

University of Louisville

**ThinkIR: The University of Louisville's Institutional Repository**

---

Electronic Theses and Dissertations

---

8-2011

# The nature of anisotropy in gain control pools.

Yeon Jin Kim

*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

---

## Recommended Citation

Kim, Yeon Jin, "The nature of anisotropy in gain control pools." (2011). *Electronic Theses and Dissertations*. Paper 750.  
<https://doi.org/10.18297/etd/750>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

# THE NATURE OF ANISOTROPY IN GAIN CONTROL POOLS

By

Yeon Jin Kim  
B.E. Kwangwoon University, 2004  
B.S. Kwangwoon University, 2004

A Dissertation  
Submitted to the Faculty of the  
Graduate School of the University of Louisville  
in Partial Fulfillment of the Requirements  
for the Degree of

Doctor of Philosophy

Department of Psychological and Brain Sciences  
University of Louisville  
Louisville, Kentucky

August 2011

Copyright 2011 by Yeon Jin Kim

All rights reserved

The Nature of Anisotropy in Gain Control Pools

By

Yeon Jin Kim

B.E. Kwangwoon University, 2004

B.S. Kwangwoon University, 2004

A Dissertation Approved on

June 24, 2011

by the following Dissertation Committee:

Edward A. Essock

Dissertation Director

Paul J. DeMarco

Heywood M. Petry

Patrick Shafto

## ACKNOWLEDGMENTS

So many people impacted this work that I can not begin to thank them all. But, I would like to particularly thank my mentor Dr. Edward A. Essock for his careful guidance, thoughtful patience, and support as well as for offering me the opportunity to engage in graduate study. I would also like to thank the rest of my committee members: Dr. Paul J. DeMarco, Dr. Martha Bickford, Dr. Patrick Shafto, and Dr. Heywood M. Petry for their advice and encouragement until this dissertation work was completed.

Many thanks to my friend Andrew M. Haun who was my lab member, for his continuous encouragement and as well as for frequently discussing my work with me, Dr. Hyung-Chul O. Li for his thoughtful advice and encouraging me to pursue my interests in vision when I was in college, Dr. Eli Brenner for his thoughtful advice, continuous encouragement and showing me a humble attitude as a vision scientist. And countless thanks to my friend Julia Chariker who always encourages me to follow my dream and shows me how a life is wonderful. Many thanks are due to my mom, sister, friends, and uncle for their loving support and advice through all these years. Finally, deeply thanks to God. You have given me more to explore than I could begin to understand after my many lifetimes, and I am richly blessed to have such wonderful people in my life to explore it with.

## ABSTRACT

### THE NATURE OF ANISOTROPY IN GAIN CONTROL POOLS

Yeon Jin Kim

June 24, 2011

When observers view a naturalistic ( $1/f$ ) broadband image, the various spatial components present in the image stimulate many detecting-mechanisms that suppress each other. This suppression is anisotropic, being relatively greater for mechanisms that detect horizontal components and least for oblique-detecting mechanisms (c.f., Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010), and leads to better sensitivity to and greater salience of obliquely-oriented structure when viewed in the presence of broadband background (the ‘horizontal effect’, Essock et al., 2003). It is thought that anisotropic suppression reflects a gain control mechanism whose output is intended to equalize the neural response to the orientation content biases in natural scenes.

Here we further investigate the dependence of this gain control anisotropy on temporal and spatial frequency by measuring tuning properties of these pools. Experiment 1 evaluates (1) the shape and the peak suppression of temporal frequency tuning functions for grating targets at 10 fixed spatio-temporal combinations by varying the temporal frequency of the  $1/f$  broadband masks, and (2) whether the observed suppression show the horizontal effect. Experiment 2 evaluates the same properties for spatial frequency tuning.

The results showed that multiple local gain control pools across the spatio-temporal locations are tuned in temporal frequency and spatial frequency, and all of these pools are stronger at horizontal. Surprisingly, results showed that multiple temporally-tuned pools are revealed with a broadband mask, whereas prior studies using very narrowband masks have shown only 2 (or 3) temporally-tuned channels (e.g., Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). We proposed that to drive these pools, the use of an appropriate broadband mask which can rise above a certain detection threshold of each of the underlying many, unknown, and lesser sensitive temporal detectors is crucial, and thereby can create strong suppression between detectors tuned to similar temporal rates which form local (*'tuned'*) pools.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	iii
ABSTRACT.....	iv-v
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii-xiii
1.0 INTRODUCTION.....	1-3
1.1 Natural Scenes.....	3-12
1.1.1 Content and Biases in Natural Scenes	
a. Spatial Frequency ( $1/f$ ).....	13-17
b. Orientation.....	17-18
1.2.1 Human Performance With Respect To Orientation.....	18-19
a. The Oblique Effect.....	20-23
b. The Horizontal Effect.....	23-24
1.3.1 Gain control models .....	25
a. Background	
i. Visual Masking.....	25-27
ii. The Origin of Gain Control Pool	
<i>Early Channel Models</i> .....	27-30
<i>Cross Orientation Masking</i> .....	30-32



iii. Foley’s Model of Pattern Masking.....	32-34
b. Orientation Biased Gain Control Model.....	34-38
i. Value of Efficiency: Whitening.....	39-41
1.2 Processing the Temporal Dimension.....	41-46
1.2.1 Background.....	46-51
a. 2 or 3 Temporal Mechanisms.....	51-54
<i>Third Channels?</i> .....	54-57
b. Implications of the Shape of the Temporal Sensitivity Curve.....	57-60
c. Current View of Temporal Processing.....	60-62
d. Separable Properties in Temporal and Spatial Frequency Dimensions?....	62-63
i. Psychophysical Hints.....	63-67
ii. Neurophysiological Reports: Spatial and Temporal Tuning.....	67-68
iii. Neurophysiological Reports: Cortical Map of Separable Tuning Properties.....	68-75
1.2.2 Temporal Content of Natural Scenes.....	75-76
a. Temporal Frequency ( $1/f$ ).....	76
b. Direction of Motion.....	77-78
1.2.3 Human Performance	
a. Perceptual Whitening of Temporal $1/f$ Spectra.....	78-80
b. Perceptual Whitening of Directional Anisotropy in Motion.....	81-82
i. Directional Anisotropy in Motion Perception: <i>The Oblique Effect vs.</i> <i>The Inverse Oblique Effect</i> .....	83-85

ii. Relation with Anisotropic Gain Control Processing.....	85-91
1.3 Current View.....	91-93
1.3.1. Anisotropy in Orientation Perception	
a. The Oblique Effect vs. the Horizontal effect.....	93-94
b. What Does the Significant Change in Orientation Perception Underlie?.....	94-96
c. Relation with Anisotropy in Direction of Motion.....	96-99
1.3.2. The Nature of Anisotropic Gain Control Mechanism.....	99-100
1.3.3. Goal of Present Experiments	
a. Temporal Tuning Property of Gain Control Mechanism (Experiment 1)	
i. Relation to Prior Studies.....	100-102
ii. Experiment 1.....	102-104
b. Spatial Tuning Property of Gain Control Mechanism (Experiment 2)	
i. Relation to Prior Studies .....	104-107
ii. Experiment 2.....	107-109
2.0 METHODS	
2.1 General.....	110-111
2.2 Experiment 1: Temporal Frequency Tuning	
a. Overview.....	111-112
b. Stimulus Generation for Experiment 1.....	112-114
c. Psychophysical Paradigm for Experiment 1.....	114
d. Data Analysis for Experiment 1.....	114-119

## 2.3 Experiment 2: Spatial Frequency Tuning

a. Overview.....	119-120
b. Stimulus Generation for Experiment 2.....	120-123
c. Psychophysical Paradigm for Experiment 2.....	124
d. Data Analysis for Experiment 2.....	124-125

## 3.0 RESULTS AND DISCUSSION

### 3.1 Experiment 1

a. Overview.....	126-129
b. Thresholds.....	129-130
c. Normalized Suppression: Temporal Frequency Tuning	
i. Amplitude ( $A$ ) with NSp Values.....	130-143
ii. Bandwidth ( $W$ ) with NSb Values.....	143-144
d. Summary.....	144-145

### 3.2 Experiment 2

a. Thresholds.....	145-146
b. Normalized suppression: Spatial Frequency Tuning	
i. Amplitude ( $A$ ) with NSp Values.....	146-153
ii. Bandwidth ( $W$ ) with NSb Values.....	153-154
c. Summary.....	154-155

4.0 GENERAL DISCUSSION.....	156-162
a. Why do the narrowband mask stimuli lead 2 (or 3) common peak masking effects.....	162-164
b. Does the naturalistic broadband mask create realistic response activities in detectors?.....	164-165
c. What does the temporally-tuned suppression indicate? Do the suppression and sensitivity mechanisms differ?.....	165-166
d. Do many and less sensitive temporal detectors have different detection thresholds?.....	166-167
e. Two pooling stages of the spatial and temporal signals lie in cascade.....	167-169
f. Two pooling stages of spatial and temporal signals are overlapped.....	169-171
5.0 CONCLUSIONS.....	172
6.0 REFERENCES.....	224-247
CURRICULUM VITAE.....	248-251

## LIST OF TABLES

TABLE	PAGE
1. Summary of the mask stimuli used in prior masking studies.....	173
2. Mask specifications used in prior works.....	174
3. Summary for the mask stimuli used in control conditions of Experiment 1.....	175

## LIST OF FIGURES

FIGURE	PAGE
1. Example stimuli to show the high spatial frequency bias.....	176
2. Anisotropic contrast gain control pool.....	177
3. The horizontal effect anisotropy.....	178
4. Hypothetical sensitivity of sustained and transient mechanisms.....	180
5. Temporal contrast sensitivity function for different spatial frequencies.....	181
6. Illustration of the test targets used in Experiment 1 and Experiment 2.....	182
7.1. Illustration of the temporal profiles used in Experiment 1 and Experiment 2.....	184
7.2. Total energy and DC distributions of the 2000 temporal-noise samples.....	186
8. Illustration of the mask stimuli used in Experiment 1 and Experiment 2 .....	187
9. Experimental conditions in Experiment 1 and Experiment 2.....	189
10. Contrast threshold with and without mask in Experiment 1.....	190
Same measurements as in <b>Figure 10</b> for a subject ‘TK’.....	192
Same measurements as in <b>Figure 10</b> for a subject ‘RE’.....	193
11. Suppression factor (the ration between masked and unmasked threshold) in	
Experiment 1.....	194
Same measurements as <b>Figure 11</b> for a subject ‘TK’.....	196
Same measurements as <b>Figure 11</b> for a subject ‘RE’.....	197
12A. Normalized suppression (NSp) for comparing magnitude in Experiment 1.....	198

12B. Plotting on different scales for condition 8 and condition 10 in Experiment 1.....	200
13. Gaussian curve fitting parameters shown in <b>Figure 12A</b> in Experiment 1.....	201
14. Control conditions: Temporal tuning plotting on the log scale in Experiment 1.....	202
15. Control conditions: Temporal tuning on the linear scale in Experiment 1.....	203
16. HEI, VEI, & H-VEI index with NSp values in Experiment 1.....	204
17. Normalized suppression (NSb) for comparing bandwidth in Experiment 1.....	205
18. Bandwidth across orientations and across conditions in Experiment 1.....	207
19. Gaussian curve fitting parameters shown in <b>Figure 18</b> in Experiment 1.....	208
20. Contrast thresholds with and without masks in Experiment 2.....	209
Same measurements as <b>Figure 20</b> for a subject ‘TK’.....	211
Same measurements as <b>Figure 20</b> for a subject ‘EJ’.....	212
21. Suppression factor in Experiment 2.....	213
22. Difference between broadband and 0.8-octave band masking in suppression.....	215
23. Normalized suppression (NSp) for comparing magnitude in Experiment 2.....	216
24. Gaussian curve fitting parameters shown in <b>Figure 23</b> in Experiment 2.....	218
25. HEI, VEI & H-VEI index with NSp values in Experiment 2.....	219
26. Normalized suppression (NSb) for comparing bandwidth in Experiment 2.....	220
27. Bandwidth across orientations and across conditions in Experiment 2.....	222
28. Gaussian fitting parameters shown in <b>Figure 27</b> in Experiment 2.....	223

## 1.0 INTRODUCTION

One of the most important concepts underlying our understanding of vision is that encoding-mechanisms<sup>1</sup> in the different visual stages selectively respond to visual features falling within a region of the visual field known as their receptive field<sup>2</sup>. In higher visual stages, the visual features that many encoding-mechanisms respond to become more complex and larger in size. In particular, encoding-mechanisms in the retina or LGN are selective for small spots of light (Kuffler, 1953). In V1 they respond highly to oriented contrast patterns at different spatial scales, motion flickers and directions of motion (Hubel & Wiesel, 1959; Hubel & Wiesel, 1968; Tolhurst & Movshon, 1975; Albright, 1984). In the later visual areas, the encoding-mechanisms process relatively complex spatial forms and motions rather than simple contrast. Specifically, in V2 the mechanisms are selective for contours (Ito & Komatsu, 2004). In V4, they are selective for outline-shapes, two-dimensional shapes and colors (Pasupathy, 2006). In IT (Inferotemporal Cortex), the mechanisms are selective for faces and objects (Logothetis, Pauls, & Poggio, 1995), and in MT (Middle Temporal Cortex) the mechanisms prefer to encode various directions of motion, pattern motions, and orientations (Van Essen & Gallant, 1994; Pack, Conway, Born, & Livingstone, 2006; Pack & Born, 2008).

---

<sup>1</sup> It is also referred to cells, analyzers, detectors, channels, filters, or detecting mechanisms, mechanisms by various scientists

<sup>2</sup> The receptive field can be defined as a small region of the visual field in which the presence of a stimulus will alter the firing of that cell. The receptive field of a single photoreceptor cell in the retina, for example, can be said to be limited to the tiny spot of light, within your field of vision, that corresponds to this photoreceptor's precise location on your retina. But in each succeeding visual pathways from retina via LGN to cortex, the receptive fields become increasingly complex, and they become even more complex when it comes to the neuron of the higher visual cortex (e.g., IT and MT) as described in text.



This cortical hierarchy increasing in receptive field complexity and size of preferred stimuli is consistent with encoding-mechanisms at each stage which refine the selectivity of their input stage by pooling (or combining) similar input (i.e., similar orientation, spatial frequency, temporal frequency, and direction of motion etc) from the preceding stages in order to remove redundancy present in visual input, but it occurs within a given stage<sup>3</sup> (Barlow, 1961). By removing redundancy at each stage, receptive fields allow the cells to efficiently encode the local (small) visual features (e.g., oriented edges of fine lines, etc) present in the visual field at the early stage (V1), and further into higher stages they become more concerned about the global visual field consisting of local visual features. As a result, their ability to encode the fine-detailed features present in the global visual field might become gradually poorer, whereas their ability to encode at least some types of complex global objects, such as faces and cars, would increase (c.f., Dakin, 2009). Through this sequence of complex encoding processes at a sequence of visual stages, we can finally get a coherent and meaningful visual perception about the natural environment.

With respect to the orientation encoding at the early cortical stage (V1), one of the well-documented efficient coding strategies would be an anisotropically-weighted gain control pooling process which can be achieved by mutual suppression between units tuned in similar orientations of content in natural scenes (i.e., Essock et al., 2003; Hansen & Essock, 2004; 2005; 2006; Essock, Haun, & Kim, 2009; Haun & Essock, 2010; Kim, Haun, & Essock, 2010; Kim, Haun, & Essock, 2011). Specifically, this suppression is relatively greater for units that detect horizontal content and least for oblique-detecting

---

<sup>3</sup> It has long reported the presence of connectivity that supports not only feed-forward, but also feedback activity in the visual system (see Sillito, Cudeiro, & Jones, 2006).

units which form a gain control mechanism (pool) whose output is intended to equalize the neural response to the orientation content bias in natural scenes (i.e., the most prevalent horizontal content in Hansen & Essock, 2004; 2005). By removing redundancy (the prevalent horizontal content), the incoming inputs, such as local oriented-contrast features within a local region of the visual field, can be efficiently pooled (combined) and then these processed signals at a given stage can be efficiently transported to the higher visual stages (V2). Here we further investigate the nature of this anisotropy in a gain control pool by measuring the temporal (Experiment 1) and spatial (Experiment 2) frequency tuning properties of these pools across the different spatio-temporal combinations, and assessing at which combinations of spatial and temporal frequency the horizontal effect exists.

### 1.1 Natural Scenes

Encoding-mechanisms underlying human vision at the early stage (i.e., the primary visual cortex or V1) respond to contrasts in natural scenes at a relatively wide range of orientations, spatial frequencies, temporal frequencies, and directions of motion within small regions of the visual field (Hubel & Wiesel, 1959; Tolhurst & Movshon, 1975; Albright, 1984). With respect to the encoding manner, Barlow (1961) hypothesized that the visual system develops in response to various visual features from the natural stimuli it is exposed to and does so in a manner that is efficient-- specifically, by removing redundancy present in natural stimuli. This enables the visual system to transmit as much visual information as possible through its limited information capacity, termed as the

efficient coding hypothesis (Barlow, 1961) which was influenced by information theory<sup>4</sup> (Shannon, 1948). This well-documented hypothesis has been widely used as a main paradigm for understanding neural encoding processes that underlie visual systems, which grants significant insight into the relationship between natural scenes and encoding processes underlying the visual mechanisms.

Recently, many of the works have shown that many encoding-mechanisms and processing strategies in the visual system are optimally suited to encode the visual features present in the natural scenes in several ways, but the agreement across works is increasing efficiency by removing redundancy in order to maximize information transfer strategies. Firstly, neural networks and computational models demonstrate that arrays of linear neurons with properties that are found in the primary visual cortex (V1) appear to maximize the 'sparse response' (i.e., removing redundancy, e.g., Barlow, 1961) to natural scenes. Specifically, it was shown by demonstrating a learning algorithm that attempts to find sparse linear codes for natural scenes (i.e., Olshausen & Field, 1996). Secondly, by removing redundant visual information in natural stimuli through the 'nonlinear gain control sensory processing' which has been termed as the gain control mechanism. According to this idea, the statistical properties of natural stimuli are thought to be efficiently processed by a bank of linear detectors (i.e., V1 simple cells). Each detector's response is half-wave rectified and squared, and then divided by a weighted sum of the squares of the rectified responses of its neighboring detectors (i.e., pooling by similar

---

<sup>4</sup> Information theory (Shannon, 1948) provided the mathematical framework for analyzing communication system. It formally defines concepts such as information, channel capacity, and redundancy. Barlow's hypothesis treats the sensory pathway as a communication channel in which sensory spiking is an efficient code for representing sensory signals. The spiking code aims to maximize available channel capacity by minimizing the redundancy between representational units. Both the efficient coding hypothesis and information theory share a key prediction of the efficient coding hypothesis is that sensory processing in the brain should be adapted to natural stimuli.

orientation and spatial frequency in order to remove redundant visual information present in natural stimuli, see Wainwright, Schwartz, and Simoncelli, 2001; Schwartz and Simoncelli, 2001; Essock et al., 2003; Hansen & Essock, 2004). Finally, the Gestalt law of 'good continuation' (i.e., it has been explained that a variety of phenomena demonstrate the importance of continuity in human perception) provides a plausible insight. This suggested that various visual features present in natural stimuli are grouped by similar features to construct meaningful shapes (i.e., Marr, 1982; Field, Hayes, & Hess, 1993; Hess & Dakin, 1997). In this notion, the responses of detectors tuned in various visual features within a small region of the visual field are combined (or grouped/integrated/pooled) by similar visual properties, such as similar orientations, movement details and spatial scales (i.e., Bex & Dakin, 2002; Bex, Simmers, & Dakin, 2003; Dakin, 2009).

More recently, by examining images or movies from the natural environment, several researchers continuously related their statistical regularities to the encoding processes underlying the performance of various visual tasks in terms of spatial frequency (Field, 1987; Brady & Field, 1995; Cass, Alais, Spehar, & Bex, 2009), orientation (Essock et al., 2003; Hansen & Essock, 2004; Essock et al., 2009; Kim, Haun, Essock, 2010; Haun & Essock, 2010; Kim, Haun, & Essock, 2011), temporal frequency (Dong & Atick, 1995), and local direction of motion (Dakin, Mareschal, & Bex, 2005) in order to understand the neural mechanisms underlying these encoding processes and perception. The common results across these works suggest that the statistical regularities of images and movies from the natural environment are especially germane to the neural encoding underlying human visual systems and visual perceptions, and it therefore

warrants wider study.

Informatively, images and movies of the natural environments are distinctive because they contain particular types of spatial and temporal structures which have been referred to as the statistical regularity. Thus, in order to understand the typical behavior (i.e., encoding processes and strategies) of these mechanisms, we must first study the structure of natural scenes. Indeed, several works on the statistics of images (i.e., Field, 1987; Burton & Moorhead, 1987; Ruderman, 1997; Tolhurst & Tadmor, 1997) and movies (i.e., Dong & Atick, 1995; Dakin et al., 2005) from the natural environment show that images and movies have remarkably constrained statistical regularities. With respect to the 'spatial frequency' (e.g., spatial scale) of typical natural scenes, it is reported that the amplitude spectrum (ignoring orientation) across images of the natural environments, which is relatively consistent across scenes, falls approximately as ' $1/f^\alpha$ ', where  $f$  is spatial frequency, and the exponent  $\alpha$  is approximately 1.0 (Field, 1987; Tolhurst & Tadmor, 1997; Simoncelli & Olshausen, 2001). This indicates that the amount of amplitude spectrum of any natural world image peaks at the larger scales (low spatial frequency) and decreases with increasing scales (high spatial frequency) of the image content. With respect to the 'temporal frequency' (motion details can be decomposed by the fundamental temporal frequency components through the Fourier analysis technique) of the natural movies, it was reported that the same  $1/f$  spectral property was also found (Dong & Atick, 1995). These findings indicate that the images and movies from the natural environments are dominated by particular spatial and temporal components which are lower spatial and temporal frequencies (i.e., the  $1/f$  spectral property indicates that the low frequency components in natural images are most redundant). Further details of these

$1/f$  properties with respect to spatial and temporal frequency contents will be described in Section 1.1.1.a and Section 1.2.2.a, respectively. In addition, with respect to the orientation (i.e., natural scenes typically contain a wide range of the 'orientation' contents), it is reported that the amplitude spectrum (ignoring spatial frequency) across natural scenes is typically anisotropic (c.f., Hansen & Essock, 2004; 2005; 2006). Specifically, the amplitude is most for horizontal, intermediate for vertical, and is least for oblique orientations. This indicates that natural scenes are indeed dominated by the horizontally oriented contents (the horizontal content bias). Finally, with respect to the local motion directions (i.e., the local motion directions, such as up/down, left/right, and off-cardinal diagonals) of the natural movies (or time-varying natural images), it was reported that the local spectral distributions at the different motion directions are typically anisotropic (Dakin et al., 2005). Specifically, the local motion power spectrum at the horizontal and vertical directions is relatively much higher and broader than at oblique directions of motion. This also indicates that the spectral distributions of the motion directions are also biased (i.e., it seems to be biased in horizontal and vertical motion directions). Note that the statistical regularities of the natural environment have been described by the power or amplitude spectra (i.e., the power spectrum is equal to the square root of the amplitude spectrum). Further details with respect to the anisotropy in the orientation contents of the natural scenes (i.e., Switkes, Mayer, & Sloan, 1978; Van der Schaaf & van Hateren, 1996; Coppola, Purves, McCoy, and Purves, 1998; Hansen & Essock, 2004; 2005; 2006), and in the motion directions of the natural movies (i.e., Dakin et al., 2005) will be described in Section 1.1.1.b and Section 1.2.2.b respectively.

Next, we can then relate these well-documented statistical regularities to the

mechanism of neural encoding underlying the visual perception by measuring behavioral performances (e.g., sensitivity or salience, etc.) for the particular stimuli resembling the statistical properties of the natural stimuli. Specifically, the  $1/f$  property is very useful because we can say confidently that these images stimulate the encoding-mechanisms in a way that is similar to natural stimulation. When you look at a natural movie, you are exposing the visual system to many spatial frequencies, orientations, temporal frequencies, and directions of motion simultaneously. Further, all of those frequencies, orientations, and directions of motion have a relatively consistent relationship in terms of their relative amplitudes (or contrasts/powers). Thus, the same would be true if you expose the visual mechanism to a stimulus resembling the statistical properties of the typical natural scenes (e.g.,  $1/f$  property), since it contains the same frequencies, orientations, and directions of motion, at the same relative amplitudes. In order to assess this issue psychophysically, several works have extensively investigated how the mechanisms of neural encoding underlying the performance of visual tasks represent these statistical regularities. One agreement across these studies is that the observers always underestimate the prevalent visual structures, such as low spatial frequencies (Cass, Alais, Spehar, & Bex, 2009), low temporal frequencies (Cass et al., 2009), cardinal orientations (Essock et al., 2003; Hansen & Essock, 2004; 2005; 2006; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim et al., 2011), and cardinal directions of motion (Dakin et al., 2005; Wilson, Loffler, Wilkinson, & Thistlethwaite, 2001) in order to compensate for the comparatively less prevalent visual structures, such as high spatial frequencies, and high temporal frequencies, and oblique orientations, and oblique directions of motion. In this view, further psychophysical studies will be described,

specifically with respect to the spatial frequency in Section 1.1.1.a, and orientation in Section 1.2.1, and temporal frequency in Section 1.2.3.a and direction of motion in Section 1.2.3.b. Critically, what occurs to cause the perceptual underestimation to the most dominant low spatial/temporal frequencies and cardinal directions of motion.

With respect to the orientation dimension, we have suggested a plausible model (Essock et al., 2009; Haun & Essock, 2010) which well describes one fundamental question of how efficiently the horizontal content bias present in natural scenes is processed by the mechanisms of neural encoding underlying the orientation perception. Much of our understanding of orientation processing comes from studies that measure behavioral performance for simple sinusoidal grating, which is very narrow in spatial frequency and orientation and present at uniform contrast (Appelle, 1972; Blackmore & Campbell, 1969; Campbell & Kulikowski, 1966; Movshon & Blackmore, 1973, Essock, 1980). The typical finding across these works is that visual performance (e.g. sensitivity) is constantly superior to the horizontal and vertical orientations relative to the oblique orientations, termed as the oblique effect (Appelle, 1972). Current understanding of orientation processing, however, comes from studies that measure the behavioral performance for visual stimuli resembling the statistical regularities of the natural scenes. It is reasonable to assume that these stimuli stimulate the encoding-mechanisms underlying the human visual perception in a way that is similar to natural stimulation (Bex et al., 2007; Olshausen & Field, 2005; Hansen & Essock, 2004) and this assumption is also supported by several neurophysiological studies (Baker & Issa, 2005; Basole et al., 2003). Thus, we can understand more realistic orientation processing which occurs in everyday viewing situation. In order to investigate this, the orientation perception was



measured using by conditionally manipulating visual stimuli either narrowband (single gratings or line) or naturalistic ( $1/f$ ) broadband stimuli and the significant change in orientation perception was observed. Specifically, when viewing a naturalistic image, visual performance is best for oblique orientations and worse for horizontal than vertical, termed as the horizontal effect (Essock et al., 2003; Hansen & Essock, 2004; 2006; Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010; Kim et al., 2011). Whereas when viewing a narrowband pattern (single grating), most observers show an oblique effect, and if more content is added to the image, the anisotropy is the horizontal effect (Hansen & Essock, 2006). In order to understand what underlies the now well-documented significant change in orientation perception (i.e., from the oblique effect to the horizontal effect), we have suggested a model of the anisotropic gain control pool (Essock et al., 2003; Hansen & Essock, 2004) whose effect is pronounced when viewing a pattern of broadband spatial content in Section 1.3.1.b. We suggested that the additional content creates more response activities in channels which contribute to the gain control pool, whereas the response activities in channels might not significant with narrowband stimuli (Hansen & Essock, 2006; Essock, et al., 2009; Kim et al., 2010; Kim et al., 2011). Such anisotropic suppression in the gain control pool could be a direct consequence of corresponding biases of cortical neurons (Li et al., 2003), indicating more neurons tuned to horizontal than other orientations and vertical than oblique orientations. In addition, the orientation biases of the typical natural scenes (Hansen & Essock, 2004) indicate that there are the most physical energy at the horizontal, least at the obliques, and intermediate at the vertical orientation. Based on this well-documented statistical property, we have suggested that the horizontal effect would be linked to the anisotropic

gain control pool which is a very efficient coding strategy of the early visual processing stage (V1) that serves to equalize the neural representation of natural scenes with respect to orientation (i.e., Essock et al., 2003; Hansen & Essock, 2004). Specifically, this anisotropic gain control pooling process equalizes the prevalent contents (i.e., the horizontal > vertical > oblique orientations bias) in typical natural scenes as increasing suppression at the horizontal orientation results in less sensitive performance at the horizontal. From this point of view, the horizontal effect anisotropy would be a good example to show that the encoding-mechanisms underlying the visual system are likely to have evolved to reflect natural stimuli, and in order to equalize the prevalent content within the natural scenes -- a specific optimizing processing strategy to “whiten” the neural response to the orientation content biases in natural scenes. By removing redundancy (the prevalent horizontal content) through the anisotropically-weighted gain control pooling process (strongest at the horizontal) at the early orientation processing stage (V1), the incoming local contrast information within a small region of the global visual field can be efficiently pooled. These pooled signals can be then efficiently transported to their higher visual stages. This conditional pooling process can be achieved by mutual suppression between channels (Foley, 1994) in Section 1.3.1.a.iii in which each channel’s response is divided by a gain control pooling signal which comes from its neighboring orientations and spatial frequencies. This is part of a process of gain control that serves to equalize the orientation biased natural scenes. From these points of view, the significant change in orientation perception would be a great example to show that the encoding-mechanisms underlying the visual system or the human visual performance are likely to have evolved to reflect natural scenes, and in order to equalize the prevalent

content (by removing redundant horizontal contents) within the natural scenes – specific optimizing processing strategy to “whiten” the neural response to natural scenes via anisotropically-weighted gain control pool. By doing this process our visual system can take two advantages: (1) the incoming input, such as local contrast features within a small region of the visual field, are conditionally pooled (i.e., with different weighted-suppression at different orientations) and then the processed signals at a given stage can be efficiently transported to the higher stages, and (2) the perceptual visibility to the comparatively less dominant oblique contents would be enhanced.

Current work will mainly explore the nature of the anisotropy in the gain control pools with respect to the temporal frequency and spatial frequency by measuring the tuning properties of these pools. Specifically, the results with respect to the temporal frequency demonstrated the existence of multiple local gain control pools at different spatial and temporal surface, and all show the horizontal effect anisotropy. Whereas prior studies (i.e., Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) have used very narrowband masks and shown only two or three temporally tuned channels in Section 1.2.1, current work (Experiment 1) shows that multiple tuned temporal mechanisms are revealed with broadband masks. In this view, further descriptions (i.e., why could prior studies only find 2 (or 3) temporal detectors?) will be specifically addressed in Section 1.3. Before moving to the Method Section, several relevant studies with the current work will be described in order to provide theoretical and general backgrounds.

### 1.1.1 Content and Bias in Natural Scenes

#### a. Spatial Frequency ( $1/f$ )

In the past two decades, numerous researchers have examined the statistics of natural images to understand their particular characteristics, using various image analysis techniques. The Fourier analysis is one of the most widely-spread image-analysis techniques which provides contrast (corresponding to amplitude or power spectrum in Fourier domain) distribution at various ranges of spatial frequencies and orientation, which is termed as the 2-dimensional amplitude spectrum (i.e., the amplitude spectrum which is defined as the square root of the power spectrum). Specifically, the Fourier domain represented through the Fourier transform has its origins in mathematical studies (i.e., Fourier theory) of periodic function (e.g., sinusoidal gratings with different amplitudes, frequencies, orientations, and phases). From the Fourier transform of an image, one obtains the 2-dimensional amplitude spectrum which clearly demonstrates amplitude distribution as a function of spatial frequency and orientation. We can, thus, compare contrast (or amplitude or power) distribution of one image with other images accurately.

The amplitude spectrum (i.e., ignoring orientation) across natural scenes, which is relatively consistent across scenes, falls approximately as ' $1/f^\alpha$ ', where  $f$  is spatial frequency and the exponent  $\alpha$  is approximately 1.0 (Field, 1987; Tolhurst & Tadmor, 1997; Simoncelli & Olshausen, 2001). The exponent, ' $\alpha = -1.0$ ' has been considered one of the typical properties of natural images. A number of studies have provided that the amplitude spectrum falls with frequency ( $f$ ) by a factor of about  $1/f$  (hence, the power spectrum is equal to the square root of the amplitude spectrum,  $1/f^{-2}$ ). However, there are

somewhat deviations for the exact exponent,  $\alpha$ , but the general agreement for the slope across natural images has been that of an observed range from -0.6 to -1.6 (Field & Brady, 1997). Interestingly, this  $1/f$  property was also found in the temporal frequency dimension (Dong & Atick, 1995) which will be discussed later.

This  $1/f$  property showing across natural scenes has been used to develop a model of spatial frequency processing of the early visual system (V1) by Field (1987). It is widely known that the human vision at the early stage consists of arrays of encoding-mechanisms (or cells, channels, detectors, sensors or filters by others) which are selective to 'local' regions of space, spatial frequency and orientation. At given knowledge, Field (1987, 1993) raised one interesting question of how this ' $1/f$  regularity is represented and processed by the human visual system. This question was approached by looking at the  $1/f$  structural regularity of natural scenes and observing how this structure relates to the visual system's encoding of spatial patterns. To predict plausible answers, Field (1987) proposed that natural scenes are approximately scale invariant with regard to their contrast across scale. For example, if an image magnifies any particular region, one is not likely to find an increase or decrease in the amount of contrast at any scale. In a photograph, the contrast can be measured by the variance in pixel luminances (here taken as the standard deviation of pixel luminances, particularly it can be described to root-mean-square (RMS)) and if the scale invariance idea is true the variance should remain roughly constant independent of the viewing distance. Field (1993) suggested that the variance is roughly constant as one magnifies any local region of the image. It thus can be expected that one should not expect to find either more or less variance in the image as one magnifies a particular region. This invariance in contrast across scales was well

described in terms of the amplitude or power spectrum of the image (Field, 1987).

Previously, Field (1987) demonstrated that the  $1/f$  pattern has equal spectral amplitude in each multiplicative spatial frequency interval by Parseval's theorem. In this theorem, each octave of spatial frequency content in a  $1/f$  structural image will have the same contrast by a root-mean-square (RMS) contrast (i.e., this idea has been used when we construct stimuli, see Essock et al., 2009; Haun & Essock, 2010). This again implies that if the amplitude of images fall-off as  $1/f$ , there is equal amplitude in equal octaves. In this sense that scaling all frequencies by a factor (e.g., moving forward or backward in an image) has little effect on the shape of the amplitude spectrum, which was initially referred to the scale invariant property (Field, 1987). For example, the total energy between 1 and 2 cpd is equal to the energy between 8 and 16cpd because these two bands have equal octave intervals.

Subsequently, this idea of scale invariant in contrast across scales (Field, 1987; 1993) showing equal energy in equal octaves has been related to the 'contrast constancy' (Georgeson & Sullivan, 1975) to describe the finding that 'high' contrast gratings of different frequencies are perceived as equal in contrast when they are roughly equal in physical contrast (Brady & Field, 1995). Brady and Field (1995) pointed out one problematic issue with respect to the contrast sensitivity function. In particular, this problem occurs from the attempt to generalize from threshold to suprathreshold sensitivity. At threshold, sensitivity to stimuli (generally sinusoidal gratings) peaks around 4cpd and shows a significant drop at both lower and higher frequencies. However, sensitivity at high contrast (suprathreshold contrast), when measured by contrast matching shows a much flatter function, termed as the 'contrast constancy' (Georgeson &

Sullivan, 1975). Brady and Field (1995) assumed that this invariance of suprathreshold sensitivity across frequency has been presumed to underlie a more interesting phenomenon. Specifically, our perception of contrast at different frequencies in the natural environment remains fairly stable across changes in viewing distances or scales. However, it has not always been made clear how the energy of such scales change with viewing distance and whether or not this idea of equal energy in the equal octave would be conserved. In this view, some apparent problems regarding our perception of natural world scenes will occur. The amplitude spectra of typical natural scenes are indeed not flat, but  $1/f$  in amplitude spectra being decreased with increasing spatial frequency (i.e., low spatial frequency bias). As described above, if the contrast energy in the natural environment is scale invariant (i.e. the contrast is independent of scale or viewing distance), the energy of natural scene would be flat. Instead, at suprathreshold contrast sensitivity in perception, however, reveals the flat, and the similar results (i.e., the flat sensitivity at suprathreshold (0.3RMS contrast) across spatial frequencies) are also demonstrated by Cass, Alais, Spehar, and Bex (2009). In addition, a flat sensitivity function can not explain why our perception of contrast remains invariant when an object scales in size.

To solve this problem, Brady and Field (1995) provided a model of broadband contrast sensitivity in which mechanism bandwidths increase with frequency (constant in octaves) to effectively equalize its perceptual response to low spatial frequency biased naturally occurring input ( $1/f$ ), a phenomenon is referred to as “whitening” (Brady & Field, 1995; 2000; Field & Brady, 1997). This implies that the high spatial frequency energy will be overestimated by our perception whereas the low spatial frequency energy

will be underestimated in order to equally perceive all spatial patterns. It can be easily demonstrated to observers. For example, consider the three images shown in **Figure 1**. On the left is a white noise (i.e.,  $1/f^0$  which is physically flat spectra), in the middle is a  $1/f^1$  noise, and on the right is a  $1/f^2$  noise. We can easily notice that the white noise pattern (physically flat spectrum) appears to be dominated by high frequency structures while the  $1/f$  pattern appears to have structures at a variety of scales (perceptually balanced). The  $1/f^2$  pattern appears to be the low spatial frequency dominant. This example clearly demonstrates that the high spatial frequency energy is overestimated, and the low frequency energy is underestimated by our perception. In this view, similar results are also reported by Cass et al. (2009). Cass et al. (2009) suggested that this spatial whitening might confer both computational and metabolic advantage by equalizing its response to the 'low frequency bias' ( $1/f$ ) present in natural images, thereby removing input redundancy (i.e., low spatial frequency bias) and increasing informational efficiency (Barlow, 1961; 2001; Simoncelli & Olshausen, 2001). Informatively, any images and movies of the natural environments exhibit  $1/f$  in their amplitude spectra in both spatial frequency and temporal frequency dimensions. This also raise a question of whether there exists a temporal whitening process analogous to that observed in the spatial frequency dimension and further details will be considered in Section 2.2.2.a.

#### b. Orientation

The amplitude spectrum (i.e., ignoring spatial frequency) across natural scenes is typically anisotropic, showing the horizontal effect anisotropy. Specifically, the amplitude



is most for horizontal, intermediate for vertical, and least for oblique orientations (i.e., the analyses of the orientation averaged amplitude spectra typically shows  $1/f$  characteristics). This anisotropy of natural scenes is firstly examined by Switkes, Mayer, and Sloan (1978). Switkes et al. (1978) examined images using the Fourier analysis and found that a significant bias of power (i.e., the power is equal to the square root of the amplitude spectrum) at the cardinal (i.e., horizontal and vertical) orientations exists, and also this anisotropy regardless of whether the image content was pure natural outdoor scenes or (indoor or outdoor) carpentered scenes. The same results were demonstrated by Coppola, Purves, McCoy, and Purves (1998) who also examined images from a similar category as the Switkes et al. (1978) and found a predominant orientation bias in the horizontal and vertical orientations. Several researchers (Van der Schaaf & van Hateren, 1996; Hansen & Essock, 2004, 2005; Oliva & Torralba, 2001) also reported the predominant horizontal and vertical spectral bias through the Fourier analysis, specifically, more horizontal than vertical (Hansen & Essock, 2004; 2005; 2006).

More recently, this strong statistical regularity demonstrated across natural scenes is used to relate neural processing and visual perception in order to predict one crucial issue of how efficiently this orientation bias is represented by encoding-mechanisms underlying the human performance of visual tasks.

### 1.2.1 Human Performance With Respect To Orientation

Orientation is one of the fundamental dimensions in natural scene perception, and has been studied for almost half a century. Much of our prior understanding of orientation processing comes from studies by measuring the behavioral performance for simple

sinusoidal gratings that are narrow in spatial frequency and orientation content and present at uniform contrasts, which can assume to create only individual detector's response (Blackmore & Campbell, 1969; Campbell & Kulikowski, 1966; Movshon & Blackmore, 1973; Essock, 1980; Daugman, 1984; Phillips & Wilson, 1984). Whereas, current understanding of orientation processing comes from studies by measuring the behavioral performance for visual stimuli that resemble the statistical regularities of natural scenes, and they can create comparatively more realistic complex responses from population detectors (i.e., natural scenes typically contain broad ranges of spatial frequencies and orientations which can assume to stimulate multiple detectors' responses). Critically, the significant change in orientation perception--i.e., from the oblique effect (Appelle, 1972) to the horizontal effect (Essock et al., 2003)-- was obtained as adding more content to the narrowband stimulus (Hansen & Essock, 2004). Specifically, when narrowband content such as a resolution-acuity stimulus is viewed, visual performance for oriented content is best for horizontal and vertical contents and worst for oblique orientation contents. If more content is added to this narrowband content stimulus, the anisotropy in orientation perception shows a horizontal effect that visual performance for oriented broadband content is best for oblique orientations, and worst for horizontal and vertical (Essock et al., 2003; Hansen & Essock, 2004; 2006). To predict underlying neural processing and visual perception, a series of studies has been provided, and one hypothetical model which well describes the significant change in orientation perception has been suggested in our lab.

### a. The Oblique Effect

The classical oblique effect (Appelle, 1972) refers to the finding that visual performance such as a resolution acuity task is much better for *narrowband* stimuli (e.g., gratings and lines) oriented horizontally or vertically than for grating stimuli oriented at oblique orientations. This anisotropy in orientation perception has been termed as the oblique effect (Appelle, 1972). Subsequently, numerous studies using various tasks demonstrated the oblique effect which implied that the oblique effects obtained on these various tasks are all considered equivalent. To clarify this, this oblique effect was specified with the tasks which divided it into two classes (Essock, 1980). One class of oblique effect was obtained with measures which reflected the basic function of the visual system (e.g., contrast sensitivity, acuity and discrimination) and termed as the class 1 oblique effect corresponding to the classical oblique effect. The term class 2 oblique effect referred to oblique effects obtained on paradigms which reflected later stimulus processing and memory by measuring reaction time (RT) for oriented line targets ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ ) with different response tasks (i.e., identification, detection, and classification). The results showed that for detection and classification tasks, observers responded equally quickly to all orientations (i.e., mean RT = 514ms). In a task requiring identification of orientation, observers didn't respond equally quickly to all orientations (i.e. mean RTs to the obliques ( $1095\text{ms}$  and  $1098\text{ms}$  for  $45^\circ$  and  $135^\circ$  respectively) was 188 ms longer than to the average of the horizontal ( $888\text{ms}$ ) and vertical ( $929\text{ms}$ ) orientations). These results indicate that the source of the class 2 oblique effect obtained in RT tasks is greater confusability between the  $45^\circ$  and  $135^\circ$  oblique lines than between the horizontal and vertical lines when identification of orientation is required (Essock, 1980). In other words,

the class 2 oblique effect seems to include the inclination to better remember and more efficiently code information relating to vertical and horizontal orientations, thus this is thought to be more cognitive or higher level in nature.

The classical oblique effect (the class 1 oblique effect, Essock, 1980) in orientation perception was systematically tested using narrowband grating stimuli with various spatial and temporal frequencies, and it was reported that the magnitude (i.e., cardinal/oblique sensitivity) of the oblique effect increased with increased spatial frequency (Campbell et al., 1966; Camisa, Blake, & Lema, 1977) and decreased with both increased temporal frequency (Camisa et al., 1977; Essock, 1982) and increased eccentricity (Berkley, Kitterle, & Watkins, 1975).

What are the bases of the anisotropy in orientation processing? It is generally agreed that the oblique effect results from neural bias, not retinal (Vogels & Orban, 1986) (i.e., the optical characteristics, such as astigmatism, dioptric effect, or diffraction) and eye-movements (Weymouth, 1959) do not explain this effect because it is still found when the contributions of these variables are eliminated (Nachmias, 1960; Campbell et al., 1966). Campbell and Maffei (1970) found that the amplitude of the visual evoked potential recorded from the occipital scalp was greater with horizontal and vertical stimuli than with oblique ones, but not in the electroretinogram. De Valois et al. (1982) carried out physiological single cell studies in the primary visual cortex (V1) of the monkey, and found that there are more cells tuned to horizontal and vertical orientations than oblique orientations. The fMRI studies on humans showed an increased response to horizontal and vertical orientations (i.e., Furmanski & Engel, 2000; Furmanski, Schuppeck, & Engel, 2004). There seems to be a good match between the oblique effect

anisotropy measured psychophysically and the response of neurons in the primary visual cortex (V1). However, opposite results indicating a decreased response to horizontal orientation, an intermediated response to vertical orientation, and the highest response to the oblique orientations—the classic horizontal effect— were recently reported (Mannion, McDonald, & Clifford, 2010). In order to (1) clarify the discrepancy between these two groups of fMRI results (i.e., Furmanski & Engel, 2000; Furmanski, Schuppeck, & Engel, 2004; Mannion et al., 2010) and (2) further speculate the cortical representation of the orientation structure of natural environments, additional follow-up fMRI study is highly encouraged.

Furthermore, Li, Peterson, and Freeman (2003) have used a database to analyze cell populations in the cat's striate cortex to determine possible difference as a function of orientation. They found that both the number of cells and the widths of orientation tuning vary as a function of preferred orientation. Specifically, more cells prefer horizontal than vertical orientations, and comparatively fewer cells prefer oblique orientations (i.e., horizontal > vertical > oblique orientation bias) (Maffei & Campbell, 1970; Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; De Valois, Yund, & Helper, 1982). Orientation tuning widths are most narrow for cells preferring horizontal orientations. These findings are most prominent for simple cells in V1 tuned to high spatial frequencies.

Surprisingly, one intriguing correspondence is that there is more information in natural scenes at horizontal and vertical orientations than oblique orientations (Coppola et al., 1998; Hansen et al., 2003; Hansen & Essock, 2004; 2005; 2006) and horizontal more prevalent than vertical (Hansen & Essock, 2004). This effect is even stronger in human

made environments, but it is still manifest in the natural environment (Hansen et al. 2003; Hansen & Essock, 2004). These series of findings clearly support that the mechanisms of neural encoding underlying the performance of visual tasks are likely to have developed to reflect natural scenes. Specifically, there are more cells tuned to cardinal orientations than oblique orientations at the early visual stage (V1) which is surprisingly matched with the image content bias presented in natural scenes, favoring the cardinal than the oblique contents.

#### b. The Horizontal Effect

Much of our understanding of orientation processing comes from studies using simple sinusoidal grating stimuli that are narrowband in spatial frequency and orientation content and present at uniform contrast (Blackmore & Campbell, 1969; Campbell & Kulikowski, 1966; Movshon & Blackmore, 1973; Lennie, 1974; Furmanski & Engel, 2000), and the typical finding is that performance is superior for horizontal and vertical orientations relative to 45° and 135° oblique orientations, termed as the ‘oblique effect’ (Appelle, 1972; Essock, 1980). To understand more realistic orientation processing occurring in everyday viewing of the world, the use of broadband spatial contents that resemble the amplitude spectrum of natural scenes (i.e.,  $1/f$ ) should be a fundamental next step (i.e., because these broadband patterns would cause more realistic complex responses from detecting-mechanisms tuned to multiple spatial components in the broadband stimuli, Basole et al., 2003; Baker & Issa, 2005; Mante & Carandini, 2005; Olshausen & Field, 2005; Bex et al., 2007).

To address this issue, the visual performance (i.e., perceptual strength or salience which can be adjusted by the magnitude of content by observers) for oriented visual stimuli of either narrowband (i.e., gratings) or  $1/f$  broadband patterns was measured by using a suprathreshold matching paradigm (Essock et al., 2003). Specifically, the logic underlying the visual performance is that if the perceptual strength or salience is equal for all orientations (horizontal,  $45^\circ$ , vertical, and  $135^\circ$ ), the magnitude of the content (or perceptual strength or salience) should be equal for all orientations. If not, the magnitude of the content should be biased in a particular orientation. The results showed two distinctive anisotropies in orientation perception in the following. The classical oblique effect was observed only with a narrowband grating pattern (single orientation and high spatial frequencies). This indicates that the perceptual strength for horizontal and vertical stimuli was better than oblique stimuli. However, with broadband stimuli (i.e., about  $20^\circ$  in orientation and 1 octave in frequency) the horizontal stimuli were seen most poorly, the oblique stimuli were seen best and vertical stimuli were seen intermediately, which was referred to as the ‘horizontal effect’ anisotropy (see Figure 4 in Essock et al., 2003). Showing the significant change in orientation perception (i.e., from the oblique effect to the horizontal effect, increasing the spatial frequency and orientation band respectively) should be counted as an important hint in order to understand more realistic orientation processing and thus develop more realistic models of orientation processing.

To understand what underlies the now well documented significant change in orientation perception, we have suggested an ‘anisotropically-weighted gain control model’ (Essock et al., 2003; Hansen & Essock, 2004; Essock et al., 2009) that provides evident neural processes to account for the change in orientation perception, and further

details underlying processes of this model are described in Section 1.3.b.

### 1.3.1 Gain Control Models

The main input signal which is fed into V1 is, as described above, a simple contrast structure (e.g., contrast at different spatial frequencies and orientations) that generally make up the content of natural scenes at any given point in time consist of a complex amalgamation of variation in contrast (i.e., difference in luminance between light and dark regions of an image). It is widely known that cells in V1 respond to local (or small) contrast within a small region of the visual field at a wide range of spatial frequencies and orientations. Specifically, orientation selectivity is the most striking feature of cells in the primary visual cortex (Hubel & Wiesel, 1959). The standard description of this stage of visual processing is that of a set of linear filter banks where each cell responds to contrast at an oriented edge at a certain location within the visual scene. From a vision scientist's point of view, we would like to understand why an orientation selective filter bank (V1) provides a useful scene encoding. The underlying neural processing of visual perception will be considered in the following section.

#### a. Background

##### i. Visual Masking

Most studies of human vision hold that the early stage of visual processing (primary visual cortex or V1) involves an array of mechanisms, each responsible for detection within a subset of the stimulus domain. The first evidence was derived from experiments on the detecting of spatial patterns that suggested the existence of mechanisms



differentially tuned to spatial frequency (Campbell & Robson, 1967). The physiological findings are that most cortical cells (corresponding to psychophysical mechanisms) highly respond to contrast structure (e.g., contrast oriented edges and spatial frequencies) (Hubel & Wiesel, 1962) that generally make up the content of the natural scenes at any given point in time consists of a complex amalgamation of variation in contrast. In particular, a given cell responds only if its receptive field is stimulated by a pattern containing components lying within a restricted range of orientations and spatial frequencies. This idea has been referred to as a channel (also called detector, unit, cell, filter, mechanisms, or pathway by various researchers). This idea has been used for many psychophysical researchers to investigate the response property of pattern-sensitive visual mechanisms (or channels) using simple grating stimuli, which are narrow in spatial frequency and orientation bands and are present at a uniform contrast (i.e., contrasts presented in natural scenes are typically extremely complex). A variety of additional psychophysical paradigms (e.g., adaptation, subthreshold summation have also been used to investigate unknown channels' functional properties, but current works mainly adapted the masking paradigm) have been used to investigate the pattern-sensitive visual mechanisms underlying an unknown mechanism, and masking, which is one of the most dominant paradigms as described in the following.

Masking is a robust psychophysical phenomenon (Legge & Foley, 1980; Wilson, McFarlane, & Phillips, 1983; Foley, 1994; Meese & Holmes, 2007; 2010) in which the presence of one stimulus (the mask) interferes with the visibility of another (the test). In a masking paradigm, the threshold for detecting a particular target pattern, including orientation and spatial frequency, is increased by the presentation of a temporally and

spatially superimposed mask. It is assumed that sensitivity to a test pattern of a particular orientation and spatial frequency will be mediated by a single tuned channel (called filter, detector, unit, mechanism, and pathway, etc) which is analogous to a set of cortical neurons. Masking effects are greatest with test and mask stimuli of the same orientation or spatial frequency. Changing the orientation of the masking stimulus or its spatial frequency alters how much test threshold is elevated. This is taken as an index of the response of the tuned pathway to these stimuli. The masking effect has a narrow spread across spatial frequency, about 1.5 octaves (Wilson, McFarlane, and Phillips, 1983) consistent with the spatial frequency bandwidth of primary visual cortex cells (e.g., DeValois, Albrecht, and Thorell, 1982; Movshon, Thompson and Tolhurst, 1978). Orientation tuning (half-width at half-height) is similarly narrow (15-30 deg, Campbell and Kulikowski, 1966; Phillips and Wilson, 1984; Blake and Holpighian, 1985; Anderson, Burr, & Morrone, 1991; Snowden, 1992; Snowden & Hammett, 1996; Cass, Stuit, Bex, & Alais, 2009), which also supports the idea that pattern masking taps mechanisms at the level of the cortex. Thus, masking studies have assumed that psychophysical masking functions directly reflect the responses of pattern-sensitive cortical mechanisms. .

## ii. The Origin of the Gain Control Mechanism

### *Early Channel Models*

Much of our understanding of early spatial vision comes from studies using narrowband grating stimuli at uniform contrasts. Based on these, it has been widely provided that the models of early spatial vision draw on an insight of Fourier's analysis<sup>5</sup> (Campbell &

---

<sup>5</sup> Fourier analysis is a mathematical technique that allows any continuous waveform to be expressed as the sum of a set of simpler waveforms,. In vision research, Fourier analysis can be used to express a stimulus

Robson, 1968; Campbell, Cooper, & Enroth-Cugell, 1969; Campbell, Carpenter, & Levinson, 1969; Maffei & Fiorentini, 1977; Thomas & Gille, 1979). According to this insight, each simple contrast pattern of a visual scene can be decomposed into fundamental contrast features at different orientations and spatial frequencies. The different features are processed by multiple spatial channels<sup>6</sup> within the visual system with some degree of independence, termed as the ‘independent channel’ model (Campbell & Robson, 1968; Graham, 1989).

