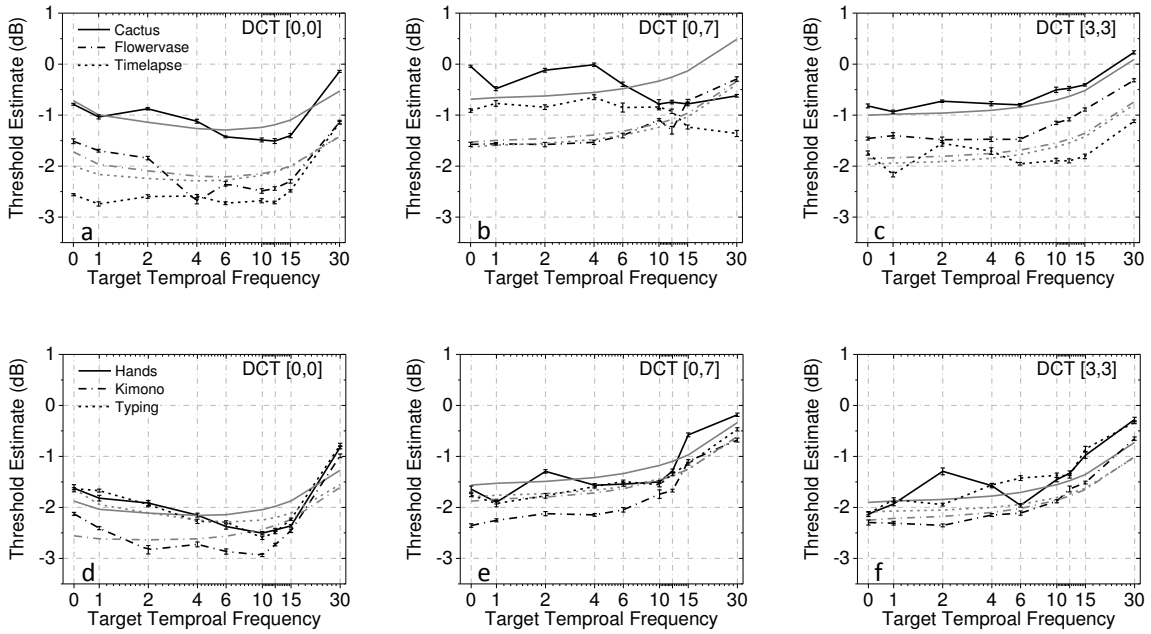

Figure 6.4: Measured and modeled target detectability contrast threshold estimates plotted over target temporal frequency, measured in Hz. This figure shows the measured target detectability contrast thresholds in black and modeled threshold estimates in gray. The specific target spatial frequency examined for each plot is listed in the upper right corner of each plot. The model predictions come from the model described in Eq. 6.1.

one mask perfectly, however the model appears to predict most of the thresholds for various masking conditions reasonably well. This is also demonstrated by the data presented in Table 6.5. Table 6.5 shows correlation scores between model predictions and measured data for individual masks.

Table 6.5: Goodness of fit between measured masked target detectability and predictions from no reference linear regression model defined by Eq. 6.1. The left column lists the masking condition. The first row of scores is for all masking conditions, while remaining rows are for individual masking conditions.

	PCC	SROCC	RMSE	slope	intercept
Overall	0.90	0.90	0.34	0.00	1.00
Flowervase	0.91	0.82	0.25	-0.56	0.68
Cactus	0.85	0.83	0.28	-0.15	0.76
Hands	0.88	0.86	0.29	-0.57	0.64
Kimono	0.90	0.89	0.27	-0.43	0.74
Lemur	0.85	0.84	0.24	0.10	1.02
Timelapse	0.90	0.82	0.34	-0.60	0.59
Typing	0.90	0.90	0.30	-0.87	0.53
Waterfall	0.90	0.73	0.27	-0.52	0.73

Observe in Table 6.5 that all values for PCC and SROCC between predictions and measurements are mostly equal. Also shown in Table 6.5, RMSE for all individual masking conditions was mostly equal. The masking condition *Timelapse* was associated with the highest RMSE, while the masking condition *Lemur* was associated with the smallest RMSE. Predictions and measurements associated with these two masking conditions can be seen in Fig. 6.3 (d)-(f). The predictions and measurements were obviously closer for the masking condition *Lemur*, however, the predictions were not perfectly matched to the measurements. Also, for the masking condition *Time-*

lapse, the predictions and measurements had similarly shaped plots for some ranges of target spatial frequencies. This suggests the modeling process did not favor any one masking condition more than others.

It is interesting that the mask *Lemur* was associated with the smaller SROCC values and at the same time, smaller RMSE values. This suggests that for some masking conditions the process was able to be closer in magnitude, but off more in rank order. Also note that for the mask *Waterfall*, while the PCC value was near the highest in Table 6.5, the SROCC value was the lowest. This may be in part due to the unique concave up shape of the plots of *Waterfall* masked target detectability thresholds, as shown in Fig. 6.3 (a) and (b). This also suggests the modeling process did not optimize coefficients for any one particular measure of goodness of fit.

The models presented in this Sect. provide a reasonable summary of the masked target detectability contrast threshold data collected for this dissertation. The modeling process discussed in this Sect. also provides some useful underlying information from the measured data. It appears that the measure of spatial standard deviation may be of significance. Further examination of masked target detectability may benefit from experiments that control mask standard deviation, which would provide more direct information about the relationships between mask content, target spatial and temporal frequency, and target detectability. Additionally, the data from this Sect. appears to suggest that the question of how to collapse video content measurements into single scores may also merit closer examination.

The seven input linear model provided more meaningful analysis of our data, and provided a concise summary of our results. However, additional research is needed in the area of target detectability prediction. Table 6.1 listed the 21 video content measurements we considered. Several different combinations of video content measurement were able to produce similar results to the ones detailed in this Sect. Additionally, different models with different numbers of inputs were also able to

produce similar results.

6.4 No-reference modeling discussion

This Sect. discusses our findings on linear regression model predictions of masked target detectability contrast thresholds. Observe from Fig. 6.3 and 6.4, and Table 6.5 that a linear regression model provided a reasonable prediction of the measured target detectability contrast thresholds. Masks associated with similar target detectability contrast thresholds had similar model predicted contrast thresholds. Targets shown with mask *Cactus* typically had higher detectability contrast thresholds, and this is also reflected in model predicted contrast thresholds. Table 6.5 shows that, although the model generalized to the group well, it did not provide an excellent fit for any particular mask. Additionally, the model did not provide exact fits for all plots of the data versus either target spatial or temporal frequencies

There are some clear differences between some of the measured and modeled target detectability contrast thresholds in Fig. 6.3 and 6.4. The most noticeable differences are for the contrast thresholds associated with the most exceptional mask tested, *Cactus*. In some cases, the difference between modeled and measured contrast thresholds suggests there may be suspect measurements due to the noisy nature of data collected from live subjects. But in general, there was no single relationship between masked target detectability contrast thresholds and target temporal frequencies or target spatial frequencies. The data in Fig. 6.3 and 6.4 show there was no simple curve to plot all target detectability contrast thresholds over all target temporal frequencies or spatial frequencies that would match well for all masks. This can also be seen in Fig. 5.3, which shows that the different masks caused different target detectability contrast threshold elevations at different target frequencies. This suggests that the target detectability contrast threshold model may need to be a more complicated function of target temporal properties, target spatial properties, and mask

properties. It is also possible that better models would be piecewise defined based on many target and mask properties.

There are many approaches to data modeling, including functional models, biologically inspired models, and physiologically plausible models. This is a reflection of the level of effort that necessary to provide the proper models used in the many different areas related to human vision and media processing. The functional model provided is only a starting point down this path. This model uses the inputs to best predict masked target detectability contrast thresholds, and does not always use the inputs in the most intuitive manner.

As important as what type of model to implement is the decision of what inputs to use for the model. It should be noted that any single measure, such as either of the useful measures provided by the video quality experts group (VQEG) to estimate video temporal and spatial content [149], may not sufficiently explain the changes in target detectability contrast thresholds due to differences in natural video masks. Figure 6.5 shows the temporal perceptual information measurement from VQEG plotted against the spatial perceptual information measurement from VQEG for each video, and provides a rough classification of the content in each natural video.

Observe from Fig. 6.5 that the eight masks have a range of spatial and temporal content. The data in Fig. 6.5 suggest that the mask *Cactus* is not unique in either spatial or temporal content. This does not appear to be in agreement with our previous results, where the mask *Cactus* appeared to possess unique masking abilities.

The data in this Sect. appear to suggest that individual measures of mask content are not yet well defined. The plots in this chapter show that masked target detectability contrast thresholds will change depending on target spatial frequencies and target temporal frequencies. Any model that does not consider the target being masked may not be able to effectively predict masking capabilities. Although future measures of mask content will help quantify effectiveness in raising target detectability contrast

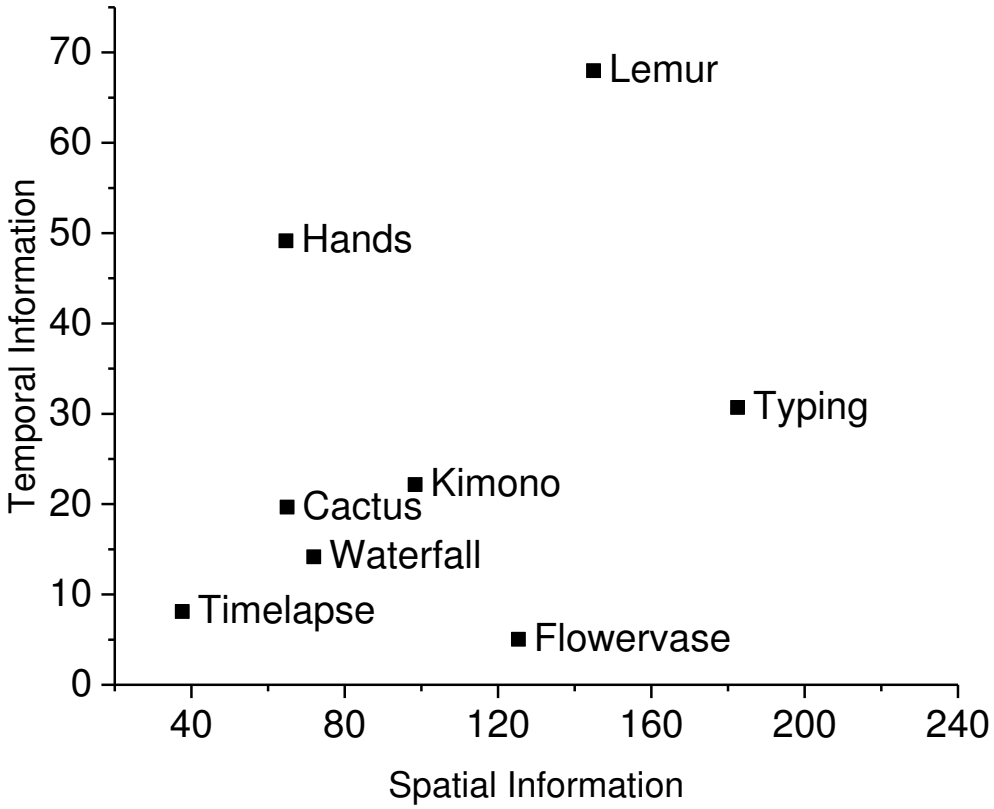

Figure 6.5: VQEG measurements of video temporal and spatial content. The horizontal axis defines the VQEG spatial perceptual classification, averaged across all frames of each mask. The vertical axis is a VQEG measure of the differences of the luminance of each video from frame to frame, averaged across all frame differences.

thresholds, due to the complexity of the interactions between the spatio-temporal target properties and the content of the natural video masks, target properties appear to always play a vital role in predicting target detectability contrast thresholds. These results suggest a need for further research in measuring mask content and modeling of natural video masked target detectability thresholds. Future work is discussed in Chapter 8.

6.5 Full reference image and video quality assessment algorithm predictions of masked target detectability contrast thresholds

Full-reference quality assessment algorithms were also used to predict target detectability contrast thresholds. The algorithms provided quality scores for distorted videos. For each algorithm, a single score was selected as a presumed quality threshold score. Using a bisection search method, a level of target contrast for each video was found that would provide the appropriate quality threshold score, and the target contrast of that distorted video was recorded as the prediction. To select the desired quality threshold scores, the algorithms were first used to provide quality scores for all masked distortion videos at the target detectability contrast threshold level, as measured by the human subjects. The quality threshold score was set as the average of all quality scores from all mask and target combinations.[1]

The bisection search had three possible terminations: if the measured quality score was within 0.5% of the threshold quality score, except for SSIM, where the limit was 0.01%; if the target contrast was less that 2% of the possible range of target contrast, usually amounting to about one or two tenths of a dB; and if the quality threshold score was outside the range of possible quality scores for the range of possible target contrast. At the lower limit of target contrast, so few pixels in so

[1]Because the tools used are not adaptive, using them two produce multiple quality scores did not result in any adaptation or tuning allowing improvement in prediction performance.

few frames had distortions that rounding errors would consume any measured target contrast. The upper limit of target contrast is when the distortions were so severe that they saturate the capabilities of the display, allowing no further increase in target contrast. To begin the bisection search, the algorithms provided quality scores for distortions at the upper and lower limits of target contrast. If the quality threshold score was not bounded by the upper and lower limit quality scores, the search was terminated, and a target contrast prediction was made using only the maximum and minimum quality scores and contrast thresholds.

Full reference quality assessment algorithms of varying complexity were examined. The mean squared error, (MSE) score was calculated according to

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} L(\grave{I} - I)\right)^2, \tag{6.4}$$

where m is the number of frames, n is the number of pixels, L signifies that the pixel differences were converted to luminance differences according to Eq. 3.10, and $\grave{I}$ is defined as $\grave{I} = B + I$, were B is the target frame and I is the mask frame. The peak signal-to-noise ratio, (PSNR) score was calculated according to

$$PSNR = \frac{1}{m}\sum_{i=1}^{m}(20 * log_{10}(245) - 10 * log_{10}(MSE)), \tag{6.5}$$

where 245 was the maximum pixel value allowed in any stimuli.[2] The scores for structural similarity, (SSIM)[3], [150] visual signal-to-noise ratio (VSNR)[4], [151] most apparent distortion, (MAD)[5], and a contrast gain control model, (CGCM)[6][69, 45], were calculated on a frame by frame basis, and then collapsed over time by averaging

[2]The calculations for MSE and PSNR were based on wikipedia.org/wiki/Peak signal to noise ratio

[3]The SSIM Matlab code is available from: https://ece.uwaterloo.ca/ z70wang/research/ssim/

[4]The VSNR Matlab code is available from: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html. Any VSNR score greater than 100 was clipped to 100.

[5]The MAD Matlab code is available from: http://vision.okstate.edu/mad/

[6]The CGCM was provided by Mushfiq Alam, as described in their recent publications. Any CGCM score less than -100 was clipped to -100.

the frame by frame scores. Due to time required to complete calculations, scores for MAD and CGCM were calculated on frames 40-50 of the 90 frame stimuli. The command line video quality metric (VQM) scores were calculated on .avi video files. For VQM, the NTIA General Model score was calculated, the meaning of which, as well as a comparison between PSNR, SSIM, and VQM, is summarized by Vranjes, Rimac-Drlje, and Zagar. [152] [7]

Table 6.6 provides a summary of the goodness of fit of the full reference quality assessment algorithm predictions and the measured target detectability contrast thresholds. The fitness of these full reference quality assessment models was assessed by measuring PCC, SROCC, and RMSE between quality assessment predictions and measured target detectability log_{10} RMS contrast energy thresholds. [8] The second column from the right of Table 6.6 details how many times the threshold score was not bounded by the QA scores for the upper and lower limits of target distortion levels. The right column of Table 6.6 details the full reference quality assessment threshold quality score used for predicting target contrasts.

What Table 6.6 shows is that some of the full reference quality assessment algorithms are able to provide an acceptable estimate of target detectability contrast thresholds. The correlation coefficients for MAD are considerably better than SSIM, however SSIM was significantly faster to compute. The current configuration of the CGCM was the slowest score to calculate. However, it should be noted that the

[7] bitmaps were converted to .avi files using software from ffmpeg.org generating raw video with a UYVY422 pixel format. VQM software was downloaded from http://www.its.bldrdoc.gov/resources/video-quality-research/guides-and-tutorials/cvqm-overview.aspx. Because the command line software requires at least four seconds of video, seven copies of the stimuli were generated, and then looped one after another.

[8] The PCC reported was a linear Pearson correlation coefficient calculation after a logistic fitting. This was based on the work by N.D. Narvekar and L. J. Karam, CPBD Sharpness Metric Software," http://ivulab.asu.edu/Quality/CPBD.

Table 6.6: Goodness of fit between measured masked target detectability contrast thresholds and predictions from full reference quality assessment algorithms. The left column lists the full reference quality assessment tool used to make the predictions. The next three columns list the goodness of fit measurements PCC, SROCC, and RMSE. The second column from the right lists how many times the desired threshold quality score was not bounded by the quality scores provided by the quality assessment algorithm for the upper and lower limits of displayable and measurable contrast. The right column lists the threshold quality score used for each algorithm.

	PCC	SROCC	RMSE	outside	threshold
MAD	0.67	0.65	0.94	0 (0%)	0.1087
CGCM	0.52	0.44	0.65	0 (0%)	1.000
SSIM	0.49	0.59	0.91	0 (0%)	0.9836
VSNR	0.12	-0.13	1.33	0 (0%)	45.5182
VQM	0.25	0.50	41.18	12 (4.9%)	0.0108
MSE	0.54	0.62	20.95	155 (63.0%)	1998.5102
PSNR	0.54	0.45	11.78	162 (65.9%)	20.0419

parameters inside the contrast gain control model could have been adjusted to perform faster, and to better fit the data. Also, the contrast gain control model has the best biological plausibility out of the full reference models considered here. Furthermore, the contrast gain control model was not tuned to incorporate any temporal information about the masks.

It should be noted for the full-reference quality assessment algorithms, as shown in Table 6.6, that the best correlation coefficients and prediction errors for full-reference quality assessment algorithms are worse than those for the two-regressor no-reference model for fitting masked data. This may suggest the importance of target spatiotemporal information in predicting target detectability. This may also suggest that models tuned to the specific task of predicting target detectability perform better at predicting target detectability than models that are tuned more for predicting more general video quality. Also, this may suggest that our data relating masked thresholds to target spatiotemporal frequencies could be used to improve the performance of full-reference video quality assessment tools. Many consumers view either medium or high quality video, in which few distortions are perceptible. It may be that the automated video quality assessment algorithm research community would put our target detectability threshold contrast data to good use.

CHAPTER 7

FURTHER INVESTIGATIONS

This chapter presents a further investigation of some previous findings from this dissertation. The chapter begins with a closer examination of how mask properties can change target detectability contrast thresholds, including mask luminance, mask contrast, and mask playback rate. Next, the chapter shows how a slight modification of the targets to make them spatially correlated with mask content changes target detectability contrast thresholds. Finally, the chapter revisits the examination of mask playback rate with respect to the detectability of targets that are spatially correlated with mask content. The data in this chapter came from: three sets of trials for one expert subject; three sets of trials from one expert subject and two sets of trials from an experienced subject; or three sets of trials from an expert subject, two sets of trials from an experienced subject, and two sets of trials from a novice subject. Weighted averages were still calculated according to Eq. 3.1 and Eq. 3.2.

The results of the modeling chapter show some correlation between mask content and target detectability. The meaning of this chapter is to find more of a cause and effect relationship between mask content and target detectability. To be clear, the modeling chapter showed how a few details about masks helped explain variations in target detectability, while this chapter examines how changing a few details about the masks varies target detectability.

7.1 Variations of mask properties and masked target detectability

This Sect. examines relationships between the individual mask properties of luminance, contrast, and playback rate and variations in masked target detectability contrast thresholds. The data in this section were collected using masks with controlled contrast and luminance, as described in Subsect. 7.1.1. Previous work by Watson [72], Chandler, Gaubatz, and Hemami [10], and Kelly [26] and Daly [24] suggest that there are a few key mask properties to examine first. The three mask properties that appear in many different models of vision are luminance, contrast, and motion. This Sect. more closely examines the relationships between mask luminance, contrast and playback rate and variations in masked target detectability contrast thresholds. The method of this examination is to control these three properties while measuring masked target detectability contrast thresholds.

7.1.1 Mask contrast and luminance adjustment

This Sect. describes how mask luminance and contrast are adjusted. Stimulus presented on a two dimensional display, such as an LCD screen, are luminance defined form. The gray scale images in the masks used in this dissertation form shapes on the display by making some points brighter than others. All stimuli used for our research were luminance defined form.

To quantify mask luminance, first the mask pixel values were converted to luminance according to equation 3.10. Next, the average luminance of each mask frame was found using the MATLAB function `mean2`. The average over all the frames for each mask was then found using the MATLAB function `mean`. This was the measure of average luminance for each mask. This information is plotted as the horizontal axis of Fig. 7.1.

A closely related property to luminance is contrast. To quantify the contrast of the masks, the RMS contrast of each frame of each mask was averaged together. The

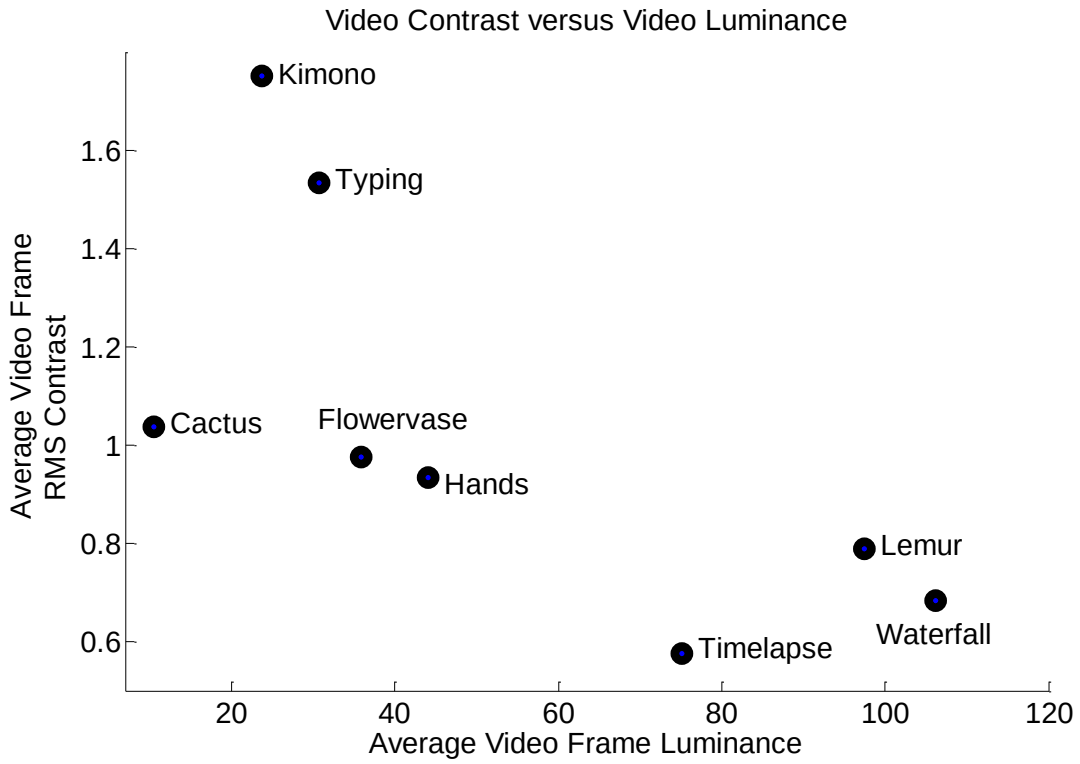

Figure 7.1: Mask average frame RMS contrast plotted over mask average frame luminance. The horizontal axis has units of cd/m^2. The vertical axis is average RMS contrast of each mask. This plot shows how mask contrast and luminance are distributed.

average contrast of each mask is plotted as the vertical axis of Fig. 7.1. The mask RMS contrast for each frame is the standard deviation of the mask frame luminance divided by the mean luminance of the mask frame. The standard deviation of the mask frame was found using the MATLAB function `std2`. The mean luminance of each mask frame was found using the MATLAB function `mean2`.

Fig. 7.1 shows that the average contrast and luminance of each video is different. The mask *Cactus* consistently reduced target detectability more than any other mask. However, Fig. 7.1 shows that the mask *Cactus* has average contrast and low luminance in comparison to the other masks. It is interesting to note that although *Lemur* and *Kimono* caused the same average elevation in detection thresholds, in Fig. 7.1 they are at opposite ends of the graph. Also, the mask *Lemur* and *Waterfall* are close in Fig. 7.1, however, *Lemur* was tied for the least average elevation, while the mask *Waterfall* caused the second highest average elevation. There were no graphs with both high luminance and high contrast.

The average contrast and average luminance of each mask from the experiment is different. These differences make it difficult to discern if either of these properties were related to masking elevations, or if there was some different factor contributing to the differences in elevations due to individual masks. To quantify the effect of mask luminance and contrast on target masking, all the masks were adjusted to have the same average luminance and contrast values.

To adjust mask luminance, a single integer constant, $\alpha_{Luminance}$, was added to the pixel value of each frame, according to

$$\text{Luminance Adjusted Mask} = \alpha_{Luminance} + \text{Original Mask}. \tag{7.1}$$

The appropriate constant for each mask and average luminance was found by a direct search method. When the constant became significant, and caused a sizable shift in pixel values, either positive or negative, some pixels would saturate, and have values outside the display capabilities of the monitor. These values would be clipped to stay

within display limits. Pixel values less than 0 were made 0, and pixel values more than 245 were made 245.

Changing average mask luminance can also change mask average contrast. Mask contrast is a known contributor to masking effectiveness, and a resulting change in thresholds may not be solely a factor of luminance unless mask contrast was also controlled. Mask average contrast was adjusted according to

$$\begin{aligned} &\text{Contrast Adjusted Mask} = \\ &\alpha_{Contrast}(\text{Original Mask} - \text{original mask average}) + \\ &\text{original mask average}, \end{aligned} \tag{7.2}$$

as described by Chandler, Gaubatz, and Hemami [10]. In this method, first, the average pixel value of each frame is calculated using the MATLAB function `mean2`. Then the average of this number over all the frames is found. This is the original mask average pixel value. To scale the contrast, first each frame pixel value has the original mask average subtracted from it. Next each frame pixel value is multiplied by the constant $\alpha_{Contrast}$. Finally, the original mask average is added back to each scaled frame pixel value. The advantage of equation 7.2 is that while contrast is scaled up or down, mean luminance will change less.

The scaling factors $\alpha_{Contrast}$ and $\alpha_{Luminance}$ were adjusted at the same time by the method of direct search. In the direct search method, a luminance constant and a contrast constant were selected. The luminance and contrast of each frame were adjusted. The pixel values outside the displayable range were clipped to the displayable limits. The luminance and contrast of the frame was calculated and stored into vectors. This was repeated for all frames in the mask. The average of the frame luminance and contrast vectors was found. If the either the luminance or contrast were too high, a smaller constant was saved for the next iteration. Likewise,

if either were too low, a larger constant was loaded for the next iteration. The step sizes of the changes in constants were reduced in later iterations of the search.