Additionally, masking was supposed to occur only when the mask excited the mechanism used to detect the test stimulus, termed as a ‘within-channel model’ (Legge & Foley, 1980). In this within-channel model, a test-detecting mechanism is stimulated by the mask stimulus, which reduces the signal-to-noise ratio for the observer, typically by compressing the detecting-mechanism’s response at high mask contrasts and accelerating response at low contrast (Nachmias & Sansbury, 1974; Legge & Foley, 1980; Wilson, 1980; Wilson McFarlane, & Philips, 1983; Phillips & Wilson, 1984).

Legge and Foley (1980) proposed a simple model of the detection of a spatial contrast pattern (test) superimposed on other presented spatial contrast patterns (masks differing from the test in spatial frequency, orientation, and other pattern dimensions). The task was to discriminate the test plus mask from the mask alone (two-alternative

---

as the sum of a set of sinusoidal contrast gratings. The utility of Fourier analysis stems from the fact that it is possible to predict performance- of a cell or psychophysical mechanism- in a task involving discrimination of luminance patterns from knowledge of the sensitivities of the cell or mechanism to the sinusoidal components of the patterns (Campbell & Robson, 1968).

<sup>6</sup> The generalized idea of a channel refers to a filtering mechanism, something which passes some, but not all, of the information that may impinge on it (De Valois & De Valois, 1988). Spatial or temporal frequency channels which transmit whatever input is present above a particular frequency are termed high-pass, those which transmit below a particular frequency are termed low-pass, and those which transmit within a restricted frequency region are termed band-pass. A system that filters with multiple channels, then, would be one with a number of different band-pass channels tuned to different frequency ranges, plus possibly low-pass and high pass channels at either end of the range.

force-choice procedure). Contrast-increment threshold (contrast discrimination) for the test was measured as a function of mask contrast and the shape of contrast response function was in the following. When mask and test patterns are simultaneously presented, contrast discrimination functions are dipper-shaped, with a region of *facilitation* at low mask contrasts and a region of *masking* at higher mask contrasts (called the dipper function). In particular, this facilitation region (the dipper which drops below the absolute detection threshold that means facilitation effect) at low contrasts is said to be due to a positively accelerating, nonlinear contrast transducer function (Nachmias & Sanbury, 1974). However, masking at higher stimulus contrasts has generally been attributed to a compressive nonlinearity (Legge & Foley 1980). This nonlinear process is referred to as the 'contrast gain control' (or nonlinear transducer<sup>7</sup>) (Note: the high contrast is corresponding to suprathreshold contrast, which is thought to stimulate the contrast gain control process). This idea owes much to observations of neural activity within the visual cortex (Burr, Morrone, & Maffei, 1981; Bonds, 1989; DeAngelis, Robson, Ohzawa, & Freeman, 1992) where, at suprathreshold contrast, suppressive interactions are thought to keep cells of limited dynamic range within the useful region of their operating characteristics through contrast gain control (Albrecht & Geisler, 1991; Heeger, 1992).

Another possibility is that the response variance or noise of V1 neurons increases with increasing contrast (Itti, Koch, & Braun, 2000; Tolhurst, Movshon, & Dean, 1983), a phenomenon shown psychophysically (Kontsevich, Chen and Tyler, 2002; Gorea & Sagi,

---

<sup>7</sup> This sigmoidal-shape shows the contrast response function of the transducer of Legge and Foley (1980) that show how 'contrast for threshold discrimination' ( $\Delta C$ ) varies with contrast level. In this function,  $\Delta C$  first decreases as contrast increases from zero, then show a large increase at high contrasts. This provides a dipper-shape (sigmoidal-shape) contrast discriminate function which is also true for V1 neurons showing sigmoidal with acceleration at low contrasts and compression or saturation at high contrasts (Albrecht & Hamilton, 1982; Tolhurst, Movshon, & Thompson, 1981).

2001). Increased noise at high mask contrast must increase the discrimination threshold, even if the transducer function were linear. It can be, thus, thought that a physiological explanation of the dipper should not rely only on considering how response amplitude depends on contrast (Chirimuuta & Tolhurst, 2005). Although we cannot rule out this possibility in current dissertation experiments, it can be sure that noise level is always constant for stimuli using entire experiments (because we measured an observer's ability to discriminate between suprathreshold contrast stimuli, a fixed contrast is only used). Thus, it is indeed true that our measurements (discrimination thresholds) should be most likely the consequence of the gain control process and also a constant noise which can be ignorable.

#### *Cross-Orientation Masking*

However, it has become widely demonstrated that, at suprathreshold contrast, a mask can raise the detecting threshold without exciting the detecting-channel, (Foley, 1994; Snowden, 1994; Meese & Holmes, 2003; Meese & Holmes, 2007; Meese & Holmes, 2010) which was against both early independent- and within-channel models. This cross-orientation masking effect, in which the detection of a target is masked by a superimposed mask stimulus at orthogonal orientation, is now widely accepted in spatial vision. This implies that cross-channel interaction between the test and mask stimuli suppress each other. Thus, the visibility of the test is reduced (Foley, 1994; Meier & Carandini, 2002; Meese & Hess, 2004; Petrov, Carandini & McKee, 2005; Meese, Summers, Holmes, & Wallis, 2007; Meese & Holmes, 2007; Meese & Holmes, 2010). At suprathreshold contrast, this nonlinear suppressive interaction between channels tuned to

different orientations is part of a process of contrast gain control that serves to increase the contrast range over which cortical neurons can effectively respond and keep stimulus selectivity (Heeger, 1992; Geisler & Albrecht, 1992; Carandini & Heeger, 1994). This cross-orientation masking would be a strong hallmark of the gain control process with respect to the orientations. Current work will mainly investigate the tuning property of the gain control mechanism with various stimulus dimensions (orientation, spatial frequency, and temporal frequency).

Subsequently, this approach has been used to account for psychophysical modeling. Foley (1994) has proposed a most influential model of masking (or normalization or nonlinear transducer model by other researchers), which shows the selectivity of the tuned spatial channels through a contrast gain control signal that suppresses responses to a particular target pattern. According to this model, the gain control signal comes from contrasts at frequencies and orientations similar to the test target. Beyond the general gain control models (Bonds, 1989; Albrecht & Geisler, 1991; Heeger, 1992), Foley's model has properties that are specific to the orientation of the mask. Specifically, Foley's model (Foley, 1994) has been used to account for the cross-orientation masking effect in the following way. There are interactions between the mask and test mechanisms. Foley's model has proposed that these interactions are inhibitory (or suppressive). The interactions in particular are specific to the orientation of the mask. For example, when the orientation of the mask and test target are both vertical, the inhibitory weight is the most, whereas, when the orientation of the mask and the target are orthogonal, the inhibitory weight is much less, but still inhibits the test target. Within this line of thinking, current work assumes that this model (Foley, 1994) captures spatial

channel (e.g., orientation and spatial frequency) interactions, in that suppressive weighting of spatial bands neighboring the target is stronger than that of channels with far away spatial bands.

### iii. Foley's Model of Pattern Masking

Foley's model (Foley, 1994) has been expanded with respect to the specific contrast gain control model of excitatory ( $E$ ) and inhibitory ( $I$ ) responses made by test and mask components respectively by considering orientation selectivity. In this model, a contrast gain control mechanism has properties that are specific to the orientation of the mask. The core idea was to add an additional term to the denominator (see Equation 1). An additional term represents suppression among the channels through a 'contrast gain control pooling process' that is a weighted ( $w_i$ ) contribution to suppression from mask and test components. Their model of orientation-selective pattern masking can be considered as follows. When an identical mask was added to the test stimulus (i.e., iso-orientation mask condition in which the test and mask pass through the same set of linear channels), threshold elevation as a function of mask contrast produced that characteristic dipper function at low mask contrast (i.e., facilitation) and masking at higher mask contrasts (Legge & Foley, 1980; Foley, 1994). The result suggests that a local 'within-channel' nonlinearity exists that facilitates at low mask contrast, but masks at high contrast. When an orthogonal mask is added to the test (i.e., cross-orientation mask condition: the test and mask no longer pass through the same orientation-selective channels). However, the facilitation dip disappeared at low mask contrasts and the masking function shifted horizontally from left to right (i.e., masking). This result

suggests that a pooled nonlinear gain control process that operates over a broad range of orientation-selective channels (i.e., ‘between-channels’) is tuned to different orientations (Bonds, 1989; Heeger, 1992; Albrecht & Giesler, 1991). By fitting the model to his data, Foley (1994) systematically demonstrated the orientation tuning bandwidth for excitatory and inhibitory terms and that the tuning bandwidth for the inhibitory term was much broader than the excitatory term (see Figure 7 in Foley, 1994). These orientation tuning properties correspond to the orientation tuning property of the gain control mechanism (i.e., *broadly* tuned for cross-orientation masking and *narrowly*-tuned for iso-orientation). In other words, this model clearly shows the tuning property of the gain control mechanisms.

If the test and mask patterns are processed by two *exclusively* distinct channels, then logically no threshold elevation would be observed in the cross-orientation mask condition. However, a substantial amount of masking was found. Thus, this result can not be explained by the previous within-channel model (or a single-channel model, Legge & Foley, 1980). To solve this problem, Foley proposed a new masking model and this model is given by

$$R = \frac{E^p}{\sum_i w_i \cdot I_i^q + Z} \quad (1)$$

where  $R$  is a response from the channel after the nonlinear gain control process.  $w_i$  is a weighted contribution to inhibition from mask components (i.e., the broadly-tuned sum of many different mask orientations).  $I$  is a divisive inhibitory term from the mask pattern that pools activity across channels (i.e., it is referred to as the contrast gain control pool) that is itself being inhibited (i.e., self-inhibition).  $E$  is the excitatory term from the test



pattern.  $Z$  is a semi-saturation constant ( $Z > 0$ ) (i.e., the model parameters  $p$  and  $q$  control the slope of the accelerating and compressive part of the nonlinearity ( $p > q$ )). The response ( $R$ ) seems to consider relatively simple kinds of nonlinear combination, such as multiplication or division, which was motivated by the divisive nonlinearity of cortical cells in V1 that have a broadband divisive input as well as the input that is due to excitation of the cells' receptive fields (Bonds, 1989; Heeger, 1992; Carandini & Heeger, 1994; Geisler & Albrecht, 1992). Note that the contrast gain control model is the generalized hypothetical model (Equation 1). The adjusting of the contrast response function of cortical cells by contrast normalization (i.e., divided by the conditionally-weighted inhibitory pooling term to the denominator in Equation 1) is a part of the process of cortical gain control. By doing this, cortical cells can efficiently respond to the contrast range over and maintain stimulus selectivity (i.e., orientation, spatial, and temporal frequency). This model has been widely and continuously used by many vision scientists to account for the unknown of spatial, orientation, and temporal pooling processing.

#### b. Orientation Biased Gain Control Model

A cross-orientation masking effect provide a strong insight that there is a considerable amount of inhibitory interaction between the test and mask stimuli even for large orientation differences of up to  $90^\circ$  (Foley, 1994). This inhibitory interaction in particular is specific to the orientation of the mask. For example, when the orientation of the mask and test are both vertical, the inhibitory interaction are the most strong, whereas when the orientation of the mask and target are orthogonal, the inhibitory interaction are much less,

but still inhibit the test target. Within this line of thinking, the present work assumes that this model (Foley, 1994) capture spatial channel interaction, in that suppressive weighting of the spatial of spatial bands neighboring the target is stronger than that of channels with far away spatial bands. The presence of inhibitory interactions between the test and mask stimuli can not be explained by prior ‘within-channel’ models (Legge & Foley, 1980) because, in this model, masking is supposed to occur only when the mask excites the mechanism used to detect the test stimulus.

To solve this discrepancy, Foley (Foley, 1994) suggested one crucial idea indicating nonlinear suppressive interactions between channels tuned to different orientations underlying the gain control pool (or mechanism). It has been widely suggested that this suppressive interaction between channels is a part of a process of contrast gain control that serves to increase the contrast range over which cortical neurons can effectively respond and maintain stimulus selectivity (Geisler & Albrecht, 1992; Heeger, 1992; Foley, 1994). In this view, changing the orientation of the mask stimuli alters how much the test threshold is elevated, and this can be taken as an index of the response of the tuned channels to these tests and mask stimuli. Logically, the threshold elevation as varying the mask orientation (Foley, 1994) should be corresponding to a part of a process of contrast gain control that serves to increase the contrast range and maintain stimulus orientation selectivity. Importantly, Foley demonstrated two orientation tuning selectivity of the gain control pool—one is a *broadly* pooled nonlinear gain control process that operates over a broad range of orientation-selective channels tuned to different orientations (cross-orientation masking). The other is a *narrowly* pooled nonlinear gain control process that operates a narrow range of

orientation-selective channels tuned similarly with the test orientation, which is a meaningful finding.

Recently, Essock and colleagues (Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim, Haun, & Essock, 2010; Kim, Haun, & Essock, 2011) have tested Foley's mode of contrast masking using  $1/f$  broadband patterns resembling the statistics of natural scenes, which can create more complex and realistic responses from channels (Note that Foley (1994) previously tested his model with a limited range of stimulus values, such as a single spatial frequency and orientation, which can only create a few channels' response) and the horizontal effect anisotropy is found in the following. When measuring the threshold for detecting a narrowband test grating masked by a  $1/f$  broadband mask stimulus, the threshold for horizontal test orientation is greatest, intermediate for vertical, and least for oblique orientations, termed as the horizontal effect (Essock et al., 2003). This fits well with the earlier findings of greatest salience and sensitivity for oblique content and least salience for horizontal in broadband patterns (Essock et al., 2003; Hansen et al., 2003; Hansen & Essock, 2004; 2005; 2006). To interpret the encoding processes underlying this anisotropy in orientation perception, we applied the core idea of gain control process to the horizontal effect anisotropy. Specifically, when a pattern of broadband spatial content is viewed, the multiple spatial components in the patterns stimulate detecting-channels that suppress each other (Foley, 1994). The suppression between channels tuned to the multiple spatial components in the broadband pattern is anisotropic, being relatively greater at horizontal and least at oblique orientations, which is termed as the 'anisotropically-weighted gain control pool', and this idea can be achieved by adding one term to the denominator in Foley's model (Foley,

1994). The general concept of the anisotropically-weighted gain control pool (Essock et al., 2003; Hansen & Essock, 2004) in Equation 2 accounts for the content-dependent orientation effect. By adding more spatial frequency and orientation content, the evident transition of the oblique effect to the horizontal effect is shown respectively (Essock et al., 2003; Hansen & Essock, 2006), and it was suggested that this change in the visual anisotropy results from the anisotropically-weighted gain control pool (strongest at horizontal), whose effect is pronounced when viewing a pattern of broadband spatial content. The additional content creates the response activity in channels; however, it is not significant with narrowband stimuli.

In order to show this, the current orientation biased model (shown in **Figure 2** below, Essock et al., 2003; Hansen & Essock, 2004), one must add an orientation biased term to the denominator ( $o_i$ , see Equation 2), in order to show that the weighted contribution to inhibition from the mask components ( $w_i$ , see Equation 1 in Foley Model) are not equal for all orientations, but are anisotropic (strong at horizontal). Specifically, we added one additional term, the orientation biased inhibitory term ( $o_i$ ) to the denominator (see Equation 2), which is weaker at oblique orientations, intermediate at vertical, and strong at the horizontal orientation, which results in the horizontal effect anisotropy. To account for this perceptual anisotropy in orientation processing, we provide a plausible model which is the same with the Foley's model (Foley, 1994) of pattern masking except the orientation bias term ( $o_i$ ) to the denominator. This model is given by

$$R = \frac{E^p}{\sum_i o_i \cdot w_i \cdot I_i^q + Z} \quad (2)$$

where  $R$  is a response from the channels after the nonlinear gain control process.  $w_i$  is a weighted contribution to inhibition from mask components.  $I$  is a divisive inhibitory term from the mask pattern that pools activity across channels (i.e., it is referred to as the contrast gain control pool) that is itself being inhibited. The core idea of this model is to add an orientation term ( $o_i$ ) to the denominator in order to account for the content-dependent orientation effect. Thus, ‘the inhibitory pooling’ term (i.e.,  $\sum_i o_i \cdot w_i \cdot I_i$ ) is biased in orientation (i.e., suppression is greatest at horizontal, intermediate at vertical, and least at oblique orientations).  $E$  is the excitatory term from the test pattern.  $Z$  is a semi-saturation constant. Therefore, the response of output channels at different orientations differs due to the anisotropic weighting of the divisive suppression, which shows the horizontal effect.

The horizontal effect anisotropy subserved by the anisotropically-weighted gain control pool has been continuously demonstrated for contrast (Haun & Essock, 2010), temporal dynamic (Essock et al., 2009; Kim, Haun, & Essock, 2011), surround/overlay suppression and context effect (Kim & Essock, 2010). These types of cortical suppression and gain control are importantly accounted for in the new standard mode of the visual cortex (see Figure 1b, Rust & Movshon, 2005).

Current work was designed to systematically investigate the nature of the anisotropy in the gain control pool with respect to temporal frequency and spatial frequency by measuring tuning properties of this anisotropic gain control pool. The results demonstrate the existence of multiple local gain control pools, and that all pools show the horizontal effect anisotropy.

### i. Value of Efficiency: Whitening

Orientation is one of the fundamental dimensions in natural scene perception and it is being explored in detail in our lab. It has been reported that typical natural scenes exhibit a bias in image content, favoring horizontal and vertical orientations relative to the oblique ones. We have suggested that this orientation bias in natural scenes drives the 'horizontal effect' anisotropy in orientation perception. In particular, visual performance (e.g., sensitivity) for oriented content is best for oblique and worse for horizontal than vertical orientations when viewing oriented naturalistic-broadband spatial contents (e.g., natural scenes, Essock, et al., 2003; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010). Most observers, however, show an oblique effect (worst performance at oblique orientations, Appelle, 1972) when narrowband content (such as a grating target) is viewed. In addition, if more content is added to the image, the anisotropy is a horizontal effect (Hansen & Essock, 2006). The oblique effect anisotropy has been attributed to a neural bias and has been related to a corresponding anisotropy in the number of early cortical neurons tuned to different orientations (i.e., Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; DeValois et al., 1982; Li, Peterson, & Freeman, 2003). We have suggested that the change in the visual anisotropy results from the anisotropic contrast gain control mechanism (strongest at horizontal), whose effect is pronounced when viewing a pattern of broadband spatial content. The additional content creates more response activity in channels contributing to the contrast gain control pool, whereas the response activity in channels is not significant with narrowband stimuli (Hansen & Essock, 2004; 2006; Essock, et al., 2009).

Such anisotropic suppression in the gain control pool (or mechanism) could be a

direct consequence of the corresponding bias of cortical neurons (Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; DeValois et al., 1982; Li et al., 2003), which indicates that more neurons are tuned to the horizontal than to other orientations and vertical than oblique orientations. In addition, the orientation bias of typical natural scenes (Switkes et al., 1978; Coppola et al., 1998; Hansen & Essock, 2004; 2005; 2006) indicates the most amplitude (or power/contrast) at the horizontal, least at the oblique ones, and intermediate at the vertical orientation. We have suggested that the horizontal effect is linked to the ‘anisotropic contrast gain control pool’, which is a very efficient coding strategy of the early visual processing stage that serves to equalize the neural representation of natural scenes with respect to orientation (Essock et al., 2003). In particular, this anisotropic gain control processing equalizes the prevalent content (i.e., the horizontal > vertical > oblique orientation bias) in the natural scenes as increasing suppression at the horizontal orientation results in less sensitive performance at the horizontal orientation, whereas as decreasing suppression at the oblique orientations results in enhanced sensitive performance at the obliques. Therefore, the horizontal effect anisotropy is a great example to show that the encoding-mechanisms underlying the visual system are likely to have evolved to reflect natural scenes, and in order to equalize the prevalent content within the natural scenes – specific optimizing processing strategy to “whiten” the neural response to natural scenes via anisotropic contrast gain control pool (see **Figure 3**).

Taken together, by removing redundancy (the prevalent horizontal content) through the anisotropically-weighted gain control pooling process (strongest at the horizontal) at the early orientation processing stage (V1), the incoming *local* contrast

features within a small region of the global visual field can be conditionally pooled. These pooled signals can, then, be efficiently transported to their higher visual stages (V2). This conditional pooling process can be achieved by mutual suppression between channels, in which each channel's response is divided by gain control pooling signals, which come from its similar orientations and spatial frequencies (see above Equation 2). This is part of a process of gain control that serves to equalize the horizontal content bias present in typical natural scenes (i.e., Essock et al., 2003; Hansen & Essock, 2004; 2005; 2006; Essock et al., 2007; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim, Haun, & Essock, 2011). Subsequently, the perception of comparatively more *global* structures (e.g., large contours, curved contours, textures, surface) can be achieved by encoding-mechanisms at preceding higher visual stages in which the global pooling occurs by integrating locally extracted visual signals from early stages (i.e., V1) over space.

## 1.2. Processing the Temporal Dimension

The natural environment in which we live is dynamic. Our visual systems used the spatio-temporal information contained in the two-dimensional images on the retina to generate a rich set of perceptions about the characteristics of our visual environment which is three-dimensional world. For example, motion in the retinal image conveys information about the shape of an object, and also about the movement of objects. These perceptions arise in a very brief time and require little conscious effort, yet they are the results of sophisticated information processing strategies that entail many stages of analysis (i.e., Van Essen & Gallant, 1994). The visual system is hierarchically organized,



but with extensive two-way information between levels, i.e., the perception of form (the shape of objects) and of motion (where things are moving) (i.e., Pack, Conway, Born, & Livingstone, 2006). This cortical hierarchy increasing in complexity is consistent with encoding-mechanisms at each visual stage that refine the selectivity of their input stage. Processing signals are fed into each processing stage where selectivity at each stage is derived by combining input from the preceding stages. The following section mainly discusses the motion pathway, specifically the early motion processing stage (i.e., V1) in which local motion information (e.g., motion details such as flickers and directions of motion) within a small region of the global visual field are mainly processed.

Several works have shown that many encoding-mechanisms and processing strategies in the visual systems are optimally suited to encode the various motion structures (e.g., flickered motion, speed and direction of motion etc) presented in the dynamic natural scenes. The strong agreements across works is increasing efficiency by removing redundancy in order to maximize information transfer strategies (Eckert & Buchsbaum, 1993; Dakin et al., 2005; Cass & Alais, 2006; Cass et al., 2009; Dakin et al., 2010). It can be achieved by integrating similar local motion details into a more meaningful global motion structure presented in the dynamic natural environments (Nishida, Ledgeway, & Edwards, 1997; Bex & Dakin, 2002; Perrone, 2004; Dakin, 2009; Bartels, Zekin, & Logothetis, 2008; Meso & Hess, 2010). In particular, the receptive fields of the early visual mechanisms (V1) are relatively small and selective for contrast at simple stimulus attributes, such as spatial frequency, orientation, temporal frequency, and direction of motion (Hubel & Wiesel, 1959; Tolhurst & Movshon, 1975; Albright,

1984). Moving from V1 to higher motion processing stages (MT), visual perception attempts to get more meaningful movement percepts (i.e., in a contrast-independent manner) other than physical contrast (a local contrast of object motions or motion details such as flickers) (i.e., in a contrast-dependent manner, V1). This indicates that a more meaningful motion perception can be obtained by the combined (or grouped) responses of a number of neighboring detectors tuned to similar local temporal details (i.e., flickers and direction of motions at limited spatial scales) making up global motion stimuli.

With respect to pooling/or grouping, the Gestalt law<sup>8</sup> of ‘good continuation’ (i.e. Marr, 1982; Field, Hayes, & Hess, 1993; Hess & Dakin, 1997) provides a plausible account. According to this notion, such local information is integrated by ‘similar’ temporal patterns (e.g., similar, temporal frequency, direction of motion or similar speed etc). Specifically, this idea has tested for moving elementary patterns (e.g., dots or Gabor elements) (Bex et al., 2003), and the same conclusions are suggested (Bex & Dakin, 2002; 2005). The aperture problem<sup>9</sup> (Marr, 1982; Pack & Born, 2001; Dakin, 2009; Kane, Bex, & Dakin, 2009; Kane, Bex, & Dakin, 2011) would be a good example, which shows this issue of how local motion signals are integrated into a coherent global motion. By integrating local information into a global structure, the visual system might efficiently

---

<sup>8</sup> Gestalt psychologists developed a list of laws to account for many of the known phenomena of perceptual grouping (e.g., Wertheimer, 1938). Among notable these, the law of good continuation played a central role. The law has been invoked to account for the perception of continuity under a variety of conditions including a number of geometric illusions. Although the phenomena which are described by this law demonstrate an important organizing principle in human vision, the law of good continuation is little more than a description of these phenomena

<sup>9</sup> To correctly perceive a realistic object motion in natural environment, the human visual system can not simply encode motion signals at different positions on an object contour. Encoding-mechanisms (or cells) in early visual system (V1) is sensitive to visual input in local (or small) part of the visual field, as if each cell is looking at the visual field through a small window (or aperture). Within an aperture, the motion direction of a contour is ambiguous because the motion component parallel to the contour cannot be inferred based on the visual input, but perpendicular to the orientation of the contours. This aperture problem implies that motion sensitive cells in primary visual cortex will always respond to a contour that crosses their receptive field, independently of its true length and orientation, as long as its direction is consistent with the preferred direction of the neuron (Lorenceanu & Shiffrar, 1992; Pack & Born, 2001).

encode motion directions (i.e., beyond V1, this pooling processing is based on long-range excitatory interactions among neighboring receptive fields with similar orientations, Field et al., 1993).

However, comparatively little is known about how the motion details (i.e., flickers/temporal frequency) at limited spatial scales are encoded by detecting-mechanisms underlying the early visual stage (V1), although many works have done beyond the early stage (V1). There still remain many unresolved issues concerning the functional architecture of temporal processing mechanisms at the early stage. Although the existence of multiple and narrowly tuned spatial frequency detecting-mechanisms is a reasonably well-established property of the spatial processing, it is not clear whether the same principle holds for temporal processing at the early stage (V1). With respect to the number of temporal channels, many prior works have suggested that there are only two (or three) and broadly tuned temporal channels (e.g., Legge, 1978; Mandler & Makous 1984; Anderson & Burr, 1985; Lehky, 1985; Hess & Snowden, 1992; Snowden & Hess, 1992; Fredericksen & Hess, 1998; Boytune & Foley, 1999; Cass & Alais, 2006; Cass et al., 2009). Much of our understanding of temporal processing at the early stage (V1) is based on the results of masking experiments employing very narrowband mask stimuli that are narrow in spatial frequency, orientation content at uniform contrast (also narrowband temporally). Unlike these narrowband flickering-mask stimuli, natural scenes have  $1/f$  characteristics in amplitude and broad distributions of spatial frequency and orientation (Field, 1987) are dynamic, and contain variable contrasts. It is reasonable to assume that these images might relatively create more realistic responses from the encoding-mechanisms underlying the human visual system than the narrowband

flickering-mask stimuli. By using the dynamic naturalistic mask image, logically we can create more realistic responses from encoding-mechanisms and then predict more realistic temporal processing. Although much analysis of the statistical property of natural scenes has suggested that encoding-mechanisms underlying the human visual system are tuned to optimize the encoding of spatial information present in natural images (Field, 1987; Simoncelli & Olshausen, 2001; Essock et al., 2003; Hansen & Essock, 2006; Essock et al., 2009), comparatively little is known about the processing of dynamic natural scenes (Dong & Atick, 1995; Dakin et al., 2005; Dakin et al., 2010).

In order to address this issue, several researchers recently examined the time varying natural images (Dong & Atick, 1995; Dakin et al., 2005) to relate statistical regularity to encoding-mechanisms underlying human visual perception (Cass, Alais, Spehar, & Bex, 2009; Dakin, Apthorp, & Alais, 2010). In particular, they examined how efficiently human visual systems encode some statistical regularities present in dynamic natural scenes in order to predict underlying neural coding processes, although this issue has been well-documented with respect to spatial properties (i.e., Field, 1987; Foley, 1994; Essock et al., 2004; Essock et al., 2009). However, little is known about this processing of dynamic natural scenes at the early stage (V1), although many studies address this processing at the higher stage (i.e., MT). Current work is focused on V1 level temporal processing by using dynamic naturalistic images ( $1/f$ , Dong & Atick, 1995) in order to predict early realistic temporal-filtering processes.

In the following section, general prior notions suggesting a limited number of temporal channels (i.e., 2 (or 3) temporal channels) are firstly considered. Then, two statistical properties of dynamic natural scenes: (1) a  $1/f$  property of amplitude on

temporal frequency in dynamic natural scenes (Dong & Atick, 1995) and (2) anisotropically-distributed energy presented in dynamic natural scenes (i.e., cardinals > oblique energy bias, Dakin et al., 2005). Finally, several psychophysical studies addressing an crucial issue of how efficiently these statistical regularities present in dynamic natural scenes are processed by the mechanisms of neural encoding underlying the human performance of visual tasks (e.g., Cass, Alais, Spehar, & Bex, 2009; Dakin et al., 2005) will be discussed. By doing this, it can be speculated whether the same principle holds for temporal processing at the early stage like spatial processing.

### 1.2.1. Background

An interesting question is how the visual system processes generally distinct temporal visual properties: ‘stationary or slow’ versus ‘flickering or rapid’, these are commonly referred to as ‘sustained’ and ‘transient’. How does the visual system process these distinctive visual properties? Human psychophysical studies in terms of sustained and transient mechanisms’ properties go back to the early 1970s. In order to determine their properties underlying human temporal sensitivity functions, various experimental paradigms have been used. Specifically, masking was one of the widely-cited paradigms which have been used to determine their characteristics (see further details about the masking paradigm above in Section 1.1.3.a.i). During the early 1970s and 1980s the dominant account of this masking phenomenon was the ‘independent channel’ (Campbell & Robson, 1968; Blakemore & Campbell, 1969; Graham, 1989) and within-channel model of masking (Wilson et al., 1983; Philips and Wilson, 1984; Harvey & Doan, 1990). Numerous psychophysical studies have provided interesting evidence of these two

distinct visual mechanisms operating within our vision. The functions subserved by these two mechanisms have been characterized in many ways: spatio-temporal tuning (Keeseey, 1972; Kulikowski & Tolhurst, 1973; Kulikowski, 1971, Tolhurst, 1973a, 1973b; Kelly, 1979; Wilson, 1980; Gorgeson, 1987; Hess & Snowden, 1992; Fredericksen & Hess, 1998), contrast gain control (Burbeck & Kelly, 1981; Wilson & Humanski, 1993; Wilson & Kim, 1998; Snowden, 2001; Hammett & Snowden, 1995; Boynton & Foley, 1999; Meese and Holmes, 2007; 2010; Essock et al., 2009; Kim et al., 2010; Kim, Haun, & Essock, 2011), temporal properties (Breitmeyer & Julesz, 1975; Keeseey, 1972; Tolhurst, 1973a, 1973b; Stromeyer et al., 1982), reaction time (Breitmeyer, 1975; Breitmeyer & Julesz, 1975; Tolhurst, 1975a; Harwerth & Levi, 1978; Vassilev & Mitov, 1976; Totev & Mitov, 2000; Mitov & Totev, 2005), temporal integration (Nachmias, 1967; Spitzberg & Richards, 1975; Breitmeyer & Ganz, 1977; Legge, 1978; Hammett & Snowden, 1995), orientation bandwidth (Campbell & Kulikowski, 1966; Blackmore & Nachmias, 1971; Movshon & Blackmore, 1973; Anderson & Burr, 1985; Blake & Holopigian, 1985; Phillips & Wilson, 1985; Anderson, Burr, & Morrone, 1991; Snowden, 1991; Snowden & Hammett, 1996; Cass, Stuit, Bex, & Alais, 2009), and the retinal locus (Sharpe & Tolhurst, 1973a, Tolhurst, Sharpe, & Hart, 1973; Snowden & Hess, 1992; Kim et al., 2010).

Although these numerous human psychophysical studies have suggested remarkable evidence of two distinctive mechanisms, there are still several unclear issues. Firstly, the general idea, that two temporal mechanisms are thought to process visual signals using a parallel, has been challenged by some researchers of the ‘cross-orientation’ masking effect (Burbeck & Kelly, 1981; Kelly & Burbeck, 1987; Foley,

1994) which is speed variant (the ration of the temporal frequency to the spatial frequency, see Meese & Holmes, 2007; 2010), and the asymmetric suppressive masking effect in which high temporal frequency mechanisms appears to suppress those selective to lower frequencies but not vice versa (Anderson & Burr, 1985; Cass & Alias, 2006; Cass, Alais, Spehar, & Bex, 2009). Secondly, one group of researchers supported that the spatial properties of the visual process from its temporal characteristics might not be separable and vice versa, (i.e., Legge, 1978; Hess & Snowden, 1992; Wilson, 1980; Tolhurst, 1973) which is termed as spatio-temporal interactions. The other group suggested that the spatial properties from its temporal characteristics can be separable, as pointing out the significant change in the shape of the temporal frequency tuning in sensitivity depend upon the spatial frequency (i.e., Robson, 1966), termed as separable properties or spatio-temporal covariation (i.e., Robson, 1966; Burbeck & Kelly, 1980; Kelly, 1966; Baker & Cynader, 1988; Baker, 1990; Foster, Gastra, Nagler, & Pollen, 1985; Baker & Issa, 2005; Yu, Verma, Yang, Tibballs, Lui, Reser, & Rosa, 2010). Although there are several arguable issues regarding the general notion addressing the limited number of temporal mechanisms on one side, numerous studies exploring their distinct properties recognize the existence of many and unknown temporal mechanisms on the other side. More recently, this discrepancy seems to be cleared by several surprising neurophysiological studies (i.e., Basole, White, & Fitzpatrick, 2003; Baker & Issa, 2005; Mante & Carandini, 2005; Shoham, Hubener, Schulze, Grinvald, & Bonhoeffer, 1997; Sun, Ueno, Waggoner, Gardner, Tanaka, & Cheng, 2007; Issa, Rosenberg, & Husson, 2008; Yu et al., 2010). Further details regarding these works will be described in the following Section 1.2.1.d.

The approach to studying sustained and transient mechanisms differs between earlier and more recent years. The different properties of them have been explained by the earlier channel models, which expect that masking was supposed to occur only when the mask excited the mechanism used to detect the test (Campbell & Robson, 1968; Blakemore & Campbell, 1969; Graham, 1989). Whereas in the recent years, one widely held view of this process supposes that interaction between the mask and test mechanisms (i.e., cross-channel interaction), and explicit models are inhibitory, which is termed as a cross-channel interaction model (i.e., Foley, 1994; Ross, Speed, & Morgan, 1993; Meier & Carandini, 2002; Chen & Foley, 2004; Meese & Hess, 2004; Petrov, Carandini, & McKee, 2005; Cass & Alias, 2006; Meese & Holmes, 2007; Meese, Summers, Holmes, & Wallis, 2007; Essock et al., 2009; Meese & Holmes, 2010). Unlike the earlier channel model, a within-channel model (Legge & Foley, 1980; Wilson, 1980; Nachmias & Sansbury, 1974; Philips & Wilson, 1984; Harvey & Doan, 1990) in which masking involves the combination of the mask and test stimulus within a single mechanism, this cross-channel inhibitory interaction model predicts that the mask should attenuate the perceived contrast of a test stimulus even for large orientation differences of up to 90° (the cross-orientation masking, Meese & Holmes, 2007; 2010). Based on these cross-channel interaction models, several human psychophysical studies have attempted to observe sustained and transient properties (Petrov, Carandini, McKee, 2006; Saarla & Horzog, 2008; Meese & Holmes, 2007; 2010; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim, Haun, & Essock, 2011).

In this new channel model (cross-channel interaction, Foley, 1994; Meese & Holmes, 2007; 2001), the focus is on the phenomenon of suppression, which has had a



long history in psychophysical vision research and has been used by numerous researchers for different purposes. The general idea of suppression is originated from the masking experiment. Masking is the psychophysical phenomenon, as described above, where the presence of one stimulus (the mask) interferes with the visibility of another (the test). There are probably several different processes involved in the various masking phenomena in the literature, but the one which is thought to underlie several of them is suppression (i.e., Foley, 1994; Boynton & Foley, 1999; Olzak & Thomas, 1999; Chen & Foley, 2004; Chen & Tyler, 2008; Meese & Holmes, 2004; Meese & Hess, 2004; 2007; 2010; Essock et al., 2009; Kim et al., 2010; Kim et al., 2011). Furthermore, researchers of most contemporary models of suppression tacitly assume that cross-orientation masking is scale invariant (Graham & Sutter, 1996; Watson & Solomon, 1997; Brady & Field, 2000), but there is surprising evidence to challenge this assumption that cross-orientation masking is scale variant (i.e., the cross-orientation masking effect is not seen at the higher spatial frequencies because the loss of sensitivity in that region means that the mask does not rise to sufficient potency, Burbeck & Kelly, 1981; Meese & Holmes, 2004; 2007; 2010).

A subsequent erosion of the earlier channel models and introducing a new channel model have changed our approach to dealing with sustained and transient mechanisms properties, considering two mechanisms' properties in terms of suppression. Some studies exploring suppression also provide very interesting evidence to support sustained and transient mechanisms' properties. The functions subserving their properties have been characterized more recently in many ways: spatio-temporal tuning (Mesese & Holmes, 2007; Meese & Baker, 2009; Kim et al., 2010), temporal properties (Wilson & Kim,

1998; Saarlea & Herzog, 2008; Essock et al. 2009), temporal integration (Yu & Levi, 1999; Petrov & McKee, 2009), orientation tuning (Petrov, Carandini, & McKee, 2005; Meese & Holmes, 2007; Cass et al., 2009; Essock et al., 2009; Meese & Baker, 2009; Mess & Holmes, 2010; Kim et al., 2011), and the retinal locus (Xing & Heeger, 2001; Petrov, Carandini, & McKee, 2005; Petrov & McKee, 2009; Kim et al., 2010).

a. 2 or 3 mechanisms

The sustained and transient distinction was introduced by Keesey (1972) who proposed that the different threshold level may represent separate pattern and motion perceptual systems. Subsequently, this idea was reinforced in studies by Tolhurst (1973) and Kulikowski and Tolhurst (1973). Before beginning, it should be made clear that the sustained and transient dichotomy derived largely from studies, in which observers view ‘very narrowband grating stimuli’ in the presence of uniform flicker mask stimuli which is very narrowband in spatial frequency, orientation, and temporal frequency. The terms ‘sustained’ and ‘transient’ refer to the inferred perceptual or psychophysical mechanisms here. These two temporal mechanisms generally differed in several respects. First, the sustained and transient systems were presumed to have different spatio-temporal tuning. The sustained system is optimally sensitive to stationary grating of moderate spatial frequency. That is, it is ‘lowpass’ in the temporal frequency dimension, but ‘bandpass’ in the spatial frequency dimension. The transient system on the other hand, is tuned to low spatial frequency (lowpass spatially) and moderate temporal frequencies (bandpass temporally) (see **Figure 4**).

In addition to sensitivity differences, these two mechanisms differ functionally.

The sustained system signals the presence of the stationary spatial pattern, whereas activation of the transient system results in the perception of temporal change (flicker or motion). Later, these studies were found to be consistent with a model proposed by Legge (see Figure 10 in Legge, 1978). He viewed the transient system as a single lowpass filter with an upper limit of about 4 cpd. He theorized that the sustained system operates from as low 0.375 cpd to high spatial frequencies, with the upper cutoff being the acuity limit of about 60 cpd. This indicates that there is some overlap between two mechanisms around low spatial frequencies. A study by Stromeyer, Klein, Dawson, and Spillman (1982) also suggested that the sustained system may extend even to low spatial frequencies. In Legge's experiment (Legge, 1978), detection was mediated by the transient system up to a spatial frequency as high as 1.5 cpd where the sensitivity profiles of the two systems cross (or overlap). From this, it is reasonable to assume that a transient mask selectively depressed the transient system so that the sustained system could mediate detection at lower spatial frequencies because of overlap between two mechanisms around low spatial frequencies (Legge, 1978; Stromeyer et al., 1982). Kulikowski and Tolhurst (1973) indicated that the transient system extends to at least 20cpd, the highest used in their study. Murray, MacCana, & Kulikowski (1983) showed upper limits ranging from 20 to 28 cpd, depending on temporal factors.

In most of these studies, the conclusion that the transient mechanism had sensitivity at high spatial frequencies depends entirely on subjective pattern and motion thresholds. Based on these studies, we might guess that we could not separate the spatial properties of the visual process from its temporal characteristics: the effect of the spatial characteristics of the stimulus depends on its temporal characteristics and vice versa

(Legge, 1978; Lehky, 1985; Snowden & Hess, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006). However, it should be made clear that this idea indicating nonseparable properties between the spatial and temporal characteristics derived largely from studies, in which observers viewed sinusoidal grating target probes in the presence of spatially and temporally narrowband mask stimuli in prior works (i.e., further details of mask stimuli across several psychophysical works, see **Table 1**). In other words, the use of spatially narrowband mask stimuli modulated in time by temporally narrow frequency would create a few individual detectors' responses (i.e., Shoham, Hubener, Schulze, Grinvald, & Bonhoeffer, 1997; DeAngelis, Ghose, Ohzawa, & Freeman, 1999; Sun, Ueno, Waggoner, Gardner, Tanaka, & Cheng, 2007). It is reasonable to assume that the responses generated by narrowband mask (a single grating) would not strong enough to evoke unknown, less sensitive, and many temporal frequency-tuned detectors, but would have strong enough to evoke only well-known two (or three) most sensitive temporal frequency tuned detectors with maximal sensitivity around 8-12 Hz (i.e., Basole et al., 2003; Baker & Issa, 2005). In order to create more realistic responses from unknown and many temporal frequency-tuned detectors at the early visual stage (V1) and thus to develop more realistic models of the human temporal processing, the use of more realistic pattern (i.e., naturalistic scenes) seems to be fundamental. Critically, we can not totally exclude the reasonable possibility that there are indeed many temporal detectors underlying the human temporal sensitivity function, and prior mask stimuli (narrowband) was not adequate to generate these many and less sensitive temporal detectors, From these points of view, it is reasonable to think that this nonseparable decision obtained using spatially and temporally narrowband mask stimuli should be an oversimplification.

Here, it seems to be needed more special cares to verify this issue and further details with respect to these views will be described in the following Section 1.2.1.d.

Furthermore, Anderson & Burr (1985) provided an interesting asymmetric masking in which masks of high temporal frequency (i.e. most likely processed by transient, bandpass mechanism) only mask the low temporal frequency test (i.e., most likely processed by sustained, lowpass mechanism), but not vice versa (i.e., Cass & Alias, 2006; Cass et al., 2009). When considering the data of Robson (Robson, 1966) shown in **Figure 5**, we can observe that there is some *overlap* between two temporal mechanisms around low temporal frequencies. This overlap would be one of the plausible reasons of why the asymmetric masking (Anderson & Burr, 1985) are occurred. The similar results were reported by Cass and Alias (2006) and Cass et al., (2009). This result is interestingly interpreted by them that this asymmetric suppressive architecture (i.e., low temporal frequency target are strongly masked by the high temporal frequency mask, but not vice versa) may serve an equalizing function to compensate for the preponderance of low temporal frequencies in natural movies (which have  $1/f$  amplitude spectrum, Dong & Atick, 1995; Cass et al., 2009), and in this view several relevant works will be described in Section 1.2.3.a.

### *Third Channel?*

Although the terms 'sustained' and 'transient' are generally used to describe electrophysiological response, numerous attempts have been made to link particular visual percepts to sustained or transient to neural mechanisms. For example, percepts of patterns and flickers (or forms or motions) were thought to arise specifically from the

responses of sustained and transient mechanisms respectively. Another type of visual channel in the high velocity region (i.e., low spatial and high temporal frequencies) has been termed as a ‘third channel’ (Burbeck & Kelly, 1981; Kelly & Burbeck, 1987; Hess & Snowden, 1992; Langley & Bex, 2007) relating to sustained and transient mechanisms properties. Burbeck and Kelly (1981) found that masking by an orthogonal grating of the same spatial and temporal frequency as the test grating produced substantial contrast threshold elevation when both the test and mask were of ‘high velocities’ (the ratio of the temporal frequency to the spatial frequency), but not in any other frequency region. In that study, they interpreted this to mean that there were two types of mechanisms: one type in the high-velocity region, and one type elsewhere in the spatio-temporal frequency domain. The fact that orthogonal masking has a profound effect in the high velocity region suggested that the underlying mechanism in that region is not as narrowly tuned for orientation as are the mechanisms responsible for thresholds in the rest of the spatio-temporal frequency domain. They also found that the mechanism in the high-velocity region is more broadly tuned for spatial frequency than the other mechanisms that have been revealed in masking studies. The results were confirmed and extended by Ferrara and Wilson (1985). Legge (1978) suggested that there was a single mechanism that is most sensitive to high temporal and low spatial frequencies – a so-called transient mechanism—and multiple sustained mechanisms handling the rest of the spatio-temporal frequency domain. In this model, each sustained mechanism has narrower tuning for spatial frequency than does the transient mechanism. Orientation tuning is not considered. Hess and Snowden (1992) proposed that a third temporal channel is apparent only when testing at very low spatial frequencies (i.e., see figure 4-7 and 12 in Hess & Snowden,

1992). A third channel idea is supported by previous studies using a variety of different techniques (Mandler and Makous, 1984; Hess & Plant, 1985; Lehky, 1985; Hess and Snowden, 1992; Snowden and Hess, 1992; Waugh & Hess, 1994; Fredericksen & Hess, 1998; Ascher & Grzywacz, 2000; Langely & Bex, 2007). However, it is still cloudy whether or not the third channel is independent from them, or belongs to transient mechanisms, because of scant psychophysical studies.

Taken together, it is well documented that the overall temporal sensitivity function is composed of a limited number of broadly tuned temporal mechanisms (Keeseey, 1972; Kulikowski & Tolhurst, 1973; Legge, 1978; Pantle, 1983; Breitmeyer, Levi, & Harwerth, 1981; Holiday & Rudock, 1983; Mandler & Makous, 1984; Anderson & Burr, 1985; Hess & Plant, 1985; Lehky, 1985; Hess and Snowden, 1992; Snowden & Hess, 1992; Waugh & Hess, 1994; Fredericksen & Hess, 1997; Ascher & Grzywacz, 2000; Cass & Alias, 2005; Langely & Bex, 2007; Cass et al., 2009), although the estimated number of temporal mechanisms varies among ‘two’ (Keeseey, 1972; Kulikowski & Tolhurst, 1973; Legge, 1978; Pantle, 1978; 1983; Breitmeyer, Levi, & Harwerth, 1981; Holiday & Rudock, 1983; Mandler & Makous, 1984; Anderson & Burr, 1985; Cass & Alias, 2006; Cass et al., 2009), ‘three’ (Mandler & Makous, 1984; Anderson & Burr, 1985; Hess & Plant, 1985; Lehky, 1985; Hess & Snowden, 1992; Snowden & Hess, 1992; Waugh & Hess, 1994; Fredericksen & Hess, 1997; Ascher & Grzywacz, 2000; Langely & Bex, 2007).

Critically, these well-documented notions that indicate only 2 (or 3) temporal channels (i.e., Lehky, 1985; Hess & Snowden, 1992; Snowden & Hess, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) are typically based on

the results of experiments employing narrowband mask stimuli, which are very ‘narrow’ in spatial frequency and orientation content and present at uniform contrasts. Based on this stimulus condition, only 2 (or 3) common temporally tuned detectors (i.e., the peak masking effect constantly appeared at one of the 2 (or 3) common temporal frequencies which are around 1-2 Hz, 8-12 Hz, or a very high temporal frequency (i.e., around 20 Hz, see figure 12 in Hess & Snowden, 1992)) are observed. From this, we still can not totally exclude the possibility that the narrowband mask stimuli is not adequate to stimulate the many, unknown, less sensitive temporal detectors whose detection threshold are different. We propose that to drive many of them, an appropriate mask stimulus which can rise above a certain detection threshold of each of many temporal detectors is crucial. To asses this idea, we need to measure the tuning of the masking function of each detector by naturalistic ( $1/f$ ) broadband mask stimuli which would stimulate the encoding-mechanisms underlying the visual system in everyday viewing situation, and might create more realistic response from many temporal detectors underlying the early stage of the human temporal processing.

#### b. Implications of the Shape of Temporal Sensitivity Curve

The general notion of human spatial vision holds that the early stage of visual processing involves an array of multiple linear channels in which each process contrasts at different orientations and spatial frequencies at a small location within the visual field. In this view, the first evidence was derived from experiments on the detection of spatial stimuli that suggested the existence of multiple channels differently tuned to the spatial frequency (Campbell & Robson, 1967) and orientation (Blakemore & Campbell, 1969; Campbell &



Kulikowski, 1966; Hubel & Wiesel, 1959). A considerable amount of studies soon followed showing that these spatial channels can be characterized by their temporal properties (Robson, 1966; Kelly, 1966; Kulikowski & Tolhurst, 1973; Koenderink & van Doorn, 1979). These studies showed that the shape of the human temporal sensitivity function typically varies with the spatial frequency. Typically, it was reported that at low spatial frequencies the temporal sensitivity function has a bandpass temporal characteristic with a peak sensitivity around 10Hz, while lowpass characteristics, with a peak sensitivity around 1Hz, are found at high spatial frequencies.

Specifically, Robson (1966) demonstrated the effect of spatial frequency upon the temporal contrast sensitivity function. The temporal sensitivity curve has bandpass temporal characteristics at a low spatial frequency pattern with a peak sensitivity around 10 Hz (a 0.5 cpd pattern in Figure 5), which shows reduced sensitivity at low temporal frequencies. Conversely, lowpass temporal characteristics at high spatial frequency pattern with a peak around 1 Hz (4, 15 and 22 cpd patterns in **Figure 5**) shows enhanced sensitivity at low temporal frequencies. This clear reverse relationship between the temporal frequency and spatial frequency preferences indicates that the temporal characteristics of each of underlying mechanism would be independent of the spatial frequency at which they are measured. This implicates an evident separable property in the spatial and temporal frequency dimensions (i.e., If they are not separable, the shape of temporal sensitivity curves should not depend upon the spatial frequency, such as lowpass temporal characteristics at low spatial frequencies or bandpass characteristics at high spatial frequencies, but didn't).

The significant change in the shape of temporal sensitivity curves with different

spatial frequencies may be attributed to a 'slower time course of inhibition' from the surround mechanisms than that of excitation from the center mechanism (i.e., Wilson, 1980; Tolhurst & Movshon, 1975). Specifically, the 'lateral inhibitory surround mechanism' (i.e., relatively insensitive to low spatial frequencies) is assumed to be slower than that for excitation from the center mechanism, so that the response to high spatial frequency patterns should increase the sensitivity to low temporal frequencies (i.e., Tolhurst & Movshon, 1975; Wilson, 1980). In this view, Stromeyer, Klein, Dawson, and Spillmann (1982) demonstrated that poor sensitivity for the low spatial frequency pattern (i.e., 0.12- 1 cpd) can be enhanced by modulating itself in time. These seem to emphasize the importance of the lateral inhibitory mechanisms (i.e., the non-classical receptive field that is the surround region).

Based on these lines of thought, it can be logically assumed that through the use of an appropriate mask stimulus, we would be able to observe multiple single peaks from the masking function when temporal frequency of the test and mask matches (i.e., indicating many temporal frequency-tuned detectors which was already well-documented in spatial frequency dimension), not only one of the 2 (or 3) common peaks from the masking function centered at one of the 2 (or 3) common categories. Typically, the peak masking effect for grating targets appeared at one of the 2 (or 3) temporal frequencies in which one is at 1 Hz (sustained or lowpass mechanism), another (transient or bandpass mechanism) is around 10 Hz, and the other is observed at a very high temporal frequency region (or termed as third channel) (i.e., see Lehky, 1985; Hess & Snowden, 1992; Fredericksen, 1998; Cass & Alias, 2006; Cass et al., 2009). Based on this, it would be carefully suggested that these appropriate mask stimuli (i.e., more realistic and complex

mask stimuli) which can rise above a certain detection threshold of each of many, unknown, less sensitive temporal detectors is crucial. This particular mask patterns which can be seen in everyday viewing situation and create more realistic response activities from these temporal detectors might be adequate to stimulate many, unknown, and less sensitive temporal-detectors, and thereby would elicit significant mutual suppressive interaction between temporal-detectors tuned to similar temporal frequencies. Current work (Experiment 1) examines this idea with a particular mask pattern which is naturalistic ( $1/f$ ) broadband spatially and but narrowband temporally which might be adequate enough to elicit the responses from many temporal frequency tuned detectors.

### c. Current Approach of Temporal Processing

The approach to exploring the mechanisms of temporal processing significantly changed since the cross-orientation masking (the cross-channel interaction can be considered as part of a process of contrast gain control process, Foley, 1994) is tuned to the speed (i.e., the ratio of the temporal frequency to the spatial frequency) (i.e., Meese & Holmes, 2007; 2010). Boynton and Foley (1999) raised a question of what underlies the well documented dependence of temporal sensitivity on spatial frequency, such that the temporal dependence is bandpass at low spatial frequencies and lowpass at high spatial frequencies (Robson, 1966; Kelly, 1966). They assumed that this well known dependence of temporal sensitivity on spatial frequency might be subserved by the contrast gain control mechanism with different contrast gains. This assumption was achieved by measuring the TvC function for each test and mask in different spatio-temporal combinations and the same vertical orientation. One notable finding among their results

is that when test and mask stimuli have 'low spatial and high temporal frequencies', the rising portion of the TvC function is steeper (high contrast gains) than for 'high spatial and low temporal frequencies' (see Table 3 in Boynton & Foley, 1999). These results strongly suggest that at least two classes of gain control mechanisms exist. Specifically, the large slope (high contrast gains) occurs when the low spatial and high temporal frequencies are used and the small slope (low contrast gains) occurs when the high spatial and low temporal frequencies are used. Previously, the similar results are also reported that the contrast response function (sigmoidal contrast response function, Legge & Foley, 1980) for low spatial/high temporal frequency mechanisms shows greater response suppressive compression (high contrast gains) than for high spatial/low temporal frequency mechanisms (low contrast gains).

More recently, this idea indicating at least two nonlinear gain control mechanisms at low spatial/high temporal frequencies (high-speed) and high spatial/low temporal frequencies (low-speed) (e.g., Burbeck & Kelly, 1980; Hess & Snowden, 1992; Boynton & Foley, 1999) has been demonstrated by cross-orientation masking experiments. This experiment showed that the cross-channel suppression indicating part of a contrast gain control process is tuned to the speed (Meese & Holmes, 2007) by measuring the cross-orientation masking effect for each test and mask in different spatio-temporal combinations. The important finding (Meese & Holmes, 2007) was that the cross-orientation masking effect is systematically increased with the ratio of temporal to spatial frequency (i.e., strongest at high speed). By fitting a gain control model to their data, the results showed that the weight of suppression (indicating gain control pool) increased with the ratio of temporal to spatial frequency. This result indeed indicates that the gain

control pool (i.e., suppressive interaction between orientation-tuned channels) is tuned to the ratio of the temporal frequency and spatial frequency (or speed).

#### d. Separable Properties in Temporal and Spatial Frequency Dimensions?

Psychophysically, it is well documented that at high spatial frequencies, the human temporal contrast sensitivity has ‘lowpass’ temporal characteristics (i.e., a peak temporal sensitivity around 1Hz), while ‘bandpass’ temporal characteristics (i.e., a peak temporal around 8-12 Hz) are found at low spatial frequencies, using only narrowband grating stimuli in spatial and temporal frequency dimensions (Robson, 1966). This well-documented dependence of temporal sensitivity on spatial frequency provides a plausible insight that the individual properties of temporal detectors at the early stage (V1) might covary with spatial frequency. In other words, it is reasonable to infer that the individual characteristics of each underlying temporal mechanism are independent of the spatial frequency. If the spatial and temporal frequency dimensions are not separable, the shape of the temporal sensitivity curves (Robson, 1966) with different spatial frequencies should be the same shape. However, the shape of temporal sensitivity curve is evidently distinctive across the spatial frequencies, such as the lowpass temporal characteristics at high spatial frequencies and the bandpass temporal characteristics at low spatial frequencies. Taken together, we might logically assume is based on the significant change in the shape of temporal frequency sensitivity curves that the individual properties of temporal detectors underlying temporal mechanisms would be separable with spatial frequency (Koenderink & van Doorn, 1979; Watson & Nachmias, 1977). If these are indeed covary (indicating separable properties of the spatial and temporal frequency), the

prior assumption that there are only two or three temporal detectors would not be logically true. Further special cares seem to be strongly needed to clarify this issue.

In the following sub-sections, several influential psychophysical (Ascher & Grzywacz, 2000) works provide a plausible insight in terms of the separable properties of the spatial and temporal frequency dimensions. In addition, several neurophysiological backgrounds (Basole et al., 2003; Baker & Issa, 2005; Mante & Carandini, 2005; Yu et al., 2010) provide a new prospective of how the organization of spatial and temporal information in the primary visual cortex are encoded. The common agreement across neurophysiological works is that any arbitrary combinations of stimulus features within a complex image are efficiently represented in small receptive fields of cortex.

#### i. Psychophysical Hints

The prior studies (Anderson & Burr, 1985) that assumed only two or three temporal channels would be challenged (i.e., Ascher and Grzywacz, 2000) in the following ways. Specifically, several detection models of temporal processing are based on the assumption that detection occurs when the activity in the channels most sensitive to the stimulus rises above a certain threshold. If both of these assumptions are true, then one would expect to see ‘discrete jumps’ in the temporal tuning of the masking function as the function of the mask is shifted from being close to the peak temporal frequency of one channel (most effective/or sensitive) to the peak temporal frequency of another channel (least effective/or sensitive). When considering the data of Anderson and Burr (1985) (e.g., their figure 4), however, one does not observe such discrete jumps, but smooth transitions instead. What indeed underlies such a smooth transition in the masking

temporal frequency?

In this view, two plausible explanations were provided by Ascher and Grzywacz (2000). Firstly, the assumption about only 2 (or 3) temporal channels could be wrong, and the smooth shift in peak masking could then reflect shifts between the peaks of many temporal channels. For example, the similar idea can come from the orientation tuning curve obtained by a masking paradigm (c.f., Phillips & Wilson, 1983). Masking effects are constantly greatest with the test and mask stimuli of the same orientation. By changing the orientation of mask stimuli, one can measure how much a test threshold is elevated. This is taken as an index of the response of the orientation-tuned channel to this stimulus. A peak masking position in the orientation tuning indeed points out the peak sensitivity of each of the underlying orientation channels. When considering the data of Phillips and Wilson (1983) (e.g., their figure 1), one did not also observe such discrete jumps, but smooth transition instead. This result has been interpreted that the smooth shift in peak masking reflects shifts between the peaks of many orientation tuned-channels. This strongly indicates that there is a considerable amount of suppressive interaction between channels tuned to the test and mask orientations. Similarly, the finding of the smooth shift in peak masking (Anderson & Burr, 1985) in the temporal tuning would be possibly interpreted by the same logic in the following way. By changing the temporal frequency of the mask, one can measure how much a test threshold is elevated. This would be taken as an index of the response of the temporal-frequency-tuned mechanisms to this stimulus. Thereby, it is reasonable to think that a peak masking in the temporal tuning of the masking function would correspond to the peak sensitivity of each of the underlying temporal frequency-tuned channels.

Specifically, when the temporal frequency of the test and mask match, the detection threshold for a test would be greatest (i.e., but prior works show that the peak masking effect constantly observed at one of the 2 (or 3) common temporal frequency positions when narrowband mask stimuli are used to measure the tuning). By increasing the difference between the test and mask temporal frequency, the threshold for a test is ‘smoothly’ decreased (i.e., this provide a strong possibility that the smooth shift in peak masking might reflect shifts between the peaks of many temporal frequency tuned-channels). Secondly, the assumption that detection is due to the suprathreshold response of the most sensitive channels would be flawed. However, this second explanation suggested by Ascher and Grzywacz (2000) would be ruled out by a prior psychophysical study (Hess & Snowden, 1992). Hess and Snowden (1992) undertake such problematic assessment using a novel temporal masking paradigm, which evaluate the temporal properties of the most sensitive mechanisms. A just suprathreshold test of fixed contrast and spatio-temporal combination (4 dB above detection threshold) is masked by a series of variable contrast mask of the narrowband spatial (i.e., sum of 4 sinusoidal gratings) modulated by single temporal frequencies. When considering the data of Hess and Snowden (1992) one did not also observe such discrete jumps, but smooth transitions were observed.

From these points of views, it might be carefully inferred that the use of narrowband mask stimuli which has been used by prior studies (e.g., Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Snowden & Hess, 1992; Fredicksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) might only stimulate most sensitive (the conventional) temporal frequency-tuned channels with maximum sensitivity at



approximately 8-12 Hz, not the unknown and many hypothetical temporal frequency-tuned channels which are lesser sensitive than the conventional temporal channels. Thereby, it leads to only detect the most sensitive temporal frequency-tuned channels which show the peak sensitivity at a 1-2 Hz (lowpass temporal characteristics), 8-10 Hz (bandpass temporal characteristics), or a very high temporal frequency-tuned mechanism (i.e., around 20 Hz, see figure 12 in Hess & Snowden, 1992). In order to develop more realistic models of the temporal processing and verify whether this hypothesis is true, the use of more realistic dynamic mask stimuli, that can create the more realistic responses from many, unknown and lesser sensitive temporal channels, seems to be crucial. Specifically, it is reasonable to assume that the use of narrowband mask stimuli would not be strong enough to evoke unknown many temporal frequency-tuned channels, but would have strength to evoke only well-known 2 (or 3) most sensitive temporal frequency-tuned channels. Taken together, in order to create more realistic responses from the many hypothetical temporal frequency-tuned channels and develop more realistic models of the human temporal processing, the use of the more realistic and complex mask stimuli, such as dynamic broadband patterns, is inevitable.

In this view, several neurophysiological studies (i.e., Basole et al., 2003; Baker & Issa, 2005, Mante & Carandini, 2005) strongly emphasized the importance of the use of the more realistic, and complex visual patterns. According these works, these moving, complex visual stimuli can create more realistic responses from population cells at the early visual stage (V1). Then, one can develop more realistic models of the spatial and temporal visual processing. In the following section, recent crucial neurophysiological reports that provide an important neurophysiological framework at the early cortical stage

(V1) will be considered. Finally, one plausible reason of why the prior works (Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Langly & Bex, 2007; Cass et al., 2009) indeed failed to observe these many, unknown, less sensitive temporal channels would be provided.

## ii. Neurophysiological Reports: Spatial and Temporal Tuning

Much of our physiological understanding of spatial and temporal visual processing comes from studies (Movshon, Thompson, & Tolhurst, 1978; Foster, Gastra, Nagler, & Pollen, 1983; Baker & Cynader, 1988) that measure an individual cell's response to a combination of the spatial and temporal frequency of very narrowband grating stimuli. The common agreement across works is that a considerable amount of cortical cells exhibit the separable tuning properties for the spatial and temporal frequency (Movshon, Thompson, & Tolhurst, 1978; Foster, Gastra, Nagler, & Pollen, 1983). It also appears to be the case that the majority of cells in the striate cortex exhibit lowpass temporal properties (Foster et al., 1985; Tolhurst & Movshon, 1975), whereas, in some areas of the prestriate cortex the majority of cells exhibit bandpass temporal properties (Foster et al., 1985; Movshon et al., 1978). This indicates that there seems to be two distinctive temporal mechanisms which are referred to a lowpass and bandpass dichotomy. This lowpass and bandpass dichotomy has been strongly supported by several neurophysiological studies (Shoham et al., 1997; DeAngelis, Ghose, Ohzawa, & Freeman, 1999) and a fMRI study (Sun et al., 2007).

More recently, Yu et al. (2010) assessed one issue of how the separable property between the spatial and temporal frequency dimensions is changed systematically across

the retinal locations by using very narrowband gratings. Yu et al. (2010) reported that most V1 cells (marmoset) had separable spatial and temporal frequency tuning near central representations of cortical areas. However, the some proportion of cells exhibit significant spatiotemporal interaction increased in the representation of far peripheral vision ( $>50^\circ$ ). They interpreted these results with respect to the efficient coding point of view in the following way. Cells representing the far periphery of the visual field are crucial for behaviors, such as defensive reaction, posture and locomotion, which demand precise determination of retinal speeds at relatively short latencies. Thus, visual processing in the periphery may have been shaped by a distinctive set of evolutionary pressures, resulting in a more sophisticated degree of ‘pre-processing’ of speed at the level of V1, as well as a particularly massive projection from peripheral V1 to the topographically equivalent regions of the superior temporal motion processing (Palmer & Rosa, 2006). On the other hand, it is also possible that the degree of spatiotemporal interaction observed in different regions of the cortex depends on receptive field size. Accurate representations of motion, including speed, may benefit from integration over a larger spatial aperture (Grzywacz & Yulle, 1990; Grossberg & Mingolla, 1993), and perhaps larger receptive fields constitute one of the factors that facilitate a more sophisticated level of spatiotemporal interaction in the peripheral representations.

### iii. Neurophysiological Reports: Cortical Map of Separable Tuning Properties

How are natural scenes or complex images encoded at the early stage of the visual system? One plausible explanation could come from single-unit studies (Hubel & Wiesel, 1959; 1977). According to the findings of Hubel and Wiesel, it is well-documented that

the primary visual cortex has a modular structure. Specifically, it is composed of multiple hypercolumns, and each one contains orientation minicolumns, in which encoding cells are selectively responsive to the line of all orientations from a particular region of the visual space and for both eyes. Based on this neurophysiological notion, we might be able to provide a plausible answer about the question in the following way. When viewing a natural scene, our visual system extracts information at different orientations and spatial scales. Typically, encoding-mechanisms underlying human vision at the early visual stage (V1) respond selectively to fine detail contrast oriented-features within natural scenes (Hubel & Wiesel, 1959; 1977) (Note that we don't know the whole story of how hypercolumns work). Although this spatio-temporal model does not account for all responses within V1, it does, with reasonable accuracy, predict population responses to a variety of complex stimuli.

Much of our understanding of the organization of the primary visual cortex has been explored by using very 'narrowband grating stimuli', which can create only the individual cell's response (i.e., Shoham et al., 1997; DeAngelis, Ghose, Ohzawa, & Freeman, 1999; Sun et al., 2007). The common agreement across these neurophysiological works is that there are two segregated temporal frequency mechanisms (i.e., the sustained and the transient). This finding surprisingly corresponds to the psychophysical finding of two temporal-frequency mechanisms (i.e., lowpass ('sustained') and bandpass ('transient') in Robson, 1966; Keeseey, 1966; Kulikowski & Tolhurst, 1973; Legge, 1978; Koenderink & van Doorn, 1979). Unlike these simple gratings, the natural scenes that contain the broad orientation and spatial frequency contents can create more realistic and complex responses from population cells. It is

reasonable to think that cells with different tuning properties respond to the same location of an image simultaneously, but not always in a manner that is predictable from their responses to individual simple bars or gratings (Gallant, Connor, & Van Essen, 1998; Ringach, Hawken, & Shapley, 2002; Basole et al., 2003; David, Vinje, & Gallant, 2004). Although many researchers have examined response properties of cortical single cells with very narrowband grating stimuli, relatively little is known about how the different response tuning properties are affected by the complex and static/or dynamic naturalistic stimuli. In the early cortical stage (V1) of visual processing, a natural scene is thought to be represented (or encoded) in a different functional dimension that is selective for specific features, such as orientation, spatial frequency, direction of motion, or temporal frequency. Thus if we know the response properties of the primary visual cortex which are important determinants, we should be able to predict how particular images are perceived. In order to predict this, previously cortical maps<sup>10</sup> of orientation and spatial frequency preference have been considered in the primary visual cortex by using simple sinusoidal grating stimuli (Hubel & Wiesel, 1959; 1977). However, more recent imaging studies (Basole et al., 2003) suggest that the maps of these two spatial parameters (i.e., orientation and spatial frequency) are not sufficient to describe patterns of activity in different orientation domains generated in response to complex moving stimuli. In order

---

<sup>10</sup> The information necessary to represent visual scenes resides in the spatial and temporal properties of a distributed pattern of neural activity in the primary visual cortex (White & Fitzpatrick, 2007). This activity arises from the aggregate responses of individual neurons that are differentially tuned to features of visual stimuli, such as their position in visual space (Bonhoeffer & Grinvald, 1991) or the energy engendered by their orientation, spatial frequency, and direction of motion (Basole et al., 2003; Mante & Carandini, 2005). For primates, cells with similar preferences are clustered into radial columns, which are arrayed in a systematic fashion across the cortical surface. The arrangement of response preferences in V1 into so-called functional map was firstly recognized by Hubel and Wiesel, who probed the organization of visual cortex with microelectrodes and neuroanatomical tracers (Hubel & Wiesel, 1959, 1977; 2005). Two groups of researchers have argued that this functional map should best be considered a map of orientation in space-time, i.e., a map of spatiotemporal energy (Basole et al., 2003; Mante & Carandini, 2005), rather than a map of contour orientation per se (Bonhoeffer & Grinvald, 1991).

to directly assess this issue, Baker and Issa (2005) investigated how the cortical organization of spatial information (i.e., the orientation and spatial frequency) is involved in encoding complex motion stimuli in terms of the linear filters of the theoretical spatio-temporal energy model (Adelson & Bergen, 1985) which has been used as a framework for characterizing the response properties of individual cortical cells (Basole, 2003; Mante & Carandini, 2005). Baker and Issa (2005) showed that the distributed pattern activity in the cortex generated in response to complex motion images can be characterized by three maps of separable tuning properties: orientation, spatial frequency, and temporal frequency tuning. The maps of these three tuning properties are sufficient to describe the activity in orientation dimensions that have been measured in response to drifting complex images. This finding provides a plausible insight that the classical description of cortical organization that has been developed by using narrowband images might not be sufficient to describe the more realistic response tuning properties of the cortex to the complex motion stimuli (Basole et al., 2003; Baker & Issa, 2005; Mante & Carandini, 2005).

Based on these findings, one can logically assume that, in order to develop a more realistic model of temporal processing, the use of the more realistic images is crucial. Specifically, the broadband patterns (i.e., naturalistic and  $1/f$ ) would create more realistic responses from population cells in the V1 stage, which is adequate to elicit the separable tuning properties between the spatial and temporal information from population cells. In only this case, we can expect that the many temporal frequency-tuned detectors could be observed based on these neurophysiological findings. However, the narrowband grating stimuli would create an individual cell's response, which might not be adequate to lead

the separable tuning properties (Baker & Issa, 2005), and thus only 2 (or 3) temporal mechanisms (i.e., these are comparatively sensitive than other unknown temporal mechanisms underlying the human vision at the early stage) are observed by prior psychophysical masking studies (e.g., Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009).

After exploring much of the work, one might say that one mechanism, ‘sustained’, is most likely sensitive to stationary, or slow, temporal visual properties. The other mechanism, ‘transient’, is most likely sensitive to flickering or rapid visual information. For example, the fact that low spatial frequencies are processed by a relatively transient mechanism and high spatial frequencies in a more sustained one can be inferred from several studies, as described above. From a demonstration of this well-known fact, however, it does not logically follow that there must exist two different underlying mechanisms. This conclusion follows only if we also assume that these two mechanisms have separable spatial and temporal characteristics (i.e., that the temporal characteristics of each underlying mechanism are independent of the spatial frequency at which they are measured). Further crucial neurophysiological studies provided a strong insight into how visual information about spatial and temporal information of a natural scene is encoded in the primary visual cortex. Several neurophysiological studies have developed a spatio-temporal energy model (Basole et al., 2003; Baker & Issa, 2005; Mante & Carandini, 2005) that addresses how a variety of complex motion stimuli are encoded within the primary visual cortex.

These crucial neurophysiological backgrounds can be used to account for two

psychophysical plausible hypotheses of the human temporal processing, which are documented by Hess and Snowden (1992). Two psychophysical hypotheses provide a crucial perceptual framework that show how the individual properties of initial spatial and temporal detectors are organized. One of the two hypotheses is the sensitivity scaling hypothesis, in which the dependence of temporal sensitivity on spatial frequency (e.g., the lowpass and bandpass dichotomy) may be due to the fact that the peak sensitivity of each of the underlying temporal detectors exhibits a strong relationship with spatial frequency (nonseparable properties between the spatial and temporal frequency). This idea would be used to account for the prior studies, which have shown a limited number of temporal mechanisms (Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). The common factors across prior studies are that detection thresholds were typically measured for single grating targets with a wide range of spatio-temporal (all targets were sinusoidally contrast reversing in time) properties (the target stimuli). Then, each target's detection thresholds were measured in the presence of narrowband masking stimuli (i.e., Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass, Alias, Spehar, & Bex, 2009) and further details of the mask stimulus are summarized in **Table 1**. Then, the ratio of the mask threshold to the unmask threshold is plotted as a function of the mask temporal frequency, which is referred to as the temporal frequency tuning curve. Most of the results showed that a single peak masking effect position underlying the temporal frequency tuning curve is observed constantly around either 1 Hz or 10 Hz (also or a very high temporal frequency, 20 Hz from Hess & Snowden, 1992). If there are many temporal frequency tuned



detectors, we should be able to observe the single peaks masking effects when the temporal frequency of the test and mask stimuli match (like narrowly tuned spatial frequency detectors), however, 2 (or 3) common peak masking effects are observed in prior masking studies. According to the neurophysiological findings (Shoham et al., 1997; DeAngelis et al., 1999; Sun et al., 2007), the narrowband mask stimuli can only stimulate a limited number of the most sensitive individual temporal detectors, and the response signals created by the narrowband masks is not strong enough to elicit comparatively less sensitive and many temporal detectors. Thus, the peak masking positions always appear at around 1 Hz, 10 Hz, or a high temporal frequency (i.e., around 20 Hz, see figure 12 in Hess & Snowden, 1992). In other words, the response patterns created by very narrowband mask stimuli seem to lead the nonseparable properties between the spatial and temporal frequency dimensions (Baker & Issa, 2005).

The other hypothesis (Hess & Snowden, 1992) is the covariation hypothesis, in which the dependence of temporal sensitivity on spatial frequency (i.e., the lowpass and bandpass dichotomy) may be due to a spatio-temporal covariation in the temporal properties of either individual detectors or the population of detectors. Specifically, this idea can describe that the detectors tuned to the low spatial frequencies prefer the high temporal frequencies, detector tuned to the middle range of spatial frequencies prefer the middle range of temporal frequencies, and detectors tuned to the high spatial frequencies prefer low temporal frequencies. These indicate that each detector's preferred temporal frequency depends on the spatial frequency and separable properties in the temporal and spatial frequency dimensions (see figure 1 in Hess & Snowden, 1992). However, this idea can not describe the detectors tuned to the low spatial /low temporal frequencies or the

high spatial frequency/low temporal frequencies. With respect to this separable property of the spatial and temporal frequency dimensions, several recent neurophysiological works (Basole et al., 2003; Baker & Issa, 2005; Mante & Carandini, 2005) suggested interesting results that the more realistic and complex motion stimuli generate more realistic cortical maps of separable tuning properties of orientation, spatial frequency and temporal frequency. From this insight, when the naturalistic ( $1/f$ ) broadband mask patterns are used, we might observe the many single peaks masking effects when the temporal frequency of the test and mask matches, not just one of the common 2 (or 3) peak masking positions around 1-2 Hz or 8-12 Hz (or very high temporal frequency, see figure 4-7 and 12 in Hess & Snowden, 1992), from current dissertation work (Experiment 1). This provides a strong insight that there might many and unknown temporal detectors whose sensitivity is lesser than the peak sensitivity at 8-12 Hz. We propose that to drive them the appropriate mask stimuli which can rise above a certain detection threshold of each of these many and less sensitive temporal detectors.

### 1.2.2. Temporal Content of Natural Scenes

Natural scenes are also time varying images (not always static), and several researches (Dong & Atick, 1995; Bex, Dakin, & Mareschal, 2005; Dakin, Mareschal, & Bex, 2005; Dakin, Athorp, & Alais, 2010) have examine statistical regularities of these time varying images in order to relate them to the human visual perception. To address this issue, the temporal statistics need to be firstly considered, and we can then use visual stimuli resembling statistics of natural scenes (i.e., it can be easily generated in Fourier domain) to test responses of encoding-mechanisms underlying the performance of visual

tasks to naturalistic visual stimuli. Finally, we can predict encoding strategies – i.e., how efficiently regularities in statistics of time varying images affect visual perception.

However, measuring and characterizing the temporal statistics of natural scenes are difficult because there is almost always some component of image motion that results from self-motion occurred by eye, head, and body movements (i.e., the spatial structure of a retinal image typically change over time due to self motion and to motion in the environment, Geisler, 2008; Price & Born, 2009). In this view, further special care might be needed.

#### a. Temporal Frequency ( $1/f$ )

Recently, several researchers have examined time-varying natural scenes (i.e., natural videos or movies) in order to get the *global* statistics using the Fourier analysis technique. In particular, Dong and Atick (1995) demonstrated that *global* power spectra across natural videos, which is relatively consistent across videos, falls approximately as ‘ $1/f^\alpha$ ’, where  $f$  is temporal frequency and the exponent  $\alpha$  is approximately 1.0, which is corresponding to the  $1/f$  property of ‘global’ power (or amplitude) on spatial frequency in static natural scenes (Field, 1987; Tolhurst & Tadmor, 1997; Simoncelli & Olshausen, 2001). In relation to the  $1/f$  property that indicates the low temporal frequency energy bias (i.e., the magnitude of energy is peak at low temporal frequency), one interesting question was raised about how this low temporal frequency energy bias is represented by human visual perception. In order to predict plausible answers, Cass et al. (2009) used visual stimuli having  $1/f$  dependence in order to test the human visual performance to visual stimuli which was manipulated for particular purposes (see further details below

section).

#### b. Direction of Motion

Subsequently, this *global* statistic of time-varying images has advanced by examining the *local* power spectrum at various directions. Specifically, Dakin, Mareschal, and Bex (2005) examined the *local* statistics (i.e., the local motion structure of natural scenes, such as power distribution at various motion directions) of time-varying image sequences, such as commercial and recorded-naturalistic movies. To do this, the various local directions of motion in natural movies were recorded by using a hand-held digital camera at walking speed and tracking approximately the head movements of the experimenter. The results showed that local directional motion energy at the cardinal directions (i.e., horizontal and vertical directions) is relatively higher and broader than oblique directions, showing the oblique effect (i.e., it was referred to as the oblique effect for motion (OEM), Dakin et al., 2005). With respect to the static natural scenes, similar results are also demonstrated (i.e., Hansen et al., 2003; Hansen & Essock, 2004, 2005; 2006). Noticeably, dominated downwards motion was obtained because of the downwards motion of the ground plane (i.e., generated by ego/self motion), which tends to have more texture and higher contrast than the upper visual fields was reported in their movies. Furthermore, in order to show that the directional anisotropies didn't result from compression artifacts, an analysis of the movie sequences filmed from both a tilted camera yielded results that were, when corrected for the 45° physical tilt, identical to those reported above. Thus, it is obvious that the anisotropies in local direction statistics that they observed are a

confident property of time-varying sequences of natural. Furthermore, the distribution of local direction energy on oblique directions tends to be broader, and frequently more asymmetric, than on the cardinal directions by fitting local energy profiles with a Gaussian model (see Figure 3h in Dakin et al., 2005). Similarly, these regularities are corresponding to the evidence for both a reduction in number of cells that are selective to oblique orientation as well as a broadening of their bandwidth (Li et al., 2003). This clearly shows that encoding-mechanisms underlying the human visual system are developed to reflect the regularity of the natural environment.

### 1.2.3. Human Performance

#### a. Perceptual Whitening of Temporal $1/f$ Spectra

In the past few decades, much of well-documented spatial statistics of the static natural scenes have suggested that encoding-mechanisms underlying the performance of visual tasks is efficiently suited to code the spatial information present in natural images (Field, 1987; Tolhurst & Tadmor, 1997; Simoncelli & Olshausen, 2001; Esock et al., 2003). However, comparatively little is known about the temporal processing of the dynamic natural images. Several studies have examined how such statistical regularities of the dynamic natural scenes can affect the mechanisms of neural encoding underlying the human perception (Dong & Atick, 1995; Dakin et al., 2005; Bex et al., 2005). It was reported that the statistical properties of the dynamic natural scenes would affect the mechanisms of neural encoding underlying the human perception. Specifically, the natural stimuli are highly dominated by the contrast at the low temporal frequency (i.e.,

$1/f$  property, Dong & Atick, 1995) and low spatial frequency (i.e.,  $1/f$  property, Field, 1987). How do the mechanisms of neural encoding underlying the human visual perception process this low temporal frequency bias (i.e.,  $1/f$  amplitude in temporal frequency) present in dynamic natural stimuli?

To answer this question, Cass et al. (2009) measured the perceptual equalization point of a series of temporal noise stimuli (temporally broad but spatially narrow—i.e., 0.5-octave above and below center 1.6 cpd spatial frequency) in which the slope of temporal spectra varied from trial to trial and asking them to report whether the trial was perceived to be either ‘low’ (slow) or ‘high’ (fast) temporal frequency dominant. Observers consistently reported that stimuli composed of unbiased temporal spectra (the slope of ‘0’) contained a greater proportion of ‘fast’ (high temporal frequency) compared with ‘slow’ (low temporal frequency) luminance modulation, which indicates the strong transient response. Thereby, when observers viewed the stimuli composed of unbiased temporal spectra (the slope of ‘0’), the transient response bias was observed at only the proportion of the high (fast) temporal frequencies, not the proportion of the low temporal frequencies). Moreover, the slope ranging from between  $\approx -0.7$  and  $-1.5$  (i.e., mean =  $-0.97$ ) were reported to contain an equal (i.e., unbiased) proportion of low and high temporal frequency modulation. This latter range of the slope corresponds to the slope of the temporal amplitude spectra measured in natural scenes (Dong & Atick, 1995; van Hateren & van der Schaaf, 1996). This result provides the first direct demonstration that the human perception is not only whiten its temporal input (i.e.,  $1/f$  including low temporal frequency bias), but it does so via a transient response bias of a magnitude that appears to compensate for the low temporal frequency bias present in natural scenes.

Specifically, this experiment has investigated how the visual system responds to broadband temporal frequencies within a narrow (1-octave) range of spatial frequencies centered around 1.6 cpd (low spatial frequencies). The results showed that the transient response bias was only observed around the high (fast) temporal frequencies. In natural viewing conditions, it was well-documented that the small fixational eye-movements improve discrimination of high (fine spatial scales) spatial frequency stimuli, but not of low spatial stimuli (i.e., Rucci, Iovin, Poletti, & Santini, 2007; Rucci & Casile, 2005; Ahissar & Arieli, 2001), although there are some debates regarding this issue (Martinez-Conde, Macknik, Troncoso, & Dyar, 2006). The key idea is that the small fixational instability generates proportionally more temporal luminance variation at fine spatial scales (high spatial frequencies) than at more coarse scales (lower spatial frequencies, see figure 1c in Rucci et al., 2007).

Taken together, the two key findings reported by Cass et al. (2009)—i.e., (1) the transient response bias only showed the *high* (fast) temporal frequencies within a narrow range of *low* spatial frequencies, and (2) the small fixational instability (or the *low* (slow) temporal frequency) improves the *high* (fine) spatial frequencies— would be consider as good examples to show that the visual system is efficiently suited to process the dynamic ( $1/f$ ) natural stimuli. Furthermore, these two findings are thought to be related to the response properties of the sustained and transient mechanisms. Typically, sustained mechanisms are described as being slow acting of low contrast sensitivity, and selectively sensitive to high spatial frequencies and the steady-state intensity level. Transient mechanisms are described as fast acting of high contrast sensitivity, and selectively sensitive to low spatial frequencies and temporal changes.

## b. Perceptual Whitening of a Directional Anisotropy in Motion

A considerable proportion of cells in the early cortical stage (V1) is not only orientation, spatial frequency and temporal frequency selective, but also motion direction selective (Hubel & Wiesel, 1959; De Valois et al., 1982; Tolhurst & Movshon, 1975; Peterson et al., 2004). By recording responses to moving bar stimuli in particular directions from single cells in macaque V1 in order to make direct comparison with area MT (Albright, 1984), one examined the relationship between the optimal direction and the optimal orientation for MT cells. The results showed that 61% of cells had an orientation preference nearly perpendicular to the preferred direction of motion, as is the case for all V1 cells (i.e., another 29% of MT cells had an orientation preference roughly parallel to the preferred direction). This response property of V1 cells (i.e., the preferred direction and preferred orientation, are approximately orthogonal) results in the aperture problem (Pack & Born, 2001). Because motion-selective cells only encode directions perpendicular to their preferred orientation, the local motion arising from over extended contours is ambiguous within an aperture. In order to overcome this ambiguity, local motion signals should be pooled across space (i.e., Kane, Bex, & Dakin, 2009; Dakin, 2009). In addition, it is widely-accepted that there are more V1 cells tuned to cardinal than oblique orientations. Therefore, it is logically inferred that there are more cells tuned to cardinal than oblique motion directions. Furthermore, the aperture problem indicating that the local motion signals are pooled across space (Dakin, 2009) is highly suggest that there would exist mutual suppression between neighboring motion-selective cells, as is the case for orientation-selective cells in V1 (i.e., mutual suppression between neighboring



orientation-tuned detectors, Essock et al., 2009). In this view, an influential psychophysical study (Kim & Wilson, 1997) addressed the idea of suppressive interactions between detector tuned different directions of motion and presenting models. Taken together, (1) there would be more cells in V1 tuned to cardinal directions of motion than oblique directions, and (2) there exists suppressive interaction between detectors tuned to different directions of motion (Kim & Wilson, 1997). When considering these two strong hints, we can not completely exclude the suppressive interaction between detectors tuned similar direction of motion which would be a strong hallmark of the gain control process with respect to different directions of motion.

To address this issue, several researchers (e.g., Wilson et al., 2001; Dakin et al., 2005) have tested the perception of motion direction using random-dot kinematograms (RDKs) with varied direction variability (systematically varied the noise level) (i.e., Dakin et al., 2005) or translational structure in random dot patterns (Glass pattern, Glass, 1969) with a fixed noise level (i.e., 30% noise dot randomly oriented, Wilson et al., 2001). In the following section, several interesting studies showed the significant change in motion direction perception by conditionally manipulating the noise level which is the transition of the oblique effect (Dakin et al., 2005; Gros et al., 1998) to the inverse oblique effect (Dakin et al., 2005; Wilson et al., 2001) as directional variability (or noise) increased. In this view, the same idea was demonstrated in orientation perception by conditionally manipulating the noise level (i.e., Essock et al., 2003; Hansen & Essock, 2006; Haun & Essock, 2010; Kim et al., 2010; Kim et al., 2011). By addition more noise (spatial contents), the significant change in orientation perception—i.e., from the oblique effect to the horizontal effect are observed.

## i. Directional Anisotropy in Motion Perception

### *The Oblique Effect vs. The Inverse Oblique Effect*

The directional anisotropy in motion perception was systematically investigated by measuring direction discrimination thresholds for small dot patterns (i.e., 64 dots) moving in various directions. The directional variance in the stimulus (i.e., standard deviation of motion directions of small dots which corresponds the noise; high standard deviation indicate high noise) was varied (Dakin, Marseshal, & Bex, 2005). Two experiments were provided. Experiment 1 used an equivalent noise (EN) model to examine how the human visual system pools local elements of direction across space in order to encode global direction by measuring the smallest discriminable change in the mean direction for stimuli (i.e., small patterns moving with a mean direction of  $90^\circ$  and  $45^\circ$ ) as a function of the amount of directional variability. It was reported that the direction discrimination threshold for a pattern moving in the oblique was significantly higher than the vertical at small directional variability, which indicates the ‘oblique effect’ in motion direction perception. Also, by fitting threshold values to the EN model, one is able to infer that a precision limitation in local motion processing was  $0.5^\circ$  to  $16^\circ$ , but it then converges (i.e., whitens) at large discrimination variability, which was  $32^\circ$  to  $90^\circ$  (higher internal noise). Taken together, as the direction variability (noise level are increased) was increased, the direction anisotropy in motion perception changed from the oblique effect to the whiten oblique effect (i.e., in order to check the threshold increment between two conditions, the ratio of the ‘threshold value obtained with large direction variability to the threshold value obtained with a small direction variability might be necessary, but Dakin et al.

(2005) demonstrated the results using thresholds not the ratio of them). Experiment 2 measured direction discrimination thresholds for a range of reference directions of motion in order to more systematically explore the effect of motion direction perception. The results showed that a robust oblique effect was obtained at a small direction variability condition (i.e.,  $0.5^\circ$ ), whereas this oblique effect in motion direction perception was abolished at a medium direction variability condition (i.e.,  $32^\circ$ ).

Taken together, both experiments (Dakin et al., 2005) demonstrated the robust change in directional anisotropy of motion perception, i.e., from the ‘oblique effect’ with a small direction variability to the ‘whiten oblique effect’ with a large directional variability. Similarly, with small directional variability ( $0.5^\circ$ - $16^\circ$  direction variability) (Dakin et al., 2005) this significantly superior visual performance for cardinal than oblique directions of motion was also observed by Gros, Blake, and Hiris (1998) using a small direction variability (i.e., 100, 50, 20% coherence levels, see Figure 5 in Gros et al., 1998). However, as the direction variability of the stimuli increased, this oblique effect diminished, until a direction variability of  $32^\circ$  to  $90^\circ$ , it disappeared (Dakin et al., 2005), and the inverse oblique effect was observed (Wilson et al., 2001) by measuring detection threshold for translational structures in random dot patterns.

Additionally, Wilson et al. (2001) also ran control experiments that investigated this effect in terms of total dot density and signal dot spacing. The results showed that the inverse oblique effect was obtained for densities from 6 to 12% and for signal dot spacings from 7 to 14 arc min, although the inverse oblique effect weakened at lower dot densities. This indicates that there is a significant relationship between the dot density/dot spacing and the motion direction perception (i.e., as the direction variability increased,

the significant transition of the oblique effect to the inverse oblique effect motion perception). What does the evident change in motion direction perception underlie? In the following section, a plausible answer will be described by using the idea of gain control process.

## ii. Relation with Anisotropic Gain Control Processing

It has been demonstrated that dynamic natural scenes typically exhibit a bias in motion details favoring horizontal and vertical directions of motion components relative to the oblique ones (Dakin et al., 2005). When viewing a pattern having a large direction variability (high noise), direction discrimination performance for oblique directions is more precise than for horizontal and vertical directions (the inverse oblique effect for motion, Wilson et al., 2001; Dakin et al., 2005). When viewing a pattern having a small direction variability (low noise), direction discrimination performance for the horizontal and vertical directions (up/down, left/right) is more precise than for the oblique directions (the oblique effect for motion, Dakin et al., 2005; Gros et al., 1998; Ball & Sekuler, 1982), and if more directional movement detail (increasing the direction variability of motion direction of the individual dots) is added to the stimulus, the anisotropy is the inverse oblique effect (Dakin et al., 2005; Wilson et al., 2001).

The classical oblique effect obtained static oriented line (Appelle, 1972) has been attributed to a neural bias and has been related to a corresponding anisotropy in the number of early cortical neurons tuned to different orientation, which indicates more V1 cells tuned to horizontal and vertical orientations than oblique orientations (Li et al., 2003; Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; DeValois,

Yund & Helper, 1982). In addition, V1 cells are highly selective for direction of motion (Hubel & Wiesel, 1959). Specifically, the preferred direction and preferred orientation are approximately orthogonal in macaque V1 cells, based on responses to moving bar stimuli (Albright, 1984), resulting in the aperture problem. Therefore, it is reasonable to assume that there are more V1 cells tuned to cardinal directions than oblique ones. Furthermore, Kim and Wilson (1997) presented several models consisting of lateral inhibitory interaction between detectors tuned to different directions of motion which would be considered as a strong signature of part of the gain control process. Based on these two crucial insights, it can be logically inferred that the robust change in motion direction perception (the oblique effect and the inverse oblique effect for motion perception, Gros et al., 1998; Dakin et al., 2005; Wilson et al., 2001) might result from the dynamic anisotropically-weighted gain control pool (i.e., strong at cardinal directions) whose effect is pronounced when viewing a pattern having a large direction variability (high noise). The additional pattern motions moving various directions create more response activities in direction-sensitive channels contributing to the gain control pool which results in the inverse oblique effect anisotropy in motion direction perception, whereas the response activities in channels is not significant with a pattern moving of one direction (a small direction variability) which results in the oblique effect in motion direction perception. Interestingly, the crucial finding of the anisotropically-distributed motion-direction energy present in dynamic natural scenes (i.e., Dakin et al., 2005) indicates the most physical energy (amplitude or contrast) at the cardinal and least at the oblique directions of motion. This is a hypothetical idea based on previous works and to check this hypothesis further research is highly encouraged in the future.

Based on these three crucial information—i.e., (1) more cells in V1 are tuned to cardinal directions of motion than obliques, (2) suppressive interaction between channels tuned to different directions of motion indeed exists (Kim & Wilson, 1997), and (3) dynamic natural scenes are biased in cardinal directions of motion (i.e., cardinal > oblique direction bias, Dakin et al., 2005), it can be logically suggested that the significant change in motion direction perception might be due to the dynamic anisotropically-weighted gain control pool which can be thought to a very efficient coding strategy of the early visual motion processing that serves to equalize the neural representation of dynamic natural scenes with respect to direction of motion. Specifically, this dynamic anisotropic gain control processing would equalizes the prevalent motion direction energy (i.e., cardinal > oblique motion direction bias, Dakin et al., 2005) in the dynamic natural scenes as increasing suppression at the cardinal directions results in less accuracy or sensitive performance at the cardinal directions, showing the inverse oblique effect (Dakin et al., 2005; Wilson et al., 2001). Therefore, the transition of the oblique effect to the inverse oblique effect in motion perception would be a good example to show that the mechanism of neural encoding underlying the human perception are likely to have evolved to reflect the natural environment, and in order to equalize the prevalent information (i.e., energy in cardinal directions, Dakin et al., 2005) within the dynamic natural environment – a specific optimizing processing strategy of “whiten” the neural response to dynamic natural environments.

In this view, several researches have provided a potential clue to support this idea. In particular, it has been widely demonstrated that the responses of temporal channels tuned to flicker details or directions of motion within local (or small) regions of the visual

field are pooled by similar orientation, spatial frequency, temporal frequency, and direction of motion in order to efficiently encode the presence of a moving oriented contour (Bex & Dakin, 2002; Bex, Simmers, & Dakin, 2003; Bex & Dakin, 2004; Kan et al., 2009; Dakin, 2009). This idea was originally from the study of contour integration, using the Gestalt law of ‘good continuation’ that has been used to describe a variety of phenomena demonstrating the importance of continuity in human perception (Field, Hayes, & Hess, 1993). With respect to spatial dimension, it has been widely demonstrated that suppressive interaction between channels is greater among channels that are similarly tuned for contrast at spatial frequency and orientation than across differently-tuned channels (Foley, 1994; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim et al., 2011).

This idea could be used to account for the low level local motion processing. As Dakin et al. (2005) suggested, for example, the anisotropy in human motion perception is thought to involve two stages. The first stage involves the encoding process of the local motion details, such as flickering motion or direction of motion which is thought to be carried out by cortical cells in V1. The second stage involves pooling these local motion detail signals across space to estimate overall or global direction. By pooling local motion features across space, the global and meaningful motion (e.g., moving objects) present in dynamic natural world would be efficiently processed. Similarly, Pack and Born (2001) suggested that a crucial step in the neural interpretation of the complex dynamic visual world is the integrating of the various local motion signals. As described above, any local motion encoding detectors are selective only to the component of motion perpendicular to a contour that extends beyond its field of

view. This aperture problem is specifically relevant to direction-selective cells in the early visual stage, where small receptive fields permit only a limited view of moving object. To overcome this limitation, the idea of the pooling has been widely suggested. In this view, several psychophysical studies of motion summation in human visual perception also showed that the motion of single drifting Gabor patches is integrated by detectors with small receptive fields that scale with the target wavelength (from 2 arcmin at high spatial frequencies to as large as 7 deg at low frequencies, see Anderson & Burr, 1987). Additionally, many psychophysical studies (Dakin, 2009; Verghese & Stone, 1996; Bex, Metha, & Makous, 1998; Bex, Dakin, & Simmers, 2003; Wilson & Wilkinson, 1998; Wilson, Wilkinson, & Asaad, 1997) have shown that later stages (V2) of visual processing involve channels with receptive fields of increasing size that specialize in the detection of complex global pattern.

One crucial finding across these works addressing the motion processing (i.e., flicker motion/temporal frequency and direction of motion etc) is pooled (or grouped/integrated) by similar local motion details (e.g., similar direction of motion or similar flickering rate etc) in order to optimally encode the global motion objects. With respect to the pooling, the Gestalt law of good continuation (i.e., Marr, 1982; Field et al., 1993; Hess & Dakin, 1997) provides a plausible account. Specifically, this idea has tested for moving elementary patterns (e.g., dots or Gabor elements) (Bex, Simmers, & Dakin, 2003), and the same conclusions are suggested (Bex & Dakin, 2002; 2005). The aperture problem<sup>11</sup> (Marr, 1982; Pack & Born, 2001; Kane, Bex, & Dakin, 2009; Dakin, 2009;

---

<sup>11</sup> To correctly perceive a realistic object motion in natural environment, the human visual system can not simply encode motion signals at different positions on an object contour. Encoding-mechanisms (or cells) in early visual system (V1) is sensitive to visual input in local (or small) part of the visual field, as if each cell is looking at the visual field through a small window (or aperture). Within an aperture, the motion



Kane, Bex, & Dakin, 2011) might be a good example, which shows this issue of how local motion signals are integrated into a coherent global motion interpretation. By integrating local information into a global structure, the visual system might efficiently encode motion directions (i.e., beyond V1, this pooling processing is based on long-range excitatory interactions among neighboring receptive fields with similar orientations, Field et al., 1993).

From this point of view, the main idea of both the ‘anisotropic gain control pool’ and the ‘Gestalt law of good continuation’ share the same rule, which is, pooling by the similar visual properties in order to maximize efficiency. According to the idea of orientation-biased gain control pool, by removing redundancy (the prevalent horizontal content) through the anisotropically-weighted gain control pooling process (strongest at the horizontal) at the early orientation processing stage (V1), the incoming *local* contrast features within a small region of the global visual field can be conditionally pooled. These pooled signals can, then, be efficiently transported to their higher visual stages (V2). This conditional pooling process can be achieved by mutual suppression between channels in which each channel’s response is divided by a gain control pooling signal, which comes from its neighboring orientations and spatial frequencies (see above Equation 2). This is part of a process of gain control that serves to equalize the orientation biased natural scenes (Essock et al., 2003; Hansen & Essock, 2004; 2005; 2006; Essock et al., 2007; Essock et al., 2009; Kim et al., 2010; Haun & Essock, 2010; Kim, Haun, & Essock, 2011). According to the Gestalt law of good continuation point of view, our

---

direction of a contour is ambiguous because the motion component parallel to the contour cannot be inferred based on the visual input, but perpendicular to the orientation of the contours. This aperture problem implies that motion sensitive cells in primary visual cortex will always respond to a contour that crosses their receptive field, independently of its true length and orientation, as long as its direction is consistent with the preferred direction of the neuron (Lorenceau & Shiffrar, 1992; Pack & Born, 2001).

visual system can efficiently process local visual information by grouping similar visual features, such as orientation, motion details, spatial scales, or direction of motion, within a small region of the global visual field. By removing redundancy through this grouping process, we can get a more meaningful perception about objects, contours, or moving objects. Since the local visual features in dynamic natural environments can be efficiently processed (i.e., local pooling at the early V1 stage), the processed visual signals can be efficiently transported to later stages (IT: form pathway and MT: motion pathway), in which global pooling occurs again to get a more meaningful object or global motion structure, such as smiling faces, moving cars, dancing trees, rather than just a piece of physical contrast information.

### 1.3. Current View

Images and movies from the natural environments show statistical regularities in the presence of contrast at different orientations, spatial frequencies, temporal frequencies, and directions of motion. Specifically, the natural stimuli are highly dominated by the contrast at the low spatial/and temporal frequency (i.e.,  $1/f$  property, Field, 1987; Dong & Atick, 1995), and cardinal orientations (i.e., the horizontal content bias, see Hansen & Essock, 2004; 2004; 2005) and cardinal directions of motion (Dakin et al., 2005). How do the mechanisms of neural encoding underlying the human visual perception process these statistical regularities presented in the natural stimuli? According to the efficient coding hypothesis (Barlow, 1961), the encoding-mechanisms underlying the human visual perception develop in response to various visual features from the natural stimuli in order to remove the redundancy. Based on this hypothesis, several works have extensively

investigated how the mechanisms of neural encoding underlying the human visual perception represent these statistical regularities in terms of spatial frequency (Field, 1987; Brady & Field, 1995; Cass, Alais, Spehar, & Bex, 2009), temporal frequency (Dong & Atick, 1995), orientation (Essock, 2003; Hansen & Essock, 2004; 2006; Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010) and local direction of motion (Dakin et al., 2005). One agreement across these studies is that observers constantly underestimate the prevalent visual structures, such as low spatial frequencies and low temporal frequencies (Cass et al., 2009), horizontal orientations (Essock et al., 2009; Hansen & Essock, 2004; 2005; 2006), and cardinal directions of motion (Dakin et al., 2005; Wilson, Loffler, Wilkinson, & Thistlethwaite, 2001) in order to compensate for the comparatively less prevalent visual structures, such as high spatial frequencies, high temporal frequencies, oblique orientations, and oblique directions of motion. From the efficient coding hypothesis point of view, it is reasonable to assume that through whitening (or equalizing) by underestimating most prevalent visual features, comparatively less prevalent (or salient) visual information (i.e., low frequency, cardinal orientations, and cardinal directions of motion) can be seen against the visual environment relatively more perceptually salient. However, these descriptions didn't provide the specific mechanisms of neural encoding underlying the human visual perception. Specifically, what occurs to cause the perceptual underestimation to the most dominate low spatial/temporal frequency (Cass & Alias, 2006; Cass et al., 2009), and cardinal directions of motion (Dakin et al., 2005).

With respect to the orientation dimension, we have suggested a plausible model (Essock et al., 2003; Hansen & Essock, 2004; Essock et al., 2009; Haun & Essock, 2010)

which well describes one fundamental question of how efficiently the horizontal content biased present in natural scenes (Hansen & Essock, 2004; 2005; 2006) is processed by the mechanism of neural encoding underlying the orientation perception. These issues are described in the following.

### 1.3.1. Anisotropy in Orientation Perception

#### a. The Oblique Effect vs. the Horizontal Effect

Much of our understanding of orientation processing comes from studies using simple grating stimuli that are narrowband in spatial frequency and orientation content and present at uniform contrast (Appelle, 1972; Blackmore & Campbell, 1969; Campbell & Kulikowski, 1966; Movshon & Blackmore, 1973). The typical finding across these works is that visual performance (i.e., sensitivity) is constantly superior to the horizontal and vertical orientations relative to the oblique orientations, termed as the oblique effect (Appelle, 1972). Current understanding of orientation processing, however, comes from studies that measure the visual performance for more realistic and complex visual stimuli which resemble the statistical regularity of the natural scenes. It is reasonable to assume that these stimuli stimulate the encoding-mechanisms underlying the human visual perception in a way that is similar to natural stimulation (Bex et al., 2007; Olshausen & Field, 2005) and this assumption is also supported by several neurophysiological studies (i.e., Baker & Issa, 2005; Basole et al., 2003). Thus, we can understand more realistic orientation processing which occurs in everyday viewing of the world. In order to investigate this, the orientation perception was measured using by conditionally manipulating visual stimuli either a narrowband (e.g., simple gratings or lines) or a

broadband (e.g.,  $1/f$  and naturalistic stimuli) and we found the significant change in orientation perception. Specifically, when viewing a naturalistic images, visual performance is best for oblique orientation and worse for horizontal than vertical, termed as the horizontal effect (Essock et al., 2003). Whereas when viewing a narrowband pattern (e.g., grating stimuli), most observers show an oblique effect, and if more content is added to the image, the anisotropy in orientation perception is the horizontal effect (Hansen & Essock, 2006).

#### b. What Does the Significant Change in Orientation Perception Underlie?

To understand what exactly underlies the now well-documented significant change in orientation perception (i.e., the oblique effect vs. the horizontal effect), we have suggested that an anisotropic gain control model (Essock et al., 2003; Hansen & Essock, 2004) whose effect is pronounced when viewing a pattern of broadband spatial content. The additional content creates more response activity in channels which contributes to the gain control pool, whereas the response activity in channels is not significant with narrowband stimuli (Hansen & Essock, 2006; Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010; Kim et al., 2011). Such anisotropic suppression in the gain control pool (Foley, 1994) could be a direct consequence of the corresponding bias of cortical neurons (Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; DeValois, Yund, & Helper, 1982; Li et al., 2003), which indicates that more neurons are tuned to the horizontal than to other orientations and vertical than oblique orientations. In addition, the orientation bias of typical natural scenes (Switkes et al., 1978; Coppola et al., 1998; Hansen & Essock, 2004; 2005; 2006) indicates the most amplitude (or contrast) at the

horizontal, least at the obliques, and intermediate at the vertical orientations. Based on these, we have suggested that the horizontal effect anisotropy is linked to the anisotropic contrast gain control pool which is a very efficient coding strategy of the early visual processing stage that serves to equalize the neural representation of natural scenes with respect to orientation (Essock et al., 2003; Hansen & Essock, 2004; Essock et al., 2009). In particular, this anisotropic gain control processing equalizes the prevalent content (i.e., the horizontal > vertical > oblique orientation bias) in the natural scenes as increasing suppression at the horizontal orientation results in less sensitive performance at the horizontal orientation, whereas as decreasing suppression at the oblique orientations results in enhanced sensitive performance at the obliques.

Therefore, the horizontal effect is a great example to show that the mechanism of neural encoding underlying the human orientation perception are likely to have evolved to reflect natural scenes, and in order to equalize the prevalent content within the natural scenes – specific optimizing processing strategy to “whiten” the neural response to natural scenes via anisotropic contrast gain control pool. By removing redundancy (the prevalent horizontal content) through the anisotropically-weighted gain control pooling processes (strongest at the horizontal) at the early orientation processing stage (V1), the incoming *local* contrast information within a small region of the global visual field can be efficiently pooled. These pooled signals can be, then, efficiently transported to their higher visual stages. This conditional pooling process can be achieved by mutual suppression between channels, in which each channel’s response is divided by a gain control pooling signal which comes from its neighboring orientations and spatial frequencies. This is part of a process of gain control that serves to equalize the orientation

biased natural scenes. From these points of view, the significant change in orientation perception would be a great example to show that the encoding-mechanisms underlying the visual system or the human visual performance are likely to have evolved to reflect natural scenes, and in order to equalize the prevalent content (by removing redundant horizontal contents) within the natural scenes – specific optimizing processing strategy to “whiten” the neural response to natural scenes via anisotropically-weighted gain control pool. By doing this process our visual system can take two advantages: (1) the incoming input, such as local contrast features within a small region of the visual field, are conditionally pooled (i.e., with different weighted-suppression at different orientations) and then the processed signals at a given stage can be efficiently transported to the higher stages, and (2) the perceptual visibility to the comparatively less dominant oblique contents would be enhanced.

### c. Relation with Anisotropy in Direction of Motion

It has been demonstrated that dynamic natural scenes typically exhibit a bias in motion details favoring horizontal and vertical directions of motion components relative to the oblique ones (Dakin et al., 2005). When viewing a pattern having a large direction variability (high noise), direction discrimination performance for oblique directions is more precise than for horizontal and vertical directions (the inverse oblique effect for motion, Wilson et al., 2001; Dakin et al., 2005). When viewing a pattern having a small direction variability (low noise), direction discrimination performance for the horizontal and vertical directions (up/down, left/right) is more precise than for the oblique directions (the oblique effect for motion, Dakin et al., 2005; Gros, Blake, & Hiris, 1998; Ball &

Sekuler, 1982), and if more directional movement detail (increasing the direction variability of motion direction of the individual dots) is added to the stimulus, the anisotropy is the inverse oblique effect (Dakin et al., 2005; Wilson, Loffler, Wilkinson, & Thistlethwaite, 2000., 2005).