In initial experiment setup, it was found that over adjusting mask luminance and contrast can make the masks seem artificial, losing many finer spatial details, and appearing cartoon like. This was most pronounced when contrast and luminance were above average for the masks. Increasing the luminance of a mask caused less of this unnatural distortion, and masks could still look mostly natural even with mask average luminance values adjusted past 120 cd/m^2. However, the same was not true for contrast. Adjusting masks to a lower contrast level left the masks looking washed out, but not cartoon like. Adjusting masks to a higher contrast level left the mask looking like a binary cartoon of the original mask, made of only white and dark pixels, with no pixel values in between 0 and 245. Also, when the masks were adjusted to a higher luminance level, the contrast level resulting in the cartoon appearance was lower. After an initial subjective evaluation of all masks at different contrast and luminance levels, an acceptable set of contrast and luminance values were selected for both the luminance and contrast experiments. All masks for the luminance experiment were adjusted to an average luminance of 7.5, 15, 30, 60, and 120 candles per meter squared (cd/m^2), with an average contrast of 0.3. All masks for the contrast experiment were adjusted to an average contrast of 0.075, 0.15, 0.30, 0.60, and 0.120, and had a luminance of 30 cd/m^2. All masks for for the playback rate experiment had a luminance of 30 cd/m^2 and a contrast of 0.3. Target detectability contrast thresholds were measured for basis functions of DCT [0,0], [0,7], and [3,3], with temporal frequencies of 0 Hz, 6 Hz, and 30 Hz.

It should be noted that there are other ways to measure and adjust mask luminance and contrast. Also, summarizing the visual properties of an entire video with only two numbers is a crude way to boil down a significant amount of information. However, these are the methods that have been used by previous researchers, and are simple

and effective ways to help find answers to the direct questions of how mask luminance and contrast can change artifact detectability.

7.1.2 Mask luminance and masked target detectability

This Sect. presents our measurements of the relationship between mask luminance and masked target detectability contrast thresholds. Two popular models on human vision, Daly's visual difference predictor, [153] and Watson's DCT quantization optimizer [72], include luminance masking and adaptation components. Daly incorporated a luminance adaptive nonlinearity as the first component of the visible differences predictor [153], stating that it was well known that visual sensitivity varies with luminance. These assumptions were based on previous work measuring photoreceptor responses [154]. It should be noted that in that work, Normann *et. al* [154] used an adaptation study, and the this study was not on compression artifact masking. Watson [72] also included luminance in a model for quantization matrices for images. This was based off of data from the detectability of unmasked DCT blocks. [77]

Several image compression standards (JPEG, MPEG, H.261) are based on the Discrete Cosine Transform (DCT). However, these standards do not specify the actual DCT quantization matrix. Both Ahumada and Watson have provide mathematical formulae to compute a perceptually lossless quantization matrix. The data presented by Watson [75] showed that increasing the luminance of the background made DCT distortions more difficult to see. Specifically, a brighter background should be associated with higher target detectability contrast thresholds. This data and model were attributed to Ahumada and Peterson [77] and Peterson, Ahumada, and Watson [16]. Ahumada and Peterson [77] used data from Peterson *et. al* [70], which described a forced choice experiment where observers stated if they could detect a single magnified DCT basis function or not. This experiment measured detectability of DCT

blocks in an unmasked condition.

Chandler [155] summarized previous findings on luminance masking [17, 156]. The general expectation is that when luminance of the mask increases, the target detectability contrast thresholds should increase. The previous works lead to the expectation that mask luminance and DCT compression artifact like target detectability contrast thresholds should be positively correlated. However, none of the previous data found measured how dynamic DCT noise detectability changes as natural video mask luminance changes.

This Sect. provides the result of an experiment to quantify the relationship between mask luminance and masked target detectability contrast thresholds. Fig. 7.1 shows that the average luminance of the different masks ranged from about 10 cd/m^2 to nearly 110 cd/m^2. The data presented by Watson [75] ranged from darkness to 100 cd/m^2. To examine if these differences in luminance contributed to masking effectiveness, the masks were adjusted in luminance to five levels with a logarithmic spacing. All masks for the luminance experiment were adjusted to an average luminance of 7.5, 15, 30, 60, and 120 cd/m^2. At the same time, the average RMS contrast of these masks was adjusted to 0.30. Detection thresholds were measured for masks *Cactus*, *Waterfall*, *Kimono*, and *Timelapse*. The unmasked condition was also tested for these luminance values. Subject J.E. recorded three sets of 32 trials each for each data point shown in Fig. 7.2. Subject K.J. recorded two sets of 32 trials each for each data point shown in Fig. 7.3.

The testing procedures described in Chapter 3 were used again in this experiment, with the modification that the masks were changed in contrast and luminance. For the masks *Cactus*, *Waterfall*, *Kimono*, and *Timelapse*, as well as the unmasked condition, thresholds were measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 0 Hz. For the masks *Cactus* and *Waterfall*, thresholds were also measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 6 and 30 Hz.

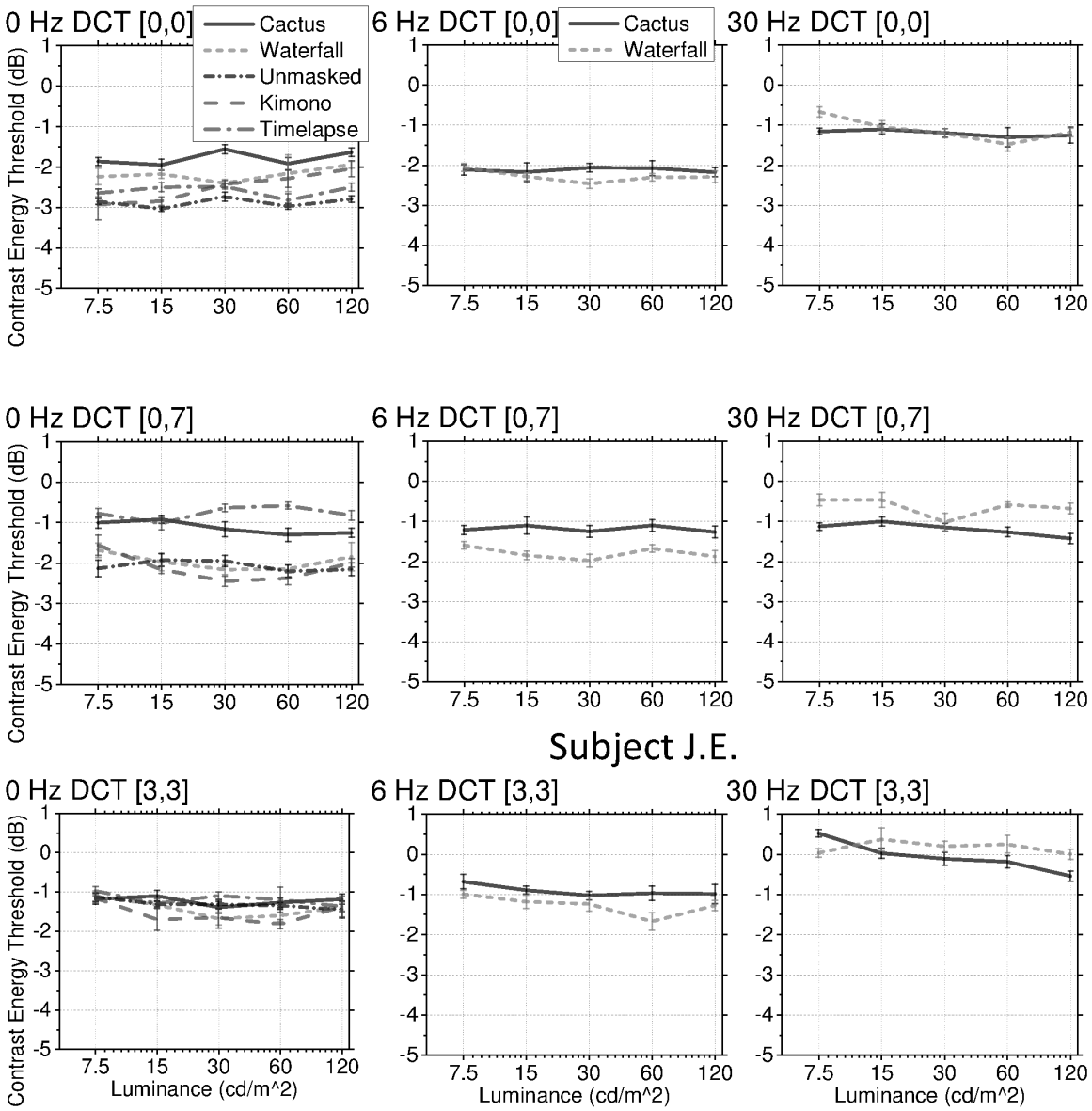

Figure 7.2: Contrast detection thresholds versus luminance for subject J.E. All masks for the luminance experiment were adjusted to an average luminance of 7.5, 15, 30, 60, and 120 cd/m^2. At the same time, the average RMS contrast of these masks was adjusted to 0.30. Detection thresholds were measured for masks *Cactus*, *Waterfall*, *Kimono*, and *Timelapse*. The unmasked condition was also tested for these luminance values. Subject J.E. recorded at least three sets of 32 trials each for each data point. The results of these sets were combined using equation 3.1. Detection thresholds were measured for targets of DCT [0,0], [0,7], and [3,3] and a target temporal frequency of 0 Hz for all masks and the unmasked condition. To understand how the target temporal frequency interacts with luminance, target temporal frequencies of 6, and 30 Hz were presented with masks *Cactus* and *Waterfall* and target detectability contrast thresholds were measured.

To show the agreement between the two subjects, their data has been plotted separately. Figure 7.2 and Fig. 7.3 show that the results of the two subjects are in basic agreement. Most of the error bars in the both plots appear to be reasonable. Most of the data for individual masks and targets appear to fall in straight lines with little slope.

From previous research, it was expected that as luminance of the mask increased, target detectability contrast thresholds should increase. This would mean that the slope of the plots in Fig. 7.2 and Fig. 7.3 should be positive. The masked target detectability contrast thresholds for targets presented with high luminance masks should be higher than those presented with low luminance masks. To quantify this, Table 7.1 shows the difference when the threshold measured with the lowest luminance masks subtracted from the threshold measured with the highest luminance masks. In Table 7.1, positive values represent a positive slope. This was calculated for each subject and each target, then the difference for the two subjects was combined using a simple average. According to the previous research, all numbers in Table 7.1 should all be positive numbers.

In Table 7.1, most of the numbers are negative. The data in Table 7.1 suggests that increasing mask luminance lowers target detectability contrast thresholds in most cases. For the mask *Cactus*, as target spatial and temporal frequency increased, this difference increased. This was true for most other masks, as well as the unmasked condition.

The exception to this was the mask *Kimono* and the target DCT [0,0] at 0 Hz. This combination was one of only three positive elevations due to increased mask luminance, and by far the largest elevation. For negative elevations, the largest came from the combination of the mask *Cactus* and target DCT [3,3] at 30 Hz.

Another way to quantify the relationship between mask luminance and masked target detectability contrast thresholds is with a linear model of normalized target

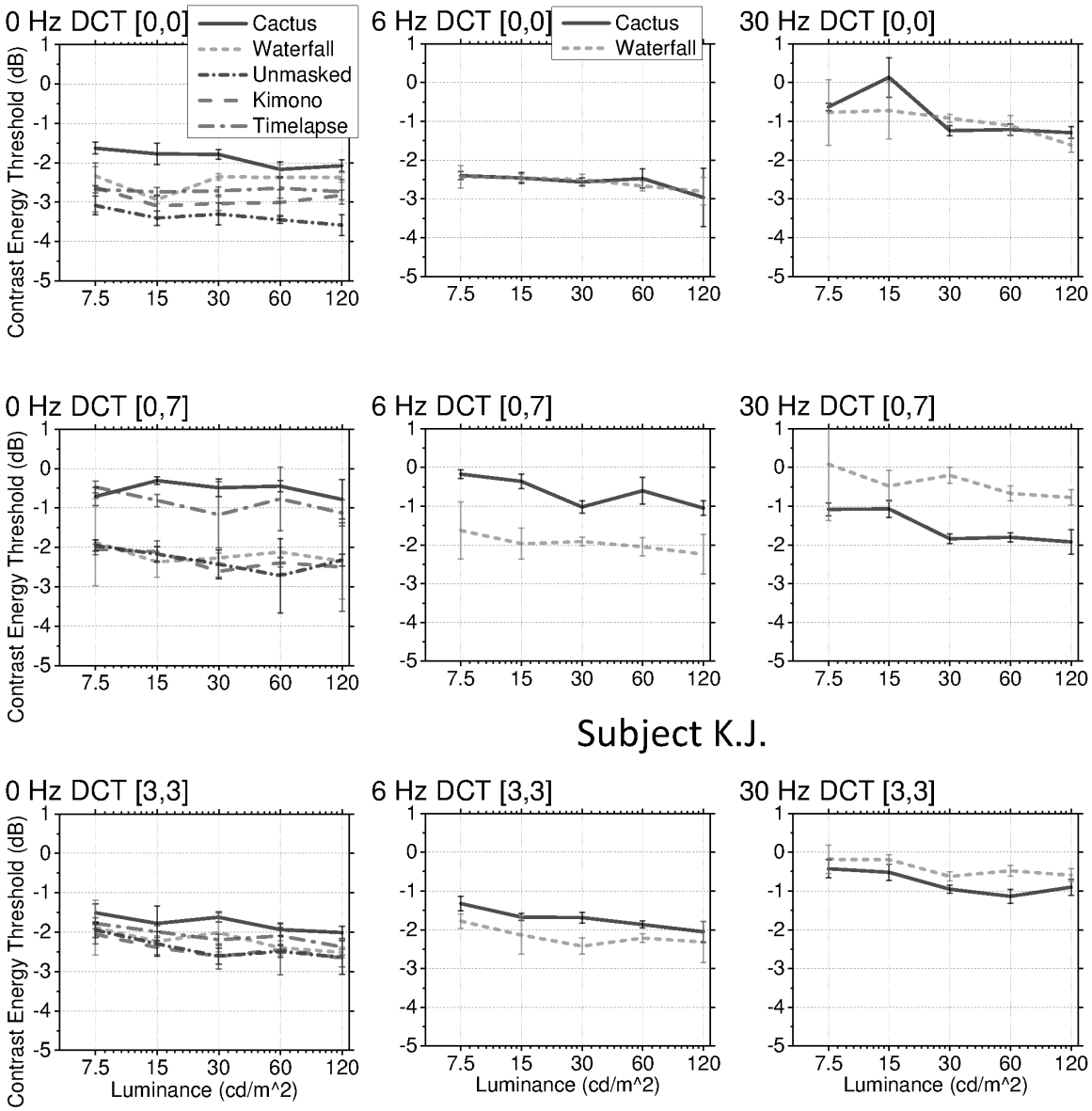

Figure 7.3: Contrast detection thresholds versus luminance for subject K.J. This set of plots is the same axis and experiments as presented in Fig. 7.2, except for subject K.J.

Table 7.1: Target detectability contrast threshold elevation due to significantly increasing mask luminance from 7.5 to 120 cd/m^2. This table shows the differences in target detectability contrast thresholds for targets presented with the highest mask luminance and lowest mask luminance. To generate the data in this table, first the results from the two subjects was combined with a simple average. Next, the elevation for each masking condition, as well as the unmasked condition was calculated for each target. This elevation was calculated by subtracting the threshold associated with the lowest luminance from the threshold associated with the highest luminance for each mask. Negative numbers in this table signify that increasing the luminance of the mask made the target easier to see. The average elevations across masks or target frequencies are shown in italics. The average of all elevations is shown in the lower right hand corner in bold italics.

	Cactus	Waterfall	Unmasked	Kimono	Timelapse	*Average*
0 Hz DCT [0,0]	-0.11	0.13	-0.22	**0.34**	0.04	*0.04*
0 Hz DCT [0,7]	-0.16	-0.32	-0.20	-0.44	-0.35	*-0.30*
0 Hz DCT [3,3]	-0.25	-0.45	-0.50	-0.41	-0.50	*-0.42*
6 Hz DCT [0,0]	-0.31	-0.30				*-0.31*
6 Hz DCT [0,7]	-0.46	-0.44				*-0.45*
6 Hz DCT [3,3]	-0.52	-0.41				*-0.46*
30 Hz DCT [0,0]	-0.38	-0.68				*-0.53*
30 Hz DCT [0,7]	-0.57	-0.54				*-0.55*
30 Hz DCT [3,3]	***-0.77***	-0.22				*-0.50*
Average	*-0.39*	*-0.36*	*-0.31*	*-0.17*	*-0.27*	***-0.30***

spatial and temporal frequencies and mask luminance. There are a total of 116 thresholds to fit from this data set. Simple models of the form

$$VCT_{Luminance} = C + \alpha_1 \times TSF + \alpha_2 \times TTF + \alpha_3 \times (TSFxTTF) + \alpha_4 \times P_1, \quad (7.3)$$

where P_1 was mask luminance, were fit to the luminance adjusted masked target detectability data. For the three input model, α_4 and P_1 were omitted. k-fold cross validation was employed to chose between at least forty different sets of coefficients for each model form. Table 7.2 provides the goodness of fit measures and model coefficients for the winning models. Again, all model inputs were normalized to range from 1 to 2.

Table 7.2: Goodness of fit between measured masked target detectability using luminance controlled masks and predictions from a three and four input no reference linear regression model. The 3 input column reflects a model using only target spatial and temporal frequencies to predict the data. The 3 + P_1 column represents a four input model, including three target spatio-temporal property inputs and mask luminance, for predicting target detectability contrast thresholds.

		3 input	3+P_1 input
fit	PCC	0.790	0.794
	SROCC	0.758	0.768
	RMSE	0.456	0.452
coefficient	constant	-4.298	-4.058
	TSF	1.009	1.009
	TTF	1.640	1.642
	(TSF x TTF)	-0.774	-0.774
	P		-0.179

Observe from Table 7.2 that the addition of mask luminance information did not result in a significant increase in model prediction performance over the model that

was a function of only target properties. This suggests that, although the masks were adjusted to have a range of luminance levels, the differences in target spatial and temporal frequency account for most of the variation in target detectability contrast thresholds. Also note from Table 7.2 that the coefficient for the luminance measurement is close to zero. This also suggests that the mask luminance did not have a significant role in predicting variations in target detectability contrast thresholds.

It is interesting to note from Table 7.2 that the coefficient for the luminance measurement is negative. This is not in line with previous research, which suggested that masks with higher luminance should result in higher target detectability contrast thresholds. However, for these data, there is little correlation between mask luminance and target detectability contrast thresholds on average. The model coefficients are in agreement with the data from Table 7.1 which suggested that for extreme increases in mask luminance, target detectability contrast thresholds also decreased.

This data was not expected. One possible explanation for this is that the previous studies were not looking at the complex question of masking compression like artifacts with natural videos. Specifically, because this is the first study to measure target detectability contrast thresholds in the presence of natural videos, the responses may be primarily driven by some other factors not accounted for. Another possibility is that the way luminance and contrast were adjusted resulted in changes in the images that were not accounted for. Specifically, natural-video mask contrast and luminance had to be adjusted at the same time. This was not a consideration for the previous research cited.

The expectation from Sect. 6.1 was that a simple model should be able to perform reasonably well, but the addition of a meaningful measurement should substantially increase the goodness of fit between model predictions and measured thresholds. In Table 6.3, a three regressor model produced a PCC of 0.690 and RMSE of 0.550, and the addition of the measurement of mask spatial standard deviation increased

the goodness of fit to a PCC of 0.818 and RMSE of 0.437. In this subsection, we see a the addition of luminance as a regressor for the model improves PCC and RMSE scores from 0.790 and 0.456 to 0.794 and 0.452. The target property only model provides a better fit of the data in this subsection. However, the additional mask property measurement input does not significantly improve the model fit of the data. This may suggest that for this data set, mask luminance is not an effective regressor. Table 7.1 suggested that large changes in mask luminance resulted in little change in target detectability contrast thresholds. Considering that mask luminance does not help improve predictions of masked target detectability, and that mask luminance does not appear to effect masked target detectability, the data appear to suggest that for the range of mask luminance examined, mask luminance is not that significant to the perception of masked target detectability contrast thresholds. It should also be noted that all of the masks had the same luminance. Unless the mask luminance was dominating target detectability contrast thresholds, and all masks produced similar target detectability, the addition of mask luminance may not have been helpful to the prediction. Specifically, a plot of the four input model would generate only 1 line for all five plots in Fig. 7.3. Because the plots show a range of target detectability that depends on what mask was used, the addition of a property that assumes all plots should be the same is not significantly beneficial.

This Sect. examined the effect of mask luminance on masked target detectability contrast thresholds. The data suggests that increasing mask luminance results in little change in target detectability, and if any, lower masked target detectability contrast thresholds. This was not what was predicted by previous research. However, none of the previous research was measuring the effectiveness of natural videos in masking dynamic DCT noise. The differences between the unadjusted masks and the luminance adjusted masks may be possibly due to changes in the mask contrast, which was adjusted at the same time as the luminance. This was to ensure that all

masks in the luminance experiment had the appropriate luminance and fixed contrast. The next Sect. explores the effect of mask contrast on dynamic DCT noise detection.

7.1.3 Mask contrast and masked target detectability

This subsection examines how mask contrast can change masking effectiveness. Contrast sensitivity functions and contrast masking have been a part of many masking models. Daly's visual difference predictor, [153] included a contrast sensitivity function as an integral part. The Watson Solomon classic model on the human visual system [45] also takes into consideration contrast sensitivity and masking. From the classic masking experiment by Legge and Foley [38], to the more specific publication by Chandler, Gaubatz, and Hemami [10] on masking of compression artifacts with still images, the expectation is that as the mask contrast increases past a certain point, the target detectability contrast thresholds should also increase. It should be noted that none of the previous data found measured how dynamic DCT noise detectability changes as video mask contrast changes.

To explore if the differences in mask contrast contributed to masking effectiveness, the masks were adjusted in contrast to five levels with a logarithmic spacing. All masks for this experiment were adjusted to an average RMS contrast of 0.075, 0.15, 0.30, 0.60, and 1.20. Average RMS contrast means that the RMS contrast for each frame were calculated, and then that number was averaged over all frames for each individual mask. At the same time, the average luminance of these masks was adjusted to 30 cd/m^2. For the masks *Cactus*, *Waterfall*, *Hands*, *Kimono*, and *Time-lapse* thresholds were measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 0 Hz. For the masks *Cactus*, *Waterfall*, thresholds were also measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 6 and 30 Hz. Detection thresholds were also measured by subject J.E. for targets masked by random frames of gray scale *Pink Noise*, an ideal mask considered to have a power spectral

density that is more similar to natural images. Subject J.E. recorded three sets of 32 trials each for each data point shown in Fig. 7.4. Subject K.J. recorded two sets of 32 trials each for each data point shown in Fig. 7.5. Data for Fig. 7.4 and Fig. 7.5 were combined for individual subjects according to equation 3.1. The testing procedures used in Chapter 3 were used again in this experiment. Only the masks were changed to the new contrast and luminance adjusted masks.

The previous works by Chandler, Gaubatz, and Hemami [10] lead to the expectation that when the mask has higher contrast, dynamic DCT noise targets should have higher detectability contrast thresholds. The graph of target detectability contrast thresholds plotted versus mask contrast should have a positive slope. Again, the individual results from subjects J.E. and K.J. have been plotted separately to allow visual comparison of their results.

Fig. 7.4 and Fig. 7.5 show that the results of the two subjects are in basic agreement. Most of the error bars in both Figs. appear to be reasonable, and not too large. Most of the data for individual masks and targets appear to fall in straight lines, with few exceptions.

Observe in Fig. 7.4 (a) the thresholds associated with the targets presented with the masks of *Timelapse* and *Pink Noise*. At a contrast of 0.075, when the patterns in the images are faint, there is a little over a log unit of difference between the two masks. However, when mask contrast is increased to 1.2, the difference between the two masks is nearly three log units. This may suggest that the contrast of the mask is an indicator of how much the content of the mask is going to effect target detectability. When the masks have very low contrast, there is little difference in what is shown in the mask. However, when the mask has high contrast, the image that forms the mask is more important.

The correlation between mask contrast and differences in masks is repeated for other target spatial and temporal frequencies, but at a reduced level. This may

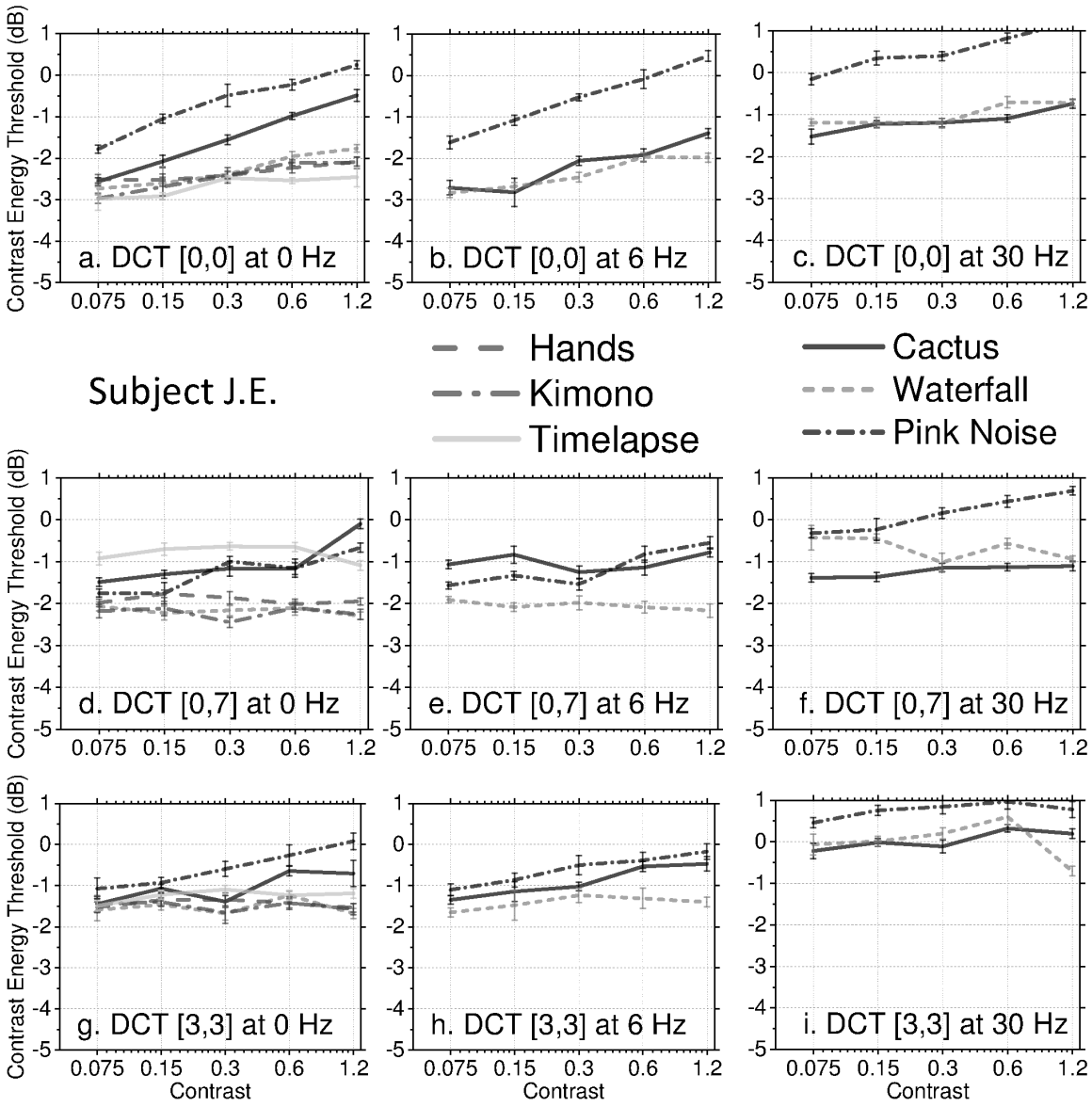

Figure 7.4: Contrast detection thresholds versus mask video contrast for subject J.E. These plots show how target detectability changes with mask contrast. Data in each plot was combined from three sets of trials by J.E. according to equation 3.1. The horizontal axis shows the average RMS contrast of the mask. This was calculated by first finding the RMS contrast of each frame in the mask and taking the average of that number over all frames in each mask. The luminance value of each mask was adjusted to 30 cd/m^2. Luminance and contrast were adjusted according to equation 7.1 and equation 7.2

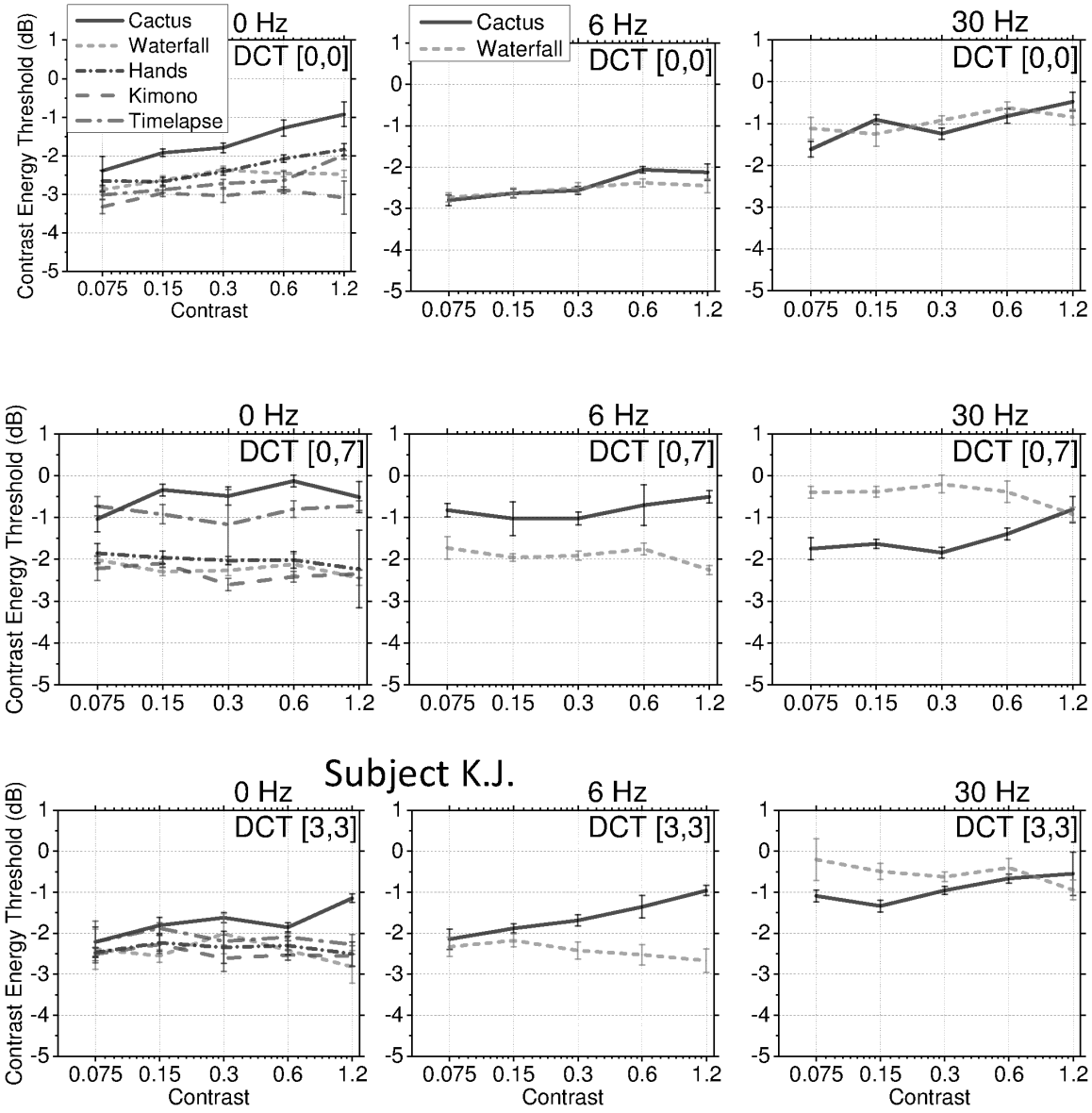

Figure 7.5: Contrast detection thresholds versus mask video contrast for subject K.J. This figure is very similar to Fig. 7.4 except that the subject for this figure is K.J. Subject K.J. completed two sets of trials for each threshold. The data from those sets of trials were combined according to equation 3.1.

suggest that unmasked targets with higher unmasked target detectability thresholds will tend to be less influenced by natural video masking. It is interesting to note that targets with 0 Hz temporal frequencies appear to still have divergent plots of thresholds versus contrast for all spatial frequencies examined. However, when target temporal frequency is increased to 30 Hz, the plots of thresholds versus contrast are more parallel. This may suggest that sufficiently large target temporal frequencies may be somewhat disconnecting or disassociating for the relationship between target detectability and mask contrast.

From previous research, it was expected that as contrast of the mask increased, target detectability contrast thresholds should increase. This would mean that the slope of the plots in Fig. 7.4 and Fig. 7.5 should be positive. The data in Fig. 7.4 and Fig. 7.5 mostly met these expectations. Target detectability contrast thresholds for targets presented with high contrast masks should be higher than those presented with low contrast masks. To quantify this, Table 7.3 shows the difference between detection thresholds when the target was presented with the masks with the highest and lowest contrast. In Table 7.3, positive values signify that the higher contrast mask had a higher resulting detection threshold, as expected. The data in Table 7.3 was calculated using a simple average of the thresholds of the two subjects.

According to the previous research, these should all be positive numbers, and mostly, they are. What Table 7.3 shows is that, in general, increasing contrast of the mask makes the target more difficult to see. The average effect of increasing average RMS contrast of the masks from 0.075 to 1.2 was an increase of 0.32 log units. The largest elevations were seen by adjusting the contrast of the mask *Cactus*, resulting in an average of 0.89 log units. The single combination of target frequency and mask resulting in the largest increase was for *Cactus* and DCT [0,0] at 0 Hz, resulting in an elevation of 1.77 log units. The mask *Waterfall*, had the least elevation due to increasing contrast, which, when averaged over all targets was -0.07 log units. The

Table 7.3: Target detectability contrast threshold elevation due to significantly increasing mask contrast from 0.075 to 1.20. This table shows target viability contrast thresholds elevations when average RMS mask contrast was changed from the lowest level for this experiment, 0.075 to the highest level, 1.2. The values in this table are calculated using simple averages of the data from both subjects. The row headings list the target spatial and temporal frequency tested. The column headings list the mask that was adjusted. Pink noise was not included in this table, as it is not a natural video mask, and only an ideal mask.

	Cactus	Waterfall	Hands	Kimono	Timelapse	*Average*
0 Hz DCT [0,0]	**1.77**	0.69	0.62	0.55	0.78	*0.88*
0 Hz DCT [0,7]	0.96	-0.33	-0.17	-0.10	-0.08	*0.05*
0 Hz DCT [3,3]	0.90	-0.26	-0.02	-0.09	0.12	*0.13*
6 Hz DCT [0,0]	1.00	0.57				*0.78*
6 Hz DCT [0,7]	0.30	-0.39				*-0.04*
6 Hz DCT [3,3]	1.03	-0.04				*0.50*
30 Hz DCT [0,0]	0.96	0.37				*0.67*
30 Hz DCT [0,7]	0.61	-0.52				*0.04*
30 Hz DCT [3,3]	0.48	***-0.69***				*-0.11*
Average	*0.89*	*-0.07*	*0.14*	*0.12*	*0.28*	***0.32***

single combination of target frequency and mask resulting in the largest decrease was for *Waterfall* and DCT [3,3] at 30 Hz, resulting in an elevation of -0.69 log units. On average, the target DCT [0,0] at 0 Hz had the largest elevation of all targets tested, which may suggest that mask contrast had the largest influence over targets that had the lowest unmasked detectability thresholds. Said differently, the easier targets are to see when they are unmasked, the more likely they are to be made much harder to see by masks with high contrast.

Table 7.3 suggests that increasing the contrast of the mask will usually increase target detectability contrast thresholds on average. It is interesting to note that the target DCT [0,0] always was paired with larger positive numbers in Table 7.3. However, other than the mask *Cactus*, targets DCT [0,7] and DCT [3,3] were paired with negative numbers in all but one case. This suggests that, although on average, the higher contrast levels resulted in higher target detectability contrast thresholds, this was not always the case. This was not expected. One important consideration is that increasing contrast made the lower frequency targets have higher target detectability contrast thresholds. That is, the targets that had lower unmasked target detectability contrast thresholds were the ones that had the largest increases in target detectability contrast thresholds due to increasing contrast. However, the targets with higher unmasked target detectability contrast thresholds showed a negative correlation between their detectability contrast thresholds and mask contrast. Said differently, it is possible that high spatial and temporal frequency targets are more likely to experience facilitation with high contrast masks.

Another way to quantify the relationship between mask contrast and masked target detectability contrast thresholds is with a linear model of normalized target spatial and temporal frequencies and mask contrast. There are a total of 180 thresholds to fit from this data set. Simple models of the form

$$VCT_{Contrast} = C + \alpha_1 \times TSF + \alpha_2 \times TTF + \alpha_3 \times (TSFxTTF) + \alpha_4 \times P_1, \quad (7.4)$$

where P_1 was contrast, were fit to the contrast adjusted data. For the three input model, α_4 and P_1 were omitted. k-fold cross validation was employed to chose between at least forty different sets of coefficients for each model form. Table 7.4 provides the goodness of fit measures and model coefficients for the models with the best performance. Again, all model inputs were normalized to range from 1 to 2.

Table 7.4: Goodness of fit between measured masked target detectability using contrast controlled masks and predictions from a three and four input no reference linear regression model. The left column of numbers details the model fit using only target properties, while the right column represents a model that includes mask contrast as an input.

		3 input	3+P_1 input
fit	PCC	0.628	0.669
	SROCC	0.594	0.634
	RMSE	0.726	0.693
coefficient	constant	-3.640	-4.415
	TSF	0.538	0.538
	TTF	1.697	1.689
	(TSF x TTF)	-0.553	-0.542
	P		0.573

The expectation from Sect. 6.1 was that a simple model should be able to perform reasonably well, but the addition of a meaningful measurement should substantially increase the goodness of fit between model predictions and measured thresholds. In Table 6.3, a three regressor model produced a PCC of 0.690 and RMSE of 0.550, and the addition of the measurement of mask spatial standard deviation increased the goodness of fit measurements PCC and RMSE to 0.818 of 0.437. In this Sect., we see the addition of contrast as a regressor for the model improves PCC and RMSE scores

from 0.628 and 0.726 to 0.669 and 0.693. The target property only model provided a better fit of the main data set presented earlier. This may suggest that the quality of data in this set is reduced, and more noisy.

The additional mask property measurement does not significantly improve the model fit of the data. This may suggest that for this data set, mask contrast is not an effective regressor. This may also suggest that this set of data is a representation of data collection error and not the response of mask contrast adjustment on masked target detectability contrast thresholds. Another possibility is that RMS contrast and luminance are not the most meaningful measures of mask content. As all the masks were adjusted to the same contrast levels, how close the plots of target detectability versus mask contrast collapse into a single line, determines how effective it is to add mask contrast as a target detectability predictor. When most of the determining factors are controlled for all masks, they will all have the same plots of target detectability versus whatever measure of mask content. Until all mask properties that cause variance in target detectability are controlled, and all masks produce the same target detectability thresholds, target detectability predictions based only on the controlled experiment parameters will be less than perfect.

It is interesting to note how the model coefficients change when mask contrast is included as a regressor. The target property coefficients are essentially unchanged, and only the constant is shifted. However, the magnitude of the coefficient for the contrast regressor is larger than the regressor for target spatial frequencies. The inclusion of contrast in the regression does not improve the goodness of the fit, however, the size of the coefficient suggests that mask contrast is significant in predicting target detectability thresholds.

The contrast coefficient is positive, suggesting that increasing mask contrast is likely to result in increasing target detectability contrast thresholds. Table 7.3 suggested that large changes in mask contrast resulted in some change in target de-

tectability contrast thresholds.

Observe from Table 7.4 that the addition of mask contrast information did not result in a significant increase in model performance over the model that was a function of only target properties. Granted, either the model including or excluding contrast as a regressor did not perform very well on this data set. This suggests that, although the masks were adjusted to have a range of contrast levels, the differences in target spatial and temporal frequency account for as much of the variation in target detectability contrast thresholds as the mask contrast. Also note from Table 7.4 that the coefficient for the contrast measurement is about the same as the target spatial frequency coefficient. This would also suggest that the mask contrast did have a significant role in predicting variations in target detectability contrast thresholds, but the model fit was not improved by placing more emphasis on mask contrast and less emphasis on target spatial frequencies.

Note from Table 7.4 that the coefficient for the contrast measurement is positive. This is in line with previous research, and suggested that masks with higher contrast should result in higher target detectability contrast thresholds. However, for these data, there is not as much correlation between mask contrast and target detectability contrast thresholds on average. The model coefficients are in agreement with the data from Table 7.4 which suggested that for extreme increases in mask contrast, target detectability contrast thresholds also increased.

The poor model fitness score data was not expected. One possible explanation is that by controlling mask contrast, some of the more obvious differences in videos have been stripped away, leaving behind data that shows true differences in the content of these natural videos, and not only differences in how they are presented. Another possibility is that the way luminance and contrast were adjusted resulted in changes in the images that were not accounted for. Specifically, contrast and luminance had to be adjusted at the same time. This was not a consideration for the previous research

cited. A third possibility is that this data set is too small to work with, and there was too much disagreement between subjects.

7.1.4 Mask lower playback rate and masked target detectability

This subsection examines relationships between mask temporal content and masked target detectability. For this experiment, mask playback rate was controlled. This is a crude measure of motion in mask content. Other measures of mask temporal content are available. For example, Watson discuses apparent motion velocity. This is a definition of how fast an object moves across a scene, that is, how fast something appears to be moving. This measurement is straight forward for a single target moving across a scene, such as a drifting sine wave grating. However, for a mask such as *Cactus*, there are multiple points contributing to the motion in the scene. Each tip of each spine of the cactus would have its own apparent motion velocity. Because the cactus is turning on a pedestal, each spine's apparent velocity across the viewing plane would change as the spine motion vector changed through the course of its revolution. Thus, to quantify the apparent motion velocity of the mask *Cactus* with a single number would at best have to be some sort of an average after lengthy calculations for each pixel of each object in the video.

Daly [24] discuses stimulus retinal velocity. Like the apparent motion velocity discussed by Watson, this is also based on calculations of individual objects moving across a scene. However, this measurement is slightly more complicated because it requires not only knowing how each pixel of each object is moving, but also requires knowing what the subject is looking at in the scene at all times during testing. This information is not available for the current data set.

The goal of visual psychophysics is to explore the human visual system and quantify what is detectable and what the eye is most sensitive to. To facilitate this exploration, and to reduce errors in measurements, researchers have employed controllable

stimulus that are readily defined by mathematical properties [28, 30]. One question visual psychophysical research explores is what is detectable to the eye. This is the quantifiable detectability of a target without the presence of a mask. For example, Robson [8] documented the detectability of unmasked vertical sine wave gratings. Robson showed that stimulus detectability is dependent on the stimulus spatial frequency, that is, how broad the bright and dark lines from the sine wave grating are, as well as the stimulus temporal frequency, or how fast the sine wave gratings flickered. Kelly [25] provided a summary of this information. As Daly [24] summarized, the general shape of these results suggests the general sensitivity of the eye. The eye is most sensitive to targets that are neither too fine or too coarse, and don't flicker too quickly.

Kelly [26] then extended this work to ask how detectability changed as a function of velocity. Specifically, Kelly built an apparatus to control the velocity of a stimulus across the retina. This allowed Kelly to measure how sensitivity to stimulus changes as a function of retinal velocity. Daly [24] also provided a summary of this data. In general, when the stimulus is stopped or nearly stopped in front of the eye, it is more difficult to see. However, when the stimulus has movement, starting from as little as what is common when the eye is looking at a stationary object, the general shape of the sensitivity curve has about the same shape. As the velocity of the stimulus across the retina increases, the sensitivity curve tends to peak over more coarse targets. That is, the faster the target moves across the eye, the harder it is to see fine details.

Laird *et al.* extended Kelly's work [100]. The work of Kelly asked how sensitivity changed as the target moves across the eye. The work of Laird *et al.* asked how sensitivity changed as the eye moved to track the target. Laird *et al.* measured the velocity of an eye with an eye tracker as an observer watched a target move across a monitor. This work confirmed that sensitivity was a function of retinal velocity, however, eye movement did not affect sensitivity. Specifically, the observers could

track a target moving up to 7.5 degrees per second with smooth eye movement, resulting in no change in sensitivity compared to a stationary target.

A more simple way to quantify the temporal properties of a mask is to simply discuss its playback rate. So if a video is modified to play back at half the frame rate as the original, the objects in the modified video should have half the apparent motion velocity as in the original video. Also, this experiment assumes that, because the spatial information of the mask is not edited, the subjects should look in mostly the same locations in each scene. This assumption suggests that by cutting the frame rate in half, the apparent motion velocity and retinal velocity should be halved as well.

This Sect. describes an experiment to quantify how video playback rate effects target detectability. For this experiment, the original masks were first modified in contrast to an average RMS contrast of 0.3, and in luminance to an average luminance of 30 cd/m^2. Then the masks were modified to appear to play back slower through frame duplication. The stimulus duration is 0.75 seconds. Ninety frames are presented in that time. For the data reported in Chapter 4, the forty five images of masks *Cactus*, *Waterfall*, and *Timelapse* would display in those ninety frames. Each image of the source frame would be displayed for two frames. For the other masks, twenty three images would be displayed in the same time, where each image was repeated over four frames. In this Sect., for the first experiment, this playback rate is categorized as 30 frames per second (fps), which was the playback speed used in all other experiments.

The set of masks for the second playback rate, 15 fps, was created by doubling the number of frames each image was repeated. Because doubling how many frames an image is held for means that fewer images can be shown, the 15 fps set of images is a subset of the images used from the 30 fps set of masks. That is, only the middle images were kept. The images at the start and the end were not displayed. Likewise, the set of masks for the third playback rate, 7.5 fps, was created by doubling

the number of frames each image was repeated. Again, the images in the middle were repeated to ensure that the same basic spatial features were shown at all mask temporal frequencies. For the mask playback rate of 1 fps, only the middle image from each mask was repeated for all 90 frames.

Figure 7.6, Fig. 7.7, and Fig. 7.8 show the results for this experiment. All masks for this playback rate experiment were adjusted to an average RMS contrast of 0.30. As described previously at the start of this section, average RMS contrast means that the RMS contrast for each frame was calculated, and then that number was averaged over all frames for an individual mask. At the same time, the average luminance of these masks was adjusted to 30 cd/m^2. This calculation was made only once using all 90 frames. After setting the mask contrast and luminance, the masks were then edited for playback rate and length. For the masks *Cactus*, *Waterfall*, *Kimono*, and *Timelapse* thresholds were measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 0, 6, and 30 Hz. Subject J.E. recorded three sets of 32 trials each for each data point shown in Fig. 7.6. Subject K.J. recorded two sets of 32 trials each for each data point shown in Fig. 7.7. Subject J.P. recorded two sets of 32 trials each for each data point shown in Fig. 7.8. Data for Fig. 7.6, Fig. 7.7, and Fig. 7.8 were combined for individual subjects according to equation 3.1. The testing procedure described in Chapter 3 was used again in this experiment, with the exception that the masks were changed to the new contrast and luminance adjusted masks, and for the masks at 15, 7.5, and 1, fps playback rates were also adjusted.

First, Fig. 7.6, Fig. 7.7, and Fig. 7.8 show that all three subjects are in relatively good agreement. This data is summarized in Table 7.5. The data in table 7.5 are based off a simple average for all seven threshold estimates for each data point in Fig. 7.6, Fig. 7.7, and Fig. 7.8.

What table 7.5 shows is that slowing down the playback rate does not always make a consistent change in elevations. From table 7.5, on average across all targets,

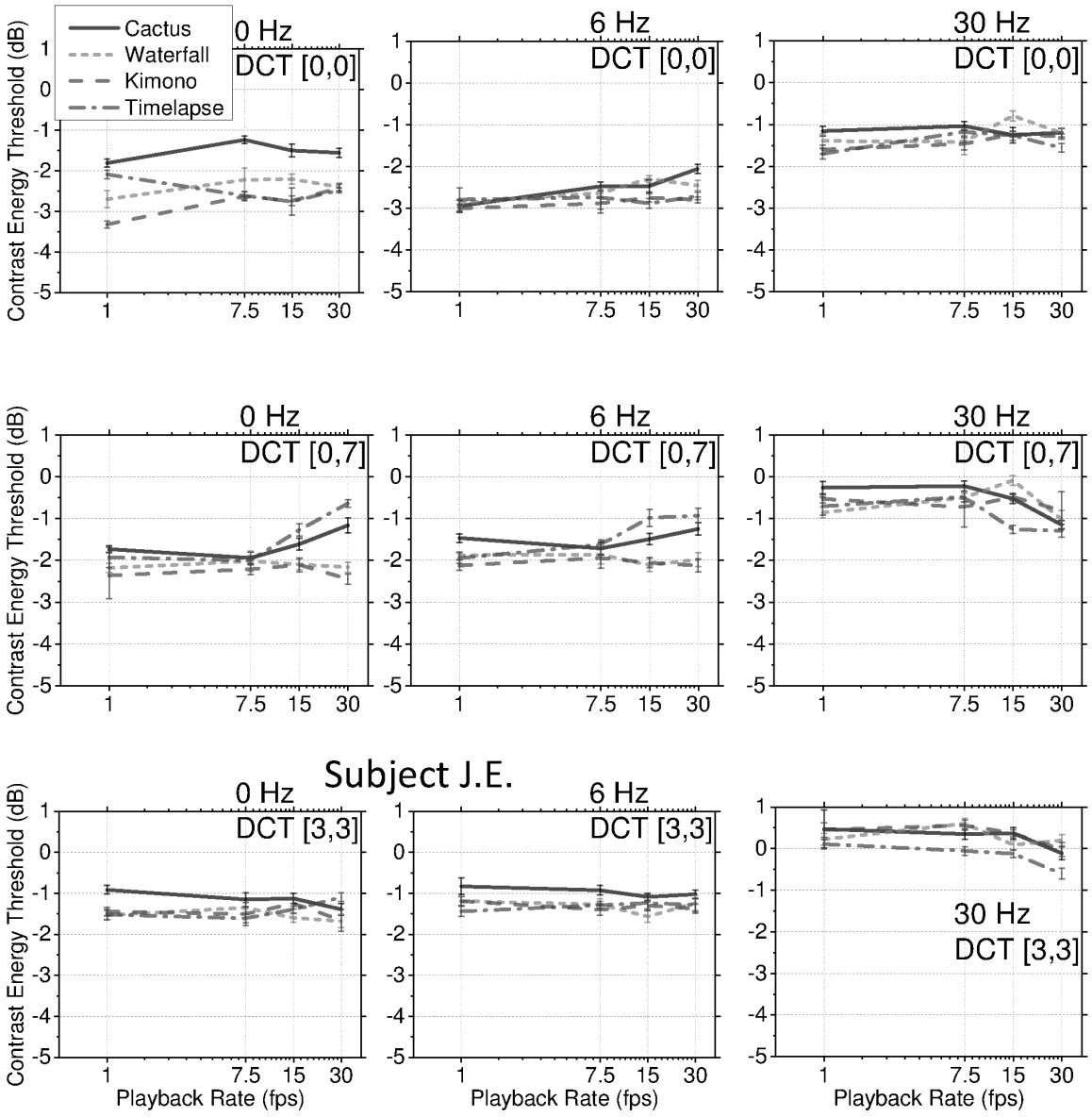

Figure 7.6: Target detection thresholds plotted versus mask playback rate for subject J.E. This figure shows how target detectability changes as the playback rate of the mask changes. The data at 30 fps shows the same mask speed as used in all previous experiments. The data at 1 fps represents target detectability against a stationary image. For this experiment, all masks were adjusted in average luminance to 30 cd/m^2 and average RMS contrast of 0.30.

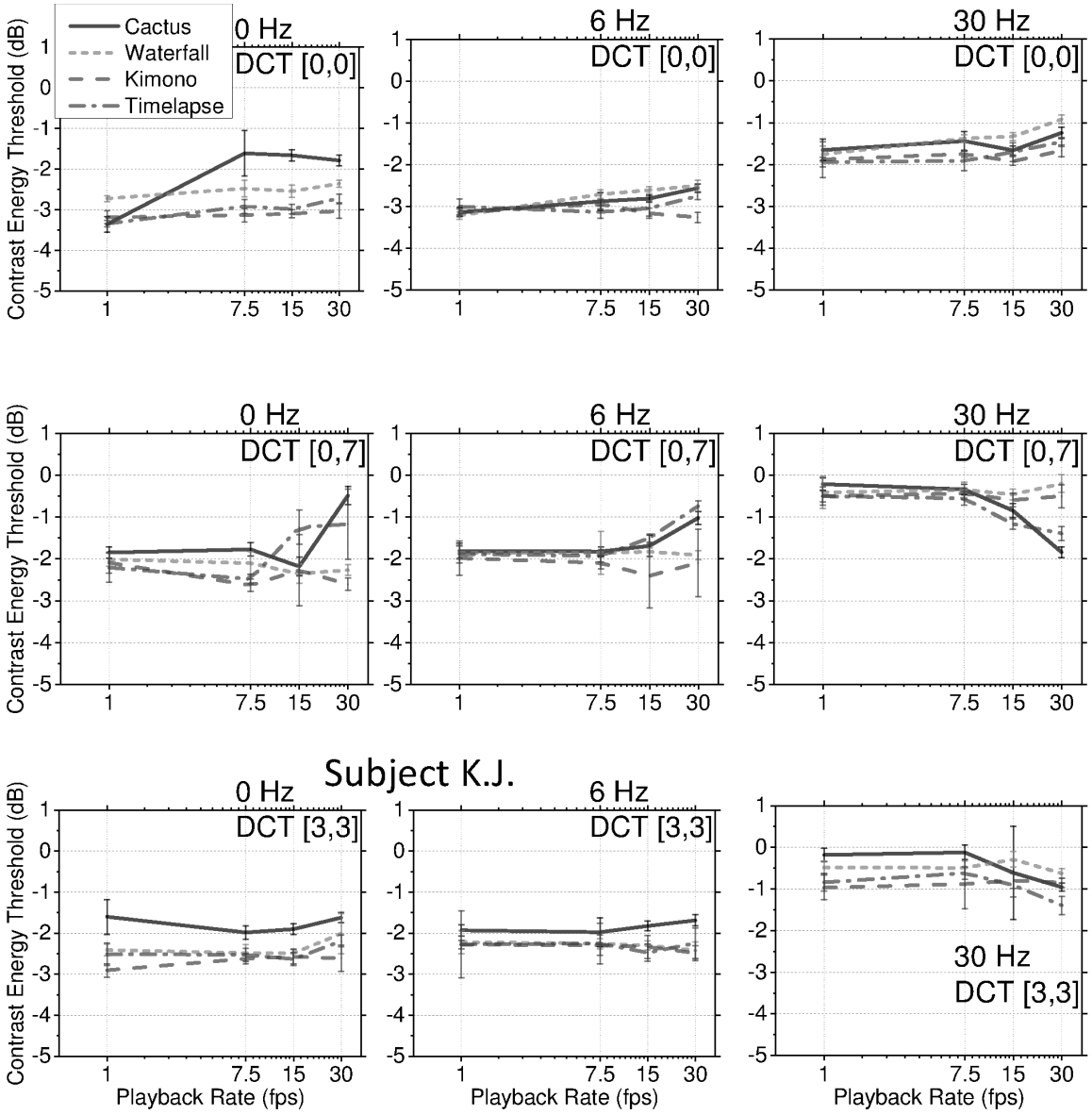

Figure 7.7: Target detection thresholds plotted versus mask playback rate for subject K.J. This figure is similar to Fig. 7.6 except this data is from subject K.J.