The classical oblique effect obtained static oriented line (Appelle, 1972) has been attributed to a neural bias and has been related to a corresponding anisotropy in the number of early cortical neurons tuned to different orientation, which indicates more V1 cells tuned to horizontal and vertical orientations than oblique orientations (Li et al., 2003; Mansfield, 1974; Mansfield & Ronner, 1978; Orban & Kennedy, 1980; DeValois, Yund & Helper, 1982). In addition, V1 cells are highly selective for direction of motion (Hubel & Wiesel, 1959), and particularly most of V1 cells had a particular orientation preference nearly perpendicular to the preferred direction of motion (Albright, 1984), resulting in the aperture problem (i.e., thus, the local motion arising from over extended contours is ambiguous within an aperture. In order to overcome this ambiguity, it has suggested that local motion signals should be pooled across space, see Kane, Bex, & Dakin, 2009; Dakin, 2009; Pack & Born, 2001). Therefore, it is reasonable to assume that there are more V1 cells tuned to cardinal directions than oblique directions. Furthermore several researchers presented several models consisting of lateral inhibitory interaction between channels tuned to different pattern motions which would be considered as a part of the gain control process (e.g., Kim & Wilson, 1997; Boynton & Foley, 1999; Meese & Holmes, 2007). Based on these two crucial information, it can be logically inferred that the robust change in motion direction perception (the oblique effect and the inverse oblique effect for motion perception, Dakin et al., 2005; Wilson et al., 2001) might result



from the dynamic anisotropically-weighted gain control pool (i.e., strong at cardinal directions) whose effect is pronounced when viewing a pattern having a large direction variability (high noise). The additional pattern motions moving various directions creates more response activity in channels contributing to the gain control pool which results in the inverse oblique effect anisotropy in motion direction perception, whereas the response activity in channels is not significant with a pattern moving one direction (a small direction variability) which results in the oblique effect anisotropy. In addition, the anisotropically distributed motion direction energy of dynamic natural scenes (Dakin et al., 2005) indicates the most physical energy (amplitude or contrast) at the cardinal and least at the oblique directions of motion (i.e., more energy at cardinal directions of motion than oblique directions also provide a strong insight that there are possibly more cells tuned to cardinal directions of motion than oblique directions of motion at the early visual stage).

Based on these three crucial information—i.e., (1) it can be logically inferred that more cells are tuned to cardinal directions of motion, (2) some suppressive interaction between channels tuned to different pattern motion indeed exists (Kim & Wilson, 1997; Boynton & Foley, 1999; Meese & Holmes, 2007), and (3) dynamic natural scenes are biased in cardinal directions of motion (i.e., cardinal > oblique direction bias, Dakin et al., 2005), it can be logically suggested that the significant change in motion direction perception might be due to the dynamic anisotropically-weighted gain control pool which can be thought to a very efficient coding strategy of the early visual motion processing stage that serves to equalize the neural representation of dynamic natural scenes with respect to direction of motion. To verify this assumption, further physiological researches

are highly encouraged. Specifically, this dynamic anisotropic gain control processing would equalizes the prevalent motion direction energy (i.e., cardinal > oblique motion direction bias, Dakin et al., 2005) in the dynamic natural scenes as increasing suppression at the cardinal orientation results in less accuracy or sensitive performance at the cardinal directions, showing the inverse oblique effect (Dakin et al., 2005; Wilson et al., 2001). To verify this assumption, further psychophysical and physiological researches are highly encouraged.

### 1.3.2. The Nature of Anisotropic Gain Control Mechanism

It is clearly that the anisotropic gain control pool efficiently encodes the most prevalent image structure with respect to the orientation dimension. In other words, gain control mechanism is local (*'tuned'*) in the orientation dimension that causes the most suppression at horizontal and least at obliques. Subsequently, this well-documented efficient coding mechanism, which indeed provides the mechanism of neural encoding underlying the human orientation perception, has been explored with respect to temporal frequency (Kim et al., 2010) and spatial frequency (Haun & Essock, 2010). However, we still didn't seem to know the whole story of this processing. Specifically, whether multiple anisotropic gain control mechanisms indeed exist across the different spatio-temporal surface, if exist whether all mechanisms are tuned to the temporal and spatial frequency dimensions, and if does so what about their tuning shape.

In order to clarify these issues, in this dissertation research we provide two masking experiments which would provide plausible answers. Before moving to the main experiments, we firstly provide a brief description of prior masking studies (Boynton &

Foley, 1999; Meese & Holmes, 2007; Kim et al., 2010) which show the temporal frequency tuning property of the gain control mechanism in order to estimate what is the current knowledge with this topic. Finally, we compare the temporal tuning shape (i.e., one crucial issue of how systematically the peak masking position of the target change with the mask temporal frequency) with one obtained by prior masking studies (Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). Specifically, one crucial issue of how the peak masking position of the target systematically changes with the mask temporal frequency in order to clearly understand whether only two (or three, Hess & Snowden, 1992) temporal frequency-tuned mechanisms exist underlying the human temporal processing. If they are different, we should consider why they should be different. Secondly we provide a brief description of our prior masking study which show the spatial tuning property of the anisotropic gain control mechanism (Haun & Essock) and a plausible reason why the anisotropic suppression (the horizontal effect) is observed at only the particular spatial frequency (i.e., around 8 cpd, Haun & Essock, 2010).

### 1.3.3. Goal of Present Experiments

#### a. Temporal Tuning Property of Gain Control Mechanism (Experiment 1)

##### i. Relation to Prior Studies

The nature of the anisotropy in the gain control pool has been investigated with respect to two distinctive temporal properties (i.e., at low-speed (8 cpd/1 Hz) and high-speed (1 cpd/16 Hz), Kim, Haun, & Essock, 2010) by measuring broadband masking for both low-speed and high-speed test gratings at either 1 or 8 cpd. The results showed that the

broadband masking are significantly higher when the temporal properties of the mask and test stimuli matched, and in this case the broadband masking effect are significantly higher for the horizontal test and least for the oblique orientations (the horizontal effect). In addition, very little cross-channel suppression on the low-speed test was caused by the high-speed mask, and vice versa. These results provide a strong insight that the orientation-biased gain control pools are tuned to two distinctive temporal properties—i.e., the low speed (low spatial/high temporal frequencies) and high-speed (high spatial/low temporal frequencies). Similar results are also reported by Boynton and Foley (1999) by measuring the TvC function for each test and mask in different spatial and temporal combinations (i.e., the orientation and spatial frequency of the test and mask are the *same*, but the mask contrast is only varied). It was reported that when the test and mask stimuli have the low spatial and high temporal frequencies (high-speed), the rising portion of the TvC function is steeper (high contrast gain) than for the high spatial and low temporal frequencies (low-speed). In addition, by measuring a cross-orientation masking effect, the similar results was reported that the strongest cross-orientation masking effect at suprathreshold was obtained at the high-speed (the low spatial/high temporal frequencies), and was also observed across different speed regions with different strength (see Figure 3 in Meese & Holmes, 2007). Taken together, one agreement across these works (Boynton & Foley, 1999; Meese & Holmes, 2007; Kim et al., 2010) provides a *critical* hint that the gain control pool is *tuned* to the different speed, and at least two classes of the speed tuned-gain control pools exist at different spatial and temporal locations.

However, the prior studies which show the temporal tuning property of the gain

control pool (i.e., Boynton & Foley, 1999; Meese & Holmes, 2007) have been tested with only limited stimulus dimensions (i.e., both of these studies used simple stimuli, such as a fixed spatial frequency and orientation, that create limited responses from encoding-mechanisms underlying the human visual system). With more realistic visual stimuli such as found in natural environments, on the other hand, these stimuli stimulate the encoding-mechanisms in a way that is similar to natural stimulation (Bex et al., 2007; Felsen & Dan, 2005; Baker & Issa, 2005; Basole et al., 2003). By doing this, we can test more realistic temporal tuning property of this gain control pool. Experiment 1 systematically investigates whether this gain control pool is tuned to different temporal rates using more realistic visual stimuli which can create more realistic response patterns from many encoding-mechanisms underlying the human early visual stage (Baker & Issa, 2005; Basole et al., 2003). Particularly, a prior study (Kim et al., 2010) reported that there are at least two distinct gain control pools and these pools show the horizontal effect anisotropy. However, these results have not been tested with a range of temporal frequencies (only two speeds were used in our prior study), and comparatively little is known of broadband masking in terms of temporal frequency.

## ii. Experiment 1

In order to clarify this issue, Experiment 1 was conducted. Experiment 1 examines whether the anisotropic masking (showing the horizontal effect) which was thought to be elicited by the orientation biased gain control pool, (Essock et al., 2009; Kim et al., 2010) for each test at three orientations at each 10 spatio-temporal combinations by using

spatially broadband mask across varied temporal rates of the mask (in order to get temporal frequency tuning curves for each test condition). Specifically, each test is a single sinusoidal grating at 10 spatial and temporal combinations (i.e., test is narrowband in space and time) at one of the three orientations ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ). The mask stimuli are oriented  $1/f$  broadband patterns at the same orientation as the test, which is modulated in time with conditionally chosen single temporal frequencies which judged to be most important. The orientation band is  $15^\circ$  and includes spatial frequencies from 1 to 16 cpd (i.e., the mask is, thus, broadband in space, but narrowband in time). This mask stimulus specification is much broader than the mask stimuli that were used by prior studies (Boynton & Foley, 1999; Meese & Holmes, 2007), and thus we can test more realistic temporal tuning property of the gain control pool. We evaluate (1) whether broadband masking, and specifically anisotropic masking, is temporally tuned (i.e., whether the peak masking effect would appear at the same temporal frequency as the test stimulus for a variety of test temporal rates), and (2) whether this anisotropic broadband masking, and the horizontal effect is observed at all of the 10 spatio-temporal combinations of the test (i.e., whether the multiple orientation-biased gain control pools showing the horizontal effect exist at the 10 spatial and temporal combinations). Three temporal frequency tuning curves (one at each of three test orientations) at the 10 spatial and temporal combinations are. If many orientation-biased gain control pools (driving the horizontal effect) indeed exist across different spatio-temporal locations, and all pools are tuned in the temporal frequency dimension, then we should be able to observe single peaks in masking when the temporal frequency of the test and mask stimuli match. At these many peaks, we should be able to observe the significant horizontal effect anisotropy.

Finally, in the results section we compare the temporal tuning shape (i.e., one crucial issue of how systematically the peak masking portion of the target changes with the mask temporal frequency) obtained in the current Experiment 1 with the temporal tuning shape obtained by prior masking studies (Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009), and if they are different, we provide a plausible reason why they should be different.

## b. Spatial Tuning Property of Gain Control Mechanism (Experiment 2)

### i. Relation to Prior Studies

The spatial frequency tuning property of this anisotropic gain control pool has been investigated. Previously, when the observer detected an oriented target, broadband contrast increment (i.e., in a 1-octave; 8-16 cpd spatial frequency band and a 20° orientation band) superimposed on a background of  $1/f$  isotropic noise (i.e., from 0.2 to 17cpd and *all* orientations), thresholds were lowest for obliquely oriented stimuli and highest for horizontally oriented stimuli (the horizontal effect, see Figure 4c in Essock et al., 2003). Continuously, we intent to understand what particular spatial frequencies within the broadband increment (Essock et al., 2003) were relied on by observers in performing the original task (i.e., determine which spatial frequencies most likely contribute to the anisotropic performance). To test this, contrast discrimination threshold for each of the five oriented-narrowband-spatial contents (i.e., in a 1-octave centered at 0.65, 1.50, 3.44, 7.90, and 18.25 cpd, and 45° orientation band) was measured with a background of  $1/f$  isotropic broadband noise (i.e., from 0.2 to 25.6 cpd, and all orientations) and without the mask (Haun & Essock, 2010). The results show that the

discrimination thresholds are lowest for target content around 8 cpd, and that at this spatial frequency a horizontal effect is seen that closely resembles the horizontal effect anisotropy observed in broadband masking (i.e., Essock et al., 2009; Kim et al., 2010). The magnitude of the horizontal effect decreased at both lower and higher spatial frequencies (see Figure 5(b) in Haun & Essock, 2010). Here, one might raise an intriguing question. Why was the horizontal effect anisotropy observed at only the particular spatial frequencies (around 8 cpd, Haun & Essock, 2010)? What aren't the low and high spatial frequency horizontal effects observed?

To answer these questions, it is necessary to understand the nature of the anisotropy in suppression at the local level. So far, it is well documented that the horizontal effect anisotropy is a perceptual bias in orientation perception, and the encoding processing underlying this anisotropy indicates that the suppressive interaction between orientation-tuned channels is strongest at horizontal, and least at oblique orientations. In the prior two works (Essock et al., 2003; Haun & Essock, 2010), one common factor is that this anisotropic suppression was measured by using particular 1-octave target contents on broadband background of  $1/f$  isotropic noise and further detail of the stimulus specifications from both works are provided in **Table 2**. Indeed, these two works seems to measure the *global* anisotropic suppression for a 1-octave target pattern caused by the broadband noise mask (i.e., this should be global when comparing the suppression for a single grating caused by broadband noise mask). In order to clearly understand this global nature of this anisotropy in suppression, at the local level understanding of the nature of the anisotropic suppression is fundamental. Specifically, we need to understand the local level's nature of this anisotropic suppression which can



be assessed by measuring the anisotropic suppression for a *narrowband* target (i.e., a single grating) by various spatial frequencies of the mask. By doing this, we can understand how systematically this local suppression changes across the different spatial frequencies of the mask. Furthermore, it can be checked whether the anisotropic suppression is indeed observed at particular target spatial frequency (i.e., around 8 cpd in Haun & Essock, 2010) and whether this anisotropic suppression is observed across the different target spatial frequencies. Based on this local level's knowledge of the anisotropic suppression, we might provide some plausible answer about these questions.

For this reason, Experiment 2 measure this *local* anisotropic suppression for very narrowband (single grating) targets by various spatial frequencies of the mask stimuli. This is indeed distinctive from the prior two works which measured the *global* anisotropic suppression for comparatively broad (1-octave) targets by broadband noise masks. Further description of Experiment 2 is addressed in below section.

With respect to the absence of the low spatial frequency horizontal effect (Haun & Essock, 2010), a strong hint, which supports this plausible explanation, has been provided by previous studies (Campbell & Robson, 1968; Stromeyer, Klein, Dawson, & Spillmann, 1982). One study (Campbell & Robson, 1968) showed that adapting to a low spatial frequency grating (below 1.5 cpd) caused maximal spatial adaptation at a significantly higher spatial frequency, not the same spatial frequency. Based on this result, Campbell and Robson (1968) suggested that there might not be adaptable spatial frequency channels tuned to below 1.5 cpd. Contrary to this view, Stromeyer et al. (1982) demonstrated that the poor sensitivity or adaptation for the low spatial frequency patterns (0.12-1 cpd) is enhanced by modulating them in time (e.g., flickering or moving). By

modulating low spatial frequency targets in time, one can increase the visibility of these low spatial frequency patterns. Further, one can observe that adapting or masking to the low spatial frequency gratings can cause maximal threshold elevations at the same low spatial frequency. This indicates that there are flicker-sensitive low spatial frequency-tuned channels and channel's tuning property would be elicited by using appropriate stimuli. Based on these two studies, it can be logically inferred that, in order to strongly stimulate low spatial frequency-tuned mechanisms, the use of an appropriate pattern is crucial.

## ii. Experiment 2

From these points of view, it can be carefully assumed that, if one uses an appropriate pattern which is strong enough to evoke the low spatial frequency tuned detecting-mechanisms, we should be able to obtain the low spatial frequency horizontal effect anisotropy. For instance, by modulating a low spatial frequency target in time, we might obtain the low spatial frequency horizontal effect, which was not demonstrated by Haun & Essock in 2010. In addition, the detection threshold for a high spatial frequency target pattern (18.3 cpd) was pretty high (around -38dB, see Figure 5(a) in Haun & Essock, 2010). This indicates that the observers might not see the oriented pattern itself well. In order to carefully discriminate whether the measured thresholds (Figure 5(b) in Haun & Essock, 2010) result from (1) the noise effect (the observers can not see the high spatial frequency target well, thus resulting in the higher discrimination thresholds) or (2) the pure suppression effect by neighboring orientation-tuned detectors, the follow-up control experiment seems to be needed in the following. Specifically, in order to rule out the

possibility that is that the observers can not see the high spatial frequency target well (the noise effect), one should re-measure the discrimination threshold for the same target (around 18.3 cpd) with an increased contrast. By doing this, the visibility of this target can be enhanced and one can measure the pure suppression effect by neighboring orientation-tuned detectors.

Based on these points of view, Experiment 2 re-evaluates the previous findings (Haun & Essock, 2010; Essock et al., 2003) by using a novel idea that is modulating visual stimuli in time (i.e., Sromeyer et al., 1982). Here, we assume that the static low spatial frequency stimuli used in a prior study (Haun & Essock, 2010) would not be strong enough to evoke the orientation-biased gain control pool at low spatial frequencies. By modulating the low spatial frequency tests in time, we would be able to enhance the visibility of the low frequency-tuned detector (test). Also, by using appropriate mask stimuli (i.e., spatially narrowband but temporally broadband), one would directly suppress the flicker-sensitive low spatial frequency-tuned detecting-mechanism (test), and that, at this low spatial frequency, a horizontal effect would be obtained. In order to systematically asses this issue, Experiment 2 systematically investigates the issue of how the anisotropic (the horizontal effect) masking across the 9 spatio-temporal combinations changes as a function of spatial frequency. In order to investigate this, the spatial frequency band of the mask stimuli is varied as a parameter. The full spatial frequency range of the  $1/f$  broadband noise image is the same broadband (4-octave) mask as in Experiment 1. The broadband mask was split into five 0.8-octave sub-band which centered at approximately 1, 2, 4, 8, 12 cpd. A set of narrow bands of spatial noise (0.8-octave) and the same broadband (4-octave) mask as in Experiment 1 were used to probe

spatial tuning. In addition, one additional low spatial frequency 0.8-octave sub-band, the 0.8 center spatial frequency was added only for condition 1 and condition 2. The other additional high spatial 0.8-octave sub-band, the 22 cpd center spatial frequency was added only for condition 6. Each mask pattern was, then, modulated by a 4.5-octave broadband temporal noise. We evaluate (1) whether the narrowband masking is spatially tuned (i.e., whether the peak masking effect would appear at the same spatial frequency of the test and mask stimuli), and (2) whether this anisotropic masking is observed at all of the 9 spatial and temporal combinations (i.e., whether the multiple orientation-biased gain control pools (showing the horizontal effect) exist across the 9 spatial and temporal combinations). Three spatial frequency tuning curves (one at each of three test orientations) at the 9 spatio-temporal combinations are measured. Therefore, one can indeed verify the hypothesis that anisotropic gain control pools are tuned to the spatial frequency dimension and how systematically this anisotropic suppression change across the different spatial frequencies.

## 2.0 METHODS

### 2.1 General

#### Apparatus

All stimuli were displayed on a monochrome (white P104 phosphor) M21LMAX Image Systems, Inc. CRT monitor. Mean luminance of the display was  $26 \text{ cd/m}^2$ , resolution was set to 800 by 600 pixels, and the refresh rate was 200Hz. A Vision Research Graphics, Inc. grayscale expander (Pelli & Zhang, 1991) was used to produce 12.4 bits of grayscale resolution. Stimuli were generated on a PC computer using Matlab 7.5.0 (MathWorks, Inc.) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). All stimuli were circular,  $4^\circ$  and 384 pixels in diameter viewed at a distance of 240 cm in a dark room. Stimuli were viewed through a circular surround in front of the monitor ( $20^\circ$  in diameter), which obscured the monitor bezel and other nearby contours.

#### Participants

There were two observers in each experiment. The author (TK) and one of the two naïve observers (RS and EJ) participated in each experiment. The author (TK) and one naïve (RS) observer participated in Experiment 1, and the author (TK) and the other naïve (EJ) observer participated in Experiment 2. All subjects had normal resolution acuity at all meridians and wore any necessary optical correction. A vision test was carried out to assure that they did not have any residual astigmatism. The age of the participants ranged between 22 and 31. Participants in all experiments provided informed consent as

approved by the University of Louisville's IRB. All naïve participants were paid \$10 per session. In order to familiarize the participants with the experiments, practice sessions were performed initially.

## Procedure

Each experiment used a 40-trial two-interval force-choice (2IFC) QUEST procedure (Watson & Pelli, 1983) to estimate the 82%-correct contrast threshold for Gaussian-windowed sinewave grating tests (Gabor). Each trial consisted of two stimulus intervals, both containing an identical mask. One interval (randomly selected) also contained the target presented concurrently with the mask. Both stimulus intervals contained either identical noise mask images (i.e., mask condition) or, in the baseline condition, an unpatterned mean-luminance background. The two sequential 1000 ms stimulus intervals of each trial were separated by a 500 ms ISI (Inter Stimulus Interval). All stimuli for both experiments were viewed foveally. The subjects were asked to fixate on a small circular centered spot present between trials and during the ISI (a 2-pixel-wide ring with an outer diameter of approximately 0.13deg). Participants were instructed to indicate the interval containing the test by pressing the '1' or '2' keyboard buttons. Auditory feedback was provided following incorrect responses.

All stimuli for both experiments were oriented, with the orientation specified in degrees in common polar coordinates, where 0° is horizontal, 45° is an up-and-to-the right oblique, and 90° is vertical.

## 2.2 Experiment 1: Temporal Frequency Tuning

## Overview

Experiment 1 investigated the anisotropic (horizontal effect) broadband masking for each test at three orientations at each of 10 spatial and temporal combinations. For each test condition, masking was examined by using a spatially broadband mask across varied temporal rates of the mask. We evaluated (1) whether broadband masking is temporally tuned (i.e., whether the peak masking effect would appear at the same temporal frequency of the test stimulus for a variety of test temporal rates), and (2) whether this broadband masking, and specifically anisotropic masking (the horizontal effect) is observed at all 10 spatial and temporal combinations of the test (i.e., whether the multiple orientation-biased gain control pools showing the horizontal effect exist at the 10 spatial and temporal combinations). Three temporal frequency tuning curves (one at each of three test orientations) at 10 spatial and temporal combinations were measured (i.e., 3 (temporal tuning curves at three test orientations) x 10 (10 spatio-temporal combinations) = 30 total temporal frequency tuning curves). If many orientation-biased gain control pools (driving the horizontal effect) indeed exist across different spatio-temporal location, and all pools are tuned in the temporal frequency dimension, then we should be able to observe many single peaks in masking when the temporal frequency of the test and mask stimuli match. At these many peaks, we should be able to observe the significant horizontal effect anisotropy.

### b. Stimuli Generation for Experiment 1

#### Test Stimuli

The tests were Gaussian windowed ( $1^\circ$  full width at half-height) sinewave gratings of

either 2, 8, or 12 cpd at each of three orientations ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ) (see B, D, and E in **Figure 6**). The tests were also Gabors in time, modulated within a Gaussian envelope (see A in Figure 2-1) that had a full width at half-height of 400 ms ( $\sigma = 170$  ms). In order to have a reasonable number of test conditions, the spatial and temporal combinations judged to be most important were used (see below). A test was modulated by one of the six (i.e., 2, 4, 10, 14, 16, or 20 Hz) different temporal frequencies in the following 10 key combinations (a total of 10 target conditions, see left panel in **Figure 8**) were selected. Specifically, the 2 cpd test was used at five temporal frequencies (4, 10, 14, 16, or 20 Hz). The 8 cpd test was used at three temporal frequencies (2, 4, and 10 Hz). The 12 cpd target was used at temporal frequencies (2 and 10 Hz). Additional control conditions were run at 2 cpd/16 Hz and 2 cpd/20 Hz. Note that the test stimuli were spatially and temporally narrowband patterns (a single frequency).

### Mask Stimuli

The masks consisted of oriented  $1/f$  broadband spatial noise images centered at the same orientation as the test (see 'A' in **Figure 8**). The spatial frequency band used was four octaves, including spatial frequencies from 1 to 16 cpd, and the orientation band was  $15^\circ$ . The 384 by 384 pixel images were created via an inverse Fourier transform of  $1/f$  amplitude spectra, with randomized phase coefficients generated on each trial (but the same on the two intervals of a single trial), which then were multiplied by a rectangular bandpass filter in orientation (see Essock et al., 2003 or Hansen and Essock, 2003 for more details). Contrast of the mask in Experiment 1 was set to the standard deviation of normalized pixel luminances, or root-mean-square contrast (RMS), of 0.10 (with pixel



luminance normalized to range from 0 to 1). The spatially broadband stimulus contrast then was temporally modulated by various (temporal) cosine waveforms (see B-I in **Figure 7-1**) with the same temporal Gaussian envelope (see ‘A’ in **Figure 7-1**) as the test. Thus, each mask had a temporal frequency of either 1, 2, 4, 10, 14, 16, 20 or 25 Hz. Note that the mask stimuli were spatially broadband (i.e., 4-octaves in spatial frequency band and 15° in orientation band), but temporally narrowband (a single temporal frequency).

#### Mask Contrast

Broadband images used in Experiment 1 were constructed as numerical matrices with means of zero, so they contained both positive and negative deviations from the mean. Then, the average standard deviation across a large set of randomly generated matrices was measured, and this value was used as a ‘normalization constant’. Then, each noise image generated during the experiment was divided by this constant, which resulted in a matrix with, on average, a standard deviation of 1.0. This matrix was then multiplied by the RMS contrast value desired (0.10 in Experiment 1), and the mean luminance value was added to the matrix.

#### c. Psychophysical Paradigm for Experiment 1

The masking paradigm, as described above in Section 1.3.1.a.i, was used through the experiments.

#### d. Data Analysis for Experiment 1

In each test condition, contrast thresholds were measured for the test both with and

without the mask. A measure of masking (i.e., suppression factor, Petrov, Carandini, & McKee, 2005; Essock et al., 2009) was obtained by dividing the masked threshold by the unmasked threshold for each condition. Then, in order to directly compare the magnitude of the horizontal effect across spatio-temporal conditions, normalized suppression (i.e., each suppression value was divided by the single maximum suppression value for each target condition irrespective of orientation) was calculated which is referred to a normalized suppression for peak (NSp). The tuning of masking with normalized suppression for peak (NSp) values across temporal frequency was assessed by fitting curves to the masking functions. A modified Gaussian function (see below Equation 3, i.e., Cass & Alias, 2006; Cass, Stuit, & Bex, 2009) with four parameters (i.e., amplitude ( $A$ ), vertically shift constant ( $C$ ) and bandwidth ( $W$ : Full Width at Half Height (FWHH), center position/or frequency ( $x_0$ )) provided a best fit to full data (NSp values), which well describes the shape of the temporal frequency tuning curves at the 10 spatial and temporal combinations. In symbols, a modified Gaussian function (of  $x$ ) can be written

$$G = C + A \cdot \exp \left[ -(-\log(0.5)) \cdot 4 \cdot \left( \frac{x - x_0}{W} \right)^2 \right] \quad (3)$$

where  $A$  is the amplitude of modified Gaussian function.  $C$  is vertical shift constant, which is also termed to an isotropic amplitude component (i.e., Cass, Stuit, & Bex, 2009).  $W$  is the bandwidth (FWHH). Note that the constant (i.e.,  $-\log(0.5) \cdot 4 \approx 2.77$ ) is used to make the measure of width  $W$  be the full width at half height, see Graham, 1989).  $x_0$  is the position of the center of the peak. The shape of temporal frequency tuning curves across the 10 spatial and temporal combinations was compared with respect to the amplitude ( $A$ ) + vertical shift constant ( $C$ ), bandwidth ( $W$ ), and center frequency ( $x_0$ ). In

each of these models peak magnitude is described by the sum of the amplitude ( $A$ ) and vertical constant ( $C$ ) parameters. Specifically, the use of a vertical shift constant ( $C$ ) parameter (i.e., termed as the un-tuned suppression in Dakin & Mareschal, 2000; Cass & Alias, 2006; Cass, Stuit, & Bex, 2009) provides improved fits. Specifically, this un-tuned part of suppression functions have been reported previously with respect to the temporal frequency (see Figure 4 in Cass & Alias, 2006), orientation (see Figure 3 and 5 in Cass, Stuit, & Bex, 2009;) and spatial frequency (see Figure 12 in Dakin & Mareschal, 2000). In order to check whether the anisotropic broadband masking is indeed tuned in temporal frequency dimension, the center frequency ( $x_0$ ) parameter was set to vary. If multiple temporally-tuned suppressive pools indeed exist across different spatio-temporal target locations, this parameter will vary with the target temporal frequency.

In order to provide a metric of the horizontal effect, an index, Horizontal Effect Index (HEI, see Kim, Haun, & Essock, 2010) was created to express the horizontal effect in normalized suppression as a proportion of the general amount of suppression (taken as suppression at the oblique) in a given condition. Horizontal Effect Index (HEI), which was calculated by taking the difference in normalized suppression at the horizontal and the oblique orientation and then dividing by the normalized suppression at the oblique orientation. In addition, Vertical Effect Index (VEI) was calculated by taking the difference in suppression at the vertical and the oblique orientation and then dividing by the suppression at the oblique orientation. Also, H-V Effect Index (H-VEI) was calculated by taking the difference in suppression at the horizontal and vertical orientation and then dividing by the suppression at the oblique orientation. Furthermore, in order to compare tuning bandwidth (FWHH) across three test orientations within each

spatio-temporal combination and across the spatio-temporal combinations, suppression was normalized separately for each orientation (in each test condition) which is referred to a normalized suppression for bandwidth (NSb). Then, these values are used to fit a Gaussian function again, and from this fitting the bandwidth (FWHH) parameter will be used to compare across three test orientations within each spatio-temporal combination and across the spatio-temporal combinations in order to estimate the temporal tuning width of the temporally-tuned gain control pools across the orientations and across the spatio-temporal conditions.

An 'lsqnonlin' function (Optimization Toolbox in Matlab 7.5.0 from Math Works, Inc.) was used to find the best fitting parameters by minimizing the sum of the squares residuals (the Sum of Squared Errors, SSE) between the data and the Gaussian function fit. Specifically,  $\text{bestparameters} = \text{lsqnonlin}(@\text{function}, \text{paramest}, \text{lowerbound}, \text{upperbound})$  starts at the 'paramest' which is a starting position in the parameter space, and defines a set of lower and upper bounds on the design variables in 'bestparameters' so that the solution is always in the range 'lowerbound'  $\leq$  'bestparameters'  $\leq$  'upperbound'. Finally, the function (i.e., @function) calls lsqnonlin in order to find a minimum of the sum of squares of the functions described in function (i.e., a Gaussian function with four parameters in current work). In addition, this algorithm starts with estimated values (i.e., starting points) for each parameter (i.e., amplitude ( $A$ ), vertical shift constant ( $C$ ), bandwidth ( $W$ ), and center frequency ( $x_0$ )). The estimation for starting points was made by using a heuristic approach to obtain reasonable starting values. Specifically, one can look at tuning curves of the full set of the data in an Excel spreadsheet, and one can estimate the maximum and minimum values for each parameter, and

see how they influence the shape of the tuning curves.

If the multiple orientation-biased gain control pools (the horizontal effect) indeed exist across the spatial and temporal combinations, and all pools are tuned in the temporal frequency dimension, then we should observe a single peak of the masking function centered at the test's temporal rate. Specifically, if the multiple temporal frequency-tuned detecting-mechanisms indeed exist, broadband masking effect would be greatest with the test and mask stimuli of the same temporal frequency. By changing the temporal frequency of broadband mask stimuli, we can measure how much a test threshold is elevated. This is taken as an index of the response of the temporally-tuned detecting-mechanisms to this stimulus. Therefore, a single peak masking position in a temporal frequency tuning curve points out the peak sensitivity of each of the underlying temporal frequency detecting-mechanisms. Furthermore, based on prior study (Kim et al., 2010), we showed that the anisotropic (the horizontal effect) broadband masking is tuned in two distinctive temporal properties (i.e., at low-speed (gradual) and high-speed (16 Hz flickered)). This provides a strong insight that the anisotropic broadband masking would be tuned in different temporal properties. Within these lines of thought, at these multiple peaks in which the masking effect is greatest with the test and mask stimuli of the same temporal frequency, we should be able to observe the significant horizontal effect. In order to check whether anisotropic suppression (the horizontal effect) is significant, one-way ANOVA (i.e., one-factor for the orientation in which there are 3 levels (vertical, 45, and horizontal)) will be first applied to the full set of the normalized suppression values (NSp). When the main effect of orientation is significant, the follow-up statistic, a planned contrast comparison, will be performed. Furthermore, in order to show how the

temporally-tuned suppression is tightly tuned in temporal frequency dimension, the center position ( $x_0$ ) parameter will be provide to compare this with the target temporal frequency.

In addition, in order to investigate how the bandwidth is significantly changed across three test orientations within each spatio-temporal combination and across the 10 spatio-temporal combinations, one-way ANOVA (i.e., one-factor for the bandwidth at three orientations in which there are 3 levels) will be first applied to the full set of the measured bandwidth values at three orientations from the Gaussian fitting in order to check whether the main effect of the bandwidth across the 10 spatio-temporal combinations is significant. And the follow-up statistics, a planned contrast comparison will be performed to show whether the bandwidth effect at three orientations within each spatio-temporal condition is significant.

## 2.3 Experiment 2: Spatial Frequency Tuning

### a. Overview

Experiment 2 systematically investigated the anisotropic (the horizontal effect) masking across 9 spatial and temporal combinations as a function of spatial frequency. In order to investigate how the tuning along spatial dimension changes across the different target spatio-temporal combinations, the spatial frequency band of the mask stimuli is varied as a parameter. The same broadband (4-octave) mask as in Experiment 1 was split into five 0.8-octave sub-bands. The spatial frequency bandwidth of each sub-band centered at approximately 1, 2, 4, 8, and 12 cpd. Each narrow sub-band of spatial noise (0.8-octave)

and the same broadband (4-octave) as in Experiment 1 masks were used to probe spatial tuning. In addition, one additional low spatial frequency 0.8-octave sub-band, the 0.8 cpd center spatial frequency was added only for condition 1 and condition 2. The other additional high spatial frequency 0.8-octave sub-band, the 22 cpd center spatial frequency was added only for condition 9. A temporally broadband mask was also used. Each mask noise pattern was modulated by a 4.5-octave broadband temporal noise. We evaluated (1) whether the masking is spatially tuned (i.e., whether the peak masking effect would appear at the same spatial frequency as the test), and (2) whether this masking is anisotropic and observed at all of the 9 spatial and temporal combinations (i.e., whether the multiple orientation-biased gain control pools (showing the horizontal effect anisotropy) exist across the 9 spatial and temporal combinations). Three spatial frequency tuning curves (one at each of three test orientations) at the 9 spatio-temporal locations were measured (i.e., 3 (spatial frequency tuning curves at three test orientations) x 9 (9 spatio-temporal locations) = 27 total spatial frequency tuning curves). Therefore, we can indeed verify the hypothesis that anisotropic gain control pools are tuned to the spatial frequency dimension and how systematically this anisotropic suppression changes across the different spatial frequencies.

## b. Stimulus Generation for Experiment 2

### Test Stimuli

All stimuli settings were identical to those used in Experiment 1, except that they differed in the following ways. Five test gratings of either 1, 2, 4, 8, or 12 cpd at each of three orientations were used (see A-E in **Figure 6**). Each was tested at various temporal rates (a

Gaussian envelope as in Experiment 1 was used; see **Figure 7-1**). In order to make a reasonable number of test conditions, particular spatial and temporal combinations were used here. Each test was modulated by different temporal frequency in the following 9 key combinations: at 1 cpd flickered at 1 Hz or 16 Hz; at 2 cpd flickered at 10 Hz or 16 Hz; at 4 cpd flickered at 1 Hz or 10 Hz; at 8 cpd flickered at 1 or 16 Hz; at 12 cpd flickered at 10 Hz was used.

### Mask Stimuli

The same broadband mask as in Experiment 1 was used ( $1/f$  broadband mask stimulus was from 1 to 16cpd, and a  $15^\circ$  orientation band) (see A in **Figure 8**), except that it was split into five 0.8-octave sub-bands whose center spatial frequency was approximately 1, 2, 4, 8 and 12 cpd (see B-F in **Figure 8**). Specifically, the spatial frequency range of each sub-band was from 1 to 1.74 cpd, from 1.74 cpd to 3.03, from 3.03 to 5.28, from 5.28 cpd to 9.19, from 9.19 to 16 cpd. Furthermore, one additional low spatial frequency 0.8-octave sub-band, the 0.8 cpd center spatial frequency was added only for condition 1 and condition 2 (i.e., in order to avoid low spatial frequency artifact, the viewing distance for a particular 0.8-octave sub-band (including spatial frequency ranging from 1.15 to 2 cpd), and the stimulus size and a target frequency were adjusted in order to keep the properties of display constant). The other additional high spatial frequency 0.8-octave sub-band, the 22 cpd center spatial frequency (including spatial frequency ranging from 16 to 27.86 cpd) was added only for condition 6. Each test spatial frequency was centered at one of the five (or six for condition 6) sub-band masks. Each sub-band had 0.045 RMS contrast and a broadband mask was 0.1 RMS.



Each mask stimulus was modulated in time with a broadband temporal-noise sample, which had a  $1/f$  temporal amplitude spectrum (i.e., Cass, Alais, Spehar, & Bex, 2009; Dong & Atick, 1995). Specifically, each temporal-noise sample was generated by inverse Fourier transform of constructed  $1/f$  amplitude spectrum, with randomized phase coefficients generated on each trial (but the same on the two intervals of a single trial). The mask included temporal frequencies from 1 to 22.22 Hz. Then, each temporal-noise sample was presented in time modulated by the same temporal Gaussian envelope as the test stimulus (see A in **Figure 7-1**; and the same as that used in Experiment 1).

#### Mask contrast

In order to keep the local contrast of the five (or six)  $1/f$  0.8-octave sub-bands and the 4-octave  $1/f$  mask constant, the contrast of each sub-band was calculated in the following way. Broadband contrast ( $BBcontrast$ ) can be expressed as:

$$\begin{aligned}
 BBcontrast &= \sqrt{S \times NBcontrast^2} \\
 \left( \begin{aligned}
 \leftrightarrow BBcontrast^2 &= S \times NBcontrast^2 \\
 \leftrightarrow NBcontrast^2 &= \frac{BBcontrast^2}{S} \\
 \leftrightarrow NBcontrast &= \sqrt{\frac{BBcontrast^2}{S}} \\
 \leftrightarrow NBcontrast &= \frac{BBcontrast}{\sqrt{S}}
 \end{aligned} \right)
 \end{aligned}$$

where  $S$  is the number of sub-bands,  $NBcontrast$  is the contrast of the sub-bands, and  $BBcontrast$  is the contrast of the broadband image. Thus, the narrowband contrast ( $NBcontrast$ ) is equal to the broadband contrast divided by the square root of a number of sub-bands ( $S$ ). Consequently, the contrast of each of the five sub-bands will be 0.045 RMS. If added together the contrast will be the same as the broadband contrast of 0.1

RMS. The contrast of broadband mask was already described above in Experiment 1.

In order to keep the total energy of each temporal-noise sample the same, the total energy ( $E_t$ ) of each temporal-noise sample was calculated in the following way. The total energy ( $E_t$ ) of each temporal-noise sample can be expressed as:

$$E_t = \left( \frac{\left( \frac{S_t}{\sigma_{S_t}} \right)}{\sqrt{2}} \right) = \left( \frac{S_t}{\sqrt{2} \cdot \sigma_{S_t}} \right)$$

where  $S_t$  is the  $i$ th temporal-noise sample, which is generated on each trial ( $i = 1, 2, 3, \dots$ ), and  $\sigma_{S_t}$  is the standard deviation of each temporal-noise sample. Each temporal-noise sample ( $S_t$ ) generated on each trial in Experiment 2 was constructed as a numerical array with a mean of zero, so they contain both positive and negative deviation from the mean. Then, the standard deviation ( $\sigma_{S_t}$ ) of each temporal-noise sample was measured, and each temporal-noise sample ( $S_t$ ) was divided by this standard deviation ( $\sigma_{S_t}$ ), which resulted in an array with a standard deviation of 1.0. Then, this temporal-noise sample having a standard deviation of 1.0 was modulated by a temporal Gaussian envelope (see A in **Figure 7-1**), and then finally multiplied by a constant value ( $1/\sqrt{2}$ ) in order to keep the total energy of a windowed cosine same as the temporal noise sample. Specifically, the energy of a windowed temporal cosine for a target is around 0.35 which is similar to the value of the energy of the temporal noise sample for a mask (i.e., the energy of a windowed temporal cosine of the target ( $\approx 0.35$ ) is close to the value we get for the windowed temporal noise samples ( $\approx 0.39$ ), see **Figure 7-2A**. Without the multiplied by this constant value ( $1/\sqrt{2}$ ), the total energy of the windowed each temporal-noise sample is constantly larger ( $\approx 0.53$ ) which is larger than the energy of a windowed cosine ( $\approx 0.35$ ).

### c. Psychophysical Paradigm for Experiment 2

The masking paradigm, as described above in Section 1.3.1.a.i, was used through the experiments.

### d. Data Analysis for Experiment 2

A Gaussian function with four parameters (see above in Section 2.2.d) provided a best fit to full data, which describes the shape of the spatial frequency tuning curves across the 9 spatio-temporal locations. The shape of spatial frequency tuning curves across the spatio-temporal locations was compared with respect to the magnitude and bandwidth.

If the multiple orientation-biased gain control pools (the horizontal effect) indeed exist across all of the spatio-temporal locations, and all pools are tuned in the spatial frequency dimension, then we should be able to observe the single peaks masking positions when the temporal frequency of the test and mask stimuli match. At these multiple peaks, we should be able to observe the significant horizontal effect. In order to check whether anisotropic suppression (the horizontal effect) is significant, one-way ANOVA (i.e., one-factor for the orientation in which there are 3 levels (vertical, 45, and horizontal)) was first applied to the full set of the peak normalized suppression values (NSp). When the main effect of the orientation was significant, the follow-up statistic, which is a planned contrast comparison, was performed. Specifically, we checked whether suppression at the horizontal orientation is significantly greater than at other orientations.

In addition, in order to investigate how the bandwidth is significantly changed across three test orientations within each spatio-temporal combination and across the 10

spatio-temporal combinations, one-way ANOVA (i.e., one-factor for the bandwidth at three orientations in which there are 3 levels) will be first applied to the full set of the measured bandwidth values at three orientations from the Gaussian fitting in order to check whether the main effect of the bandwidth across the 10 spatio-temporal combinations is significant. And the follow-up statistics, a planned contrast comparison will be performed to show whether the bandwidth effect at three orientations within each spatio-temporal condition is significant.

### 3.0 RESULTS AND DISCUSSION

#### 3.1 Experiment 1

##### a. Overview

Although the existence of the multiple (or many) and narrowly tuned spatial frequency (Campbell & Robson, 1968) and orientation detectors (Blakemore & Campbell, 1969; Campbell & Kulikowski, 1966) is a reasonably well established property of spatial processing, it is still not clear whether the same principle holds for temporal processing underlying the human vision at the early cortical stage (V1). In the temporal processing point of view, it has widely-documented that there are no multiple and narrowly tuned temporal frequency detectors, but probably a few (i.e., 2 or 3) and broadly tuned temporal detectors (Robson, 1966; Kelly, 1972; Koenderink & van Doorn, 1979, Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). In spite of the similarity between the psychophysical contrast sensitivity functions on these two dimensions (i.e., the smooth transition in spatial and temporal frequency sensitivity functions), why did prior studies assume that these two dimensions are entirely different? Are there indeed a limited number of the temporal frequency tuned detectors underlying the human temporal processing? If there are indeed a few and broadly tuned temporal detectors, how could we perceive all of the different temporal information in everyday viewing situations? If only a few temporal frequency-tuned detectors at the early stage of the human temporal processing exist, it should be logically true that detectors tuned to the temporal frequency dimension are much less selective than

channels tuned to the spatial-frequency dimension. Is that really true? If it is really true, as Graham (1989) was described in a her excellent book, the temporal frequency selectivity might be a by-product of the existence of multiple detectors on the spatial frequency dimension coupled with some *interaction* between temporal and spatial frequency (i.e., the similar description was also provided by using a sensitivity scaling hypothesis, see figure 1 in Hess & Snowden, 1992). Specifically, the detectors sensitive to low spatial frequencies are quite sensitive to high as well as to low temporal frequencies, unlike the detectors sensitive to high spatial frequencies are sensitive to low temporal frequencies (i.e., the bandpass and lowpass dichotomy, see Robson, 1966).

In this view, most prior masking studies (i.e., Lehky, 1985; Hess & Snowden, 1992; Snowden & Hess, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Langly & Bex, 2007; Cass et al., 2009) have used very narrowband mask stimuli and have shown only two (or three) temporally tuned detectors (i.e., the peak masking effect constantly appears at one of the 2 (or 3) common temporal frequencies which are approximately at 1-2 Hz and 8-12 Hz (or a very high temporal frequency which is termed as a third channel, see figure 4-7 and 12 in Hess & Snowden, 1992). Here, one plausible question is, what would we be observed the temporal masking function when we use more realistic (i.e.,  $1/f$  or broadband) patterns as mask stimuli instead of a narrowband mask stimuli? Are the same results (i.e., one of the 2 or 3 temporal peak masking effects) observed? In this view, several interesting neurophysiological findings (i.e., Basole et al., 2003; Baker & Issa, 2005) provide an interesting prospective of how the organization of spatial (including spatial scale and orientation) and temporal information are encoded by the primary visual cortex (V1). The common agreement across neurophysiological findings is

that any arbitrary combinations of stimulus features within a complex image are efficiently represented in small receptive fields of cortex. Further studies using more realistic stimuli resembling the statistical properties of natural scenes are highly encouraged in order to further investigate this issue. Although the main purpose of Experiment 1 is not to answer these questions, we would be able to answer these interesting questions at the end.

Previously, we have suggested that a lower-level suppression exist (Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010; Kim, Haun, & Essock, 2011) that is anisotropic and becomes significant when driven by many spatial components (i.e., content broadband in spatial frequency and orientation). However, this anisotropic low-level suppression is apparently not revealed by a narrowband mask stimulus as in prior studies (i.e., Lehky, 1985; Hess & Snowden, 1992; Fredricken & Hess, 1998; Cass & Alias, 2006). Here, we assess the temporal frequency tuning property of these anisotropic pools. In order to test this, Experiment 1 is designed. The main purpose of Experiment 1 is to investigate the anisotropic (the horizontal effect) masking for each test at three orientations at the 10 fixed spatial and temporal combinations. For each test condition, masking was examined by using a spatially broadband mask across varied temporal frequencies of the  $1/f$  broadband mask. Then, we evaluate (1) whether the broadband masking is temporally tuned (i.e., whether the peak masking effect would appear at the same temporal frequency of the test grating for a variety of test temporal frequencies), and (2) whether this broadband masking, and specifically anisotropic masking (the horizontal effect) is observed at all of the 10 fixed spatial and temporal combinations of the test grating (i.e., whether the multiple anisotropic gain control pools exist across the

10 fixed spatial and temporal combinations). Three temporal frequency tuning curves (one at each of three test orientations) at the 10 spatio-temporal combinations were measured.

#### b. Thresholds

Contrast thresholds in each condition for the two subjects are averaged and are shown with mask in **Figure 10A** and without mask (baseline) in **Figure 10B**. As to be expected, the baseline (no mask) conditions (**Figure 10B**) show an oblique effect of contrast threshold at the higher spatial frequencies (i.e., with 8 cpd and 12 cpd targets, highest thresholds are at oblique orientations). The oblique effect is not present at the lower spatial frequency (i.e., 2 cpd, equal thresholds across orientations). As can be seen in **Figure 10A**, masked contrast thresholds are seen to be highest when the temporal frequency of the test and mask matches. In these conditions, a horizontal effect of contrast threshold is shown in approximately all of the conditions except two (i.e., condition 1 and condition 5 which show the oblique effect). Furthermore, Condition 8 (i.e., a 2 cpd/16 Hz test) shows two peak masking effects at 10 Hz and at 16 Hz which is the same temporal frequency of the target. Also, Condition 10 (i.e., a 2 cpd/20 Hz test) shows two peaks masking effects at 10Hz and at 20 Hz which is the same temporal frequency of the target. To show the effect of the  $1/f$  broadband masks more directly, suppression factor (i.e., Petrov et al., 2005; Essock et al., 2009) was calculated by dividing the masked threshold by the unmasked threshold for each target condition in **Figure 11**. As can be seen in **Figure 11**, the broadband masking is constantly the greatest when the temporal frequency of the test and mask matches except two conditions (i.e.,



condition 8 and condition 10). Specifically, condition 8 and condition 10 show two peak masking effects at 10 Hz and at the target's temporal frequency (i.e., for condition 8 at 16 Hz and for condition 10 at 20 Hz).

c. Normalized suppression: Temporal frequency tuning

i. Amplitude ( $A$ ) with NSp values

In order to directly compare the magnitude of the horizontal effect across the 10 spatio-temporal conditions, the tuning of broadband masking with the normalized suppression values (NSp) across temporal frequency was assessed by fitting curves to the broadband masking functions. As can be seen in **Figure 12A**, the tuning of broadband masking with the normalized suppression (NSp) plotting on a log temporal frequency scale is always greatest when the frequency of the test and mask matches except two conditions (i.e., condition 8 and condition 10). In these temporally-matched conditions, the broadband masking is highly anisotropic ( $F(2,27) = 23.95, p = 0.00$ ) showing the significant horizontal effect pattern in the following. The broadband masking for horizontal targets is significantly greater than for oblique targets ( $t_{(27)} = 6.89, p = 0.00$ :  $0^\circ$  vs  $45^\circ$ ) and than for vertical targets ( $t_{(27)} = 3.99, p = 0.00$ :  $0^\circ$  vs  $90^\circ$ ). This indicates that the anisotropic broadband suppression (showing the horizontal effect) is tuned in the temporal frequency dimension. Each of the model parameters (i.e., bandwidth ( $W$ ), amplitude ( $A$ ), vertical shift constant ( $C$ ) and the center frequency ( $x_0$ )) are shown in **Figure 13**.

Interestingly, in the two target conditions whose temporal frequency is higher than 14 Hz (i.e., 2 cpd/16 Hz (condition 8) and 2 cpd/20 Hz (condition 10)), a common single peak masking effect is observed at 10 Hz, not each target's temporal frequency (see

**Figure 12A**). Informatively, the temporal tuning of the broadband masking (**Figure 12A**) is plotted on a logarithmic scale which is typically more compressed at the high temporal frequency side of the masking curves than the low temporal frequency side. Thus, even if the peak masking effect at high temporal frequency indeed exists, it is easy to neglect the peak masking effect at high temporal frequency which is clearly shown in a linear scale in the bottom part in **Figure 12B**. To address this issue, the same data for the condition 8 and condition 10 whose temporal frequency is higher than 14 Hz (see **Figure 12A**) are re-plotted on a linear scale and shown in **Figure 12B**. As can be seen a bottom in **Figure 12B**, when the same data (i.e., these two conditions) are re-plotted on a linear scale, two evident peak masking effects are shown at 10 Hz and at each target's temporal frequency. With respect to selecting an appropriate scale (i.e., log or linear etc), we need to put a more serious attention in order to escape misinterpretation of the given data.

In addition, in these two conditions plotting on the linear scale shown in the bottom part in **Figure 12B**, the broadband masking effect is significantly higher at 10 Hz than at each target's temporal frequency (i.e., the masking for a 2 cpd/16 Hz target is significantly higher at 10 Hz than at 16 Hz:  $t_{(2)} = 9.770, p = 0.010$ , and the masking effect for a 2 cpd/20 Hz target is significantly higher at 10 Hz than 20 Hz:  $t_{(2)} = 8.693, p = 0.013$  ). The temporally-tuned suppression as reflecting gain control pools whose output is combined and weighted to reflect the conventional temporal detector is greater at 10 Hz than each target's temporal frequency (i.e., each target's temporal frequency is higher than 10 Hz). This might be due to the maximal sensitivity of the conventional temporal channel which is typically appeared at around 8-12 Hz. Although the temporally-tuned suppression at 10 Hz is greater than each target's temporal frequency, there still exists the

second peak masking effect at each target's frequency. This indicates the existence of the high temporal frequency-tuned suppression as reflecting the high temporal frequency-tuned local gain control pool. However, this second peaking masking effect in the bottom part in **Figure 12B** is disappeared when plotted on a log scale in the top part in **Figure 12B**.

In this view, similar results showing the two single peaks of the masking function at 10 Hz and each target's temporal frequency (i.e., 28 Hz which is higher than 10-14 Hz) can be seen in the data of Fredericksen and Hess (1998). Their masking function (i.e., figure 5 in Fredericksen & Hess, 1998) was obtained by measuring masking for a test probe (i.e., 1 cpd/28 Hz) by using a noise samples were generated by two dimensional box-pass filtering with boxes centered at 1 cpd and at particular temporal frequencies. The box-pass was 0.5-octaves wide in spatial frequency (from 0.841 to 1.189) and 4 Hz wide in temporal frequency. In addition, when considering the masking function of Hess and Snowden (1992) (e.g., their figure 5D), the similar two peak masking effects can be also seen at 10 Hz and the target's temporal frequency (i.e., 25 Hz which is higher than 10-14 Hz). This masking function was obtained by measuring threshold for a target (3 cpd/25 Hz) by using a mask pattern (i.e., sum of 4 sinusoidal gratings: 0.3+1+3+15 cpd) across varied single temporal frequencies of the mask. Critically, one common factor across these two works and the current Experiment 1 is that the temporal tuning of the masking functions for a low spatial frequency target (i.e., 1 cpd for Fredericksen & Hess, 1998; 3 cpd for Hess & Snowden, 1992; 2 cpd for the current Experiment 1) were measured by the particular mask stimuli which were not very narrowband (i.e., not a single grating), but with a particular mask containing a range of spatial frequencies (i.e.,

including a range of spatial frequencies: 0.5-octaves wide in spatial frequency from Fredericksen & Hess, 1998, the sum of 4 gratings from Hess & Snowden, 1992, and 4-octaves wide in spatial frequency from the current Experiment 1, see **Table 1**). However, when considering a particular prior temporal masking function for a target grating by very narrowband masks (i.e., single gratings, see figure 5-7 in Lehky, 1985), the peak masking effect always appeared at one of the 2 (or 3) common temporal frequencies, not the target's temporal frequency. Based on these findings, it is reasonable to assume that the responses generated by a narrowband mask (Lehky, 1985) would not be adequate to evoke many (unknown) temporal frequency-tuned detectors whose sensitivity is lesser than the well-known and most sensitive conventional temporal detector, but would be adequate to evoke only these well-known 2 (or 3) most sensitive temporal frequency-tuned detectors. (Note that we can not totally exclude a possibility of the existence of the multiple temporal detectors underlying the human temporal sensitivity function. In this view, a plausible account suggested by Ascher & Grzywacz (2000) will be described in General Discussion. In this view, a recent neurophysiological work (i.e., figure 8 in Alitto & Usrey, 2004) supports this idea by demonstrating the distribution of the temporal frequency for cortical cells.) Therefore, in order to create more realistic response from many temporal detectors underlying the human vision at the early stage, the use of more realistic pattern, such as naturalistic ( $1/f$ ) broadband, seems to be fundamental. We assume that (1) there are many temporal detectors underlying the human vision at the early stage whose detection threshold is differ, and (2) the naturalistic broadband mask pattern would rise above a certain threshold of each of the underlying many temporal detectors. Thereby, naturalistic broadband pattern is adequate to drive temporally-tuned

suppression as reflecting multiple local gain control pools whose output is combined and weighted by neighboring temporal detectors (i.e., detectors tuned in the similar temporal frequencies).

From this point of view, the tuning of the broadband masking for particular grating targets shown in **Figure 12A** could be considered as an index (or measurement) which shows the shape of the tuning of the temporally-tuned suppression as reflecting the multiple local gain control pool whose output is combined and weighted by neighboring temporal detectors (i.e., detector tuned in similar temporal frequency). The current Experiment 1 (**Figure 12A**) clearly shows that there are multiple temporally-tuned suppressions as reflecting multiple local gain control pools and these pools are only revealed with the naturalistic ( $1/f$ ) broadband mask patterns (i.e., the peak masking effect is only observed when the temporal frequency of the test and mask matches). As described above, naturalistic ( $1/f$ ) broadband mask pattern which is adequate to rise above a certain detection threshold of each of the underlying many (unknown) temporal detectors and thereby many local pools whose output is combined and weighted by neighboring detectors (detectors tuned in similar temporal rate) are able to drive, whereas narrowband content pass through relatively unaltered with performance reflect the weightings that from the conventional temporal channel with maximal sensitivity at 8-12 Hz. In order to drive multiple temporally-tuned pools, the significant stimulation at the spatial pooling stage is crucial.

This idea can be used to account for the prior masking works using very narrowband mask have shown only 2 (or 3) temporally-tuned mechanisms. Specifically, when considering a particular prior temporal masking function for a target by very

narrowband masks (i.e., single gratings, see Lehky, 1985), the peak masking effect always appeared at one of the 2 (or 3) common temporal frequencies, not the target's temporal frequency. This temporal tuning of the narrowband masking (e.g., figure 5-7 in Lehky, 1985) could be considered as the tuning shape of the conventional temporal channel which is pretty distinct from the tuning of the broadband masking in the following way. The narrowband (a single grating) mask used in Lehky (1985) can only stimulate a few spatial detectors which is not adequate to create combined and weighted suppression by neighboring spatial detectors (detectors tuned in similar spatial frequency and orientation) at the spatial pooling stage. However, the temporal tuning of the broadband masking shown in **Figure 12A** and **12B** would be considered an index of the temporally-tuned suppression as reflecting multiple local gain control pools whose output is combined and weighted by the neighboring temporal detectors with maximum sensitivity around 8-12 Hz. The naturalistic broadband mask contains a wide range of spatial frequencies and orientations which can stimulate many spatial detectors and thereby the strong suppressive interaction by neighboring spatial detectors tuned to similar spatial frequency and orientation, and then these strong suppressive signals in the spatial pooling stage are fed into the next temporal pooling stage. In this temporal pooling stage, we assume that the multiple temporal frequency tuned detectors whose sensitivity is lesser than the maximal sensitive temporal detector appeared around at 8-12 Hz. If these lesser sensitive (many) temporal detectors are strongly stimulated by a particular mask (here naturalistic broadband mask pattern which is able to rise above a certain detection threshold of each of lesser sensitive many temporal detectors), temporal interaction between detector tuned to similar temporal frequency (i.e., indicating

temporally-tuned pool) can be occurred. **Figure 12A** show the tuning property of the temporally-tuned suppression as reflecting multiple local gain control pools. Thus, it can reasonable to suggest that in order to drive multiple temporally-tuned suppression, the use of the particular mask pattern which rise above a certain detection threshold of each of many and lesser sensitive temporal detectors and thereby create the strong suppression between temporally tuned the conventional temporal channel is crucial. The narrowband content passes through relatively unaltered with performance reflecting the weightings that form the conventional temporal channel which is most sensitive temporal channels.

In order to asses this idea, three control conditions are performed for the conditions in which two temporal peaks are observed (Control 8 and Control 10, **Figure 12A** and **12B**) by using conditionally varied spatial frequency of the mask pattern (see **Table 3**). First, Control 1 measures the temporal frequency tuning for the 2 cpd/16 Hz target which is the same as condition 8 (**Figure 12B**) by using a spatially narrowband mask. The mask includes a 1-octave wide band in spatial frequency (from to 2 cpd) and  $1^\circ$  orientation band which are presented at 0.045RMS contrast. Based on Control 1, we should be able to verify one crucial assumption that the broadband mask pattern drives these local temporally-tuned suppression as reflecting multiple local gain control pools whose output is combined and weighted to reflect the conventional temporal channel with maximal sensitivity at 8-12 Hz, whereas narrowband content which stimulate a few detectors is not adequate to drive these pools. Second, Control 2 measure tuning for the same target as Control 8 with the same mask as the Control 1 except the contrast whose contrast is 0.1RMS which is approximately two times higher than the Control 1. Based on Control 2, one should be able to check that the contrast information is also one of the

crucial factors for contributing the gain control pool. Finally, Control 3 measures the tuning for the same target as condition 10 (2 cpd/20 Hz target) with a comparatively narrowband (i.e., 2-octave wide in spatial frequency (ranging from 1 to 4 cpd) and 7° orientation band which are presented at 0.18 RMS contrast). Although the mask is spatially narrower than the mask used in the condition 10, its contrast is approximately 2 times higher contrast than the condition 10. **Table 3** summarizes the entire mask specifications used in Control conditions and the original conditions.

**Figure 14** shows the results of these Control conditions with the results obtained in original condition 8 and 10. Normalized suppressions (NSp) for comparing the magnitude, and normalized suppressions (NSb) for comparing the bandwidth in each condition for the two subjects are averaged and are shown in **Figure 14A** and **14B** respectively. Also, the same data (in **Figure 14A** and **14B**) are re-plotted on linear scale in **Figure 15A** and **15B**. Because the highly compressed scale at the high frequency side, the tuning of the masking functions plotted on the linear scale (**Figure 15A**) are mainly used to describe the results of the Control conditions. As can be seen in **Figure 15A**, when the temporal frequency of the target and mask matches, the single peak masking effect across each Control condition appears at 10 Hz, not each target's temporal frequency (16 Hz). When the same data in **Figure 14A** is re-plotted on the linear scale, the two single peak masking effects appear at 10 Hz and at each target's temporal frequency for both Control condition 2 and 3. However, in Control 1, the single peak masking effect only appears at 10 Hz, not the target temporal frequency (16 Hz). These Control conditions show that the broadband mask pattern including multiple spatial components and the contrast information are crucial keys to drive multiple temporally-



tuned pools. As can be seen in **Figure 14B**, the bandwidth estimation little varies across Control conditions and across different orientations within each Control condition (i.e., see the bandwidth parameters below the tuning curves in **Figure 14B**).

*How does the peak masking position systematically change with the target frequency?  
(the center frequency ( $x_0$ ))*

The single peak masking effect constantly appears at the same temporal frequency of the test for a variety of test temporal frequencies up to 14 Hz, and but not beyond 14 Hz (see **Figure 12A**). Beyond 14 Hz, the single peak masking effect is observed at 10 Hz, not the same temporal frequency of the test. **Figure 12A** shows the tuning of the broadband masking which is plotted on the log space. However, when each tuning of the masking function is plotted on the linear scale in **Figure 12B**, the two single peak masking effects are observed at 10 Hz and each target's temporal frequency (i.e., condition 8: target 2 cpd/16 Hz and condition 10: target 2 cpd/20 Hz). In other words, when the target temporal frequency is beyond 14 Hz, the two single masking effects are observed—specifically, the first peak at 10Hz and the second peak at each target's temporal frequency. The masking effect at 10 Hz is significantly greater than one at each target's temporal frequency. This might be due to the maximum sensitivity of the conventional temporal frequency channels which typically appears at 8-12 Hz (i.e., the temporally-tuned suppression as reflecting the multiple local gain control pools whose output is combined and weighted to reflect to the conventional temporal channel with maximum sensitivity at 8-12 Hz). In addition, when the temporal frequency of the test and mask matches, the masking is highly anisotropic showing the significant horizontal effect anisotropy. The

center frequency ( $x_0$ ) model parameter which was estimated by plotting the log scale is shown in **Figure 13**. Again, when plotting on a linear scale for targets up to 14 Hz, the multiple peak masking effects are shown when the temporal frequency of the target and mask matches. When plotting on a linear scale for targets beyond 14 Hz, two peak masking effects appear at 10 Hz and at each target's temporal frequency.

Based on these findings, it is reasonable to suggest that (1) the anisotropic (horizontal effect) broadband masking is indeed temporally tuned and (2) this broadband masking is indeed anisotropic (i.e., showing the significant horizontal effect).

#### *Untuned-suppression (vertical shift (C))*

We fit a Gaussian function on a log scale, centered on the temporal frequency of highest suppression and with the asymptote as a third parameter (vertical shift (C)). A non-zero asymptote suggests un-tuned suppression (i.e., like un-masked proportion in spatial frequency tuning, see in figure 12 in Dakin & Mareschal, 2000, and un-tuned proportion/or isotropic masking in orientation tuning, see in figure 3 in Cass, Stuit, Bex, & Alais, 2009). As can be seen in **Figure 13**, non-zero un-tuned suppression (vertical shift parameter (C)) across the different spatio-temporal surface is found. Specifically, the vertical shift parameter (C) value shown in **Figure 13** is the greatest for a target (i.e., a 12 cpd/10 Hz target in which average the parameter values across orientations is approximately 0.64), and is least for target gratings (i.e., 2 cpd/10 Hz and 2 cpd/14 Hz targets in which average the parameter value across orientations is 0.09). In the current Experiment 1, this un-tuned suppression (vertical shift parameter values (C) in **Figure 13**) in temporal frequency tuning is highly correlated with the baseline (unmasked)

contrast threshold values (i.e.,  $R^2 = 0.92$ ). From this high correlation, it is reasonable to suggest that this un-tuned suppression is significantly due to the baseline contrast threshold values.

*How do the horizontal and vertical biases change across the spatio-temporal conditions?*

To answer this subtle question, each normalized suppression index (HEI, VEI, and H-VEI in suppression) was calculated and was shown in **Figure 16**. With the mask (a right column in **Figure 16**), the HEI values are demonstrated in each target condition, where the HEI with normalized suppression values (NSp) can be seen vary between 0.12 and 0.70 indicating that NSp values at horizontal are 12-70% greater than NSp values at oblique. The VEI values are demonstrated in each target condition where the VEI with NSp values can be seen vary between -0.01 and 0.36 indicating that NSp values at vertical are -1-36% greater than NSp values at oblique. The H-VEI values are demonstrated in each target condition where the H-VEI with NSp values can be seen vary between 0.06 and 0.39 indicating that NSp values at horizontal are 6-39% greater than these values at vertical. In summary, the NSp values at horizontal across the target conditions are constantly greater than both vertical and oblique, and the NSp values at vertical across the conditions are constantly greater than both vertical and oblique orientations.

With the baseline sensitivity (a left column in **Figure 16**), the HEI with sensitivity values can be also vary between -0.03 and 0.96 indicating that sensitivity values at horizontal are -3-96% greater than sensitivity values at oblique. The VEI with sensitivity can be seen vary between -0.03 and 0.93 indicating that sensitivity values at vertical are -

3-93% greater than sensitivity values at oblique. The H-VEI can be also seen vary between 0 and 0.16 indicating that sensitivity values at horizontal are 0-16% greater than sensitivity values at vertical. The difference between horizontal and vertical in sensitivity seems to vary little, indicating almost no difference. In summary, the horizontal and vertical sensitivity are pretty similar (see H-VEI of a left column in **Figure 16**), and seem to be increasing with increasing spatial frequency.

Each normalized suppression index value (HEI, VEI, and H-VEI in NSp) seems to be related to each corresponding sensitivity index value (HEI, VEI, and H-VEI in sensitivity) in order to investigate how the proportional horizontal or vertical biases in temporally-tuned suppression (i.e., caused by broadband masks) in each target condition are correlated with the proportional horizontal and vertical biases in sensitivity (i.e., caused by single gratings) in each target condition. This could be a reasonable way to check whether the orientation-biased temporally-tuned suppression (stimulated by broadband masks) and sensitivity mechanism (stimulated by single grating) are the same mechanism. According to the masking paradigm, masking effect is the greatest with the test and mask stimuli of the same spatial frequency (or orientation), but not temporal frequency. In temporal masking, prior studies showed that the maximum masking effect is not constantly observed when the temporal frequency of the test and mask matches, but is constantly observed at one of the 2 (or 3) common temporal frequencies which are approximately at 1-2 Hz or 8-12 Hz (or very high temporal frequency which is termed as third channel, see Hess & Snowden, 1992). If this conventional notion of temporal processing (Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) is indeed true and the proportional horizontal or vertical

biases in temporally-tuned suppression (a right column in **Figure 16**) would be mainly due to the horizontal or vertical biases in sensitivity (a left column in **Figure 16**), we should be able to observe the strong correlation between them. However, the correlation between these two sets of values is not strong (i.e., NSp vs. sensitivity in HEI:  $R^2 = 0.64$ , NSp vs. sensitivity in VEI:  $R^2 = 0.12$ , NSp vs. sensitivity in H-VEI:  $R^2 = 0.24$ ).

Based on this weak correlation, it can be logically inferred that the multiple local temporally-tuned suppressions are revealed by the different mechanisms, not the only most sensitive conventional temporal mechanisms. Specifically, these multiple temporally-tuned suppressions reflect the multiple local (*'tuned'*) gain control pools whose output is combined and weighted to reflect the plausibly many and unknown temporal channels (i.e., see figure 8B in Alitto & Usrey, 2004) with maximum sensitivity around 8-12 Hz. From this point of view, the temporal tuning obtained by the naturalistic ( $1/f$ ) broadband masking function would be considered as an index which shows the temporally-tuned pools (i.e., driven by strong mutual suppression between detectors tuned similar temporal rates). From this point of view, the temporal tuning function obtained by the broadband masking indeed distinct from the prior temporal tuning function which was obtained by very narrowband mask pattern in the following. The narrowband mask pattern would be thought to be not adequate enough to rise above a certain threshold of the underlying many (unknown) temporal detectors whose sensitivity is lesser than the maximal sensitivity at around 8-12 Hz, but seems to be adequate to rise above a certain threshold of the most sensitive temporal detector at 8-12 Hz. Thus, the peak masking effect demonstrated in prior works is constantly observed at around the peak sensitivity 8-12 Hz (i.e., see Control condition 1 in **Figure 16**) and prior masking

works (Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alais, 2006; Cass et al., 2009).

Taken together, the narrowband mask is not adequate to stimulate many (unknown) temporal detectors whose sensitivity is lesser than the maximal sensitivity at around 8-12 Hz, but is adequate to stimulate a few (2 or 3) most sensitive conventional temporal detectors (i.e., further details with respect to this view will be described in General Discussion). In order to drive the many, unknown and lesser sensitive temporal detectors, the use of the particular mask pattern which can rise above a certain detection threshold of these lesser sensitive temporal detectors, and thereby create strong mutual suppression between temporal detectors tuned in similar temporal frequencies which forms a local (*'tuned'*) pool. We assumed that there are amount of strong suppressive interactions between the mask and test temporal frequency-tuned detectors. This suppressive interaction (or weight) is the most when the temporal frequency of the mask and target matches (see **Figure 12A and 12B**), whereas the suppressive interaction is much less when the temporal frequency of the mask and test are different. Within this line of thinking, current work assumes that this idea captures a temporal channel interaction, in that suppressive weighting of temporal frequencies neighboring the test is stronger than that of temporal detectors with far away temporal frequencies.

#### ii. Bandwidth ( $W$ ) with NSb values

It is customary to fit Gaussian functions, usually on a log scale (i.e., a log spatial frequency: Baker, Thompson, Smyth, & Tholurst, 1998; Dakin & Mareschal, 2000; Henriksson, Nurminen, Hyvärinen, & Vanni, 2008; Holub & Morton-Gibson, 1981;

Movshon & Tolhurst, 1978c; Tolhurst & Thompson, 1981; a log temporal frequency: Robson, 1966; Anderson, Burr, & Morrone, 1991; Hess & Snowden, 1992; Waung & Hess, 1995; Cass & Alias, 2006; Cass, Alais, Spehar & Bex, 2009). We also found a log Gaussian functions to fit quite well (i.e., the Sum of Squared Errors or SSE average across 10 conditions = 0.043).

As can be seen in **Figure 17**, bandwidths vary little across the 10 spatio-temporal conditions (i.e., the main effect of the bandwidth is not significant:  $F(2,27) = 0.026, p = 0.974$ ) and the bandwidth is not significantly different across three orientations ( $t_{(27)} = 0.112, p = 0.912$ :  $90^\circ$  vs  $45^\circ$ ,  $t_{(27)} = -0.117, p = 0.908$ :  $0^\circ$  vs  $45^\circ$ ,  $t_{(27)} = -0.229, p = 0.821$ :  $90^\circ$  vs  $0^\circ$ ). **Figure 18** shows the bandwidth estimations and the average of bandwidth across conditions is 1.52 octaves. Indeed all data in different conditions are fit fairly well (SSE = 0.043) by the single Gaussian function plotted in the graph in **Figure 17**. Model parameters with normalized suppression (NSb) are provided in **Figure 19**.

#### d. Summary

Suppression of each grating target across the 10 spatio-temporal combinations by a naturalistic ( $1/f$ ) broadband mask is shown to be maximal when the temporal frequency of the broadband mask and target matches (see **Figure 12A, 12B** and **15A**). In these matched conditions across the 10 spatio-temporal combinations, the naturalistic ( $1/f$ ) broadband suppression is highly anisotropic, showing the horizontal effect. We assume that a narrowband mask used in prior works would not be adequate enough to stimulate many and unknown temporal detectors whose sensitivity is lesser than the maximal sensitivity at around 8-12 Hz, but is only adequate enough to stimulate a few (2 or 3)

most sensitive conventional temporal detectors at around 1 Hz and 8-12 Hz. We propose that to drive these many, unknown, and less sensitive temporal detectors, the use of an particular mask pattern, naturalistic ( $1/f$ ) broadband which would be able to rise above a certain detection threshold of each of the underlying these many temporal detectors, and thereby would be able to create strong mutual suppression between detectors tuned in similar temporal frequencies which forms a local (*'tuned'*) temporal pool. There exist multiple temporally-tuned (*'local'*) pools across the 10 spatio-temporal locations, and each of the pools is highly anisotropic, showing the horizontal effect. In addition, the bandwidth of each pool (see **Figure 17** and **18**) varies little across the 10 spatio-temporal conditions and across the different orientations within each target condition which are around 1.52 octaves.

### 3.2. Experiment 2

#### a. Thresholds

Contrast thresholds in each condition for the two subjects are averaged are shown without mask (baseline) in **Figure 20A** and with mask in **Figure 20B**. As can be seen in **Figure 20A**, the baseline (no mask) conditions show an oblique effect of contrast thresholds at the higher spatial frequencies (i.e., 8 cpd and 12 cpd target, highest thresholds are at oblique orientations). The oblique effect anisotropy seems to be disappeared at the lower spatial frequencies as expected (i.e., 1, 2, and 4 cpd). **Figure 20A** shows the masking effect of each target across the 9 spatio-temporal combinations as a function of center spatial frequency at each of the five (or six) 0.8-octave sub-bands and a broadband mask which are modulated by the temporal noise (i.e., 4.5-octave in temporally). The masked



contrast thresholds are seen to be greatest when the spatial frequency of the test and mask matches. Broadband masking effects (bb in **Figure 20A**) seems to be larger than this matched condition across the 9 spatio-temporal target conditions. To show the effect of the masks more directly, suppression factor was calculated by dividing the masked threshold by the unmasked threshold for each of the 9 spatio-temporal test conditions in **Figure 21**. The suppression is the greatest when the spatial frequency of the test and mask matches. **Figure 22** shows the difference between the suppression obtained by a broadband mask (bb) and the suppression obtained by the 0.8-octave narrowband (nb) mask. The broadband suppression seems to be larger than suppression obtained by the 0.8-octave narrowband conditions in which the spatial frequency of the test and mask matched except two target conditions (i.e., 8 cpd/16Hz and 12 cpd/ 10Hz).

#### b. Normalized suppression: Spatial frequency tuning

##### i. Amplitude ( $A$ ) with NSp values

In order to directly compare the magnitude of the horizontal effect across the 9 spatio-temporal target conditions, the tuning of masking with the normalized suppression values (NSp) across the spatial frequency was assessed by fitting curves to the masking functions. A Gaussian function with four parameters provided a best fit to full normalized suppression data (i.e., SSE averaging across the 9 spatio-temporal conditions = 0.026). As can be seen in **Figure 23**, the tuning of masking with the normalized suppression (NSp) plotting on the log scale is constantly larger when the spatial frequency of the test and mask matches across the all of the 9 spatio-temporal conditions. In these spatial frequency matched conditions, the masking effects is highly anisotropic (i.e., the main

effect of the orientation:  $F(2, 24) = 56.733, p = 0.00$ ) and the significant horizontal effect pattern are shown in the following. The masking for horizontal targets is significantly greater than for oblique targets ( $t_{(24)} = 10.63, p = 0.00 : 0^\circ$  vs  $45^\circ$ ) and than for vertical targets ( $t_{(24)} = 5.95, p = 0.00 : 0^\circ$  vs  $90^\circ$ ).

As can be seen within each of the 9 spatio-temporal conditions in **Figure 23**, the masking effect is smooth showing spatial tuning and peaks at the test frequency. This strongly indicates that the suppression between channels tuned to different spatial frequency is indeed local (*'tuned'*) (i.e., the suppressive interaction seems to be most from the neighboring spatial channels tuned to the similar target spatial frequency which would be approximately 0.8-octave width spatially). Based on these results, it could be suggested that (1) the narrowband masking is spatially tuned (i.e., the peak masking effect always appears at the same spatial frequency of the test and mask stimuli) and (2) in these conditions, the significant anisotropic masking, anisotropic suppression (the horizontal effect), is observed (i.e., when the spatial frequency of the test and mask matches, the significant horizontal effect are constantly observed at all of the conditions). This result clearly shows that the anisotropic (the horizontal effect) suppression is tuned in the spatial frequency dimension.

Furthermore, when considering the masking effect obtained by the 4-octave spatially broadband patterns modulated by a 4.5-octave temporal noise with the normalized suppression (NSp), the similar significant horizontal effect anisotropy is also obtained (the main effect of orientation:  $F(2, 24) = 29.489, p = 0.00$ , and  $t_{(24)} = 7.52, p = 0.00 : 0^\circ$  vs  $45^\circ$  and  $t_{(24)} = 5.11, p = 0.00 : 0^\circ$  vs  $90^\circ$ ). Interestingly, the masking effect for the broadband (4-octaves) pattern with the normalized suppression (NSp) is significantly

greater than masking for the narrowband (0.8-octaves) matched condition in which the spatial frequency of the test and mask matched (i.e., the peak masking in **Figure 23**) (i.e.,  $t_{(26)} = -2.509, p = 0.019$ ) at all target conditions. This indicates that the spatially-tuned suppression as reflecting the multiple local ('*tuned*') gain control pools and the output of each of these pools is weighted and combined by response activities in neighboring spatio-temporal detectors. Thereby, these pools are only pronounced with spatially and temporally broadband mask pattern. This broadband pattern contains a wide range of spatial frequencies, orientations, and temporal frequencies which can stimulate many spatio-temporal detectors and thereby the strong suppressive interaction would be created.

These results are indeed distinctive from the prior work (i.e., Haun & Essock, 2010) which demonstrated that the anisotropic masking (the horizontal effect) is largest at the spatial frequency target band around 8 cpd. With respect to the low spatial frequency horizontal effect, Experiment 2 showed the horizontal effect anisotropy for 1 cpd target flickered at either 1 or 16 Hz by spatially narrow(0.8-octave) and temporally broad (4.5-octave) mask noise. In this view, previous studies (i.e., Campbell & Robson, 1968; Stromeyer et al., 1982) showed that the modulating low spatial frequency targets in time, masking effect to the low spatial frequency targets can cause maximal masking effect at the same low spatial frequency, but the masking to the static low spatial frequency target can cause maximal masking effect at a significantly higher spatial frequency of the mask, not the same spatial frequency. Based on these works, it can be logically inferred that in order to strongly stimulate the low spatial frequency-tuned channels, the used of an appropriate mask patterns (i.e., temporally modulated) seems to be crucial. This would be a main reason why the low spatial frequency horizontal effect was not observed in the

prior work (Haun & Essock, 2010). The static mask would not be strong enough to stimulate the low spatial frequency orientation-tuned channels, and in this reason the strong suppressive interactions between these low spatial frequency-tuned channels would not be created, and thereby the anisotropic suppression (the horizontal effect) would not be observed in prior work (Haun & Essock, 2010). However, current Experiment 2 measured the spatially-tuned suppression of each target by particular mask stimuli, such as five (or six) 0.8-octave bands and each of them is modulated by a 4.5-octave temporal noise, and a 4-octave band modulated by a 4.5-octave temporal noise. This indicates that the spatially-tuned suppression (i.e., the multiple peak masking effects shown in **Figure 23**) could be considered as multiple local (*'tuned'*) pools and the output of each of these pools is combined and weighted by neighboring detector tuned in similar spatial and temporal properties. Therefore, it is reasonable to suggest that in order to stimulate the local (*'tuned'*) low spatial frequency-tuned anisotropic suppression (or anisotropic gain control pool), the use of the temporally modulated mask pattern is fundamental. Each model parameters are demonstrated in **Figure 24**.

Taken together, in order to drive multiple spatially-tuned pools across spatial frequencies, the use of particular mask pattern which can rise above a certain detection threshold of each of the low spatial frequency- or high spatial frequency-tuned pools is crucial. In the current Experiment 2, the use of the temporally broadband mask pattern is thought to rise above a certain detection threshold of the many low spatial frequency-tuned detectors, thus the low spatial frequency-tuned (*'local'*) anisotropic pool (showing the horizontal effect) is observed. However, the static mask pattern used in prior work (Haun & Essock, 2010) is thought to be not adequate enough to rise above a certain

threshold of the many low spatial frequency tuned detectors, and thereby the low spatial frequency tuned anisotropic pool (horizontal effect) is not observed. Therefore, in order to drive the multiple local (*'tuned'*) gain control pools across spatial frequencies, the use of an appropriate mask pattern which can be able to rise above the certain threshold of each of these many detectors and thereby can create strong anisotropic suppression (the horizontal effect) between detectors tuned in similar spatial frequencies is essential.

*How does the peak masking position systematically change with the target spatial frequency? (center position ( $x_0$ ))*

As can be seen in **Figure 25**, the center position of the model parameter ( $x_0$ ) indicating the peak masking spatial frequency in the tuning of the masking function is indeed matched with the peak masking spatial frequency demonstrated in **Figure 23**. This indicates that the anisotropic suppression (the horizontal effect) is indeed tuned (*'local'*) in spatial frequency dimension.

*Untuned-suppression (vertical shift ( $C$ ))*

As can be seen in **Figure 24**, a non-zero asymptote suggests un-tuned-suppression (i.e., like un-masked proportion in spatial frequency tuning, see in figure 12 n Dakin & Mareschal, 2000). This non-zero un-tuned suppression (a vertical shift parameter ( $C$ ) in **Figure 24**) which varies across the different spatio-temporal surface is found. Specifically, this parameter values is the greatest for a target (i.e., 12 cpd/10 Hz in which average parameter values across orientation is 0.58) and is least for a target (i.e., 4 cpd/1 Hz in which average parameter values is 0.12). In the current Experiment 2, this un-tuned

suppression (vertical parameter (C) in Figure 24) in spatial frequency tuning is highly correlated with the baseline (unmasked) contrast threshold values (i.e.,  $R^2 = 0.92$ ). This high correlation would indicate that this un-tuned suppression is highly due to the baseline contrast threshold.

*How do the horizontal and vertical biases change across the spatio-temporal locations?*

With the mask (a right column in **Figure 25**), the HEI values are demonstrated in each target condition, where the HEI with normalized suppression values (i.e., NSp: each suppression values was divided by the single maximum suppression values for each target condition irrespective of orientation) can be seen vary between 0.31 and 0.87 indicating that NSp values at horizontal are 31-87% greater than NSp values at obliques. The VEI values are demonstrated in each target condition where the VEI with NSp values can be seen vary between 0.10 and 0.56 indicating that NSp values at vertical are 10-56% greater than NSp values at obliques. The H-VEI are demonstrated in each target condition where the H-VEI with NSp values can be seen vary between 0.10 and 0.58 indicating that NSp values at horizontal are 10-57% greater than these values at vertical. In summary, the NSp values at horizontal across the target conditions are constantly greater than both verticals and obliques, and the NSp values at vertical across the conditions are also constantly greater than the NSp values at obliques. Interestingly, a maximum NSp value across the 9 target spatio-temporal conditions is shown at an 8 cpd/ 1Hz target condition (i.e., HEI = 0.87 and VEI = 0.56). Similarly, the strongest horizontal bias at an 8 cpd target band (1-octave spatial frequency band and 45° orientation band) was also reported (see figure 6 in Haun & Essock, 2010).

With the baseline sensitivity (a left column in **Figure 25**), the HEI with sensitivity values can be seen vary between -0.01 and 1.20 indicating that sensitivity values at horizontal are -1-120% greater than sensitivity values at obliques (i.e., at a 2 cpd/16 Hz target condition, HEI sensitivity values are -0.01 indicating the oblique sensitivity is a little better than horizontal, but the value is pretty small). The VEI with sensitivity values can be seen vary between -0.04 and 1.02 indicating that sensitivity values at vertical are -4%-102% greater than sensitivity values at obliques (i.e., at a 2 cpd/10 Hz condition, VEI sensitivity values are -0.04 indicating the oblique sensitivity is a little better than vertical, but this values is still pretty small). The H-VEI can be also seen vary between -0.02 and 0.18 indicating that sensitivity values at horizontal are -2%-18% greater than sensitivity values at verticals, but these values seem to be so small. The horizontal and vertical sensitivity seem to be similar. Interestingly, a maximum HEI and VEI value across the 9 target spatio-temporal conditions is shown at an 8 cpd/1 Hz target condition (i.e., HEI = 1.20 and VEI = 1.02). Taken together, the sensitivity values at horizontal and vertical are pretty similar and seem to be increasing with increasing the spatial frequency of the target.

Each normalized suppression index is used to relate each corresponding sensitivity index value demonstrated in **Figure 25** in order to consider whether the orientation-biased spatially-tuned suppression (driven by temporally broadband mask) and sensitivity mechanisms (stimulated by single grating at different spatio-temporal locations) are the same mechanism. The correlation between these two set of values is not strong (i.e., NSp vs. sensitivity in HEI:  $R^2 = 0.16$ , NSp vs. sensitivity in VEI:  $R^2 = 0.34$ , NSp vs. sensitivity in H-VEI:  $R^2 = 0.08$ ). Based on this weak correlation, it can be logically inferred that the multiple local spatially-tuned suppressions (i.e., the peak

making positions in **Figure 23**) are driven by the different mechanisms, not the same sensitivity mechanism which is not weighted and combined responses by neighboring detectors tuned in similar spatial frequencies (i.e., unaltered response). We assume that the multiple spatially-tuned suppressions reflect the multiple local (*'tuned'*) gain control pools and each of the multiple local pools is combined and weighted by responses coming from neighboring spatial detectors (tuned in similar spatial frequencies and orientations).

#### ii. Bandwidth ( $W$ ) with NSb values

As can be seen in **Figure 26**, bandwidth plotting on a log scale varies little across the 9 spatio-temporal conditions (i.e., the main effect of the bandwidth is not significant:  $F(2,24) = 0.02, p = 0.98$ ) and the bandwidth across three orientations within each of the 9 spatio-temporal target conditions is not significantly different ( $t_{(24)} = -0.163, p = 0.87$ :  $90^\circ$  vs  $45^\circ, t_{(24)} = 0.022, p = 0.983$ ;  $0^\circ$  vs  $45^\circ, t_{(24)} = -0.185, p = 0.855$ ;  $90^\circ$  vs  $0^\circ$ ). **Figure 27** shows the bandwidth estimations across the 9 spatio-temporal target conditions. The average of the bandwidth is 1.35 octaves and all data are fit fairly well (i.e., SSE averaging across the 9 spatio-temporal conditions = 0.043) by using the single Gaussian function. Each model parameters are shown in **Figure 28**.

Physiologically, general nature of spatial tuning of cells in the LGN are generally very broadly tuned, showing sharp high frequency attenuation but only a very gentle drop in sensitivity to low spatial frequencies (Campbell et al., 1969). By contrast, the majority of cortical cells are considerably more narrowly tuned, showing sharp low spatial frequency as well as high spatial frequency attenuation. Most cortical cells (i.e., simple



and complex cells) thus have a distinct bandpass characteristic (see for example the cells shown in figure in DeValois, Albrecht, & Thorell, 1982; see figure 3 in Maffei & Fiorentini, 1973). It was reported that the range of bandwidth of cortical cell population is very large. Most of the cortical cells have bandwidth between 1.0 and 1.5 octaves, but there are a number of very narrowly tuned cortical cells with bandwidths of less than an octave and a sizable population of broadly tuned cells with bandwidths as large as 2.0 octaves. The median bandwidth for simple cells (both foveal and parafoveal) is about 1.4 octaves which is very close to that predicted from the Blackmore and Campbell (1969). Specifically, prior estimations of channel bandwidth from various types of psychophysical investigations have varied from less than 0.5 octaves (i.e., Sachs, Nachmias, & Robson, 1971) through 1.2 octaves (i.e., Blakemore & Campbell, 1969) and 1.5 octaves (Wilson, McFarlane, & Phillips, 1983) to 2.0 octaves (i.e., Wilson, 1978). As can be seen in **Figure 27**, the average of the bandwidth for the suppressive pool across each target location seems to be equally spaced on an octave scale is 1.35 which are not distinctively different from the prior bandwidth estimations.

### c. Summary

The suppression of each target by a temporally broadband mask is spatially tuned (i.e., the peak masking effect always appears at the same spatial frequency as the test), and this suppression is generally anisotropic (showing the horizontal effect) which is observed at all of the 9 spatio-temporal combinations (i.e., the multiple anisotropic gain control pools showing the horizontal effect indeed exist across the 9 spatial and temporal combinations). The key is that to drive multiple spatially-tuned suppression as reflecting

multiple local (*tuned*) gain control pools across the different spatio-temporal combinations, the use of an appropriate mask pattern which can drive strong mutual suppression between neighboring detectors is crucial. In addition, the bandwidth of each pool (see **Figure 26** and **27**) varies little across the different orientations within each target condition which are around 1.35 octaves.

#### 4.0 GENERAL DISCUSSION

The prior masking studies by using very narrowband masks have shown 2 (or 3) temporal detectors by measuring the temporal tuning of grating targets at different spatio-temporal combinations with and without a very narrowband (i.e., grating) mask across various temporal rates of the mask. Then, the ratio of the mask threshold to the unmask threshold was plotted as a function of the mask temporal frequency which is referred to as the temporal frequency tuning function. This tuning of the masking functions show how the narrowband masking effect of each target at the different spatio-temporal combinations systematically changes as a function of the temporal frequency of the very narrowband mask. When considering the peak masking frequency at which the masking is most effective (i.e., indicating the most sensitive temporal frequency in the masking function) seems to depend upon the spatial frequency, not the temporal frequency of the target. Specifically, the peak masking effect constantly appears at one of the 2 (or 3) common temporal frequencies which are approximately at 1-2 Hz, 8-12 Hz, or a high temporal frequency (i.e., this is referred to a third channel, see figure 4-7 and 12 in Hess & Snowden, 1992), not the same temporal frequency of the target and mask. These 2 (or 3) well-known peak masking temporal frequencies in the masking function have been considered as the most sensitive temporal detector's temporal frequencies underlying the human temporal sensitivity function. Again, one should make sure that the prior masking works shown 2 (or 3) temporal detectors (or mechanisms, channels and filter, etc)

have been developed by using very narrowband mask pattern (i.e., a single grating which has been mainly used to measure the human temporal sensitivity function).

Interestingly, when considering the human temporal frequency sensitivity function (Robson, 1966) one should be able to see that these 2 (or 3) peak masking positions (Lehky, 1985) which well match the peak sensitivity positions (Robson, 1966). This provides a strong insight that the very narrowband mask stimulus used in prior mask works (Lehky, 1985; Cass & Alias, 2006; Cass et al., 2009) is only adequate enough to rise above a certain detection threshold of the most sensitive temporal frequency tuned detectors (i.e., conventional temporal detectors), but is not adequate enough to rise above a certain threshold of each of the underlying many and lesser sensitive temporal detectors than the conventional temporal ones. Thereby, the peak masking effect constantly appears at one of the 2 (or 3) common temporal frequencies indicating the well-documented 2 (or 3) temporal frequency tuned detectors in the following. Briefly, one is transient temporal detector (showing the bandpass temporal tuning characteristics) with maximum sensitivity around 8-12 Hz which typically prefers the low spatial frequencies. Another is the sustained temporal detector (showing the lowpass temporal tuning characteristic) with maximum sensitivity around 1-2Hz which prefers the high spatial frequencies. The other is very high temporal frequency tuned detectors which has been referred to 'third detector (or channel)' (i.e., around 20 Hz, see figure 12 in Hess & Snowden, 1992) which prefers the very low spatial frequencies. Based on these, it is reasonable to infer that the narrowband mask pattern most likely reduce the sensitivity of the 'most' sensitive temporal detectors, not many (lesser sensitive) temporal detectors when measuring the tuning of each target by a very narrowband (e.g., grating) mask across varied temporal

rates of the mask.

Therefore, it is reasonable to suggest that the use of the narrowband mask is not adequate to probe many and unknown temporal detectors whose sensitivity is lesser than the peak sensitivity at 8-12 Hz, and thereby relatively uncombined and un-weighted response activities that form the conventional temporal detector are obtained. Based on this line of thinking, the prior works using a very narrowband mask (Lehky, 1985; Cass & Alias, 2006; Cass et al., 2009) have shown the tuning of each grating target and this tuning shape in which the peak masking effect constantly appears at one of the 2 (or 3) common temporal frequencies. Thereby, this tuning of the narrowband masking function could be considered as an index which shows the tuning shape of the most sensitive 2 (or 3) temporal detectors. However, the peak masking effect did not always appear at one of the 2 (or 3) common temporal frequencies from the tuning functions shown in two prior works (i.e., Hess & Snowden, 1992; Fredericksen & Hess, 1998) which used a mask pattern that included a range of spatial frequencies. When considering the data of the several prior masking studies (see figure 5D in Hess & Snowden, 1992, figure 5 in Fredericksen & Hess, 1998), the two peak masking effects indeed appear at 10 Hz and at each target's temporal frequency whose frequency is higher than 10 Hz. This provides a plausible insight that the use of a particular mask pattern might drive the temporally-tuned suppression (i.e., the peak masking effect appears at the target temporal frequency when using the appropriate mask pattern). As can be seen in **Table 1**, these two prior works (i.e., Hess & Snowden, 1992; Fredericksen & Hess, 1998) used the particular mask patterns which included a range of the spatial frequencies. From this view, current Experiment 1 showed these multiple temporally-tuned (*'local'*) pools by using the

naturalistic ( $1/f$ ) broadband mask pattern at superthreshold contrast (Note that both spatial components and contrast seem to be crucial factors to drive these multiple temporally-tuned suppressions as reflecting local (*'tuned'*) gain control pools as reported above in control conditions, see **Figure 15A**). Specifically, each of these pools indicates local (*'tuned'*) suppression which can be achieved by strong mutual suppression between detectors tuned in similar temporal rates. We assume that the naturalistic broadband pattern is adequate enough to rise above a certain detection threshold of each of the underlying many, unknown, less sensitive temporal detectors. Thereby, strong local (*'temporally-tuned'*) suppression which comes from the pooling response to contrast at temporal frequencies similar to the target was observed in the current Experiment 1. We assume that this local suppression can be achieved by mutual suppression between neighboring temporally tuned detectors which form the local (*'temporally-tuned'*) pools (see **Figure 12A** and **Figure 12B**).

Interestingly, one plausible hypothesis which assumed the existence of the multiple temporal detectors underlying the human temporal sensitivity function was also addressed by Ascher and Grzywacz (2000). Specifically, the typical models of detection are based on the assumption that detection (masking) occurs when the activity in the detectors most sensitive to the stimulus rises above a certain threshold. If (1) this assumption and (2) the general assumption addressing a limited number of the temporal detectors (i.e., 2 or 3 detectors) underlying the human temporal processing are indeed true, then we would expect to see 'discrete jumps' in the temporal tuning of the masking function as the function of the mask is shifted from being close to the peak temporal frequency of one detector (most sensitive) to the peak temporal frequency of another

detector (least sensitive). However, when considering the masking data of the prior masking works (i.e., Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) and the current Experiment 1 (i.e., **Figure 12A**), one does not observe such discrete jumps, but smooth transitions instead. Based on this, it could be reasonably assumed: (1) many temporal detectors would exist underlying the human vision at the early vision (V1) and each of them are tuned in varied temporal rates, and (2) each of the many temporal detectors might have a different detection criterion. We propose that to drive each of these many temporal detectors the use of the particular mask stimulus which can rise above a certain detection threshold of each of the many and unknown temporal detectors seems to be crucial. The narrowband mask patterns used in prior masking might not be adequate enough to drive these many (unknown and less sensitive) temporal detectors, but is only adequate enough to drive the most sensitive (a few) temporal detectors (i.e., 2 (or 3) conventional temporal detectors).

This idea can be also supported by considering the current Control conditions shown in **Figure 15**. When considering the current work's Condition 1 (see top in **Figure 15**), the bandpass tuning shape with the peak masking effect at 10 Hz can be observed in **Figure 15A** and **15B**. Again, we didn't see such discrete jumps, but smooth transition instead. We assume that this smooth shift in peak masking would reflect shifts between the peaks of many temporal detectors which is lesser sensitive than the temporal detector at around 10 Hz. This result is similar with the prior masking results. Why does the peak masking position appear at 10 Hz, not target temporal frequency (16 Hz)? To answer this question, we need to consider the mask specification used in Control 1. The mask used in Control 1 was comparatively narrower spatially and low in contrast (i.e., one-octave

spatially ranging from 1 to 2 cpd and single orientation band with 0.045RMS contrast, see **Table 3**) compared to the mask used in the original condition 8 (i.e., 4-octave spatially at 0.1RMS contrast). We assume that this mask used in Control 1 is not adequate to rise above a certain detection threshold of many temporal detectors whose sensitivity is lesser than the temporal detector at 10 Hz. Thus, the peak masking effect appears at a most sensitive common temporal frequency (i.e., around 10 Hz), not 16 Hz.

Next, with the same target used in Control 1 we measured the tuning of the masking function by the same mask as Control 1, but higher in contrast (i.e., 0.1RMS contrast). The results (see middle row in **Figure 15A**) is very interesting in that the two single peak masking effects appear at 10 Hz and at 16 Hz which is the target's temporal frequency. However, the second peak masking effect at 16 Hz in the tuning function is disappeared when the same data is re-plotted on a log scale in **Figure 15B**. Because of the property of the log scales (i.e., highly compressed scale at the high frequency side), the second peak masking effect at 16 Hz would be disappeared. Several prior masking studies (see figure 5D in Hess & Snowden, 1992, figure 5 in Fredericksen & Hess, 1998) have also demonstrated similar results. (Note that the tuning functions were plotted on a log scale. If these tuning are re-plotted on a linear scale, we would be able to see the two clear peak masking effects.) Based on this result of Control 2, it could be true that the contrast information is also crucial factor to drive these less sensitive temporal detectors (see also Control 3). **Figure 15A** show how the peak masking effect systematically changes with the spatial broadness of the mask and also contrast. These two factors would be considered as the crucial information to drive these less sensitive temporal detectors.



Based on these, it can be logically inferred that the tuning of the mask function for a single grating target by very narrowband mask stimuli across varied temporal frequency of the mask should be considered as an index which shows the tuning shape of the conventional temporal channel which is relatively un-combined and un-weighted (unaltered) suppression that form the conventional temporal detector. Whereas the tuning of the broadband masking function shown in **Figure 12A** can be considered an index which shows the temporally-tuned suppression as reflecting multiple local (*'tuned'*) gain control pools whose output is combined and weighted (altered) by neighboring temporal detectors (temporal detectors tuned to similar temporal rates) . This suggested that the naturalistic broadband mask pattern used in current Experiment 1 would be a good stimulus which can rise above a certain detection thresholds of each of many and less sensitive temporal detectors underlying the human temporal sensitivity function. Thereby temporally-tuned suppression which reflects local (*'tuned'*) gain control pools whose output is combined and weighted by neighboring temporal detectors is only observed when the particular, naturalistic mask pattern are used. It is reasonable to assume that there are possibly many temporal detectors and each temporal detector's detection threshold would be different and to drive the peak sensitivity of each of them the use of appropriate mask which can rise above a certain detection threshold of each of many temporal detectors is fundamental. Therefore, it is reasonable to suggest that the prior notion addressing human temporal processing (i.e., 2 (or 3) temporally-tuned channels) would not be the whole story.

a. Why do the narrowband mask stimuli lead 2 (or 3) common peak masking effects?

One plausible answer would be obtained by recent several neurophysiological studies. According to these neurophysiological studies (Shoham et al., 1997; DeAngelis et al., 1999; Sun et al., 2007), the narrowband mask stimuli are only adequate enough to stimulate a limited number of the most sensitive (a few) temporal detectors (i.e., sustained (lowpass) vs transient (bandpass) which are the well-known conventional temporal detectors). Based on this, it is reasonable to infer that the response signals caused by narrowband mask would not be strong enough to rise above a certain detection threshold of each of many and unknown temporal detectors which form the temporally tuned-suppression (or pool), but is adequate to rise a few most sensitive (conventional) temporal detectors. Thereby, strong mutual suppression was not pronounced, and the peak making positions of the temporal masking functions constantly appear at the most sensitive temporal detector's frequencies which are either 1-2 Hz or 8-12 Hz (or a very high temporal frequency detector, termed as a third channel, Hess & Snowden, 1992). These temporal frequency tuned detectors would be considered as the well-known lowpass (sustained) and bandpass (transient) temporal detectors.

From this neurophysiological insight, it is reasonable to suggest in the following. When the narrowband mask stimuli are used to probe the temporal frequency tuned detectors, the peak masking position is always toward to one of the 2 (or 3) temporal frequency categories in the following (i.e., the narrowband mask seems to only stimulate a few most sensitive temporal detectors underlying the human temporal sensitivity function which match the peak sensitivity of each of the underlying 2 (or 3) temporal detectors). One is the bandpass detector (transient) with a peak sensitivity around 8-12 Hz and the other is the lowpass mechanism detector (sustained) with peak sensitivity around

1-2 Hz). Thereby, it is reasonable to suggest that a general notion of the human temporal processing employed by the narrowband mask stimuli would not be a whole story.

b. Does the naturalistic broadband mask create realistic response activities in detectors?

According to a recent neurophysiological finding (Baker & Issa, 2005), more realistic and complex motion stimuli generate more realistic cortical response properties of separable tuning properties of orientation, spatial frequency and temporal frequency. This finding is thought to be a plausible when considering the tangential organization of the primary visual cortex in which the visual information about the orientation, spatial frequency, temporal frequency and drift direction of image components is thought to be encoded by overlapping group of encoding-mechanisms within each hypercolumn (i.e., Bonhoeffer & Grinvald, 1991; Hubel & Wiesel, 1977; Hubener et al., 1997; Issa et al., 2000; Shmuel & Grinvald, 1996; Weliky et al., 1996). The realistic and complex dynamic visual stimuli would evoke the realistic response activities from numerous encoding-mechanisms at the early cortical stage which lead the separable response properties of orientation, spatial frequency, and temporal frequency dimensions.

Based on these findings, it is reasonable to suggest that the naturalistic and complex stimuli would generate realistic response activities from many encoding-mechanisms underlying the human vision at the early cortical stage (V1) which lead the separable tuning response properties of the spatial (including spatial frequency and orientation) and temporal dimensions. As can be seen in **Figure 12**, when naturalistic ( $1/f$ ) broadband pattern is used to stimulate each detecting-mechanism across the different spatio-temporal locations, the realistic multiple temporally-tuned suppressions as

reflecting multiple local gain control pools whose output is combined and weighted by neighboring temporally-tuned detectors are observed. We assume that there would be many temporally-tuned detectors (see figure 8B in Alitto & Usrey, 2004). In order to drive these many temporally-tuned detector, the use of a naturalistic mask pattern which can rise above a certain detection threshold of each of many and unknown temporal detector seems to be crucial.

c. What does the temporally-tuned suppression indicate? Do the suppression and sensitivity mechanisms differ?

The temporally-tuned suppression shown in **Figure 12** reflects multiple local gain control pools whose output is combined and pooled by neighboring temporal detectors (i.e., detectors tuned in similar temporal frequency). This nonlinear pooling process can be achieved by mutual suppression between detectors tuned in similar spatial and temporal properties, which is indeed distinct from the temporal tuning of the masking function demonstrated in prior work. This prior temporal tuning of the masking function is mainly obtained by very narrowband mask stimuli which would not be adequate enough to rise above a certain detection threshold of each of the underlying many and unknown temporal detectors, but would be adequate to rise above a certain threshold of the most sensitive (conventional) temporal detectors (i.e., 2 or 3) which is relatively unaltered output (i.e., un-combined and un-weighted output by neighboring temporal detectors). As described above in Results section, the suppression measured with the naturalistic ( $1/f$ ) broadband mask pattern seem to differ from the sensitivity measured without mask pattern when considering the correlation between these two sets of the values in **Figure**

16. Specifically, the correlation is pretty weak (i.e., NSp vs. sensitivity in HEI:  $R^2 = 0.64$ , NSp vs. sensitivity in VEI:  $R^2 = 0.12$ , NSp vs. sensitivity in H-VEI:  $R^2 = 0.24$ ). This weak correlation between them indicates that the suppression and sensitivity mechanisms would be driven by different mechanisms.

d. Do many and less sensitive temporal detectors have different detection thresholds?

Based on the results in Experiment 1, it is reasonable to suggest that there are possibly many temporally-tuned detectors as suggested in spatial frequency dimension.

Informatively, several detection models are based on the assumption that detection occurs when the activity in the detectors most sensitive to the stimulus rise above a certain threshold. If each of the underlying many temporal detectors has different detection threshold criterion, we should use different mask stimuli which can drive each of many and less sensitive temporal detectors. Thus, to drive these multiple lesser sensitive temporal detectors, the use the particular stimuli which can rise above a certain threshold of these many and lesser sensitive temporal detectors seem to be crucial. As demonstrated in Experiment 1, the naturalistic ( $1/f$ ) broadband masking pattern would be a good stimuli which can rise above a certain detection threshold of each of many temporally-tuned detectors, and thereby the temporally-tuned (*'local'*) suppressions which can be formed by strong mutual suppression between similar detectors. That would be a good reason why the prior masking works used in very narrowband masks are failed to show these multiple temporally-tuned suppressions. These narrowband mask pattern is not adequate enough to create strong mutual suppression between similar detectors (i.e., tuned in similar temporal rates), and thereby the peak masking effects constantly appears at one of

the most sensitive temporal frequencies which are around 1 Hz and 8-12 Hz.

e. Two pooling stages of the spatial and temporal signals lie in cascade

This idea is provided based on the results from prior masking works which have been employed based on the very narrowband mask pattern (i.e., Lehky, 1985; Cass & Alias, 2006) and three Control conditions in Experiment 1 which conditionally varied the spatial broadness of the mask pattern.

Multiple local (*'tuned'*) gain control pools across the spatio-temporal surface are tuned in temporal frequency (Experiment 1) and spatial frequency (Experiment 2) dimensions, and all of these pools show the significant horizontal effect. This indicates that each detecting mechanism across the spatio-temporal locations has its own pool which includes two pooling stage of the spatial and temporal signals which lies in cascade fashion (i.e., the spatial pooling stage precedes the temporal pooling stage). In the first spatial pooling stage, each spatial detector across the different spatio-temporal surface has its own pool whose output is combined and weighted by neighboring spatial detectors (i.e., tuned in similar spatial frequency and orientations) which reflects the many spatially tuned detectors with maximum sensitivity around 4 cpd. In the second temporal pooling stage, the processed signals at the prior spatial pooling stage are fed into the next temporal pooling stage in which each temporally tuned detector has its own pool whose output is combined and weighted by neighboring temporal detectors (tuned in similar temporal rates). Critically, in order to drive these multiple local spatially- and temporally-tuned anisotropic suppressions (showing the horizontal effect), the use of the particular mask stimulus (i.e., naturalistic ( $1/f$ ) and complex) which can rise above a

certain detection threshold of each of the many temporal detectors is essential. Specifically, prior masking studies using very narrowband masks have shown only 2 (or 3) temporally-tuned channels (Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). This very narrowband mask stimuli might pass through relatively unaltered with performance reflecting the weightings that form the conventional spatial channel (i.e., this narrowband pattern stimulates a few spatial detectors which is not adequate enough to create pooled activity from the similar spatial detectors). Then, these un-weighted and un-altered signals indicating a weak stimulation at the spatial pooling stage (which is not adequate to drive many temporally-tuned pools) is fed into the next temporal pooling stage. Thus, the only 2 (or 3) peak masking positions indicating the peak sensitivity of each of the conventional temporal detectors are observed at 1-2 Hz or 8-12 Hz (or very high temporal frequency around 20 Hz, see Hess & Snowden). Control conditions in Experiment 1 also addressed this issue (see **Figure 15**).

In summary, two plausible accounts can be logically deduced. First, each detecting-mechanism across the spatio-temporal surface has its own anisotropic gain control pool which consists of two separate spatial and temporal pooling stages which lies in cascade. The spatio-temporal signals are initially encoded by the spatial pooling stage (in which static spatial visual information are mainly encoded) and then the processing signals at this spatial stage are fed into the next temporal pooling stage (in which dynamic temporal information are mainly encoded). Thus, in order to drive the multiple temporally tuned detectors, the significant stimulation at the prior spatial pooling stage is crucial and this can be obtained by the broadband spatial pattern.