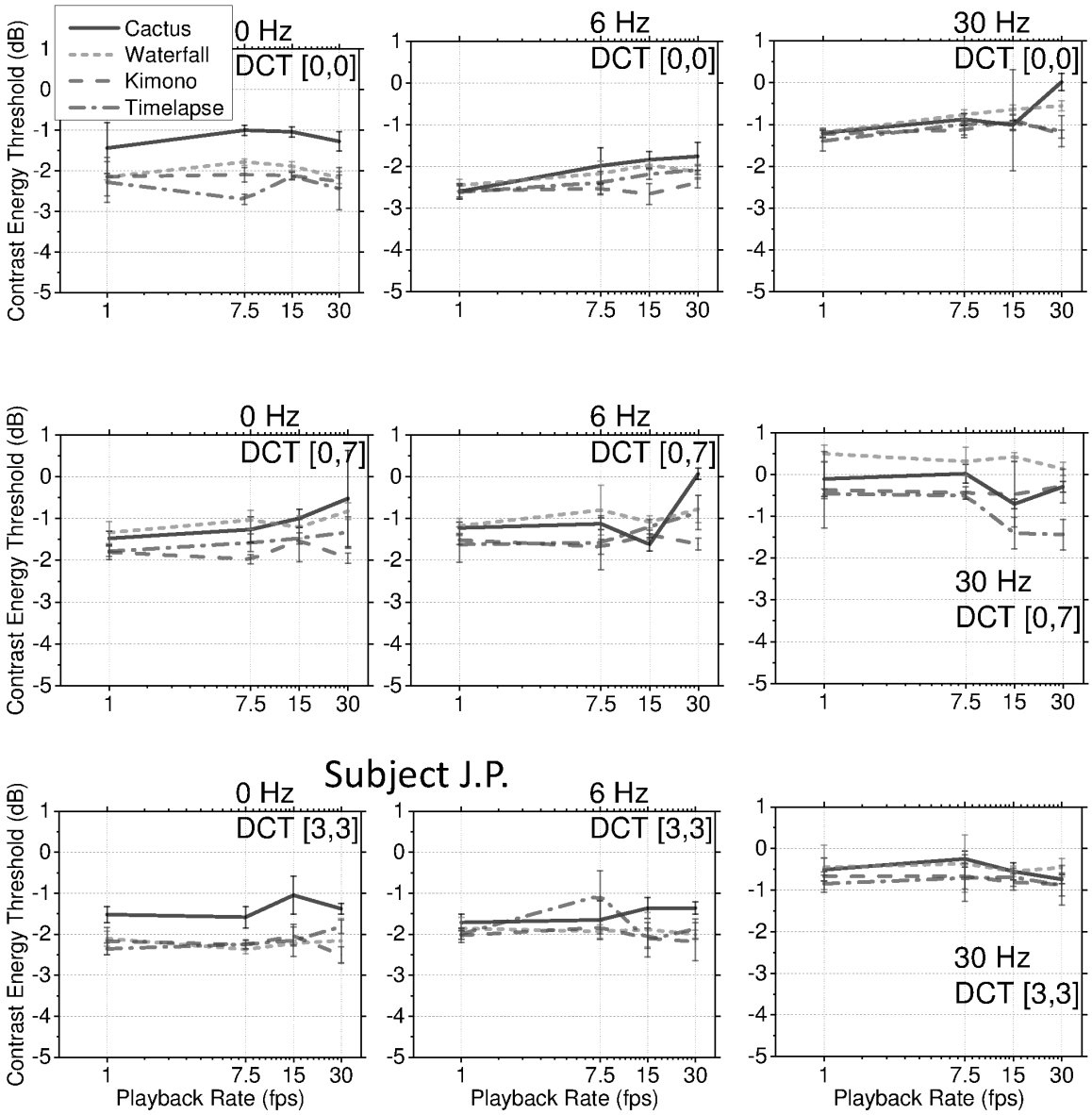

Figure 7.8: Target detection thresholds plotted versus mask playback rate for subject J.P. This figure is similar to Fig. 7.6 except this data is from subject J.P. It should be noted that this visual psychophysics data is the first set of data subject J.P. had ever collected.

Table 7.5: Target detection threshold elevation due to change in playback rate from 1 fps to 30 fps. This table shows the target detection threshold when the mask had a playback rate of 30 fps minus the threshold when the mask was at 1 fps. These elevations were based on a simple average on the seven estimates available from the three subjects.

Target	*Cactus*	*Waterfall*	*Kimono*	*Timelapse*	Average
0 Hz DCT [0,0]	0.67	0.22	0.49	0.06	*0.36*
0 Hz DCT [0,7]	0.91	-0.04	-0.21	**1.04**	*0.42*
0 Hz DCT [3,3]	-0.22	0.06	-0.12	0.40	*0.03*
6 Hz DCT [0,0]	0.75	0.52	0.12	0.27	*0.41*
6 Hz DCT [0,7]	0.65	0.07	-0.03	1.01	*0.43*
6 Hz DCT [3,3]	0.03	-0.09	-0.16	0.14	*-0.02*
30 Hz DCT [0,0]	0.34	0.49	0.22	0.26	*0.33*
30 Hz DCT [0,7]	***-1.01***	-0.02	-0.10	-0.85	*-0.49*
30 Hz DCT [3,3]	-0.52	-0.10	-0.25	-0.62	*-0.37*
Average	*0.18*	*0.12*	*0.00*	*0.19*	***0.12***

masks, and subjects, changing the playback rate from 1 fps to 30 fps results in an elevation of 0.12 log units. The maximum elevation is for the mask *Timelapse* with the target DCT [0,7] at 0 Hz at 1.04 log units. The minimum elevation, mostly equal in magnitude, is for the mask *Cactus* with the target DCT [0,7] at 30 Hz -1.01 log units.

Looking at the averages for each mask across target spatial and temporal frequencies, the largest jump is associated with the mask *Timelapse*, which is 0.19 log units. However, for the average across all targets shown with the mask *Kimono*, there was essentially no change in target detectability when the playback rate was changed from a still image to a video at normal playback rate. This suggests that, considering all the targets presented with the mask at the same time, increasing the playback rate does not seem to make a very consistent difference.

Linear regression can also examine the data. Table 7.6 shows the results of fitting two linear regression models to the playback rate manipulated data. The first model considers only target properties in predicting the data. The second model uses four inputs, including mask playback rate, to predict target detectability thresholds.

Observe from Table 7.6 that the three regressor model provides a reasonable fit of the data. Table 7.6 shows that the three coefficients are in line with previous models in that the higher target spatial and temporal frequencies are associated with higher target detectability contrast thresholds, and that the third regressor, (TSF x TTF), which acts as a limiting factor at high target spatiotemporal frequencies is negative.

Table 7.6 shows that extending the model to include the property of mask playback rate was able to provide a significant improvement in the model fit in both correlation coefficients and RMSE. The coefficient for the mask playback rate regressor is positive. This suggests that there is a positive correlation between faster mask playback rates and higher target detectability contrast thresholds.

It is interesting to note that the coefficient for the third regressor for (TSF x TTF)

Table 7.6: Goodness of fit between measured masked target detectability using playback rate controlled masks and predictions from a three and four input no reference linear regression model. Mask playback rates for the data modeled were 1, 7.5, 15, and 30 Hz. The left column of numbers details the model fit using only target properties, while the right column represents a model that includes mask playback rate as an input.

		3 input	3+P input
fit	PCC	0.790	0.870
	SROCC	0.758	0.830
	RMSE	0.456	0.396
coefficient	constant	-4.296	-4.932
	TSF	1.010	0.851
	TTF	1.641	1.265
	(TSF x TTF)	-0.777	0.130
	P		0.126

had a significant change. For the three regressor model, this coefficient was negative and only slightly smaller in magnitude than the other two target property regressor coefficients. For the four regressor model, this coefficient was positive and significantly smaller in magnitude than the other two target property regressor coefficients. The magnitude for the other two property spatiotemporal regressor coefficients were also slightly reduced in magnitude. This may suggest a unique relationship between mask playback rates and target detectability contrast thresholds.

On average, across all targets and masks, the data does not suggest that there is much difference when the playback rate goes from a normal speed to a still image. When looking at individual masks, over all targets, again, there is not a consistently significant increase in target detectability when the mask is moving at normal speed compared to when the mask is stationary. However, looking at the coefficients of linear regression models for this data, there is a slight positive correlation between mask playback rates and target detectability contrast thresholds. A slightly different question is if these elevations would change if playback rates were increased above normal rates. This question is explored in the next Sect.

7.1.5 Mask higher playback rate and masked target detectability

In Sect. 7.1.4, it was seen that detection thresholds are mostly similar when targets are presented with masks that are either stationary or moving at normal speeds. But what happens if videos are played back at a faster rate? This experiment was set up to ask this different question. Do these thresholds change when the mask is moving faster than normal playback rates? Although this is a slightly unnatural viewing condition, it does help explore the temporal properties of vision.

To explore this, the mask *Cactus* was modified to appear to move faster than normal. The playback rates measured for this experiment were 60, 120, 180, 240, and 300 frames per second. To achieve the normal 30 fps second rate, each image of the

mask was repeated in two frames. To build the 60 fps mask, each image was shown in only one frame. The 120 fps mask would skip every other image from frame to frame. That is, the first image of the source video would show in the first frame of the mask, and then the third image of the source video would show on the second frame of the mask. The 180 fps mask would skip two images from the source video, 240 fps would skip three, and 300 fps would skip 4 images of the source video between frames of the mask. This data is summarized in Table 7.7.

Table 7.7: Image and frame information for mask *Cactus* for high speed playback rate experiment. This table shows how each different playback rate was obtained. For the slowest playback rate, a single frame was held. For the fastest playback rate, four images from the original mask were skipped between frames. Faster playback rates for this mask were not possible because of the limited number or frames available from the original video. Column [b] lists the range of source video frames covered in the 0.75 second video. Column *c* lists this number as an effective frame rate. However, because the frames for the mask *Cactus* were shown twice, column **d** shows the recorded frame rate, where 30 fps corresponds to the normal frame rate used for previous data collection.

Name	[b]	*c*	**d**	Video modification
Still	[1]	*1.33*	**1**	Repeat one image for ninety frames
Quarter speed	[11]	*14.67*	**7.5**	Repeat each image for eight frames
Half Speed	[22]	*29.33*	**15**	Repeat each image for four frames
Normal Speed	[44]	*58.67*	**30**	Repeat each image for two frames
Double speed	[89]	*118.67*	**60**	Each image is its own frame
4 x speed	[178]	*237.33*	**120**	Skip one image between frames
6 x speed	[267]	*356.00*	**180**	Skip two images between frames
8 x speed	[356]	*474.67*	**240**	Skip three images between frames
10 x speed	[445]	*593.33*	**300**	Skip four images between frames

Because of the limited number of frames available in the original video, it was not possible to examine a playback rate faster than 300 fps. But is this fast enough to test the visual capabilities of the eye? Some previous research provides guidance on this matter. As discussed in subsection 7.1.4, Kelly [26], Daly [24], and Laird *et al.* [100] have looked at related questions. Kelly [26] began this discussion by asking how detectability changed as a function of target velocity across the retina. Kelly found that when the stimulus has movement, starting from as little as what is common when the eye is looking at a stationary object, the general shape of the target detectability curve has about the same shape. As the velocity of the stimulus across the retina increases, the detectability curve tends to peak over more coarse targets. That is, the faster the target moves across the eye, the harder it is to see fine details. Laird *et al.* extended Kelly's work [100]. The work of Kelly asked how sensitivity changed as the target moves across the eye. The work of Laird *et al.* asked how sensitivity changed as the eye moved to track the target. Laird *et al.* measured the velocity of an eye with an eye tracker as an observer watched a target move across a monitor. This work confirmed that sensitivity was a function of retinal velocity, however, eye movement did not affect sensitivity. Specifically, the observers could track a target moving up to 7.5 degrees per second with smooth eye movement, resulting in no change in sensitivity compared to a stationary target.

So how fast is the apparent motion velocity when the mask *Cactus* is played back at 300 fps? Does this mask have components that exceed the range tested by Laird *et al.*? A rough calculation of apparent motion velocity will help answer this. The stimulus used in all experiments is 128 pixels wide. The viewing distance is 32 pixels per degree. So the current stimulus is 4 degrees of viewing wide. To be close the threshold of 7.5 degrees per second, an object would have to go from one side of the scene to the other in half a second. The current playback rate is 120 frames per second, so in under 60 frames, an object would need to move from one side of the

scene to the other. Fig. 7.9 provides a view of frame 30 through 52 of the mask *Cactus* when played at 300 fps. In Fig. 7.9, the leading edge of a turning cacti is highlighted with a black line.

Fig. 7.9 shows one of the difficulties in calculating apparent motion velocity for this mask. Because the leading edge of the cactus begins by moving to the front of the scene, and then moves across it, due to the fact that it is revolving about the center of the pedestal, the velocity across the visual plane is not constant. To further complicate the issue, the difference in distance from an object in the scene to the pivot point of the pedestal will dictate the angular velocity, which sets the velocity across the viewing plane. Even though the leading edge appears to move across nearly 4 degrees of visual plane in about a twelfth of a second, or an apparent motion velocity of nearly 48 degrees per second, other parts of the scene will be moving faster still, while others may have effectively no apparent motion, similar to the leading edge of the cactus from frame 32 to 36. Also, it should be noted from the end of the conclusions section by Laird *et al.* [100], that they only tested up to 7.5 degrees per second for eye tracking abilities, but suggested that higher rates be tested to find the limit of human vision. Young [157] had suggested that smooth pursuiant eye movement could be up to 30 c/deg. Daly [24] mentioned that this velocity was often for perfect tracking, and that in tracking objects that occur in nature, and about 80 c/deg is a better estimate of what the eye can track. So it is possible that, although parts of the scene played back at 10 times its normal speed may be above the ability of the human eye to track, there will be points in an objects motion that will be slow enough that the eye may still track it accurately. In which case, as suggested by Laird *et al.* [100], the contrast sensitivity will not be changed by the motion of the video because, if the eye is able to track the motion, the retinal velocity stays low enough that the contrast sensitivity function is not shifted to favor sensitivity to larger objects.

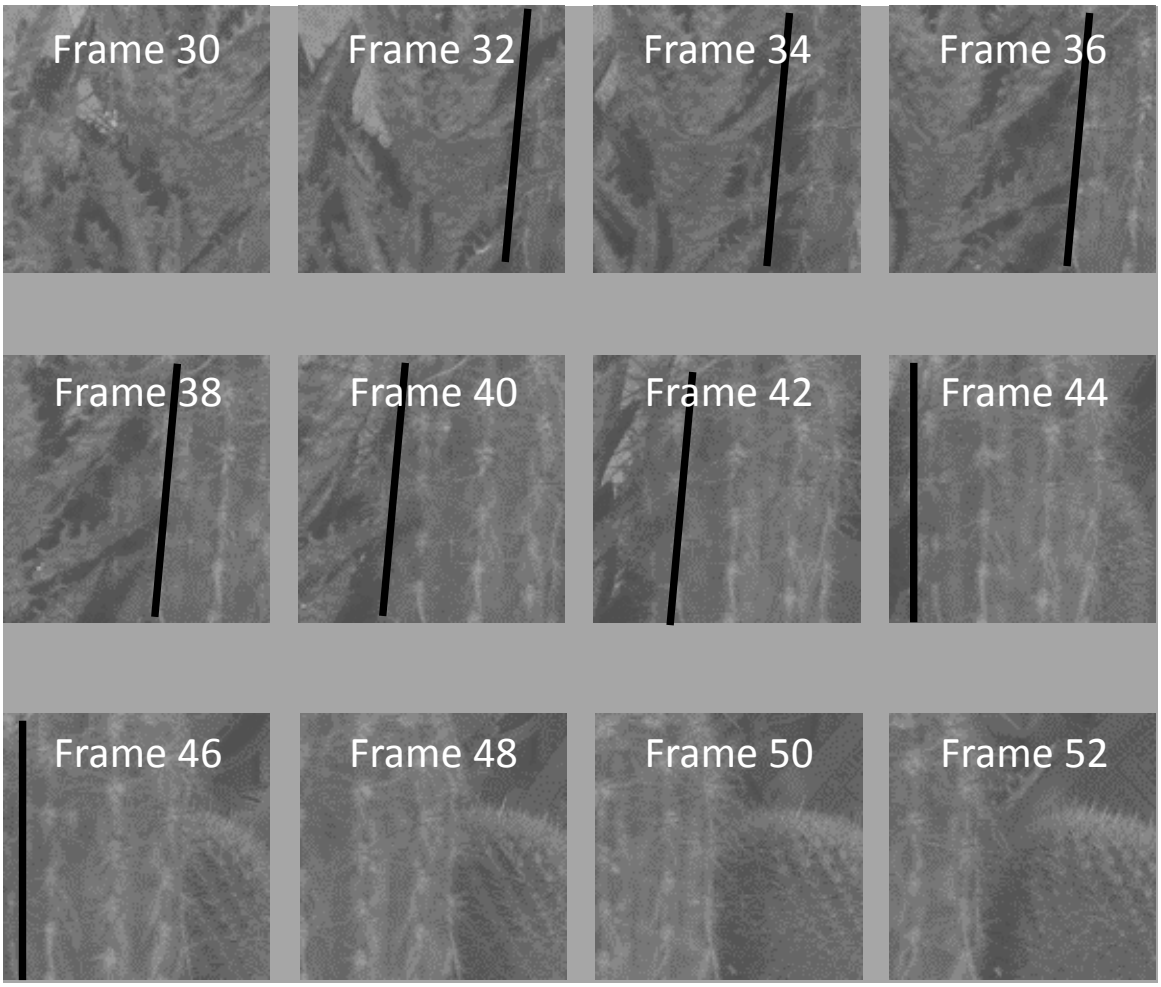

Figure 7.9: Frame 30 through 52 of mask *Cactus* when played back at 300 fps. This image shows a black line on the leading edge of one of the cacti spinning on a pedistal in the scene. Near frame 36 or 38, the leading edge stops moving toward the front of the scene and begins its motion to the left. This is the beginning of its sideways movement which would result in apparent motion velocity. Near frame 46, the leading edge has moved out of view. So in about 10 frames, the leading edge has moved across the scene. Given the playback rate of 120 frames per second, 10 frames will pass in a twelfth of a second. Given the viewing distance of 32 pixels per degree, and the stimulus is 128 pixels wide, the leading edge of the cactus covers nearly 4 degrees of the viewing plane in a twelfth of a second, resulting in an apparent motion velocity of nearly 48 degrees per second.

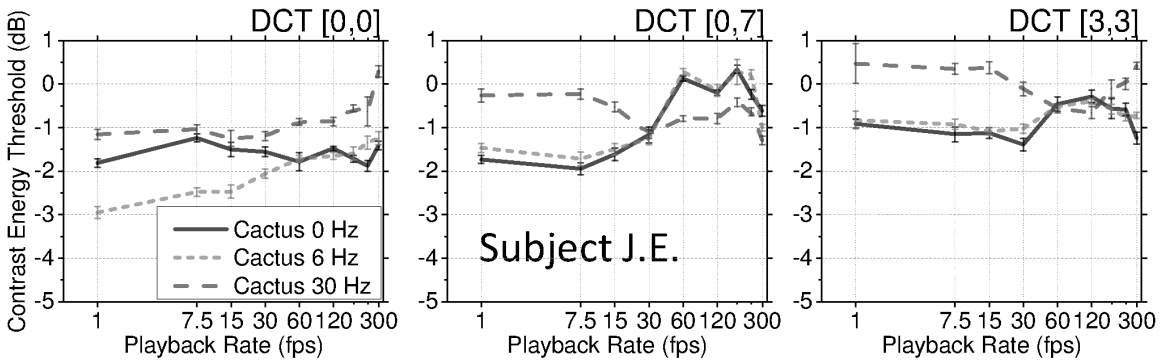

Figure 7.10: Contrast detection thresholds versus mask video playback rate for subject J.E. for higher playback rates than normal viewing conditions.

The mask for this playback rate experiment was adjusted to an average RMS contrast of 0.30. At the same time, the average luminance of this mask was adjusted to 30 cd/m^2. For the masks *Cactus*, thresholds were measured for targets DCT [0,0], [0,7], and [3,3] with a temporal frequency of 0, 6, and 30 Hz. Subject J.E. recorded three sets of 32 trials each for each data point shown in Fig. 7.10. The data for subject J.E. in Fig. 7.6 has been repeated for ease of comparison. Because only one mask was measured for this experiment, three different target temporal frequencies are shown in each plot.

Looking at the plots in Fig. 7.10, no apparent trends are obvious. That is, as the playback rate of the mask increases, there does not appear to be a consistent result. Table 7.8 details some of the measurable information about this data set.

What Table 7.8 shows is how changing the playback rate of the mask changes target detectability. The second column of Table 7.8 shows the how much target detectability decreases when the mask playback rate goes from a stationary image to 300 fps. Averaged across all targets, this causes an increase of 0.43 log units. However, because the plots are not monotonic, the biggest increases do not occur when going from one extreme to the other.

The last column of table 7.8 is the slope of a line of best fit for plot of each target

Table 7.8: Elevations due to change in mask playback rate. This table shows how changing the mask playback rate changes target detectability. The first column lists the target spatial and temporal frequency. The second column lists the elevation due to changing the mask playback rate from 1 fps to 300 fps. That is, the second column is the elevation corresponding to the mask played at 300 fps minus the elevation corresponding to the mask played at 1 fps. The third column is the slope of the least squares estimate for a line of best fit for the detection threshold versus mask playback rate. Because this slope is so small, the slope was multiplied by 1,000 for ease of comparison.

Target	Elevation	Slope ×1000
DCT [0,0] at 0 Hz	0.40	-0.26
DCT [0,0] at 6 Hz	**1.71**	4.60
DCT [0,0] at 30 Hz	1.45	3.95
DCT [0,7] at 0 Hz	1.11	**4.82**
DCT [0,7] at 6 Hz	0.47	4.02
DCT [0,7] at 30 Hz	***-1.03***	***-1.60***
DCT [3,3] at 0 Hz	-0.33	0.74
DCT [3,3] at 6 Hz	0.10	0.88
DCT [3,3] at 30 Hz	-0.05	-0.18
Average	*0.43*	*1.89*

frequency. Because the rise in elevations was small and the change in playback rate was large, the slope of this line was mostly zero for all plots. For ease of comparison, the slope of the best fitting line was multiplied by 1,000. What this last column shows is that, on average over all target frequencies, the plots generally have an upward trend. The largest slope was for the target frequency that also had the largest difference. There are three negative slopes, however, there magnitudes are not as large as some of the positive slopes. About four of the slopes were close to zero. This would suggest for nearly half the target frequencies, changing the playback rate of the mask did not make a significant difference.

Linear regression can also examine the data. Table 7.9 shows the results of fitting two linear regression models to the playback rate manipulated data. The left column of numbers present the fit of a target property only model, while the right column presents the goodness of fit for a model that included the mask playback rate as an input. All inputs were normalized to range from one to two. This allows an apples to apples comparison of the contributions of target spatial and temporal frequencies to target detectability with the contributions of mask playback in predicting target detectability.

Observe from Table 7.9 that the three regressor model uses the target spatiotemporal properties to provide a useful fit of the data, and the regressor coefficients are in line with most of the previous three and four regressor models shown in this section. Table 7.9 also shows the coefficients for the four regressor model of the faster playback rate data. Observe that the target property coefficients are mostly unchanged by the addition of the mask property regressor. However, the additional regressor accounting for mask playback rate does result in some improvement of the model fit of the measured data.

It is also interesting that the coefficient for the mask playback rate regressor is positive and significant in magnitude. This suggests a positive correlation between

Table 7.9: Goodness of fit between measured masked target detectability using playback rate controlled masks and predictions from a three and four input no reference linear regression model. The data modeled was from the use of masks with playback rates of 1, 7.5, 15, 30, 60, 120, 180, 240, and 300 Hz. The left column of numbers details the model fit using only target properties, while the right column represents a model that includes mask playback rate as an input.

		3 input	3+P input
fit	PCC	0.660	0.744
	SROCC	0.611	0.723
	RMSE	0.537	0.478
coefficient	constant	-3.258	-4.111
	TSF	1.114	1.109
	TTF	1.341	1.333
	(TSF x TTF)	-1.019	-1.012
	P		0.639

faster mask playback rates and higher target detectability contrast thresholds. However, examining the plots of the data in Fig. 7.10 and in the second column of Table 7.8, it appears that the significance given to mask playback rate by the modeling effort is not so obvious to see in the data.

Observe from Table 7.9 that the goodness of the model fit for this data set is lower than what has been seen with the slower playback rate set, as shown in Table 7.6, where the three regressor RMSE was 0.46. The high speed data was only collected by one subject, J.E., and for only one mask, *Cactus*. To better understand the relationship between mask playback rate and target detectability contrast thresholds, this data set may need to be expanded to include results from more subjects, as well as additional masks.

There are still noticeable differences in the plots showing the relationship between mask playback rate and target detectability, and the data in these plots do not fall in straight lines. The analysis in this Sect. suggests that the changes in the plots may not be related only to changes in the mask playback rate. One possible explanation for these differences is entropy masking. Watson, Borthwick, and Taylor [51], present the idea of entropy masking. Watson, Borthwick, and Taylor describe this as the amount of unknown in the mask. Watson, Borthwick, and Taylor show that this should be included into calculations on masking. Watson, Borthwick, and Taylor point out that this idea has been shown in the past. Swift and Smith describe a learning process in their experiments that supports the theory of entropy masking [50]. Daly also incorporated this notion of mask learning [153].