Note that this idea is only acceptable only if it is true that spatial and temporal signals are separately processed in these two pooling detectors.

f. Two pooling stages of spatial and temporal signals are overlapped

The other possible idea is that the spatial and temporal pooling stages are overlapped as resembling the spatial and temporal encoding processes by overlapping group of encoding-mechanisms within each hypercolumn at the cortical stage (i.e., Bonhoeffer & Grinvald, 1991; Hubel & Wiesel, 1977; Hubener et al., 1997; Issa et al., 2000; Shmuel & Grinvald, 1996; Weliky et al., 1996). In this view, a recent neurophysiological work (Baker & Issa, 2005) provides a plausible insight in the following. They proposed that to create more realistic response activities from many (population) encoding-mechanisms at the early cortical stage (V1), the use of the realistic and complex visual stimuli is highly required. Specifically, this work demonstrated that the distributed pattern activity in the cortex generated in response to complex motion images can be characterized by three maps of separable tuning properties: orientation, spatial frequency, and temporal frequency tuning.

According to this neurophysiological insight, it can be logically inferred that the spatial and temporal pooling stages would be overlapped as resembling the spatial and temporal encoding processes by overlapping group of encoding-mechanisms within each hypercolumn at the early cortical stage (V1). This overlapping process of spatial (including orientation and spatial frequency) and temporal pooling would be used to account for the mechanisms of neural encoding underlying the visual performance if in the following two assumptions are true. One is that there are many temporal detectors,



each of them is tuned in different temporal frequency (see figure 8B in Alitto & Usrey, 2004), and there exists mutual suppression between detectors tuned in similar temporal frequency. Second is that each of the many temporal detectors has different detection criterion, and thereby to stimulate these many temporal detectors the use of the appropriate visual stimuli is crucial. Note that several detection models are based on assumption that the detection occurs when the activity in the detector most sensitive to the stimulus rises above a certain threshold. Naturalistic ( $1/f$ ) broadband patterns used in Experiment 1 and complex motion stimuli (Baker & Issa, 2005) are thought to be an appropriate visual stimuli which would be adequate enough to rise above a certain detection criterion of the many spatio-temporally tuned encoding-mechanisms at the early cortical stage (V1), and thereby the strong spatially- or temporally-tuned suppression as reflecting multiple local (*'tuned'*) gain control pools whose output is altered by the response signals of the neighboring spatio-temporally tuned encoding-mechanisms.

Taken together, it can be logically assumed that, in order to develop more realistic model of the human spatial and temporal processing, the use of the realistic (or naturalistic ( $1/f$ )) broadband mask stimulus is essential. Specifically, the naturalistic ( $1/f$ ) broadband pattern which can be seen in everyday (naturalistic) viewing situations would create more realistic response activities from numerous detecting-mechanisms at the early cortical stage (V1) which would be adequate enough to drive the pooled activities which can be achieved by mutual suppression between similar spatio-temporal detectors. However, the narrowband grating stimulus would create a few and most sensitive temporal detectors' responses which are not adequate enough to lead the separable tuning properties (Baker & Issa, 2005), and thereby the only most sensitive 2 (or 3) temporal

detectors are observed in prior masking studies (e.g., Lehky, 1985; Anderson & Burr, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009) and Control conditions (i.e., **Figure 14** and **15**) in the current Experiment 1.

## 5.0 CONCLUSIONS

A local (*'tuned'*) gain control pool reflects local suppression, which can be achieved by mutual suppression between neighboring (similar) spatio-temporal detectors. Specifically, this suppression is anisotropic, being relatively greater at horizontal than oblique orientations. There exist multiple pools across the different spatio-temporal locations, and all of these pools are stronger at horizontal (showing the horizontal effect). All of these pools are tuned in the temporal frequency (Experiment 1) and spatial frequency (Experiment 2) dimensions. The pools have the same standard Gaussian shape (however, some pools have an asymmetric shape) and size in both temporal (i.e., 1.52 octaves in **Figure 17** and **18**) and spatial (i.e., 1.35 octaves in **Figure 26** and **27**) dimensions. Critically, current results showed that multiple temporally-tuned pools are revealed with a naturalistic ( $1/f$ ) broadband mask (i.e., **Figure 15**), whereas prior studies using very narrowband masks have shown only 2 (or 3) temporally-tuned detectors (e.g., Lehky, 1985; Hess & Snowden, 1992; Fredericksen & Hess, 1998; Cass & Alias, 2006; Cass et al., 2009). We propose that to drive these pools, the use of an appropriate broadband mask which can rise above a certain detection threshold of each of the underlying many, unknown, and lesser sensitive temporal detectors is crucial, and thereby can create strong mutual suppression between detectors tuned to similar temporal rates which form local (*'tuned'*) pools.

**Table 1:** Summary of the mask stimuli used in prior temporal masking studies

	Test		Mask	
	Spatial	Temporal	Spatial	Temporal
Lehky (1985)	Single frequency	Single frequency	Single gratings	Single frequency
Anderson & Burr (1985)	Single frequency	Single frequency	Unfiltered one dimensional vertical noise (not clearly described in text)	Single frequency
Hess & Snowden (1992)	Single frequency	Single frequency	Sum of 4 gratings (i.e., 0.3 + 1 + 3 + 15 cpd)	Single frequency
Snowden & Hess (1992)	Single frequency	Single frequency	Sum of 4 gratings (i.e., 0.3 + 1 + 3 + 15 cpd)	Single frequency
Fredericksen & Hess (1998)	Single frequency	Single frequency	0.5-octave wide in spatial frequency	4 Hz wide in temporal frequency
Cass & Alias (2006)	Single frequency	Single frequency	Single grating	Single frequency
Cass, Alias, Spehar & Bex (2009)	Single frequency	Single frequency	0.5-octave above and below centered at 2.4 cpd	Single frequency

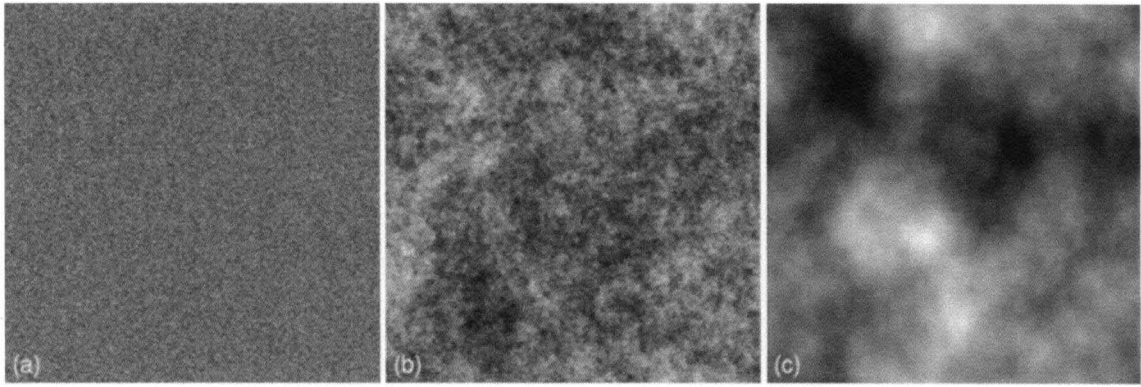
Note that these works showed the temporal frequency tuning of single grating targets at different spatio-temporal combinations by using the masking paradigm which is the same paradigm used in the current Experiment 1.

**Table 2:** Mask specifications used in prior works

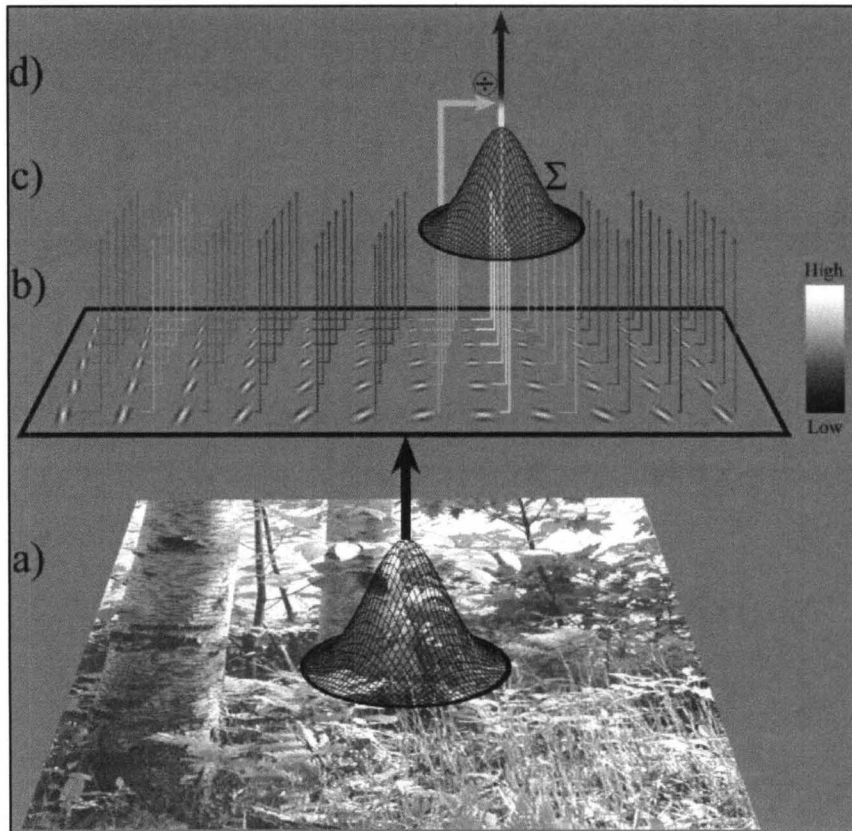
	Test	Mask
Essock et al., (2003)	A 1-octave (8-16 cpd) spatial frequency band & a 20° orientation band	A 1/f broadband noise including spatial frequency ranging from 0.2-17 cpd & all orientations
Haun & Essock (2010)	A 1-octave centered at 0.65, 1.50, 3.44, 7.90, and 18.25 cpd, & 45° orientation band	A 1/f broadband noise including spatial frequency ranging from 0.2-25.6 cpd & all orientations

**Table 3:** Summary for the mask stimuli used in Control conditions for Experiment 1

	Spatial frequency	Orientation band	Contrast
Control 1	1-octave (1-2 cpd)	1°	0.045 RMS
Control 2	1-octave (1-2 cpd)	1°	0.1 RMS
condition 8	4-octaves (1-16 cpd)	15°	0.1 RMS
Control 3	2-octaves (1-4 cpd)	7°	0.18 RMS
condition 10	4-octaves (1-16 cpd)	15°	0.1 RMS

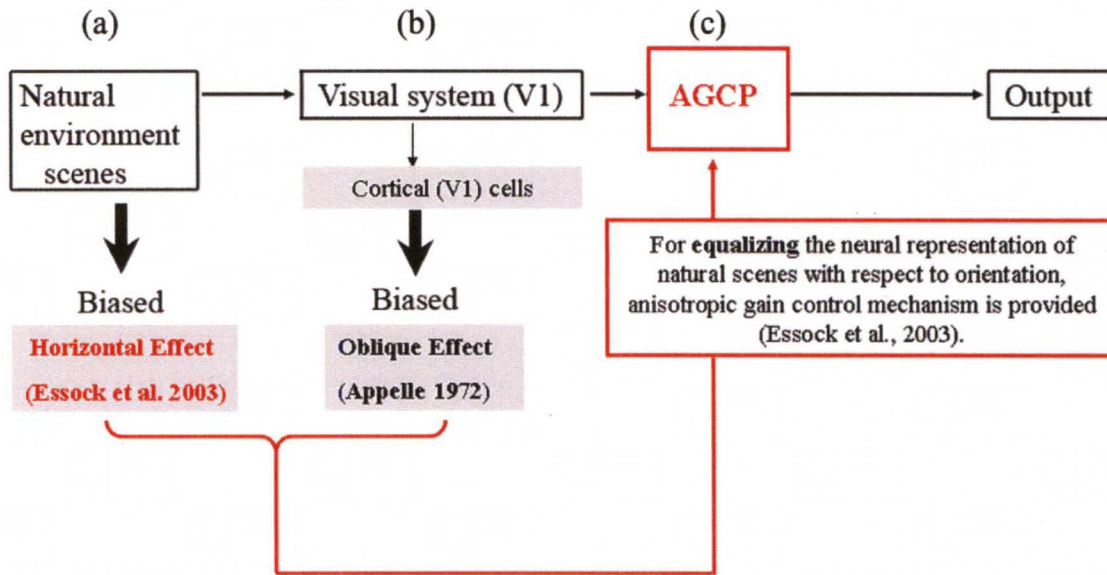


**Figure 1.** Example stimuli to show the high spatial frequency bias (a) Gaussian white noise ( $1/f^0$ ), (b)  $1/f^{-1}$  noise, and (c)  $1/f^{-2}$  noise (images from Chandler & Field, 2007)



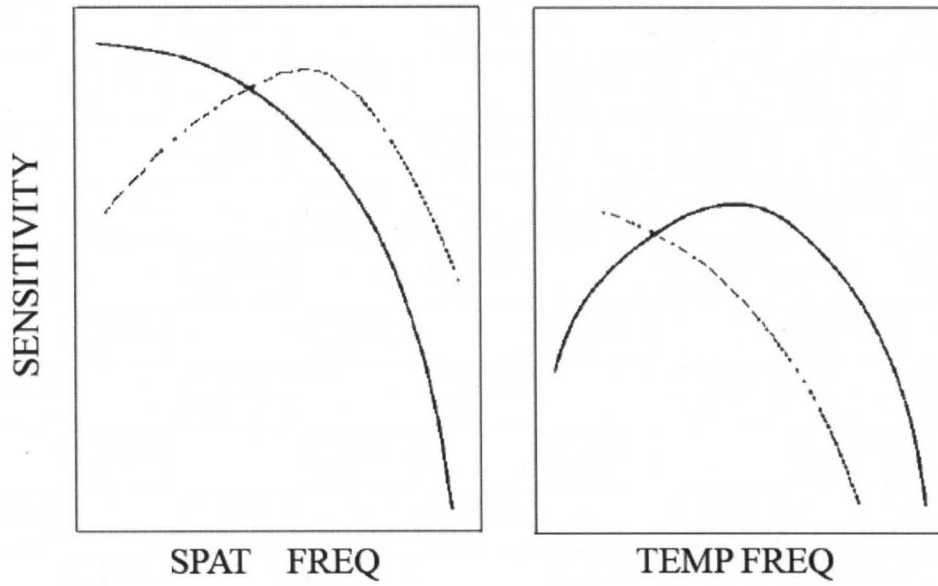
**Figure 2.** Anisotropic contrast gain control pool: (a) For a given region within a scene, (b) structure (contrast at different spatial scales and orientation) is extracted by oriented filters of different size. (c) Output of similar filters is pooled, and (d) combined divisively with a filter's output. Final output determines perceptual salience. Specifically, the inhibitory pooling signal ( $\sum o \cdot w \cdot I$ ) to the denominator in Equation 2 is biased in orientation dimension which is greatest at horizontal, intermediate at vertical, and least at oblique orientations (Horizontal > Vertical > Oblique orientation bias) present in natural scenes. This “un-does” the natural bias and whiten the orientation-biased natural scenes (original illustration from Hansen & Essock, 2004).



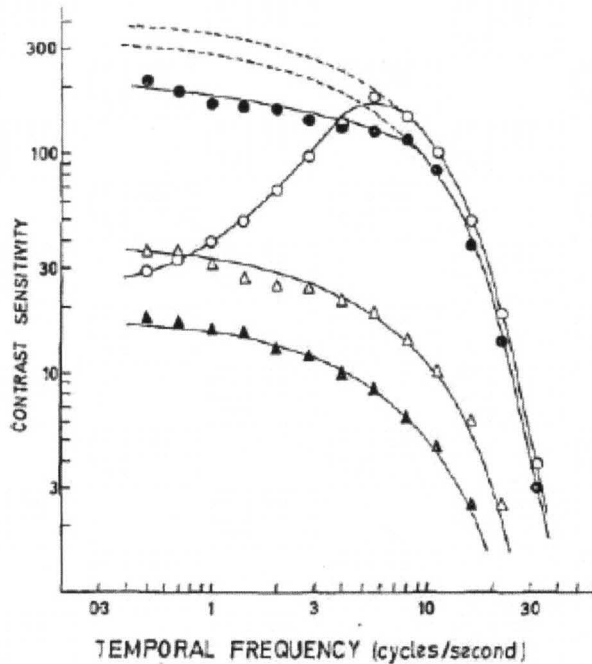


**Figure 3.** Encoding mechanisms underlying the human vision at the early stage (V1) respond to contrast structures in natural scenes at a wide range of spatial frequencies, temporal frequencies and orientation in a particular manner that is efficient (Barlow, 1961). Thereby, it is worthy to know the statistical properties of natural scenes, and then we can relate these statistical properties to the encoding processes underlying the visual system by measuring the visual performance (e.g., sensitivity etc) for particular stimuli resembling the statistical properties present in natural scenes. (a) Orientation is one of the fundamental dimensions in natural scene perception and is being explored in detail in our lab. Much of our understand of orientation processing comes from studies by measuring the behavioral performance for simple gratings, whereas current understanding of orientation processing comes from the studies by measuring the behavioral performance for visual stimuli resembling the statistical regularities of natural scenes. Critically, the significant change in orientation perception—i.e., from the oblique effect (Appelle, 1972) to the horizontal effect (Essock et al., 2003)—was obtained as adding more content to the narrowband stimulus (Hansen & Essock, 2004). (c) The significant change in the

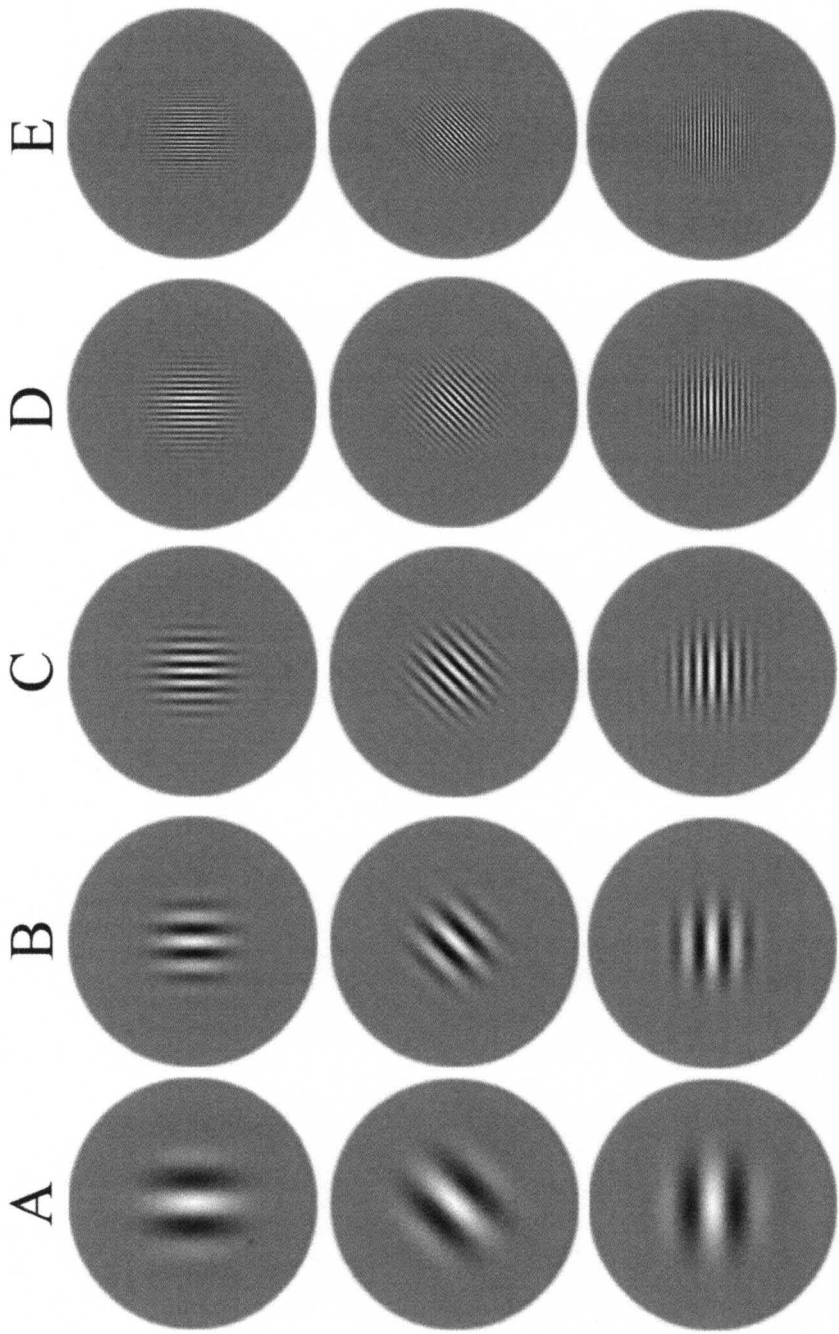
orientation perception results from the anisotropic gain control mechanisms (strongest at horizontal), whose effect is pronounced when viewing a pattern of broadband spatial content resulting in the horizontal effect. The additional content creates more response activities in channels contributing to gain control pools whereas the response activity in channels is not significant with narrowband stimuli, resulting in the oblique effect (Hansen & Essock, 2006; Essock et al., 2009; Haun & Essock, 2010; Kim et al., 2010). Specifically, we have suggested that the horizontal effect is linked to the anisotropic gain control pool that causes most suppression at horizontal, least at obliques, and intermediate at vertical orientations. Through this orientation-biased pooling process, the horizontal content bias present in typical natural scenes are equalized. From this point of view, the horizontal effect could be a great perceptual hallmark which shows that the encoding mechanisms underlying the visual systems are likely to have evolved to reflect natural scenes, and in order to equalize the prevalent content bias present in natural scenes—specific optimizing processing strategy to 'whiten' the neural response to natural scenes via anisotropic gain control pool (**c**). Such anisotropic suppression in the gain control mechanism could be direct (**b**) consequence of corresponding bias of cortical cells (Li et al., 2003), indicating more cells tuned to cardinal than other oblique orientations. In addition, (**a**) the orientation bias of typical natural scenes (Hansen & Essock, 2004; 2005; 2006) indicates the most content at the horizontal, least content at the obliques and intermediate at vertical.



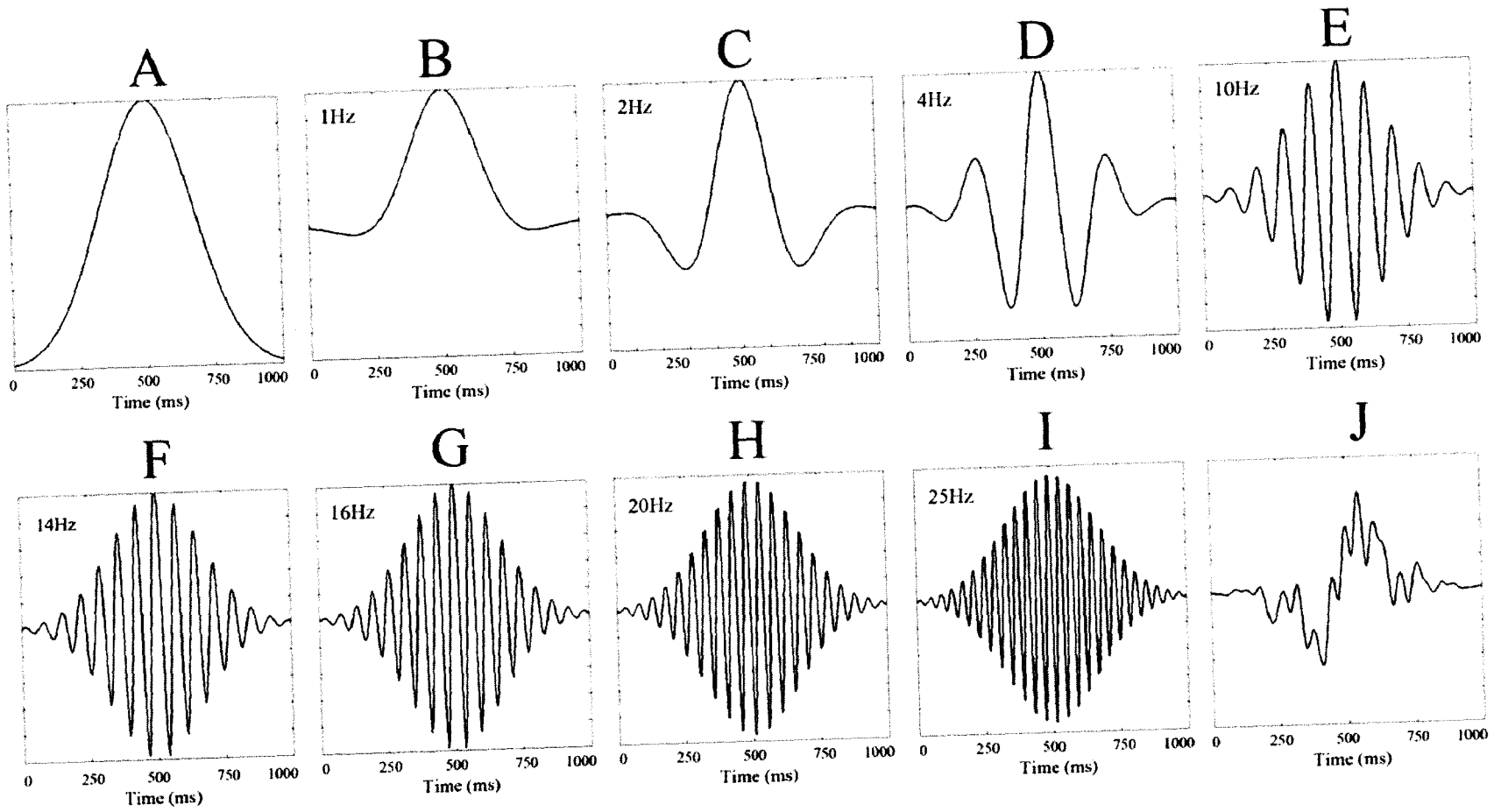
**Figure 4.** Hypothetical sensitivity of sustained and transient mechanisms. Left panel shows sensitivity for sustained (dashed lines) and transient systems (solid lines) as a function of spatial frequency. Right panel represents sensitivity as a function of temporal frequency (from Green, 1984).



**Figure 5.** Temporal contrast sensitivity functions for different spatial frequencies. The test was a sinusoidal grating with spatial frequency of  $\circ$  0.5 cpd,  $\bullet$  4,  $\Delta$  16,  $\blacktriangle$  22 cpd. Points are averages of four adjustments thresholds. The curves (including dashed sections) differ only in vertical position. Spatial and temporal sensitivity seems to be separable at high temporal (i.e.,  $>10$  Hz) and spatial frequencies (i.e.,  $\Delta$  16,  $\blacktriangle$  22 cpd), but at low spatial frequencies ( $\circ$  0.5 cpd) sensitivity at low temporal frequencies (i.e.,  $<10$  Hz) seems to be reduced (from Robson, 1966). Specifically, the tuning of a low spatial frequency target ( $\circ$  0.5 cpd) is overlapped with the tuning of high spatial frequency targets ( $\Delta$  16,  $\blacktriangle$  22 cpd) at low temporal frequency regions.

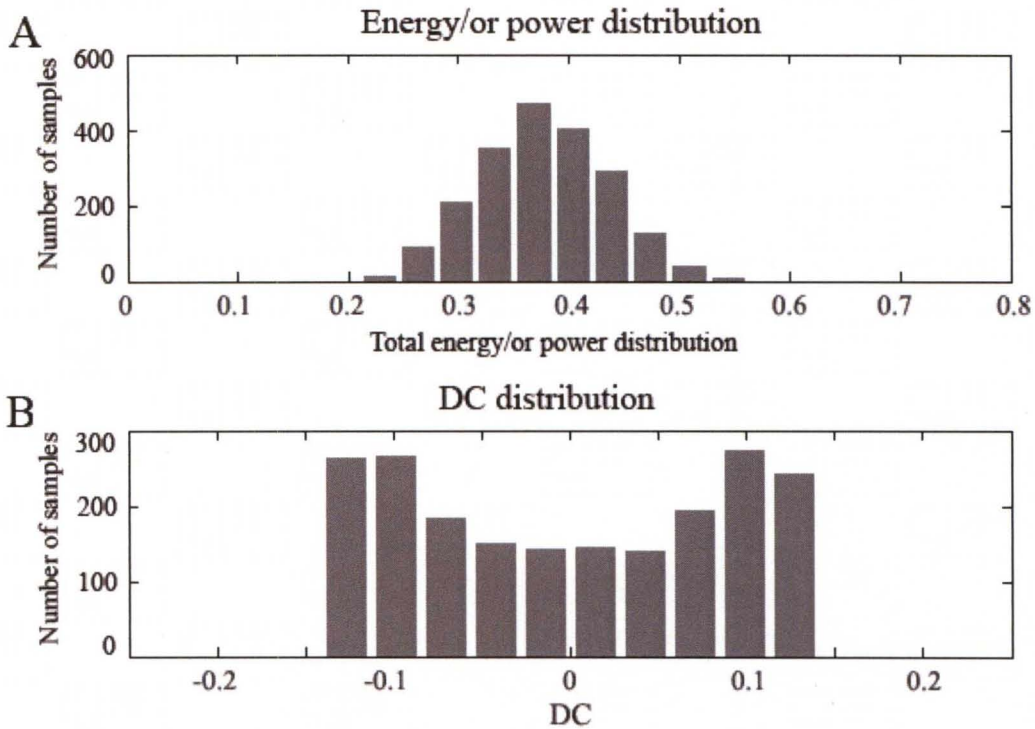


**Figure 6.** Illustration of the test targets used in Experiment 1 and Experiment 2. Test stimuli were cosine grating, either 1, 2, 4, 8, or 12 cpd from left to right column (**A-E**), Gaussian windowed to a  $1^\circ$  full-width at half-height at one of the three orientations ( $90^\circ$ ,  $45^\circ$ , and  $0^\circ$ ) from top to bottom row.

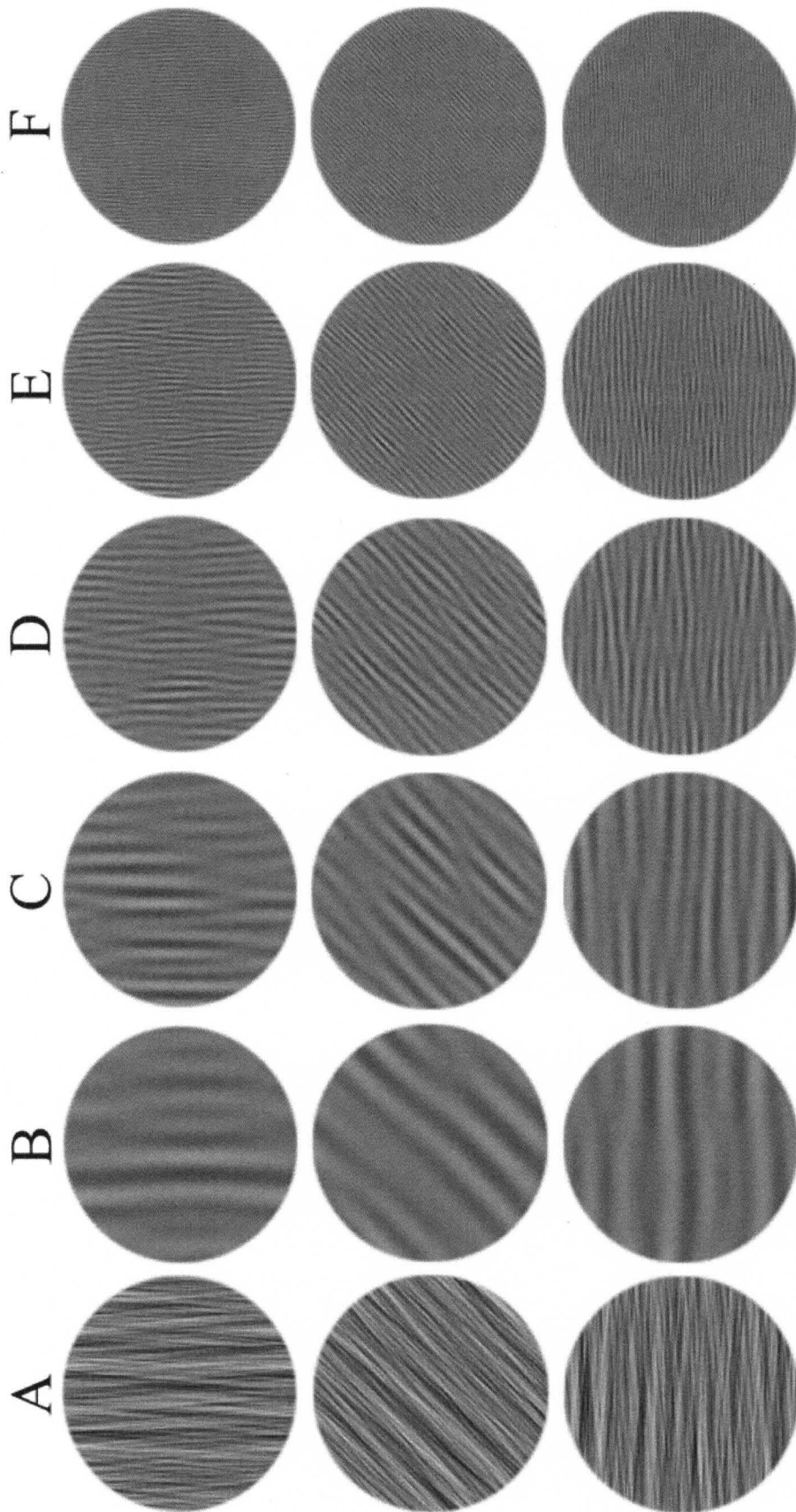


**Figure 7.1.** Illustration of the temporal profiles used in Experiment 1 and Experiment 2. A temporal Gaussian envelope (**A**) used to modulate the contrast of the test and mask stimuli is 1000ms in duration, with 400ms at half-height of the Gaussian. This temporal envelope is used to be multiplied by cosine temporal waveforms resulting in the Gabors shown (**B-I**) for Experiment 1 and a broadband temporal-noise sample (**J**) with randomized phase for Experiment 2.

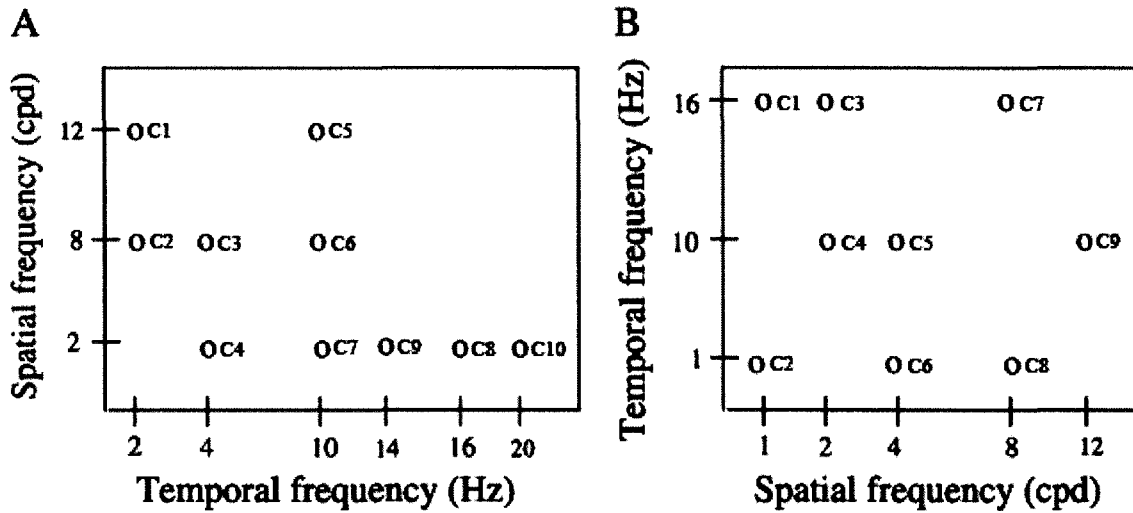




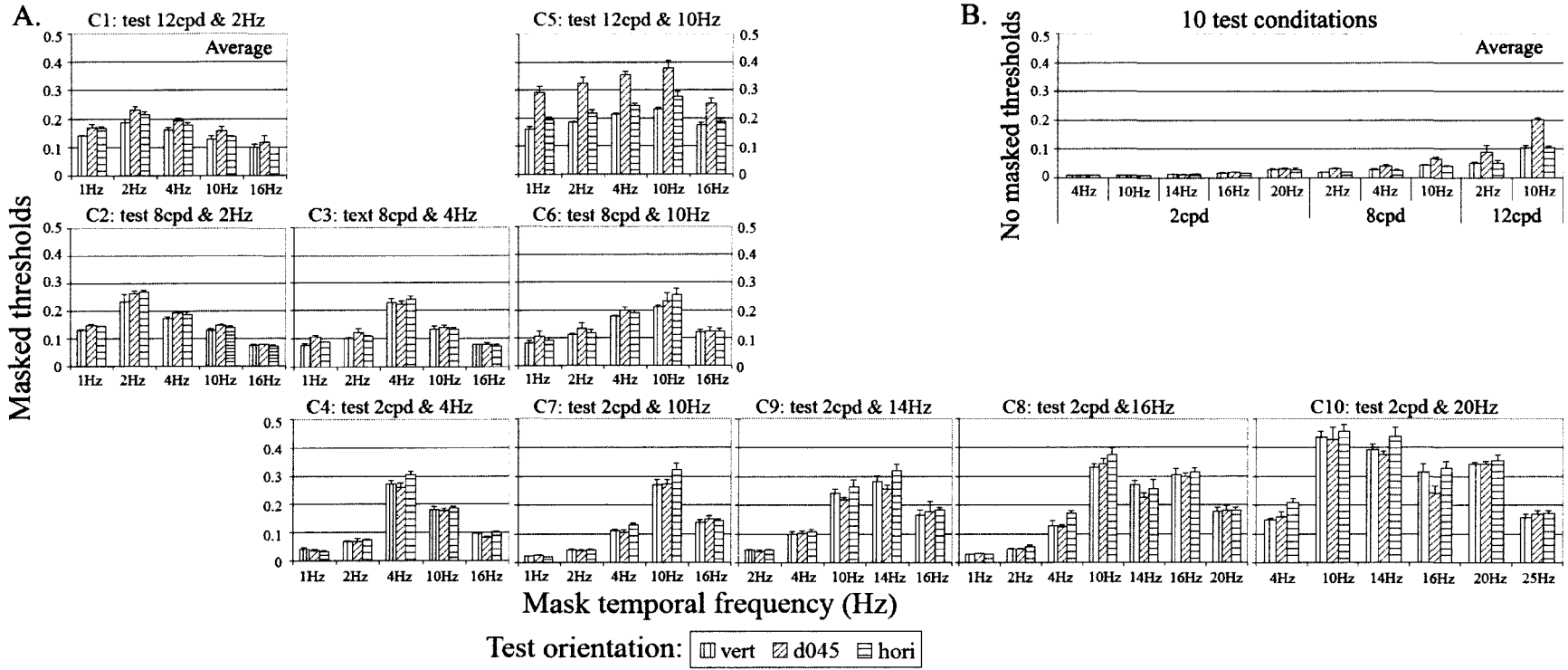
**Figure 7.2.** Total energy (or contrast/power) and DC components of 2000 temporal-noise samples used in Experiment 2 were plotted 'A' and 'B', respectively. The total energy of each temporal-noise sample can be calculated by the standard deviation of each temporal-noise sample. The DC value of each temporal-noise sample can be calculated by the average of each temporal-noise sample. In order to demonstrate that the total energy of each temporal-noise sample is constant ( $\approx 0.39$ ), one randomly takes the 2000 temporal-noise samples, and then calculates the standard deviation (corresponding to the total energy) and the average (corresponding to the DC value) of each temporal-noise sample which are plotted. The total energy (**A**) is close to 0.39 and DC deviation (**B**) is close to zero.



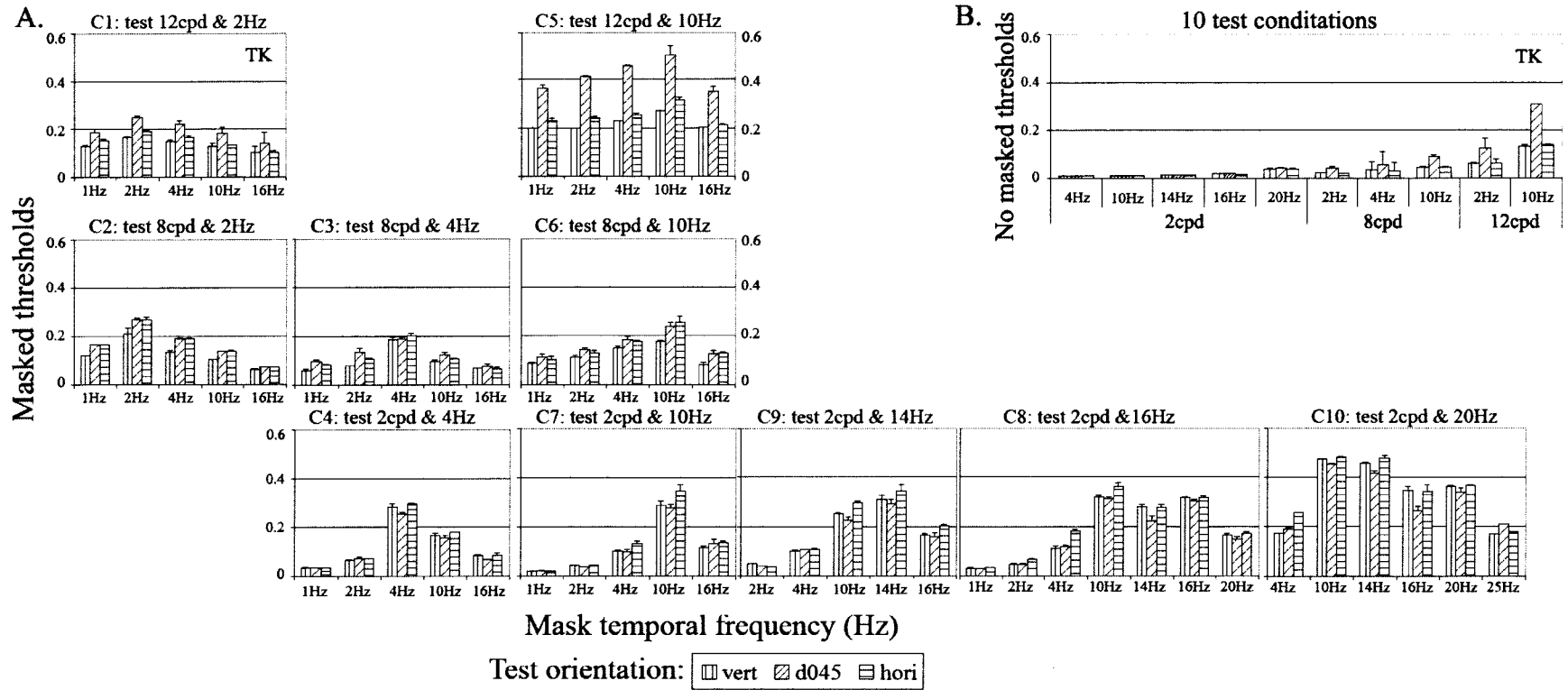
**Figure 8.** Illustration of the mask stimuli used in Experiment 1 (**A**) and Experiment 2 (**A-F**). Mask stimuli with  $1/f$  spatial and randomized phase at one of the three orientations from top to bottom row ( $90^\circ$ ,  $45^\circ$ , and  $0^\circ$ ). The spatial frequency band is varied at **A-F**, and the orientation band is fixed as  $15^\circ$ . (**A**) A broadband mask, the spatial frequency band used is four octaves, ranging from 1 to 16cpd. **B-F**) The broadband mask (**A**) split into five 0.8-octave sub-bands. From left to right column (**B-F**), each 0.8-octave sub-band's center spatial frequency is 1, 2, 4, 8 and 12 cpd (22 cpd center frequency band only used in condition 6 at Experiment 2 which was not shown here because of not enough resolution) respectively.



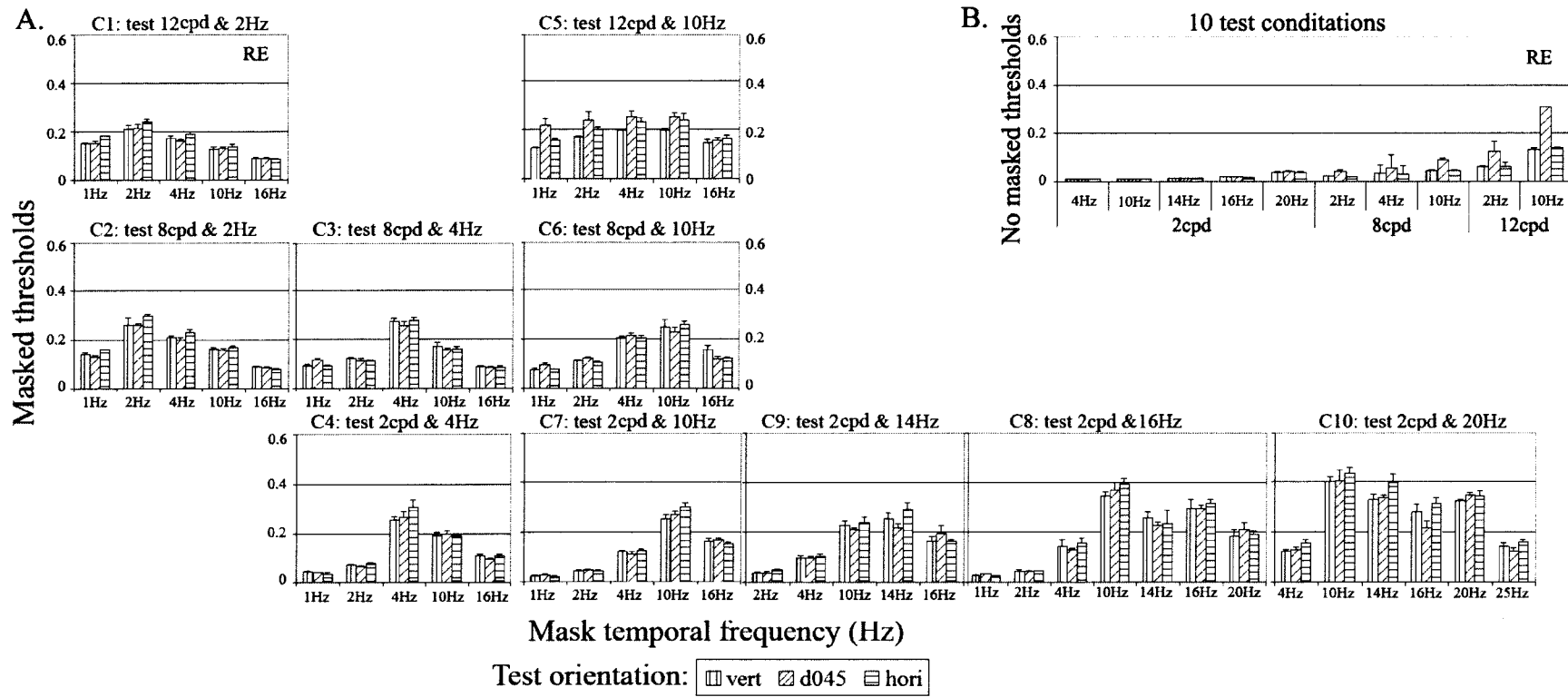
**Figure 9.** Experimental conditions in Experiment 1 (A) and Experiment 2 (B). In Experiment 1, each target across the 10 spatial and temporal combinations denoted by open circles was used. Threshold of each target at different locations was measured without and with spatially broadband mask across varied temporal frequency of the mask. In Experiment 2, each target across the 9 spatio-temporal combinations denoted by open circles was used. Threshold for each target was measured without and with temporally broadband mask across varied spatial frequency of the mask.



**Figure 10.** Contrast thresholds for each grating target at each of three test orientations (vertical, 45° and horizontal) across the 10 spatial and temporal combinations with the temporally varied naturalistic ( $1/f$ ) broadband mask stimuli (**A**) and without the mask (baseline) (**B**). For each test condition, masking was measured by using a spatially broadband masking across varied temporal rates of the mask. The ‘average’ of two subjects is shown. Error bars are  $\pm$  standard error of the group mean.

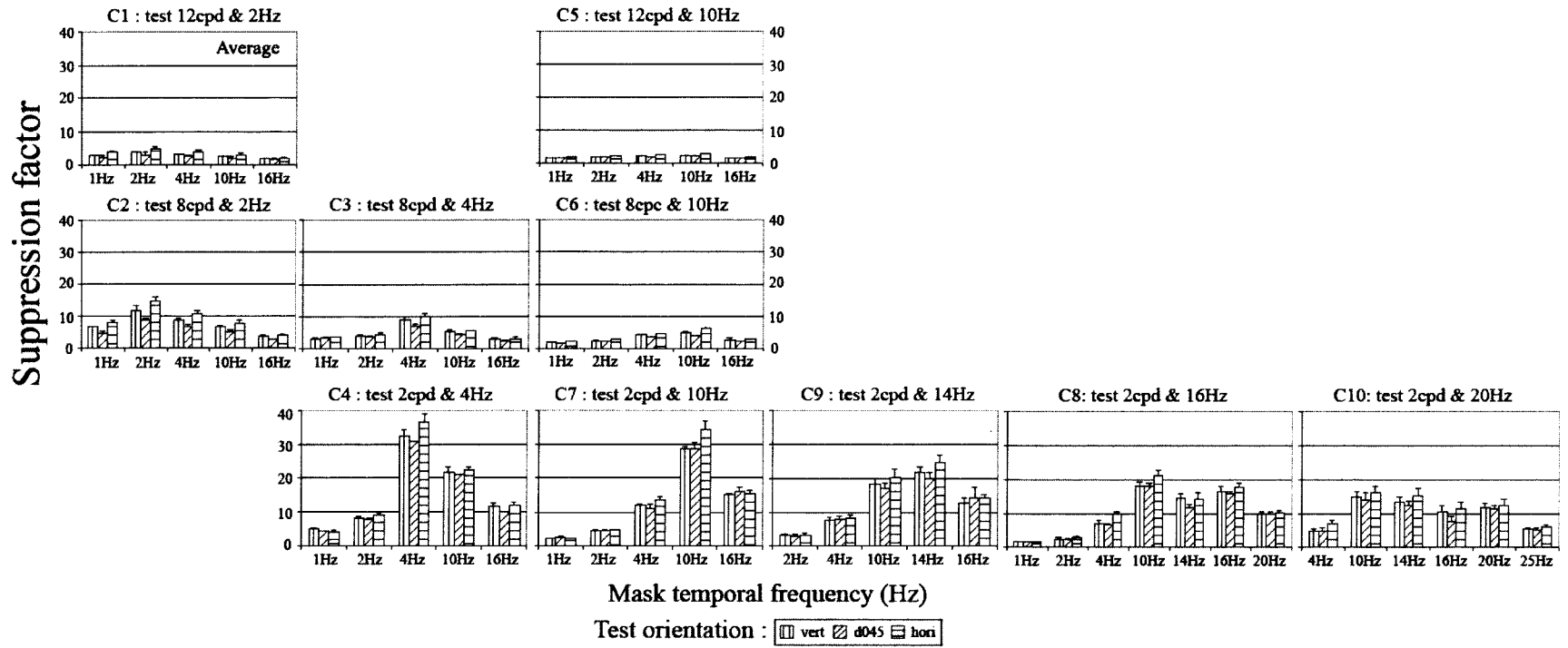


Same measurements as in **Figure 10** for a subject 'TK'

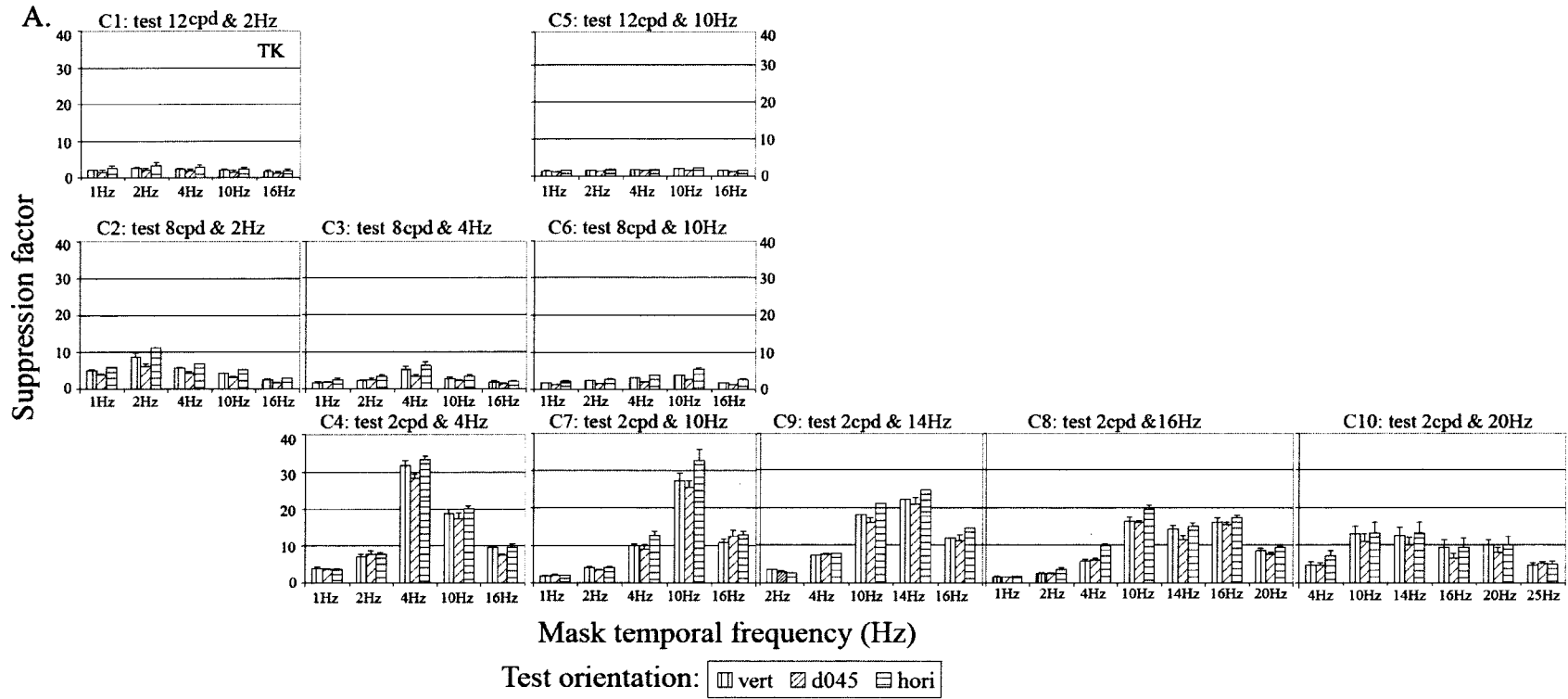


Same measurements as in **Figure 10** for a subject 'RE'

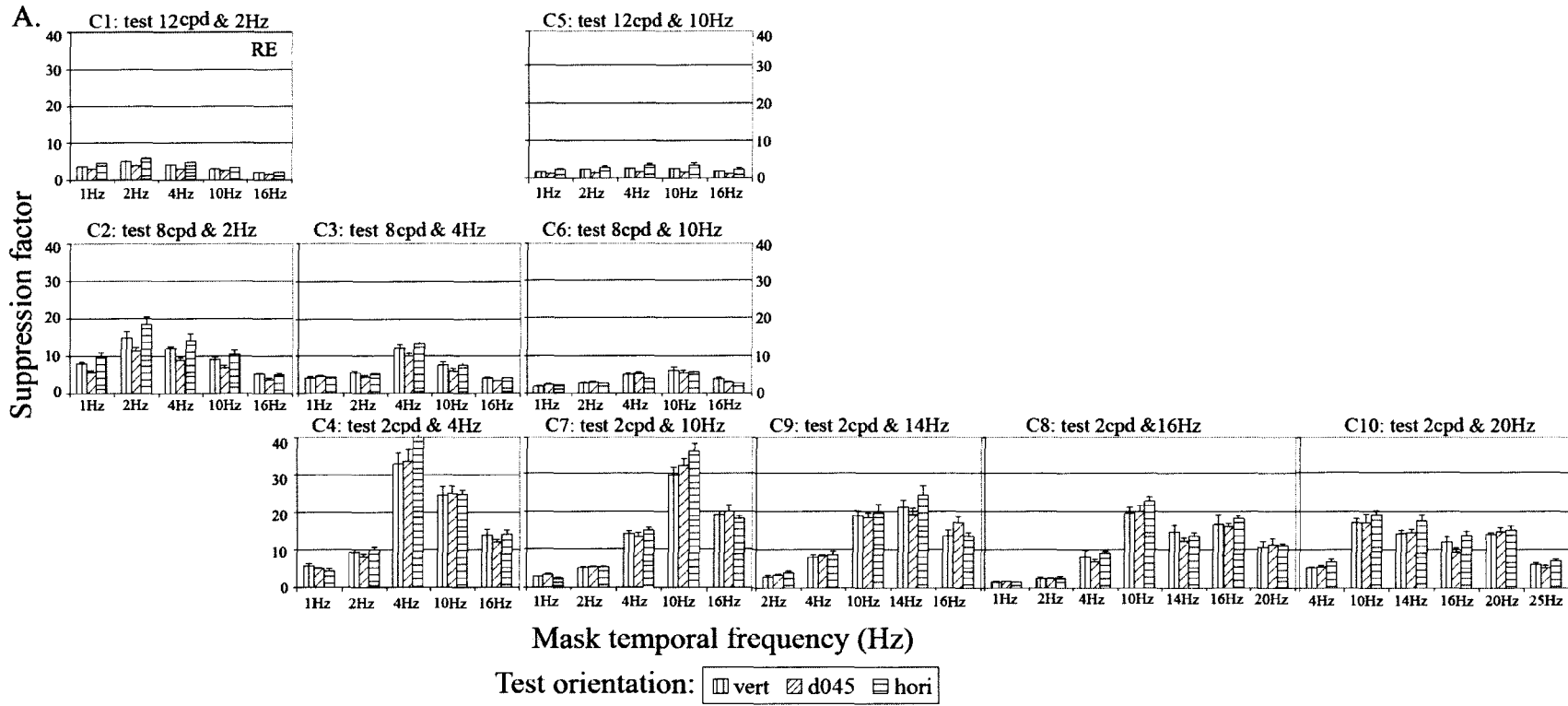




**Figure 11** Suppression factor (the ration between masked and unmasked thresholds) of each target across three orientations (vertical, 45° and horizontal) and across the 10 spatio-temporal combinations is shown for mean thresholds (in Figure 10). The standard errors were calculated in terms of Gaussian error propagation (Lo, 2005) and the mean of these standard errors are shown.