For this experiment, subject J.E. would watch each stimulus twice, and decide which mask was presented with dynamic DCT noise. This would happen 32 times for each trial. Each measurement represented three trials. And all of this was repeated for the 9 target frequency combinations. This means for a single threshold, subject J.E. could watch the mask as many as 1,728 times for the 300 fps playback rate. This

may seem like a large number of times to watch the same 0.75 second clip, and should have been ample time to learn the mask at that rate enough to not be surprised by it. However, there were 8 other temporal frequencies explored for the mask *Cactus*. Also, subject J.E. watched the mask *Cactus* at 30 fps as many as 1,728 times for each of the five contrast levels, each of the other 4 luminance levels, as well as all the times for the data from Chap. 4. It is likely that for every one time subject J.E. watched the mask *Cactus* at 300 fps, they watched it at least 10 times at 30 fps. This means that subject had to unlearn the expectations from the 30 fps mask and then learn the 300 fps mask. This, perhaps, is a special case of entropy masking, as described by Watson, Brothwick, and Taylor, Swift and Smith, and Daly, where the subject has learned to expect one action, and is not able to get their mind ready for the different action, no matter how many times they see this. Because learning, relearning, and unlearning occurred throughout the experiment, it is possible that the differences in thresholds only represent how difficult the masks were to learn. The upward slope may simply be accounted for by the fact that subject J.E. was able to learn the slower moving masks faster than they could learn the faster moving masks.

All this being considered, it does not appear that changing the mask playback rate has any consistent effect on target detectability. This is consistent with the findings of the previous subsection, where three subjects were tested. It should be noted that subject J.P. had not participated in any other experiments, and had seen the slow moving masks an equal number of times to the normal moving masks, and the data suggests that the results from subject J.P. were similar to the results from subject J.E. who had seen the mask moving at 30 fps thousands of times for other experiments as well as developing the data collection tool tools. Table 7.8 does not show any clear pattern for elevations, differences, or slopes. Some of these measures are larger than others, but it is not consistent across target frequencies. Generally, increasing the mask playback rate does appear to mostly make targets more difficult

to see.

7.2 Detectability of targets spatially correlated with mask content

As seen in the previous sections, one consistent way to increase target detection thresholds is to change the target properties of spatial and temporal frequency. This Sect. explores the relationships between target detectability and a different target property, target spatial correlation with the mask. Typically in DCT based compression, the artifacts are correlated in space with the image. During compression, after an image is transformed to the DCT domain, some rounding occurs for some DCT coefficients, resulting in slight errors and imperfections when the image is transformed back to the spatial domain. This is usually done in a block wise fashion, and each block has different imperfections or artifacts, based on what spatial content was present in the scene before compression. The resulting artifact or distortion is then spatially correlated with the image being compressed and localized to each block. If a particular DCT band is not present in a particular block, there will be no component to round, and thus, no artifact that looks like that band. However, when a block has a certain DCT band present, then and only then is it possible for an artifact that looks like that DCT band to show up after quantization or rounding. If there is nothing to round, there can be no rounding error. Thus common DCT artifacts or distortions are spatially correlated with the image, and distortions can only appear where that DCT band is present in the image.

The advantage of using uncorrelated dynamic DCT noise for previous studies is that our results could be tied back to the results from Watson, Hu, and McGowan [6]. The disadvantage is that correlated targets are more similar to the compression artifacts seen in many modern video compression algorithms. This section expands our results by examining this more realistic target, and compares our results using uncorrelated dynamic DCT noise as targets with correlated target detectability contrast

thresholds.

Dynamic DCT noise, as described by Watson, Hu, and McGowan [6], as well as in Chapter 3, is evenly distributed over the entire stimulus. In Chapter 3, building dynamic DCT noise began in the DCT domain with all DCT bands for an 8×8 pixel block set to zero. The appropriate DCT band was selected for the spatial frequency of the target, and set to a thousand. Then the spatial DCT block was formed by taking the inverse DCT transform. This 8×8 pixel block then had its amplitude normalized to one. This normalized block was then repeated over the entire 128×128 target image. A random phase offset was added to each block while it was modulated in time with a Gabor function. The random phase offset reduced the likelihood that blocks closer together would change target amplitude in phase with each other. Then a scaling factor changes the amplitude of each target frame to ensure the desired contrast energy for each trial in the experiment.

The targets for this subsection are slightly modified to make the targets spatially correlated with the spatial content of the mask. First, the mask frame is separated into 8×8 pixel blocks. Then each block is transformed to the DCT domain. The appropriate DCT band for the selected spatial frequency of the target is set to 0. The inverse DCT transform of this block produces the difference block. The original 8×8 pixel block is then subtracted from the difference block to produce the target block. Then the target blocks are stitched back together to from the entire 128×128 target image. This process is repeated for each frame of the mask to form a unique set of target frames for each mask. After this point, the rest of the process is the same as was used for the previous dynamic DCT noise. A random phase offset is added to each block, and the amplitude of the block is controlled with a Gabor function. A scaling factor stored in a look up table ensures that at each trial the target amplitude is correct to get the desired contrast energy.

The result of this change is that the distortions will be correlated with the spatial

content of the mask. For example, in the mask *Timelapse*, where most of the sky has no texture, resulting in nothing in certain DCT bands to round, and thus no distortions will appear in those areas of the sky. However, along the leading edge of the clouds in the mask *Timelapse*, most of the distortions will appear in that localized area. In one way, the correlated dynamic DCT noise is more difficult to see because the artifacts only appear in areas of the mask that contain the same spatial frequency. However, the correlated target can be easier to see because now there is less area of the patch where the artifact can show up. Fig. 7.11 shows a few sample frames from the mask *Timelapse*, as well as an example corresponding correlated target and the combined stimulus consisting of both the mask and the correlated target.

Fig. 7.11 shows how different the new targets are from the previously used dynamic DCT noise. In Chapter 3, Fig. 3.1c showed the original dynamic DCT noise for the target DCT [3,3] at 0 Hz. The difference between these two targets is that in Chapter 3, Fig. 3.1c, the dynamic DCT noise is equally likely to appear in all 8×8 pixel blocks, but in Fig. 7.11, the noise is more likely to show up in blocks that have noticeable content. Areas with more noticeable features, such as edges in Fig. 7.11 tend to have more pronounced artifacts, while smooth regions tend to be left undistorted. However, as with many other examples in this paper, correlated targets are also something that changes from mask to mask. Fig. 7.12 shows a few frames with correlated targets for mask *Cactus*.

In Fig. 7.12, the targets are more evenly distributed. The mask *Cactus* has more going on all over the scene, and thus more blocks of content that would have DCT [3,3] band content that can be rounded off to generate artifacts. Note that the targets in Fig. 7.12 look more like the target from Chapter 3 Fig. 3.1c.

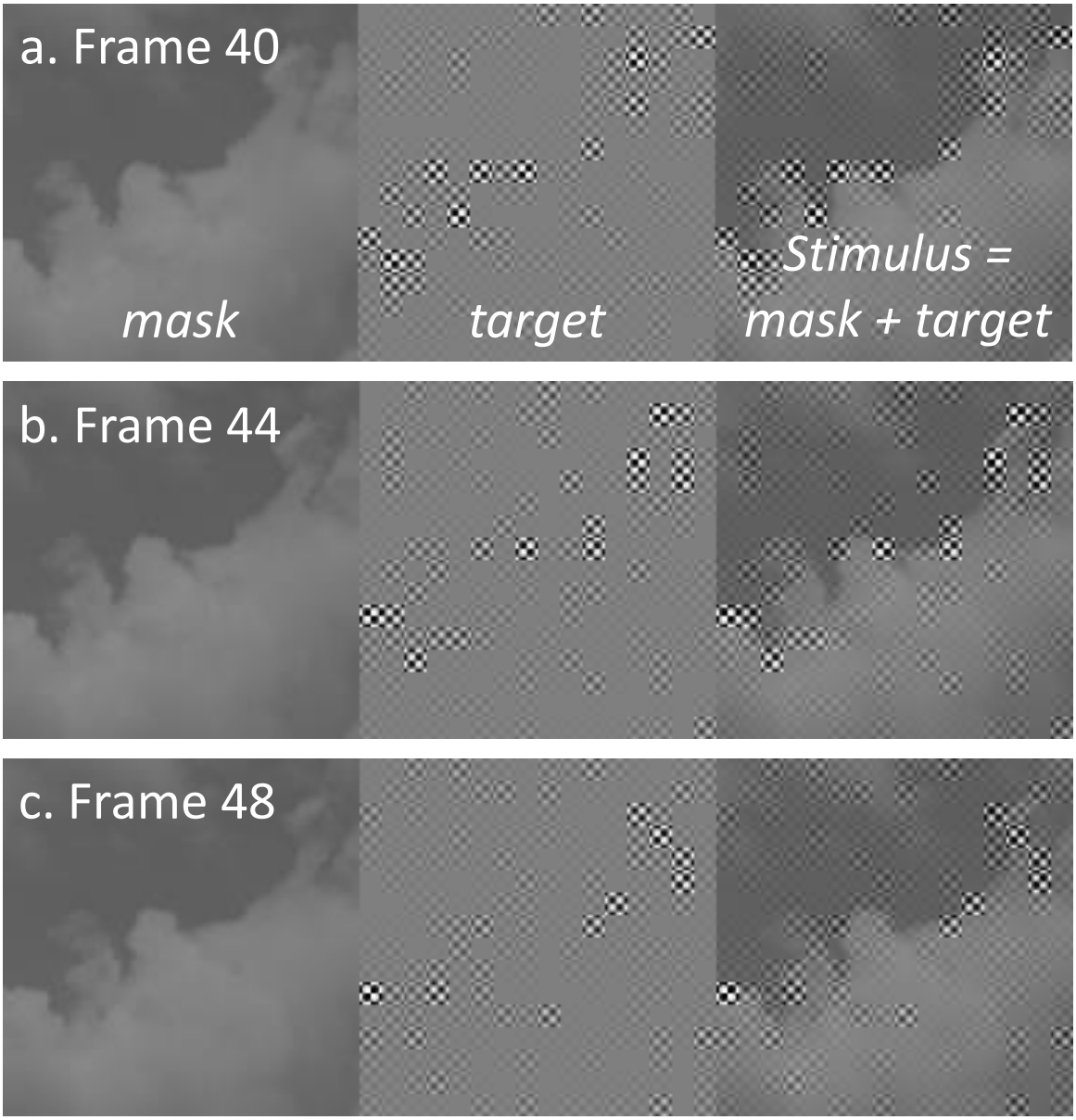

Figure 7.11: Frames from the mask *Timelapse*, along with example correlated targets presented both in the unmasked condition, as well as with the mask. This figure provides an example of correlated target frames. The target shown is for DCT [3,3] at a temporal frequency of 0 Hz. Moving from panel a through c, the figure shows how the target changes to match the spatial content of the mask. Note in this figure that the targets now only appear in the areas that have clouds in the mask *Timelapse*.

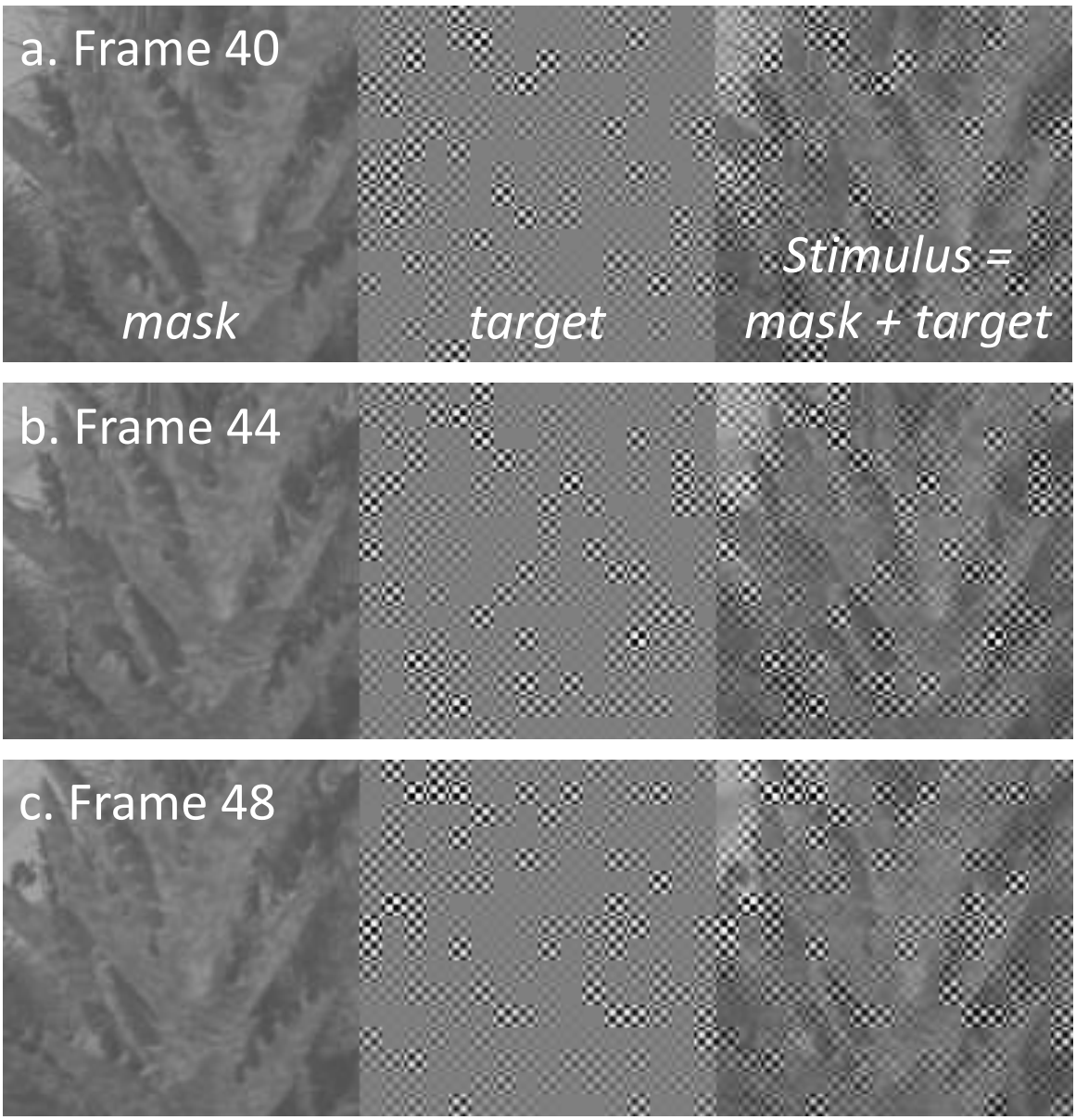

Figure 7.12: Frames from the mask *Cactus*, along with example correlated targets presented both in the unmasked condition, as well as with the mask. This figure provides an example of correlated target frames. This figure is similar to Fig. 7.11, except that this figure employs the mask *Cactus*. The target shown is for DCT [3,3] at a temporal frequency of 0 Hz. Moving from panel a through c, the figure shows how the target changes to match the spatial content of the mask. Note in this figure that the targets now appear in more areas, and are most pronounced in the areas that have spines in the mask *Cactus*.

7.2.1 Detectability of unmasked targets spatially correlated with mask content

One of the first questions relating a change in the target is how unmasked target detectability contrast thresholds change because the difference in targets. As Fig. 7.11 and Fig. 7.12 show, the differences are unique to each mask. An experiment was conducted to quantify how much of a difference in target detectability can be attributed to a change in targets. This experiment was designed to measure the detectability of this new target in an unmasked condition. A set of correlated targets was presented against the gray background used for the unmasked condition to examine how different detection thresholds are for the new target. To generate these unmasked stimuli, a set of target images was created that was correlated with masks *Typing*, *Timelapse*, and *Cactus*, and then presented against the gray background used for the unmasked conditions. The results of this experiment are shown in Fig. 7.13 and Fig. 7.14. Fig. 7.13 shows the results for subject J.E. Fig. 7.14 shows the results for subject K.J.

For this measurement, subject J.E. completed three sets of 32 trials each for each threshold. Subject K.J. completed two sets of 32 trials each for each threshold. Because this is a target spatial property, additional target spatial frequencies were explored.

Fig. 7.13 and Fig. 7.14 show good agreement between subjects. The error bars in Fig. 7.13 and Fig. 7.14 are reasonable. Fig. 7.13 and Fig. 7.14 also show good agreement between the two types of targets, when both are presented in the unmasked condition. The general shapes and trends of the plots of the new correlated targets match those of the previous uncorrelated targets in the unmasked condition. The plots of the correlated targets have slightly different threshold elevations than the uncorrelated targets for some target spatiotemporal frequencies, but not for all.

The unmasked target detectability contrast threshold differences between uncorrelated and correlated targets is quantifiable. Table 7.10 presents fitness scores between

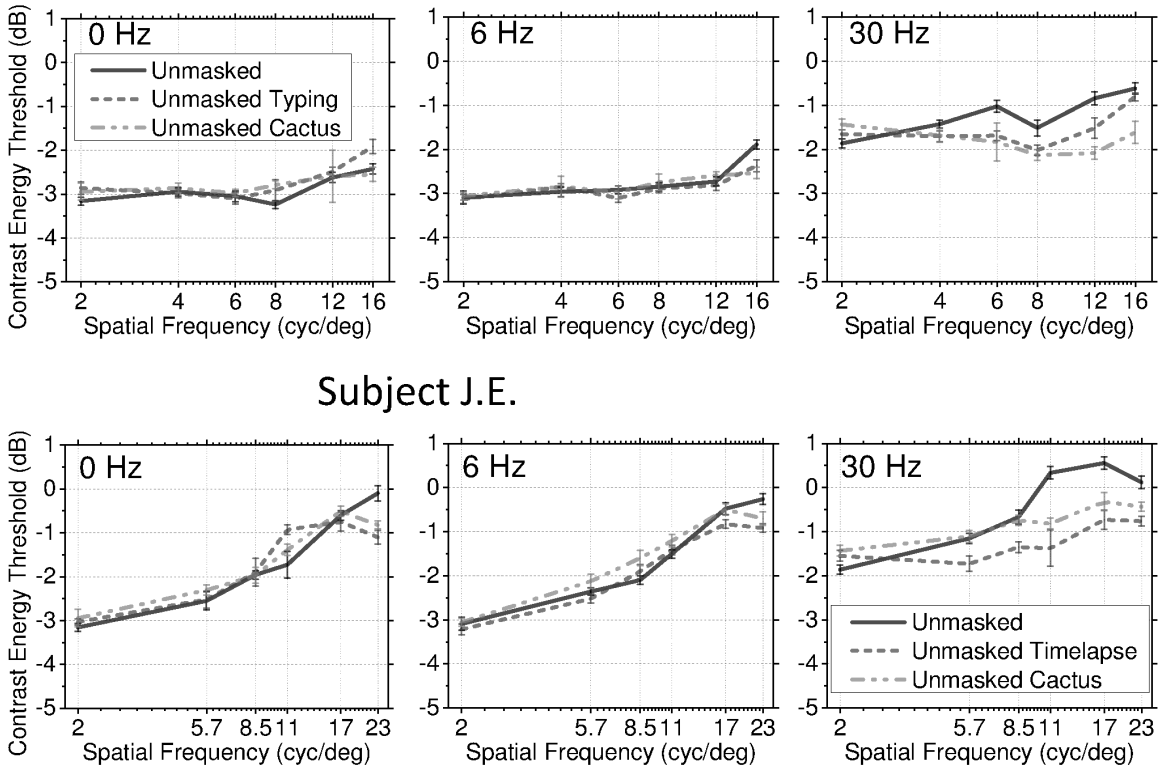

Figure 7.13: Detection contrast thresholds versus target spatial frequency for subject J.E. for spatially correlated unmasked targets. This plot shows that changing the target from uncorrelated to spatially correlated did not significantly change the general trends seen in unmasked target detection.

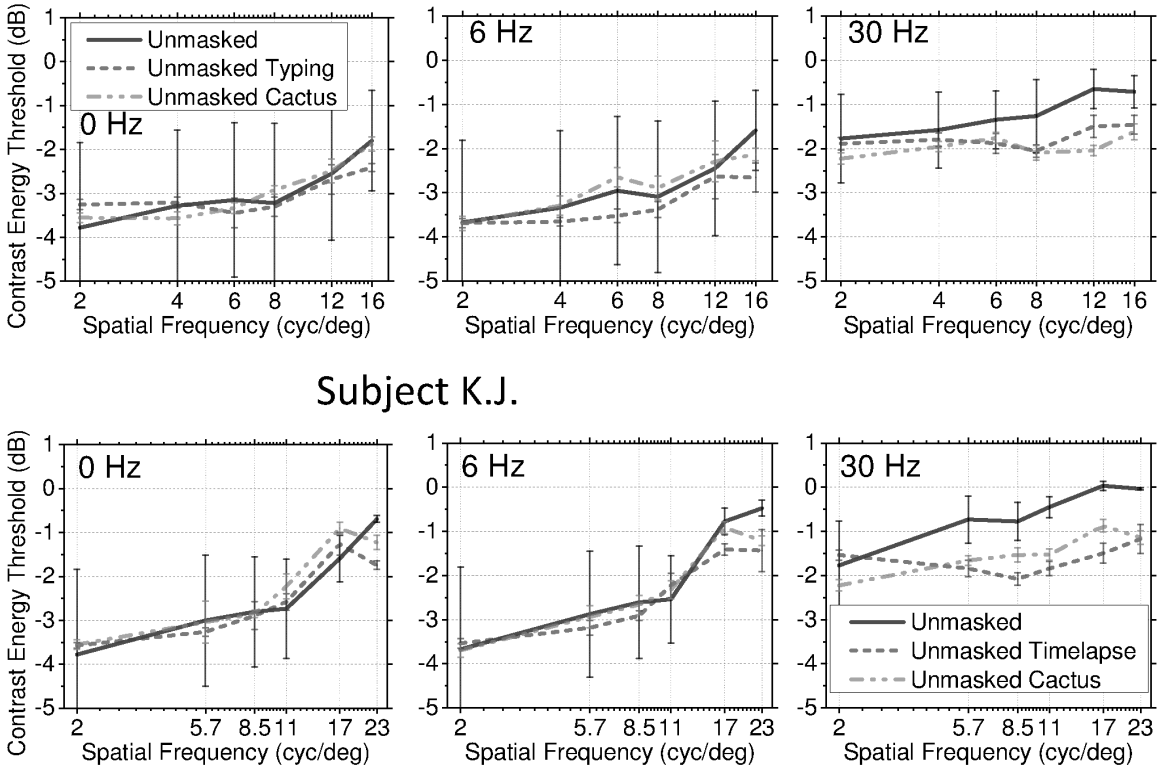

Figure 7.14: Detection contrast thresholds versus spatial frequency for subject K.J. This plot shows that changing the target did not significantly change the general trends seen in unmasked target detection.

the unmasked correlated and unmasked uncorrelated target detectability contrast thresholds. The PCC reported was a linear Pearson correlation coefficient calculation after a logistic fitting. This was based on the work by N.D. Narvekar and L. J. Karam, CPBD Sharpness Metric Software," http://ivulab.asu.edu/Quality/CPBD. For both PCC and SROCC, the best possible score is 1. Root of the mean squared errors (RMSE) provide a summary of the differences between groups of data. For RMSE, the best possible score is 0.

Table 7.10: Correlation between unmasked correlated and unmasked uncorrelated target detectability contrast thresholds. The data was broken into targets with vertical and diagonal alignments. For the vertically aligned targets, the masks *Cactus* and *Typing* were used as templates for the targets. For the diagonally aligned targets, the masks *Cactus* and *Timelapse* were used as templates for the targets. Elevation was the average of correlated target detectability contrast thresholds minus the uncorrelated target detectability contrast thresholds.

Unmasked Vertical			Unmasked Diagonal		
	Cactus	Typing		Cactus	Timelapse
PCC	0.974	0.982	PCC	0.926	0.945
SROCC	0.899	0.858	SROCC	0.889	0.913
RMSE	0.211	0.179	RMSE	0.406	0.353
Elev.	-0.503±0.532	-0.445±0.329	Elev.	-0.250±0.493	-0.459±0.424

Observe from Table 7.10 that there is a strong correlation between the uncorrelated and correlated target detectability contrast thresholds. The goodness of fit scores show the match between the two types of unmasked targets is reasonable. The correlation coefficients are similar to the correlations shown between subjects. It is interesting to note from the elevations in Table 7.10 that the correlated targets generally had higher detectability contrast thresholds. It is also important to note that the

elevations due to changing the target from uncorrelated to correlated were different between the two target spatial orientations, as well as the two different masks the targets were correlated with.

The data reported in this Sect. suggests that the change in the target, from being evenly distributed to being present only where the masks have spatial content at the target spatial frequency, had little effect on target detectability contrast thresholds. The spatially correlated targets tend to have slightly higher contrast detectability contrast thresholds, however the differences are small. The differences between correlated and uncorrelated target detectability contrast thresholds do appear to be dependent on both the target spatial orientation as well as which mask the targets are correlated with.

7.2.2 Detectability of masked targets spatially correlated with mask content

This section examines how spatial correlation changes target detectability for masked targets. The previous subsection showed that, in general, correlated targets are easier to see than uncorrelated targets. The differences between target visibilities were small, but appear to dependent on which mask targets are correlated with, as well as the target spatial orientation. Figures 7.15 and 7.16 provide plots of experiment results, showing correlated target detectability contrast thresholds versus target spatial frequencies.

Observe in Fig.s 7.15 and 7.16 that the two subjects appear to provide consistent results. In general, the correlated target detectability contrast thresholds appear to be similar to uncorrelated target detectability contrast thresholds. The error bars from both subjects appear to be reasonable. Note that, as in Section 5.7, targets presented with the mask *Cactus* had significantly higher target detectability contrast thresholds than unmasked targets.

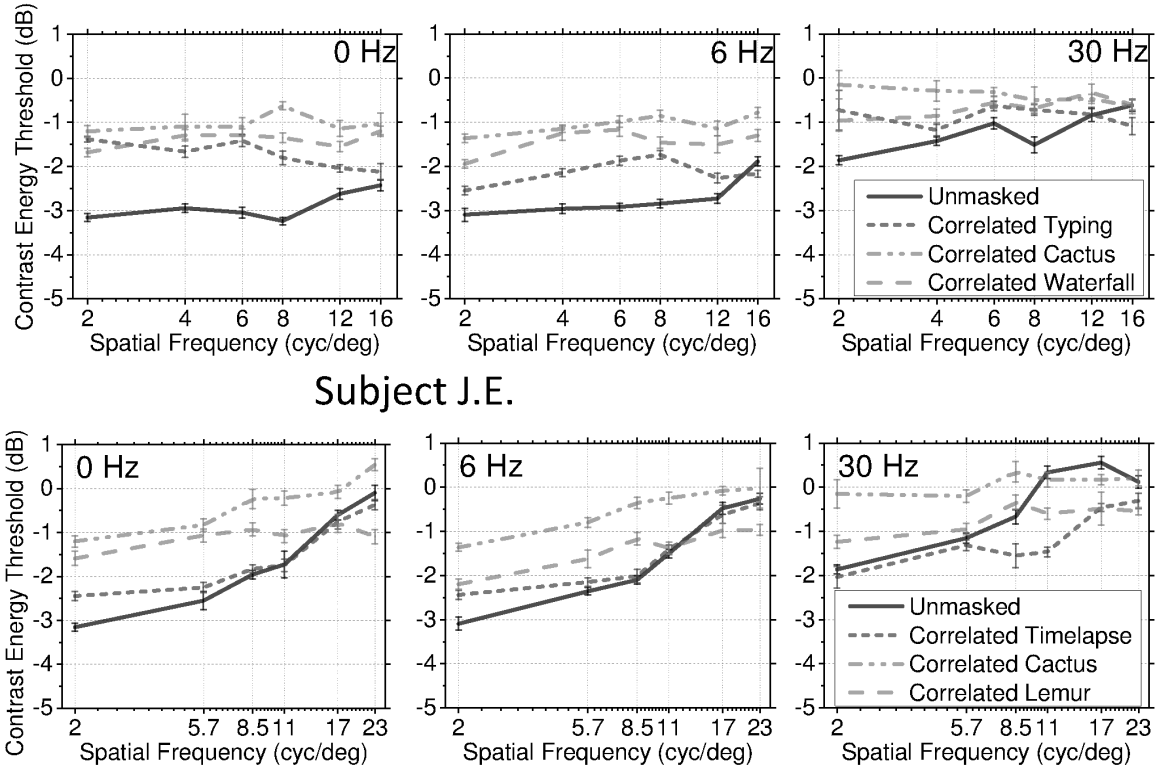

Figure 7.15: Detection contrast thresholds versus target spatial frequency for subject J.E. This shows masked detection thresholds for correlated targets.