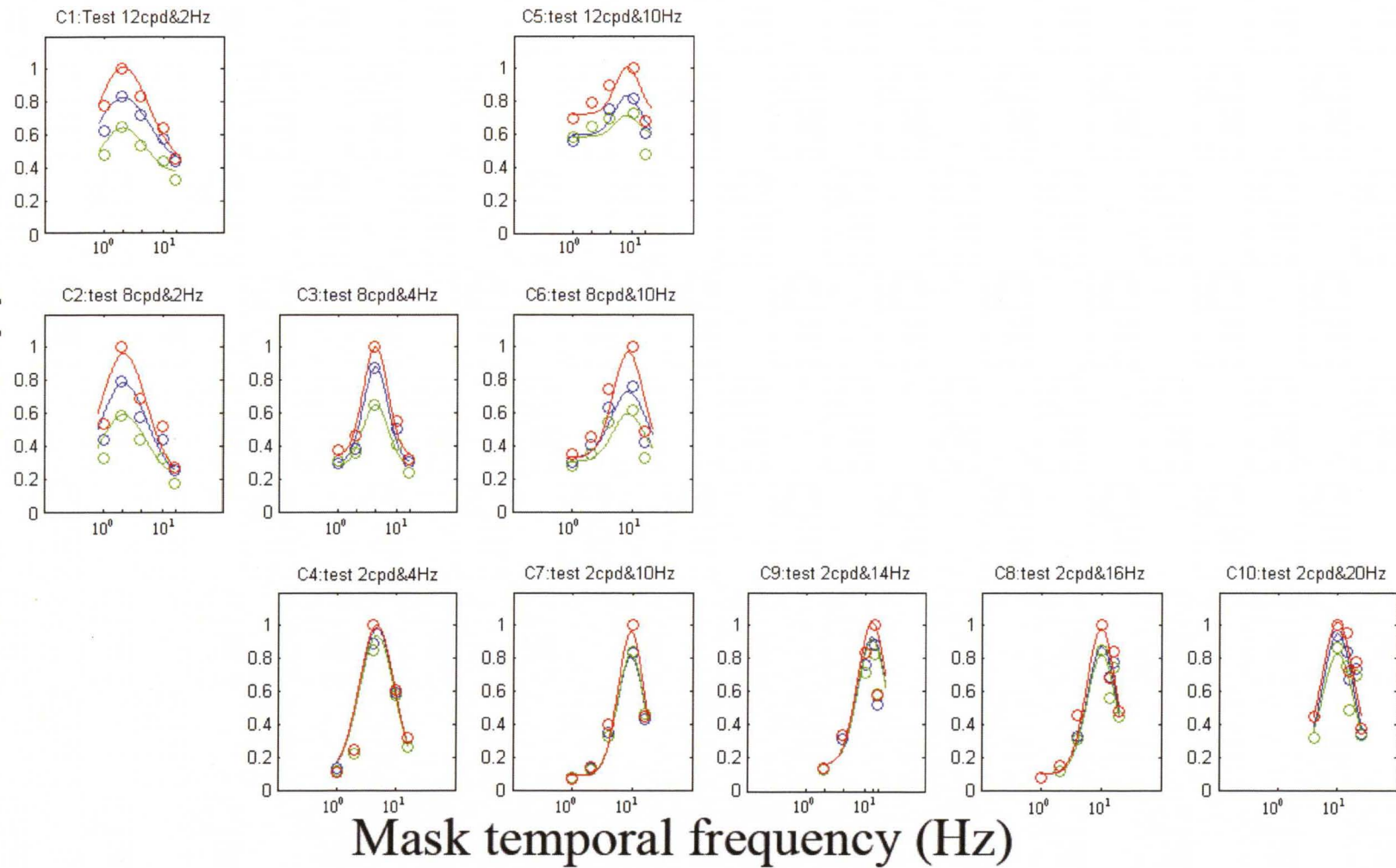


Same measurements as in **Figure 11** for a subject 'TK'



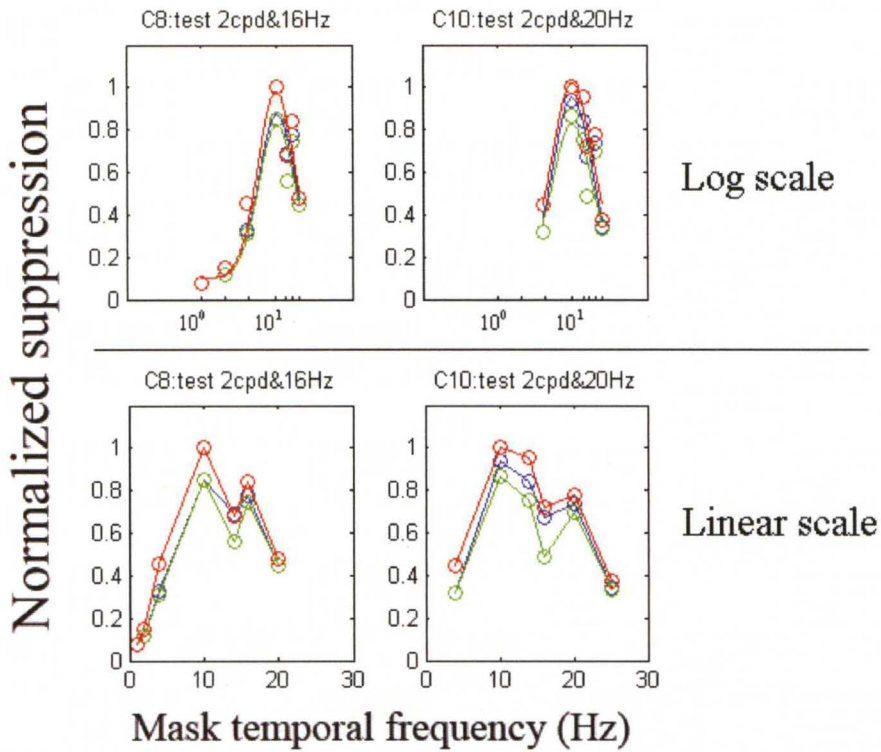
Same measurements as in **Figure 11** for a subject 'RE'

# Normalized suppression



**Figure 12A.** Normalized suppression (NSp: each suppression value was divided by the single maximum suppression value in that condition) for comparing magnitude of the horizontal effect anisotropy in temporal frequency tuning curves across 10 spatio-temporal locations. Each tuning curve in each condition is plotted on log scale. The Gaussian function (smooth curve) provides a Best-fit (open circles, SSE average across conditions = 0.043) describing temporal frequency tuning curves. Test orientation: Horizontal (red circles), oblique (green circles), and vertical (blue circles). The peak masking effect of the masking function appears at the same temporal frequency of the test up to 14 Hz. Beyond 14 Hz, the peak masking effect appears at 10 Hz, not the target temporal frequency when plotting on log scale.





**Figure 12B** Plotting on different scales for condition 8 and 10. The same data for condition 8 and 10 is re-plotted on log scales in top, and the linear scale in bottom. As can be seen a top in **Figure 12B**, the log scale is a little compressed at the high temporal frequency sides of the masking function, and thereby the peak masking effect at each target's temporal frequency (beyond 14 Hz) appears at 10 Hz. The use of an appropriate scale seems to be crucial in order to escape misinterpretation of the given data.



12cpd/2hz	VER	D45	HOR
$W$	1.71	1.57	1.71
$A$	0.78	0.28	0.58
$C$	0.23	0.37	0.40
$x_0$	2.14	2.14	2.14

12cpd/10hz	VER	D45	HOR
$W$	1.42	1.39	1.39
$A$	0.24	0.13	0.29
$C$	0.60	0.59	0.72
$x_0$	9.03	9.03	9.03

8cpd/2hz	VER	D45	HOR
$W$	1.85	1.85	1.85
$A$	0.64	0.44	0.82
$C$	0.11	0.11	0.11
$x_0$	2.14	2.14	2.14

8cpd/4hz	VER	D45	HOR
$W$	1.40	1.42	1.40
$A$	0.59	0.38	0.66
$C$	0.29	0.27	0.34
$x_0$	4.21	4.21	4.21

8cpd/10hz	VER	D45	HOR
$W$	1.59	1.52	1.51
$A$	0.40	0.29	0.64
$C$	0.33	0.30	0.33
$x_0$	9.33	9.33	9.33

2cpd/4hz	VER	D45	HOR
$W$	1.45	1.45	1.45
$A$	0.87	0.83	0.89
$C$	0.11	0.11	0.11
$x_0$	4.50	4.50	4.50

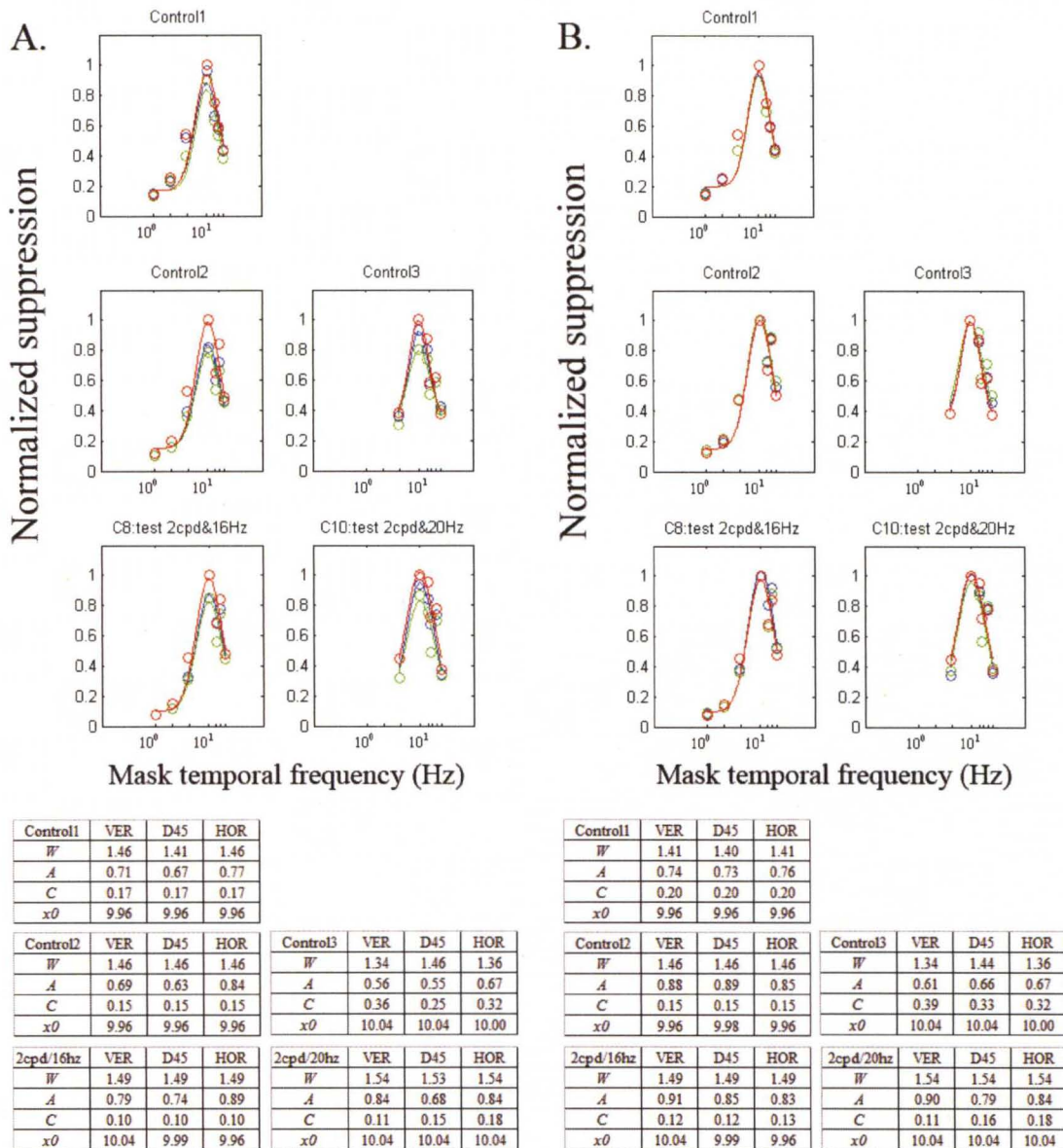
2cpd/10hz	VER	D45	HOR
$W$	1.45	1.44	1.42
$A$	0.72	0.73	0.87
$C$	0.09	0.09	0.09
$x_0$	9.73	9.73	9.73

2cpd/16hz	VER	D45	HOR
$W$	1.49	1.49	1.49
$A$	0.79	0.74	0.89
$C$	0.10	0.10	0.10
$x_0$	10.04	9.99	9.96

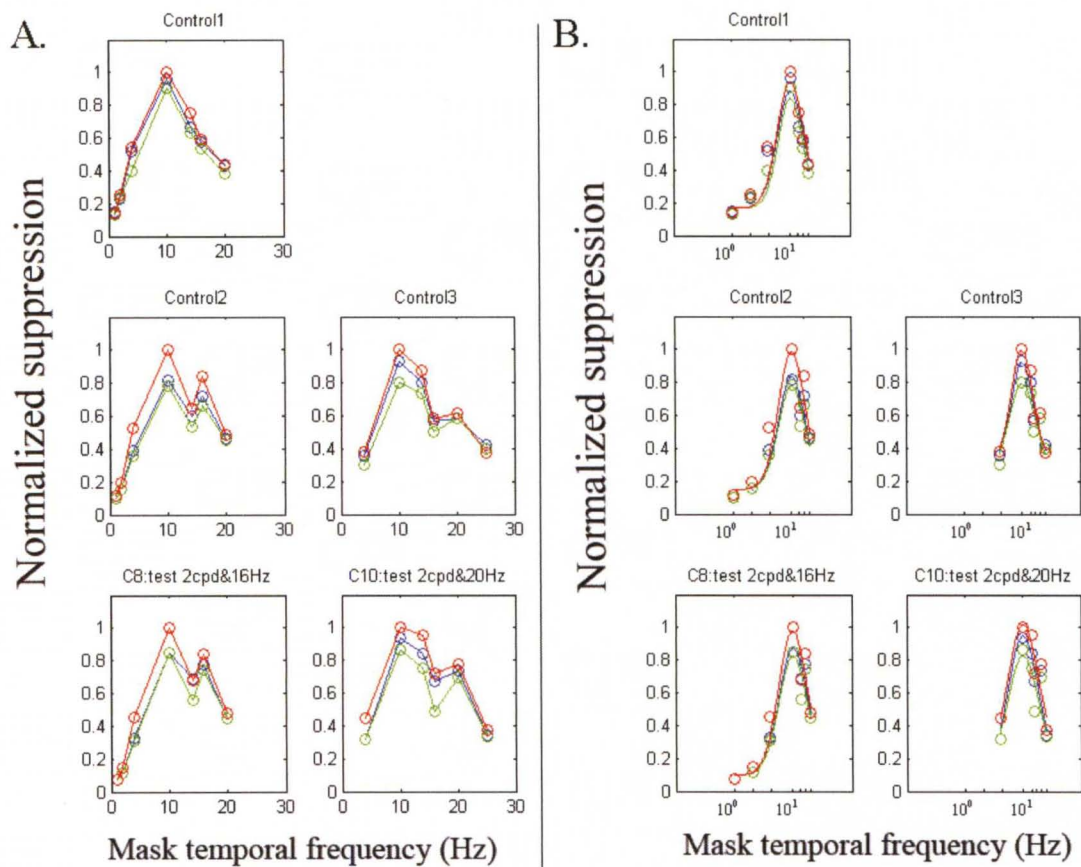
2cpd/14hz	VER	D45	HOR
$W$	1.48	1.48	1.48
$A$	0.78	0.76	0.86
$C$	0.09	0.09	0.09
$x_0$	13.65	13.65	13.65

2cpd/20hz	VER	D45	HOR
$W$	1.54	1.53	1.54
$A$	0.84	0.68	0.84
$C$	0.11	0.15	0.18
$x_0$	10.04	10.04	10.04

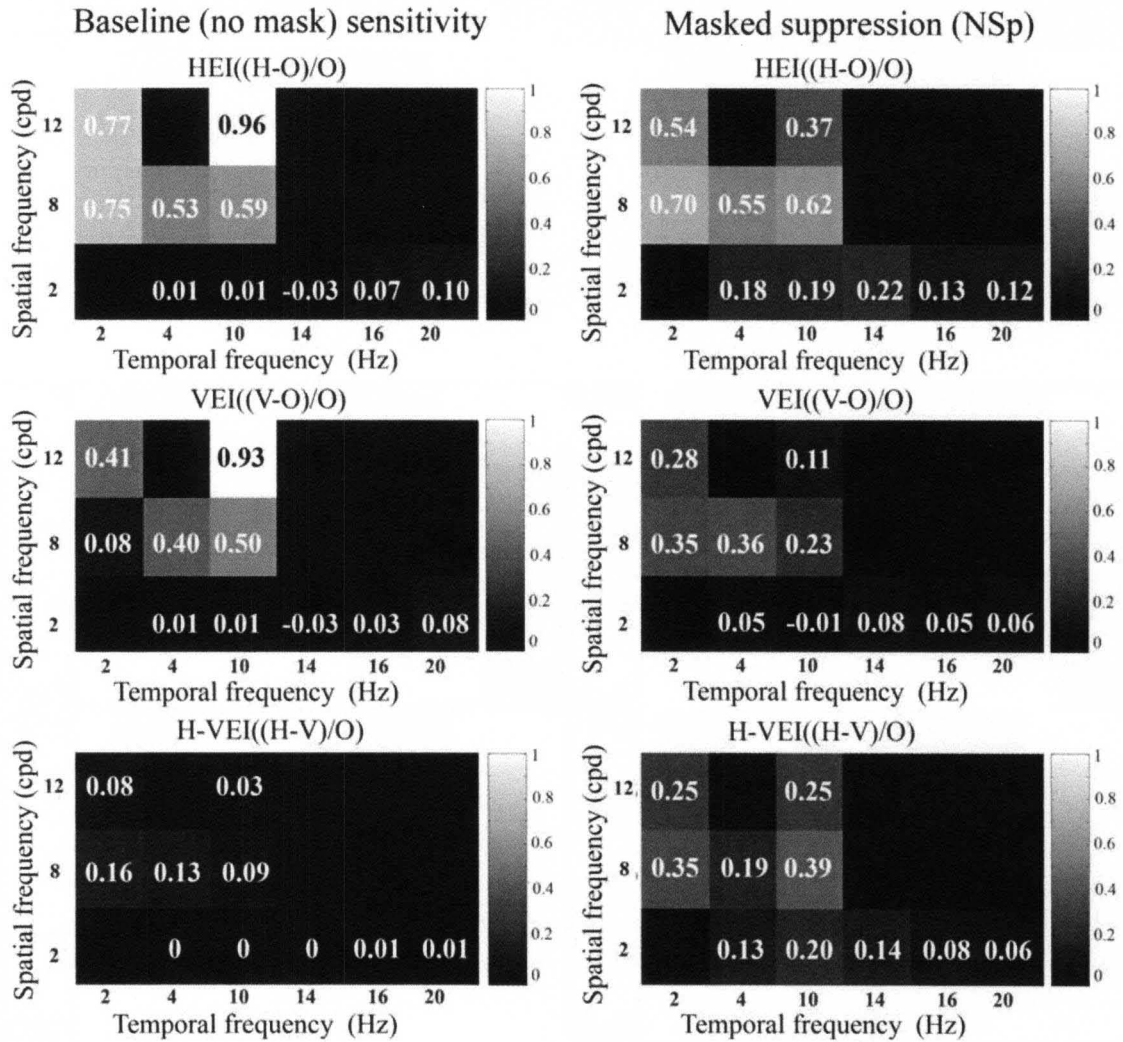
**Figure 13.** Gaussian curve fitting parameter values for each condition for the curves shown in **Figure 12A**. The tuning of masking with normalized suppression (NSp) values across the temporal frequency of the mask across 10 spatio-temporal conditions (Figure 12A) was assessed by fitting Gaussian functions with four parameters (i.e., bandwidth ( $W$ : Full width at Half Height (FWHH)), amplitude ( $A$ ), vertical shift constant ( $C$ ), and center position ( $x_0$ )) to the masking functions.



**Figure 14 .** Three Control conditions with two original conditions (condition 8 and condition 10) are shown. Normalized suppression (NSp) for comparing magnitude of the horizontal effect in temporal tuning curves is shown with Gaussian curve fitting parameter values in **Figure 14A**. Normalized suppression (NSb: each suppression value is divided separately for each orientation in each condition) for comparing bandwidth are shown with parameter values in **Figure 14B**.

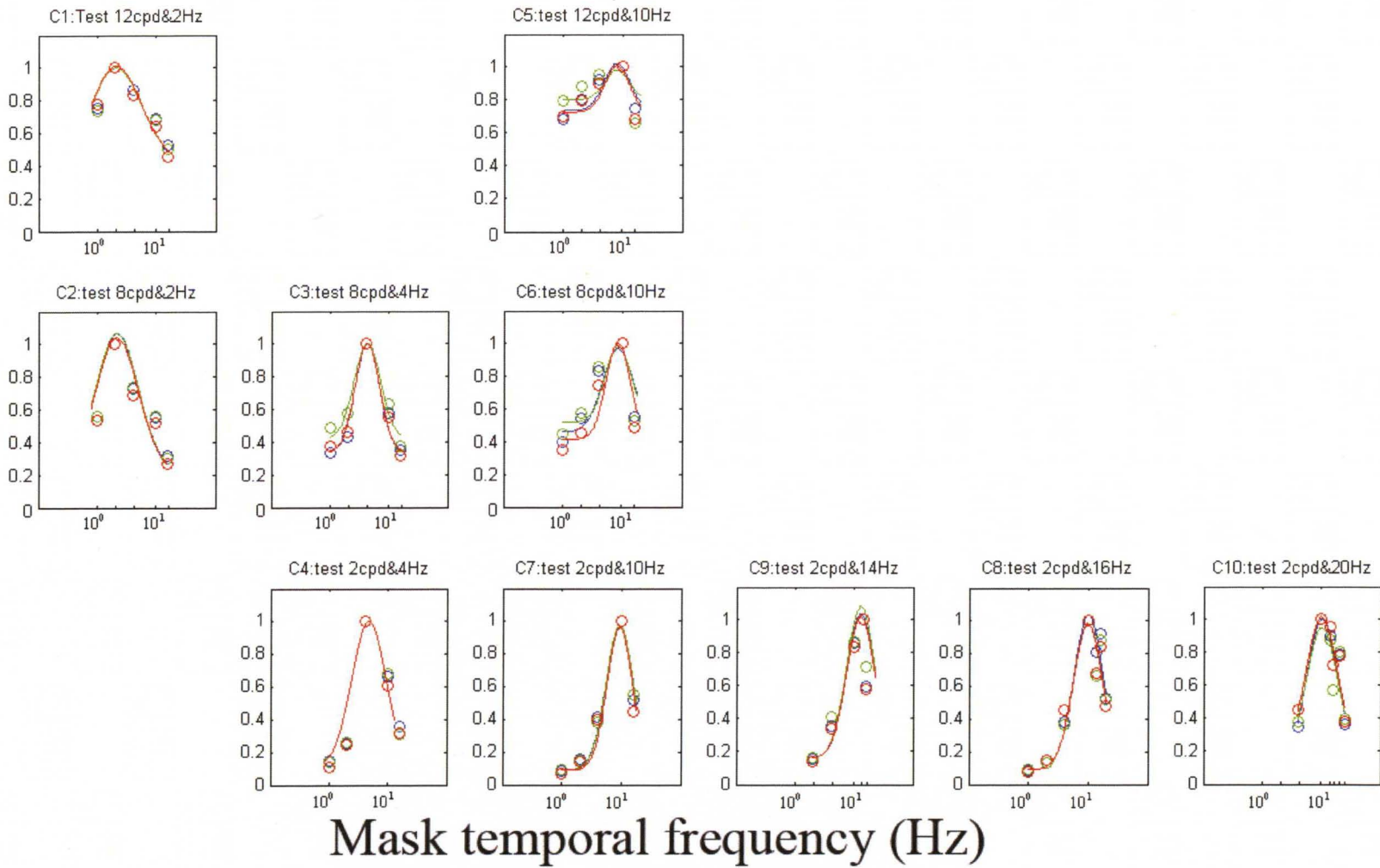


**Figure 15.** The tuning of masking with normalized suppression (NSp) values for comparing magnitude shown in **Figure 14A** is re-plotted on a linear scale in **(A)** and a log scale in **(B)**. As can be seen in **Figure 15A**, the log scale is a little compressed at the high frequency side of the mask function, and thereby the peak masking effect at high temporal frequency (the peak masking at 16 Hz) shown in Control 2 and Control 3, and condition 8 and condition 10 in **(A)** is easily disappeared in **(B)** in which the same data are re-plotting on log scale.

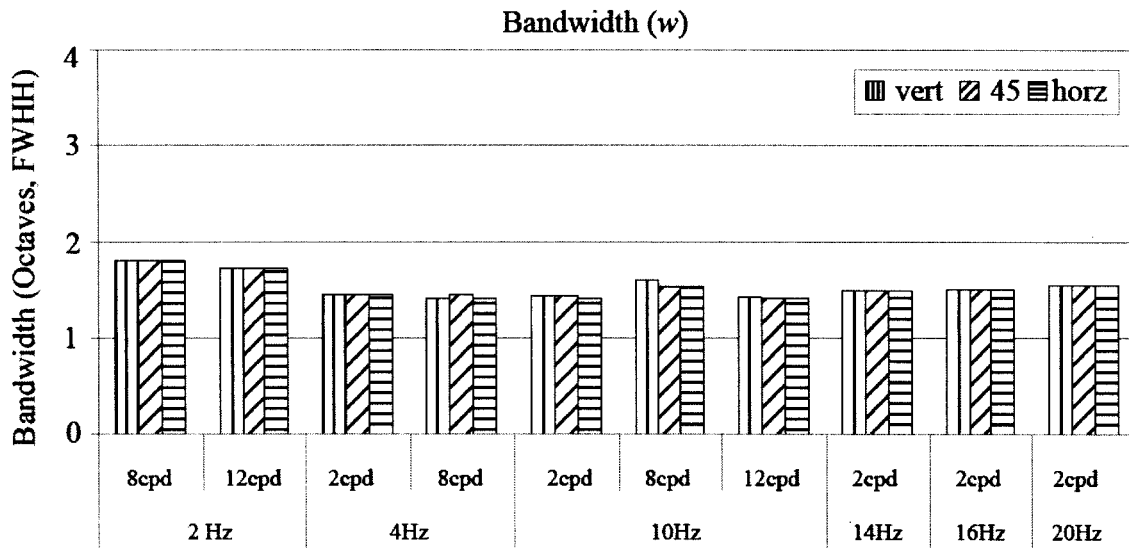


**Figure 16.** HEI, VEI, & H-VEI index across the 10 spatio-temporal locations with the baseline sensitivity (a left column) and with normalized suppression values (NSp) (a right column). A gray scale is used to indicate the magnitude of each index values across the 10 spatio-temporal locations.

# Normalized suppression



**Figure 17.** Normalized suppression (NSb: each suppression value is divided separately for each orientation in each condition) for comparing bandwidth in temporal frequency tuning curves across 10 spatio-temporal locations. Each tuning curve in each condition is plotted on the log scale. Best-fit Gaussian function shown (SSE average across conditions = 0.043). Test orientation: Horizontal (red circles), oblique (green circles), and vertical (blue circles).



**Figure 18.** Bandwidth estimations across different orientations and conditions. The bandwidth estimation of the Full Width at Half Height (FWHH in octaves) of the temporal frequency tuning functions across the 10 spatio-temporal target locations which is shown in **Figure 17** are plotted. As can be seen, the bandwidth varies little across orientation or condition; average bandwidth estimation was 1.52 octaves.

12cpd/2hz	VER	D45	HOR
W	1.71	1.71	1.71
A	0.58	0.57	0.58
C	0.43	0.43	0.43
x0	2.14	2.14	2.14

12cpd/10hz	VER	D45	HOR
W	1.42	1.39	1.39
A	0.29	0.18	0.29
C	0.73	0.80	0.72
x0	9.03	9.03	9.03

8cpd/2hz	VER	D45	HOR
W	1.80	1.80	1.80
A	0.85	0.85	0.82
C	0.12	0.12	0.12
x0	2.28	2.28	2.28

8cpd/4hz	VER	D45	HOR
W	1.40	1.44	1.40
A	0.67	0.58	0.66
C	0.34	0.42	0.34
x0	4.21	4.21	4.21

8cpd/10hz	VER	D45	HOR
W	1.59	1.52	1.52
A	0.40	0.29	0.64
C	0.33	0.30	0.33
x0	9.33	9.33	9.33

2cpd/4hz	VER	D45	HOR
W	1.45	1.45	1.45
A	0.89	0.89	0.89
C	0.11	0.11	0.11
x0	4.50	4.50	4.50

2cpd/10hz	VER	D45	HOR
W	1.43	1.43	1.39
A	0.84	0.86	0.85
C	0.13	0.13	0.12
x0	9.73	9.73	9.73

2cpd/16hz	VER	D45	HOR
W	1.49	1.49	1.49
A	0.91	0.85	0.83
C	0.12	0.12	0.13
x0	10.04	9.99	9.96

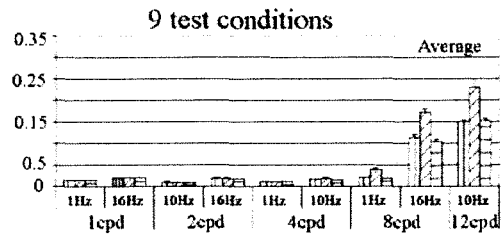
2cpd/14hz	VER	D45	HOR
W	1.48	1.48	1.48
A	0.88	0.93	0.87
C	0.00	0.00	0.00
x0	13.65	13.65	13.65

2cpd/20hz	VER	D45	HOR
W	1.54	1.54	1.54
A	0.90	0.79	0.84
C	0.11	0.16	0.18
x0	10.04	10.04	10.04

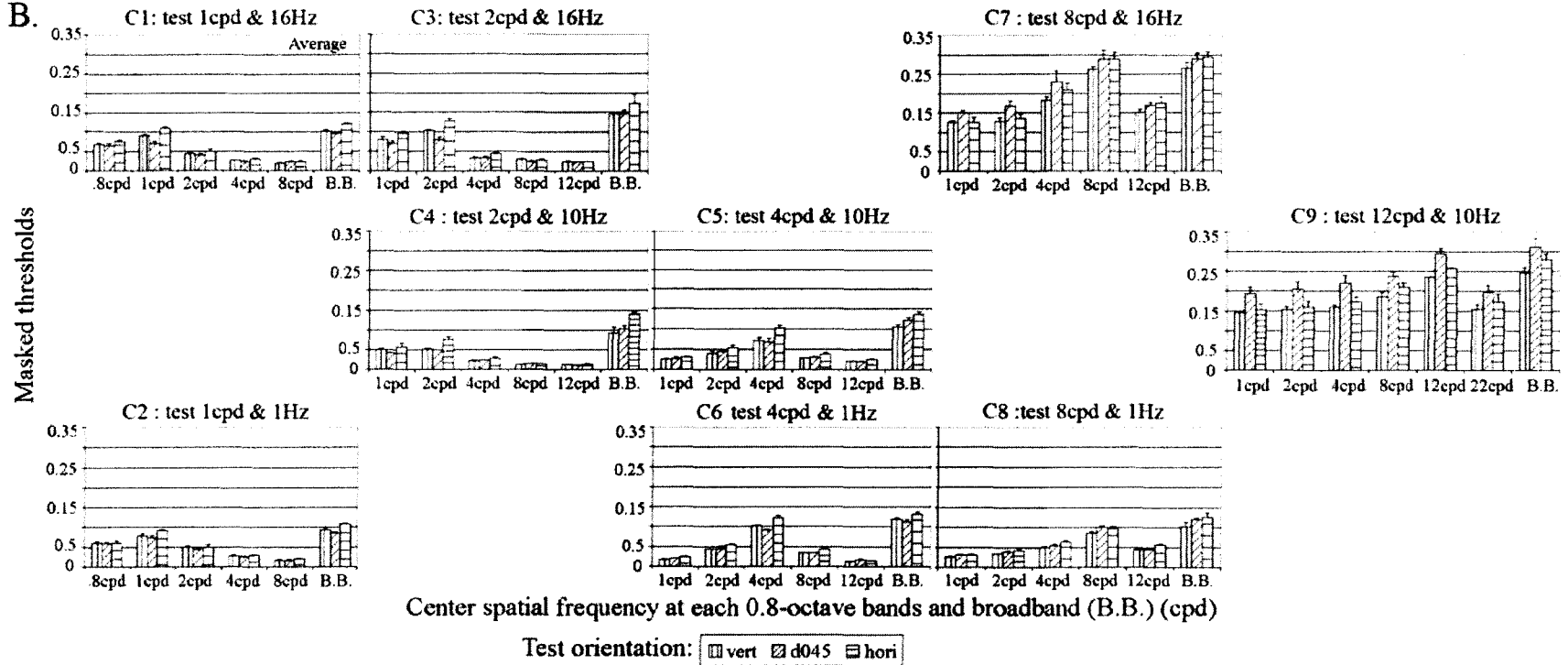
Figure 19. Gaussian curve fitting parameters for bandwidth shown in Figure 18 are provided.



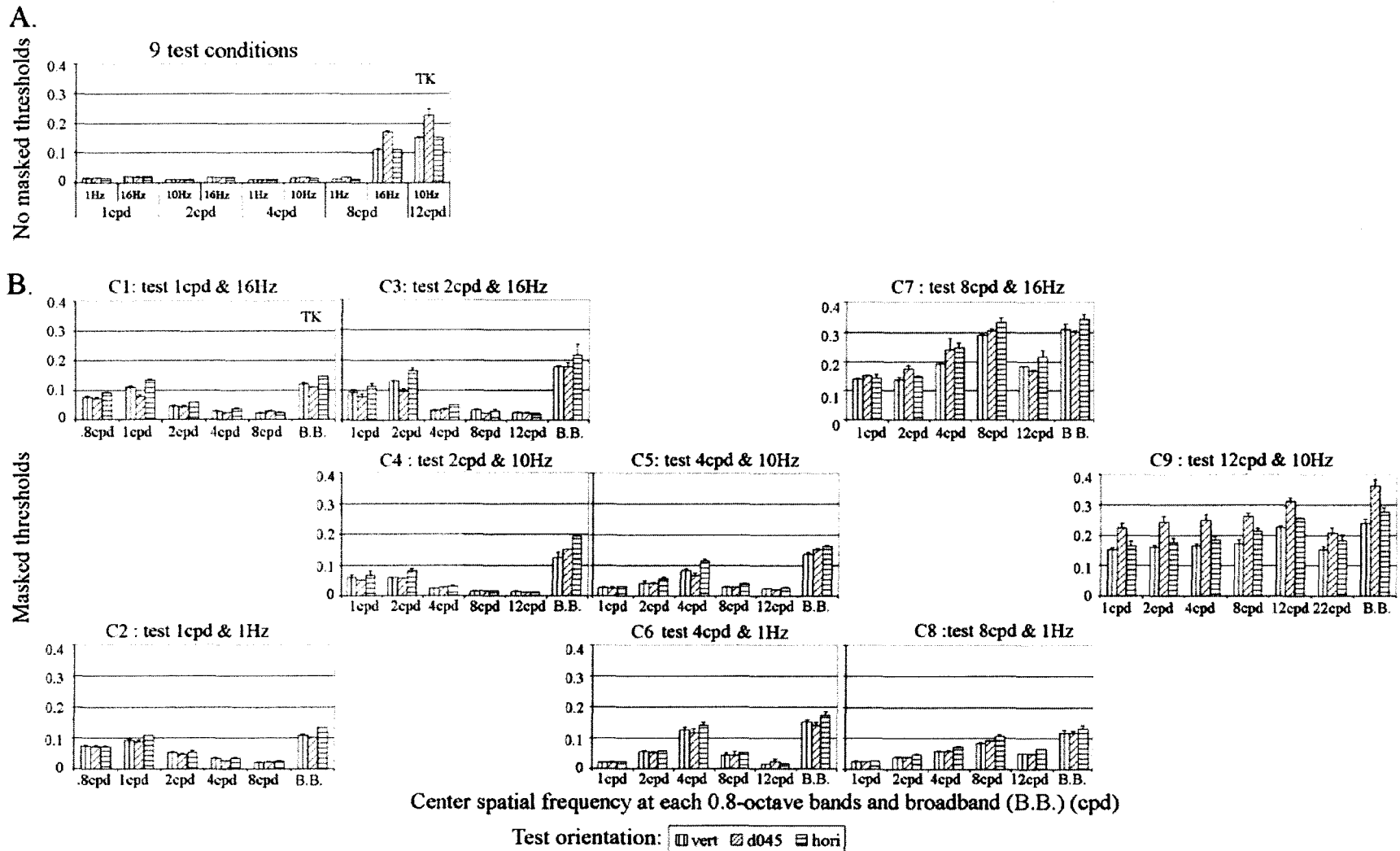
A.



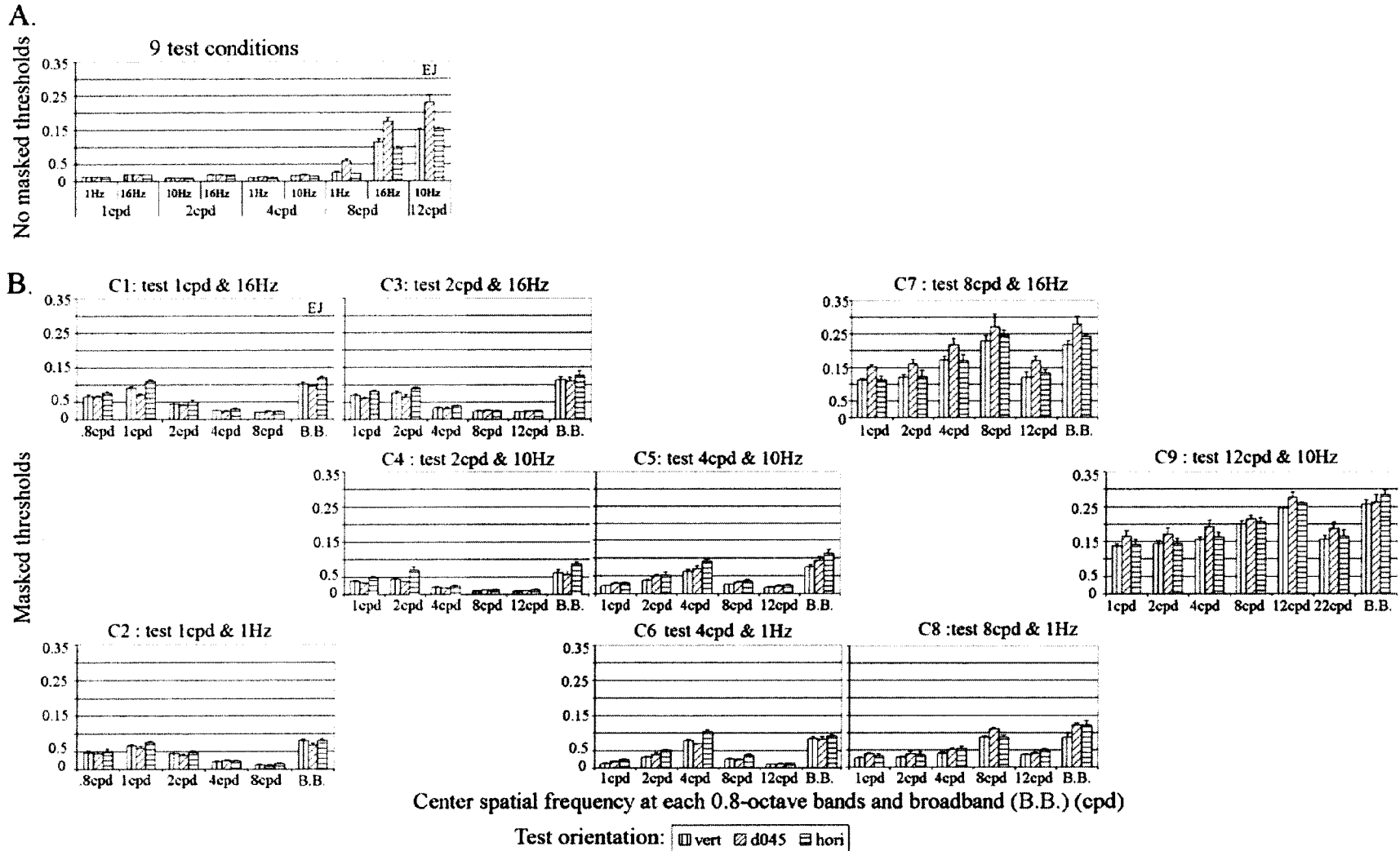
B.



**Figure 20.** Contrast thresholds for each grating target at each of three test orientations (vertical, 45° and horizontal) across the 9 spatial and temporal combinations without the mask (baseline) (**A**) and with spatially varied mask stimuli (**B**). To understand how the tuning along spatial dimension changes across the different target spatio-temporal combinations, the spatial frequency band of the mask stimuli is varied as a parameter. The broadband mask (4-octave) mask as used in Experiment 1 was split into five 0.8-octave sub-bands. The spatial frequency bandwidth of each sub-band centered at approximately 1, 2, 4, 8, and 12 cpd. Each sub-band of spatial noise (0.8-octave) and the broadband (4-octave) masks were used to probe spatial tuning for each target across different spatio-temporal surface. Additionally, one additional low spatial frequency sub-band, the 0.8 cpd center spatial frequency was added only for conditions 1 and 2. The other additional high spatial frequency sub-band, the 22 cpd center spatial frequency was added only for condition 9. Each mask noise pattern was modulated by a broadband temporal noise (4.5-octave). The average of two subjects is shown. Error bars are  $\pm$  standard error of the group mean.

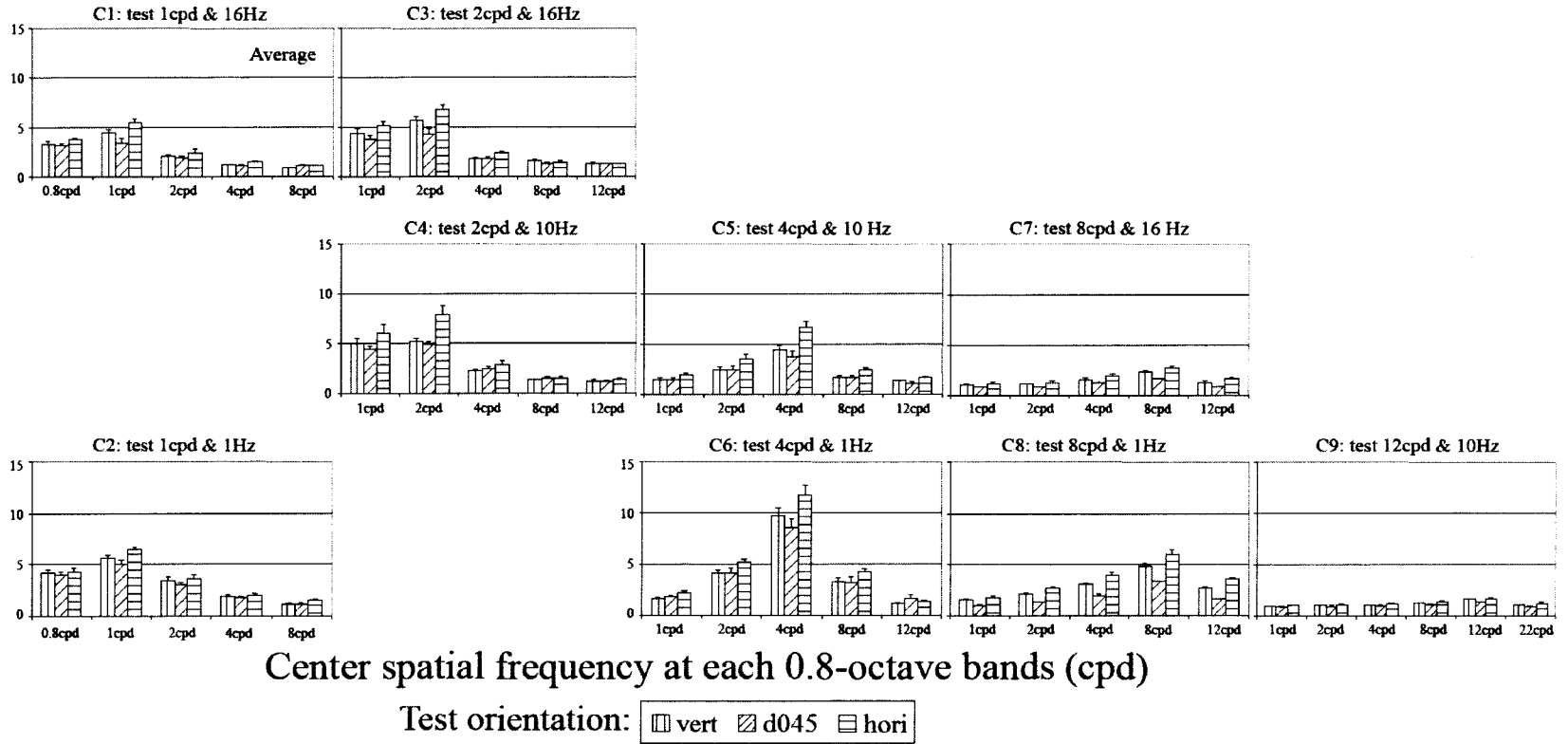


Same measurements as in **Figure 20** for a subject 'TK'



Same measurements as in **Figure 20** for a subject 'EJ'

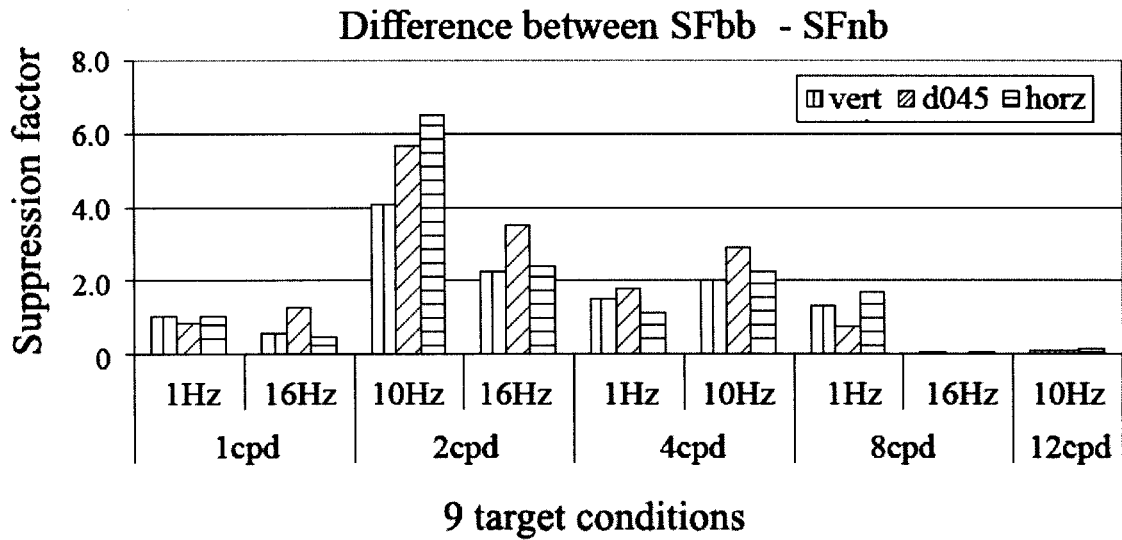
Suppression factor



Center spatial frequency at each 0.8-octave bands (cpd)

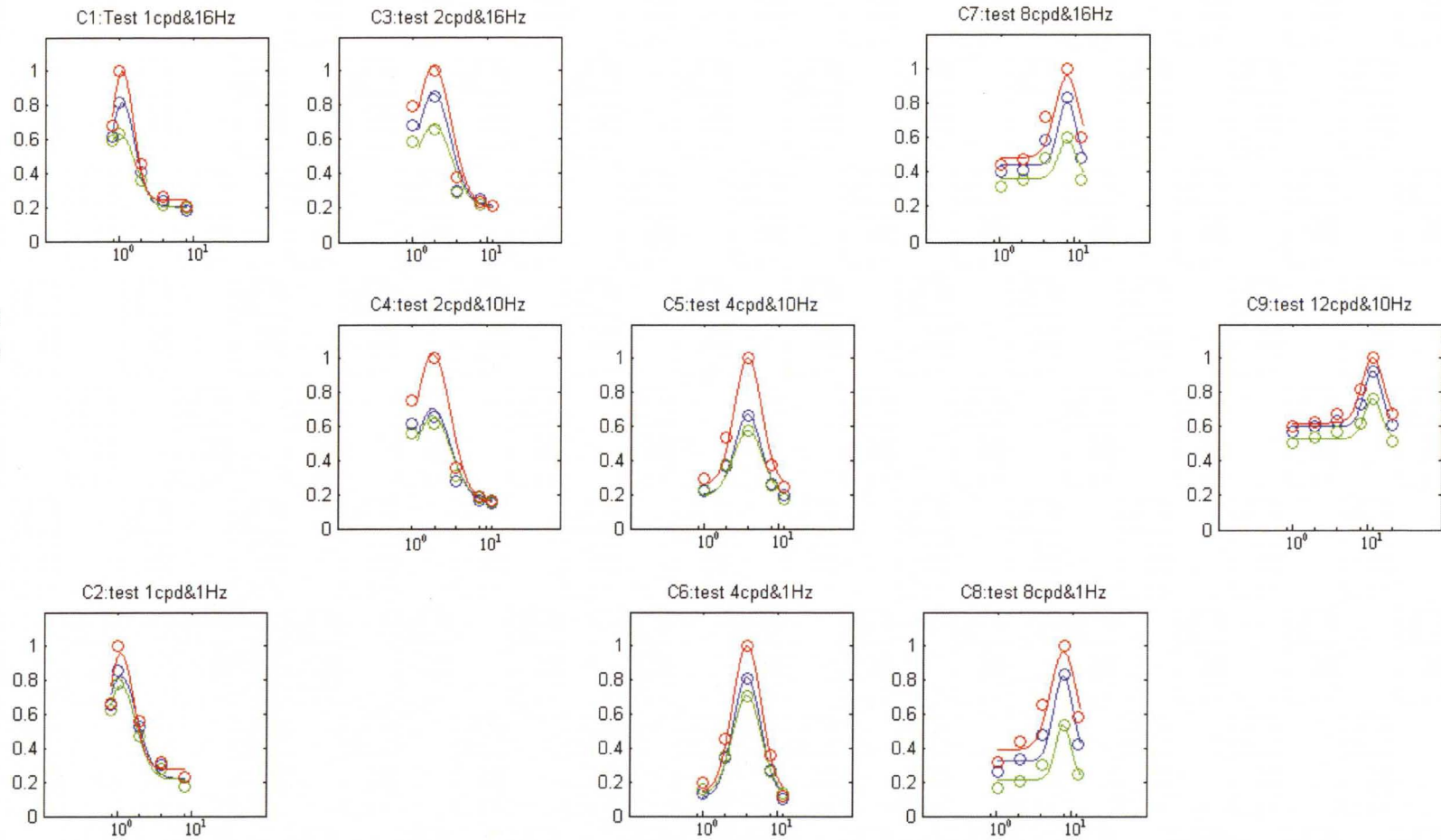
Test orientation: 
 vert
  d045
  hori

**Figure 21.** Suppression factor (the ration between masked and unmasked thresholds) of each target across three orientations (vertical, 45° and horizontal) and across the 9 spatio-temporal combinations is shown for mean thresholds (in **Figure 20**). As can be seen here, suppression is greatest when the mask matches target's spatial frequency. The standard errors were calculated in terms of Gaussian error propagation (Lo, 2005) and the mean of these standard errors are shown.