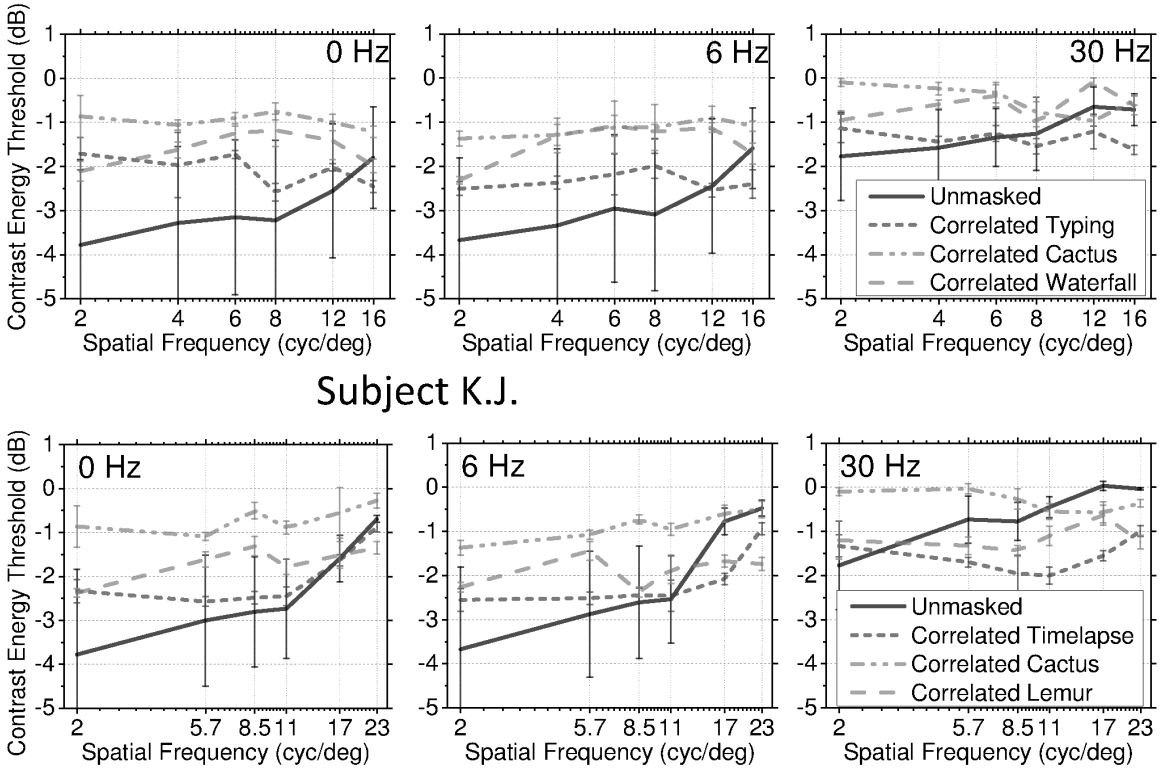

Figure 7.16: Detection contrast thresholds versus target spatial frequency for subject K.J. This shows masked detection thresholds for correlated targets.

One question is how similar the results from correlated targets are to results from uncorrelated targets. Table 7.11 presents fitness scores between the two datasets. Scores are presented for the entire data set, as well as for individual masking conditions.

Table 7.11: Similarities between masked target detectability contrast thresholds for correlated and uncorrelated targets. Elevation is the average of the correlated target detectability target contrast thresholds minus the uncorrelated target detectability contrast thresholds.

	PCC	SROCC	RMSE	elevation (±)	
Overall	0.833	0.830	0.451	0.147	(0.460)
Cactus	0.795	0.729	0.338	0.027	(0.360)
Lemur	0.722	0.686	0.321	0.252	(0.373)
Timelapse	0.950	0.928	0.251	-0.039	(0.335)
Typing	0.887	0.787	0.372	0.102	(0.588)
Waterfall	0.892	0.834	0.280	0.492	(0.503)

Observe in Table 7.11 that the correlated and uncorrelated target detectability contrast thresholds have a strong similarity. Note from Table 7.11 that the mask *Timelapse* showed the strongest similarity between the two target types. Recall from Fig. 5.3 in Sect. 5.7 that presenting uncorrelated targets with the mask *Timelapse* resulted in target detectability contrast thresholds similar to unmasked uncorrelated target detectability contrast thresholds.

Figure 5.3 in Sect. 5.7 showed that the mask *Cactus* had a stronger influence on masked uncorrelated target detectability contrast thresholds. Table 7.11 shows the second lowest PCC and SROCC scores between correlated and uncorrelated target detectability was for targets presented with the mask *Cactus*. This may suggest that the difference in masked target detectability between correlated and uncorrelated

targets is proportional to the ability of the mask to effect target detectability.

The lowest PCC and SROCC scores between correlated and uncorrelated target detectability contrast thresholds came from targets presented with the mask *Lemur*. Figure 5.3 in Sect. 5.7 did not suggest that the mask *Lemur* had any outstanding ability to effect target detectability contrast thresholds. This may suggest that the difference in masked target detectability between correlated and uncorrelated targets is not proportional to the ability of the mask to effect target detectability. However, it is clear that the differences between uncorrelated and correlated masked target detectability contrast thresholds is mask dependent.

The data in Table 7.11 suggests there is significant similarity between the masked uncorrelated and correlated target detectability contrast thresholds. The elevation of the uncorrelated target detectability contrast thresholds over the correlated target detectability contrast thresholds was 0.147 ± 0.460. This suggests that on average, the masked correlated targets had slightly lower target detectability contrast thresholds, however, that difference was usually small. The data in Table 7.11 also suggests the target detectability contrast threshold differences between correlated and uncorrelated targets varies from mask to mask.

7.2.3 Discussion of targets spatially correlated with mask content

The results of this section present a new data point to help map the landscape of video compression artifact detectability understanding. From previous sections, unmasked uncorrelated target detectability contrast thresholds are similar to other unmasked target detectability contrast thresholds for more controlled targets which are more common in visual Psychophysics. Presenting natural videos with uncorrelated targets resulted in target detectability contrast thresholds that were reasonable extrapolations from the unmasked data, and masked target detectability contrast thresholds were generally higher than unmasked target detectability contrast thresholds. The

changes in uncorrelated target detectability contrast thresholds varied from mask to mask, and were dependent on target spatio-temporal frequencies.

From this section, unmasked correlated target detectability contrast thresholds are mostly similar to unmasked uncorrelated target detectability contrast thresholds, and the differences appear to be dependent on both which mask the targets are correlated with, as well as the target spatial properties. In general, correlated targets had lower target detectability contrast thresholds than uncorrelated targets. When correlated targets were presented with natural video masks, target detectability contrast thresholds showed strong similarity with masked uncorrelated target detectability contrast thresholds. In general, when targets are presented with natural video masks, correlated target detectability contrast thresholds tended to be slightly lower than uncorrelated target detectability contrast thresholds. However, the differences appear to vary from mask to mask.

The modeling results in Sect. 5.2 showed that a simple model could capture most of the variations in unmasked target detectability contrast thresholds due to changes in target spatial and temporal frequencies. In that Sect., in the same Table 5.2, it was shown that using only target spatial and temporal frequencies to predict masked target detectability contrast thresholds was not nearly as effective, but still contributed useful information. Now, in this Sect., it is shown that changing the target from uncorrelated to correlated does not result in a large change in target detectability contrast thresholds, but the changes in detectability vary from mask to mask.

There are two next logical extensions of the current work. One extension would be to move towards compression artifacts with multiple target spatial frequencies. Many modern compression algorithms compress multiple spatial frequencies at different levels. Another reasonable extension would be to measure the detectability of artifacts due to prediction error. Based on the results from this chapter, it is reasonable to

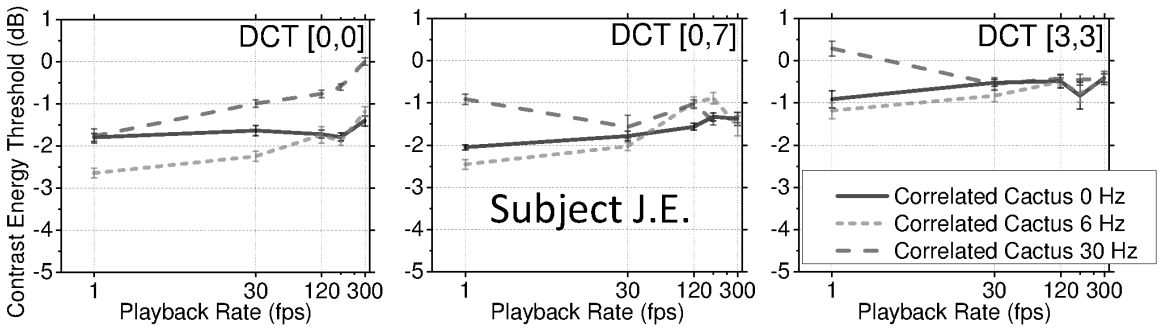

Figure 7.17: Detection thresholds versus mask video playback rate for subject J.E. for higher playback rates than normal viewing conditions, using a correlated target.

assume that the results from both areas of research can be summarized as follows: Changing the target to be more like artifacts experienced by the common consumer produced results similar to those seen with more controlled targets, however, some differences are significant, and the level of difference varies depending on the mask content and the target spatio-temporal properties.

7.3 Detectability of masked targets spatially correlated with mask content at higher mask playback rates

In Sect. 7.1, it was shown that controlling mask playback rates had some effect on target detectability contrast thresholds. In Sect. 7.2, it was shown that targets correlated with mask content are slightly different than uncorrelated targets. This section examines the relationship between mask playback rates and correlated target detectability contrast thresholds. Subject J.E. completed three sets of trials measuring correlated target detectability contrast thresholds for targets presented with masks with playback rates from 1 to 300 frames per second. Fig. 7.17 shows the results of this experiment.

Observe from Fig. 7.17 that as target spatial frequencies increased, target detectability contrast thresholds increased. This is in line with previous observations from this chapter. Similarly, as target temporal frequencies increased, target de-

tectability contrast thresholds increased. Although changing target temporal frequencies from 0 Hz to 6 Hz resulted in little difference, target temporal frequencies of 30 Hz were generally associated with higher target detectability contrast thresholds.

As in Sect. 7.1.5, a linear regression model was formed around the data. Model coefficients were found using only target properties as regressors, as well as including mask playback rate as a regressor. Table 7.12 presents measures of model goodness of fit as well as model coefficients.

Table 7.12: Model fit and coefficients of data from the experiment using masks controlled for playback rate and correlated targets. The left column of numbers details the model fit using only target spatial and temporal properties, while the right column represents a model that includes mask playback rate as an input.

		3 input	3+P input
fit	PCC	0.457	0.569
	SROCC	0.422	0.557
	RMSE	0.568	0.525
coefficient	constant	-2.267	-3.114
	TSF	0.437	0.442
	TTF	1.136	1.145
	(TSF x TTF)	-0.884	-0.896
	P		0.590

Observe from Table 7.12 that the correlation between model predictions and measured thresholds is poor. This suggests that conclusions based on this data are merely speculative. The coefficients for the three and four regressor models are similar to those listed in Table 7.9, as well as other models listed in the dissertation. The coefficient for target spatial frequency is smaller than the coefficient for target temporal frequency. This may suggest that for playback controlled masks and correlated tar-

gets, target temporal frequency now has a larger role in predicting target detectability contrast thresholds. The data in Table 7.12 suggests that including mask playback rate as a regressor improves the correlation between model predictions and measured target detectability contrast thresholds. However, even the four regressor model does not provide a prediction with strong correlation with measured thresholds.

The results from Table 7.12 can be compared to the results from Table 7.9. The plots shown in Fig.s 7.10 and 7.17 may not suggest a clearly monotonic relationship between target detectability contrast thresholds and mask playback rate. The correlations between model predictions and measured data reported in Tables 7.9 and 7.12 may not suggest the models perfectly reflect the variations in target detectability contrast thresholds due to changes in mask playback rates.

The data from this section, as well as from Sect. 7.1 suggest that increasing mask playback rate will generally result in increased target detectability contrast thresholds. Neither the results in Table 7.9 or 7.12 have such strong correlations with measured data as to suggest either model is perfectly capturing relationships between mask playback rates and target detectability contrast thresholds. However, both models appear to suggest that mask playback rate can effect target detectability contrast thresholds. The noise of the data plotted in Fig.s 7.10 and 7.17 do not appear to make this point as clearly as the model coefficients.

Finally, changing the target type, from uncorrelated to correlated, does not appear to have a significant influence on the relationship between target detectability contrast thresholds and mask playback rate. Although the size of the data sets are small, and the clarity of their meanings is limited, neither set suggests that mask playback rate can be eliminated from a list of factors that effects target detectability contrast thresholds. The data does appear to support further examination of this topic.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

This chapter presents the conclusions supported by our work. This chapter also lists promising future work based on our research. The chapter ends with a summary of results from our research.

8.1 Conclusions

We extended the work of Watson, Hu, and McGowan [6] by investigating the masking of dynamic DCT distortions by natural videos. We measured the log_{10} of the contrast energy of detectability thresholds for compression like artifacts that ranged from 0-30 Hz in temporal frequency and from 2.8-22.6 cyc/deg in spatial frequency. Target detectability contrast thresholds were measured in the unmasked condition as well as masked by eight gray-scale videos, 0.75 seconds long. Later, a subset of those videos were modified in luminance, contrast, and playback rate, and target detectability contrast thresholds were measured again. Additionally, the target was later modified to be spatially correlated with mask content.

The conclusions from our research are as follows:

1. Masking targets with natural videos can impact target detectability. For targets with low unmasked detectability thresholds, natural video masking is more likely to elevate target detectability thresholds; however, the amount of elevation is dependent on mask content. For targets with high unmasked detectability thresholds, especially due to high target spatial frequencies, natural video masking is most likely to have little influence on target detectability thresholds;

however, some mask content was associated with facilitation for targets with higher spatial frequencies, and the amount of facilitation is dependent on mask content.

2. Target spatial frequency has an important role in determining target detectability. Increasing target spatial frequency tends to increase target detectability thresholds significantly. The elevations in target detectability due to increasing target spatial frequency can also be significantly reduced by changing the target masking condition or the target temporal frequency. Targets with the highest spatial frequencies examined seem to be more susceptible to facilitation.
3. Target temporal frequency has an important role in determining target detectability. Increasing target temporal frequency tends to increase target detectability thresholds. The elevations in target detectability due to increasing target temporal frequency can be significantly reduced by changing the masking condition or the target spatial frequency. In general, a target with a temporal frequency of 30 Hz is going to have a higher detectability threshold than a target with a temporal frequency of 0 Hz.
4. Target spatial and temporal frequencies can be used to predict most variation in unmasked target detectability contrast thresholds and most variation in natural video masked target detectability contrast thresholds. The addition of video content measurements as model inputs can significantly improve predictions of target detectability contrast thresholds. No reference models tuned to predict natural video masked dynamic DCT noise target detectability thresholds predict variations in masked target detectability better than more general full reference quality assessment algorithms tuned to provide general quality assessment scores.
5. Some properties of natural-video masks can influence masked target detectability thresholds. The level of influence is dependent on target spatial and tempo-

ral frequencies, as well as mask content. Increasing mask luminance appears to cause a slight decrease in masked target detection thresholds. Increasing mask playback rate appears to cause a slight increase in masked target detection thresholds. Increasing mask contrast appears to cause a considerable increase in masked target detection thresholds.

6. Changing the target, from dynamic DCT noise evenly distributed over the entire frame to spatially correlated dynamic DCT noise, only present in regions of the mask that contained spatial content at the frequency of the target, resulted in some significant changes in masked target detectability thresholds; however, changes in target detectability varied from natural video to natural video.

8.2 Future research

In order to support the complicated field video compression and the messy world it captures, additional research is required. We feel there are several other experiments necessary to tie these data more closely to modern video viewing experiences. Other extensions of this work would to be to measure: target sensitivity above detectability, at the supra-threshold level; the interaction of multiple target spatial frequencies in a summation study; the relationship between block size and target detectability; and the interaction of color masks and color targets.

Professor Le Callet of the IRCCyN lab with Polytech'Nantes of the University de Nantes suggested another possible direction for this research. Professor Le Callet suggested that the current data in this paper was a necessary first step for other researchers to have. [146] Professor Le Callet suggests the type of data collected so far only considers part of the artifacts possible during compression. Another piece of the human vision and video compression puzzle is motion prediction. Professor Le Callet suggested the measurement of the detectability thresholds of errors due to incorrect motion prediction, as this is a key part of modern video compression.

This data, as well as the data presented by Robson [8] and Watson, Hu, and McGowan [6], have shown that target temporal frequencies have played an important role in effecting target detectability contrast thresholds. Although artifact temporal frequency is not commonly controlled in video compression, this could be a valuable tool for either improving compressed video fidelity, or easing aggravations related to more aggressive compression rates. The maximum target temporal frequency displayable is limited by the refresh rate of the display, and the maximum target temporal frequency precipitable is limited by the mechanics of the eye. In the future we hope to quantify how these limits relate to the ability to mask targets with natural video masks. Tables 5.3 and 5.4 show that large changes in target temporal frequency cause significant changes in masked target detectability contrast thresholds. Table 5.6 also shows that small changes in target temporal frequency make masked target detectability contrast thresholds lower half the time. We plan to measure what refresh rates are necessary to make target temporal frequencies high enough to maximize target detectability contrast thresholds when targets are masked by natural video masks. As with all applications of research, to bear fruit, such findings would then need to be implemented in the real world, which is often messy and complicated. The question would then become if the cost of necessary changes in video compression technology would be worth the benefit.

There are many approaches to data modeling, including functional models, biologically inspired models, and physiologically plausible models. This is a reflection of the level of effort that necessary to provide the proper models used in the many different areas related to human vision and media processing. The functional model provided in Chapter 6 is only a starting point down this path. This model uses the regressors to best predict masked target detectability contrast thresholds, and does not always use the regressors in the most intuitive manner.

After gathering additional data, modeling efforts should be revisited. It appears

that the measure of spatial standard deviation may be of significance. Further examination of masked target detectability may benefit from experiments that control mask standard deviation, which would provide more direct information about the relationships between mask content, target spatial and temporal frequency, and target detectability. Additionally, the data from Chapter 6 appears to suggest that the question of how to collapse video content measurements into single scores may also merit closer examination. These additional data should be most useful in understanding target detectability contrast thresholds for normal video viewing and compression research.

8.3 Summary of results

The following list details our main results from our data analysis of our main data set:

1. Changing from the CRT monitor used by Watson, Hu, and McGowan [6] to a LCD monitor did not change the trends observed with dynamic DCT noise targets presented in the unmasked condition.
2. Unmasked targets higher in spatial and temporal frequencies had higher detectability contrast thresholds, as suggested by the results from Robson [8] and Watson, Hu, and McGowan [6].
    (a) Large changes in unmasked target spatial frequencies resulted in large detectability elevations when target temporal frequencies were small.
    (b) Large changes in unmasked target spatial frequencies resulted in reduced detectability elevations when target temporal frequencies were near 30 Hz.
    (c) Large changes in unmasked target temporal frequencies resulted in large detectability elevations when target spatial frequencies were small.
    (d) Large changes in unmasked target temporal frequencies resulted in reduced

detectability elevations when target spatial frequencies were large.

3. Masked target detectability trends had reasonable similarity to unmasked target detectability trends, however:
   (a) Presenting targets with natural video masks can reduce or eliminate elevations due to large changes in target spatial or temporal frequencies.
   (b) Different masks caused unique changes in relationships between target detectability and target spatiotemporal properties at different target spatial and temporal frequencies.
   (c) Large increases in target spatial frequencies can sometimes result in negative elevations in target threshold detectability for some masked targets.
   (d) Smaller increases in target spatial frequencies are more likely to result in negative elevations for masked target detectability.
   (e) Large increases in target temporal frequencies can sometimes result in little to no elevation in target threshold detectability for some masked targets.
   (f) Smaller increases in target temporal frequencies are more likely to result in negative elevations for masked target detectability.
4. Natural video masks appear to be most effective in reducing detectability thresholds for targets with lower spatial frequencies.
5. Natural video masks were also effective in reducing detectability thresholds for targets with lower temporal frequencies.
6. All natural video masks examined were capable of producing facilitation, however:
   (a) Some masks caused significant elevations in target detectability thresholds most of the time, and rarely resulted in facilitation.
   (b) Some masks were most likely to cause facilitation, and rarely caused significant elevations in target detectability thresholds.

The following list details our main results from our modeling efforts using our main data set:

1. A no reference linear model with only two regressors was able to predict unmasked target detectability thresholds with a PCC of 0.961.
2. Including a third regressor, based on observations from our data, as well as from Robson [8] only resulted in a PCC increase of 0.003 for unmasked thresholds.
3. The two regressor model to was able to predict masked target detectability thresholds with a PCC of 0.684, and the third regressor improved PCC by 0.006.
4. Including a fourth regressor measuring mask spatial standard deviation, averaged over all frames, improved the prediction PCC from 0.690 to 0.818.
5. The normalized model coefficients showed that mask spatial standard deviation was possibly only a best candidate for a data fitting exercise, and not a intuitive indicator of natural video masking ability.
6. A model with twelve regressors measuring mask content had a prediction to measured threshold PCC increase of 0.102 over a model with only one mask measurement regressor.
7. A seven regressor model provided an acceptable prediction of all masked target detectability thresholds that was on par with how well one subject could produce the results of another subject during data collection.
8. Full reference models were less effective in predicting masked target detectability contrast thresholds than the no reference linear models.

The following list details our main results from our expanded investigations using natural video masks controlled for luminance, contrast, and playback rate:

1. Our data appears to suggest that mask luminance has a limited influence over target detectability thresholds, and that a significant increase in mask luminance

is correlated with only a slight decrease in target detectability thresholds. Although modeling coefficients predicting target detectability thresholds based on target spatial and temporal frequencies and mask luminance suggest a weak relationship, plots of detectability thresholds versus mask luminance suggest this relationship is more difficult to discern.

2. Our data appears to suggest that mask playback rate has a limited influence over target detectability thresholds, and that a significant increase in mask playback rate is correlated with only a slight increase in target detectability thresholds. Although modeling coefficients predicting target detectability thresholds based on target spatial and temporal frequencies and mask playback rate suggest a weak relationship, plots of detectability thresholds versus mask playback rate suggest this relationship is questionable. Because of the poor correlation between mask playback rates and target detectability thresholds, data from additional subjects may be necessary to quantify this relationship more clearly. From our data, it appears that detectability thresholds for targets with lower spatial frequencies can be increased by increasing mask playback rate; however, thresholds for targets with higher spatial frequencies are more likely to have lower detectability thresholds for masks with higher playback rates.
3. Our data suggests that mask contrast has considerable influence over target detectability thresholds, and that a significant increase in mask contrast is correlated with an increase in target detectability thresholds. It appears that detectability thresholds for targets with lower spatial frequencies can be increased significantly by increasing mask contrast. Thresholds for targets with higher spatial frequencies are less likely to have higher detectability thresholds for masks with higher contrast.
4. Our data appears to suggest that changing the target from dynamic DCT noise evenly distributed over the entire frame to spatially correlated dynamic DCT

noise, only present in regions of the mask that contained spatial content at the frequency of the target, resulted in some significant changes in target detectability thresholds for masked targets. Correlated unmasked targets had similar detectability thresholds to uncorrelated unmasked targets, however correlated targets tended to have slightly higher detectability thresholds. The masked target detectability of correlated targets was slightly different from masked uncorrelated targets. Masked correlated targets tended to have slightly lower detectability thresholds, however the elevations due to changing target correlation varied from natural video to natural video.

5. Our data appears to suggest that when targets are spatially correlated with masks, significant increases in mask playback rates are correlated with minor increases in target detectability thresholds.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[2] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[3] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. 100, no. 1, pp. 90–93, 1974.

[4] A. B. Watson, "Image compression using the discrete cosine transform," *Mathematica journal*, vol. 4, no. 1, p. 81, 1994.

[5] J. A. Solomon, A. B. Watson, and A. Ahumada, "detectability of dct basis functions: Effects of contrast masking," in *Data Compression Conference, 1994. DCC'94. Proceedings*, pp. 361–370, IEEE, 1994.

[6] A. B. Watson, J. Hu, and J. F. McGowan, "Digital video quality metric based on human vision," *Journal of Electronic imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[7] J. Nachmias and B. E. Rogowitz, "Masking by spatially-modulated gratings," *Vision Research*, vol. 23, no. 12, pp. 1621–1629, 1983.

[8] J. G. ROBSON, "Spatial and temporal contrast-sensitivity functions of the visual system," *J. Opt. Soc. Am.*, vol. 56, pp. 1141–1142, Aug 1966.

[9] M. M. Alam, K. P. Vilankar, and D. M. Chandler, "A database of local masking thresholds in natural images," in *IS&T/SPIE Electronic Imaging*, pp. 86510G–86510G, International Society for Optics and Photonics, 2013.

[10] D. M. Chandler, M. D. Gaubatz, S. S. Hemami, *et al.*, "A patch-based structural masking model with an application to compression," *EURASIP Journal on Image and Video Processing*, vol. 2009, 2009.

[11] D. M. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *JOSA A*, vol. 20, no. 7, pp. 1164–1180, 2003.

[12] M. J. Nadenau, J. Reichel, and M. Kunt, "Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli," *Signal processing: Image communication*, vol. 17, no. 10, pp. 807–823, 2002.

[13] T. Caelli and G. Moraglia, "On the detection of signals embedded in natural scenes," *Perception & psychophysics*, vol. 39, no. 2, pp. 87–95, 1986.

[14] J. Evert and D. Chandler, "On the effectiveness of natural videos in masking dynamic dct noise," in *Signals, Systems and Computers, 2013 Asilomar Conference on*, pp. 1339–1345, IEEE, 2013.

[15] J. Evert and D. Chandler, "Visual masking of dynamic discrete-cosine-transform noise by natural videos: Experiment, analysis, and modeling," *Journal of Electronic Imaging*, 2015. In review.

[16] H. Peterson, A. Ahumada, and A. B. Watson, "The detectability of dct quantization noise," in *SID International Symposium Digest of Technical Papers*, vol. 24, pp. 942–942, Citeseer, 1993.

[17] D. Regan, *Human perception of objects*. Sinauer Associates Sunderland, MA, 2000.

[18] G. Mather, *Foundations of perception*. Taylor & Francis, 2006.

[19] R. L. De Valois and K. K. De Valois, "Spatial vision," *Annual review of psychology*, vol. 31, no. 1, pp. 309–341, 1980.

[20] H. Schober and R. Hilz, "Contrast sensitivity of the human eye for square-wave gratings," *JOSA*, vol. 55, no. 9, pp. 1086–1090, 1965.

[21] F. W. Campbell and J. Robson, "Application of fourier analysis to the detectability of gratings," *The Journal of physiology*, vol. 197, no. 3, p. 551, 1968.

[22] H. A. Peterson, A. J. Ahumada Jr, and A. B. Watson, "Improved detection model for dct coefficient quantization," in *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pp. 191–201, International Society for Optics and Photonics, 1993.

[23] F. Campbell and D. Green, "Optical and retinal factors affecting visual resolution.," *The Journal of Physiology*, vol. 181, no. 3, pp. 576–593, 1965.

[24] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," *Human Vision and Electronic Imaging III*, vol. 3299, pp. 180–191, 1998.

[25] D. H. KELLY, "Frequency doubling in visual responses," *J. Opt. Soc. Am.*, vol. 56, pp. 1628–1632, Nov 1966.

[26] D. H. Kelly, "Motion and vision. ii. stabilized spatio-temporal threshold surface," *J. Opt. Soc. Am.*, vol. 69, pp. 1340–1349, Oct 1979.

[27] A. B. Watson *et al.*, "Temporal sensitivity," *Handbook of perception and human performance*, vol. 1, pp. 6–1, 1986.

[28] H. D. L. Dzn, "Experiments on flicker and some calculations on an electrical analogue of the foveal systems," *Physica*, vol. 18, no. 11, pp. 935 – 950, 1952.

[29] D. J. Tolhurst and J. A. Movshon, "Spatial and temporal contrast sensitivity of striate cortical neurones," 1975.

[30] S. OTTO H. SCHADE, "Optical and photoelectric analog of the eye," *J. Opt. Soc. Am.*, vol. 46, pp. 721–738, Sep 1956.

[31] D. Tolhurst, "Separate channels for the analysis of the shape and the movement of a moving visual stimulus," *The Journal of Physiology*, vol. 231, no. 3, pp. 385–402, 1973.

[32] J. Kulikowski and D. Tolhurst, "Psychophysical evidence for sustained and transient detectors in human vision," *The Journal of Physiology*, vol. 232, no. 1, pp. 149–162, 1973.

[33] J. J. Koenderink, A. J. van Doorn, *et al.*, "Spatiotemporal contrast detection threshold surface is bimodal," *Optics Letters*, vol. 4, no. 1, pp. 32–34, 1979.

[34] S. J. Cropper and A. M. Derrington, "Detection and motion detection in chromatic and luminance beats," *JOSA A*, vol. 13, no. 3, pp. 401–407, 1996.

[35] G. E. Legge, D. Kersten, and A. E. Burgess, "Contrast discrimination in noise," *JOSA A*, vol. 4, no. 2, pp. 391–404, 1987.

[36] J. M. Foley and C.-C. Chen, "Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: Threshold measurements and a model," *Vision research*, vol. 39, no. 23, pp. 3855–3872, 1999.

[37] C.-C. Chen and J. M. Foley, "Pattern detection: Interactions between oriented and concentric patterns," *Vision Research*, vol. 44, no. 9, pp. 915–924, 2004.

[38] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *JOSA*, vol. 70, no. 12, pp. 1458–1471, 1980.

[39] F. W. Campbell and J. Kulikowski, "Orientational selectivity of the human visual system," *The Journal of physiology*, vol. 187, no. 2, pp. 437–445, 1966.

[40] S. R. Lehky, "Temporal properties of visual channels measured by masking," *JOSA A*, vol. 2, no. 8, pp. 1260–1272, 1985.

[41] R. E. Fredericksen and R. F. Hess, "Temporal detection in human vision: dependence on stimulus energy," *J. Opt. Soc. Am. A*, vol. 14, pp. 2557–2569, Oct 1997.

[42] D. G. Pelli, *Effects of visual noise*. PhD thesis, University of Cambridge, 1981.

[43] D. G. Pelli and B. Farell, "Why use noise?," *JOSA A*, vol. 16, no. 3, pp. 647–653, 1999.

[44] I. Oruç, M. S. Landy, and D. G. Pelli, "Noise masking reveals channels for second-order letters," *Vision research*, vol. 46, no. 8, pp. 1493–1506, 2006.

[45] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *JOSA A*, vol. 14, no. 9, pp. 2379–2391, 1997.

[46] J. M. Foley and G. E. Legge, "Contrast detection and near-threshold discrimination in human vision," *Vision research*, vol. 21, no. 7, pp. 1041–1053, 1981.

[47] G. E. Legge, "A power law for contrast discrimination," *Vision research*, vol. 21, no. 4, pp. 457–467, 1981.

[48] J. Ross and H. D. Speed, "Contrast adaptation and contrast masking in human vision," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 246, no. 1315, pp. 61–70, 1991.

[49] A. Dorais and D. Sagi, "Contrast masking effects change with practice," *Vision Research*, vol. 37, no. 13, pp. 1725–1733, 1997.

[50] D. J. Swift and R. A. Smith, "Spatial frequency masking and weber's law," *Vision Research*, vol. 23, no. 5, pp. 495–505, 1983.

[51] A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," in *Electronic Imaging'97*, pp. 2–12, International Society for Optics and Photonics, 1997.

[52] O. O. Imade and D. M. Chandler, "Image-adaptive contrast and entropy based model of regions of visible distortion," in *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*, pp. 65–68, IEEE, 2010.

[53] O. Braddick, "A short-range process in apparent motion," *Vision research*, vol. 14, no. 7, pp. 519–527, 1974.

[54] S. A. Klein, T. Carney, L. Barghout-Stein, and C. W. Tyler, "Seven models of masking," in *Electronic Imaging'97*, pp. 13–24, International Society for Optics and Photonics, 1997.

[55] P. W. Jones, S. J. Daly, R. S. Gaborski, and M. Rabbani, "Comparative study of wavelet and discrete cosine transform (dct) decompositions with equivalent quantization and encoding strategies for medical images," in *Medical Imaging 1995*, pp. 571–582, International Society for Optics and Photonics, 1995.

[56] W. Zeng, S. Daly, and S. Lei, "An overview of the visual optimization tools in jpeg 2000," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 85–104, 2002.

[57] J. M. Foley and G. M. Boynton, "Forward pattern masking and adaptation: Effects of duration, interstimulus interval, contrast, and spatial and temporal frequency," *Vision research*, vol. 33, no. 7, pp. 959–980, 1993.

[58] A. B. Watson, J. A. Solomon, A. J. Ahumada Jr, and A. Gale, "Discrete cosine transform (dct) basis function detectability: effects of viewing distance and contrast masking," in *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pp. 99–108, International Society for Optics and Photonics, 1994.

[59] D. M. Chandler and S. S. Hemami, "Contrast based quantization and rate control for wavelet coded images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3, pp. III–233, IEEE, 2002.

[60] D. M. Chandler, M. A. Masry, and S. S. Hemami, "Quantifying the visual quality of wavelet-compressed images based on local contrast, visual masking, and global precedence," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1393–1397, IEEE, 2003.

[61] D. M. Chandler, N. L. Dykes, and S. S. Hemami, "Visually lossless compression of digitized radiographs based on contrast sensitivity and visual masking," in *Medical Imaging*, pp. 359–372, International Society for Optics and Photonics, 2005.

[62] D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *Image Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 397–410, 2005.

[63] C. Bird, G. Henning, and F. Wichmann, "Contrast discrimination with sinusoidal gratings of different spatial frequency," *JOSA A*, vol. 19, no. 7, pp. 1267–1273, 2002.

[64] S. Jiménez, X. Otazu, V. Laparra, and J. Malo, "Chromatic induction and contrast masking: similar models, different goals?," in *IS&T/SPIE Electronic Imaging*, pp. 86511J–86511J, International Society for Optics and Photonics, 2013.

[65] E. Peli, J. Yang, and R. B. Goldstein, "Image invariance with changes in size: The role of peripheral contrast thresholds," *JOSA A*, vol. 8, no. 11, pp. 1762–1774, 1991.

[66] D. J. Field *et al.*, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[67] Y. Petrov, "Luminance correlations define human sensitivity to contrast resolution in natural images," *JOSA A*, vol. 22, no. 4, pp. 587–592, 2005.

[68] R. A. Frazor, W. S. Geisler, *et al.*, "Local luminance and contrast in natural images," *Vision research*, vol. 46, no. 10, pp. 1585–1598, 2006.

[69] M. M. Alam, K. P. Vilankar, D. J. Field, and D. M. Chandler, "Local masking in natural images: A database and analysis," *Journal of vision*, vol. 14, no. 8, p. 22, 2014.

[70] H. A. Peterson, H. Peng, J. H. Morgan, and W. B. Pennebaker, "Quantization of color image components in the dct domain," 1991.

[71] S. J. Daly, "Application of a noise-adaptive contrast sensitivity function to image data compression," *Optical Engineering*, vol. 29, no. 8, pp. 977–987, 1990.

[72] A. B. Watson, "Dct quantization matrices visually optimized for individual images," in *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pp. 202–216, International Society for Optics and Photonics, 1993.

[73] A. B. Watson, "Visual optimization of dct quantization matrices for individual images," *Proc. AIAA Computing in Aerospace*, vol. 9, pp. 286–291, 1993.

[74] A. B. Watson, "Dctune: A technique for visual optimization of dct quantization matrices for individual images," *Sid International Symposium Digest of Technical Papers*, vol. 24, pp. 946–946, 1993.

[75] A. B. Watson, "Perceptual optimization of dct color quantization matrices," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 1, pp. 100–104, IEEE, 1994.

[76] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "detectability of wavelet quantization noise," *Image Processing, IEEE Transactions on*, vol. 6, no. 8, pp. 1164–1175, 1997.

[77] A. J. Ahumada Jr and H. A. Peterson, "Luminance-model-based dct quantization for color image compression," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pp. 365–374, International Society for Optics and Photonics, 1992.

[78] T. Carney, C. W. Tyler, A. B. Watson, W. Makous, B. Beutter, C.-C. Chen, A. M. Norcia, and S. A. Klein, "Modelfest: Year one results and plans for future years," in *Electronic Imaging*, pp. 140–151, International Society for Optics and Photonics, 2000.

[79] A. B. Watson, M. Taylor, and R. Borthwick, "Dctune perceptual optimization of compressed dental x-rays," *Medical Imaging, SPIE Proceedings*, vol. 3031, 1997.

[80] M. P. Eckstein, A. J. Ahumada, and A. B. Watson, "Visual signal detection in structured backgrounds. ii. effects of contrast gain control, background variations, and white noise," *JOSA A*, vol. 14, no. 9, pp. 2406–2419, 1997.

[81] D. M. Chandler, K. H. Lim, and S. S. Hemami, "Effects of spatial correlations and global precedence on the visual fidelity of distorted images," in *Electronic Imaging 2006*, pp. 60570F–60570F, International Society for Optics and Photonics, 2006.

[82] S. S. Hemami, D. M. Chandler, B. G. Chern, and J. A. Moses, "Suprathreshold visual psychophysics and structure-based visual masking," in *Electronic Imaging 2006*, pp. 60770O–60770O, International Society for Optics and Photonics, 2006.

[83] D. M. Chandler and S. S. Hemami, "Additivity models for suprathreshold distortion in quantized wavelet-coded images," in *Electronic Imaging 2002*, pp. 105–118, International Society for Optics and Photonics, 2002.

[84] M. A. Webster and E. Miyahara, "Contrast adaptation and the spatial structure of natural images," *JOSA A*, vol. 14, no. 9, pp. 2355–2366, 1997.

[85] A. J. Ahumada Jr and H. A. Peterson, "A visual detection model for dct coefficient quantization," in *Computing in Aerospace*, vol. 9, pp. 314–318, 1993.

[86] A. J. Ahumada and A. B. Watson, "Visible contrast energy metrics for detection and discrimination," in *IS&T/SPIE Electronic Imaging*, pp. 86510D–86510D, International Society for Optics and Photonics, 2013.

[87] D. Kelly, "Visual contrast sensitivity," *Journal of modern optics*, vol. 24, no. 2, pp. 107–129, 1977.

[88] C. A. Burbeck and D. Kelly, "Contrast gain measurements and the transient/sustained dichotomy," *JOSA*, vol. 71, no. 11, pp. 1335–1342, 1981.

[89] A. J. Pantle, "Temporal determinants of spatial sine-wave masking," *Vision Research*, vol. 23, no. 8, pp. 749–757, 1983.

[90] B. Breitmeyer, D. M. Levi, and R. S. Harwerth, "Flicker masking in spatial vision," *Vision Research*, vol. 21, no. 9, pp. 1377–1385, 1981.

[91] M. Green, "Psychophysical relationships among mechanisms sensitive to pattern, motion and flicker," *Vision Research*, vol. 21, no. 7, pp. 971–983, 1981.

[92] R. A. Smith, "Studies of temporal frequency adaptation in visual contrast sensitivity," *The Journal of physiology*, vol. 216, no. 3, pp. 531–552, 1971.

[93] G. B. Henning, "Spatial-frequency tuning as a function of temporal frequency and stimulus motion," *JOSA A*, vol. 5, no. 8, pp. 1362–1373, 1988.

[94] R. Hess and R. Snowden, "Temporal properties of human visual filters: Number, shapes and spatial covariation," *Vision research*, vol. 32, no. 1, pp. 47–59, 1992.

[95] M. P. Eckstein, J. S. Whiting, and J. P. Thomas, "Role of knowledge in human visual temporal integration in spatiotemporal noise," *JOSA A*, vol. 13, no. 10, pp. 1960–1968, 1996.

[96] Z.-L. Lu and G. Sperling, "Contrast gain control in first-and second-order motion perception," *JOSA A*, vol. 13, no. 12, pp. 2305–2318, 1996.

[97] R. Fredericksen and R. Hess, "Estimating multiple temporal mechanisms in human vision," *Vision Research*, vol. 38, no. 7, pp. 1023–1040, 1998.

[98] G. M. Boynton and J. M. Foley, "Temporal sensitivity of human luminance pattern mechanisms determined by masking with temporally modulated stimuli," *Vision Research*, vol. 39, no. 9, pp. 1641–1656, 1999.

[99] L. Meier and M. Carandini, "Masking by fast gratings," *Journal of Vision*, vol. 2, no. 4, p. 2, 2002.

[100] J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly, "Spatio-velocity csf as a function of retinal velocity using unstabilized stimuli," in *Electronic Imaging 2006*, pp. 605705–605705, International Society for Optics and Photonics, 2006.

[101] R. W. Sekuler and L. Ganz, "Aftereffect of seen motion with a stabilized retinal image," *Science*, vol. 139, no. 3553, pp. 419–419, 1963.

[102] A. B. Watson, "Toward a perceptual video-quality metric," in *Photonics West'98 Electronic Imaging*, pp. 139–147, International Society for Optics and Photonics, 1998.

[103] A. Watanabe, T. Mori, S. Nagata, and K. Hiwatashi, "Spatial sine-wave responses of the human visual system," *Vision Research*, vol. 8, no. 9, pp. 1245–1263, 1968.

[104] R. J. Snowden, R. F. Hess, and S. J. Waugh, "The processing of temporal modulation at different levels of retinal illuminance," *Vision Research*, vol. 35, no. 6, pp. 775–789, 1995.

[105] N. Graham, "Spatial frequency channels in the human visual system: Effects of luminance and pattern drift rate," *Vision research*, vol. 12, no. 1, pp. 53–68, 1972.

[106] D. Kelly, "Motion and vision. l. stabilized images of stationary gratings," *JOSA*, vol. 69, no. 9, pp. 1266–1274, 1979.

[107] E. Levinson and R. Sekuler, "The independence of channels in human vision selective for direction of movement.," *The Journal of Physiology*, vol. 250, no. 2, pp. 347–366, 1975.

[108] A. B. Watson, P. G. Thompson, B. J. Murphy, and J. Nachmias, "Summation and discrimination of gratings moving in opposite directions," *Vision Research*, vol. 20, no. 4, pp. 341–347, 1980.

[109] A. Van Doorn and J. Koenderink, "Temporal properties of the visual detectability of moving spatial white noise," *Experimental Brain Research*, vol. 45, no. 1-2, pp. 179–188, 1982.

[110] D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," *Vision research*, vol. 22, no. 4, pp. 479–484, 1982.

[111] A. B. Watson, A. J. Ahumada Jr, *et al.*, "Model of human visual-motion sensing," *JOSA A*, vol. 2, no. 2, pp. 322–341, 1985.

[112] A. B. Watson and J. Nachmias, "Patterns of temporal interaction in the detection of gratings," *Vision Research*, vol. 17, no. 8, pp. 893–902, 1977.

[113] A. B. Watson and A. J. Ahumada, "A standard model for foveal detection of spatial contrast," *Journal of Vision*, vol. 5, no. 9, 2005.

[114] M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial-frequency channels in human vision," *JOSA*, vol. 61, no. 9, pp. 1176–1186, 1971.

[115] M. Georgeson and G. Sullivan, "Contrast constancy: deblurring in human vision by spatial frequency channels.," *The Journal of Physiology*, vol. 252, no. 3, pp. 627–656, 1975.

[116] G. Sperling, "Model of visual adaptation and contrast detection," *Perception & Psychophysics*, vol. 8, no. 3, pp. 143–157, 1970.

[117] R. Quick Jr, "A vector-magnitude model of contrast detection," *Kybernetik*, vol. 16, no. 2, pp. 65–67, 1974.

[118] D. G. Pelli, "Uncertainty explains many aspects of visual contrast detection and discrimination," *JOSA A*, vol. 2, no. 9, pp. 1508–1531, 1985.

[119] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision research*, vol. 22, no. 5, pp. 545–559, 1982.

[120] S.-H. Lee and R. Blake, "Detection of temporal structure depends on spatial structure," *Vision research*, vol. 39, no. 18, pp. 3033–3048, 1999.

[121] M. Carrasco, C. Penpeci-Talgar, and M. Eckstein, "Spatial covert attention increases contrast sensitivity across the csf: support for signal enhancement," *Vision research*, vol. 40, no. 10, pp. 1203–1215, 2000.

[122] J. R. Jarvis and C. M. Wathes, "Mechanistic modeling of vertebrate spatial contrast sensitivity and acuity at low luminance," *Visual neuroscience*, vol. 29, no. 03, pp. 169–181, 2012.

[123] A. B. Watson and J. Malo, "Video quality measures based on the standard spatial observer," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3, pp. III–41, IEEE, 2002.

[124] A. B. Watson, "31.1: Invited paper: The spatial standard observer: A human vision model for display inspection," in *SID Symposium Digest of Technical Papers*, vol. 37, pp. 1312–1315, Wiley Online Library, 2006.

[125] A. Watson and C. Ramirez, "A standard observer for spatial vision.," in *Investigative Ophthalmology & Visual Science*, vol. 41, pp. S713–S713, ASSOC

RESEARCH VISION OPHTHALMOLOGY INC 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998 USA, 2000.

[126] S. M. Wuerger, A. B. Watson, and A. J. Ahumada Jr, "Towards a spatio-chromatic standard observer for detection," in *Electronic Imaging 2002*, pp. 159–172, International Society for Optics and Photonics, 2002.

[127] A. B. Watson and A. J. Ahumada, "The spatial standard observer," *Journal of Vision*, vol. 4, no. 8, pp. 51–51, 2004.

[128] T. Carney, S. A. Klein, C. W. Tyler, A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C.-C. Chen, *et al.*, "Development of an image/threshold database for designing and testing human vision models," in *Electronic Imaging'99*, pp. 542–551, International Society for Optics and Photonics, 1999.

[129] J. M. Foley, "Human luminance pattern-vision mechanisms: masking experiments require a new model," *JOSA A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[130] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pp. 127–141, International Society for Optics and Photonics, 1994.

[131] A. B. Watson, Q. J. Hu, J. F. McGowan III, and J. B. Mulligan, "Design and performance of a digital video quality metric," in *Electronic Imaging'99*, pp. 168–174, International Society for Optics and Photonics, 1999.

[132] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Visually lossless h. 264 compression of natural videos," *The Computer Journal*, vol. 56, no. 5, pp. 617–627, 2013.

[133] P. R. Bevington, *Data reduction and error analysis for the physical sciences, ise*. 1969.

[134] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, no. 4, pp. 433–436, 1997.

[135] D. G. Pelli, "The videotoolbox software for visual psychophysics: Transforming numbers into movies," *Spatial vision*, vol. 10, no. 4, pp. 437–442, 1997.

[136] A. B. Watson and D. G. Pelli, "Quest: A bayesian adaptive psychometric method," *Attention, Perception, & Psychophysics*, vol. 33, no. 2, pp. 113–120, 1983.

[137] Tangopaso, "Waterfall." Wikipedia File:Saut de l'Ognon.ogv. Online; accessed 21 March 2015. Public domain.

[138] RAI and EBU, "Cactus." ftp://ftp.tnt.uni-hannover.de/testsequences. These sequences and all intellectual property rights therein remain the property of the RAI. These sequences may only be used for the purpose of developing, testing and promulgating technology standards. RAI and EBU Technical make no warranties with respect to the sequences and expressly disclaims any warranties regarding their fitness for any purpose.

[139] N. L. of Tokyo Institute of Technology, "Kimono." ftp://ftp.tnt.uni-hannover.de/testsequences. Individuals and organizations extracting sequence from this archive agree that the sequences and all intellectual property rights therein remain the property of Nakajima Laboratory of Tokyo Institute of Technology. This material may only be used for the purpose of developing, testing and promulgating technology standards. The material cannot be distributed with charge. Nakajima Laboratory of Tokyo Institute of Technology makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose.

[140] Mike.lifeguard, "Hands." Wikimedia File:Two-hand manual.ogg, October 2006. Online; accessed 21 March 2015. Public domain permission.

[141] Xiph.org, "Timelapse." http://media.xiph.org/basilgohar/timelapse/kt-clouds-1/. Online; accessed 21 March 2015.

[142] S. E. Patel, "Lemur." Wikimedia: Propithecus candidus ground feeding 001.ogv. Online; accessed 21 March 2015.

[143] NotFromUtrecht, "Typing." Wikimedia File: Typing example.ogv. Online; accessed 21 March 2015.

[144] N. DOCOMO, "Flowervase." ftp://ftp.tnt.uni-hannover.de/testsequences. These sequences and all intellectual property rights therein remain the property of the NTT DOCOMO, INC. These sequences may only be used for the purpose of developing, testing and promulgating technology standards. NTT DOCOMO, INC. makes no warranties with respect to the sequences and expressly disclaims any warranties regarding their fitness for any purpose.

[145] A. B. Watson, H. Barlow, J. G. Robson, *et al.*, "What does the eye see best?," *Nature*, vol. 302, no. 5907, pp. 419–422, 1983.

[146] "Personal conversation with professor lecallet during asilomar ssc 2013."

[147] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *Signal Processing Letters, IEEE*, vol. 19, no. 7, pp. 423–426, 2012.

[148] C. T. Vu and D. M. Chandler, "S3: a spectral and spatial sharpness measure," in *Advances in Multimedia, 2009. MMEDIA'09. First International Conference on*, pp. 37–43, IEEE, 2009.

[149] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.

[150] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error detectability to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[151] D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 2284–2298, 2007.

[152] M. Vranješ, S. Rimac-Drlje, and D. Žagar, "Objective video quality metrics," in *49th International Symposium ELMAR-2007 focused on Mobile Multimedia*, 2007.

[153] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pp. 2–15, International Society for Optics and Photonics, 1992.

[154] R. A. Normann, B. S. Baxter, H. Ravindra, and P. J. Anderton, "Photoreceptor contributions to contrast sensitivity: applications in radiological diagnosis," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 5, pp. 944–953, 1983.

[155] D. M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *ISRN Signal Processing*, vol. 2013, 2013.

[156] H. R. Blackwell, "Contrast thresholds of the human eye.," *Journal of the Optical Society of America*, 1946.

[157] L. R. Young, "Pursuit eye tracking movements," *The control of eye movements*, pp. 429–443, 1971.

[158] H. Barlow, "The coding of sensory messages," *Current problems in animal behaviour*, pp. 331–360, 1961.

[159] H. Barlow, "Understanding natural vision," in *Physical and biological processing of images*, pp. 2–14, Springer, 1983.

[160] M. S. Beauchamp, K. E. Lee, J. V. Haxby, and A. Martin, "Fmri responses to video and point-light displays of moving humans and manipulable objects," *Journal of Cognitive Neuroscience*, vol. 15, no. 7, pp. 991–1001, 2003.

[161] A. C. Bovik, *Handbook of image and video processing*. Academic Press, 2010.

[162] D. Bowker, "Suprathreshold spatiotemporal response characteristics of the human visual system," *JOSA*, vol. 73, no. 4, pp. 436–440, 1983.

[163] N. Brady and D. J. Field, "What's constant in contrast constancy? the effects of scaling on the perceived contrast of bandpass patterns," *Vision Research*, vol. 35, no. 6, pp. 739–756, 1995.

[164] W.-H. Chen and W. Pratt, "Scene adaptive coder," *Communications, IEEE Transactions on*, vol. 32, no. 3, pp. 225–232, 1984.

[165] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, 2011.

[166] K. K. De Valois, *Seeing*. Academic Press, 2000.

[167] J. Fiser, P. J. Bex, and W. Makous, "Contrast conservation in human vision," *Vision Research*, vol. 43, no. 25, pp. 2637–2648, 2003.

[168] N. V. S. Graham, *Visual pattern analyzers*. No. 16, Oxford University Press on Demand, 2001.

[169] S. F. Gull, "Developments in maximum-entropy data analysis," in *Maximum Entropy and Bayesian Methods* (J. Skilling, ed.), pp. 53–71, Kluwer Academic, Dordrecht, 1989.

[170] K. M. Hanson, "Introduction to Bayesian image analysis," in *Medical Imaging: Image Processing* (M. H. Loew, ed.), vol. 1898 of *Proc. SPIE*, pp. 716–731, Feb. 1993. [doi:10.1117/12.154577].

[171] A. Harris, J. J. Sluss, Jr., H. H. Refai, and P. G. LoPresti, "Free-space optical wavelength diversity scheme for fog migration in a ground-to-unmanned-aerial-vehicle communications link," *Opt. Eng.*, vol. 45, p. 086001, 2006. [doi:10.1117/1.2338565].

[172] E. Hiris, D. Humphrey, and A. Stout, "Temporal properties in masking biological motion," *Perception & psychophysics*, vol. 67, no. 3, pp. 435–443, 2005.

[173] P. Le Callet, A. Saadane, and D. Barba, "Interactions of chromatic components on the perceptual quantization of the achromatic component," in *SPIE Human Vision and Electronic Imaging*, vol. 3644, 1999.