**Figure 22.** Difference between a broadband (bb) and a 0.8-octave band (nb: whose spatial frequency is centered at the target spatial frequency) masking across three orientation and across the 9 spatio-temporal combinations in suppression (shown in **Figure 21**). As can be seen, suppression obtained by naturalistic ( $1/f$ ) broadband mask is greatest for all of the conditions except two target conditions (i.e., 8 cpd/16 Hz and 12 cpd/10 Hz) in which no difference between them are shown.

Normalized suppression



Center spatial frequency at each 0.8-octave band (cpd)



**Figure 23.** Normalized suppression (NSp) for comparing magnitude of the horizontal effect anisotropy in spatial frequency tuning curves across 9 spatio-temporal locations. Each spatial tuning curve in each condition in each orientation is plotted on log scale. Best-fit Gaussian function shown (SSE averaging across conditions = 0.026). Test orientation: Horizontal (red circles), oblique (green circles), and vertical (blue circles). The masking effect is greatest when mask matches target's spatial frequency.

1cpd/16Hz	VER	D45	HOR
$W$	1.32	1.41	1.28
$A$	0.61	0.43	0.76
$C$	0.21	0.20	0.24
$x_0$	1	1	1

2cpd/16Hz	VER	D45	HOR
$W$	1.45	1.48	1.48
$A$	0.66	0.48	0.83
$C$	0.21	0.20	0.19
$x_0$	2	2	2

8cpd/16Hz	VER	D45	HOR
$W$	1.21	1.21	1.30
$A$	0.37	0.22	0.47
$C$	0.44	0.36	0.48
$x_0$	8	8	8

2cpd/10Hz	VER	D45	HOR
$W$	1.48	1.48	1.48
$A$	0.55	0.49	0.87
$C$	0.16	0.17	0.16
$x_0$	2	2	2

4cpd/10Hz	VER	D45	HOR
$W$	1.33	1.38	1.35
$A$	0.47	0.39	0.75
$C$	0.21	0.19	0.26
$x_0$	4	4	4

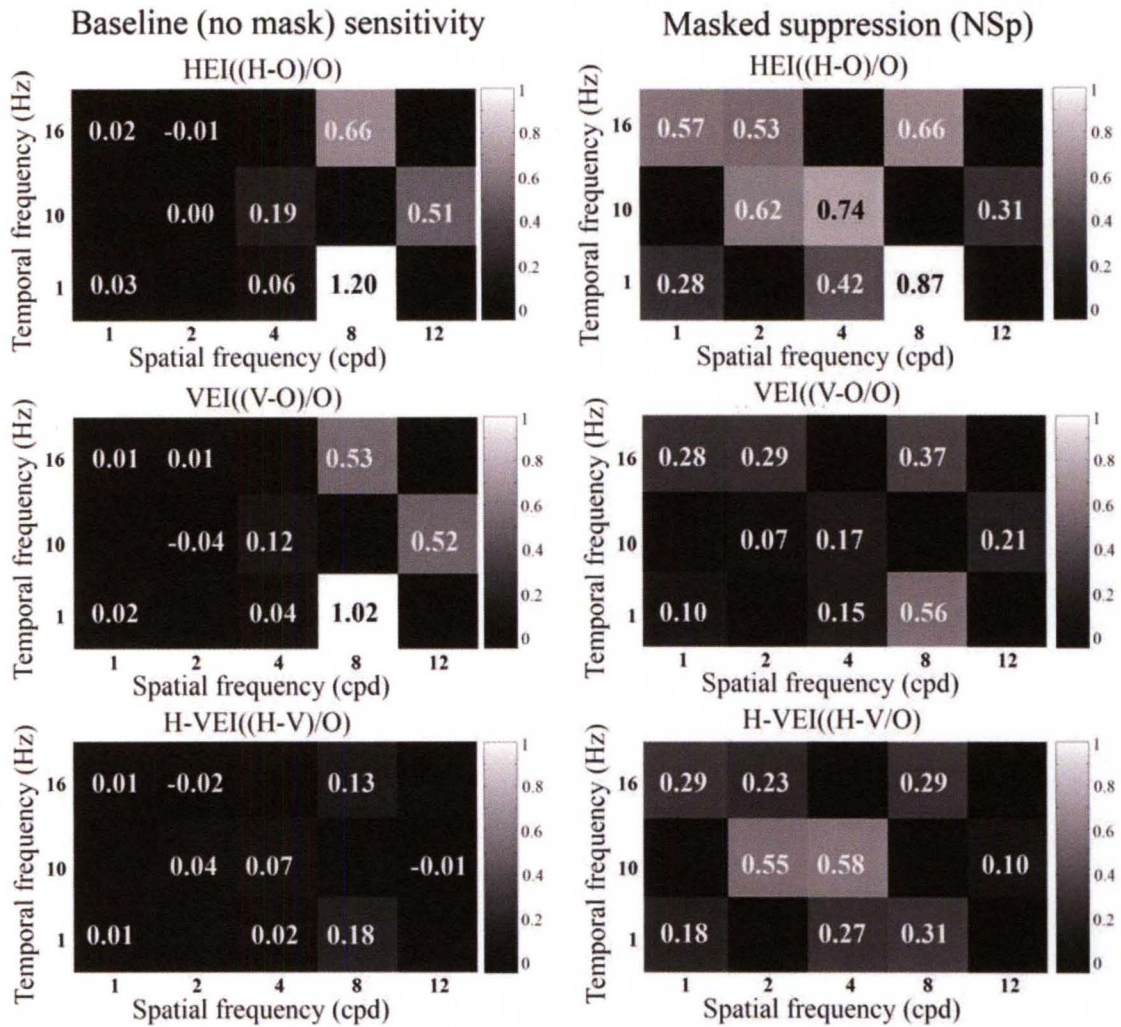
12cpd/10Hz	VER	D45	HOR
$W$	1.22	1.20	1.27
$A$	0.33	0.24	0.38
$C$	0.60	0.53	0.62
$x_0$	12	12	12

1cpd/1Hz	VER	D45	HOR
$W$	1.44	1.40	1.33
$A$	0.60	0.55	0.68
$C$	0.22	0.22	0.28
$x_0$	1	1	1

4cpd/1Hz	VER	D45	HOR
$W$	1.36	1.37	1.37
$A$	0.71	0.58	0.86
$C$	0.11	0.13	0.14
$x_0$	4	4	4

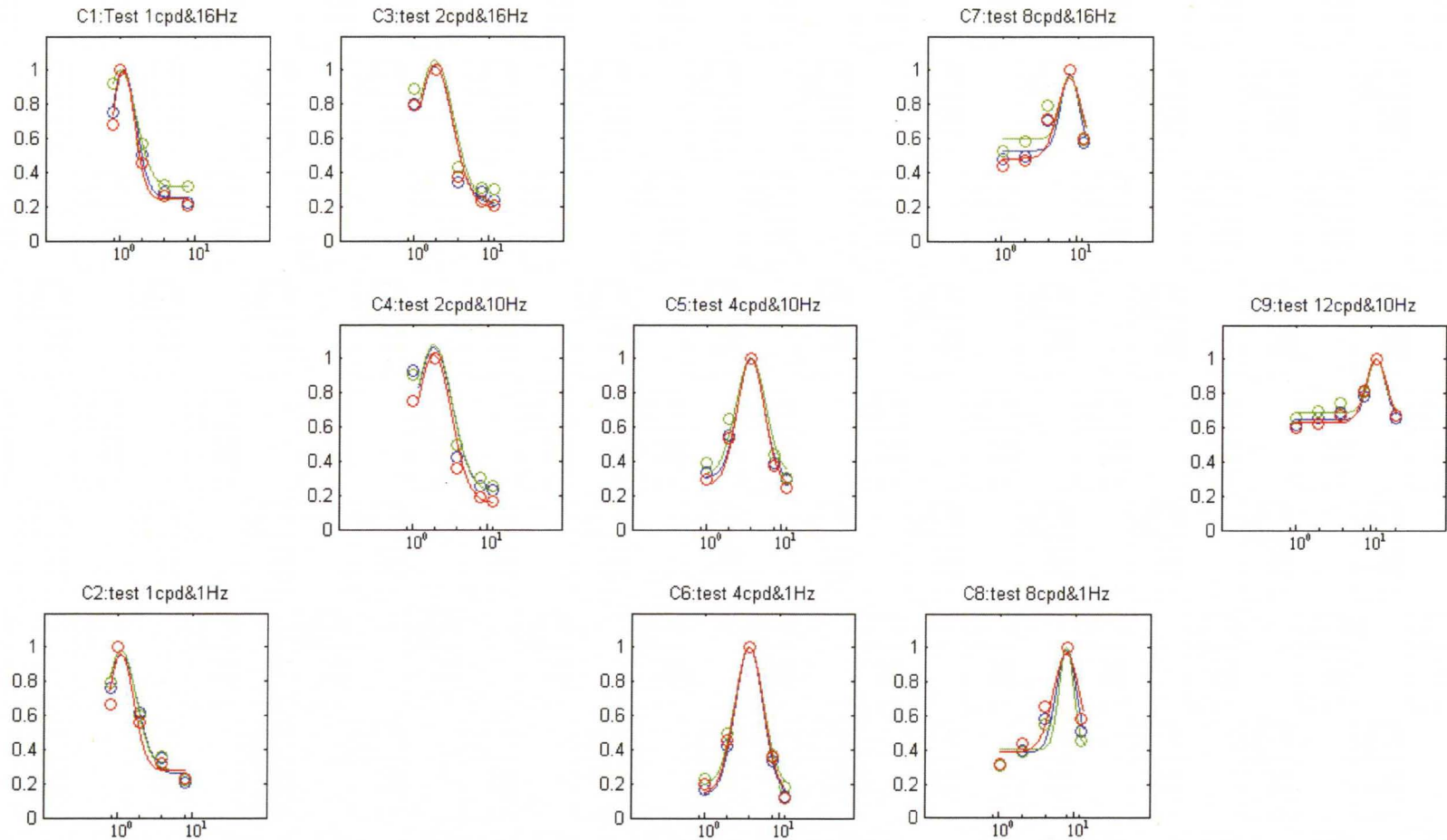
8cpd/1Hz	VER	D45	HOR
$W$	1.24	1.18	1.33
$A$	0.50	0.32	0.58
$C$	0.32	0.21	0.39
$x_0$	8	8	8

**Figure 24.** Gaussian curve fitting parameter values shown in **Figure 23**. The tuning of masking with normalized suppression (NSp) values across the spatial frequency of the mask across 10 spatio-temporal conditions (Figure 12A) was assessed by fitting Gaussian function with four parameters (i.e., bandwidth ( $W$ : Full width at Half Height (FWHH)), amplitude ( $A$ ), vertical shift constant ( $C$ ), and center position ( $x_0$ )) to the masking functions. A set of parameter values in each orientation within in each target condition are provided.



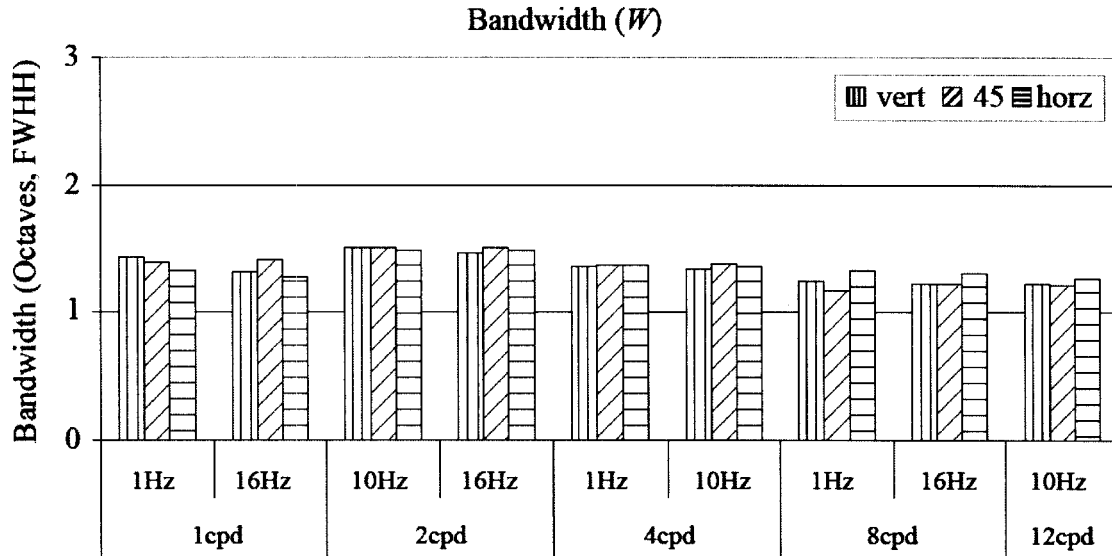
**Figure 25.** HEI, VEI, & H-VEI index across the 9 spatio-temporal locations with the baseline sensitivity (a left column) and with normalized suppression values (NSp) (a right column). A gray scale is used to indicate the magnitude of each index value across the 9 spatio-temporal locations.

Normalized suppression



Center spatial frequency at each 0.8-octave band (cpd)

**Figure 26.** Normalized suppression (NSb: each suppression value is divided separately for each orientation in each condition) for comparing bandwidth in temporal frequency tuning curves across 9 spatio-temporal locations. Each tuning curve in each condition is plotted on the log scale. Best-fit Gaussian function shown (SSE averaging across conditions = 0.026). Test orientation: Horizontal (red circles), oblique (green circles), and vertical (blue circles).



**Figure 27** Bandwidth estimations across orientations and conditions. The bandwidth estimation of the Full Width at Half Height (FWHH in octave) of the spatial frequency tuning functions across the 9 spatio-temporal target locations shown in **Figure 26** are plotted. As we can be seen, the bandwidth estimation little varies and the average bandwidth estimation was 1.35 octaves.

1cpd/16Hz	VER	D45	HOR
<i>W</i>	1.32	1.41	1.28
<i>A</i>	0.75	0.68	0.76
<i>C</i>	0.25	0.32	0.24
<i>x0</i>	1	1	1

2cpd/16Hz	VER	D45	HOR
<i>W</i>	1.46	1.51	1.49
<i>A</i>	0.80	0.84	0.83
<i>C</i>	0.22	0.22	0.19
<i>x0</i>	2	2	2

8cpd/16Hz	VER	D45	HOR
<i>W</i>	1.21	1.21	1.30
<i>A</i>	0.45	0.37	0.47
<i>C</i>	0.53	0.60	0.48
<i>x0</i>	8	8	8

2cpd/10Hz	VER	D45	HOR
<i>W</i>	1.51	1.51	1.48
<i>A</i>	0.85	0.86	0.87
<i>C</i>	0.22	0.22	0.16
<i>x0</i>	2	2	2

4cpd/10Hz	VER	D45	HOR
<i>W</i>	1.33	1.38	1.35
<i>A</i>	0.70	0.69	0.75
<i>C</i>	0.31	0.32	0.26
<i>x0</i>	4	4	4

12cpd/10Hz	VER	D45	HOR
<i>W</i>	1.22	1.20	1.26
<i>A</i>	0.35	0.31	0.37
<i>C</i>	0.65	0.69	0.63
<i>x0</i>	12	12	12

1cpd/1Hz	VER	D45	HOR
<i>W</i>	1.43	1.40	1.33
<i>A</i>	0.70	0.70	0.68
<i>C</i>	0.26	0.28	0.28
<i>x0</i>	1	1	1

4cpd/1Hz	VER	D45	HOR
<i>W</i>	1.36	1.37	1.37
<i>A</i>	0.87	0.82	0.86
<i>C</i>	0.13	0.19	0.14
<i>x0</i>	4	4	4

8cpd/1Hz	VER	D45	HOR
<i>W</i>	1.24	1.17	1.33
<i>A</i>	0.60	0.60	0.58
<i>C</i>	0.39	0.40	0.39
<i>x0</i>	8	8	8

Figure 28 Gaussian curve fitting parameters for Bandwidth shown in Figure 27 are provided.

## REFERENCES

- Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52, 1106-1130.
- Albrecht, D.G. & Geisler, W.S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neurosciences*, 7, 531-546.
- Adelson, E.H. & Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America-A*, 2, 284-29.
- Ahissar, E. & Arieli, A. (2001). Figuring space by time. *Neuron*, 32, 185-201.
- Anderson, S.J. & Burr, D.C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research*, 25, 1147-1154.
- Anderson, S.J. & Burr, D.C. & Morrone, M.C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of Optic Society of America-A*, 8, 1340-1351.
- Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The oblique effect in man and animals. *Psychological Bulletin*, 78, 266-278.
- Ascher, D. & Grzywacz, N.M. (2000). A Bayesian model of temporal frequency masking. *Vision Research*, 49, 2219-2232.
- Bartels, A., Zeki, S. & Logothetis, N.K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral Cortex*, 18, 705-717.



- Baker, T.I. & Issa, N.P. (2005). Cortical maps of separable tuning properties predict population responses to complex visual stimuli. *Journal of Neurophysiology*, 94, 775-787.
- Baker, C.L. Jr & Cynader, M.S. (1988). Space-time separability of direction selectivity in cat striate cortex neurons. *Vision Research*, 28, 239-246.
- Baker, C.L. Jr (1990). Spatial and temporal frequency selectivity as a basis for velocity preference in cat striate cortex. *Visual Neuroscience*, 4, 101, 113.
- Basole, A., White, L.E., & Fitzpatrick, D. (2003). Mapping multiple features in the population response of visual cortex. *Nature*, 424, 986-990.
- Barlow, H.B. (1961). Possible principles underlying the transformation of sensory messages. In: W.A. Rosenblith (Ed), *Sensory Communication* (pp.217-234), MIT Press.
- Barlow, H.B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12, 241-253.
- Ball, K. & Sekuler, R. (1982). Models of stimulus uncertainty in motion perception. *Psychological Review*, 87, 435-469.
- Blake, R. & Holopigian, K. (1985). Orientation selectivity in cats and human assessed by masking. *Vision Research*, 25, 1459-1468.
- Blakemore, C. & Campbell, F.W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 203, 237-260.
- Bex, P.J., Metha, A.B., & Makous, W. (1999). Psychophysical evidence for a functional hierarchy of motion processing mechanisms. *Journal of Optic Society*, 15, 769-

776.

Bex, P.J. & Dakin, S.C. (2002). Comparison of the spatial-frequency selectivity of local and global motion detectors. *Journal of Optic Society*, *19*, 670-677.

Bex, P.J., Simmers, A.J., & Dakin, S.C. (2003). Grouping local directional signals into moving contours. *Vision Research*, *43*, 2143-2153.

Bex, P.J., Dakin, S.C., Mareschal, I. (2005). Critical band masking in optic flow. *Network: Computation in Neural Systems*, *16*, 261-284.

Bex, P.J., Mareschal, I., & Dakin, S.C. (2007). Contrast gain control in natural scenes. *Journal of Vision*, *7*, 1-12.

Berkley, M.A., Kitterle, F. & Watkins, D.W. (1975). Grating visibility as a function of orientation and retinal eccentricity. *Vision Research*, *15*, 239-244.

Brady, J. & Field, D.J. (1995). Wh6.at's constant in contrast constancy? The effects of scaling on the perceived contrast of band-pass patterns. *Vision Research*, *35*, 739-7

Brady, N. & Field, D.J. (2000). Local contrast in natural images: normalization and coding efficiency. *Perception*, *29*, 1041-1055.

Breitmeyer, B.G. & Julesz, B. (1975). The role of on and off transients in determining the psychophysical spatial frequency response. *Vision Research*, *15*, 411-415.

Breitmeyer, B.G. & Ganz, L. (1977). Temporal studies with flashed gratings: inferences about human transient and sustained channels. *Vision Research*, *17*, 861-865.

Breitmeyer, B.G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, *15*, 111-1412.

Breitmeyer, B.G., Levi, D.M. & Harwerth, R.S. (1981). Flicker masking in spatial vision.

- Vision Research*, 21, 1377-1385.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.
- Burton, G.J. & Moorhead, I.R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26, 157-170.
- Burr, D.C. Morrone, C. & Maffei, L. (1981). Intracortical inhibition prevents simple cells from responding to textured patterns. *Experimental Brain Research*, 43, 455-458.
- Burbeck, C.A. & Kelly, D.H. (1981). Contrast gain measurements and the transient/sustained dichotomy. *Journal of the Optical Society of America*, 71, 1335-1342.
- Bonds, A.B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2, 41-55.
- Boynton, G.M. & Foley, J.M. (1999). Temporal sensitivity of human luminance pattern mechanism determined by masking with temporally modulated stimuli. *Vision Research*, 39, 1641-1656.
- Cass, J. & Alias, D. (2006). Evidence for two interacting temporal channels in human visual processing. *Vision Research*, 46, 2859-2868.
- Cass, J., Alias, D., Spehar, B., Bex, P.J. (2009). Temporal whitening: Transient noise perceptually equalizes the  $1/f$  temporal amplitude spectrum. *Journal of Vision*, 9, 1-19.
- Cass, J., Stuit, S., Bex, P.J., & Alais, D. (2009). Orientation bandwidths are invariant across spatiotemporal frequency after isotropic components are removed. *Journal of Vision*, 9, 1-14.
- Campbell, F.W. & Kulikowski, J.J. (1966). Orientation selectivity of the human visual

- system. *Journal of Physiology*, 187, 437-445.
- Campbell, F.W. & Maffei, T.N. (1970). Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *The Journal of Physiology*, 207, 635-652.
- Campbell, F.W. and Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197, 551-566.
- Campbell, F.W., Cooper, G.F. & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *The Journal of Physiology*, 203, 223-235.
- Campbell, F.W., Carpenter, R.H.S. & Levinson, J.Z. (1969). Visibility of aperiodic patterns compared with that of sinusoidal gratings. *The Journal of Physiology*, 204, 283-298.
- Camisa, J.M., Blake, R., & Lema, S. (1977). The effects of temporal modulation on the oblique effect in human. *Perception*, 6, 165-171.
- Carandini, M. & Heeger, D.J. (1994). Summation and division by neurons in visual cortex. *Science*, 264, 1333-1336.
- Chandler, D.M. & Field, D.J. (2007). Estimates of the information content and dimensionality of natural scenes from proximity distributions. *Journal of Optical Society of America-A*, 24, 922-941.
- Chirimuuta, M. & Tolhurst, D.J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, 45, 2943-2959.
- Chen, C-C. & Foley, J.M. (2004). Pattern detection: Interactions between oriented and concentric patterns. *Vision Research*, 44, 915-924.

- Chen, C-C. & Tyler, C.W. (2008). Excitatory and inhibitory interaction fields of flankers revealed by contrast-masking function. *Journal of Vision*, 8, 1-14.
- Coppola, D.M., Purves, H.R., McCoy, A.N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 4002-4006.
- Dakin, S.C., Mareschal, I. & Bex, P.J. (2005). An oblique effect for local motion: Psychophysics and natural movie statistics. *Journal of Vision*, 5, 878-887.
- Dakin, S.C. (2009). Vision: Thinking Globally, Acting Locally. *Current Biology*, 19, 851-854.
- Dakin, S.C., Athorp, D. & Alais, D. (2010). Anisotropies in judging the direction of moving natural scenes. *Journal of Vision*, 10, 1-19.
- David, S.V., Vinje, W.E., & Gallant, J.L. (2004). Natural stimulus statistics alter the receptive field structure of V1 neurons. *Journal of Neuroscience*, 24, 6991-7006.
- DeValois, R.L., Yund, E.W., & Helper, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22, 531-544.
- DeValois, R.L., Albrecht, D.G., and Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545-559.
- DeAngelis, G.C., Ohzawa, I., & Freeman, R.D. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, 69, 144-163.
- DeAngelis, G.C., Ohzawa, I., & Freeman, R.D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69, 1091-1117.

- DeAngelis, G.C., Ghose, G.M., Ohzawa, I. & Freeman, R.D. (1999). Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *Journal of Neurophysiology*, *19*, 4046-4064.
- Dong, D.W. & Atick, J. (1995). Statistics of time-varying natural images. *Network*, *6*, 345-358.
- Daugman, J.G. (1984). Spatial visual channels in the Fourier plane. *Vision Research*, *24*, 891-910.
- Eckert, M.P. & Buchsbaum, G. (1993). Efficient coding of natural time varying images in the early visual system. *Philosophical Transactions: Biological Sciences*, *339*, 385-395.
- Essock, E.A. (1980). The oblique effect of stimulus identification considered with respect to two classes of oblique effects. *Perception*, *9*, 37-46.
- Essock, E.A. (1982). Anisotropies of perceived contrast and detection speed. *Vision Research*, *22*, 1185-1191.
- Essock, E.A., DeFord, J.K., Hansen, B.C., & Sinai, M.J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research*, *43*, 1329-1335.
- Essock, E.A., Haun, A.M., & Kim, Y.J. (2009). An anisotropy of orientation-tuned suppression matches the anisotropy of typical natural scenes. *Journal of Vision*, *9*, 1-15.
- Enroth-Cugell, C. & Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, *187*, 517-552.
- Ferrara, V.P. and Wilson, H.R. (1985). Influence of the contrast sensitivity function on

- the reaction time. *Vision Research*, 33, 2461-2466.
- Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America-A*, 4, 2379-2394.
- Field, D.J., Hayes, A., & Hess, R.F. (1993). Contour integration by the human visual system: Evidence for a local “association field”. *Vision Research*, 33, 173-193.
- Field, D.J. (1993). Scale-invariance and self-similar ‘wavelet’ transforms: an analysis of natural scenes and mammalian visual systems. In: Farge, M. Hunt, J. and Vassilicos (Eds.). *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*. Oxford: Oxford University Press.
- Field, D.J. & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37, 3367-3383.
- Foley, J.M. (1994). Human luminance pattern-vision mechanisms-masking experiments require a new model. *Journal of Optical Society of America-A*, 11, 1710-1719.
- Foley, J.M., & Legge, G.E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, 21, 1041-1053.
- Foster, K.H., Gastra, J.P., Nagler, M. & Pollen, D.A. (1983). Spatial and temporal frequency selectivity of neurons in V1 and V2 of the macaque monkey. (Abstract) *Neuroscience*, 9, 618.
- Foster, K.H., Gastra, J.P., Nagler, M. & Pollen, D.A. (1985). Spatial and temporal frequency selectivity of neurons in visual cortical areas V1 and V2 of the macaque monkey. *The Journal of Physiology*, 365, 331-363.
- Fredericksen, R.E. & Hess, R.F. (1998). Estimation multiple temporal mechanisms in

- human vision. *Vision Research*, 38, 1023-1040.
- Furmanski, C.S. & Engel, S.A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, 3, 535-536.
- Furmanski, C.S., Schuppeck, D. & Engel, S.A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, 14, 573-578.
- Gallant, J.L., Connor, C.E., & Van Essen, D.C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, 9, 2153-2158.
- Geisler, W.S. & Albrecht, D.G. (1992). Cortical neurons: Isolation of contrast-gain control. *Vision Research*, 32, 1409-1410.
- Geisler, W.S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59, 167-192.
- Georgeson, M.A. & Sullivan, G.D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *The Journal of Physiology*, 252, 627-656.
- Georgeson, M.A. (1987). Temporal properties of spatial contrast vision. *Vision Research*, 27, 765-780.
- Glass, L. (1969). "Moire effects from random dots". *Nature (London)*, 243, 578-580.
- Green, D.M. (1984). Masking by light and the sustained-transient dichotomy. *Perception & Psychophysics*, 35, 519-535.
- Graham, N. (1989). *Visual Pattern Analyzers*. New York: Oxford University Press.
- Graham, N. & Sutter, A. (1996). Effect of spatial scale and background luminance on the intensive and spatial nonlinearities in texture segregation. *Vision Research*, 36, 1371-1390.



- Grzywacz, N.W. & Yulle, A.I. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London-B*, 239, 129-161.
- Grossberg, S. & Mingolla, E. (1993). Neural dynamics of motion perception: direction fields, apertures, and resonant grouping. *Perception and Psychophysics*, 53, 243-278.
- Gorea, A. & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neurosciences*, 14, 897-919.
- Gros, B.L., Blake, R., & Hiris, E. (1998). Anisotropies in visual motion perception: A fresh look. *Journal of the Optical Society of America, A, Optics Image Science, and Vision*, 15, 2003-2011.
- Hammett, S.T. & Snowden, R.J. (1995). The effect of contrast adaptation on briefly presented stimuli. *Vision Research*, 35, 1721-1725.
- Harwerth, R.S. & Levi, D.M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision Research*, 18, 1579-1586.
- Hansen, B.C. & Essock, E.A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, 4, 1044-060.
- Hansen, B.C. & Essock, E.A. (2005). Influence of scale and orientation on the visual perception of natural scenes. *Visual Cognition*, 12, 1199-1234.
- Hansen, B.C. & Essock, E.A. (2006). Anisotropic local contrast normalization: The role of stimulus orientation and spatial frequency bandwidths in the oblique and horizontal effect perceptual anisotropies. *Vision Research*, 46, 4398-4415.

- Haun, A.M. & Essock, E.A. (2010). Contrast sensitivity for oriented patterns in  $1/f$  noise: Contrast response and the horizontal effect. *Journal of Vision*, *10*, 1-21.
- Harvey, L.O. & Doan, V.Y. (1990). Visual masking at different polar angles in the two-dimensional Fourier plane. *Journal of Optical Society of America-A*, *7*, 116-126.
- Hess, R.F. & Plant, G.T. (1985). Temporal frequency discrimination in human vision: Evidence for an additional mechanism in the low spatial and high temporal frequency region. *Vision Research*, *25*, 1493-1500.
- Hess, R.F. & Snowden, R.J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Research*, *32*, 47-59.
- Hess, R.F. & Dakin, S.C. (1997). Absence of contour linking in peripheral vision. *Nature*, *390*, 602-604.
- Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181-197.
- Hubel, D.H., & Wiesel, T.N. (1959). Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, *148*, 574-591.
- Hubel, D.H., & Wiesel, T.N. (1962). Receptive field, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106-154.
- Hubel, D.H., & Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*, 215-243.
- Hubel, D.H., & Wiesel, T.N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *198*, 1-59.
- Holub, R.A. & Morton-Gibson, M. (1981). Response of visual cortical neuron of the cat

- to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. *Journal of Neurophysiology*, *46*, 1244-1259.
- Holiday, I.E. & Rudock, K.H. (1983). Two spatio-temporal filters in human vision. I. Temporal and spatial frequency response characteristics. *Biological Cybernetics*, *47*, 173-190.
- Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *The Journal of Neuroscience*, *24*, 3313-3324.
- Itti, L., Koch, C., & Braun, J. (2000). Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America-A*, *17*, 1899-1917.
- Issa, N.P., Rosenberg, A., & Husson, T.R. (2008). Models and measurements of functional maps in V1. *Journal of Neurophysiology*, *99*, 2745-2754.
- Kane, D. Bex, P.J. & Dakin, S.C. (2009). The aperture problem in contoured stimuli. *Journal of Vision*, *9*, 1-17.
- Kane, D. Bex, P.J. & Dakin, S.C. (2011). Quantifying “the aperture problem” for judgments of motion direction in natural scenes. *Journal of Vision*, *11*, 1-20.
- Kelley, D.H. (1972). Adaptation effects on spatio-temporal sine-wave thresholds. *Vision Research*, *12*, 89-101.
- Kelly, D.H. (1979). Motion and vision. I. Stabilized images of stationary gratings. *Journal of Optical Society of America*, *69*, 1266-1274.
- Keeseey, U.T. (1972). Flicker and pattern detection: A comparison of threshold. *Journal of the Optical Society of America*, *62*, 446-448.
- Kim, Y.J., Haun, A.M., & Essock, E.A. (2010). The horizontal effect in suppression:

- Anisotropic overlay and surround suppression at high and low speeds. *Vision Research*, 50, 838-849.
- Kim, Y.J., Haun, A.M., & Essock, E.A. (2011). Time-course of anisotropic masking at high and low spatial frequencies. (Abstract) *Vision Sciences Society*, in press.
- Kim, Y.J. & Essock, E.A. (2011). Anisotropic gain control pools are tuned in temporal frequency as well as spatial frequency and orientation. (Abstract) *Journal of Vision*, 9, in press.
- Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37-68.
- Kulikowski, J.J. & Tolhurst, D.J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *The Journal of Physiology*, 232, 149-162.
- Kulikowski, J.J. (1971). Some stimulus parameters affecting spatial and temporal resolution of human vision. *Vision Research*, 11, 83-93.
- Kontsevich, L.L., Chen, C-C., & Tyler, C.W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research*, 42, 1771-1784.
- Koenderink, J.J. & van Doorn, A.J. (1979). Spatiotemporal contrast detection threshold surface is bimodal. *Optics Letters*, 4, 845-849.
- Langely, K. & Bex, P.J. (2007). Contrast adaptation implies two spatiotemoral channels but three adaptating processes.
- Legge, G.E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Research*, 18, 69-81.
- Lehky, S.R. (1985). Temporal properties of visual channels measured by masking. *Journal of Optical Society of America-A*, 2, 1260-1272.

- Li, B.W., Peterson, M.R. & Freeman, R.D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology*, 90, 204-217.
- Logothetis, N.K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5, 552-563.
- Lorenceanu, J. & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, 32, 263-273.
- Lo, E. (2005). Gaussian error propagation applied to ecological data: post-ice-storm-down woody biomass. *Ecological Monographs*, 75, 451-466.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Mandler, J.M. & Makous, K.T. (1984). A three channel model of temporal frequency perception. *Vision Research*, 26, 609-619.
- Mansfield, R.J.W. (1974). Neural basis of orientation perception in primate vision. *Science*, 186, 1133-1135.
- Maffei, R.J.W. & Campbell, F.W. (1970). Neurophysiological localization of the vertical and horizontal visual coordinates in man. *Science*, 167, 386-38.
- Maffei, L. & Fiorentini, A. (1977). Spatial frequency rows in the striate visual cortex. *Vision Research*, 17, 257-264.
- Mannion, D.J., McDonald, J.S., & Clifford, C.W.G. (2010). The influence of global form on local orientation anisotropies in human visual cortex. *Neuroimage*, 52, 600-605.
- Mansfield, R.J.W. & Ronner, S.P. (1978). Orientation anisotropy in monkey visual cortex. *Brain Research*, 149, 229-234.
- Mante, V. & Carandini, M. (2005). Mapping of stimulus energy in primary visual cortex.

*Journal of Neurophysiology*, 94, 788-798.

Martinez-Conde, S., Macknik, S.L., Troncoso, X.G., & Dyar, T.A. (2006).

Microsaccades counteract fading during fixation. *Neuron*, 49, 297-305.

Meier, L. & Carandini, M. (2002). Masking by fast gratings. *Journal of Vision*, 2, 293-301.

Meese, T.S. and Holmes, D.J. (2003). Orientation masking: Suppression and orientation bandwidth. (Abstract) *Perception*, 32, 388.

Meese, T.S. and Hess, R.F. (2004). Low spatial frequencies are suppressively masked across spatial scale, orientation, field position and eye of origin. *Journal of Vision*, 4(10), 2, 843-859.

Meese, T.S. & Holmes, D.J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceeding of the Royal Society B: Biological Sciences*, 274, 127-136.

Meese, T.S. & Baker, D.H. (2009). Cross-orientation masking is speed invariant between ocular pathways but speed dependent within them. *Journal of Vision*, 9, 1-15.

Meese, T.S., Summers, R.J., Holmes, D.J., & Wallis, S.A. (2007). Contextual modulation involves suppression and facilitation from the center and the surround. *Journal of Vision*, 7, 1-21.

Meese, T.S. & Holmes, D.J. (2010). Orientation masking and cross-orientation suppression (XOS): Implications for estimates of filter bandwidth. *Journal of Vision*, 10, 1-20.

Meso, A.I. & Hess, R.F. (2010). Visual motion gradient sensitivity shows scale invariant spatial frequency and speed tuning properties. *Vision Research*, 50, 1475-1485.

- Mitov, D. & Totev, T. (2005). How many pathways determine the speed of grating detection? *Vision Research*, 45, 821-825.
- Movshon, J.A. & Blakemore, C. (1973). Orientation specificity and spatial selectivity in human vision. *Perception*, 2, 53-60.
- Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology London*, 283, 53-77.
- Murray, I., MacCana, F. & Kulikowski, J.J. (1983). Contribution of two movement detecting mechanisms to central and peripheral vision. *Vision Research*, 23, 151-159.
- Nachmias, J. (1960). Meridional variations in visual acuity and eye movements during fixation. *Journal of Optical Society America*, 50, 569-571.
- Nachmias, J. (1967). Effect of exposure duration on visual contrast sensitivity with square-wave gratings. *Journal of Optical Society of America*, 57, 421-427.
- Nachmias, J. & Sansbury, R.V. (1974). Contrast-discrimination may be better than detection. *Vision Research*, 14, 1039-1042.
- Nishida, S. Ledgeway, T. & Edwards, M. (1997). Dural multiple-scale processing for motion in the human visual system. *Vision Research*, 37, 2685-2698.
- Olshausen, B.A. & Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural image. *Natural*, 381, 607-609.
- Olshausen, B.A. & Field, D.J. (2005). How close are we to understanding V1? *Neural Computation*, 17, 1665-1699.
- Olzak, L.A. & Thomas, J.P. (1999). Neural recoding in human pattern vision: model and

- mechanisms. *Vision Research*, 39, 3983-3987.
- Orban, G.A. & Kennedy, H. (1980). Evidence for meridional anisotropies in orientation selectivity of visual cortical neurons. *Archives Internationales de Physiologie et de Biochimie*, 88, 13-14.
- Oliva, A. & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145-175.
- Palmer, S.M. & Rosa, M.G.P. (2006). A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. *European Journal of Neuroscience*, 24, 2389-2405.
- Pasupathy, A. (2006). Neural basis of shape representation in the primate brain. *Progress in Brain Research*, 154, 293-313.
- Pack, C.C. & Born, R. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409, 1040-1042.
- Pack, C.C., Conway, B.R., Born, R.T., & Livingstone, M.S. (2006). Spatiotemporal structure of nonlinear subunits in macaque visual cortex. *The Journal of Neuroscience*, 26, 893-907.
- Pack, C.C. & Born, R.T. (2008). Cortical mechanisms for the integration of visual motion. In: A. I. Basbaum, A. Kaneko., G. M. Shepherd & G. Westheimer (Eds.), *The Senses: A comprehensive references* (pp. 189-218). San Diego: Academic Press.
- Pantle, A. (1983). Temporal determinants of spatial sine-wave masking. *Vision Research*, 23, 749-757.
- Petrov, Y., Carandini, M. & McKee, S.P. (2005). Two distinct mechanisms of



- suppression in human vision. *Journal of Neuroscience*, 25, 8704-8707.
- Petrov, Y., Carandini, M. & McKee, S.P. (2006). The time course of contrast masking reveals two distinct mechanisms of human surround suppression. (Abstract) *Journal of Vision*, 6, 808a.
- Petrov, Y. & McKee, S.P. (2009). The effect of spatial configuration on surround suppression of contrast sensitivity. *Journal of Vision*, 6, 224-238.
- Perrone, J.A. (2004). A visual motion sensor based on the properties of V1 and MT neurons. *Vision Research*, 44, 1733-1755.
- Price, N.S.C. & Born, R.T. (2009). Representation of movement. *Encyclopedia of Neuroscience*, 8, 107-114.
- Phillips, G.C. & Wilson, H.R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of Optical Society of America-A*, 1, 226-232.
- Ringach, D.L., Hawken, M.J., & Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2, 12-24.
- Robson, J.G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of Optical Society of America*, 56, 1141-1142.
- Ross, J., Speed, H.D. & Morgan, M.J. (1993). The effects of adaptation and masking on incremental thresholds for contrast. *Vision Research*, 33, 2051-2056.
- Rucci, M. & Casile, A. (2005). Fixational instability and natural image statistics: implications for early visual representations. *Network: Computation in Neural Systems*, 16, 121-138.
- Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance

- fine spatial detail. *Nature*, 447, 851-854.
- Ruderman, D. (1997). Origins of scaling in natural images. *Vision Research*, 23, 3385-3398.
- Rust, N.C. & Movshon, J.A. (2005). In praise of artifice. *Nature Neuroscience*, 8, 1647-1650.
- Saarla, T.P. & Horzog, M.H. (2008). Time-course and surround modulation of contrast masking in human vision. *Journal of Vision*, 8, 1-10.
- Schwartz, O. and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience* 4, 819-825.
- Shapley, R.M. & Victor, J.D. (1979). The contrast gain control of the cat retina. *Vision Research*, 19, 431-434.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, pp. 379-423, 623-656.
- Sharpe, C.R. & Tolhurst, D.J. (1973). Orientation and spatial frequency channels in peripheral vision. *Vision Research*, 13, 2013-2112.
- Shoham, D., Hubener, M., Schulze, S., Grinvald, A., & Bonhoeffer, T. (1997). Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, 385, 529-533.
- Snowden, R.J. & Hess, R.F. (1992). Temporal frequency filters in the human peripheral visual field. *Vision Research*, 32, 61-72.
- Snowden, R.J. (1992). Orientation bandwidth—the effect of spatial and temporal frequency. *Vision Research*, 32, 1965-1974.
- Snowden, R.J. (1994). Adaptability of the visual system is inversely related to its

- sensitivity. *Journal of the Optical Society of America-A*, 11, 25-32.
- Snowden, R.J. & Hammett, S.T. (1996). The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research*, 38, 1935-1945.
- Snowden, R.J. & Hess, R.F. (1992). Temporal frequency filters in the human peripheral visual field. *Vision Research*, 32, 61-72.
- Snowden, R.J. (2001). Contrast gain mechanism or transient channels? Why the effects of a background pattern alter over time. *Vision Research*, 41, 1879-1883.
- Spitzberg, R. & Richards, W. (1975). Broad band spatial filters in the human visual system. *Vision Research*, 15, 837-841.
- Sun, P., Ueno, K., Waggoner, R.A., Gardner, J.L., Tanaka, K., & Cheng, K. (2007). A temporal frequency-dependent functional architecture in human V1 revealed by high-resolution fMRI. *Nature Neuroscience*, 10, 1404-1406.
- Sillito, A.M., Cudeiro, J., & Jones, H.E. (2006). Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in Neurosciences*, 29, 307-316.
- Simoncelli, E.P. & Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193-1216.
- Stromeyer, C.F., Klein, S., Dawson, B.M., & Spillmann, L. (1982). Low spatial-frequency channels in human vision: adaptation and masking. *Vision Research*, 22, 225-233.
- Switkes, E., Mayer, M.J., & Sloan, J.A. (1978). Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Research*, 18, 1393-1399.

- Tolhurst, D.J. (1973a). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *The Journal of Physiology*, 231, 385-402.
- Tolhurst, D.J. (1973b). The analysis of the drift rate of moving sinusoidal gratings. *Vision Research*, 13, 2545-2555.
- Tolhurst, D.J., Sharpe, C.R. & Hart, G. (1973). The analysis of the drift of moving sinusoidal gratings. *Vision Research*, 13, 2545-2555.
- Tolhurst, D.J. (1975). Sustained and transient channels in human vision. *Vision Research*, 15, 1151-1155.
- Tolhurst, D.J. & Movshon, J.A. (1975). Spatial and temporal contrast sensitivity of striate cortical neurons. *Nature*, 257, 674-675.
- Tolhurst, D.J. & Tadmor, Y. (1997). Band-limited contrast in natural images explains the detectability of change in the amplitude spectra. *Vision Research*, 37, 3203-3215.
- Tolhurst, D.J., Movshon, J.A., & Dean, A.F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23, 775-785.
- Tolhurst, D.J., Tadmor, y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & Physiological Optics*, 12, 229-232.
- Totev, C. & Mitov, D. (2000). Temporal integration and reaction time to grating-onset detection. *Perception*, 29 (Suppl), 61.
- Thomas, J.P. & Gille, J. (1979). Bandwidth of orientation detector in human vision. *Journal of the Optical Society of America*, 69, 652-660.
- van der Schaaf, A. & van Hateren, J.H. (1996). Modeling the power spectra of natural images: statistics and information. *Vision Research*, 36, 2759-2770.
- Vassilev, A. & Mitov, D. (1976). Perception time and spatial frequency. *Vision Research*,

16, 89-92.

- Van Essen, D.C. & Gallant, J.L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13, 1-10.
- Vergheze, P. & Stone, L.S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, 381, 161-163.
- Vogels, R. & Orban, G.A. (1986). Decision process in visual discrimination of line orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 115-132.
- Watson, A.B. & Nachmias, J. (1977). Patterns of temporal interaction in the detection of gratings. *Vision Research*, 17, 893-902.
- Watson, A. and Solomon, J. (1997). A model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, pages 2379-2391.
- Waugh, S.J. & Hess, R.F. (1994). Suprathreshold temporal-frequency discrimination in the fovea and the periphery. *Journal of Optical Society of America-A*, 11, 1199-1212.
- Wainwright, M.J., Schwartz, O. & Simoncelli, E.P. (2001). Natural image statistics and divisive normalization: Modeling nonlinearities and adaptation in cortical neurons. In R Rao, B, Olshausen, & M. Lewicki (Eds), *Probabilistic models of the brain: Perception and neural function*. Cambridge, MA: MIT Press.
- Weymouth, F.W. (1959). Stimulus orientation and threshold: an optical analysis. *American Journal of Ophthalmology*, 48, 6-10.
- Wilson, H.R. (1980). Spatiotemporal characterization of a transient mechanisms in the human visual system. *Vision Research*, 20, 443-452.

- Wilson, H.R., McFarlane, D.K., & Phillips, G.C. (1983). Spatial-frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 873-882.
- Wilson, H.R. & Humanski, R. (1993). Spatial-frequency adaptation and contrast gain-control. *Vision Research*, 33, 1133-1149.
- Wilson, H.R. Wilkinson, F. & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, 37, 2325-2330.
- Wilson, H.R. & Kim, J. (1998). Dynamics of divisive gain control in human vision. *Vision Research*, 38, 2735-2741.
- Wilson, H.R. & Wilkinson, F. (1998). Spatial channels in vision and spatial pooling. In: L.M. Chalupa & J.S. Werner (Eds.), *The Visual Neurosciences*, MIT Press: Cambridge, MA.
- Wilson, H.R., Loffler, G., Wilkinson, F. & Thistlethwaite, W.A. (2001). An inverse oblique effect in human vision. *Vision Research*, 41, 1749-1753.
- Xing, J. & Heeger, D.J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Research*, 41, 571-583.
- Yu, C. & Levi, D.M. (1999). The time course of psychophysical end-stopping. *Vision Research*, 39, 2063-2073.
- Yu, C., Klein, S.A., & Levi, D.M. (2003). Cross- and Iso-oriented surrounds modulate the contrast response function: The effect of surround contrast, *Journal of Vision*, 3, 527-540.
- Yu, H-H., Verma, R., Yang, Y., Tibballs, H.A., Lui, L.L., Reser, D.H., & Rosa, M.G.P. (2010). Spatial and temporal frequency tuning in striate cortex: functional

uniformity and specializations related to receptive field eccentricity. *European Journal of Neuroscience*, 31, 1043-1062.

## CURRICULUM VITAE

### Yeon Jin Kim

University of Louisville  
Department of Psychological & Brain Sciences  
317 Life Sciences Building  
Louisville, KY 40292  
Lab: (502) 852-7710  
E-mail: [y0kim009@louisville.edu](mailto:y0kim009@louisville.edu)

### Education

- |   |              |             |
|---|--------------|-------------|
| University of Louisville, Louisville, KY, USA<br>Dept. of Psychological & Brain Sciences<br>Major Area: Experimental Psychology<br>Specialization: Vision Science<br>Research Advisor: Dr. Edward A. Essock | <b>Ph.D.</b> | <b>2011</b> |
| Kwangwoon University, Seoul, South Korea<br>Dept. of Information & Control Engineering  | <b>B.E.</b>  | <b>2004</b> |
| Kwangwoon University, Seoul, South Korea<br>Dept. of Industrial Psychology<br>Major Area: Experimental Psychology<br>Specialization: Vision Science<br>Research Advisor: Dr. Hyung-Chul O. Li.              | <b>B.S.</b>  | <b>2004</b> |

### Published Articles

- 1) Kim, Y.J., Haun, A.M., & Essock, E.A. (2011). Time-Course of Anisotropic Masking at High and Low Spatial Frequencies (In preparation).
- 2) Kim, Y.J. & Essock, E.A. (2010). The Nature of Anisotropy in Gain Control Pools (In press).
- 3) Kim, Y.J., Haun, A.M., & Essock, E.A. (2010). The horizontal effect in Suppression: Anisotropic overlay and surround suppression at high and low Speeds. *Vision Research*, 50, 838-849.
- 4) Essock, E.A., Haun, A.M., & Kim Y.J. (2009). Anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *Journal of Vision*, 9, 1-15.



## Published Abstracts

- 1) Kim, Y.J., Haun, A.M., & Essock, E.A. (2011). Time-Course of Anisotropic Masking at High and Low Spatial Frequencies [Abstract]. *Vision Sciences Society, 2011*.
- 2) Schweinhart, A.M., Kim, Y.J., & Essock, E.A. (2011). Anisotropy in Paintings: A reflection of structural anisotropy in natural scenes? [Abstract]. *Vision Sciences Society, 2011*.
- 3) Kim, Y.J., & Essock, E.A. (2010). Anisotropic Gain Control Pools Are Tuned In Temporal Frequency As Well As Spatial Frequency And Orientation [Abstract]. *Journal of Vision 10*, 1253.
- 4) Kim, Y.J., Haun, A.M., & Essock, E.A. (2009). Overlay and Surround Suppression Both Show a Horizontal Effect Anisotropy [Abstract]. *Journal of Vision, 9*, 1009a.
- 5) Kim, Y.J., Haun, A.M., & Essock, E.A. (2008). The effect of sustained/transient temporal modulation on the horizontal effect of contrast masking [Abstract]. *Journal of Vision 8*, 276a.
- 6) Kim, Y.J., Haun, A.M., & Essock, E.A. (2007). Anisotropic Contrast Sensitivity During Viewing of Broadband Stimuli: Timing and Tuning [Abstract]. *Journal of Vision 7*, 592a.
- 7) Kim, Y.J., Haun, A.M., & Essock, E.A. (2007). Anisotropic Contrast Sensitivity During Viewing of Broadband: Timing and Tuning [Abstract]. The Louisville Chapter, Society for Neuroscience.
- 8) Haun, A.M., Hansen, B.C., Kim, Y.J., & Essock, E.A. (2005). Sequential effects and stimulus-response dependencies in an orientation identification task: Characterization of the class 2 oblique effect [Abstract]. *Journal of Vision 5*, 158a.
- 9) Kim, Y.J., Yoon, J.H., & Li, H. –C. O. (2004). Interaction of binocular disparities and Pulfrich Effect in the perception of depth and rotation direction of transparent cylinder [Abstract]. *Journal of Vision 4*, 469a.

## Awards

2004-2011	Graduate Student Travel Fund, University of Louisville
2002-2003	Student Academic Scholarship, Kwangwoon University

## Research Experience

2004- 2011	Research Assistant Department of Psychological & Brain Sciences, University of Louisville, KY, USA Research advisor: Dr. Edward A. Essock
August, 2003	Research Assistant

University Medical Center Groningen,  
University of Groningen, Groningen, The Netherlands

- Primary investigation involved the eye movement measurement (used by Eye tracker) to pursue a small moving object with their eyes.

Research advisor: Dr. Hyung-Chul O. Li, Dr. Eli Brenner &  
Dr. Frans W. Cornelissen

2002-2003

Research Assistant for Human Factor Project  
Department of Industrial Psychology,  
Kwangwoon University, Seoul, South Korea

- Primary Investigation involved 3-dimensional space perception and Pulfrich Effect.

Research advisor: Dr. Hyung-Chul O. Li

### **Teaching Experience**

2009-2011

Course-Instructor  
(Psychology 302: Experimental Psychology)  
Dept of Psychological & Brain Sciences,  
University of Louisville

2009-2010

Graduate Teaching Assistantships  
(Psychology 201: Introduction to Psychology)  
Dept of Psychological & Brain Sciences,  
University of Louisville

2008-2009

Graduate Teaching Assistantships  
(Psychology 332: Cognitive Process)  
Dept of Psychological & Brain Sciences,  
University of Louisville

2004-2007

Graduate Teaching Assistantships  
(Psychology 331: Sensation & Perception)  
Dept of Psychological and Brain Sciences,  
University of Louisville

### **Technical Skills**

Programming Languages: C, MATLAB, OpenGL  
Proficient in MATLAB extension including the most current Psychophysics, Image Processing, Signal Processing, Optimization and Statistics toolboxes.

Operating Systems: Windows Window7/XP/2000/ME/98/95  
Macintosh e-Mac/Power Mac OS X

Software Platforms: All components of MS Office XP, SPSS, SAS, Adobe  
PhotoShop, Adobe Illustrator, Vision Research Graphics,  
nVidia

Imaging/Sensor Systems: Video Eye Tracker, Pelli-Zhang video attenuator.