[174] Z.-L. Lu and G. Sperling, "The functional architecture of human visual motion perception," *Vision research*, vol. 35, no. 19, pp. 2697–2722, 1995.

[175] H. Oh, A. Bilgin, and M. Marcellin, "detectability thresholds for quantization distortion in jpeg2000," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 228–232, IEEE, 2009.

[176] B. A. Olshausen and D. J. Field, "What is the other 85% of v1 doing," *Problems in Systems Neuroscience*, pp. 182–211, 2004.

[177] T. N. Pappas, "The rough side of texture: texture analysis through the lens of hvei," in *IS&T/SPIE Electronic Imaging*, pp. 86510P–86510P, International Society for Optics and Photonics, 2013.

[178] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *Broadcasting, IEEE Transactions on*, vol. 50, no. 3, pp. 312–322, 2004.

[179] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *JOSA A*, vol. 18, no. 10, pp. 2385–2397, 2001.

[180] K. Seshadrinathan, R.-j.-i. Soundararaj an, A. C.-n.-r.-a. Bovik, and L. .-K. Cormack, "Study of subjective and objective quality assessment of video," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1427–1441, 2010.

[181] .-R. .-a.-B. .-A. C. .-a.-C. .-L. K. Seshadrinathan, Kalpana andSoundararajan, "A subjective study to evaluate video quality assessment algorithms," in *IS&T/SPIE Electronic Imaging*, pp. 75270H–75270H, International Society for Optics and Photonics, 2010.

[182] A. T. Smith and R. J. Snowden, *Visual detection of motion*. Academic Pr, 1994.

[183] C. Vu and D. M. Chandler, "Main subject detection via adaptive feature refinement," *Journal of Electronic Imaging*, vol. 20, March 2011.

[184] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1398–1402, IEEE, 2003.

[185] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.

[186] A. B. Watson *et al.*, "Visual detection of spatial contrast patterns: Evaluation of five simple models," *Optics Express*, vol. 6, no. 1, pp. 12–33, 2000.

[187] S. Winkler, "Perceptual distortion metric for digital color video," in *Electronic Imaging'99*, pp. 175–184, International Society for Optics and Photonics, 1999.

[188] S. Winkler, *Digital video quality: vision models and metrics*. Wiley. com, 2005.

[189] S. Winkler, "Analysis of public image and video databases for quality assessment," vol. 6, pp. 616–625, Oct. 2012.

APPENDIX A

INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL LETTER

Oklahoma State University Institutional Review Board

Date	Thursday, May 22, 2014	**Protocol Expires:**	**5/21/2015**
IRB Application No:	EG096		
Proposal Title:	Content-Based Strategies of Image and Video Compression and Quality Assessments		
Reviewed and Processed as:	Expedited **Continuation**		

Status Recommended by Reviewer(s): **Approved**

Principal Investigator(s) :

Damon Chandler
202 Engineering South
Stillwater, OK 74078

Approvals are valid until the given expiration date, after which time a request for continuation must be submitted. Any modifications to the research project approved by the IRB must be submitted for approval with the advisor's signature. The IRB office MUST be notified in writing when a project is complete. Approved projects are subject to monitoring by the IRB. Expedited and exempt projects may be reviewed by the full Institutional Review Board.

☒ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

Signature :

Sheila Kennison, Chair, Institutional Review Board

Thursday, May 22, 2014
Date

CONSENT TO PARTICIPATE IN A RESEARCH STUDY
OKLAHOMA STATE UNIVERSITY

PROJECT TITLE: Content-Based Strategies of Image and Video Compression and Quality Assessment

INVESTIGATOR: Damon Chandler, Ph.D., School of Electrical and Computer Engineering

PURPOSE:

The purpose of this study is to examine how digital alteration affects the visual quality of images and video. This study is part of a five-year project to research new ways of automatically assessing whether certain digital manipulations are good or bad for images and video.

PROCEDURES:

In this experiment, you will view a series of original and digitally altered images and videos. For each altered image, you are asked to provide a score that indicates the image's visual quality relative to the original. During this time, you will wear special goggles with small cameras attached to them, which allow us to track your eye movements. The goggles are lightweight and comfortable; they are similar to wearing eyeglasses without lenses.

The experiment will be conducted in the basement of Engineering South on the Oklahoma State University main campus. Each session will entail viewing 3-5 original images and several altered versions of each original image/video. For each altered image/video, you will provide a rating of visual quality relative to the original using the keyboard/mouse. Your quality ratings will be on a scale from 0-200, where a higher value denotes greater quality, and where 100 denotes equal quality to the original. Your answers will be recorded via the computer.

After your ratings are complete, we will test your ability to visually detect compression artifacts in the images. On the computer screen, you will view a subset of the same images presented in sets of three at a time. For each set of three, you will indicate, via the computer keyboard, which of the three contains compression artifacts. This task will be repeated 30 times for each set.

Next, you will be asked to categorize a collection of patches taken from the images. On the computer screen, you will be shown an original image, and a small red box will also be shown to highlight a particular patch. For the highlighted patch, you will indicate, via the computer keyboard, to which of several predefined categories the patch belongs. You will be asked to categorize approximately 200 patches.

Each experimental session is designed to last approximately 60 minutes. A short break will be given after 30 minutes.

As a participant, you will be expected to complete at least one (1) experimental session. If you desire, you can participate in up to five (5) experimental sessions. We will do our best to schedule each session to best accommodate your available times. However, to minimize fatigue, no more than two sessions can be scheduled on the same day.

RISKS OF PARTICIPATION:

There are no risks associated with this project, including stress, psychological, social, physical, or legal risk which are greater, considering probability and magnitude, than those ordinarily encountered in daily life. If, however, you begin to experience discomfort or stress in this project, you may end your participation at any time.

BENEFITS OF PARTICIPATION:

While there are no direct benefits to you from this research, you may find the experiment interesting. Your participation will ultimately help to improve how digital images and video are compressed and enhanced.

CONFIDENTIALITY:

The data gathered in this experiment will be treated with confidentiality. Your answers will be recorded via the computer and will be stored on a computer in the Image Coding and Analysis Lab. All information about you will be kept confidential and will not be released. Your answers will have identification numbers, rather than names, on them. Research records will be stored securely and only researchers and individuals responsible for research oversight will have access to the records. It is possible that the consent process and data collection will be observed by research oversight staff responsible for safeguarding the rights and wellbeing of people who participate in research. The data will be saved as long as it is scientifically useful; typically, such information is kept for five years after publication of the results. Results from this study may be presented at professional meetings or in publications. You will not be identified individually.

COMPENSATION:

Compensation for your participation in this experiment will be made in one of the following two ways:

Option 1: You will receive $7 for each experimental session to be paid on the day of the experiment. You may earn up to a maximum of $35 for completion of five experimental sessions. If you withdraw from the study, compensation will be issued for a percentage of this amount that matches your time of participation.

Option 2: Alternatively, if you are enrolled in a participating ECE course, you may elect to receive compensation in the form of bonus points. Completion of each experimental session will be compensated with one (1) bonus point added to your final grade. You may earn up to a maximum of five bonus points for completion of five experimental sessions. Please note that an alternative way to earn such bonus points is available: If you desire, you may complete the additional problem sets that will be made available throughout the semester. Successful completion of each problem set will earn one (1) bonus point (up to a maximum of five bonus points for completion of five problem sets).

CONTACTS:

You may contact Dr. Damon Chandler at the following address and phone number, should you desire to discuss your participation in the study and/or request information about the results of the study: Damon Chandler, Ph.D., School of Electrical and Computer Engineering, 202 Engineering South, Oklahoma State University, Stillwater, OK 74078, (405) 744-9924 or damon.chandler@okstate.edu.

If you have questions about your rights as a research volunteer, you may contact Dr. Shelia Kennison, IRB Chair, 219 Cordell North, Stillwater, OK 74078, 405-744-3377 or irb@okstate.edu.

PARTICIPANT RIGHTS:

Your participation in this research is voluntary. There is no penalty for refusal to participate, and that you are free to withdraw your consent and participation in this project at any time, without penalty.

CONSENT DOCUMENTATION:

I have been fully informed about the procedures listed here. I am aware of what I will be asked to do and the benefits of my participation. I also understand the following statements:

- I affirm that I am 18 years of age or older.
- I have read and fully understand this consent form. I sign it freely and voluntarily. A copy of this form will be given to me. I hereby give permission for my participation in the study.

_____ _____

Signature of Participant Date

I certify that I have personally explained this document before requesting that the participant sign it.

_____ _____

Signature of Researcher Date

In-class announcement to be made to potential participants

Good Morning/Afternoon,

I would like to make an announcement to seek your participation in a research experiment. Research is currently underway to help develop automated methods of determining image and video quality. This study is being conducted by Dr. Damon Chandler at the Image Coding and Analysis Lab at Oklahoma State University. The project is being funded by the National Science Foundation.

I'm sure all of you have seen images and video streamed over the Internet. Some images and video appear to be of high quality, while others appear severely degraded. In this research, Dr. Chandler's lab is studying how digital compression affects the visual quality of images and video. The goal is to develop new techniques of compression that can ultimately lead to higher-quality media over the Internet.

To achieve this goal, participants are needed to provide ratings of quality for various altered images and video. The images and video will contain commonplace subject matter—nothing offensive. During this time, you may be asked to wear special goggles to track your eye movements. After you provide your quality ratings, we will measure your ability to visually detect compression artifacts, and we will ask you to categorize some image patches. Each experimental session will require about an hour of your time. You can participate in 1-5 sessions, and the sessions can be scheduled to accommodate your own work schedule.

For each completed session, one bonus point added to your final grade. You may participate in up to five sessions for a maximum compensation of five bonus points. Note that you may alternatively complete the additional problem sets that will be handed out throughout the semester to earn the bonus points. So, please don't feel that participation in this experiment is required to earn the bonus points.

If this experiment sounds to be of interest to you, please contact me via phone or e-mail, and I would be happy to provide further details. <*Announcer will provide his/her phone and e-mail address.*> Thanks very much.

APPENDIX B

MODEL FIT PERFORMANCE FOR FOUR INPUT MODELS

Please see Sect. 6.1 and Table 6.2 for additional details. This section presents quantification of four input model performance. Tables in this section provide the PCC, SROCC, and RMSE between model predictions and measured masked target visibility contrast thresholds. Each table has twelve rows of data, corresponding to the three performance measures for four different methods to collapse multiple measures into a single measure. In each table, the four left columns correspond to the four measurement treatments applied before including the measure in the model.

For the spatial video content measurements, each frame was measured, and then the frame measurements were collapsed over time into a single measurement. Likewise, for the temporal video content measures, statistics were calculated on a pixel by pixel basis and then collapsed over all pixels. The four methods to collapse measurements, were a simple average, the 2-Norm, 5-Norm, and finally, selecting the maximum measure. Selecting the maximum over all frames was suggested by the VQEG. [149]

Table B.1: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure VQEG Spatial Perceptual Information.

Clock Time (sec)		2.50			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.751	0.725	0.713	0.693
	SROCC	0.749	0.722	0.713	0.688
	RMSE	0.501	0.523	0.533	0.547
2-Norm	PCC	0.751	0.725	0.712	0.693
	SROCC	0.749	0.723	0.714	0.688
	RMSE	0.502	0.523	0.533	0.547
5-Norm	PCC	0.750	0.725	0.712	0.693
	SROCC	0.748	0.723	0.713	0.687
	RMSE	0.502	0.523	0.533	0.548
Max	PCC	0.747	0.724	0.709	0.690
	SROCC	0.745	0.720	0.709	0.685
	RMSE	0.505	0.524	0.536	0.550

Table B.2: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure VQEG Temporal Perceptual Information.

Clock Time (sec)		1.65			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.689	0.692	0.685	0.682
	SROCC	0.677	0.679	0.675	0.673
	RMSE	0.550	0.548	0.553	0.555
2-Norm	PCC	0.709	0.700	0.689	0.682
	SROCC	0.699	0.690	0.684	0.666
	RMSE	0.535	0.543	0.551	0.555
5-Norm	PCC	0.708	0.695	0.689	0.683
	SROCC	0.698	0.686	0.683	0.665
	RMSE	0.537	0.546	0.551	0.554
Max	PCC	0.704	0.693	0.688	0.684
	SROCC	0.695	0.684	0.682	0.665
	RMSE	0.540	0.547	0.551	0.554

Table B.3: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Standard Deviation.

Clock Time (sec)				1.58	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.812	0.755	0.742	0.702
	SROCC	0.806	0.750	0.740	0.700
	RMSE	0.444	0.498	0.509	0.541
2-Norm	PCC	0.811	0.756	0.741	0.701
	SROCC	0.804	0.749	0.738	0.699
	RMSE	0.444	0.497	0.510	0.541
5-Norm	PCC	0.809	0.755	0.738	0.699
	SROCC	0.800	0.749	0.736	0.696
	RMSE	0.447	0.498	0.512	0.543
Max	PCC	0.797	0.753	0.727	0.693
	SROCC	0.789	0.745	0.727	0.688
	RMSE	0.458	0.500	0.521	0.548

Table B.4: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Skewness.

Clock Time (sec)		2.78			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.722	0.738	0.700	0.738
	SROCC	0.724	0.739	0.705	0.744
	RMSE	0.525	0.513	0.543	0.512
2-Norm	PCC	0.725	0.747	0.698	0.736
	SROCC	0.725	0.747	0.700	0.742
	RMSE	0.523	0.505	0.544	0.514
5-Norm	PCC	0.711	0.725	0.691	0.718
	SROCC	0.707	0.723	0.685	0.725
	RMSE	0.534	0.523	0.549	0.529
Max	PCC	0.693	0.691	0.685	0.698
	SROCC	0.682	0.681	0.674	0.700
	RMSE	0.547	0.549	0.553	0.544

Table B.5: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Kurtosis.

Clock Time (sec)		2.74			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.797	0.813	0.718	0.763
	SROCC	0.796	0.812	0.722	0.769
	RMSE	0.459	0.442	0.528	0.491
2-Norm	PCC	0.782	0.802	0.709	0.750
	SROCC	0.778	0.798	0.711	0.757
	RMSE	0.473	0.454	0.536	0.503
5-Norm	PCC	0.742	0.754	0.694	0.726
	SROCC	0.737	0.751	0.689	0.733
	RMSE	0.509	0.499	0.547	0.523
Max	PCC	0.707	0.705	0.686	0.701
	SROCC	0.696	0.692	0.674	0.705
	RMSE	0.537	0.539	0.553	0.541

Table B.6: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Edge Density.

Clock Time (sec)				15.33	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.740	0.732	0.699	0.751
	SROCC	0.742	0.733	0.701	0.752
	RMSE	0.511	0.518	0.543	0.501
2-Norm	PCC	0.745	0.737	0.701	0.754
	SROCC	0.747	0.739	0.702	0.755
	RMSE	0.507	0.513	0.542	0.499
5-Norm	PCC	0.754	0.750	0.705	0.758
	SROCC	0.755	0.752	0.707	0.759
	RMSE	0.499	0.503	0.539	0.495
Max	PCC	0.784	0.790	0.715	0.770
	SROCC	0.787	0.793	0.721	0.773
	RMSE	0.472	0.465	0.531	0.485

Table B.7: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Entropy.

Clock Time (sec)		1.73			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.682	0.685	0.684	0.687
	SROCC	0.672	0.673	0.675	0.684
	RMSE	0.555	0.554	0.554	0.552
2-Norm	PCC	0.682	0.684	0.684	0.687
	SROCC	0.672	0.671	0.675	0.684
	RMSE	0.555	0.554	0.554	0.552
5-Norm	PCC	0.682	0.684	0.684	0.688
	SROCC	0.672	0.671	0.675	0.685
	RMSE	0.555	0.554	0.554	0.551
Max	PCC	0.684	0.683	0.685	0.692
	SROCC	0.674	0.671	0.678	0.692
	RMSE	0.554	0.555	0.553	0.548

Table B.8: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Local Entropy.

Clock Time (sec)				17.69	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.687	0.683	0.687	0.697
	SROCC	0.681	0.676	0.683	0.700
	RMSE	0.552	0.555	0.552	0.545
2-Norm	PCC	0.687	0.683	0.687	0.697
	SROCC	0.682	0.675	0.683	0.700
	RMSE	0.552	0.554	0.552	0.544
5-Norm	PCC	0.687	0.684	0.687	0.699
	SROCC	0.682	0.676	0.685	0.702
	RMSE	0.552	0.554	0.552	0.543
Max	PCC	0.690	0.686	0.689	0.703
	SROCC	0.685	0.679	0.688	0.708
	RMSE	0.550	0.553	0.550	0.540

Table B.9: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Magnitude Slope.

Clock Time (sec)				7.77	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.720	0.713	0.694	0.681
	SROCC	0.720	0.715	0.694	0.661
	RMSE	0.527	0.532	0.547	0.556
2-Norm	PCC	0.720	0.713	0.693	0.681
	SROCC	0.720	0.715	0.694	0.661
	RMSE	0.527	0.533	0.547	0.556
5-Norm	PCC	0.719	0.713	0.693	0.681
	SROCC	0.720	0.715	0.694	0.661
	RMSE	0.527	0.533	0.547	0.556
Max	PCC	0.704	0.702	0.690	0.682
	SROCC	0.704	0.702	0.690	0.666
	RMSE	0.539	0.541	0.550	0.555

Table B.10: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Magnitude Intercept.

Clock Time (sec)		7.75			
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.686	0.688	0.684	0.696
	SROCC	0.673	0.673	0.674	0.694
	RMSE	0.552	0.551	0.554	0.545
2-Norm	PCC	0.686	0.688	0.684	0.696
	SROCC	0.673	0.673	0.674	0.694
	RMSE	0.552	0.551	0.554	0.545
5-Norm	PCC	0.686	0.688	0.684	0.696
	SROCC	0.673	0.673	0.674	0.694
	RMSE	0.552	0.551	0.554	0.545
Max	PCC	0.686	0.688	0.684	0.696
	SROCC	0.673	0.673	0.674	0.694
	RMSE	0.552	0.551	0.554	0.545

Table B.11: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial FISH Sharpness.

Clock Time (sec)				4.52	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.707	0.710	0.690	0.679
	SROCC	0.698	0.703	0.683	0.669
	RMSE	0.537	0.535	0.550	0.557
2-Norm	PCC	0.707	0.710	0.690	0.679
	SROCC	0.698	0.702	0.683	0.669
	RMSE	0.537	0.535	0.550	0.557
5-Norm	PCC	0.707	0.710	0.690	0.679
	SROCC	0.698	0.703	0.683	0.669
	RMSE	0.537	0.535	0.550	0.557
Max	PCC	0.703	0.706	0.688	0.681
	SROCC	0.695	0.697	0.679	0.672
	RMSE	0.540	0.538	0.551	0.556

Table B.12: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial S3 Sharpness.

Clock Time (sec)				918.54	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.758	0.719	0.723	0.700
	SROCC	0.753	0.720	0.725	0.696
	RMSE	0.496	0.528	0.525	0.542
2-Norm	PCC	0.755	0.718	0.720	0.698
	SROCC	0.750	0.720	0.722	0.694
	RMSE	0.498	0.528	0.527	0.544
5-Norm	PCC	0.750	0.718	0.716	0.695
	SROCC	0.746	0.719	0.717	0.691
	RMSE	0.502	0.529	0.530	0.546
Max	PCC	0.744	0.719	0.710	0.690
	SROCC	0.740	0.717	0.710	0.685
	RMSE	0.507	0.528	0.535	0.550

Table B.13: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial Michaelson Contrast.

Clock Time (sec)				1.37	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.683	0.683	0.683	0.713
	SROCC	0.677	0.677	0.669	0.713
	RMSE	0.555	0.555	0.555	0.532
2-Norm	PCC	0.683	0.683	0.683	0.713
	SROCC	0.677	0.677	0.669	0.713
	RMSE	0.555	0.555	0.555	0.532
5-Norm	PCC	0.683	0.683	0.683	0.713
	SROCC	0.677	0.677	0.669	0.713
	RMSE	0.555	0.555	0.555	0.532
Max	PCC	0.683	0.683	0.683	0.712
	SROCC	0.678	0.678	0.672	0.713
	RMSE	0.555	0.555	0.554	0.533

Table B.14: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial RMS Contras.

Clock Time (sec)				1.58	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.693	0.768	0.685	0.687
	SROCC	0.678	0.765	0.672	0.674
	RMSE	0.547	0.487	0.553	0.552
2-Norm	PCC	0.765	0.765	0.730	0.720
	SROCC	0.765	0.760	0.728	0.718
	RMSE	0.489	0.489	0.519	0.527
5-Norm	PCC	0.759	0.763	0.724	0.712
	SROCC	0.760	0.759	0.723	0.712
	RMSE	0.495	0.491	0.524	0.533
Max	PCC	0.683	0.745	0.683	0.681
	SROCC	0.672	0.745	0.673	0.673
	RMSE	0.555	0.506	0.555	0.556

Table B.15: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial DCT Band RMS Contras.

Clock Time (sec)				432.20	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.684	0.684	0.684	0.689
	SROCC	0.674	0.670	0.683	0.687
	RMSE	0.554	0.554	0.554	0.550
2-Norm	PCC	0.684	0.684	0.684	0.689
	SROCC	0.674	0.670	0.681	0.686
	RMSE	0.554	0.554	0.554	0.550
5-Norm	PCC	0.684	0.684	0.684	0.689
	SROCC	0.674	0.669	0.681	0.686
	RMSE	0.554	0.554	0.554	0.550
Max	PCC	0.684	0.684	0.684	0.689
	SROCC	0.675	0.669	0.681	0.684
	RMSE	0.554	0.554	0.554	0.550

Table B.16: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Spatial DCT Band Kurtosis.

Clock Time (sec)				432.20	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.725	0.726	0.688	0.713
	SROCC	0.726	0.728	0.682	0.718
	RMSE	0.523	0.523	0.551	0.532
2-Norm	PCC	0.723	0.724	0.687	0.709
	SROCC	0.722	0.722	0.682	0.716
	RMSE	0.525	0.524	0.552	0.536
5-Norm	PCC	0.714	0.713	0.686	0.702
	SROCC	0.710	0.708	0.677	0.708
	RMSE	0.532	0.532	0.552	0.541
Max	PCC	0.723	0.730	0.689	0.712
	SROCC	0.717	0.725	0.684	0.722
	RMSE	0.525	0.519	0.550	0.533

Table B.17: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure DCT Band RMS Contrast Nearest Neighbor.

Clock Time (sec)				432.20	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.683	0.684	0.684	0.688
	SROCC	0.674	0.669	0.684	0.687
	RMSE	0.555	0.554	0.554	0.551
2-Norm	PCC	0.684	0.684	0.683	0.686
	SROCC	0.674	0.669	0.681	0.686
	RMSE	0.554	0.554	0.554	0.552
5-Norm	PCC	0.684	0.684	0.684	0.686
	SROCC	0.673	0.669	0.681	0.686
	RMSE	0.554	0.554	0.554	0.552
Max	PCC	0.683	0.684	0.682	0.687
	SROCC	0.674	0.670	0.678	0.684
	RMSE	0.554	0.554	0.555	0.552

Table B.18: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure DCT Band Kurtosis Nearest Neighbor.

Clock Time (sec)				432.20	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.716	0.718	0.688	0.706
	SROCC	0.703	0.706	0.678	0.706
	RMSE	0.531	0.529	0.551	0.538
2-Norm	PCC	0.714	0.717	0.686	0.702
	SROCC	0.701	0.704	0.674	0.700
	RMSE	0.532	0.529	0.552	0.541
5-Norm	PCC	0.710	0.711	0.685	0.698
	SROCC	0.697	0.699	0.675	0.694
	RMSE	0.535	0.534	0.553	0.544
Max	PCC	0.715	0.712	0.688	0.703
	SROCC	0.702	0.699	0.678	0.698
	RMSE	0.531	0.533	0.551	0.540

Table B.19: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Temporal Standard Deviation.

Clock Time (sec)				2.54	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.705	0.693	0.693	0.685
	SROCC	0.698	0.686	0.691	0.678
	RMSE	0.539	0.547	0.548	0.553
2-Norm	PCC	0.741	0.711	0.708	0.689
	SROCC	0.735	0.702	0.709	0.685
	RMSE	0.510	0.534	0.536	0.551
5-Norm	PCC	0.782	0.752	0.719	0.689
	SROCC	0.776	0.745	0.719	0.685
	RMSE	0.474	0.501	0.528	0.550
Max	PCC	0.798	0.783	0.716	0.686
	SROCC	0.798	0.783	0.715	0.682
	RMSE	0.458	0.472	0.530	0.552

Table B.20: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Temporal Skewness.

Clock Time (sec)				6.25	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.756	0.799	0.720	0.764
	SROCC	0.749	0.788	0.725	0.767
	RMSE	0.497	0.457	0.527	0.490
2-Norm	PCC	0.687	0.689	0.684	0.693
	SROCC	0.675	0.675	0.668	0.689
	RMSE	0.552	0.551	0.554	0.547
5-Norm	PCC	0.702	0.700	0.688	0.682
	SROCC	0.698	0.693	0.683	0.671
	RMSE	0.541	0.543	0.551	0.555
Max	PCC	0.694	0.695	0.688	0.685
	SROCC	0.694	0.694	0.684	0.677
	RMSE	0.546	0.546	0.551	0.553

Table B.21: Goodness of fit between measured masked target detectability and predictions from a four input no reference linear regression model with inputs of target spatial and temporal properties, as well as mask content measure Temporal Kurtosis.

Clock Time (sec)				6.22	
		x	x^2	x/TSF	$x/(TSF+TTF)$
Average	PCC	0.695	0.696	0.684	0.714
	SROCC	0.691	0.693	0.674	0.716
	RMSE	0.546	0.545	0.554	0.532
2-Norm	PCC	0.685	0.685	0.683	0.690
	SROCC	0.668	0.669	0.673	0.690
	RMSE	0.553	0.553	0.554	0.550
5-Norm	PCC	0.687	0.688	0.686	0.682
	SROCC	0.680	0.681	0.679	0.668
	RMSE	0.552	0.551	0.552	0.555
Max	PCC	0.682	0.683	0.685	0.685
	SROCC	0.672	0.675	0.677	0.674
	RMSE	0.555	0.554	0.554	0.553

VITA

JEREMY PAUL EVERT

Candidate for the Degree of

Doctor of Philosophy

Thesis: MEASUREMENT AND ANALYSIS OF NATURAL VIDEO MASKED DYNAMIC DISCRETE COSINE TRANSFORM NOISE DETECTABILITY

Major Field: Electrical and Computer Engineering

Biographical:

- Education:

  Completed the requirements for the Doctor of Philosophy in Electrical and Computer Engineering at Oklahoma State University, Stillwater, Oklahoma in May, 2015.

  Completed the requirements for the Master of Science in Electrical and Computer Engineering at Oklahoma State University, Stillwater, Oklahoma in December, 2010.

  Completed the requirements for the Bachelor of Science in Mechanical Engineering at Kansas State University, Manhattan, Kansas, in December 2003.

- Experience

  Researcher, Computational Perception and Image Quality Laboratory (Damon Chandler) *Oklahoma State University*, 08/2012 - 05/2015

  Staff Officer/ Developmental Engineer/ Acquisitions Officer, *Oklahoma City Air Logistics Center/LR, Tinker Air Force Base, OK*, 04/2005 - 04/2008